Exact skewness-kurtosis tests for multivariate normality and goodness-of-fit in multivariate regressions with application to asset pricing models *

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

First version: June 2002
This version: March 2003
Compiled: February 10, 2005, 11:45am

This paper has been published in the *Oxford Bulletin of Economics and Statistics*, 65 (2003), 891-906.

* This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Alexander-von-Humboldt Foundation (Germany), the Institut de finance mathématique de Montréal (IFM2), 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, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec).This paper was also partly written at the Centre de recherche en Économie et Statistique (INSEE, Paris), the Technische Universität Dresden (Fakultät Wirtschaftswissenschaften) and the Finance Division at the University of British Columbia.

† 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: 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

‡ Centre interuniversitaire de recherche en économie quantitative (CIREQ), and GREEN, Université Laval, Pavillon J.-A. De Sève, Ste-Foy, Québec, Canada, G1K 7P4. TEL: (418) 656 2131-2409; FAX: (418) 656 7412. email: lynda.khalaf@ecn.ulaval.ca

§ Centre Interuniversitaire de Recherche sur les Politiques Économiques et l'Emploi (CIRPÉE), Département de finance et assurance, Pavillon Palasis-Prince, local 3632, Ste-Foy, Université Laval, Québec, Canada G1K 7P4. TEL: (418) 656-2926, FAX: (418) 656-2624. email: Marie-Claude.Beaulieu@fas.ulaval.ca

ABSTRACT

We study the problem of testing the error distribution in a multivariate linear regression (MLR) model. The tests are functions of appropriately standardized multivariate least squares residuals whose distribution is invariant to the unknown cross-equation error covariance matrix. Empirical multivariate skewness and kurtosis criteria are then compared to simulation-based estimate of their expected value under the hypothesized distribution. Special cases considered include testing multivariate normal, Student t, normal mixtures and stable error models. In the Gaussian case, finite-sample versions of the standard multivariate skewness and kurtosis tests are derived. To do this, we exploit simple, double and multi-stage Monte Carlo test methods. For non-Gaussian distribution families involving nuisance parameters, confidence sets are derived for the nuisance parameters and the error distribution. The procedures considered are evaluated in a small simulation experiment. Finally, the tests are applied to an asset pricing model with observable risk-free rates, using monthly returns on New York Stock Exchange (NYSE) portfolios over five-year subperiods from 1926-1995.

Key-words: multivariate linear regression; goodness-of-fit; normality test; multivariate normality; multinormality; Student t; normal mixture; stable distribution; specification test; diagnostics; exact test; Monte Carlo test; bootstrap; nuisance parameter; asset pricing model; CAPM.

Journal of Economic Literature classification: C3; C12; C33; C15; G1; G12; G14.

RÉSUMÉ

Dans cet article, nous proposons des tests sur la forme de la distribution des erreurs dans un modèle de régression linéaire multivarié (RLM). Les tests que nous développons sont fonction des résidus obtenus par moindres carrés multivariés, lesquels sont standardisés de façon à ce que leur distribution soit invariante à la matrice de covariance, inconnue, des erreurs. Notre approche utilise des mesures empiriques d'asymétrie et d'aplatissement de la distribution des erreurs, que nous comparons à des estimations engendrées par simulation de ces caractéristiques sous cette même hypothèse distributionnelle. Les cas spécifiques que nous étudions comprennent des tests sur les erreurs du modèle dans le cadre des lois normale, t de Student, mélange de normales et stable. Dans le cas gaussien, nous obtenons des versions exactes de tests d'ajustement standards sur l'asymétrie et l'aplatissement des erreurs dans le cas multivarié. À cette fin, nous utilisons des tests de Monte Carlo simples, doubles et multiples. Dans les cas non-gaussiens, comme les familles de lois dépendent de paramètres de nuisance, nous proposons des régions de confiance pour ces derniers et la distribution des erreurs. Les procédures introduites dans cet article sont alors évaluées par une simulation de petite taille. Finalement, les tests proposés sont appliqués à un modèle d'évaluation d'actifs impliquant un taux d'intérêt sans risque observable et utilisant les rendements de portefeuilles mensuels de titres inscrits à la bourse de New York, sur des sous-périodes de cinq ans allant de janvier 1926 à décembre 1995.

Mots-clefs: modèle de régression multivarié; test d'ajustement; test de normalité; normalité multivariée; t de Student; mélange de lois normales; distribution stable; test de spécification; diagnostic; test exact; test de Monte Carlo; bootstrap; paramètre de nuisance; modèle d'évaluation d'actifs financiers; CAPM.

Classification du Journal of Economic Literature: C3; C12; C33; C15; G1; G12; G14.

Contents

List of Propositions and Theorems **iv**

1. Introduction **1**

2. Framework **4**

3. Multivariate standardized residuals **5**

4. Skewness-kurtosis goodness-of-fit tests **7**
4.1. Basic test statistics 7
4.2. Extension to testing non-Gaussian distributions 9
4.3. Nonstandard null distributions and multi-stage MC tests 10
4.4. The unknown nuisance parameter case 13

5. Simulation experiment **13**

6. Empirical application **17**

7. Conclusion **23**

A. Appendix: Monte Carlo tests **24**

List of Tables

1 Size of multinormality tests 14
2 Power of multinormality tests: Design I 15
3 Power of multinormality tests: Design II 16
4 Power of multinormality tests: Design III 16
5 Portfolio definitions 17
6 Multinormality tests 18
7 Multivariate t distributions: combined-tests based confidence sets for the degrees-of-freedom parameter 19
8 Multivariate symmetric stable distributions: confidence sets for the kurtosis parameter based on combined tests 20
9 Multivariate t and stable distributions: skewness-based confidence sets for distributional parameters 21
10 Multivariate t and stable distributions: kurtosis-based confidence sets for distributional parameters 22

List of Definitions, Propositions and Theorems

3.1 **Theorem :** Invariance of Cholesky standardized multivariate residuals 6
3.2 **Theorem :** Invariance of Mahalanobis residual matrix 7
4.1 **Theorem :** Distribution of JB-type statistics in MLR 9
4.2 **Theorem :** Distribution of Mardia-type skewness and kurtosis 9

1. Introduction

Drawing inference on the parameters of multivariate linear regression (MLR) models is a basic statistical problem. Such models, which can combine both cross-section and time series data, are common in various fields of statistics and econometrics; see Rao (1973, Chapter 8), Anderson (1984, chapters 8 and 13), Kariya (1985), Stewart (1997), Dufour and Khalaf (2002*c*, 2002*b*, 2003), and the references therein. Important cases include consumer and factor demand systems, reduced forms derived from linear simultaneous equation models, and various asset pricing models in finance. In particular, familiar MLR-based applications in finance include *market-models,* such as the capital asset pricing model (CAPM) which may be traced back to Gibbons (1982) and Fama and French (1993, 1995). The associated empirical literature which has evolved from Gibbons' seminal work is enormous; for reviews, the reader may consult Campbell, Lo and MacKinlay (1997) and Shanken (1996).[1]

Inference procedures (such as tests and confidence sets) for MLR models tend to be heavily influenced by the disturbance distribution and the assumptions made on the latter. Under standard conditions, usual asymptotic distributions are often distribution-free, but it is well known that the finite-sample reliability of large-sample approximations tends to be quite bad; see, for example, Dufour and Khalaf (2002*b*, 2002*c*) for simulation evidence. Some finite-sample procedures have been proposed in the statistical literature, but these are almost entirely restricted to the case where the disturbance vectors follow a Gaussian distribution. Another avenue consists in using simulation-based tests, as described in Dufour and Khalaf (2002*c*). The latter approach allows one to relax the normality assumption and provides provably exact tests in finite samples, but still requires the formulation of a parametric model on the errors. In particular, heavy-tailed distributions that would be important in financial modelling may easily be accommodated in this way.

