Testing the CAPM with possibly non-Gaussian error distributions: an exact simulation-based approach *

Marie-Claude Beaulieu † Jean-Marie Dufour ‡ Lynda Khalaf §

March 2002
Compiled April 2, 2002, 10:40am

*This work was supported by the Institut de Finance Mathématique de Montréal, the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds FCAR (Government of Québec). This paper was also partly written at the Centre de recherche en Économie et Statistique (INSEE, Paris) and the Technische Universität Dresden (Fakultät Wirtschaftswissenschaften). The authors thank Christian Gouriéroux and Raymond Kan for several useful comments as well as seminar participants at the 2000 EC^2 meetings, CREST (Paris), the University of British Columbia, the University of Toronto, the 2001 Canadian Economic Association meetings, and CIRANO.

†CRÉFA, 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. email: Marie-Claude.Beaulieu@fas.ulaval.ca

‡ Canada Research Chair Holder (Econometrics). Centre de recherche et développement en économique (C.R.D.E.), Centre interuniversitaire de recherche en analyse des organisations (CIRANO), 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: 1 514 343 2400; FAX: 1 514 343 5831; e-mail: jean.marie.dufour@umontreal.ca. Web page: http://www.fas.umontreal.ca/SCECO/Dufour .

§GREEN, Université Laval, Pavillon J.-A. De Sève, St. Foy, Québec, Canada, G1K 7P4. TEL: (418) 656 2131-2409; FAX: (418) 656 7412. email: lynda.khalaf@ecn.ulaval.ca

ABSTRACT

In this paper we propose exact likelihood-based mean-variance efficiency tests of the market portfolio, allowing for a wide class of error distributions which include normality as a special case. These tests are developed using 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 Capital Asset Pricing Model [CAPM] tests. In non-Gaussian contexts, we re-consider mean-variance efficiency tests allowing for multivariate Student-*t* and normal mixtures errors. Our framework allows to cast more evidence on whether the normality assumption is too restrictive when testing the CAPM. We provide 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) normality is rejected in most subperiods, (ii) residual checks reveal no significant departures from the *i.i.d.* assumption, and (iii) the CAPM is not rejected as frequently once it is allowed for the possibility of non-normal errors.

Contents

List of Definitions, Propositions and Theorems

List of Tables

1. Introduction

The capital asset pricing model (CAPM) is by far the most commonly used mean-variance model in asset-pricing practice. Empirical tests of the CAPM conducted within the Multivariate Linear Regression (MLR) framework may be traced back to Gibbons (1982). The associated empirical literature which has evolved since Gibbons' seminal work is enormous; recent references may be found in Campbell, Lo and MacKinlay (1997) and Shanken (1996). The purpose of this paper is to propose exact finite-sample tests for possibly non-Gaussian versions of the CAPM.

In this context, applying exact tests is important because test results that are only approximate and/or do not take into account the non-normality of asset returns can cause spurious empirical interpretations of the CAPM. These problems are due to serious discrepancies between asymptotic and finite-sample distributions or specification errors related to the hypothesized distribution of fundamentals. Indeed, a number of studies from the econometrics and finance literature have revealed that: (i) standard asymptotic theory provides a poor approximation to the distribution of MLR-based tests [see Dufour and Khalaf (2002*b*) and the references cited therein], and (ii) as a result, the conclusions of financial MLR-based empirical studies can be strongly affected, even if sample sizes encountered in finance are typically large [see Campbell et al. (1997, Chapter 5) and Shanken (1996, Section 3.4.2))]. These difficulties find their origin in the fact that the relevant null distributions typically depend on unknown (nuisance) parameters - *e.g.* the error covariance matrix - whose number increases rapidly with the dimension of the system.

Even though some exact tests have been proposed for a few specific hypotheses, all available provably exact distributional results use the normality assumption, which is consistent with the CAPM. Well known procedures include the exact market portfolio mean-variance efficiency test proposed by MacKinlay (1987) and by Gibbons, Ross and Shanken (1989, henceforth GRS) for the observable risk-free rate case. Specifically, to test the joint significance of the MLR-CAPM intercepts, these authors use Hotelling-T^2 statistic, which may be transformed into an F-distributed statistic. See also Stewart (1997) for more recent work on exact F tests in finance.

Nonetheless, it has long been recognized that financial returns do not exhibit normality [see Fama (1965)]. Theoretical work on distributional assumptions consistent with the CAPM show that elliptical distributions could be used to derive and test the model [see Ingersoll (1987))]. Zhou (1993) reconsidered the GRS problem under elliptical distributions and provided simulation-based test procedures which exploit exact invariance results for these distributions. Although nuisance parameters are not completely accounted for by Zhou (1993), to the best of our knowledge, no other non-asymptotic result is available which does not impose Gaussianity.

In this context, an important research question is whether imposing normality of stock market returns is a restrictive assumption to test the CAPM. For instance, different authors have studied the properties of GRS's test using the empirical distributions of asset returns. Affleck-Graves and McDonald (1989) present simulation evidence which indicates that the multivariate tests are robust to deviations from disturbance normality, except in periods of exceptional market variability. MacKinlay and Richardson (1991) report a sensitivity to conditional heteroskedasticity. On the other hand, Zhou (1993) finds no important differences between test decisions using normal and elliptical distributions. A recent study by Groenwold and Fraser (2001) using Australian data reports similar results. It is evident that a decisive answer to this research question may not be reached, in the absence of a provably exact procedure in possibly non-Gaussian contexts. To present and apply such a procedure is one of our objectives in this paper.

In a different vein, we have recently proposed several general exact test procedures for MLR models [see Dufour and Khalaf (2002*b*)]. In particular, we considered the Wilks statistic, which is defined as the ratio of the determinants associated with the constrained and unconstrained sum of squared error matrices. This statistic is a monotonic transformation of the commonly used Gaussian quasi-likelihood-ratio [QLR]. For a specific class of hypotheses on the regression coefficients which take the Uniform Linear (UL) form, we proposed an exact simulation-based Wilks test without the normality assumption.[1] To do this, we showed, given a wide class of error distributions which include normality as a special case, that the distribution of the Wilks statistic under UL hypotheses does not depend on any unknown parameter including the error covariance. In this paper, we extend these results to single and multi-*beta* portfolio mean-variance efficiency and CAPM tests, when the risk free interest rate is observable. We also consider further inference problems not studied by Dufour and Khalaf (2002*b*), including exact distributional lack-of-fit tests, and further exact specification tests (e.g. tests for multivariate GARCH).

On these issues, this paper makes five main contributions. First, we extend exact portfolio efficiency tests beyond the Gaussian model. Our statistical methodology holds given a general distributional hypothesis which includes the elliptical family as a special case. Primarily, we focus on the multivariate Student-t distribution to conform with standard and recent mean-variance efficiency theoretical setups. We next consider multivariate mixtures of normals which allow modelling more extreme kurtosis.

