2002s-85

Testing Mean-Variance Efficiency in CAPM with Possibly Non-Gaussian Errors: an Exact Simulation-Based Approach

Marie-Claude Beaulieu, Jean-Marie Dufour,
Lynda Khalaf

Série Scientifique
Scientific Series

Montréal
Novembre 2002

CIRANO
Centre interuniversitaire de recherche
en analyse des organisations

Testing Mean-Variance Efficiency in CAPM with Possibly Non-Gaussian Errors: an Exact Simulation-Based Approach[*]

Marie-Claude Beaulieu[†], Jean-Marie Dufour[‡], Lynda Khalaf[§]

Résumé / *Abstract*

Dans cet article, nous proposons des tests exacts, basés sur la vraisemblance de l'efficience du portefeuille de marché dans l'espace moyenne-variance. Ces tests, utilisés ici dans le contexte du modèle du CAPM (Capital Asset Pricing Model), permettent de considérer diverses classes de distributions incluant la loi normale. Les tests sont développés dans le cadre de modèles de régression linéaires multivariés (RLM). Il est, par ailleurs, bien établi que, malgré leur structure simple, les écart-types et tests usuels asymptotiques de ces modèles ne sont pas fiables. En économétrie financière, des tests en échantillons finis ont été proposés pour quelques hypothèses spécifiques, lesquels dépendent pour la plupart de l'hypothèse de normalité [Jobson et Korkie (Journal of Financial Economics, 1982), MacKinlay (Journal of Financial Economics, 1987), Gibbons, Ross et Shanken (Econometrica, 1989), Zhou (Journal of Finance 1993)]. Dans le contexte gaussien, nos tests d'efficience correspondent à ceux de Gibbons, Ross et Shanken. Dans un contexte non-gaussien, nous reconsidérons l'efficience moyenne-variance du portefeuille de marché en permettant des distributions multivariées de Student et des « mélanges de lois normales ». Notre démarche nous permet d'évaluer si l'hypothèse de normalité est trop restrictive lorsque l'on teste le CAPM. Nous proposons aussi des tests diagnostiques multivariés (incluant des tests pour les effets GARCH multivariés et une généralisation multivariée des tests de ratio de variance), des tests de spécification ainsi qu'un estimateur ensembliste pour les paramètres de nuisance pertinents. Nos résultats montrent que i) l'hypothèse de normalité multivariée est rejetée sur la plupart des sous-périodes, ii) les tests diagnostiques appliqués aux résidus de nos estimations ne montrent pas de différences importantes par rapport à l'hypothèse des erreurs i.i.d. multivariées, et iii) les tests d'efficience du portefeuille de marché dans l'espace moyenne-variance ne rejettent aussi fréquemment l'hypothèse d'efficience lorsqu'on s'autorise à considérer des lois non normales sur les erreurs.

[†] Centre Interuniversitaire de Recherche sur les Politiques Économiques et l'Emploi (CIRPÉE) et Département de finance et assurance, Pavillon Palasis-Prince, local 3632, Université Laval, Québec, Canada G1K 7P4. Tel: (418) 656-2131-2926, fax: (418) 656-2624, e-mail: Marie-Claude.Beaulieu@fas.ulaval.ca.

[‡] Canada Research Chair Holder (Econometrics). Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. Tel: (514) 343-2400; fax: (514) 343-5831, e-mail: jean.marie.dufour@umontreal.ca, Web page: http://www.fas.umontreal.ca/SCECO/Dufour.

[§] Centre interuniversitaire de recherche en économie quantitative (CIREQ), and GREEN, Université Laval, Pavillon J.-A. De Sève, St. Foy, Québec, Canada, G1K 7P4. Tel: (418) 656 2131-2409, fax: (418) 656 7412, e-mail: lynda.khalaf@ecn.ulaval.ca.

In this paper we propose exact likelihood-based mean-variance efficiency tests of the market portfolio in the context of Capital Asset Pricing Model (CAPM), allowing for a wide class of error distributions which include normality as a special case. These tests are developed in the framework of multivariate linear regressions (MLR). It is well known however that despite their simple statistical structure, standard asymptotically justified MLR-based tests are unreliable. In financial econometrics, exact tests have been proposed for a few specific hypotheses [Jobson and Korkie (Journal of Financial Economics, 1982), MacKinlay (Journal of Financial Economics, 1987), Gibbons, Ross and Shanken (Econometrica, 1989), Zhou (Journal of Finance 1993)], most of which depend on normality. For the gaussian model, our tests correspond to Gibbons, Ross and Shanken's mean-variance efficiency tests. In non-gaussian contexts, we reconsider mean-variance efficiency tests allowing for multivariate Student-t and gaussian mixture errors. Our framework allows to cast more evidence on whether the normality assumption is too restrictive when testing the CAPM. We also propose exact multivariate diagnostic checks (including tests for multivariate GARCH and multivariate generalization of the well known variance ratio tests) and goodness of fit tests as well as a set estimate for the intervening nuisance parameters. Our results [over five-year subperiods] show the following: (i) multivariate normality is rejected in most subperiods, (ii) residual checks reveal no significant departures from the multivariate i.i.d. assumption, and (iii) mean-variance efficiency tests of the market portfolio is not rejected as frequently once it is allowed for the possibility of non-normal errors.

Contents

List of Propositions and Theorems

List of Tables

1. Introduction

The capital asset pricing model (CAPM) is one of the most commonly used models in asset-pricing theory and practice. The associated empirical literature is enormous: reviews and references may be found in Campbell, Lo and MacKinlay (1997, Chapters 5 and 6) and Shanken (1996). Since the work of Gibbons (1982), empirical tests of the CAPM are usually conducted within the *multivariate linear regression* (MLR) framework. Statistical inference for the MLR model in econometrics and empirical finance is usually based either on asymptotic approximations or on finite-sample distributional theory.

Concerning the first approach, several studies in the econometrics and finance literatures have shown that standard asymptotic theory provides a poor approximation to the finite-sample distribution of MLR-based tests, even with fairly large samples; see Shanken (1996, Section 3.4.2), Campbell et al. (1997, Chapter 5), Dufour and Khalaf (2002*d*), and the references therein. In particular, test size distortions grow quickly as the number of equations increases. As a result, the conclusions of financial MLR-based empirical studies on the CAPM can be strongly affected and may lead to spurious rejections.

In this context, as emphasized by Shanken (1996) and Campbell et al. (1997), applying finite-sample statistical methods appears to be important. As a result, a number of authors have proposed tests based on a finite-sample distributional theory in order to assess the CAPM; see Jobson and Korkie (1982), MacKinlay (1987), Gibbons, Ross and Shanken (1989, henceforth GRS), Zhou (1991) and Stewart (1997). These proposals are based on exploiting results from the statistical literature on multivariate analysis-of-variance, in particular Hotelling's T^2 statistic which may be transformed into a statistic with a Fisher distribution [see, for example, Anderson (1984, chapters 8 and 13) and Rao (1973, chapter 8)]. These results, however, are based on the assumption that the errors in the CAPM follow a gaussian distribution. This may provide valid tests of a gaussian CAPM satisfying mean-variance efficiency but not of alternative similarly restricted non-gaussian CAPM. Rejections of mean-variance efficiency by such tests may be interpreted as a rejection of normality, instead of the CAPM relationships. Further, the CAPM can be derived from an expected utility maximization under a variety of distributional assumptions on the cross-sectional distribution of the returns, which include the gaussian distribution, but also several non-gaussian distributions such as the multivariate Student-t, normal mixtures, elliptically symmetric distributions, etc.[1]

It has long been recognized that financial returns do not exhibit normality [see, for example, Fama (1965), Baillie and Bollerslev (1998) and Beaulieu (1998)]. Consequently, it appears that the appropriateness of the multinormal assumption should be tested in statistical analysis of the CAPM and, eventually, valid tests for non-gaussian CAPM should be put forward. Of course, the normality assumption can usually be relaxed by considering large-sample procedures based on the central limit theorem (such as GMM procedures) or bootstrap techniques (to get better finite-sample behavior); for some examples in the context of the CAPM, see Affleck-Graves and McDonald (1989), Fama

---

[1]For discussions of the class of return distributions compatible with the CAPM, the reader may consult Ross (1978), Chamberlain (1983), Ingersoll (1987, Chapter 4), Nielsen (1990), Allingham (1991) and Berk (1997). Another possibility would consist in considering stable Paretian laws; see Samuelson (1967). However, since stable distributions other the normal distribution do not have finite second moments, this requires replacing the variance of a portfolio by another measure of risk.

and French (1993) and Groenwold and Fraser (2001). But asymptotically-based procedures can only be a solution of last resort, which is always fundamentally unsatisfactory: there is no guarantee that the *level* or the *size* of the test is controlled and what can be legitimately rejected by such a procedure typically remains unclear [for further discussion of these issues, see Dufour (1997) and Dufour and Khalaf (2001)].

Clearly, it would be much more satisfactory to have finite-sample tests that allow for non-gaussian errors. But there has been very little work on this issue. The main contribution on this important problem appears to be due to Zhou (1993). In a seminal paper, this author reconsidered the GRS problem under elliptical distributions and, on observing some invariance properties of the Hotelling statistic for testing mean-variance efficiency, suggested the use of a simulation procedure to approximate the appropriate p-values. To the best of our knowledge, no other non-asymptotic result that does not impose normality appears to be available for the CAPM model with an observable risk-free return.[2] Zhou (1993) also provided formal tests of multinormality and other elliptically symmetric distributions: these are based on the statistics proposed by Mardia (1970) for the multivariate location-scale model, *i.e.* a MLR where the regressor matrix reduces to a vector of ones _ without however providing finite-sample adjustments to deal with the fact that the tests were applied to regression residuals.[3]

In this paper, we pursue the work of Zhou (1993) in several directions. The objective is to propose exact finite-sample tests for possibly non-gaussian versions of the CAPM. The procedures suggested involve testing both the usual cross-equation parameter restrictions entailed by the CAPM as well as various specification tests, including tests of normality and distributional goodness-of-fit, tests against the presence of ARCH effects and serial dependence, which all take into account the multivariate nature of the model. To be more specific, our contributions can be classified in five categories.

First, we reconsider the problem of building exact tests of mean-variance efficiency in non-gaussian set-ups, using Monte Carlo (MC) test techniques. The procedures we propose for that purpose are based on the finite-sample methods described in Dufour and Khalaf (2002*d*) for hypothesis testing in MLR models. The latter allow one to test linear restrictions on MLR models using a wide array of multivariate test statistics _ such as the gaussian LR criterion [or the Wilks (1932) statistic], the Lawley-Hotelling (LH) trace criterion, the Bartlett-Nanda-Pillai (BNP) trace criterion, or the maximum root (MR) criterion _ under general parametric distributional assumptions, which include, besides the gaussian distribution, a wide spectrum of non-gaussian distributions, both elliptically symmetric and non-elliptical. In particular, in the case of uniform linear (UL) restrictions,[4] all standard test statistics have null distributions which are free of nuisance parameters as long the error distribution is specified up to an unknown linear transformation, so that Monte Carlo (MC) test techniques [originally proposed by Dwass (1957) and Barnard (1963)] can be applied to obtain

---

[2]There has also been work on finite-sample inference for the CAPM model without an observable risk-free asset (Black's CAPM); see Shanken (1986), Zhou (1991, 1995) and Velu and Zhou (1999). This problem goes beyond the scope of the present paper, but we reconsider it in Beaulieu, Dufour and Khalaf (2001).

[3]This limitation is pointed out by Zhou (1993, page 1935, footnote 5).

[4]Examples of UL hypotheses include: (i) identical transformations of the regression coefficients (within or across equations) are equal to given values, (ii) the coefficients of the same regressor are zero across equations, and (iii) a single parameter equals zero; see also Berndt and Savin (1977) and Stewart (1997).

provably exact tests even with a very small number of MC replications [such as 19 replications for a test of level 5%; see Dufour and Khalaf (2001)]. The fact that the relevant analytical distributions are complicated is not a problem: the only requirement is the possibility of simulating the test statistic under the null hypothesis. On observing that standard CAPM mean-variance efficiency restrictions take the UL form when the risk-free rate is observable, we show that the latter can be tested in this way under quite general distributional assumptions, which include all elliptical distributions. Both single and multi-beta portfolios are covered by these results.