This situation underscores the importance of testing disturbance normality as well as other parametric distributional assumptions in the context of MLR models. Another motivation comes from the fact that relatively specific distributional assumptions may be required by important economic or financial hypotheses, e.g. mean-variance efficiency in the context of the CAPM model.[2]

In multivariate regression contexts, relatively little work has been done on testing distributional goodness-of-fit (GF) tests compared to the univariate case. This holds even when the hypothesized null distribution is multivariate normal; see the reviews of Mardia (1980), D'Agostino and Stephens (1986) and Thode (2002). Indeed, system diagnostic tests raise problems not encountered in the analysis of univariate models. In particular, an important difficulty comes from cross-equation disturbance correlations. Whereas it is highly desirable to use test procedures that take account of these correlations, the fact remains that these parameters can easily constitute (unknown) nuisance

[1]Well known financial applications include: (i) portfolio efficiency tests in e.g. CAPM contexts [see, for example, Shanken (1986), MacKinlay (1987), Gibbons, Ross and Shanken (1989, GRS), Affleck-Graves and McDonald (1989), Shanken (1990), Zhou (1991), Zhou (1993), Zhou (1995), Fama and French (1993, 1995), Stewart (1997), Velu and Zhou (1999), Chou (2000), Groenwold and Fraser (2001) and Beaulieu, Dufour and Khalaf (2001*b*, 2001*a*)]; (ii) spanning tests [see for example Jobson and Korkie (1982, 1989), Kan and Zhou (2001)]; and (iii) event studies tests [see Binder (1985), Schipper and Thompson (1985)].

[2]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).

parameters. The typical approach to this problem is to consider statistics whose distribution is asymptotically free of nuisance parameters; see, for example, Mardia (1970), Richardson and Smith (1993), Kilian and Demiroglu (2000), Fiorentini, Sentana and Calzolari (2003), and the review of Thode (2002). Although this leads to convenient test procedures, in systems with many equations, it is likely that the number of nuisance parameters will be quite large relative to the sample size, so again asymptotic results will provide poor approximations in finite samples; see Horswell and Looney (1992, 1993) and Holgersson and Shukur (2001).

It is worth noting that the statistical literature on GF tests has focused mainly on the location-scale model, which may be seen as a special case of the MLR model where the regressors reduce to a vector of ones. This is clearly the case, for example, for the multivariate skewness and kurtosis coefficients suggested by Mardia (1970); e.g., see Mardia (1980, 1974), Baringhaus and Henze (1992), Lütkepohl and Theilen (1991), Horswell and Looney (1992, 1993) and Henze (1994). Indeed, the presence of covariates considerably complicates the testing problem and related (exact and asymptotic) distributional theory, even in univariate regressions; see Dufour, Farhat, Gardiol and Khalaf (1998), Bontemps and Meddahi (2002) and the references therein. Furthermore, despite the widespread recognition of such problems, our review of the statistics and econometrics literature has revealed that exact multivariate GF tests are unavailable, even for the Gaussian hypothesis or the location-scale model.

In this paper, we propose a general exact method for GF testing in MLR models. Our results can be summarized as follows. *First*, we address the distributional complications arising from the presence of covariates and unknown error covariances. We first state some basic finite-sample results concerning residual-based tests in general MLR models. We show that tests which use properly standardized residuals have a null distribution that does not depend on either regression coefficients, error variances or covariances, once the error distribution is parametrically specified up to an (unknown) linear transformation (or covariance matrix). More specifically, these tests are based on exploiting invariance properties for two distinct families of empirically scaled residuals: (1) a properly rescaled version of the residual matrix (using the Cholesky root of the empirical residual covariance matrix) is invariant to general triangular transformation of the error vector (across equations); (2) the projector matrix associated with the least square residual matrix from the MLR model is invariant to general linear transformations of the error vector. Corresponding pivotality properties then follow from these features. Although related pivotality results have been pointed out for the simpler Gaussian location-scale models [see Mardia (1980), Horswell and Looney (1993) and Lütkepohl and Theilen (1991)], it does not appear those presented here have been used in the earlier literature on inference in general MLR models.

Second, we exploit the above invariance results to derive finite-sample tests of multinormality for the disturbances of MLR models. We consider two categories of test statistics based on empirical multivariate skewness and kurtosis coefficients: (1) multivariate extensions of the familiar Jarque-Bera tests [Jarque and Bera (1987, henceforth JB)], obtained by combining individual residual-based JB tests computed from individual equations [as suggested by Kilian and Demiroglu (2000) for vector-autoregressive (VAR) models]; (2) Mardia-type statistics based on empirical skewness and kurtosis derived from the least squares residual projector. These statistics have finite-sample null distributions which may be very difficult to evaluate through analytical methods. However, due to the fact that their distributions are free of nuisance parameters and easy to simulate, we can exploit the technique of Monte Carlo (MC) tests [Dwass (1957), Barnard (1963), Dufour and Kiviet (1996, 1998), Dufour and Khalaf (2001)]. This simulation-based procedure yields an exact test when the distribution of the test statistic is *pivotal* under the null hypothesis: all we need is the possibility of simulating the relevant test statistic under the null hypothesis. Due to the flexibility of the MC test method, we define a number of new multinormality test statistics; in particular, these involve methods for combining excess skewness and kurtosis criteria.

Thirdly, we show that the multinormality tests proposed can easily be adapted to assess other hypothesized disturbance distributions. For that purpose, the statistics are modified in order to compare empirical multivariate skewness and kurtosis measures with *simulation-based estimates* of their expected values under the hypothesized distribution (instead of theoretical _ possibly inaccurate _ expected values that may be difficult to derive). These corrections are also applicable in the Gaussian case. The MC test method then works in this case as in the previous one to achieve perfect size control, taking account of the fact that the simulated test statistics are exchangeable (due to the presence of simulated moment estimates) rather than independent identically distributed (*i.i.d.*), leading to a *double MC test* procedure. As long as the disturbance distribution is specified up to an unknown linear transformation, there is no restriction on the form of the tested distribution. For example, the latter can be heavy-tailed and may even miss moments. The fact that a distribution does have a finite fourth moment does not preclude one to use an *empirical kurtosis* as a basis for assessing its goodness-of-fit. A *triple MC test* method is also proposed to combine several tests into an omnibus GF test.

Fourth, in view of financial applications, we focus on three classes of non-normal (possibly heavy-tailed) families: (1) multivariate Student t distributions, (2) multivariate mixtures of normal distributions, and (3) multivariate stable distributions. Our approach, however, is not restricted to these distributions. In contrast with the normal case, the non-normal families considered involve additional parameters, such as the degrees of freedom for the Student t distribution, that may be taken as unknown. The proposed MC non-Gaussian GF tests are exact when the null hypothesis sets these nuisance parameters to specific values. On assembling the nuisance parameter values which are not rejected (i.e., by "inverting" the GF tests), this yields *confidence sets* for the fitting distributions. Such confidence sets may then be used as an intermediate step in the context of other inference problems.

Fifth, we present the results of a small simulation experiment comparing the procedures considered. These show that the available asymptotic tests are completely unreliable from the viewpoint of size control, while the MC tests have the correct size (as expected). With respect to power, we find that Mardia-type tests are generally preferable to JB-type tests, sometimes by a wide margin, while the JB-type tests can perform marginally better in the case of stable distributions. This may reflect the fact that Mardia-type statistics are more directly adapted to testing multivariate (rather than univariate) normality.

Sixth, the tests proposed are applied to an asset pricing model 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 reveal the following. We first find that multivariate normality is rejected for all subperiods. This conclusion can be contrasted with earlier evidence on this issue, which is mixed: whereas the results of Campbell et al. (1997) and Affleck-Graves and McDonald (1989) suggest that normality cannot be rejected, those of Richardson and Smith (1993) indicate more rejections. So our results provide a firmer basis for rejecting normality. Then, inversion of the GF tests for Student t and stable error distributions reveals heavy kurtosis. In this empirical analysis, the Mardia-type tests appear to be much superior from the power viewpoint to those based on JB-type statistics from individual-equations. In particular, the confidence sets based on Mardia-type statistics are much tighter with those based on JB-type statistics. This observation is noteworthy, given the popularity of JB-type tests in econometrics.

The paper is organized as follows. Section 2 describes the framework. In Section 3, we define standardized residuals, discuss their invariance properties, and state our basic exact distributional results. In Section 4, we propose our multivariate GF test procedures; the associated size and power Monte Carlo studies are described in 5. Section 6 reports our empirical analysis. We conclude in Section 7.

2. Framework

Let us consider a system of correlated regression equations of the form:

$$Y = XB + U \tag{2.1}$$

where $Y = [Y_1, \ldots, Y_n]$ is a $T \times n$ matrix of observations on n dependent variables, X is a $T \times k$ full-column rank matrix, $B = [B_1, \ldots, B_n]$ is a $k \times n$ matrix of unknown fixed coefficients and $U = [U_1, \ldots, U_n] = [V_1, \ldots, V_T]'$ is a $T \times n$ matrix of random disturbances. Following Dufour and Khalaf (2002*c*), we suppose the errors have the following structure:

$$V_t = JW_t, \quad t = 1, \ldots, T, \tag{2.2}$$

$$w \equiv vec(W_1, \ldots, W_T) \sim \mathcal{F}(\nu)\,, \tag{2.3}$$

where $\mathcal{F}(\nu)$ is a known distribution, which may depend on the parameter ν, and J satisfies one of the two following conditions:

J is an unknown nonsingular lower triangular matrix; (2.4a)

J is an unknown nonsingular matrix. (2.4b)

Some of the procedures described below will be valid provided J is restricted to be triangular, while other ones only require J to be nonsingular.