Second, we propose a formal method to deal with unknown distributional parameters, namely: the degrees-of-freedom in the case of the multivariate Student-t distribution, and the probability-of-mixing and ratio-of-scale parameters for the mixtures of normals. To do this, we propose an exact confidence set for the parameters in question, which is then used to obtain an exact Monte Carlo [MC] test [see Dufour (2000)]. The MC test procedure yields an exact simulation-based p-value whenever the parameters which intervene in the null distribution of the test statistic are known. 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. In the presence of unknown intervening parameters, it is shown in Dufour (2000) that if decision is based on the largest simulated MC p-value for all nuisance parameter consistent with the null hypothesis, then the associated MC test (called a “maximized MC [MMC] test”) controls the level of the test for any sample size and any number of MC replications.[2] Here we apply a two-stage constrained consistent set MMC, as in Dufour and Kiviet (1996), to obtain the largest MC p-value over a consistent nuisance parameter set estimator. We propose to obtain the latter set estimate

[1] 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).

[2] In nuisance parameter dependent problems, a test is α-level *exact* if the largest rejection probability over the nuisance parameter space consistent with the null hypothesis is $\leq \alpha$. The MMC test is thus exact by construction; a formal proof of exactness is provided by Dufour (2000). For applications on the MMC test in econometrics, see e.g. Dufour and Khalaf (2001*a*) and Dufour and Khalaf (2001*b*).

inverting a distributional goodness of fit (GF) test. 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 apply exact multivariate GF tests [see Dufour, Khalaf and Beaulieu (2001)]. As explained in Richardson and Smith (1993) who considered tests for departures from Gaussianity, it is crucial in MLR-based financial models to consider multivariate tests of asset returns which explicitly take the error covariance into consideration. We propose an exact moments-based GF test of the hypothesized error distributions (the multivariate normal; the Student-t error and the mixture of normals, with possibly unknown parameters).[3] 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 application, 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 residual-based diagnostic tests for departures from the maintained *i.i.d.* hypothesis [see Dufour, Khalaf and Beaulieu (2001)]. As with normality tests, concerns over cross-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.[4] 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 *standardized* OLS residuals to ensure invariance to the error-covariance [see Dufour, Khalaf and Beaulieu (2001)]. Second, we combine tests across equations using an exact simulation-based procedure which does not call for Bonferroni treatment [see Dufour and Khalaf (2002*a*)]. As is well known, and emphasized in Shanken (1990), Bonferroni-based combined tests require one to divide the test's significance level overall individual tests. 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), hence large power reduction. Following the exact strategy we applied to test the significance of the CAPM intercepts, we also impose multivariate Student-t errors and multivariate mixtures of normals with possibly unknown parameters; this allows 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 consider are multivariate generalization of the popular Variance Ratio tests. Using the same residuals standardization and the same simulation-based combination strategy we proposed for the GARCH test, we

[3] 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. Our exact test for multivariate normality will allow to conclusively solve such controversies.

[4] 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).

show how these very useful tests can be applied exactly, in a multivariate perspective. We emphasize that the usefulness of these new tests extends beyond our specific applications.

Finally, 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 Gaussian based non-rejections may formally be treated as conclusive from our viewpoint (recall, of course, that we formally combine GF with efficiency tests here).

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 nuisance parameter dependent error distributions. 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 tests.

2. Framework

The fundamental finance problem we focus on 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, $t = 1, \ldots, T$, and $\widetilde{R}_{\mathrm{M}t}$ the returns on the market portfolio under consideration. The MLR statistical model underlying the well known CAPM test [Gibbons et al. (1989)] takes the following form:

$$r_{it} = a_i + b_i \widetilde{r}_{\mathrm{M}t} + u_{it},\ t = 1, ..., T,\ i = 1, ..., n, \tag{2.1}$$

where $r_{it} = R_{it} - R_t^F$, $\widetilde{r}_{\mathrm{M}t} = \widetilde{R}_{\mathrm{M}t} - R_t^F$, R^F is the riskless rate of return, and u_{it} is a random disturbance. In this context, the testable implications of the CAPM on the coefficients of this model are:

$$H_{CAPM} : a_i = 0,\ i = 1, \ldots, n, \tag{2.2}$$

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

The above model is a special case of the following MLR

$$Y = XB + U \tag{2.3}$$

where $Y = [Y_1, \ldots, Y_n]$ is $T \times n$, X is $T \times k$ with rank k and is assumed fixed, and $U = [U_1, \ldots, U_n] = [V_1, \ldots, V_T]'$ is an $T \times n$ matrix of error terms. In Appendix A.1, we summarize general exact results from Dufour and Khalaf (2002*b*) regarding MLR-based hypotheses tests. The hypotheses considered take the *uniform linear* (UL) form

$$H_0 : HBE = D \tag{2.4}$$

where H is a $h \times k$ matrix of rank h and E is a $n \times e$ matrix of rank e. This is relevant because

H_{CAPM} (2.2) is a special case of the latter. Indeed, rewriting (2.1) as in (2.3) with

$$
\begin{aligned}
Y &= [r_1,\ \ldots,\ r_n]\,,\ X = [\iota_T,\ \tilde{r}_{\mathrm{M}}]\,, \\
r_i &= (r_{1i},\ \ldots,\ r_{Ti})'\,,\ \tilde{r}_{\mathrm{M}} = (\tilde{r}_{1\mathrm{M}},\ \ ,\ \tilde{r}_{T\mathrm{M}})'\,,
\end{aligned}
$$

it is easily seen that (2.2) yields $(1,0)B = 0$, which corresponds to (2.4) where $E = I_n$, $D = 0$ and H is the row vector $(1,0)$. For further discussion of the MLR model, the reader may consult Dufour and Khalaf (2002*b*), Rao (1973, chapter 8), Anderson (1984, chapters 8 and 13), Kariya (1985) and Berndt and Savin (1977).

As is well known, the CAPM imposes further restrictions on the error distributions. In particular, the standard assumption consists in assuming that $V_1,\ \ldots\ ,\ V_T$ are *i.i.d.* multivariate normal. In this paper, we consider the more general case where

$$V_t = JW_t\ ,\ t = 1,\ \ldots\ ,\ T\,, \tag{2.5}$$

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*. At this stage, we note that the Gaussian and elliptical distributions underlying the standard CAPM are also consistent with (2.5). In our context, we therefore focus on multivariate t-distributions and mixtures of normals, which we denote $\mathcal{F}_1(W)$ and $\mathcal{F}_2(W)$ respectively, and define as follows:

$$W \sim \mathcal{F}_1(W;\kappa) \Leftrightarrow W_t = Z_{1t}/(Z_{2t}/\kappa)^{1/2}\,, \tag{2.6}$$

where Z_{1t} is multivariate normal $(0, I_n)$ and Z_{2t} is a $\chi^2(\kappa)$ variate independent from Z_{1t};

$$W \sim \mathcal{F}_2(W;\pi,\omega) \Leftrightarrow W_t = \pi Z_{1t} + (1-\pi)Z_{3t}, \tag{2.7}$$

where Z_{3t} is multivariate normal $(0, \omega I_n)$ and is independent from 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 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 [Ingersoll (1987)]. For further reference, we shall use the following notation:

$$W \sim \mathcal{F}_i(W;\nu),\ i = 1, 2 \tag{2.8}$$

where

$$
\begin{aligned}
\nu\ &= \kappa, &&if\quad W_t\ \text{satisfies (2.6)},\\
&= (\pi,\omega), &&if\quad W_t\ \text{satisfies (2.7)}.
\end{aligned}
$$

3. CAPM tests with a known normalized disturbance distribution

In this section, we study the case where the nuisance parameter ν is known. Extensions to unknown ν are presented in Section 4 below. 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, which is denoted Σ, is $JJ'$and is invertible.