Second, in conformity with standard and recent mean-variance efficiency theoretical set-ups, we focus for practical applications on two families of distributions: (1) multivariate Student-t distributions, and (2) mixtures of two normal distributions. Both these families raise a nuisance parameter problem: the number of degrees of freedom, for the multivariate Student-t; the probability-of-mixing and ratio-of-scale parameters for the normal mixtures. To deal with this problem [not considered by Zhou (1993)], we propose a two-stage confidence procedure similar to the ones used in Dufour (1990) and Dufour and Kiviet (1996). First, one builds an exact confidence set for the nuisance parameter, through the "inversion" of a distributional goodness-of-fit (GF) test (described below). Second, the MC p-value for the hypothesis of interest (which depends on the nuisance parameter) is maximized over this confidence set. Provided the significance levels used at the different steps of the procedure satisfy a simple inequality, this two-stage procedure yields an exact test that controls for the presence of the nuisance parameters considered.[5] In this way, we formally deal with the (often ignored) problem of the joint characteristic of the null hypothesis which imposes distributional constraints, in addition to the restrictions on the regression coefficients.

Third, we propose and apply exact multivariate goodness-of-fit (GF) tests. As emphasized by Richardson and Smith (1993) who considered tests for departures from normality, it is crucial in MLR-based financial models to consider multivariate tests of asset returns which explicitly take the error covariance into consideration. In order to obtain provably valid tests within the MLR framework, we adjust the tests proposed by Mardia (1970) and Zhou (1993), and we propose an exact GF test of the hypothesized error distributions (multivariate normal, Student-t error and normal mixtures, with possibly unknown parameters).[6] The test is based on comparing multivariate skewness and kurtosis criteria to a simulation-based estimate of their expected value under the hypothesized distribution and is implemented as a MC test. Although the GF test is used, as explained above, to obtain a confidence set for the intervening distributional parameters, we note that the test is new and to our knowledge, no other exact test for these distributions is available. Beside its relevance for the present paper, this test would be quite useful as a specification check in empirical finance, given the popularity of the Student-t distribution.

Fourth, we conduct exact multivariate residual-based diagnostic tests for departures from the maintained hypothesis of *i.i.d.* disturbances. As with normality tests, concerns over cross-

---

[5]This procedure may be interpreted as a special form of "maximized MC" (MMC) test [see Dufour (2002)]. For further discussions of MMC tests in econometrics, see Dufour and Khalaf (2001, 2002c). In nuisance parameter dependent problems, a test is *exact* at level α if the largest rejection probability over the nuisance parameter space consistent with the null hypothesis is not greater at α [see Lehmann (1986, sections 3.1, 3.5)].

[6]Regarding normality tests, available empirical evidence for monthly data (on which we focus) is mixed. For instance, whereas the results of Campbell et al. (1997) and Affleck-Graves and McDonald (1989) suggest that normality is not rejected often at monthly frequencies, the tests conducted by Richardson and Smith (1993) provide more firm rejections.

correlations of portfolio returns has been the subject of several studies [*e.g.*, Richardson and Smith (1993) and Shanken (1990)]. For these problems, standard multivariate approaches [including Richardson and Smith (1993) and Shanken (1990)] are asymptotic. In spite of the well known problems associated with such an approach, reliance on asymptotics is not surprising in the absence of applicable exact results.[7] In view of the fact that CAPM tests may be sensitive to the presence of GARCH-type heteroskedasticity [see MacKinlay and Richardson (1991)], we first consider tests against multivariate GARCH effects, in the spirit of those proposed in Shanken (1990). Our procedures differ from Shanken's in two basic aspects. First, our tests are based on properly *standardized* OLS residuals to ensure invariance to the error-covariance matrix. Second, we combine tests across equations using an exact simulation-based procedure [similar to the one used in Dufour and Khalaf (2002*b*)] which does not call for using Bonferroni level adjustment. Bonferroni-based combined tests require one to divide the level of each individual test by the number of tests [see Dufour (1990) or Shanken (1990)]. Although this can provide guarantees against certain types of specification errors [see Dufour and Torrès (1998)], it can also yield utterly conservative tests if the MLR includes many equations (*i.e.*, many portfolios), leading to possibly large power losses. Following the exact strategy we applied to test the significance of the CAPM intercepts, we also consider multivariate Student-t errors and multivariate normal mixtures with possibly unknown parameters; this allows one to test whether GARCH effects are still prevalent, even if fat tails are formally modelled into the error distribution. The second class of tests we study are multivariate generalizations of the popular variance ratio tests. Using the same residual standardization and the simulation-based combination strategy proposed for the GARCH test, we show how these very useful tests can be applied exactly, in a multivariate set-up. We emphasize that the usefulness of these new tests extends beyond the specific applications studied here.

Fifth, the tests proposed are applied to the CAPM with observable risk-free rates. We consider monthly returns on New York Stock Exchange (NYSE) portfolios, which we construct from the University of Chicago Center for Research in Security Prices (CRSP) 1926-1995 data base. Our results allow to compare test results for asymptotic statistics and exact tests under normality. Furthermore we can also compare exact p-values for different elliptical distributions. We explain why non-rejections obtained under a gaussian distributional assumption may formally be treated as conclusive from our viewpoint (recall, of course, that we formally combine GF with efficiency tests here). Our results (over five-year subperiods) show the following: (i) multivariate normality is rejected in most subperiods, (ii) multivariate residual checks reveal no significant departures from the *i.i.d.* assumption, and (iii) mean-variance efficiency of the market portfolio is not rejected as frequently once it is allowed for the possibility of non-normal errors.

The paper is organized as follows. Section 2 describes the statistical framework studied. In Section 3, we describe the existing test procedures and we show how extensions allowing for non-normal distributions can be obtained. In Section 4, we present extensions to error distributions involving nuisance parameters. Exact GF and diagnostic tests are proposed in Section 5. In Section 6 we report the empirical results. Section 7 concludes and discusses extensions to other asset pricing

---

[7]Indeed, this problem is not exclusive to financial applications: our review of the statistics and econometrics literature has revealed that exact multivariate specification tests which take the error covariance explicitly into consideration are quite rare; see Dufour, Khalaf and Beaulieu (2001).

tests.

2. Framework

The fundamental finance problem we focus on here involves testing the mean-variance efficiency of a candidate benchmark portfolio. Let R_{it}, $i = 1, \ldots, n$, be returns on n securities for period t, and $\widetilde{R}_{Mt}$ the returns on the market portfolio under consideration ($t = 1, \ldots, T$). If these variables satisfy a temporally stable CAPM, the following relations must hold:

$$r_{it} = b_i \widetilde{r}_{Mt} + u_{it}, \quad t = 1, \ldots, T, \ i = 1, \ldots, n, \tag{2.1}$$

where $r_{it} = R_{it} - R_t^F$, $\widetilde{r}_{Mt} = \widetilde{R}_{Mt} - R_t^F$, R_t^F is the riskless rate of return, and u_{it} is a random disturbance. Following Gibbons et al. (1989), this can be cast as a MLR model of the form

$$r_{it} = a_i + b_i \widetilde{r}_{Mt} + u_{it}, \quad t = 1, \ldots, T, \ i = 1, \ldots, n, \tag{2.2}$$

on which the following restrictions have been imposed:

$$H_C : a_i = 0, \quad i = 1, \ldots, n, \tag{2.3}$$

i.e., the intercepts a_i are jointly equal to zero.

The above model is a special case of the MLR model:

$$Y = XB + U \tag{2.4}$$

where $Y = [Y_1, \ldots, Y_n]$ is $T \times n$ matrix of dependent variables, X is a $T \times k$ full-column rank matrix of regressors, and

$$U = [U_1, \ldots, U_n] = [V_1, \ldots, V_T]' \tag{2.5}$$

is a $T \times n$ matrix of random disturbances which are independent of X (so that we can easily condition on X and take X as fixed for statistical analysis). In this framework, the CAPM-based restriction H_C belong to the class of so-called *uniform linear* (UL) restrictions, *i.e.* it has the form

$$H_0 : HBE = D \tag{2.6}$$

where H is an $h \times k$ matrix of rank h and E is an $n \times e$ matrix of rank e. This is relevant because H_C is a special case of H_0. Indeed, rewriting (2.2) as in (2.4) with

$$
\begin{aligned}
Y &= [r_1, \ldots, r_n], \quad X = [\iota_T, \widetilde{r}_M], \\
r_i &= (r_{1i}, \ldots, r_{Ti})', \quad \widetilde{r}_M = (\widetilde{r}_{1M}, \ldots, \widetilde{r}_{TM})', \quad \iota_T = (1, \ldots, 1)',
\end{aligned}
$$

we see that H_C can be written as $(1, 0)B = 0$, which corresponds to H_0 with $H = (1, 0)$, $E = I_n$ and $D = 0$. For further discussion of the MLR model, the reader may consult Anderson (1984, chapters 8 and 13), Berndt and Savin (1977), Dufour and Khalaf (2002*d*), Kariya (1985), Kariya and Kim (1997), Rao (1973, chapter 8) and Stewart (1997). In particular, it is worthwhile to note

that standard exact methods for testing hypotheses of the form H_0 rely on the assumption that the disturbance vectors are *i.i.d.* gaussian, namely:[8]

$$V_1, \ldots, V_T \text{ are } i.i.d. \ N[0, \ \Sigma] \,. \tag{2.7}$$

The CAPM can de derived from an expected utility maximization under a variety of assumptions on the cross-sectional distribution of returns, which include the gaussian distribution, the multivariate Student t, mixtures of normals, etc. [see Ingersoll (1987) and Berk (1997)]. In view of the available statistical methods, the standard assumption under which finite-sample tests have been proposed is (2.7). However, since security returns tend to follow more heavy-tailed distribution than the normal, this may be unduly restrictive.

In this paper, we use a more comprehensive statistical theory [developed in Dufour and Khalaf (2002d)] which allows one to easily obtain finite-sample tests under alternative error distributions. More precisely, we consider the general case:

$$V_t = JW_t \,, \ t = 1, \ldots, T \,, \tag{2.8}$$

where J is an unknown, non-singular matrix and the distribution of the vector $w = vec\,(W_1, \ldots, W_T)$ is either: (i) known (hence, free of nuisance parameters), or (ii) specified up to an unknown nuisance-parameter. We call w the vector of *normalized disturbances* and its distribution the *normalized disturbance distribution*. This set-up includes as a special case the *i.i.d.* gaussian or the (more general) elliptical symmetry assumption as well as many other distributional set-ups (*e.g.*, cases where $W_1, \ldots, W_T$ are *i.i.d.* according to an arbitrary heavy-tailed distribution, which may not be elliptically symmetric).

In conformity with standard and recent mean-variance efficiency theoretical set-ups, we will now focus on two families of distributions consistent with both CAPM theory and the statistical framework just described: (1) multivariate Student-t distributions, and (2) mixtures of two normal distributions. Let us denote these two distributions $\mathcal{F}_1(W)$ and $\mathcal{F}_2(W)$ respectively. We say that:

$$W_t \sim \mathcal{F}_1(\kappa) \Leftrightarrow W_t = Z_{1t}/(Z_{2t}/\kappa)^{1/2} \,, \tag{2.9}$$

where $Z_{1t} \sim N[0, \ I_n]$ _ *i.e.* Z_{1t} is multivariate normal with mean 0 and covariance matrix I_n _ and $Z_{2t} \sim \chi^2(\kappa)$ and is independent of Z_{1t}; and

$$W_t \sim \mathcal{F}_2(\pi, \ \omega) \Leftrightarrow W_t = \pi Z_{1t} + (1 - \pi)Z_{3t} \,, \tag{2.10}$$

where $Z_{3t} \sim N[0, \ \omega I_n]$ and is independent of Z_{1t}, and $0 < \pi < 1$. As mentioned in the introduction, we focus on these families of distributions for the following reasons: (i) from an empirical

---

[8]It is also possible to show that the gaussian-based distributional theory still leads to valid tests under the more general assumption of that the distribution of U is *jointly* elliptically symmetric; see Kariya (1985) and Kariya and Kim (1997). However, except in gaussian case, this assumption has the unattractive feature of precluding independence between the vectors $V_1, \ldots, V_T$, even if each vector V_t follows an elliptically symmetric distribution. For further discussion of elliptically symmetric distributions, see Kelker (1970), Chmielevsky (1981), Owen and Rabinovitch (1983) and Dufour and Roy (1985).

perspective, financial return data typically displays spikes and fat tails [Fama (1965)] and, (ii) on theoretical grounds, the multivariate Student t and this specific mixture-of-normals are return distributions consistent with expected utility maximization. For further reference, we shall use the following notation:

$$W_t \sim \mathcal{F}_i(\nu), \quad i = 1, 2, \tag{2.11}$$

where

$$\begin{aligned} \nu &= \kappa, && \text{if } W_t \text{ satisfies (2.9),} \\ &= (\pi, \omega), && \text{if } W_t \text{ satisfies (2.10).} \end{aligned}$$

3. Mean-variance efficiency tests with a known normalized disturbance distribution

In this section, we study the case where the nuisance parameter ν is specified by the null hypothesis. Extensions to unknown ν are presented in Section 4. Note that no further regularity conditions are required for most of our proposed statistical procedures, not even the existence of second moments. Yet the latter hypothesis is typically maintained in CAPM contexts. In this case, the covariance matrix of V_t is $\Sigma = JJ'$ and $\det(\Sigma) \neq 0$.