On setting $W = [W_1, \ldots, W_T]'$, the above assumptions entail that

$$W = U(J^{-1})'. \tag{2.5}$$

In particular, this condition will be satisfied for *i.i.d.* normal errors. Let

$$\Sigma = JJ' \tag{2.6}$$

which gives the covariance matrix of V_t when $Cov(W_t) = I_n$. Note the assumptions (2.4a) and (2.4a) are equivalent when W_t follows a multinormal distribution (because the covariance matrix can always be written in the form $\Sigma = JJ'$ with J lower triangular), but this may not be the case if W_t is not Gaussian.

The least squares estimate of B is

$$\hat{B} = (X'X)^{-1}X'Y \tag{2.7}$$

and the corresponding residual matrix is

$$\hat{U} = \left[\hat{U}_1, \ldots, \hat{U}_n\right] = Y - X\hat{B} = MY = MU \tag{2.8}$$

where $M = I - X(X'X)^{-1}X'$. Note that the Gaussian (quasi) maximum likelihood estimators for this model are $\hat{B}$ and $\hat{\Sigma} = \hat{U}'\hat{U}/T$. It is clear from (2.5) and (2.8) that the distribution of $\hat{U}$ in general depends on the unknown scaling matrix J (or on the covariance matrix $\Sigma = JJ'$) so that test statistics based on $\hat{U}$ may involve J as a nuisance parameter. This will be the case in particular for the off-diagonal parameters which typically determine the dependence between the disturbances in different equations.

Our empirical application focuses on the asset pricing model

$$r_{ti} = a_i + b_i\tilde{r}_{\mathrm{M}t} + u_{ti}\,, \quad t = 1, \ldots, T,\ i = 1, \ldots, n, \tag{2.9}$$

where $r_{ti} = R_{ti} - R_t^F$, $\tilde{r}_{\mathrm{M}t} = \tilde{R}_{\mathrm{M}t} - R_t^F$, R_{ti}, $i = 1, \ldots, n$, are returns on n securities for period t, $\tilde{R}_{\mathrm{M}t}$ are the returns on the market portfolio under consideration, R_t^F is the riskless rate of return ($t = 1, \ldots, T$), and u_{ti} is a random disturbance. Clearly, this model is a special case of (2.1) where

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

and u_{ti} are the elements of the matrix U.

3. Multivariate standardized residuals

We now consider the problem of building residual-based test statistics whose null distribution will not be affected by the unknown scaling matrix *J*. In order to do this, we shall now state two general invariance results ensuring that appropriately standardized residuals have distributions which do not depend on *J*. The first one applies under the assumption (2.4a) where J is restricted to be triangular, while the second one holds under the more general assumption (2.4b).

Let

$$\tilde{W} = \hat{U}\,S_{\hat{U}}^{-1} \tag{3.1}$$

where $S_{\hat{U}}$ is the Cholesky factor of $\hat{U}'\hat{U}$, i.e. $S_{\hat{U}}$ is the (unique) upper triangular matrix such that

$$\hat{U}'\hat{U} = S_{\hat{U}}' S_{\hat{U}}\,, \quad \big(\hat{U}'\hat{U}\big)^{-1} = S_{\hat{U}}^{-1}\,\big(S_{\hat{U}}^{-1}\big)'.$$

Clearly, $\tilde{W}$ may be interpreted as a standardized form of $\hat{U}$. Further, $\tilde{W}$ satisfies the following important property.

Theorem 3.1 INVARIANCE OF CHOLESKY STANDARDIZED MULTIVARIATE RESIDUALS. *Under* (2.1) *and for all error distributions compatible with* (2.2) *and* $(2.4a)$*, the standardized residual matrix* $\tilde{W}$ *defined in* (3.1) *satisfies the identity*

$$\tilde{W} = \hat{U}\, S_{\hat{U}}^{-1} = \hat{W}\, S_{\hat{W}}^{-1} \tag{3.2}$$

where $\hat{W} = MW$ *and* $S_{\hat{W}}$ *is the* $($*unique*$)$ *upper triangular matrix such that*

$$\hat{W}'\hat{W} = S_{\hat{W}}' S_{\hat{W}}\,, \quad \big(\hat{W}'\hat{W}\big)^{-1} = S_{\hat{W}}^{-1}\,\big(S_{\hat{W}}^{-1}\big)'. \tag{3.3}$$

PROOF. Using (2.5), (2.8) and (3.3), we have:

$$\tilde{W} = \hat{U}(J^{-1})'\big(J'S_{\hat{U}}^{-1}\big) = MU(J^{-1})'\big(J'S_{\hat{U}}^{-1}\big) = MW\big(J'S_{\hat{U}}^{-1}\big) \tag{3.4}$$

and

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

On observing that $J'S_{\hat{U}}^{-1}$ is lower triangular, this means that $\big(J'S_{\hat{U}}^{-1}\big)'$ is the (unique) Cholesky factor of $(W'MW)^{-1}$, hence

$$J'S_{\hat{U}}^{-1} = S_{\hat{W}}^{-1}.$$

Substituting the latter identity in (3.4), we see that

$$\tilde{W} = MW\big(J'S_{\hat{U}}^{-1}\big) = \hat{W}S_{\hat{W}}^{-1}\,.$$

□

It follows from the latter theorem that any statistic which depends on the data only through $\tilde{W}$ follows a distribution which does not involve B or J (and is thus invariant to Σ), under the assumptions (2.1), (2.2) and (2.4a).

Consider now the Mahalanobis matrix

$$\hat{D} = \hat{U}\big(\hat{U}'\hat{U}/T\big)^{-1}\hat{U}' \tag{3.5}$$

on which Mardia-type tests of multinormality will be based [see Mardia (1970)]. The elements of this matrix satisfy an even stronger invariance property given by the following theorem.

Theorem 3.2 INVARIANCE OF MAHALANOBIS RESIDUAL MATRIX. *Under* (2.1) *and for all error distributions compatible with* (2.2) *and* $(2.4b)$, *the residual-based Mahalanobis matrix* $\hat{D}$ *defined in* (3.5) *satisfies*

$$\hat{D} = T\,\hat{W}\big(\hat{W}'\hat{W}\big)^{-1}\hat{W}' \tag{3.6}$$

and thus follows a distribution which is completely determined by the distribution of W *given* X.

PROOF. Using the identities $\hat{U} = MU$ and $U = WJ'$, we see that:

$$\begin{aligned}
\hat{U}(\hat{U}'\hat{U}/T)^{-1}\hat{U}' &= T\,MU(U'MU)^{-1}U'M = TMU(J^{-1})'J'(U'MU)^{-1}JJ^{-1}U'M \\
&= T\,MU(J^{-1})'\big[(J^{-1})U'MU(J^{-1})'\big]^{-1}J^{-1}U'M \\
&= T\,MW\left(W'MW\right)^{-1}W'M = T\,\hat{W}\big(\hat{W}'\hat{W}\big)^{-1}\hat{W}'.
\end{aligned}$$

□

It follows from the latter theorem that any statistic which depends on the data only through $\hat{D}$ follows a distribution which does not depend on B and J (and is thus invariant to Σ), under the assumptions (2.1), (2.2) and (2.4b). It is worth noting that the latter result relates to Theorem **3.1** since it is easy to see that

$$\hat{D} = T\,\hat{U}(\hat{U}'\hat{U})^{-1}\hat{U}' = T\,\hat{U}S_{\hat{U}}^{-1}\,(S_{\hat{U}}^{-1})'\hat{U}' = T\,\tilde{W}\tilde{W}'.$$

Theorems **3.1**-**3.2** include as special cases several known exact invariance results in the Gaussian location-scale model; see, for example, Mardia (1970), Lütkepohl and Theilen (1991) and Thode (2002). Here we show that invariance to B and Σ holds in general MLR models, and for all error distributions (Gaussian and non-Gaussian) which satisfy assumption (2.2).