One of the most commonly used statistics to test H_{CAPM} in (2.2) [indeed, to test any UL hypothesis] is the Gaussian quasi maximum likelihood (QMLE) based criterion:

$$LR = T\ln(\Lambda), \;\; \Lambda = |\widehat{\Sigma}_{CAPM}|/|\widehat{\Sigma}| \tag{3.9}$$

where $\widehat{\Sigma} = \widehat{U}'\widehat{U}/T$, $\widehat{U} = Y - X\widehat{B}$, $\widehat{B} = (X'X)^{-1}X'Y$ and $\widehat{\Sigma}_{CAPM}$ is the Gaussian QMLE under H_{CAPM}. In Theorem **A.1** of Appendix A.1, we provide the exact null distribution of the latter statistic under (2.3) and (2.5) and the general UL hypothesis (2.4). Two results regarding this distribution are worth noting.

First, under (2.5), 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 [Dufour (2000)] 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 (2.5) is presented in Appendix B.1 and may be summarized as follows given (2.8). Conditional on υ, generate, imposing (2.8), N i.i.d. draws from the distribution of $W_1, \ldots, W_T$; these yield 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.

Second, results specific to the Gaussian special case of (2.5) lead to the F-tests used by GRS. Formally, if $\min(h, e) \leq 2$ where $h = \mathrm{rank}\,(H)$ and $e = \mathrm{rank}(E)$, 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_{CAPM} corresponds to $\min(h, e) = 1$.

For the CAPM problem, Theorem **A.1** allows one to characterize the null distribution of LR for all error distributions which satisfy (2.5), as follows.

Theorem 3.1 DISTRIBUTION OF THE QUASI-LR CAPM TEST STATISTIC. *Under* (2.1), (2.2) *and* (2.5), *the LR statistic defined by* (3.9) *is distributed like*

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

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* $W = [W_1, \ldots, W_T]'$ *is defined by* (2.5).

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 **A.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_{CAPM} (2.2) in the context of

$$r_{it} = a_i + \sum_{j=1}^{s} b_{ji}\widetilde{r}_{jt} + u_{it},\ t = 1, ..., T,\ i = 1, ..., n, \tag{3.10}$$

where $\widetilde{r}_{jt} = \widetilde{R}_{jt} - R_t^F$ and $\widetilde{R}_{jt},\ j = 1,\ \dots\ , s$ are returns on s benchmark portfolios. In this case, the null distribution of the statistic defined by (3.9) obtains as in Theorem **3.1** where $X = [\iota_T,\ \widetilde{r}_1,\ \dots\ ,\ \widetilde{r}_s]$, $\widetilde{r}_j = (\widetilde{r}_{1j},\ \dots\ ,\ \widetilde{r}_{Tj})'$, and H is the $(s+1)$ dimensional row vector $(1, 0,\ \dots\ , 0)$.

It is of interest to relate Theorem **3.1** to the available non-asymptotic CAPM tests, i.e.: (i) the GRS test, and (ii) the simulation-based test proposed by Zhou (1993). As formally stated in Appendix A.1, Theorem **3.1** and (A.4) entail, when errors are Gaussian, that

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

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

$$Q = \frac{T\widehat{a}'\left(\frac{T}{T-k}\widehat{\Sigma}\right)^{-1}\widehat{a}}{1 + \overline{r}'\widehat{\Delta}^{-1}\overline{r}} \tag{3.11}$$

where $\widehat{a}$ is the vector of intercept OLS estimates, $\frac{T}{T-k}\widehat{\Sigma}$ is the OLS-based unbiased estimator of Σ, $\overline{r}$ and $\widehat{\Delta}$ include respectively the time-series-means and sample covariance matrix corresponding to the right-hand-side portfolio returns. Under (2.2), Q follows the *Hotelling* $T^2(n,\ T - s - 1)$ distribution or equivalently,

$$\frac{Q(T - s - n)}{n(T - s - 1)} \sim F(n, T - s - n) \tag{3.12}$$

where

$$\Lambda - 1 = \frac{Q}{T - s - 1}\ ; \tag{3.13}$$

see also Stewart (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 non-Gaussian CAPM tests is due to Zhou (1993): the GRS problem is reconsidered in models with elliptical distributions, and the multivariate Student t and mixtures of normal distributions are included as examples. In this context, Zhou demonstrates exact location/scale invariance of the GRS-type efficiency test statistic and exploits this property to derive simulation based p-values; see also Zhou (1991, Section 3). In both articles, the LR statistic is expressed in terms of the roots of a determinantal equation which only depends on the constrained and unconstrained residual cross-products. From there on, location-scale invariance is proved without the normal assumption. This is the same approach underlying Theorem **A.1** [see Dufour and Khalaf (2002*b*)], with the exception that the latter result is not restricted to asset pricing tests. The "*Monte Carlo integration technique*" proposed by Zhou to obtain non-asymptotic p-values is highly related to our MC tests.

Our approach here first formalizes Zhou's test strategy, which is presented rather heuristically, as an exact procedure, [i.e. there is no formal proof for exactness, for any sample size and any number of MC replications]. We show formally how the test size is exactly controlled when no unknown parameter appears in the error distribution. For this hypothesis, our statistical framework includes Zhou's one as a special case. This serves to justify some of the results less explicated by Zhou, such as: ellipticity is not the main reason why exact tests obtain; location-scale invariance is not restricted to elliptical distributions (although, of course, the latter are important for the theory underlying the CAPM). Furthermore, and perhaps more importantly, our results generalize beyond H_{CAPM} as is shown in Appendix A.1. Finally, it is important to emphasize that the nuisance parameter problem is not completely dealt with by Zhou (1993). Indeed, if errors follow (2.8), Theorem 4.2 just ensures location-scale invariance (i.e., invariance to B and Σ), which yields pivotality (thus exactness) for known ν. It is thus evident that an unknown ν would intervene in the null distribution of the CAPM test statistic. We consider this case in the next section. Our empirical results which differ from those Zhou (1993) (refer to section 6) illustrate the importance of formally accounting for the unknown ν case.