One of the most commonly used statistics to test H_C in (2.3) [indeed, to test any UL hypothesis] is the gaussian quasi maximum likelihood (QMLE) based criterion:

$$LR = T \ln(\Lambda), \;\; \Lambda = |\hat{\Sigma}_C| / |\hat{\Sigma}|, \tag{3.12}$$

where $\hat{\Sigma} = \hat{U}'\hat{U}/T$, $\hat{U} = Y - X\hat{B}$, $\hat{B} = (X'X)^{-1}X'Y$ and $\hat{\Sigma}_C$ is the gaussian QMLE under H_C. In Theorem **A.1** of Appendix A.1, we provide the exact null distribution of the latter statistic under (2.4), (2.8) and the general UL hypothesis (2.6). Two results regarding this distribution are worth noting.

First, under (2.8), the distribution does not depend on B and Σ and thus may easily be simulated if draws from the distribution of $W_1, \ldots, W_T$ are available. This entails that a Monte Carlo exact test procedure may be easily applied based on LR. The general simulation-based algorithm which allows to obtain a MC size-correct exact p-value for all hypotheses conforming with (2.8) may be summarized as follows. Using the distributional assumption (2.11), generate N *i.i.d.* replications of the disturbance matrix $W = [W_1, \ldots, W_T]$. This yields N simulated values of the test statistic. The exact Monte Carlo p-value is then calculated from the rank of the observed LR relative to the simulated ones. Further details on Monte Carlo tests are presented in Appendix B.

Second, results specific to the gaussian special case of (2.8) lead to the F-tests used by GRS. Formally, if $\min(h, e) \leq 2$ where $h = \operatorname{rank}(H)$ and $e = \operatorname{rank}(E)$, then a (monotonic) transformation of LR follows the F-distribution with known degrees-of-freedom; see (A.4) in Appendix A.1. Obviously, this is relevant since H_C corresponds to $\min(h, e) = 1$.

For the CAPM problem, Theorem **A.1** in Appendix A.1 allows one to characterize the null distribution of LR for all error distributions which satisfy (2.8), as follows.

Theorem 3.1 DISTRIBUTION OF THE QUASI-LR CAPM TEST STATISTIC. *Under* (2.2), (2.3) *and* (2.8), *the LR statistic defined by* (3.12) *is distributed like*

$$L(W) \equiv T \ln \left(\left| W'MW \right| / \left| W'M_0W \right| \right)$$

where

$$M = I - X(X'X)^{-1}X', \; M_0 = M + X(X'X)^{-1}H'[H(X'X)^{-1}H']^{-1}H(X'X)^{-1}X',$$

H *is the row vector* $(1, 0)$ *and* $W = [W_1, \ldots, W_T]'$.

In the present case, we have $M_0 = I - \widetilde{r}_{\mathrm{M}}(\widetilde{r}_{\mathrm{M}}'\widetilde{r}_{\mathrm{M}})^{-1}\widetilde{r}_{\mathrm{M}}'$. Note that Theorem **3.1** also allows one to characterize the null distribution of LR in multi-beta efficiency tests problems. In other words, we can also use it to test H_C in the context of

$$r_{it} = a_i + \sum_{j=1}^{s} b_{ji}\widetilde{r}_{jt} + u_{it}, \quad t = 1, \ldots, T, \; i = 1, \ldots, n, \tag{3.13}$$

where $\widetilde{r}_{jt} = \widetilde{R}_{jt} - R_t^F$ and $\widetilde{R}_{jt}$, $j = 1, \ldots, s$, are returns on s benchmark portfolios. In this case, the null distribution of the statistic defined by (3.12) obtains as in Theorem **3.1** where $X = [\iota_T, \widetilde{r}_1, \ldots, \widetilde{r}_s]$, $\widetilde{r}_j = (\widetilde{r}_{1j}, \ldots, \widetilde{r}_{Tj})'$, and H is the $(s+1)$-dimensional row vector $(1, 0, \ldots, 0)$.

It is of interest to relate Theorem **3.1** to the available non-asymptotic tests of mean-variance efficiency, *i.e.*: (i) the GRS test, and (ii) the simulation-based test proposed by Zhou (1993). When errors are gaussian and $T - s - n \geq 1$, Theorem **3.1** and (A.4) entail that

$$\frac{(T - s - n)}{n}(\Lambda - 1) \sim F(n, T - s - n),$$

which yields the Hotelling T^2 test proposed by MacKinlay (1987) and Gibbons et al. (1989). Specifically, GRS suggest the following test statistic:

$$Q = \frac{T\,\hat{a}' \left[\frac{T}{T-k}\hat{\Sigma} \right]^{-1} \hat{a}}{1 + \bar{r}'\hat{\Delta}^{-1}\bar{r}} \tag{3.14}$$

where $\hat{a}$ is the vector of intercept OLS estimates, $\frac{T}{T-k}\hat{\Sigma}$ is the OLS-based unbiased estimator of Σ, $\bar{r}$ and $\hat{\Delta}$ include respectively the time-series-means and sample covariance matrix corresponding to the right-hand-side portfolio returns. Under (2.3), Q follows the *Hotelling* $T^2(n, T - s - 1)$ distribution or equivalently,

$$\frac{(T - s - n)}{n(T - s - 1)}Q \sim F(n, T - s - n) \tag{3.15}$$

8

where Q and Λ are related by a monotonic transformation:

$$\Lambda - 1 = \frac{Q}{T - s - 1} \; ; \tag{3.16}$$

see Stewart (1995, 1997). We thus see that GRS's results follow from Theorem **3.1** under the special case of normal errors.

To the best of our knowledge, the first study proposing useful finite-sample tests of mean-variance efficiency within a non-gaussian MLR is due to Zhou (1993), who reconsidered the GRS problem in models with elliptical distributions. In this context, Zhou demonstrates exact location/scale invariance of the GRS-type efficiency test statistic and exploits this property to approximate the relevant critical value by simulation (in applications, Zhou used 10000 replications). From Theorem **3.1**, we see that the required pivotality property holds under conditions much more general than elliptical symmetry. Further, from the general theory of Monte Carlo tests (Appendix B), we observe that the size of a simulation-based test can be *perfectly controlled* even with a very small number of Monte Carlo replications (provided it is implemented in the right way): for example, 19 replications are sufficient to obtain a test of size .05, although of course larger number of replications can also be employed [such as 39, 59, 79, 99, 119, . . . , for a test of size .05]. For power considerations, there is in principle an advantage in using a larger number of replication, but the power gain from using a number of replications larger than 100 or 200 is typically small; for further discussion and evidence on this issue, see Dwass (1957), Jöckel (1986), Dufour, Farhat, Gardiol and Khalaf (1998), Dufour (2002) and Dufour and Khalaf (2002*b*, 2002*d*).

Another problem _ which is not solved by Zhou (1993) or the above results _ consists in testing mean-variance efficiency when the error distribution has nuisance parameters which do not disappear from the null distribution of the test statistic. Theorem **3.1** simply ensures location-scale invariance (*i.e.*, invariance to B and Σ), which yields pivotality (for ν given) and allows the realization of an exact MC test. However, if ν is not specified, it will typically appear in the null distribution of the test statistic. We consider this case in the next section. Our empirical results illustrate the importance of formally accounting for such unknown parameters (see Section 6).

4. Mean-variance efficiency tests with an incompletely specified error distribution

In this section, we extend the above results to a case where the error distribution involves a nuisance parameter, namely (2.11). At this stage, two points deserve notice. First, for $\nu = \nu_0$ given, the distributional assumption (2.11) is covered by (2.8). Let $H_C(\nu_0)$ be the conjunction of H_C and $\nu = \nu_0$. Then the MC p-value provided by Theorem **3.1** is exact for the null hypothesis $H_C(\nu_0)$ provided the corresponding distribution $\mathcal{F}_i(\nu)$ is used to perform the simulations of the MC test. Second, whether ν is viewed (from an empirical perspective) as a parameter of interest or a nuisance parameter, it is important, for test validity, to devise a decision rule which takes this parameter into consideration. Otherwise, level control is not assured.

4.1. Two-stage constrained maximized Monte Carlo test

Here we propose a solution based on the finite-sample approach described in Dufour and Kiviet (1996). The method involves two stages: (1) an exact confidence set is built for ν, and (2) the MC p-value presented above is maximized over all the values of ν in the confidence set. We will refer to the latter test as a maximized MC [MMC] test. It is important to note that if a test with level α is desired, then the pre-test confidence set and the MMC test should be applied with levels $1 - \alpha_1$ and α_2, respectively, so that $\alpha = \alpha_1 + \alpha_2$. In the empirical application considered next, we use $\alpha_1 = \alpha_2 = \alpha/2$.

For any confidence set with level $1 - \alpha_1$ for ν which we will denote $\mathcal{C}(Y)$ where Y refers to the return data [as in (2.4)], the maximized MC algorithm proceeds as follows. On applying Theorem **3.1** and the MC algorithm in Appendix B.1, we can obtain for each $\nu \in \mathcal{C}(Y)$ a MC p-value $\hat{p}_N(\Lambda_0|\nu)$ [see (B.1), in Appendix B.1]. If we set

$$Q_U(\nu) = \sup_{\nu \in \mathcal{C}(Y)} \hat{p}_N(\Lambda_0|\nu), \tag{4.1}$$

then the critical region

$$Q_U(\nu) \leq \alpha_2 \tag{4.2}$$

has level $\alpha_1 + \alpha_2$. The associated test is conservative in the following sense: if indeed $Q_U(\nu) \leq \alpha_2$ for the sample at hand, then the test is certainly significant at level $\alpha = \alpha_1 + \alpha_2$.

Since a procedure to derive an exact confidence set for ν is not available, we provide one in what follows. The maximized MC procedure just presented is however not specific to our proposed confidence set. Observe that, in principle, the confidence set $\mathcal{C}(Y)$ may not be a bounded confidence interval. For proofs and further references, see Dufour (1990), Dufour and Kiviet (1996) and Dufour (2002).

4.2. Confidence set for error distribution parameters

We now discuss the set estimation method we propose to obtain $\mathcal{C}(Y)$. Given the recent literature documenting the dramatically poor performance of asymptotic Wald-type confidence intervals [see for example Dufour (1997), Staiger and Stock (1997), Wang and Zivot (1998)], we prefer to build a confidence set by "inverting" a test for the null hypothesis (2.11) where $\nu = \nu_0$ for known ν_0.

In Section 5.1, we describe a moment-based method for testing any given value of ν in (2.11). But our proposed set estimate for ν is however not specific to the latter test, so we present it in terms of an arbitrary test with level α_1 for (2.11) and based on a given criterion denoted $\mathcal{T}(Y)$, where Y refers to the return data [as in (2.4)].

The "inversion" of the test $\mathcal{T}(Y)$ is performed as follows. Let $\mathcal{T}_0(Y)$ denote the value of the statistic computed from the observed sample. Obtain the p-value $\hat{p}(\mathcal{T}_0(Y)|\nu_0)$ conforming with (2.11). For the GF tests described below, this is achieved on applying MC test techniques. The confidence set for ν corresponds to the values of ν_0 which are not rejected by the test, *i.e.* for which $\hat{p}[\mathcal{T}_0(Y)|\nu_0] > \alpha_1$.

It is useful to compare our confidence set MC test with Zhou (1993)'s test. This author considers

the multivariate skewness and kurtosis criteria proposed by Mardia (1970) and argues that these criteria may serve to test for departures from (2.11), if cut-off points are appropriately "approximated", *e.g.* by simulation, assuming (2.11). In view of this, he estimates ν as follows: a few values are retained by trial-and-error techniques (no further details are provided); then skewness and kurtosis tests are applied which confirm that the values retained do not yield significant lack-of-fit.

Although concerns regarding the possible conservative character of our inference procedure may not be ruled out, our proposed confidence set is definitely an improvement over available trial and error methods. From the results of our empirical analysis, we do observe that the estimated confidence sets are wide, yet the associated efficiency test decision is not adversely affected.

5. Exact diagnostic checks

In this section, we propose multivariate specification tests, including distributional goodness-of-fit tests and checks for departures from the hypothesis of *i.i.d.* errors. We present in turn exact multivariate GF tests (for a given error distribution), tests for multivariate GARCH effects and multivariate variance ratio (VR) tests. The proposed tests are formally valid for any parametric error model consistent with (2.8). Such procedures have not apparently been proposed in the earlier literature on the statistical analysis of MLR models. In conformity with our empirical model, we focus on (2.11) with possibly unknown parameters.