4. Skewness-kurtosis goodness-of-fit tests

In this section, we use the above results to derive goodness-of-fit (GF) tests based on multivariate skewness and kurtosis coefficients. The proposed tests are formally valid for any parametric null hypothesis which takes the general form (2.2). In our empirical application [see section 6], we focus on multivariate t and symmetric stable distributions, which we denote $t(\kappa)$ and $Stb(\alpha_s)$ respectively, where κ represents degrees of freedom and α_s the kurtosis parameter of the stable distribution. Let us first consider the null hypothesis (2.2) where $\nu = \nu_0$ with ν_0 known.

4.1. Basic test statistics

The GF test statistics suggested here use two popular multivariate skewness and kurtosis measures: (i) measures based on Mahalanobis distance, and (ii) measures which aggregate individual equation skewness and kurtosis criteria. Specifically, we first consider extensions of the statistics

$$SK_{\mathrm{M}} = \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\hat{d}_{st}^{3}\,, \tag{4.1}$$

$$KU_{\mathrm{M}} = \frac{1}{T}\sum_{t=1}^{T}\hat{d}_{tt}^{2}\,, \tag{4.2}$$

where the variables $\hat{d}_{st}$ are the elements of the matrix $\hat{D} = \left[\hat{d}_{st}\right]$. These criteria were introduced by Mardia (1970) to assess deviations from multivariate normality, in models where the regressor matrix reduces to a vector of ones; see also Zhou (1993).[3] Mardia further proposed the omnibus normality test:

$$MSK = \frac{T}{6}SK_{\mathrm{M}} + \frac{T\left[KU_{\mathrm{M}} - n(n+2)\right]^2}{8n(n+2)} \underset{T\to\infty}{\sim} \chi^2\big((n/6)(n+1)(n+2)+1\big) \tag{4.3}$$

where the symbol $\underset{T\to\infty}{\sim}$ refers to the asymptotic null distribution of the test statistic.

Second, we consider extensions of the aggregate skewness and kurtosis criteria applied by Kilian and Demiroglu (2000) in vector-autoregressive contexts; these criteria were originally proposed by Jarque and Bera (1987, JB) :

$$SK_{\mathrm{KD}} = (sk_1, \ldots, sk_n)'(sk_1, \ldots, sk_n)\,, \tag{4.4}$$

$$KU_{\mathrm{KD}} = (ku_1 - 3, \ldots, ku_n - 3)'(ku_1 - 3, \ldots, ku_n - 3)\,, \tag{4.5}$$

$$sk_i = \frac{T^{-1}\sum_{t=1}^{T}\tilde{W}_{ti}^{3}}{\left(T^{-1}\sum_{t=1}^{T}\tilde{W}_{ti}^{2}\right)^{3/2}}, \quad i = 1, \ldots, n, \tag{4.6}$$

$$ku_i = \frac{T^{-1}\sum_{t=1}^{T}\tilde{W}_{ti}^{4}}{\left(T^{-1}\sum_{t=1}^{T}\tilde{W}_{ti}^{2}\right)^{2}}, \quad i = 1, \ldots, n, \tag{4.7}$$

where $\tilde{W}_{ti}$ denote the elements of the matrix $\tilde{W}$ defined by (3.1); in other words, sk_i and ku_i are the individual skewness and kurtosis measures based on the standardized residuals matrix. The Jarque-Bera omnibus normality test studied by Kilian and Demiroglu (2000) is:

$$JB = \frac{T}{6}SK_{\mathrm{KD}} + \frac{T}{24}KU_{\mathrm{KD}} \underset{T\to\infty}{\sim} \chi^2(2n). \tag{4.8}$$

[3]Zhou (1993) proposed simulation-based variants of these criteria to test elliptically symmetric distributions, without however providing a finite-sample theory for their application to MLR residuals _ a limitation pointed out by Zhou (1993, p. 1935, footnote 5) himself.

4.2. Extension to testing non-Gaussian distributions

To extend the above criteria beyond the Gaussian context, we shall modify them in three ways. First, we propose to use these measures in excess of their expected values under (2.2). Second, we show that for ν given, our modified test statistics are pivotal under the null hypothesis which allows to derive an exact simulation based p-value. Finally, we propose an exact combined skewness-kurtosis test.

Our approach rests on the following invariance properties regarding residuals based skewness and kurtosis tests, which we prove not only for (2.2), but for all error distributions compatible with (2.2) and either (2.4a) or (2.4b).

Theorem 4.1 DISTRIBUTION OF JB-TYPE STATISTICS IN MLR. *Under* (2.1) *and for all error distributions compatible with* (2.2) *and* (2.4a), *the multivariate skewness and kurtosis criteria* sk_i *and* ku_i, $i = 1, \ldots, n$, *defined in* (4.6) - (4.7) *are distributed, respectively, like*

$$\widetilde{sk}_i = \frac{T^{-1}\sum_{t=1}^{T}\overline{W}_{ti}^{3}}{\left(T^{-1}\sum_{t=1}^{T}\overline{W}_{ti}^{2}\right)^{3/2}}, \quad i = 1, \ldots, n, \tag{4.9}$$

$$\widetilde{ku}_i = \frac{T^{-1}\sum_{t=1}^{T}\overline{W}_{ti}^{4}}{\left(T^{-1}\sum_{t=1}^{T}\overline{W}_{ti}^{2}\right)^{2}}, \quad i = 1, \ldots, n, \tag{4.10}$$

where the $\overline{W}_{ti}$ *are the elements of the matrix* $\hat{W}S_{\hat{W}}^{-1}$ *where* $\hat{W} = MW$ *and* $S_{\hat{W}}$ *is the Cholesky factor of* $\hat{W}'\hat{W}$ *as defined in* (3.3), $M = I - X(X'X)^{-1}X'$, *and* W *is defined by* (2.2).

Theorem 4.2 DISTRIBUTION OF MARDIA-TYPE SKEWNESS AND KURTOSIS. *Under* (2.1) *and for all error distributions compatible with* (2.2) *and* (2.4b), *the multivariate skewness and kurtosis criteria* SK_{M} *and* KU_{M} *defined in* (4.1) - (4.2) *are distributed, respectively, like* $\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}d_{st}^{3}$ *and* $\frac{1}{T}\sum_{t=1}^{T}d_{tt}^{2}$, *where the* d_{st} *are the elements of the matrix* $T\hat{W}\left(\hat{W}'\hat{W}\right)^{-1}\hat{W}'$, $\hat{W} = MW$, $M = I - X(X'X)^{-1}X'$, *and* W *is defined by* (2.2).

The proof of both theorems follows immediately from Theorems **3.1** - **3.2**. On this basis, we propose the following skewness-kurtosis based statistics to test (2.2). Let

$$ESK_{\mathrm{M}}(\nu_0) = \left|SK_{\mathrm{M}} - \overline{SK}_{\mathrm{M}}(\nu_0)\right|, \tag{4.11}$$

$$EKU_{\mathrm{M}}(\nu_0) = \left|KU_{\mathrm{M}} - \overline{KU}_{\mathrm{M}}(\nu_0)\right|, \tag{4.12}$$

$$ESK_{\mathrm{KD}}(\nu_0) = \left(esk_1(\nu_0), \ldots, esk_n(\nu_0)\right)'\left(esk_1(\nu_0), \ldots, esk_n(\nu_0)\right), \tag{4.13}$$

$$EKU_{\mathrm{KD}}(\nu_0) = \left(eku_1(\nu_0), \ldots, eku_n(\nu_0)\right)'\left(eku_1(\nu_0), \ldots, eku_n(\nu_0)\right), \tag{4.14}$$

with

$$esk_i(\nu_0) = \left(sk_i(\nu_0) - \mu_{sk_i(\nu_0)}\right)/\sigma_{sk_i(\nu_0)}, \ i = 1, \ldots, n, \tag{4.15}$$

$$eku_i(\nu_0) = \left(ku_i(\nu_0) - \mu_{ku_i(\nu_0)}\right)/\sigma_{ku_i(\nu_0)}, \ i = 1, \ldots, n, \tag{4.16}$$

where $\overline{SK}_{\mathrm{M}}(\nu_0)$ and $\overline{KU}_{\mathrm{M}}(\nu_0)$ are simulation-based estimates of the mean of SK_{M} and KU_{M} given (2.2), $\mu_{sk_i(\nu_0)}$ and $\sigma_{sk_i(\nu_0)}$ are simulation-based estimates of the mean and standard deviation of $sk_i(\nu_0)$ under (2.2), $ku_i(\nu_0)$ and $\sigma_{ku_i(\nu_0)}$ are simulation-based estimates of the mean and the standard deviation of $ku_i(\nu_0)$ given (2.2). For presentation ease, we shall call these estimates "*reference simulated moments*" (RSM). We also denote by

$$E = [ESK_{\mathrm{M}}(\nu_0),\, EKU_{\mathrm{M}}(\nu_0),\, ESK_{\mathrm{KD}}(\nu_0),\, EKU_{\mathrm{KD}}(\nu_0)]' \tag{4.17}$$

the vector whose components are the test statistics just defined.