4. CAPM tests with an incompletely specified error distribution

In this section, we extend the above results to the unknown distributional parameter case for the error families of interest, namely (2.8). At this stage, two points deserve notice. First, for ν given, the distributional hypothesis underlying (2.8) satisfies (2.5). Thus the MC p-values associated with Theorem **3.1** are exact for a given ν. Secondly, whether ν is viewed (from an empirical perspective) as a parameter of interest or a nuisance parameter, it is important, for the precision of the tests, to devise a decision rule which takes this parameter explicitly 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 test approach proposed by 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 values of ν in the latter confidence set. We will refer to the latter test as a maximized MC [MMC] test. It is important to note that if an overall α-level test 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 e.g. (2.3)), the maximized MC algorithm proceeds as follows. $\forall \nu \in \mathcal{C}(Y)$, and applying Theorem **3.1** and the MC algorithm in Appendix B.1, one obtains the MC p-values

$\widehat{p}_N(\Lambda_0|\nu)$ [see (B.1), in Appendix B.1]. Let

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

then the critical region

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

has exactly 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 most certainly significant.

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, $\mathcal{C}(Y)$ may be unbounded. For proofs and further references, see Dufour (1990), Dufour and Kiviet (1996) and Dufour (2000).

4.2. Confidence set for error distribution parameters

In this section, we 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 (2.8) where $\nu = \nu_0$ for known ν_0.

Since a procedure to derive an exact test for the distributions (2.8) is not available, we provide a moment-based one in Section 5.1. Our proposed set estimate for ν is however not specific to the latter test, so we present it in terms of any α_1 level test of (2.8) based on a given criterion denoted $\mathcal{T}(Y)$ where Y refers to the return data (as in e.g. (2.3)). Inverting $\mathcal{T}(Y)$ formally implies the following. Let $\mathcal{T}_0(Y)$ denote the value of the statistic computed from the observed sample. Obtain the p-value $\widehat{p}(\mathcal{T}_0(Y)|\nu_0)$ conforming with (2.8). For the moments-based test we propose below, this is achieved on applying MC test techniques. The confidence set for ν corresponds to the values of ν_0 for which $\widehat{p}(\mathcal{T}_0(Y)|\nu_0) > \alpha_1$.

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 (2.8), if cut-off points are appropriately "approximated", e.g. by simulation, imposing (2.8) errors. 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 application, 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 focus on multivariate specification tests including distributional lack-of-fit tests, and checks for departures from the *i.i.d.* errors hypothesis. We present in turn, exact multivariate GF tests, tests for multivariate GARCH effects and multivariate variance ratio (VR) tests. The proposed tests are formally valid for any parametric null hypothesis which takes the general form (2.5). Conforming with our empirical model, we focus on (2.8) with possibly unknown parameters.

5.1. Goodness-of-fit tests

The null hypothesis of concern here is (2.8) with unknown ν. We will solve the unknown ν problem by applying a MMC strategy, as follows. We propose a GF criterion which is pivotal if ν is known, which 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 to (jointly) assess whether e.g. Student distributions with different degrees of freedom are empirically relevant; this leads to a formal estimate for the Student distributions which best fit the data. The argument holds for multivariate mixtures.

The GF test statistics we propose use the same multivariate skewness and kurtosis criteria considered by Zhou (1993):

$$\mathrm{SK} = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{i=1}^{T}\widehat{d}_{ii}^{3}\,, \tag{5.1}$$

$$\mathrm{KU} = \frac{1}{T}\sum_{t=1}^{T}\widehat{d}_{tt}^{4}\,, \tag{5.2}$$

where $\widehat{d}_{it}$ are the elements of the matrix $\widehat{D} = \widehat{U}(\widehat{U}'\widehat{U})^{-1}\widehat{U}'$. The latter statistics were introduced by Mardia (1970) to assess deviations from multivariate normality, in models where the regressor matrix reduces to a vector of ones. Zhou (1993) observes that cut-off points consistent with (2.8) [with known ν] can be obtained for these criteria, using the same Monte Carlo integration technique proposed for the mean-variance efficiency test of the market portfolio. A note explains that the ensuing procedure is not *strictly exact* because it is based on residuals. We next show that this problem can be solved easily by recognizing the pivotal character of $\widehat{D}$ and applying the MC test methodology.

Proposition 5.1 DISTRIBUTION OF THE MULTIVARIATE SKEWNESS AND KURTOSIS TEST STATISTICS. *Under* (2.3), *and for all error distributions compatible with* (2.5), *the multivariate skewness and kurtosis criteria* (5.1) *and* (5.2) *are distributed, respectively, like* $\frac{1}{T^2}\sum_{t=1}^{T}\sum_{i=1}^{T} d_{it}^{3}$ *and* $\frac{1}{T}\sum_{t=1}^{T} d_{tt}^{4}$, *where* d_{it} *are the elements of the matrix* $MW\left(W'MW\right)^{-1}W'M$, $M = I - X(X'X)^{-1}X'$, W *is defined by* (2.5).

The proof is given in Appendix B.2. In the literature on multivariate normality tests, this property is recognized (under Gaussianity) 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 our testing strategy is valid for all distributions compatible with (2.5), including the normal.

The fact remains that the skewness and kurtosis tests are designed to estimate the moments of the multivariate normal distribution. The above invariance property does not strictly ensure their usefulness as GF tests beside the normal errors case. We thus propose the following modification: we consider alternative measures of skewness and kurtosis in excess of expected values consistent with (2.8). 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:

$$\mathrm{ESK}(\nu_0) = \left|\mathrm{SK} - \overline{\mathrm{SK}}(\nu_0)\right|, \tag{5.3}$$

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

where $\overline{\mathrm{SK}}(\nu_0)$ and $\overline{\mathrm{KU}}(\nu_0)$ are simulation-based estimates of the expected SK and KU given (2.8). These may be obtained, given ν_0, by drawing N_0 samples of T observations from (2.8), then computing the corresponding average measures of skewness and kurtosis; see Appendix B.2.1.[5]

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 on the same $\overline{\mathrm{SK}}(\nu_0)$ and $\overline{\mathrm{KU}}(\nu_0)$; 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 (2000).

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**:

$$\mathrm{CSK} = 1 - \min\left\{\widehat{p}(\mathrm{ESK}(\nu_0)|\nu_0),\ \widehat{p}(\mathrm{EKU}(\nu_0)|\nu_0)\right\} \tag{5.5}$$

where the subscript $\mathcal{M}$ (previously used in the notation for the individual MC p-values) is withheld 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 an 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 (2002*a*), Dufour, Khalaf, Bernard and Genest (2001) and Dufour and Khalaf (2001*c*).