5.1. Goodness-of-fit tests

The null hypothesis of concern here is (2.11) with ν unspecified. We will solve the unknown ν problem by applying a MMC strategy, as follows. We propose a GF criterion which is pivotal when ν is specified by the null hypothesis and thus allows to easily obtain a MC p-value given ν. The GF test is considered significant if the largest MC p-value overall relevant values of ν is less than or equal to the desired significance level. The MMC approach allows one to (jointly) assess whether Student distributions with different degrees of freedom are empirically relevant; this leads to a formal estimate for the Student distributions which are consistent with the data. The argument also holds for multivariate mixtures.

To test the goodness-of-fit of alternative disturbance distributions, we reconsider the multivariate skewness and kurtosis criteria used by Zhou (1993):

$$SK \;\; = \;\; \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T} \hat{d}_{st}^3 \, , \tag{5.1}$$

$$KU \;\; = \;\; \frac{1}{T} \sum_{t=1}^{T} \hat{d}_{tt}^2 \, , \tag{5.2}$$

where $\hat{d}_{it}$ are the elements of the matrix

$$\hat{D} = \hat{U}' \hat{\Sigma}^{-1} \hat{U} = T \, \hat{U} (\hat{U}' \hat{U})^{-1} \hat{U}' . \tag{5.3}$$

11

These statistics were introduced by Mardia (1970) to assess deviations from multivariate normality, in models where the regressor matrix reduces to a vector of ones (the location-scale model). Zhou proposed to use them as well to test elliptically symmetric distributions, without however providing a finite-sample theory for their application to least squares residuals from MLR models _ a limitation pointed out by Zhou (1993, p. 1935, footnote 5) himself.

In order to obtain provably valid GF tests in MLR models, we shall use the following characterization of the distribution of SK and KU statistics in MLR models.

Proposition 5.1 DISTRIBUTION OF THE MULTIVARIATE SKEWNESS AND KURTOSIS TEST STATISTICS. *Under* (2.4), *and for all error distributions compatible with* (2.8), *the multivariate skewness and kurtosis criteria* (5.1) *and* (5.2) *are distributed, respectively, like* $\frac{1}{T^2} \sum_{t=1}^{T} \sum_{i=1}^{T} d_{st}^3$ *and* $\frac{1}{T} \sum_{t=1}^{T} d_{tt}^2$, *where* d_{st} *are the* (s,t)-*th element of the matrix* $D = MW \left(W' \bar{M} \bar{M} W \right)^{-1} W' M$, $\bar{M} = (1/T)[I - X(X'X)^{-1}X']$ *and* $W = [W_1, \ldots, W_T]'$.

The proof of latter proposition is given in Appendix B.2. This proposition shows that the statistics SK and KU follow null distributions without nuisance parameters. In the literature on multivariate normality tests, this property is recognized (under normality) in models where the regressors reduce to a vector of ones. Proposition **5.1** entails that nuisance parameter invariance holds even though residuals (rather than observable variables) are used to construct the skewness and kurtosis statistics. This implies that the MC tests method can be straightforwardly applied to test any distributions compatible with (2.8), including the normal. Before implementing these tests, we shall, however, consider two further adjustments: (1) a simulation-based "centering" of the test statistics; (2) a formal procedure for combining them into a single test.

The statistics SK and KU are not unbiased estimators of the relevant moments, especially for non-normal distributions. To avoid corresponding biases (and power losses) in the tests based on these statistics, it appears desirable to "center" the test statistics with respect to their mean under the error distribution tested. Further, such a centering can simplify the construction of two-sided tests (which are relevant in the present problem). We thus propose the following modification: we consider alternative measures of skewness and kurtosis in excess of expected values consistent with (2.11). For ν given, our modified tests are pivotal under the null hypothesis which justifies an MMC test technique, maximizing overall ν. We next propose an exact combined skewness-kurtosis test. Our proposed modified statistics take the following form:

$$ESK(\nu_0) = \left| SK - \overline{SK}(\nu_0) \right|, \tag{5.4}$$
$$EKU(\nu_0) = \left| KU - \overline{KU}(\nu_0) \right|, \tag{5.5}$$

where $\overline{SK}(\nu_0)$ and $\overline{KU}(\nu_0)$ are simulation-based estimates of the expected SK and KU given (2.11). These may be obtained, given ν_0, by drawing N_0 samples of T observations from (2.11), then computing the corresponding average measures of skewness and kurtosis; see Appendix B.2.1.[9]

To obtain an exact test based on these criteria, we apply the MC technique [as described in Appendix B.1]. Note that the observed and simulated statistics have to be obtained conditional

---

[9]For the Gaussian case, one may use $\overline{SK} = 0$ and $\overline{KU} = n(n+2)$; see Mardia (1970).

on the same $\overline{SK}(\nu_0)$ and $\overline{KU}(\nu_0)$; see Appendix B.2.2 for further details. This ensures that they remain exchangeable, which provides, along with Proposition **5.1**, the necessary conditions for the validity of the MC p-values in B5; see Dufour (2002).

This procedure allows to obtain size correct individual p-values for each test statistic. The problem of combining the skewness and kurtosis tests remains unanswered. To avoid relying on Boole-Bonferroni rules, we propose the following combined test statistic, which may be used for all null hypotheses underlying Proposition **5.1**:

$$CSK = 1 - \min\left\{\hat{p}\big(ESK(\nu_0)|\nu_0\big),\ \hat{p}\big(EKU(\nu_0)|\nu_0\big)\right\} \tag{5.6}$$

where the subscript N (previously used in the notation for the individual MC p-values) is suppressed to simplify notation. The intuition underlying this combined criterion is to reject the null hypothesis if at least one of the individual tests is significant; for convenience, we subtract the minimum p-value from one to obtain a right-sided test. The MC test technique may once again be applied to obtain a test based on the combined statistic; details of the algorithm can be found in Appendix B.2.3. For further reference on such combined tests, see Dufour and Khalaf (2002b), Dufour, Khalaf, Bernard and Genest (2001) and Dufour and Khalaf (2002c).

5.2. Multivariate tests for GARCH effects and variance ratio tests

We now consider tests for departure from *i.i.d.* errors, specifically, tests for GARCH effects and variance ratio tests. If one pursues a univariate approach, these standard tests may be applied to each equation in the system (2.2). For instance, the Engle GARCH test statistic for equation i, which we will denote E_i is given by TR_i^2, where T is the sample size, R_i^2 is the coefficient of determination in the regression of the equation's squared OLS residuals $\hat{u}_{it}^2$ on a constant and $\hat{u}_{(t-j),i}^2$ ($j = 1,\ \dots\ ,\ q$); see Engle (1982) and Lee (1991). Lee and King (1993) proposed an alternative test which exploits the one-sided nature of H_A. The test statistic is

$$LK_i = \frac{\left\{(T-q)\sum_{t=q+1}^{T}\left[(\hat{u}_{it}^2/\hat{\sigma}_i^2 - 1)\right]\sum_{j=1}^{q}\hat{u}_{i,t-j}^2\right\} \Big/ \left\{\sum_{t=q+1}^{T}(\hat{u}_{it}^2/\hat{\sigma}_i^2 - 1)^2\right\}^{1/2}}{\left\{(T-q)\sum_{t=q+1}^{T}\left(\sum_{j=1}^{q}\hat{u}_{i,t-j}^2\right)^2 - \left(\sum_{t=q+1}^{T}\left(\sum_{j=1}^{q}\hat{u}_{i,t-j}^2\right)\right)^2\right\}^{1/2}} \tag{5.7}$$

where $\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_{it}^2$, and its asymptotic null distribution is standard normal. The variance ratio test statistic [Lo and MacKinlay (1988, 1989)] is:

$$VR_i = 1 + 2\sum_{j=1}^{J}(1 - \frac{j}{J})\hat{\rho}_{ij} \tag{5.8}$$

where

$$\hat{\rho}_{ij} = \frac{\sum_{t=j+1}^{T} \hat{u}_{it}\hat{u}_{i,t-j}}{\sum_{t=1}^{T} \hat{u}_{ti}^2}, \quad j = 1, \dots, J, \tag{5.9}$$

are empirical residual autocorrelations. The latter statistic estimates the ratio

$$\frac{\mathsf{V}(\hat{u}_{it} - \hat{u}_{i,t-J})}{J\,\mathsf{V}(\hat{u}_{it})}$$

where $\mathsf{V}(\hat{u}_{it} - \hat{u}_{i,t-J})$ is the variance of the lag differences $\hat{u}_{it} - \hat{u}_{i,t-J}$, and $\mathsf{V}(\hat{u}_t)$ is the residual variance. In a single equation perspective, under *i.i.d.* errors, $\mathsf{V}(\hat{u}_{it} - \hat{u}_{i,t-J})$ is J times $\mathsf{V}(\hat{u}_{it})$, hence deviations from a ratio of one are considered evidence against the null hypothesis. The asymptotic null distribution of this statistic is

$$VR_i - 1 \overset{asy}{\sim} N\big[0,\ 2(2J-1)(J-1)/3J\big]. \tag{5.10}$$

In Dufour, Khalaf, Bernard and Genest (2001), we showed that the Engle and Lee-King test criteria are nuisance-parameter-free under the homoskedasticity hypothesis, in a single equation setting. We establish the same property in the case of the variance ratio test in Dufour and Khalaf (2002a). This ensures that the MC versions of these tests are valid univariate tests (and preferred to the asymptotic tests). However, it is well known that such univariate tests may not be appropriate in multivariate regressions. This is mainly due to two statistical problems. First, as pointed out above, the error covariance, which appears as a nuisance parameter, is typically not taken into consideration if a series of univariate tests are applied. Second, the problem of combining test decisions overall equations is not straightforward, since the individual tests are not independent. For further useful insight on this problem in finance, see Shanken (1990).

In view of this, we consider the following multivariate modification of these tests [see Dufour, Khalaf and Beaulieu (2001)]. Let $\widetilde{W}_{it}$ denote the elements of the *standardized residuals* matrix

$$\widetilde{W} = \hat{U}(\hat{U}'\hat{U}\,)^{-1/2} \tag{5.11}$$

where $(\hat{U}'\hat{U}\,)^{-1/2}$ refers to the inverse of a Cholesky-type decomposition of $\hat{U}'\hat{U}$. Obtain standardized versions of the univariate Engle, Lee-King and variance ratio test, denoted $\widetilde{E}_i$, $\widetilde{LK}_i$ and $\widetilde{VR}_i$, replacing $\hat{u}_{it}$ by $\widetilde{W}_{it}$ in the formula for these statistics. In Dufour, Khalaf and Beaulieu (2001) we show that for all error distributions compatible with (2.8), $\widetilde{W}$ has a distribution which is completely determined by the distribution of W given X. Hence any statistic which depends on the data only through $\widetilde{W}$ has a distribution which is invariant to B and Σ, under (2.8). It follows that under (2.8), the *joint* (across equations) null distributions of $\widetilde{E}_i$, $\widetilde{LK}_i$ and $\widetilde{VR}_i$ do not depend on B and Σ.

To obtain combined inference across equation, we propose a combination method similar to the one we used in Section 5.1. The combined statistics are:

$$\widetilde{E} = 1 - \min_{1 \leq i \leq n} \big[p(\widetilde{E}_i)\big], \tag{5.12}$$

$$\widetilde{LK} = 1 - \min_{1 \leq i \leq n} \big[p(\widetilde{LK}_i)\big], \tag{5.13}$$

$$\widetilde{VR} = 1 - \min_{1 \le i \le n} \left[p(\widetilde{VR_i}) \right] , \tag{5.14}$$

where $p(\widetilde{E}_i)$, $p(\widetilde{LK}_i)$ and $p(\widetilde{VR}_i)$ refer to p-values; these may be obtained applying a MC test method, or using asymptotic null distributions (to cut execution time). Then apply a MMC test procedure to the combined statistic imposing (2.11). In Appendix B, we provide a MMC test algorithm for any criterion which is a pivotal function of X and W, where the distribution of W depends on the parameter ν. We use the same confidence set for ν as in the MMC mean-variance efficiency test. The overall procedure remains exact even if approximate individual p-values are used, if the p-value of the combined test is obtained applying the MMC technique. Indeed, the property underlying exactness is joint pivotality, which was achieved by using standardized residuals.