To obtain these RSM, one may proceed as follows:

A1. draw N_0 realizations of W following the distribution $\mathcal{F}(\nu_0)$ in (2.3), independently of the observed data;

A2. for each draw, construct the pivotal quantities $\hat{W}S_{\hat{W}}^{-1}$ and $T\,\hat{W}\big(\hat{W}'\hat{W}\big)^{-1}\hat{W}'$ which yield, applying theorems **4.1** - **4.2**, N_0 realizations of the statistics under consideration;

A3. the empirical moments of the latter simulated series yield the desired estimates.

4.3. Nonstandard null distributions and multi-stage MC tests

Obviously, our modified test criteria have nonstandard null distributions. In fact, the exact distributions are nonstandard even under normal null hypotheses. Yet these distributions are pivotal (in normal and non-normal contexts) and can be easily simulated which justifies the application of the Monte Carlo test technique [Dufour (2002)]. This simulation-based procedure yields a bootstrap-type exact test whenever the distribution of the underlying statistic is free of nuisance parameters under the null hypothesis. The fact that the associated analytical distributions are complicated is not a problem: all we need is the possibility of simulating the test statistic under the null hypothesis. The general methodology is described in Appendix A. When applied to the above GF criteria, it can be summarized as follows.

B1. We obtain the RSM (according to A1-A3), which are generated only once, so the next steps are conditional on these estimates.

B2. Using the RSM and applying the definitions (4.11) - (4.14) to the sample data, we find the observed value of E :

$$E^{(0)} = [ESK_{\mathrm{M}}^{(0)}(\nu_0),\, EKU_{\mathrm{M}}^{(0)}(\nu_0),\, ESK_{\mathrm{KD}}^{(0)}(\nu_0),\, EKU_{\mathrm{KD}}^{(0)}(\nu_0)]'\,. \tag{4.18}$$

B3. Independently of the RSM and $E^{(0)}$, we draw N *i.i.d.* realizations of W according to $\mathcal{F}(\nu_0)$ in (2.3), and for each of these draws, we compute the pivotal quantities $\hat{W}S_{\hat{W}}^{-1}$ and $T\,\hat{W}\big(\hat{W}'\hat{W}\big)^{-1}\hat{W}'$. N_1 is chosen so that $\alpha(N_1+1)$ is an integer.

B4. Using the same RSM as for the observed sample, the values of the statistics $ESK_{\mathrm{M}}(\nu_0)$, $EKU_{\mathrm{M}}(\nu_0)$, $ESK_{\mathrm{KD}}(\nu_0)$, $EKU_{\mathrm{KD}}(\nu_0)$ are calculated from each of these MC samples; in what follows, we will refer to these simulated values as the "*basic simulated statistics*" (BSS):

$$E^{(j)} = [ESK_{\mathrm{M}}^{(j)}(\nu_0),\, EKU_{\mathrm{M}}^{(j)}(\nu_0),\, ESK_{\mathrm{KD}}^{(j)}(\nu_0),\, EKU_{\mathrm{KD}}^{(j)}(\nu_0)]'\,, \quad j = 1, \ldots, N\,.$$

Using theorems **4.1** - **4.2**, it is easy to see that the $N+1$ vectors $E^{(j)},\ j = 0, 1, \ldots, N$ are exchangeable under the null hypothesis.

B5. We can then compute a simulated p-value, for any one of the test statistics in $E^{(0)}$:

$$\hat{p}_N[ESK_{\mathrm{M}}(\nu_0)],\ \hat{p}_N[EKU_{\mathrm{M}}(\nu_0)],\ \hat{p}_N[ESK_{\mathrm{KD}}(\nu_0)],\ \hat{p}_N[EKU_{\mathrm{KD}}(\nu_0)]\,,$$

where $\hat{p}_N[\cdot]$ is defined in Appendix A for each statistic in E [see (A.1)] and can be calculated from the rank of the observed statistic relative to the relevant BSS. The null hypothesis is rejected at level α by the test $ESK_{\mathrm{M}}(\nu_0)$ if $\hat{p}_N[ESK_{\mathrm{M}}(\nu_0)] \leq \alpha$, and similarly for the other tests. By the exchangeability of $E^{(j)},\ j = 0, 1, \ldots, N$, and provided E follows a continuous distribution, this procedure satisfies the size constraint, i.e.

$$\mathsf{P}\big[\hat{p}_N[ESK_{\mathrm{M}}(\nu_0)] \leq \alpha\big] = \alpha \tag{4.19}$$

under the null hypothesis, and similarly for all the other tests.

Because the above MC test procedure involves two nested simulations (a first one to get the reference simulated moments, and a second one to get the test statistics), we call it a *double MC test*. The procedure described above allows one to obtain individual simulated 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 Theorem **4.2**:

$$CSK_{\mathrm{M}}(\nu_0) = 1 - \min\big\{\hat{p}_N[ESK_{\mathrm{M}}(\nu_0)],\ \hat{p}_N[EKU_{\mathrm{M}}(\nu_0)]\big\}\,, \tag{4.20}$$

$$CSK_{\mathrm{KD}}(\nu_0) = 1 - \min\big\{\hat{p}_N[ESK_{\mathrm{KD}}(\nu_0)],\ \hat{p}_N[EKU_{\mathrm{KD}}(\nu_0)]\big\}. \tag{4.21}$$

The intuition here 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. For further reference on these combined tests, see Dufour and Khalaf (2002*a*).

The MC test technique may once again be applied in order to obtain an exact combined test. This can be done by using a *three-stage MC test* (or a *triple MC test*), which involves the estimation of the p-value functions $p_N(\cdot \mid \cdot)$ for individual test statistics, through a preliminary simulation experiment. The algorithm for implementing such a procedure can be described as follows.

C1. Generate a set of reference simulated moments (according to A1-A3), the observed value of $E^{(0)}$ in (4.18), and the N corresponding BSS (following B1-B4).

C2. For each test statistic considered, obtain the p-value functions determined by the BSS (generated at step C1): $p_N(S^{(0)}\,;\,S)$, for $S = ESK_{\mathrm{M}}(\nu_0),\ EKU_{\mathrm{M}}(\nu_0),\ ESK_{\mathrm{KD}}(\nu_0),\ EKU_{\mathrm{KD}}(\nu_0)$, where the function $p_N(S^{(0)}\,;\,S)$ is defined in Appendix A.

C3. Independently of the previous RSM, BSS and $E^{(0)}$, generate N_1 additional *i.i.d.* realizations of W according to $\mathcal{F}(\nu_0)$ in (2.3), and for each draw, compute the pivotal quantities $\hat{W}S_{\hat{W}}^{-1}$ and $T\,\hat{W}\big(\hat{W}'\hat{W}\big)^{-1}\hat{W}'$. N_1 is chosen so that $\alpha(N_1+1)$ is an integer.

C4. Using the RSM and the N_1 draws generated at steps C1 and C3, compute the corresponding simulated statistics:

$$EE^{(l)} = [ESK_{\mathrm{M}}^{(l)}(\nu_0),\ EKU_{\mathrm{M}}^{(l)}(\nu_0),\ ESK_{\mathrm{KD}}^{(l)}(\nu_0),\ EKU_{\mathrm{KD}}^{(l)}(\nu_0)]'\,, \quad l = 1,\, \ldots\,,\, N_1.$$

C5. Using the p-value functions $p_N(\cdot\,;\,\cdot)$ obtained at step C2 (and based on the BSS generated at step C1), evaluate the simulated p-values for the observed and the N_1 additional simulated statistics: $\hat{p}_N^{(l)}[S] \equiv p_N(S^{(l)}\,;\,S)$, $l = 0,\, 1,\, \ldots\,,\, N_1$, for $S = ESK_{\mathrm{M}}(\nu_0),\ EKU_{\mathrm{M}}(\nu_0),\ ESK_{\mathrm{KD}}(\nu_0),\ EKU_{\mathrm{KD}}(\nu_0)$.