5.2. Multivariate tests for GARCH and variance ratio tests

In this section we consider tests for departure from i.i.d. errors, specifically, tests for GARCH and variance ratio tests. If one pursues a univariate approach, these standard tests may be applied to each equation in the system (2.1). For instance, the Engle GARCH test statistic for equation

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

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 $\widehat{u}_{it}^2$ on a constant and $\widehat{u}_{(t-j),i}^2$ $(j = 1, \ldots, q)$; 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

$$LK_i = \frac{\big((T-q)\sum_{t=q+1}^{T}\left[(\widehat{u}_{it}^2/\widehat{\sigma}_i^2 - 1)\right]\sum_{j=1}^{q}\widehat{u}_{(t-j),i}^2\big)/\sum_{t=q+1}^{T}\big((\widehat{u}_{it}^2/\widehat{\sigma}_i^2 - 1)^2\big)^{1/2}}{\big((T-q)\sum_{t=q+1}^{T}\big(\sum_{j=1}^{q}\widehat{u}_{(t-j),i}^2\big)^2 - \big(\sum_{t=q+1}^{T}\big(\sum_{j=1}^{q}\widehat{u}_{(t-j),i}^2\big)\big)^2\big)^{1/2}} \tag{5.6}$$

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

$$VR_i = 1 + 2\sum_{j=1}^{J}(1 - \frac{j}{l})\widehat{\rho}_{ij} \tag{5.7}$$

where

$$\widehat{\rho}_{ij} = \frac{\sum_{t=j+1}^{T}\widehat{u}_{it}\widehat{u}_{i,t-j}}{\sum_{t=1}^{T}\widehat{u}_{ti}^2}, \quad j = 1, \ldots, J, \tag{5.8}$$

are empirical residual autocorrelations. The latter statistic estimates the ratio

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

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

$$(VR_i - 1) \overset{asy}{\sim} N(0, 2(2J-1)(J-1)/3J). \tag{5.9}$$

In Dufour, Khalaf, Bernard and Genest (2001), we show that Engle and Lee-King's test criteria are nuisance-parameter-free under the homoskedasticity null hypothesis, in a single equation setting. We establish the same property in the case of the variance ratio test in Dufour and Khalaf (2001*c*). 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 applicable to a multivariate regression. This is mainly due to two statistical problems. First, as pointed out above, the error covariance, which intervenes 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

$$\widetilde{W} = \widehat{U}\big(\widehat{U}'\widehat{U}\,\big)^{-1/2} \tag{5.10}$$

where $\big(\widehat{U}'\widehat{U}\,\big)^{-1/2}$ refers to the inverse of a Cholesky-type decomposition of $\widehat{U}'\widehat{U}$. 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 $\widehat{u}_{it}$ 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 (2.5), $\widetilde{W}$ 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 (2.5). It follows that under (2.5), 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. Consider the following combined statistics:

$$\widetilde{E} \;=\; 1 - \min_{1\le i\le n}\big[p\big(\widetilde{E}_i\big)\big] \tag{5.11}$$

$$\widetilde{LK} \;=\; 1 - \min_{1\le i\le n}\big[p\big(\widetilde{LK}_i\big)\big] \tag{5.12}$$

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

where $p\big(\widetilde{E}_i\big)$, $p\big(\widetilde{LK}_i\big)$ and $p\big(\widetilde{VR}_i\big)$ refer to p-values; these may be obtained applying a MC test method, or using asymptotic null distributions [to cut execution time]. Then apply an MMC test procedure to the combined statistic imposing (2.8); refer to Appendix B, where 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 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 (2.2) in the context of (2.1)] with different distributional assumptions for 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.[6] For each month the industry portfolios comprise those firms for which the return, price per common share and

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

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).

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 proxy the market return with the value-weighted NYSE returns, also available from CRSP. The risk-free rate is proxied by the one-month Treasury Bill rate, also from CRSP.

Our results are summarized in Tables 2-3. All MC tests where applied with 999 replications. As usual in this literature, we estimate and test the model over intervals of 5 years.[7] 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 Gaussianity 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 further justifies our approach to test the CAPM under non-Gaussian errors.

In columns (4)-(7) of Table 2, we present the LR and its asymptotic $\chi^2(n-1)$ p-value (p_∞), the Gaussian based and the largest Student-t based MC p-value associated with LR (respectively, $p_{\mathcal{N}}$ and Q_U). The confidence set for κ ($\mathcal{C}(Y)$) appears in column (8). These results allow us to compare rejection decisions across different distributional assumptions for the returns of the 12 portfolios. Similarly in columns (1)-(5) of Table 3, we report our set estimates of π and ω;

[7]Note that we also ran the analysis using ten year subperiods and that our results were not significantly affected.

Table 2. Normality Tests and Tests of the CAPM

	Normality tests			Tests of H_{CAPM}				
	(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	.373	3-15
1931-35	.001	.001	.001	16.257	.1798	.313	.322	2-8
1926-40	.001	.001	.001	16.018	.1904	.319	.333	4-23
1941-45	.004	.004	.013	25.869	.0112	**.045**	**.049**	≥ 4
1946-50	.001	.001	.001	37.196	.0002	.003	.004	4-26
1951-55	.001	.002	.001	36.510	.0003	.004	.005	4-34
1956-60	.024	.022	.032	43.841	.0000	.002	.002	≥ 4
1961-65	.594	.453	.588	39.098	.0001	.002	.002	≥ 7
1966-70	.011	.003	.004	36.794	.0002	.003	.003	≥ 4
1971-75	.001	.002	.001	21.094	.0490	.120	.129	4-24
1976-80	.001	.001	.001	28.373	.0049	**.023**	**.026**	3-17
1981-85	.001	.001	.001	27.189	.0073	**.033**	**.035**	4-31
1986-90	.028	.030	.038	35.747	.0007	.003	.005	≥ 4
1991-95	.177	.326	.244	16.752	.1592	.299	.305	≥ 9

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 (5.4), (5.4) and (5.5) given multivariate normal errors. p-values are MC pivotal statistics based. Column (4) presents the quasi-LR statistic defined in (3.9) to test H_{CAPM} [see (2.2)]; columns (5), (7) and (8) 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 (9). 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 (5.5) 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 the CAPM given multivariate mixture of normals