6. Empirical analysis

Our empirical analysis focuses on mean-variance efficiency tests of the market portfolio [formally, tests of (2.3) in the context of (2.2)] with different distributional assumptions on stock market returns. We use nominal monthly returns over the period going from January 1926 to December 1995, obtained from the University of Chicago's Center for Research in Security Prices (CRSP). As in Breeden, Gibbons and Litzenberger (1989), our data include 12 portfolios of New York Stock Exchange (NYSE) firms grouped by standard two-digit industrial classification (SIC). Table 1 provides a list of the different sectors used as well as the SIC codes included in the analysis.[10] For each month the industry portfolios comprise those firms for which the return, the price per common share and the number of shares outstanding are recorded by CRSP. Furthermore, portfolios are value-weighted in each month. In order to assess the testable implications of the asset pricing models, we measure the market return by the value-weighted NYSE returns, also available from CRSP. The risk-free rate is measured by the one-month Treasury Bill rate, also from CRSP.

Our results are summarized in Tables 2 and 3. All MC tests where applied with 999 replications. As usual in this literature, we estimate and test the model over intervals of 5 years.[11] We report in columns (1)-(3) of Table 2, the p-values of the exact multi-normality tests based on ESK, EKU and CSK (see Section 5.1). These tests allow us to evaluate whether observed residuals exhibit non-gaussian behavior through excess skewness and kurtosis. For most subperiods, normality is rejected. These results are interesting since, although it is well accepted in the finance literature that continuously compounded returns are skewed and leptokurtic, empirical evidence of non-normality is weaker for monthly data; for instance, Affleck-Graves and McDonald (1989) reject normality in about 50% of the stocks they study. Our results, which are exact (*i.e.*, cannot reject spuriously), indicate much stronger evidence against normality. This also confirms the results of Richardson and Smith (1993) who provide evidence against multivariate normality based on asymptotic tests; see also Fiorentini, Sentana and Calzolari (2000). Of course, this evidence provides further motivation for using our approach to test mean-variance efficiency under non-gaussian errors.

---

[10]Note that as in Breeden et al. (1989), firms with SIC code 39 (Miscellaneous manufacturing industries) are excluded from the dataset for portfolio formation.

[11]We ran the analysis with October 1987 and January returns over 5 and 10 year subperiods as well as without those observations over 10 year subperiods. Our results are not significantly affected by such modifications.

Table 1. Portfolio definitions

Portfolio number	Industry Name	Two-digit SIC codes
1	Petroleum	13, 29
2	Finance and real estate	60-69
3	Consumer durables	25, 30, 36, 37, 50, 55, 57
4	Basic industries	10, 12, 14, 24, 26, 28, 33
5	Food and tobacco	1, 20, 21, 54
6	Construction	15-17, 32, 52
7	Capital goods	34, 35, 38
8	Transportation	40-42, 44, 45, 47
9	Utilities	46, 48, 49
10	Textile and trade	22, 23, 31, 51, 53, 56, 59
11	Services	72, 73, 75, 80, 82, 89
12	Leisure	27, 58, 70, 78, 79

Note _ This table presents portfolios according to their number and sector as well as the SIC codes included in each portfolio using the same classification as Breeden et al. (1989).

In columns (4)-(7) of Table 2, we present the LR statistics for mean-variance efficiency, the corresponding asymptotic p-values obtained from the asymptotic $\chi^2(n-1)$ distribution (p_∞), the exact gaussian-based MC p-values $(p_\mathcal{N})$, and the maximized MC p-values based on the Student-t error model (Q_U). The confidence set $\mathcal{C}(Y)$ for the number of degrees of freedom κ appears in column (8). These results allow one to compare rejection decisions across different distributional assumptions on the returns of the 12 portfolios. Similarly in columns (1)-(5) of Table 3, we report our set estimates of π and ω; for presentation simplicity, the confidence region is summarized as follows: we give the confidence set for ω corresponding to five different values of π [$\pi = 0.1, 0.2, 0.3, 0.4, 0.5$]. Column (6) of Table 3 present the largest mixture-of-normals based MC p-values associated with the efficiency LR statistic [reported in column (4) of Table 2]. This empirical evidence shows that asymptotic p-values are quite often spuriously significant (*e.g.*, for 1941-55). Furthermore, the maximal p-values exceed the gaussian-based p-value. It is "easier" to reject the testable implications under normality. Conversely, recall that the gaussian model obtains as $\kappa \to \infty$. So, if $p_\mathcal{N}$ exceeds the significance level, the largest p-value *a fortiori* also exceeds the significance level. Thus the decision implied by a non-significant gaussian p-value is exactly conclusive (*i.e.*, there is no need to reconsider t-based p-values if $p_\mathcal{N}$ fails to reject). For instance, at the 5% level of confidence, we find ten rejections of the null hypothesis for the asymptotic $\chi^2(11)$ test, nine for the MC p-values under normality, six for the MC under the Student-t distribution and, as shown on the last column of Table 3, seven under the mixtures of normal distributions.[12] These

---

[12]Our tests for MC p-values under the Student and mixtures of normals distributions are joint tests for nuisance parameters consistent with the data and the mean-variance efficiency hypothesis. Since we have attributed a level of 2.5% to the construction of the confidence set, to establish a fair comparison with the MC p-values under the normality assumption or the asymptotic p-values, we must refer the p-values for the efficiency tests under the Student and the mixtures of normals distributions to 2.5%.

Table 2. Normality and mean-variance efficiency tests

	Normality tests			Tests of H_C				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample	SK	KU	CSK	LR	p_∞	$p_\mathcal{N}$	Q_U	$\mathcal{C}(Y)$
1927-30	.001	.001	.001	16.104	.1866	.364	.357	3-12
1931-35	.001	.001	.001	16.257	.1798	.313	.322	3-8
1936-40	.001	.001	.001	16.018	.1904	.319	.333	4-26
1941-45	.004	.002	.004	25.869	.0112	**.045**	**.049**	≥ 5
1946-50	.001	.001	.001	37.196	.0002	.003	.004	4-26
1951-55	.001	.002	.001	36.510	.0003	.004	.005	5-31
1956-60	.024	.003	.003	43.841	.0000	.002	.002	≥ 5
1961-65	.594	.479	.631	39.098	.0001	.002	.002	≥ 7
1966-70	.011	.002	.004	36.794	.0002	.003	.003	≥ 5
1971-75	.001	.002	.001	21.094	.0490	.120	.129	4-24
1976-80	.001	.001	.001	28.373	.0049	**.023**	**.026**	4-17
1981-85	.001	.001	.001	27.189	.0073	**.033**	**.035**	5-34
1986-90	.028	.020	.030	35.747	.0007	.003	.005	≥ 5
1991-95	.177	.311	.239	16.752	.1592	.299	.305	≥ 15

Notes _ Numbers in columns (1)-(3) represent p-values for multinormality tests: numbers in (1)-(2) pertain to the null hypothesis of respectively no excess skewness and no excess kurtosis in the residuals of each subperiod. The p-values in column (3) correspond to the combined statistic CSK designed for joint tests of the presence of skewness and kurtosis; individual and joint statistics obtain applying (5.5), (5.5) and (5.6) given multivariate normal errors. p-values are MC pivotal statistics based. Column (4) presents the quasi-LR statistic defined in (3.12) to test H_C [see (2.3)]; columns (5), (6) and (7) are the associated p-values using, respectively, the asymptotic $\chi^2(n)$ distribution, the pivotal statistics based MC test method imposing multivariate normal regression errors, and an MMC confidence set based method imposing multivariate $t(\kappa)$ errors which yields the largest MC p-value for all κ within the specified confidence set. The latter is reported in column (8). See Section 4.2 for details on the construction of the confidence set: the values of κ in this set are not rejected by the joint test GF test associated with (5.6) under multivariate Student-t errors. See Appendix B for description of MC tests. January and October 1987 returns are excluded from the dataset.

Table 3. Tests of mean-variance efficiency with multivariate normal mixture errors

| Sample | Confidence set for π, ω | | | | | LR |
| | (1) | (2) | (3) | (4) | (5) | (6) |
	$\pi = .1$	$\pi = .2$	$\pi = .3$	$\pi = .4$	$\pi = .5$	Q_U
1927-30	≥ 1.7	1.6-2.8	1.6-2.5	1.6-2.5	1.6-2.6	.382
1931-35	≥ 2.1	1.9-3.0	1.9-2.7	1.9-2.7	2.1-3.0	.313
1966-40	1.5-3.5	1.4-2.3	1.4-2.0	1.5-2.2	1.4-2.0	.328
1941-45	1.3-3.0	1.3-2.1	1.3-1.9	1.3-2.0	1.3-1.9	**.043**
1946-50	1.5-3.5	1.4-2.2	1.4-2.0	1.4-2.3	1.4-2.0	.003
1951-55	1.4-3.5	1.4-2.2	1.4-2.0	1.3-2.1	1.4-2.0	.003
1956-60	1.3-2.8	1.2-2.0	1.2-1.9	1.0-1.9	1.2-1.9	.002
1961-65	1.0-2.2	1.0-1.7	1.0-1.5	1.0-1.5	1.0-1.5	.002
1966-70	1.3-2.8	1.3-2.0	1.3-1.9	1.3-1.8	1.2-1.9	.002
1971-75	1.5-3.5	1.5-2.2	1.4-2.0	1.4-2.2	1.4-2.0	.128
1976-80	1.6-4.0	1.5-2.5	1.5-2.2	1.5-2.4	1.5-2.3	.022
1981-85	1.4-3.5	1.4-2.2	1.3-2.0	1.4-2.1	1.4-2.4	**.030**
1986-90	1.1-3.0	1.1-2.0	1.1-1.9	1.0-1.9	1.1-1.7	.004
1991-95	1.0-1.9	≤ 1.5	≤ 1.3	1.0-1.4	1.0-1.3	.306

Note _ Numbers in columns (1)-(5) represent a confidence set for the parameters (π, ω) [respectively, the probability of mixing and the ratio of scales] of the multivariate mixtures-of-normal error distribution. See Section 4.2 for details on the construction of the confidence set: the values of (π, ω) in this set are not rejected by the joint test GF test associated with (5.6) under multivariate mixture errors. The maximum of the p-value occurs in the closed interval for ω. Column (6) presents a MMC p-value relative to the quasi-LR statistic defined in (3.12) to test H_C [see (2.3)]; the observed values of this statistic are reported in Table 2, column (4). The MMC p-value is the largest MC p-value for all (π, ω) within the reported confidence set. The maximum of the p-value occurs in the closed interval for ω. See Appendix B for description of MC tests. January and October 1987 returns are excluded from the dataset.

findings differ from those of Zhou (1991), who found no change in rejection rates of mean-variance efficiency using elliptical distributions other than the normal. This difference can be explained by the fact that Zhou did not explicitly account for nuisance parameters. Interestingly, whenever the results obtained under non-gaussian distributions differ from those obtained under the gaussian distribution, the gaussian distributional assumption is strongly rejected.

Our results clearly indicate that GRS-type tests are sensitive to the hypothesized error distribution. Of course, this observation is relevant when the hypothesized distributions are empirically consistent with the data. Focusing on the t and mixture distributions with parameters not rejected by exact GF tests, we see that the decision of the MMC mean-variance efficiency test can change relative to the F-based test.

Figures 1 to 14 illustrate how the p-value varies overall $\mathcal{C}(Y)$ for the t-distribution. Although $\mathcal{C}(Y)$ is quite wide, it is evident from Figures 1-14 that restricting this set further does not have a strong influence on the decision. Specifically, the p-values do not seem to fluctuate a lot throughout $\mathcal{C}(Y)$, at least in this application.

It is usual to aggregate the efficiency test results overall subperiods, in some manner. For instance, Gibbons and Shanken (1987) propose two aggregate statistics which, in terms of our notation, may be expressed as follows:

$$GS_1 = -2 \sum_{j=1}^{14} \ln(p_{\mathcal{N}}[j]) \,, \quad GS_2 = \sum_{j=1}^{14} \Psi^{-1}(p_{\mathcal{N}}[j]) \tag{6.1}$$

where $[j]$ refers to the sub-periods, and $\Psi^{-1}(.)$ provides the standard normal deviate corresponding to $p_{\mathcal{N}}[j]$. If the mean-variance efficiency hypothesis holds across all subperiods, then $GS_1 \sim \chi^2(2 \times 14)$ whereas $GS_2 \sim N(0, 14)$. It is worth noting that the same aggregation methods can be applied to our test problem even under (2.11) by replacing, in (6.1), $p_{\mathcal{N}}[j]$ with $Q_{U[j]}$, the MMC p-values obtained imposing (2.11). Indeed, as is observed by Gibbons and Shanken (1987), the F-distribution is not necessary to obtain the null distribution of these combined statistics. All what is needed is a continuous null distribution (a hypothesis satisfied given normal, Student t or mixture errors) and, of course, independence across subperiods. Our results, under normal, Student t and mixture errors respectively, are: $GS_1 = 102.264, 101.658$ and 105.464 and $GS_2 = 28.476, 28.397$ and 28.476; all associated p-values are smaller than .0000. If independence is upheld as in Gibbons and Shanken (1987), this implies that mean-variance efficiency is jointly rejected with our data.[13]

Finally, Table 4 presents the results of our multivariate exact diagnostic checks for departures from the *i.i.d.* assumption, namely our proposed multivariate versions of the Engle, Lee-King and variance ratio tests; we use 12 months-lags.[14] The results show very few rejections of the null hy-

---

[13]Note that even if one questions independence and prefers to combine using Bonferroni-based criteria, the smallest p-value is .002 which when referred to $.025/14 \simeq .002$ comes close to a rejection. In the context of a MC with 999 replications, the smallest possible p-values are .001, .002 and so on so forth. To allow a fair Bonferroni test, it is preferable to consider the level $.028/14 = .002$. This means that in every period, the pre-test confidence set should be applied with $\alpha_1 = 2.2\%$ to allow 2.8% to the mean-variance efficiency test. The results reported in the above Tables are robust to this change of levels.