C6. From the latter, compute the corresponding values of the combined test statistics:

$$\begin{aligned} CSK_{\mathrm{M}}^{(l)}(\nu_0) &= 1 - \min\big\{\hat{p}_N[ESK_{\mathrm{M}}^{(l)}(\nu_0)],\ \hat{p}_N[EKU_{\mathrm{M}}^{(l)}(\nu_0)]\big\}\,, \quad l = 0,\, 1,\, \ldots\,,\, N_1\,, \\ CSK_{\mathrm{KD}}^{(l)}(\nu_0) &= 1 - \min\big\{\hat{p}_N[ESK_{\mathrm{KD}}^{(l)}(\nu_0)],\ \hat{p}_N[EKU_{\mathrm{KD}}^{(l)}(\nu_0)]\big\}, \quad l = 0,\, 1,\, \ldots\,,\, N_1\,. \end{aligned}$$

Again, it is easy to see that the vectors $\big(CSK_{\mathrm{M}}^{(l)}(\nu_0),\ CSK_{\mathrm{KD}}^{(l)}(\nu_0)\big)$, $l = 0,\, 1,\, \ldots\,,\, N_1$, are exchangeable.

C7. The combined test $CSK_{\mathrm{M}}(\nu_0)$ rejects the null hypothesis at level α if $\hat{p}_{N_1}[CSK_{\mathrm{M}}(\nu_0)] \equiv p_{N_1}\big(CSK_{\mathrm{M}}^{(0)}\,;\,CSK_{\mathrm{M}}(\nu_0)\big) \leq \alpha$, where the p-value function $p_{N_1}(\cdot\,|\,\cdot)$ is based on the simulated variables $CSK_{\mathrm{M}}^{(l)}(\nu_0), l = 0,\, 1,\, \ldots\,,\, N_1$; the rule is similar for the test based on $CSK_{\mathrm{KD}}(\nu_0)$.

The test with critical regions $\hat{p}_{N_1}[CSK_{\mathrm{M}}(\nu_0)] \leq \alpha$ has level α, because the variables $CSK_{\mathrm{M}}^{(l)}(\nu_0)$, $l = 0,\, 1,\, \ldots\,,\, N$, are exchangeable under the null hypothesis. The same holds for the test with critical region $\hat{p}_{N_1}[CSK_{\mathrm{KD}}(\nu_0)] \leq \alpha$.

We have also studied the following modified version of the omnibus-type tests based on the sum of the skewness and kurtosis statistics:

$$\widetilde{MD} = \begin{bmatrix} SK_{\mathrm{M}} - \overline{SK}_{\mathrm{M}}(\nu_0) \\ KU_{\mathrm{M}} - \overline{KU}_{\mathrm{M}}(\nu_0) \end{bmatrix}' \bar{\varDelta}_{\mathrm{M}}^{-1}(\nu_0) \begin{bmatrix} SK_{\mathrm{M}} - \overline{SK}_{\mathrm{M}}(\nu_0) \\ KU_{\mathrm{M}} - \overline{KU}_{\mathrm{M}}(\nu_0) \end{bmatrix}, \tag{4.22}$$

$$\widetilde{JB} = ESK_{\mathrm{KD}}(\nu_0) + EKU_{\mathrm{KD}}(\nu_0)\,, \tag{4.23}$$

where $\bar{\varDelta}_{\mathrm{M}}(\nu_0)$ is a simulation-based estimate of the covariance of SK_{M} and KU_{M}, which can obtained as outlined in A1-A3; in this case, in addition to the empirical means and standard deviations of simulated SK_{M} and KU_{M} series, we also obtain these empirical covariances. These statistics are obviously less expensive to simulate than the ones based on the smallest p-values. MC p-values can be obtained for them in a way similar to the one described in B1-B5, except that the matrix $\bar{\Delta}_{\mathrm{M}}(\nu_0)$ now belongs to the set of moments to be estimated by simulation.

4.4. The unknown nuisance parameter case

So far, we have treated the case where the distributional parameter ν is known. To account for an unknown ν, we obtain a confidence set estimate for this parameter which "inverts" the above GF tests. Specifically, the confidence set corresponds to the set of ν_0 values which are not rejected by the GF test for (2.2) where $\nu = \nu_0$ for known ν_0. This leads to a formal estimate for the distributions which best fit the data. As we will show in the next section, this estimate may prove to be very useful for other testing problems regarding the regression under consideration, for it may be easily shown that the usual test statistics for hypothesis on the regression coefficients or error terms will also depend on ν; see Beaulieu, Dufour and Khalaf (2001*b*, 2001*a*). When the confidence set for ν is empty, the distributional family (2.2) is rejected.

5. Simulation experiment

We conducted a small-scale simulation experiment to assess the performance of the GF tests. The model considered is (2.1) with three designs. The first, denoted Design I, includes $n = 12$ equations and the following regressor matrix:

$$X_I = \left[\iota_n \vdots X_{(1)}\right],$$

where $X_{(1)}$ is a $T \times 1$ standard normal variate, with $T = 60$. The second, denoted Design II, includes $n = 12$ equations and, in addition to the regressors of design I, dummy variable regressors over a window covering the 20% sample endpoints; the associated regressor matrix takes the form:

$$X_{II} = \left[\iota_n \vdots X_{(1)} \vdots X_{(2)}\right], \quad X_{(2)} = \begin{bmatrix} X_{(2,1)} \\ I_{k_1} \end{bmatrix}$$

where $X_{(1)}$ and $X_{(2,1)}$ are $T \times 1$ and $(T - k_1) \times k_1$ standard normal variates and $k_1 = INT(.02 \times T)$, with $T = 60$. The third design (III), uses the same regressor matrix X_I but includes $n = 40$ equations. In all designs, $T = 60$, $N_0 = 1000$, $N = N_1 = 999$ and the number of simulations in each experiment is 1000. Because of location-scale invariance, all the above tests were applied to the residuals generated as $\hat{U} = MW$, hence there was no need to specify values for the regression coefficients and error covariances.

We studied the following sets of hypotheses. H_0: $W_t \sim$ multivariate normal, against: (i) $W_t \sim t(\kappa)$, with $\kappa = 5$, 10, 20, 30, 40, 50, (ii) $W_t \sim Stb(\alpha_s)$, with $\alpha_s = 1.8$, 1.85, 1.9, 1.95, 1, 98, 2.0 and $\beta_s = 0$, and (iii) $W_t \sim$ multivariate mixture of normals

$$W \sim Mix(\pi, \omega) \Leftrightarrow W_t = \pi Z_{1t} + (1 - \pi) Z_{3t}\ ,$$

Table 1. Size of multinormality tests

Design	Mardia-type					
	MSK_{asy}	MSK_{MC}	ESK_{M}	EKU_{M}	CSK_{M}	$\widetilde{MD}$
I	.022	.054	.052	.047	.049	.052
II	1.00	.052	.053	.039	.043	.039
III	0.00	.044	.034	.046	.040	.039
	JB-type					
	JB_{asy}	JB_{MC}	ESK_{KD}	EKU_{KD}	CSK_{KD}	
I	.064	.053	.051	.050	.058	
II	.521	.048	.048	.051	.052	
III	.056	.056	.061	.054	.051	

Note _ This table reports the actual rejection frequencies based on 5 percent critical values under the asymptotic and finite distributions. Design I refers to 12 equations of 60 observations each and a regressor matrix including a constant and a $T \times 1$ standard normal variate. Design II refers to 12 equations of 60 observations each but in this case the regressor matrix contains a constant, a $T \times 1$ standard normal variate and a $(T - k_1) \times k_1$ standard normal variate where $k_1 = INT(.02 \times T)$. Design III is the same as Design I but includes 40 equations instead of 12. $ESK_{\mathrm{M}}, EKU_{\mathrm{M}}, ESK_{\mathrm{KD}}$ and EKU_{KD} refer to the excess skewness and excess kurtosis criteria defined in (4.11) - (4.14). CSK_{M} and CSK_{KD} refer to the min-p-value combined skewness/kurtosis criteria (4.20) - (4.21). $\widetilde{MD}$ denotes our dependence-corrected version of Mardia's tests (4.22). All the latter tests are MC tests with 999 replications ($N = N_1 = 999$, $N_0 = 1000$). MSK_{asy} and JB_{asy} refer to original tests (4.3) - (4.8) and MSK_{MC} and JB_{MC} are their MC versions. The number of simulations in each experiment is 1000.

where $Z_{1t} \sim N[0,\, I_n]$, $Z_{3t} \sim N[0,\, \omega I_n]$ and is independent of Z_{1t}, and $0 < \pi < 1$; we use $\pi = .5$ and $\omega = 3,\, 2.5,\, 2,\, 1.5$. The multivariate $t(\kappa)$ is generated as follows

$$W \sim t(\kappa) \Leftrightarrow W_t = Z_{1t}/(Z_{2t}/\kappa)^{1/2}\ , \tag{5.1}$$

where Z_{1t} is multivariate normal $(0,\, I_n)$ and Z_{2t} is a $\chi^2(\kappa)$ variate independent from Z_{1t}; stable errors are drawn componentwise, applying Weron (1996). The results are reported in tables 1 to 4.