	Confidence set for π, ω					LR
	(1)	(2)	(3)	(4)	(5)	(6)
Sample	π =.1	π =.2	π =.3	π =.4	π =.5	Q_U
1927-30	$\geq$1.7	1.6-1.8	1.6-2.5	1.6-2.5	1.6-3 & $\geq$10	.382
1931-35	2.1-5.5 $\geq$100	1.9-3.0	1.9-3.0	2.0-3.5	$\geq$2.2	.313
1966-40	1.5-3.5	1.5-2.3	1.4-2.1	1.5-2.2 $\cup$ $\geq$20	1.5-2.4 $\cup$ $\geq$100	.328
1941-45	1.5-3.0	1.3-2.1	1.3-2.0	1.3-2.0 $\cup$ $\geq$20	1.3-2.3 $\cup$ $\geq$10	**.043**
1946-50	1.5-3.5	1.4-2.3	1.4-2.1	1.4-2.3 $\cup$ $\geq$20	1.4-2.5 $\cup$ $\geq$10	.003
1951-55	1.4-3.0	1.3-2.1	1.3-2.1	1.3-2.1 $\cup$ $\geq$20	1.4-2.3 $\cup$ $\geq$10	.003
1956-60	1.0-2.8	1.0-2.0	1.0-2.0	1.0-1.9 $\cup$ $\geq$20	1.0-1.9 $\cup$ $\geq$100	.002
1961-65	1.0-2.2	1.0-1.7	1.0-1.6	1.0-1.5 $\cup$ $\geq$20	1.0-1.6 $\cup$ $\geq$100	.002
1966-70	1.3-2.8	1.3-2.0	1.3-2.0	1.3-1.8 $\cup$ $\geq$20	1.2-2.0 $\cup$ $\geq$100	.002
1971-75	1.5-3.5	1.4-2.2	1.4-2.1	1.4-2.2 $\cup$ $\geq$20	1.4-2.6 $\cup \geq 9.5$	.128
1976-80	1.6-3.5	1.5-2.5	1.5-2.3	1.5-2.4 $\cup$ $\geq$20	1.6-2.9 $\cup \geq 7.5$	.022
1981-85	1.4-3.5	1.4-2.3	1.4-2.1	1.4-2.1 $\cup$ $\geq$20	1.4-2.4 $\cup$ $\geq$10	**.030**
1986-90	1.0-3.0	$\leq$2	$\leq$1.9	1.0-1.9 $\cup$ $\geq$50	1.0-2.0 $\cup$ $\geq$50	.004
1991-95	1.0-1.9	$\leq$1.6	$\leq$1.4	1.0-1.4 $\cup$ $\geq$100	1.0-1.3 $\cup$ $\geq$200	.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 (5.5) 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 (3.9) to test H_{CAPM} [see (2.2)]; 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.

for presentation ease, the confidence region is summarized as follows: we give the confidence set for ω corresponding to five different values of π namely .1, .2., 3., .4 and .5. Column (6) of Table 3 present the largest mixture-of-normals based MC p-values associated with the efficiency LR statistic (which we reported in column (4) of 2). Our empirical evidence shows that asymptotic p-values are quite often spuriously significant (e.g. 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 with $\kappa \to \infty$. So, if $p_{\mathcal{N}}$ exceeds the significance level, then 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 re-consider 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.[8] These results differ from Zhou's (1991) which showed 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 account explicitly for nuisance parameters.

Our results indicate clearly 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 mixtures distributions with parameters not rejected by proper GF tests, we see that the decision of the MMC CAPM test can change relative to the F-based test.

Figures 1-28 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, in this literature 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:

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

where $[j]$ refers to the sub-periods, and $\Psi^{-1}(.)$ provides the standard normal deviate corresponding to $p_{\mathcal{N}}[j]$. If the CAPM null 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 (2.8) by replacing, in (6.1), $p_{\mathcal{N}}[j]$ with $Q_{U[j]}$, the MMC p-values obtained imposing (2.8). 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,

[8] 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%.