[14]We have also run univariate diagnostic checks. For space considerations, we only report the multivariate results. Note that the univariate test results are available upon request.
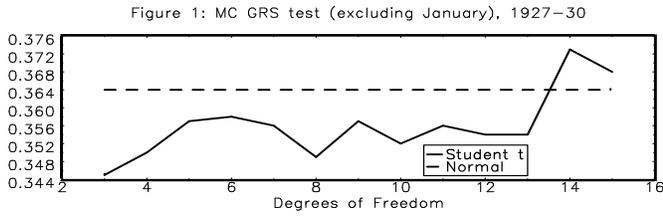
Figure 1: MC GRS test (excluding January), 1927−30
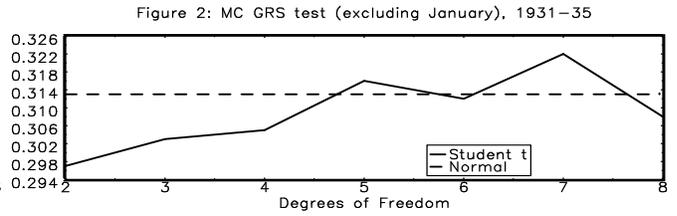
Figure 2: MC GRS test (excluding January), 1931−35
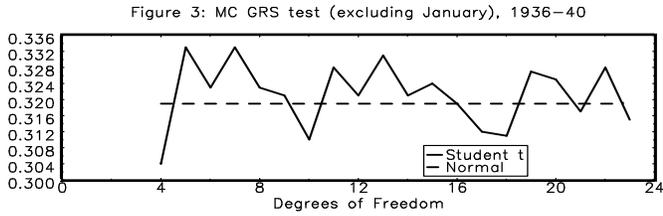
Figure 3: MC GRS test (excluding January), 1936−40
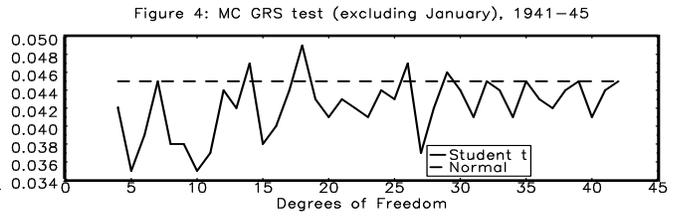
Figure 4: MC GRS test (excluding January), 1941−45
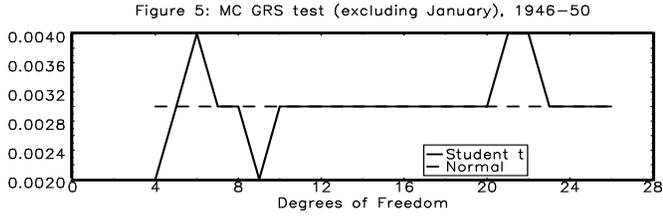
Figure 5: MC GRS test (excluding January), 1946−50
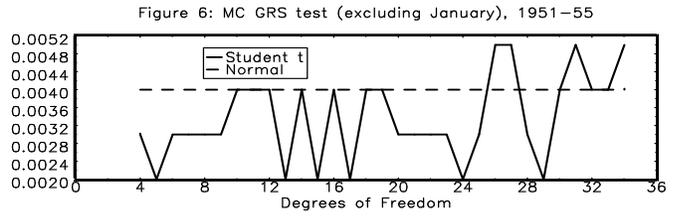
Figure 6: MC GRS test (excluding January), 1951−55
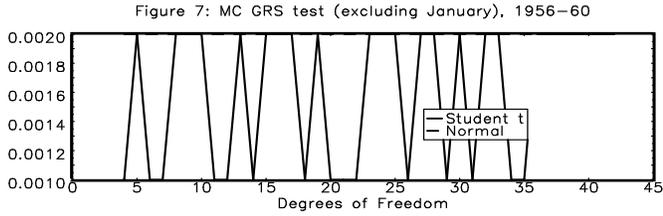
Figure 7: MC GRS test (excluding January), 1956−60
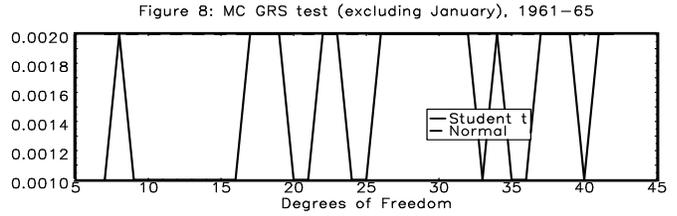
Figure 8: MC GRS test (excluding January), 1961−65
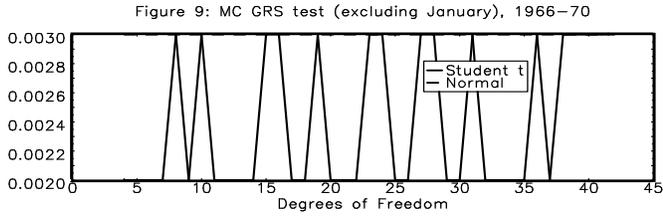
Figure 9: MC GRS test (excluding January), 1966−70
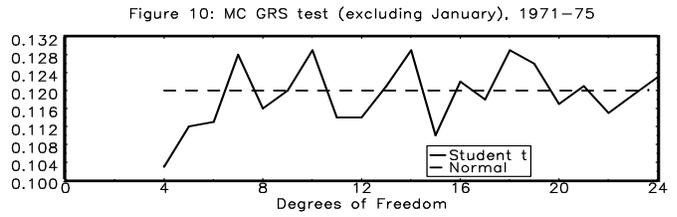
Figure 10: MC GRS test (excluding January), 1971−75
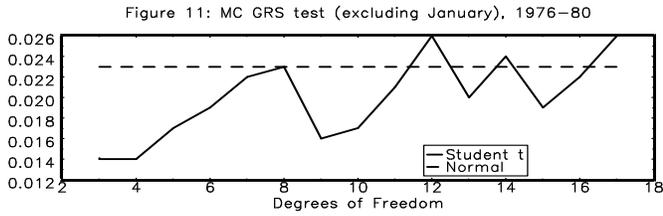
Figure 11: MC GRS test (excluding January), 1976−80
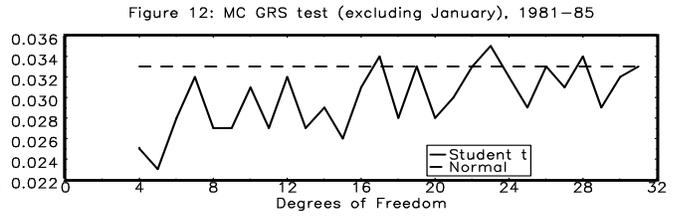
Figure 12: MC GRS test (excluding January), 1981−85
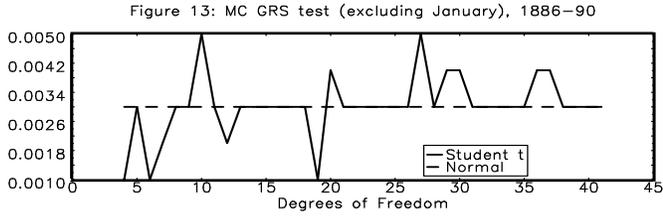
Figure 13: MC GRS test (excluding January), 1886−90
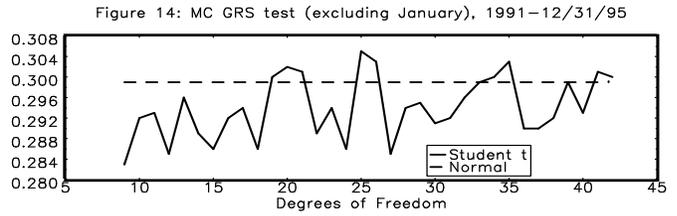
Figure 14: MC GRS test (excluding January), 1991−12/31/95

Each figure contains a plot of Degrees of Freedom on the x-axis with a continuous line (Student t) and a dashed line (Normal).

Note _ Using the Student-t distribution, each figure represents the Monte Carlo probability of rejecting the hypothesis that $H_C : a_i = 0, i = 1, \ldots, n$ associated with the degrees of freedom on the confidence set with the continuous line. The dash line represents the Monte-Carlo p-value of rejecting H_C using the normality assumption.

20

Table 4. Multivariate diagnostics

Sample	Normal errors			Student-t errors			Mixture errors		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$\widetilde{E}$	$\widetilde{LK}$	$\hat{VR}$	$\widetilde{E}$	$\widetilde{LK}$	$\widetilde{VR}$	$\widetilde{E}$	$\widetilde{LK}$	$\widetilde{VR}$
1927-30	.001	.356	.004	.013	.301	.004	.155	.335	.004
1931-35	.022	.748	.069	.082	.659	.066	.208	.671	.080
1936-40	.075	.612	.855	.124	.587	.867	.150	.672	.864
1941-45	.824	.979	.163	.843	.982	.177	.831	.979	.175
1946-50	.003	.804	.063	.017	.784	.068	.029	.772	.072
1951-55	.139	.353	.111	.168	.321	.120	.190	.326	.120
1956-60	.987	.628	.093	.994	.628	.095	.996	.625	.099
1961-65	.339	.207	.577	.375	.195	.584	.344	.207	.592
1966-70	.027	.274	.821	.043	.278	.847	.065	.288	.846
1971-75	.280	.224	.218	.316	.212	.224	.326	.207	.217
1976-80	.004	.011	.165	.016	.013	.183	.043	.009	.184
1981-85	.027	.103	.208	.050	.103	.217	.081	.095	.223
1986-90	.033	.453	.346	.077	.442	.366	.089	.455	.357
1991-95	.803	.236	.088	.821	.237	.092	.816	.252	.092

Note _ Numbers shown are p-values associated with the combined tests $\widetilde{E}$, $\widetilde{LK}$, and $\widetilde{VR}$, defined by (5.12), (5.13) and (5.14). $\widetilde{E}$ and $\widetilde{LK}$ are multivariate versions of Engle's and Lee and King's GARCH tests and $\widetilde{VR}$ is a multivariate version of Lo and MacKinlay's variance ratio tests; see Section 5.2. In columns (1)-(3), the p-values are MC pivotal statistics based; p-values in columns (4)-(9) are MMC confidence set based. The relevant 2.5% confidence set for the nuisance parameters is reported in Table 2, column (8) for the multivariate Student-t distribution, and in Table 3, columns (1)-(5) for the multivariate mixture-of-normals distribution. See Appendix B for description of MC tests. January and October 1987 returns are excluded from the dataset.

pothesis both at the 1% and 5% level of significance. This implies that, in our statistical framework, *i.i.d.* errors provide an acceptable working assumption.

7. Conclusion

We have shown that in gaussian or non-gaussian contexts, the exact test procedure proposed in Dufour and Khalaf (2002*d*) may be used to perform a mean-variance efficiency test of the market portfolio. We have specifically illustrated how to deal in finite samples with Student-t errors and multivariate mixtures of normals, with possibly unknown parameters.

Our empirical results are important for assessing the reliability and empirical performance of the CAPM. It appears that the normality assumption is too restrictive given the observed financial return data, even with monthly data. First, while our exact multivariate GF tests conclusively reject normality, Student-t or mixtures-of-normals are consistent with our data. Furthermore, we show that mean-variance efficiency exact tests which formally take these non-normal distributions

into consideration fail to reject mean-variance efficiency for 3 out of 9 subperiods for which the gaussian-based test is significant. It appears that the distributional set-up is crucial when testing mean-variance efficiency. This suggests that more work is needed from a theoretical perspective to better circumscribe the necessary and sufficient distributional hypotheses underlying fundamental asset pricing models.