First, tables 2 - 4 reveal that available asymptotic tests are completely unreliable. Indeed, in Design II, Mardia's size is 1.0 and is zero in design 3; even in Design I, the asymptotic test is undersized. The JB-type test is also seriously oversized in Design II. We thus only analyze the power of the MC tests; we note however that the size problems we observed with Mardia's asymptotic test translated into very low power, with empirical rejections not exceeding the nominal size.

In terms of power, our results over all designs presented in tables 2 - 4, can be summarized as follows. For elliptical families, Mardia-type tests are superior to the JB-type; this observation is important given the relevance of ellipticity in asset pricing applications. The JB-type test displayed better power for detecting errors whose marginal distributions are from the stable family. We note that although all alternatives studied are symmetric, the skewness tests show high power. This is because the null is parametric; the cut-off points of the skewness tests are thus derived under

Table 2. Power of multinormality tests: Design I

Design I	Mardia-type					JB-type			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	ESK_{M}	EKU_{M}	CSK_{M}	$\widetilde{MD}$	MSK	ESK_{KD}	EKU_{KD}	CSK_{KD}	JB
$t(5)$	.999	1.0	1.0	1.0	.999	.682	.825	.807	.819
$t(10)$	.920	.980	.972	.963	.939	.302	.403	.383	.390
$t(20)$	.542	.663	.636	.591	.600	.160	.190	.182	.173
$t(30)$	.320	.425	.407	.378	.373	.108	.117	.120	.113
$t(40)$	.232	.307	.279	.262	.266	.099	.099	.098	.101
$t(50)$	.180	.232	.221	.187	.231	.074	.084	.083	.081
$Mix(.5, 3)$	1.0	1.0	1.0	1.0	1.0	.768	.970	.961	.954
$Mix(.5, 2.5)$	1.0	1.0	1.0	1.0	1.0	.675	.891	.872	.873
$Mix(.5, 2)$	1.0	1.0	1.0	1.0	1.0	.466	.661	.625	.617
$Mix(.5, 1.5)$	.763	.889	.874	.851	.803	.203	.235	.227	.239
$Stb(1.8)$	.965	.971	.971	.967	.970	.991	.997	.997	.998
$Stb(1.85)$	.911	.930	.926	.914	.922	.967	.984	.980	.983
$Stb(1.90)$	.763	.782	.784	.761	.789	.873	.923	.918	.925
$Stb(1.95)$	.476	.488	.494	.474	.510	.626	.687	.678	.686
$Stb(1.98)$	.217	.215	.223	.217	.244	.331	.361	.636	.636

Note _ This table reports the actual rejection frequencies based on the 5 percent critical values under the finite-sample distributions. Design I refers to 12 equations of 60 observations and a regressor matrix including a constant and a $T \times 1$ standard normal variate. Design II refers to 12 equations of 60 observations but in this case the regressor matrix contains a constant, a $T \times 1$ standard normal variate and a $(T - k_1) \times k_1$ standard normal variate where $k_1 = INT(.02 \times T)$. Design III is the same as Design I but includes 40 equations instead of 12. t stands for the Student distribution, Mix the mixture of normal distribution and Stb the stable distribution. Numbers in parentheses present the chosen values for the nuisance parameters in these distributions. ESK_{M}, EKU_{M} ESK_{KD} and EKU_{KD} refer to the excess skewness and excess kurtosis criteria defined in (4.11) - (4.14). CSK_{M} and CSK_{KD} refer to the min-p-value combined skewness/kurtosis criteria (4.20) - (4.21). $\widetilde{MD}$ denotes our dependence-corrected version of Mardia's tests (4.22). MSK and JB refer to original tests (4.3) - (4.8) in their MC versions. All MC tests use 999 replications ($N = N_1 = 999$, $N_0 = 1000$).The number of simulations in each experiment is 1000.

Table 3. Power of multinormality tests: Design II

Design II	Mardia-type					JB-type			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	ESK_{M}	EKU_{M}	CSK_{M}	$\widetilde{MD}$	MSK	ESK_{KD}	EKU_{KD}	CSK_{KD}	JB
$t(5)$	.993	1.0	.999	.997	.999	.596	.686	.686	.748
$t(10)$	.844	.900	.896	.868	.897	.250	.314	.293	.355
$t(20)$	.433	.516	.493	.447	.510	.128	.153	.148	.163
$t(30)$	.255	.324	.307	.276	.325	.099	.102	.095	.109
$t(40)$	.178	.237	.221	.198	.240	.085	.093	.090	.098
$t(50)$	.146	.182	.163	.147	.200	.072	.080	.078	.087
$Mix(.5,\ 3)$	1.0	1.0	1.0	1.0	1.0	.706	.901	.871	.946
$Mix(.5,\ 2.5)$	1.0	1.0	1.0	1.0	1.0	.582	.772	.756	.856
$Mix(.5,\ 2)$	.994	1.0	1.0	1.0	.999	.394	.506	.485	.602
$Mix(.5,\ 1.5)$	.652	.783	.761	.729	.747	.139	.178	.176	.213
$Stb(0,\ 1.8)$	.919	.932	.932	.920	.936	.963	.982	.984	.989
$Stb(0,\ 1.85)$	.810	.842	.836	.805	.851	.909	.946	.939	.954
$Stb(0,\ 1.90)$	.626	.658	.657	.632	.674	.807	.843	.836	.856
$Stb(0,\ 1.95)$	.371	.378	.384	.354	.402	.519	.584	.563	.582
$Stb(0,\ 1.98)$	.184	.174	.173	.171	.190	.269	.303	.292	.299

Table 4. Power of multinormality tests: Design III

Design III	Mardia-type					JB-type			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	ESK_{M}	EKU_{M}	CSK_{M}	$\widetilde{MD}$	MSK	ESK_{KD}	EKU_{KD}	CSK_{KD}	JB
$t(5)$	1.0	1.0	1.0	1.0	1.0	.584	.744	.735	.735
$t(10)$	.986	.987	.988	.978	.987	.238	.326	.297	.326
$t(20)$	.738	.776	.763	.703	.771	.107	.138	.118	.134
$t(30)$	.484	.536	.509	.447	.519	.091	.100	.103	.107
$t(40)$	.346	.380	.365	.305	.381	.074	.083	.079	.070
$t(50)$	.247	.281	.271	.230	.279	.063	.074	.079	.072
$Mix(.5,\ 3)$	1.0	1.0	1.0	1.0	1.0	.844	.973	.965	.965
$Mix(.5,\ 2.5)$	1.0	1.0	1.0	1.0	1.0	.684	.868	.856	.859
$Mix(.5,\ 2)$	1.0	1.0	1.0	1.0	1.0	.401	.549	.530	.528
$Mix(.5,\ 1.5)$	.957	.968	.962	.948	.961	.132	.162	.154	.151
$Stb(1.8)$	.958	.964	.959	.943	.965	1.0	1.0	1.0	1.0
$Stb(1.85)$	.857	.686	.870	.825	.878	.999	.998	.999	1.0
$Stb(1.90)$	.637	.643	.633	.571	.663	.991	.994	.994	.994
$Stb(1.95)$	.315	.323	.326	.276	.343	.861	.909	.989	.910
$Stb(1.98)$	.129	.129	.134	.115	.142	.515	.584	.562	.580

Table 5. 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).

symmetry and normal kurtosis. This problem is well known in statistic [see Horswell and Looney (1993)] and must be emphasized, given the importance empirical practitioners attribute to the skewness coefficient. We also note that our min-p-value combined Mardia test was most powerful in many instances.

6. Empirical application

Our empirical analysis focuses on the asset pricing model (2.9) 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 5 provides a list of the different sectors used as well as the SIC codes included in the analysis.[4] For each month the industry portfolios comprise those firms for which the return, price per common share and number of shares outstanding are recorded by CRSP. Furthermore, portfolios are value-weighted in each month. The market return is measured by 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 reported in Tables 6–10.