Figure 15: MC GRS test (excluding January), 1927–30

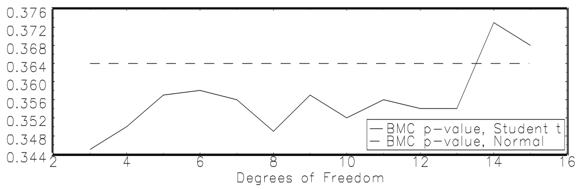


Figure 16: MC GRS test (excluding January), 1931–35

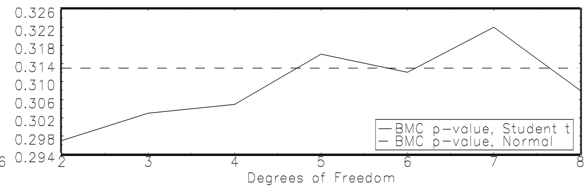


Figure 17: MC GRS test (excluding January), 1936–40

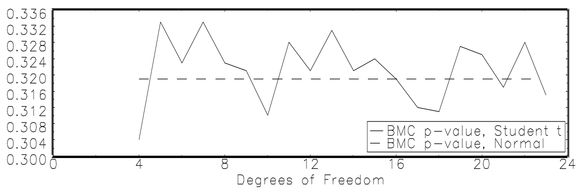


Figure 18: MC GRS test (excluding January), 1941–45

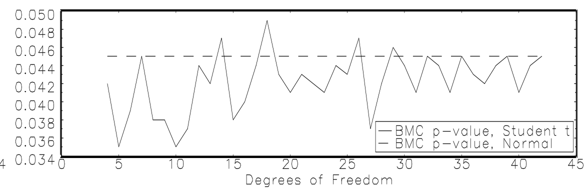


Figure 19: MC GRS test (excluding January), 1946–50

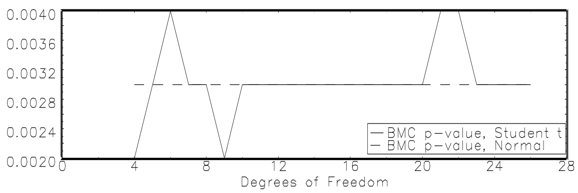


Figure 20: MC GRS test (excluding January), 1951–55

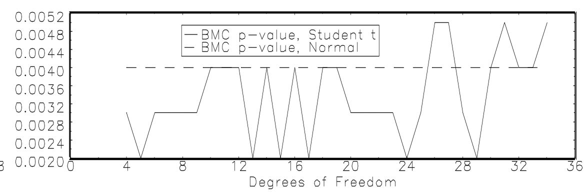


Figure 21: MC GRS test (excluding January), 1956–60

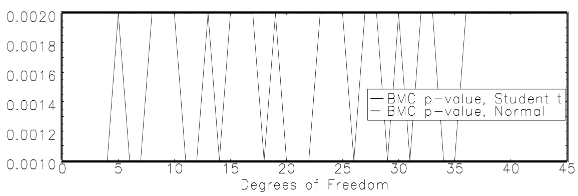


Figure 22: MC GRS test (excluding January), 1961–65

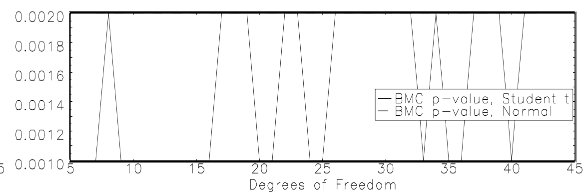


Figure 23: MC GRS test (excluding January), 1966–70

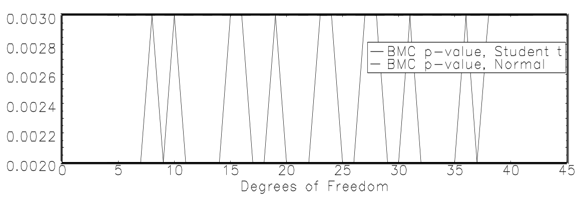


Figure 24: MC GRS test (excluding January), 1971–75

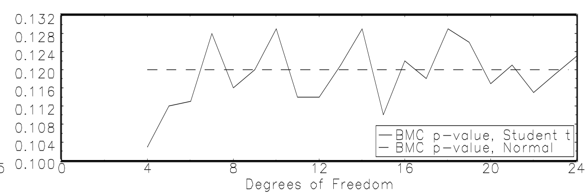


Figure 25: MC GRS test (excluding January), 1976–80

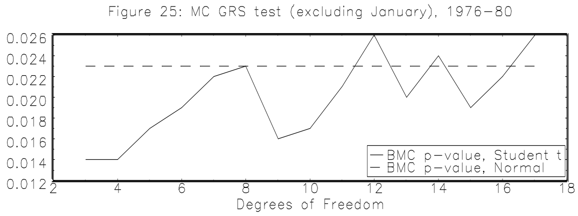


Figure 26: MC GRS test (excluding January), 1981–85

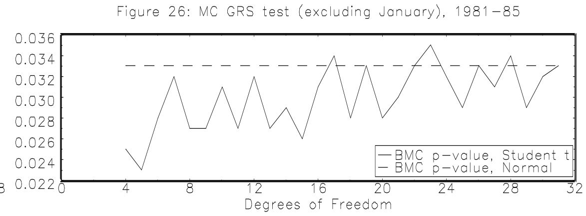


Figure 27: MC GRS test (excluding January), 1886–90

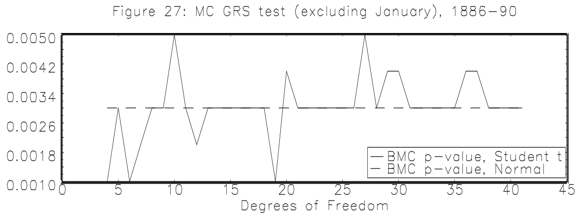
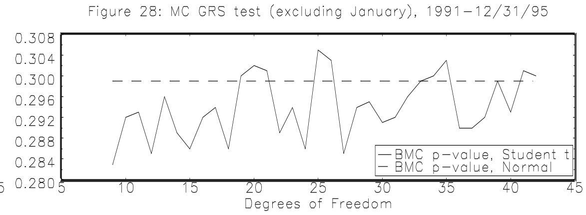


Figure 28: MC GRS test (excluding January), 1991–12/31/95



Note _ Using the Student-t distribution, each figure represents the Monte Carlo probability of rejecting the hypothesis that $H_{CAPM}\ : a_i = 0, i = 1, ..., n$ 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_{CAPM} using the normality assumption.

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 $< .0000$. If independence is upheld as in Gibbons and Shanken (1987), this implies that the CAPM is jointly rejected with our data.[9]

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.[10] The results show very few rejections of the null hypothesis 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*b*) 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 CAPM exact tests which formally take these non-normal distributions into consideration fail to reject the CAPM for 3 out of 9 subperiods for which the Gaussian-based test is significant. It appears that the distributional setup is crucial when testing the CAPM. This suggest 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 CAPM 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)); see 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

[9] 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 CAPM test. The results reported in the above Tables are robust to this change of levels.

[10] 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.

Table 4. Multivariate diagnostics

	Normal errors			Student t-errors			Mixtures errors		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sample	$\widetilde{E}$	$\widetilde{LK}$	$\widetilde{VR}$	$\widetilde{E}$	$\widetilde{LK}$	$\widetilde{VR}$	$\widetilde{E}$	$\widetilde{LK}$	$\widetilde{VR}$
1927-30	.001	.356	.004	130	.333	.004	.155	.335	004
1931-35	.022	.748	.069	124	.659	.066	.208	.671	080
1926-40	.075	.612	.855	124	.594	.867	.208	.596	866
1941-45	.824	.979	.163	843	.982	.177	.831	.979	175
1946-50	.003	.804	.063	017	.780	.068	.039	.778	072
1951-55	.139	.353	.111	189	.338	.119	.242	.328	120
1956-60	.987	.628	.093	994	.628	.095	.996	.633	099
1961-65	.339	.207	.577	375	.195	.584	.425	.207	592
1966-70	.027	.274	.821	067	.278	.847	.098	.288	846
1971-75	.280	.224	.218	316	.212	.224	.390	.207	220
1976-80	.004	.011	.165	031	.013	.183	.038	.009	184
1981-85	.027	.103	.208	075	.103	.217	.105	.095	223
1986-90	.033	.453	.346	077	.442	.366	.110	.455	357
1991-95	.803	.236	.088	821	.237	.096	.816	.252	950

Note _ Numbers shown are p-values associated with the combined tests $\widetilde{E}$, $\widetilde{LK}$, and $\widetilde{VR}$, defined by (5.11), (5.12) and (5.13). $\widetilde{E}$ 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.

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, Dufour and Khalaf (2001). Finally, we note that an apparent shortcoming of our exact tests comes from the fact that right-hand-side benchmarks are possibly observed with errors. The development of exact tests which correct for error-in-variable problems is an appealing idea for future research.

Appendix

A. Finite-sample distributional properties

The results in this Appendix pertain to any asset pricing model which may be cast in terms of the MLR given by (2.3) and (2.5).

A.1. General uniform linear restrictions tests

This section relates to testing constraints on regression coefficients of the UL form (2.4). On observing that (2.4) 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:

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

where $\widehat{\Sigma}_0$ 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 (2002*b*).

Theorem A.1 DISTRIBUTION OF THE QUASI-LR UNIFORM-LINEAR HYPOTHESIS TEST STATISTIC. *Under* (2.3), (2.5) *and* (2.4)*, the statistic*

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

is distributed like

$$\left|E'W'MWE\right| / \left|E'W'M_0WE\right| \tag{A.3}$$

where $\widehat{\Sigma}_{01}$ *and* $\widehat{\Sigma}$ *are the constrained and unconstrained MLE of* Σ,

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

and $W = [W_1,\ \ldots\ ,\ W_n]$ *is defined as in* (2.5).

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

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

where

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

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 (2.5). 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 $\widehat{U} = MU$ and $U = WJ'$, it is straightforward to see that

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

Since (2.5) entails that W has a known distribution, it follows that $\widehat{U}(\widehat{U}'\widehat{U})^{-1}\widehat{U}'$ (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. Monte Carlo tests

The Monte Carlo (MC) test procedure goes back to Dwass (1957) and Barnard (1963). Dufour (2000) analyzes formally the nuisance-parameter-dependent case. Here we summarize the underlying methodology (given a right tailed test), as it applies to the test statistics we consider in this paper.

B.1. General method

Let us first consider the pivotal statistics case, i.e. the case where the statistic at hand, say $S(y,\ X)$ can be written as a pivotal function of W (in (2.5)), formally

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

where W is defined by (2.5), 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: $W^j = [W_1^j,\ \ldots\ ,\ W_n^j]$, $j = 1, ..., N$, conforming with (2.5).