Although we focused on mean-variance efficiency tests, it is worth emphasizing that our proposed methodology applies to several interesting asset pricing tests including many problems where the Hotelling test [exploited by GRS and MacKinlay (1987)] and Rao's F test [see Stewart (1997) and (A.4) in Appendix A.1] have been used. Although, in view of its fundamental importance, mean-variance efficiency is one of the first and very few MLR-based problems which have been approached from an exact perspective, a few authors have recognized that hypotheses dealing with the joint significance of the coefficients of *two* regression coefficients across equations can also be tested exactly applying Rao's F test. Examples include inter-temporal asset pricing tests in Shanken (1990, footnote 18). Furthermore, as discussed in Shanken (1996), econometric tests of spanning fall within this class. Indeed, spanning tests [see Jobson and Korkie (1989), Kan and Zhou (2001)] may be written in terms of a model of the GRS form. The hypothesis is however more restrictive, in the sense that over and above the restriction on the intercepts, the betas for each regression are required to sum to one. These hypotheses fit into our UL framework. The results in this paper extend available exact tests of these important financial problems beyond the gaussian context.

The fact remains that the results presented in this paper are specific to UL hypotheses. Recall that not all linear hypotheses may be cast in this form. We study extensions to non-linear problems including tests of Black's version of the CAPM in Beaulieu et al. (2001). Finally, we note that an apparent shortcoming of our exact tests comes from the fact that the right-hand-side benchmark may be observed with errors. The development of exact tests which correct for error-in-variable problems is an appealing idea for future research.

Appendix

A. Appendix: Finite-sample distributional theory

The results in this Appendix pertain to any asset pricing model which may be cast in terms of the MLR given by (2.4) and (2.8).

A.1. General uniform linear restrictions tests

In this section, we discuss the testing of constraints on regression coefficients of the UL form (2.6). On observing that (2.6) corresponds to $(E' \otimes H) \, vec(B) = vec(A)$, it is clear that not all linear hypotheses can be cast in the UL form. The associated gaussian quasi-LR statistic is:

$$LR = T \ln(|\hat{\Sigma}_0|/|\hat{\Sigma}|) = T \ln(\Lambda) , \tag{A.1}$$

where $\Lambda = |\hat{\Sigma}_0|/|\hat{\Sigma}|$ and $\hat{\Sigma}_0$ is the constrained MLE of Σ. The statistic Λ corresponds to the inverse of the well known Wilks statistic. The following exact distributional results are proved in Dufour and Khalaf (2002d).

Theorem A.1 DISTRIBUTION OF THE QUASI-LR UNIFORM-LINEAR HYPOTHESIS TEST STATISTIC. *Under* (2.4), (2.8) *and* (2.6), *the statistic*

$$\Lambda = |\hat{\Sigma}_{01}|/|\hat{\Sigma}| \tag{A.2}$$

is distributed like

$$\overline{\Lambda}(W) = \left| E'W'MWE \right| / \left| E'W'M_0WE \right| \tag{A.3}$$

where $\hat{\Sigma}_{01}$ *and* $\hat{\Sigma}$ *are the constrained and unconstrained MLE of* Σ, $W = [W_1, \ldots, W_n]$ *and*

$$
\begin{aligned}
M_0 &= M + X(X'X)^{-1}H'[H(X'X)^{-1}H']^{-1}H(X'X)^{-1}X' , \\
M &= I - X(X'X)^{-1}X .
\end{aligned}
$$

For certain values of h and e and normal errors, the null distribution reduces to the F distribution. For instance, if $\min(h, e) \leq 2$, then

$$\frac{\rho\tau - 2\lambda}{he}(\Lambda^{1/\tau} - 1) \sim F(he, \rho\tau - 2\lambda) \tag{A.4}$$

where

$$
\begin{aligned}
\lambda &= \frac{he - 2}{4} , \quad \rho = T - k - \frac{(e - h + 1)}{2} , \\
\tau &= \begin{cases} \left((h^2e^2 - 4)/(h^2 + e^2 - 5)\right)^{1/2} , & \text{if } h^2 + e^2 - 5 > 0 , \\ 1 , & \text{otherwise.} \end{cases}
\end{aligned}
$$

Further, the special case $h = 1$ leads to Hotelling's T^2 criterion which is a monotonic function of Λ. If $h > 2$ and $e > 2$, then the distributional result (A.4) holds asymptotically [Rao (1973, Chapter 8)]. Stewart (1997) provides an extensive discussion of these special F tests. Of course, these results are restricted to UL hypotheses of the form (2.8). However, beside this specific hypothesis class, the null distribution of the LR statistic is not nuisance-parameter-free.

A.2.   Invariance of lack-of-fit tests

Proof of Proposition **5.1**. On observing that $\hat{U} = MU$ and $U = WJ'$, we see that

$$
\begin{aligned}
\hat{U}(\hat{U}'\hat{U})^{-1}\hat{U}' &= MU(U'MU)^{-1}U'M \\
&= MU(J^{-1})'J'(U'MU)^{-1}JJ^{-1}U'M \\
&= MU(J^{-1})'\big[(J^{-1})U'MU(J^{-1})'\big]^{-1}J^{-1}U'M \\
&= MW\,(W'MW)^{-1}\,W'M\,.
\end{aligned}
$$

Since (2.8) entails that W has a known distribution, it follows that $\hat{U}(\hat{U}'\hat{U})^{-1}\hat{U}'$ (and consequently SK and KU) are completely determined by the distribution of W (given X). This is the same method of proof which led to Theorem **A.1**.

B.   Appendix: Monte Carlo tests

The Monte Carlo (MC) test procedure goes back to Dwass (1957) and Barnard (1963). Here we summarize the underlying methodology (given a right tailed test), as it applies to the test statistics we consider in this paper. A general discussion is available in Dufour and Khalaf (2001) and Dufour (2002).

B.1.   General method

Let us first consider the case of a pivotal statistic, *i.e.* the case where the statistic at hand, say $S(y, X)$ can be written as a pivotal function of W [defined in (2.8)], formally

$$
S(y, X) = \overline{S}(W, X)\,,
$$

where W is defined by (2.8), and the distribution of the rows of W is known. This is the case where the conditional distribution of $S(y, X)$, given X, is completely determined by the matrix X and the conditional distribution of W given X.

1. Let $S_{(0)}$ denote the observed test statistic.

2. By Monte Carlo methods, draw N *i.i.d.* replications of W :

$$
W_{(j)} = [W_1^{(j)}, \,\ldots\,, W_n^{(j)}], \quad j = 1, \,\ldots\,, N.
$$

3. From each simulated error matrix $W_{(j)}$, compute the statistics

$$S^j = \bar{S}(W_{(j)}, X), \quad j = 1, \ldots, N.$$

For instance, in the case of the QLR statistic underlying Theorem **A.1**, calculate $|W'_{(j)} M W_{(j)}| / |W'_{(j)} M_0 W_{(j)}|$, $j = 1, \ldots, N$.

4. Compute the MC p-value

$$\hat{p}_N(S_{(0)}) = \frac{N \hat{G}_N(S_{(0)}) + 1}{N + 1}, \tag{B.1}$$

where

$$\hat{G}_N(x) = \frac{1}{N} \sum_{j=1}^N I_{[0,\infty]}(S_i - x), \quad I_A(x) = \left\{ \begin{array}{ll} 1, & if \ x \in A, \\ 0, & if \ x \notin A. \end{array} \right.$$

In other words, $N\hat{G}_N(S_{(0)})$ is the number of simulated criteria $\geq S_{(0)}$ and $\hat{R}_N(S_{(0)}) = N - N\hat{G}_N(S_{(0)}) + 1$ gives the rank of $S_{(0)}$ in the series $S_{(0)}, S_1, \ldots, S_N$.

5. The MC critical region is

$$\hat{p}_N(S_{(0)}) \leq \alpha, \quad 0 < \alpha < 1. \tag{B.3}$$

If $\alpha(N + 1)$ is an integer,

$$P_{(H_0)} \left[\hat{p}_N(S_{(0)}) \leq \alpha \right] = \alpha. \tag{B.4}$$

The above algorithm is valid for any fully specified distribution of W. Consider now the case where the distribution of W involves a nuisance parameter as in (2.11). In this case, given ν, (B.1) yields a MC p-value which we will denote $\hat{p}_N(S_{(0)}|\nu)$ where the conditioning on ν is emphasized for further reference. The test defined by $\hat{p}_N(S_{(0)}|\nu) \leq \alpha$ is *exactly size correct* (in the sense of (B.4)) for known ν. Treating ν as a formal nuisance parameter, the test based on

$$\sup_{\nu \in \Phi_0} [\hat{p}_N(S_{(0)}|\nu)] \leq \alpha \tag{B.5}$$

where Φ_0 is a nuisance parameter set consistent with H_0, is *exact at level* α; see Dufour (2002). Note that no asymptotics on the number N of MC replications is required to obtain the latter result; this is the fundamental difference between the latter procedure and the (closely related) parametric bootstrap method, which in this context would correspond to test based on $\hat{p}_N(S_{(0)}|\hat{\nu}_0)$, where $\hat{\nu}_0$ is any *point* estimate of ν. In Dufour and Khalaf (2002*c*), we call the test based on simulations using a point nuisance parameter estimate a *local* MC (LMC) test. The term *local* reflects the fact that the underlying MC p-value is based on a specific choice for the nuisance parameter. Furthermore, we show that LMC non-rejections are *exactly* conclusive in the following sense: if $\hat{p}_N(S_{(0)}|\hat{\nu}_0) > \alpha$, then the exact test MMC test is clearly not significant at level α.

B.2. MC skewness and kurtosis tests

The algorithm for implementing the MC skewness and kurtosis tests can de decomposed in three wide steps.

B.2.1. Estimating expected skewness and kurtosis

A1. Draw N_0 *i.i.d.* replications, $\bar{W}_{(i)} = [\bar{W}_1^{(i)}, \ldots, \bar{W}_n^{(i)}]$, $i = 1, \ldots, N_0$, conforming with the hypothesized distribution with $\nu = \nu_0$.

A2. From each simulated error matrix $\bar{W}_{(i)}$, compute

$$\bar{M}\bar{W}_{(i)}\left[\bar{W}'_{(i)}\bar{M}\bar{W}_{(i)}\right]^{-1}\bar{W}'_{(i)}\bar{M}, \quad i = 1, \ldots, N_0.$$

These provide N_0 replications of SK and KU, applying (5.1) and (5.2), namely: $\overline{SK}_{(i)}$ and $\overline{KU}_{(i)}$, $i = 1, \ldots, N_0$.

A3. Calculate the average values:

$$\overline{SK}(\nu_0) = \sum_{i=1}^{N_0} \overline{SK}_{(i)}/N_0, \quad \overline{KU}(\nu_0) = \sum_{i=1}^{N_0} \overline{KU}_{(i)}/N_0.$$

Two questions arise at this stage: (i) how to obtain exact cut-off points for (5.4) and (5.5), and (ii) how to obtain a size-correct simultaneous test which combines (5.4) and (5.5). Let us first address the individual p-values issue, which may be run as in Appendix B.1 above.

B.2.2. Individual excess skewness and kurtosis tests

B1. Let $ESK_{(0)}$ and $EKU_{(0)}$ denote the observed test statistics.

B2. For a given number N_1 of replications, and independently from the simulation performed to obtain $\overline{SK}(\nu_0)$ and $\overline{KU}(\nu_0)$ (*i.e.* step A1 above), draw $W_{(j)} = [W_1^{(j)}, \ldots, W_n^{(j)}]$, $j = 1, \ldots, N_1$, conforming with (2.9).

B3. From each simulated error matrix $W_{(j)}$, compute

$$\bar{M}W_{(j)}\left[W'_{(j)}\bar{M}W_{(j)}\right]^{-1}W'_{(j)}\bar{M}, \quad j = 1, \ldots, N_1,$$

and, on applying (5.1) and (5.2), N_1 replications of SK and $KU : SK_{(j)}$ and $KU_{(j)}$, $j = 1, \ldots, N_1$.

B4. Conditioning on $\overline{SK}(\nu_0)$ and $\overline{KU}(\nu_0)$ [generated only once as in steps A1-A3], obtain, applying (5.4) and (5.5), N_1 replications of ESK and EKU :

$$ESK_{(j)} \text{ and } EKU_{(j)}, j = 1, \ldots, N_1.$$

B5. Obtain (respectively) the ranks of $ESK_{(0)}$ and $EKU_{(0)}$ in the series $\{ESK_{(0)}, ESK_{(1)}, \ldots, ESK_{(N_1)}\}$ and $\{EKU_{(0)}, EKU_{(1)}, \ldots, EKU_{(N_1)}\}$ respectively; applying (B.1), these yield the MC p-values $\hat{p}_{N_1}(ESK_{(0)}|\nu_0)$ and $\hat{p}_{N_1}(EKU_{(0)}|\nu_0)$.