3. From each simulated error matrix W^j, compute the statistics

$$S^j = \overline{S}(W^j,\ X),\ j = 1\ ,\ \ldots\ ,\ N.$$

For instance, in the case of the QLR statistic underlying Theorem **A.1**, calculate $|W^{j\prime}\widetilde{M}W^j|/|W^{j\prime}\widetilde{M}_0 W^j|\ ,\ j = 1\ ,\ \ldots\ ,\ N.$

4. Compute the MC p-value

$$\widehat{p}_N(S_0) = \frac{N\widehat{G}_N(S_0\ ) + 1}{N+1}, \tag{B.1}$$

where

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

In other words, $N\widehat{G}_N(S_0)$ is the number of simulated criteria $\geq S_0$ and $\widehat{R}_N(S_0) = N - N\widehat{G}_N(S_0) + 1$ gives the rank of S_0 in the series $S_0,\ S_1,\ \ldots\ ,\ S_N$.

5. The MC critical region is

$$\widehat{p}_N(S_0) \leq \alpha, \quad 0 < \alpha < 1. \tag{B.3}$$

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

$$P_{(H_0)}\left[\widehat{p}_N(S_0\ ) \leq \alpha\right] = \alpha. \tag{B.4}$$

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 (2.8). In this case, given ν, (B.1) yields a MC p-value which we will denote $\widehat{p}_N(S_0|\nu)$ where the conditioning on ν is emphasized for further reference. The test defined by $\widehat{p}_N(S_0|\nu) \leq \alpha$ is *exactly size correct* (in the sense of (B.4)) for known ν. Treating ν as a formal nuisance parameter, the test based on

$$\sup_{\nu\ \in\ \Phi_0}\left[\widehat{p}_N(S_0|\nu)\right] \leq \alpha \tag{B.5}$$

where Φ_0 is a nuisance parameter set consistent with H_0, is *exact at level* α; see Dufour (2000). 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 $\widehat{p}_N(S_0|\widehat{\nu}_0)$, where $\widehat{\nu}_0$ is any *point* estimate of ν. In Dufour and Khalaf (2001*a*)-Dufour and Khalaf (2001*b*), 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 $\widehat{p}_N(S_0|\widehat{\nu}_0) > \alpha$, then the exact test MMC test is clearly not significant at level α.

B.2. MC skewness and kurtosis tests

B.2.1. Estimating expected skewness and kurtosis

A1. Draw N_0 *i.i.d.*replications, $W^j = [W_1^j, \ldots, W_n^j]$, $j = 1, ..., N_0$, conforming with the hypothesized distribution with $\nu = \nu_0$.

A2. From each simulated error matrix W^j, compute

$$MW^j \left(W^{j\prime} MW^j\right)^{-1} W^{j\prime} M,$$

$j = 1, \ldots, N_0$. These provide N_0 replications of sk and ku, applying (5.1) and (5.2), namely sk_j and ku_j.

A3. Then calculate

$$\overline{\mathrm{SK}}(\nu_0) = \sum_{j=1}^{N_0} \mathrm{sk}_j / N_0, \qquad \overline{\mathrm{KU}}(\nu_0) = \sum_{j=1}^{N_0} \mathrm{KU}_j / N_0$$

Two questions arise at this stage: (i) how to obtain exact cut-off points for (5.3) and (5.4), and (ii) how to obtain a size-correct simultaneous test which combines (5.3) and (5.4). Let us first address the individual p-values issue, which may be run as in Appendix B 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 $\mathcal{M}$ of replications, and independently from the simulation performed to obtain $\overline{\mathrm{SK}}(\nu_0)$ and $\overline{\mathrm{KU}}(\nu_0)$ (i.e. step A1 above), draw $W^m = [W_1^m, \ldots, W_n^m]$, $m = 1, ..., \mathcal{M}$, conforming with (2.6).

B3. From each simulated error matrix W^m, compute

$$MW^m \left(W^{m\prime} MW^m\right)^{-1} W^{m\prime} M,$$

$m = 1, \ldots, \mathcal{M}$. Conformably, derive, applying (5.1) and (5.2), $\mathcal{M}$ replications of SK and KU, SK_m and KU_m.

B4. Conditioning on $\overline{\mathrm{SK}}(\nu_0)$ and $\overline{\mathrm{KU}}(\nu_0)$ (generated only once as in steps (A1-A3)), obtain, applying (5.3) and (5.4), $\mathcal{M}$ replications of ESK and EKU, ESK_m and EKU_m.

B5. Obtain (respectively) the ranks of ESK_0 and EKU_0 in the series $\{\mathrm{ESK}_0, \mathrm{ESK}_1, \ldots, \mathrm{ESK}_{\mathcal{M}}\}$ and $\{\mathrm{EKU}_0, \mathrm{EKU}_1, \ldots, \mathrm{EKU}_{\mathcal{M}}\}$ respectively; these yield the MC p-values (applying B.1) $\widehat{p}_{\mathcal{M}}(\mathrm{ESK}_0 \,|\nu_0)$ and $\widehat{p}_{\mathcal{M}}(\mathrm{EKU}_0 \,|\nu_0)$.

B.2.3. Combined excess skewness and kurtosis test

C1. Derive the observed value of the test statistic from (5.5), which we will denote csk_0. To do so, run steps B1-B5 (of the preceding section).

C2. Obtain a simulated value for esk and eku, which we will denote ESK_{01}, EKU_{01} using the same value for $\overline{\mathrm{SK}}(\nu_0)$ and $\overline{\mathrm{KU}}(\nu_0)$. Run steps B2-B5, using the same simulated series in B2 and replacing ESK_{01}, EKU_{01} for ESK_0, EKU_0. This would yield one simulated value of the combined statistic.

C3. Repeat step C2 to complete the simulated series. To do so, draw ESK_{0m}, EKU_{0m}, $m = 2, 3, ...,$ maximum number of replications desired.

C4. Obtain the rank of the observed statistic CSK_0, within the simulated series, and derive the corresponding p-value, which we will denote $\widehat{p}(\mathrm{CSK}_0 \,|\nu_0)$.

References

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

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

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., 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.

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.

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. (2000), Économétrie, théorie des tests et philosophie des sciences, *in* ‘Présentations de l’Académie des lettres et des sciences humaines’, Vol. 53, Royal Society of Canada/Société royale du Canada, Ottawa, pp. 166–182.

Dufour, J.-M. and Khalaf, L. (2001*a*), ‘Exact tests for contemporaneous correlation of disturbances in seemingly unrelated regressions’, *Journal of Econometrics* **forthcoming**.

Dufour, J.-M. and Khalaf, L. (2001*b*), Finite sample tests in seemingly unrelated regressions, Technical report, C.R.D.E., Université de Montréal, and GREEN, Université Laval. 25 pages.

Dufour, J.-M. and Khalaf, L. (2001*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*a*), ‘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*b*), ‘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, Département d'économique, Université Laval, and CRDE, Université de Montréal.

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 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.

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.

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.

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.

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 sampes: 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.

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.

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

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

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.