B.2.3. Combined excess skewness and kurtosis test

C1. Using the function $\hat{p}_{N_1}(x|\nu_0)$ obtained from the simulations performed over the steps B.2.1 and B.2.2, compute the observed value of the combined test statistic:

$$CSK_{(0)} = 1 - \min\left\{\hat{p}_{N_1}(ESK_{(0)}|\nu_0),\ \hat{p}_{N_1}(EKU_{(0)}|\nu_0)\right\}.$$

C2. Generate N *i.i.d.* replications of SK and KU, independently of the values simulated over the steps B.2.1 and B.2.2, and compute the corresponding excess skewness and kurtosis statistics:

$$ESK_{1(l)},\ EKU_{1(l)}, \quad l = 1,\, 2,\, \ldots,\, N. \tag{B.6}$$

For each replication, the centering values $\overline{SK}(\nu_0)$ and $\overline{KU}(\nu_0)$ are identical to the ones generated in step B.2.1.

C3. Using the function $\hat{p}_{N_1}(x|\nu_0)$ already defined, compute the combined test statistics associated with the statistics generated in the previous step:

$$CSK_{(l)} = 1 - \min\left\{\hat{p}_{N_1}(ESK_{1(l)}|\nu_0),\ \hat{p}_{N_1}(EKU_{1(l)}|\nu_0)\right\}, \quad l = 1,\, 2,\, \ldots,\, N. \tag{B.7}$$

C4. Obtain the rank of the observed statistic $CSK_{(0)}$, within the simulated series, and derive the corresponding p-value, which we will denote $\hat{p}_N(CSK_{(0)}|\nu_0)$.

The null hypothesis is rejected if $\hat{p}_N(CSK_{(0)}|\nu_0) \leq \alpha$. This test has level α because the variables $CSK_{(l)}$, $l = 0,\, 1,\, \ldots,\, N$, are exchangeable under the null hypothesis. In the applications presented in this paper, we used the values $N_0 = N_1 = N$.

References

Affleck-Graves, J. and McDonald, B. (1989), 'Nonnormalities and tests of asset pricing theories', *Journal of Finance* **44**, 889–908.

Allingham, M. (1991), 'Existence theorems in the capital asset pricing model', *Econometrica* **59**, 1169–1174.

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, second edn, John Wiley & Sons, New York.

Baillie, R. and Bollerslev, T. (1998), 'Time to maturity in the basis of stock market indices: Evidence from the S&P 500 and the MMI', *Journal of Empirical Finance* **7**, 297–305.

Barnard, G. A. (1963), 'Comment on 'The spectral analysis of point processes' by M. S. Bartlett', *Journal of the Royal Statistical Society, Series B* **25**, 294.

Beaulieu, M.-C. (1998), 'Time to maturity in the basis of stock market indices: Evidence from the S&P 500 and the MMI', *Journal of Empirical Finance* **5**, 177–195.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2001), Testing Black's CAPM with possibly non-Gaussian error distributions: An exact simulation-based approach, Technical report, Département d'économique, Université Laval, and CRDE, Université de Montréal.

Berk, J. B. (1997), 'Necessary conditions for the CAPM', *Journal of Economic Theory* **73**, 245–257.

Berndt, E. R. and Savin, N. E. (1977), 'Conflict among criteria for testing hypotheses in the multivariate linear regression model', *Econometrica* **45**, 1263–1277.

Breeden, D. T., Gibbons, M. and Litzenberger, R. H. (1989), 'Empirical tests of the consumption based CAPM', *Journal of Finance* **44**, 231–262.

Campbell, Y. Y., Lo, A. W. and 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.

Chmielevsky, M. A. (1981), 'Elliptically symmetric distributions: A review and bibliography', *International Statistical Review* **49**, 75–93.

Dufour, J.-M. (1990), 'Exact tests and confidence sets in linear regressions with autocorrelated errors', *Econometrica* **58**, 475–494.

Dufour, J.-M. (1997), 'Some impossibility theorems in econometrics, with applications to structural and dynamic models', *Econometrica* **65**, 1365–1389.

Dufour, J.-M. (2002), 'Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics', *Journal of Econometrics* **forthcoming**.

Dufour, J.-M., Farhat, A., Gardiol, L. and Khalaf, L. (1998), 'Simulation-based finite sample normality tests in linear regressions', *The Econometrics Journal* **1**, 154–173.

Dufour, J.-M. and Khalaf, L. (2001), Monte Carlo test methods in econometrics, *in* B. Baltagi, ed., 'Companion to Theoretical Econometrics', Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, U.K., chapter 23, pp. 494–519.

Dufour, J.-M. and Khalaf, L. (2002*a*), Exact simulation based muliple hypothesis tests, Technical report, CIRANO and CIREQ, Université de Montréal, and Département d'économique, Université Laval.

Dufour, J.-M. and Khalaf, L. (2002*b*), 'Exact tests for contemporaneous correlation of disturbances in seemingly unrelated regressions', *Journal of Econometrics* **106**(1), 143–170.

Dufour, J.-M. and Khalaf, L. (2002*c*), Finite sample tests in seemingly unrelated regressions, *in* D. E. A. Giles, ed., 'Computer-Aided Econometrics', Marcel Dekker, New York. Forthcoming.

Dufour, J.-M. and Khalaf, L. (2002*d*), 'Simulation based finite and large sample tests in multivariate regressions', *Journal of Econometrics* **forthcoming**.

Dufour, J.-M., Khalaf, L. and Beaulieu, M.-C. (2001), Exact invariant tests for multivariate normality, Technical report, CIRANO and CIREQ, Université de Montréal, and Département d'économique, Université Laval.

Dufour, J.-M., Khalaf, L., Bernard, J.-T. and Genest, I. (2001), Simulation-based finite-sample tests for heteroskedasticity and ARCH effects, Technical report, C.R.D.E., Université de Montréal, and GREEN, Université Laval. 48 pages.

Dufour, J.-M. and Kiviet, J. F. (1996), 'Exact tests for structural change in first-order dynamic models', *Journal of Econometrics* **70**, 39–68.

Dufour, J.-M. and Roy, R. (1985), 'Some exact results on sample autocorrelations and tests of randomness', *Journal of Econometrics* **29**, 257–273. Corrigendum 41(1989), 279-281.

Dufour, J.-M. and Torrès, O. (1998), Union-intersection and sample-split methods in econometrics with applications to SURE and MA models, *in* D. E. A. Giles and A. Ullah, eds, 'Handbook of Applied Economic Statistics', Marcel Dekker, New York, pp. 465–505.

Dwass, M. (1957), 'Modified randomization tests for nonparametric hypotheses', *Annals of Mathematical Statistics* **28**, 181–187.

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', *Econometrica* **50**, 987–1007.

Fama, E. F. (1965), 'The behaviour of stock prices', *Journal of Business* **60**, 401–424.

Fama, E. F. and French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**, 3–56.

Fiorentini, G., Sentana, E. and Calzolari, G. (2000), The score of conditionally heteroskedastic dynamic regression models with Student t innovations, and an LM test for multivariate normality, Technical report, University of Alicante, CEMFI, and University of Florence.

Gibbons, M. R. (1982), 'Multivariate tests of financial models: A new approach', *Journal of Financial Economics* **10**, 3–27.

Gibbons, M. R., Ross, S. A. and Shanken, J. (1989), 'A test of the efficiency of a given portfolio', *Econometrica* **57**, 1121–1152.

Gibbons, M. R. and Shanken, J. (1987), 'Subperiod aggregation and the power of multivariate tests of portfolio efficiency', *Journal of Financial Economics* **19**, 389–394.

Groenwold, N. and Fraser, P. (2001), 'Tests of asset-pricing models: How important is the iid-normal assumption', *Journal of Empirical Finance* **8**, 427–449.

Ingersoll, J. (1987), *Theory of Financial Decision Making*, Rowman & Littlefield, NJ.

Jobson, J. D. and Korkie, B. (1989), 'A performance interpretation of multivariate tests of asset set intersection, spanning, and mean-variance efficiency', *Journal of Financial and Quantitative Analysis* **24**, 185–204.

Jobson, J. and Korkie, B. (1982), 'Potential performance and tests of portfolio efficiency', *Journal of Financial Economics* **10**, 433–466.

Jöckel, K.-H. (1986), 'Finite sample properties and asymptotic efficiency of Monte Carlo tests', *The Annals of Statistics* **14**, 336–347.

Kan, R. and Zhou, G. (2001), Tests of mean-variance spanning, Technical report, Rotman School of Management, University of Toronto, Toronto, Canada.

Kariya, T. (1985), *Testing in the Multivariate General Linear Model*, number 22 *in* 'Economic Research Series, The Institute of Economic Research, Hitotsubashi University, Japan', Kinokuniya Company Ltd., Tokyo.

Kariya, T. and Kim, P. (1997), Finite sample robustness of tests: An overview, *in* G. S. Maddala and C. R. Rao, eds, 'Handbook of Statistics 15: Robust Inference', North-Holland, Amsterdam, chapter 22, pp. 645–660.

Kelker, D. (1970), 'Distribution theory of spherical distributions and a location-scale parameter generalization', *Sankhyā, Series A, Indian Journal of Statistics* **32**, 419–430.

Lee, J. H. H. (1991), 'A Lagrange multiplier test for GARCH models', *Economics Letters* **37**, 265–271.

Lee, J. H. and King, M. L. (1993), 'A locally most mean powerful based score test for arch and garch regression disturbances', *Journal of Business and Economic Statistics* **11**, 17–27. Correction 12 (1994), 139.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses, 2nd edition*, John Wiley & Sons, New York.

Lo, A. and MacKinlay, C. (1988), 'Stock prices do not follow random walks: Evidence from a simple specification test', *Review of Financial Studies* **1**, 41–66.

Lo, A. and MacKinlay, C. (1989), 'The size and power of the variance ratio test in finite samples: A Monte Carlo investigation', *Journal of Econometrics* **40**, 203–238.

MacKinlay, A. C. (1987), 'On multivariate tests of the Capital Asset Pricing Model', *Journal of Financial Economics* **18**, 341–372.

MacKinlay, A. C. and Richardson, M. P. (1991), 'Using generalized method of moments to test mean-variance efficiency', *The Journal of Finance* **46**, 511–527.

Mardia, K. V. (1970), 'Measures of multivariate skewness and kurtosis with applications', *Biometrika* **57**, 519–530.

Nielsen, L. T. (1990), 'Existence of equilibrium in CAPM', *Journal of Economic Theory* **52**, 223–231.

Owen, J. and Rabinovitch, R. (1983), 'On the class of elliptical distributions and their applications to the theory of portfolio choice', *Journal of Finance* **XXXVIII**(3), 745–752.

Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, second edn, John Wiley & Sons, New York.

Richardson, M. and Smith, T. (1993), 'A test for multivariate normality in stock returns', *Journal of Business* **66**, 295–321.

Ross, S. A. (1978), 'Mutual fund separation in financial theory - The separating distributions', *Journal of Economic Theory* **17**, 254–286.

Samuelson, P. (1967), 'On multivariate tests of the Capital Asset Pricing Model', *Journal of Financial and Quantitative Analysis* **2**, 107–122.

Shanken, J. (1986), 'Testing portfolio efficiency when the zero-beta rate is unknown: A note', *Journal of Finance* **41**, 269–276.

Shanken, J. (1990), 'Intertemporal asset pricing: An empirical investigation', *Journal of Econometrics* **45**, 99–120.

Shanken, J. (1996), Statistical methods in tests of portfolio efficiency: A synthesis, *in* G. S. Maddala and C. R. Rao, eds, 'Handbook of Statistics 14: Statistical Methods in Finance', North-Holland, Amsterdam, pp. 693–711.

Staiger, D. and Stock, J. H. (1997), 'Instrumental variables regression with weak instruments', *Econometrica* **65**, 557–586.

Stewart, K. G. (1995), 'The functional equivalence of the W, LR and LM statistics', *Economics Letters* **49**, 109–112.

Stewart, K. G. (1997), 'Exact testing in multivariate regression', *Econometric Reviews* **16**, 321–352.

Velu, R. and Zhou, G. (1999), 'Testing multi-beta asset pricing models', *Journal of Empirical Finance* **6**, 219–241.

Wang, J. and Zivot, E. (1998), 'Inference on structural parameters in instrumental variables regression with weak instruments', *Econometrica* **66**, 1389–1404.

Wilks, S. S. (1932), 'Certain generalizations in the analysis of variance', *Biometrika* **24**, 471–494.

Zhou, G. (1991), 'Small sample tests of portfolio efficiency', *Journal of Financial Economics* **30**, 165–191.

Zhou, G. (1993), 'Asset-pricing tests under alternative distributions', *The Journal of Finance* **48**, 1927–1942.

Zhou, G. (1995), 'Small sample rank tests with applications to asset pricing', *Journal of Empirical Finance* **2**, 71–93.