Regarding normality tests, Table 6 reveals the following. Although we are dealing with monthly data, normality is definitely rejected except in the last subsample (1990-95) where the smallest p-value is 9.1%. Furthermore, both excess skewness and excess kurtosis are evident. The MC version

[4] 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 6. Multinormality tests

Sample	Mardia-type					JB-type			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	ESK_{M}	EKU_{M}	MSK	CSK_{M}	$\widetilde{MD}$	ESK_{KD}	EKU_{KD}	JB	CSK_{KD}
1927-30	.001	.001	.001	.001	.001	.004	.006	.007	.008
1931-35	.001	.001	.001	.001	.001	.001	.001	.001	.001
1936-40	.001	.001	.001	.001	.001	.005	.048	.020	.011
1941-45	**.004**	**.002**	**.004**	**.004**	.004	**.378**	**.092**	**.199**	**.141**
1946-50	.001	.001	.001	.001	.001	.003	.009	.005	.004
1951-55	.001	.001	.001	.001	.001	.002	.003	.003	.005
1956-60	**.024**	**.003**	**.015**	**.003**	.016	**.700**	**.333**	**.603**	**.474**
1961-65	**.594**	**.479**	**.736**	**.631**	.151	**.037**	**.014**	**.008**	**.029**
1966-70	**.011**	**.002**	**.011**	**.004**	.005	**.632**	**.559**	**.759**	**.728**
1971-75	**.001**	.002	.001	.001	.001	**.554**	.015	.060	.029
1976-80	.001	.001	.001	.001	.001	.013	.015	.012	.030
1981-85	**.001**	**.002**	**.001**	**.001**	.002	**.932**	**.096**	**.305**	**.154**
1986-90	.028	.020	.024	.030	.061	.006	.024	.009	.007
1991-95	.177	.311	.917	.239	.408	.065	.425	.127	.091

Note _ Numbers shown are MC p-values. Columns (1), (2), (6) and (7) refer to the excess skewness and excess kurtosis criteria defined in (4.11) - (4.14). Columns (4) and (9) refer to the min-p-value combined skewness/kurtosis criteria (4.20) - (4.21). Columns (3) and (8) refer to MC versions of the original Mardia and JB-type tests (4.3) - (4.8). Column (5) reports our dependence-corrected version of Mardia's tests (4.22). The p-values in bold highlight cases where the various tests yield conflicting decisions at the 5% level. All MC tests use 999 replications ($N = N_1 = 999$, $N_0 = 1000$).

of the omnibus tests (based on adding up the skewness and kurtosis criteria) and the *min p-value* based combined tests seem to yield the same decision. It is however noteworthy that the Mardia-type and the Kilian-Demiroglu JB-type tests yield conflicting decisions in several cases: for the 1941-50, 1956-60, 1966-70 and 1981-85, JB and CSK_{KD} are not significant, whereas our Mardia-type tests are significant; conversely, in 1961-65, both JB and CSK_{KD} are significant yet all of the Mardia-type tests fail to reject the normal null.

These results seem to suggest that it is worthwhile to consider strategies which combine both type of tests.[5] For example, exact MC joint tests may be obtained using a criterion of the form:

$$CSK_{\mathrm{M/KD}} = 1 - \min\{\widehat{p}(CSK_{\mathrm{M}}),\ \widehat{p}(CSK_{\mathrm{KD}})\}\,. \tag{6.1}$$

Indeed, the flexibility of the MC test method allows one to consider combinations that would be hard to justify applying standard asymptotic strategies; for further references on combining non-independent tests, see Dufour and Khalaf (2002*b*), Dufour, Khalaf, Bernard and Genest (2003) and Dufour and Khalaf (2002*a*).

Let us now turn to the GF tests for Student t and stable distributions. Tables 7 to 10 report the test results in the form of confidence sets for the distributional parameters; these correspond to the

[5]In this regard, see Horswell and Looney (1992) on the cost (in terms of size) of combining normality tests.

Table 7. Multivariate t distributions:
combined-tests based confidence sets for the degrees-of-freedom parameter

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample	Student t, 2.5% level				Student t, 5% level			
	CSK_{M}	CSK_{KD}	$\widetilde{MD}$	$\widetilde{JB}$	CSK_{M}	CSK_{KD}	$\widetilde{MD}$	$\widetilde{JB}$
1927-30	$3-12$	$1-17$	$2-14$	$1-18$	$3-12$	$2-14$	$3-12$	$1-11$
1931-35	$3-8$	$2-6$	$3-8$	$1-6$	$3-7$	$2-3$	$3-8$	$1-3$
1936-40	$4-29$	$2-29$	$4-33$	$1-30$	$5-24$	$2-22$	$5-26$	$1-29$
1941-45	≥ 5	≥ 2	≥ 5	≥ 1	$6-40$	≥ 2	≥ 5	≥ 1
1946-50	$4-31$	$2-20$	$5-34$	$1-27$	$4-24$	$2-13$	$5-29$	$1-19$
1951-55	$5-34$	$2-14$	$4-39$	$1-13$	$5-29$	$2-9$	$5-31$	$1-7$
1956-60	≥ 5	$2-34$	≥ 5	$1-34$	≥ 6	$2-34$	≥ 5	$1-34$
1961-65	≥ 7	≥ 2	≥ 6	≥ 1	$12-42$	≥ 2	≥ 7	$1-26$
1966-70	≥ 5	≥ 2	≥ 4	≥ 1	$6-42$	≥ 2	≥ 5	≥ 1
1971-75	$4-28$	≥ 2	$5-29$	≥ 1	$5-21$	≥ 2	$6-22$	≥ 1
1976-80	$4-17$	≥ 2	$3-19$	≥ 1	$4-16$	≥ 2	$4-18$	$1-41$
1981-85	$5-33$	≥ 2	$5-41$	≥ 1	$5-26$	≥ 2	$6-31$	≥ 1
1986-90	$5-41$	$2-30$	$5-41$	$1-38$	$7-41$	$2-19$	$6-41$	$1-26$
1991-95	≥ 15	≥ 2	≥ 9	≥ 1	$24-42$	≥ 2	≥ 14	≥ 1

Note _ Numbers shown are values of the degrees-of-freedom parameter κ not rejected by the MC-GF tests. Columns (1), (3), (5) and (7) pertain to our combined Mardia-type statistic (4.20) - (4.22). Columns (2), (4), (6) and (8) are based on the combined Kilian-Demiroglu JB-type statistic (4.21) - (4.23).

Table 8. Multivariate symmetric stable distributions:
confidence sets for the kurtosis parameter based on combined tests

		(1)	(2)	(3)	(4)
Sample	Test level	Stable distributions with $\beta_s = 0$			
		CSK_{M}	CSK_{KD}	$\widetilde{MD}$	$\widetilde{JB}$
1927-30	2.5%	$1.38 - 1.96$	$1.1 - 1.99$	$1.28 - 1.88$	$.9 - 1.99$
1931-35	2.5%	$1.34 - 1.92$	$1.1 - 1.99$	$1.30 - 1.90$	$.9 - 1.99$
1936-40	2.5%	$1.56 - 1.98$	$1.1 - 1.99$	$1.46 - 1.98$	$.9 - 1.99$
1941-45	2.5%	$1.58 - 1.98$	$1.1 - 1.99$	$1.66 - 1.98$	$.9 - 1.99$
1946-50	2.5%	$1.56 - 1.98$	$1.1 - 1.99$	$1.58 - 1.98$	$.9 - 1.99$
1951-55	2.5%	$1.56 - 1.98$	$1.1 - 1.99$	$1.64 - 1.98$	$.9 - 1.99$
1956-60	2.5%	$1.56 - 1.98$	$1.1 - 1.99$	$1.66 - 1.98$	$.9 - 1.99$
1961-65	2.5%	$1.66 - 2.0$	$1.1 - 2.0$	$1.56 - 2.0$	$.9 - 2.0$
1966-70	2.5%	$1.56 - 1.98$	$1.1 - 1.99$	$1.48; 1.98$	$.9 - 1.99$
1971-75	2.5%	$1.56 - 1.98$	$1.1 - 1.99$	$1.54 - 1.98$	$.9 - 1.99$
1976-80	2.5%	$1.5 - 1.98$	$1.1 - 1.99$	$1.44 - 1.98$	$.9 - 1.99$
1981-85	2.5%	$1.56 - 1.98$	$1.1 - 1.99$	$1.54 - 1.98$	$.9 - 1.2$
1986-90	2.5%	$1.62 - 2.0$	$1.1 - 2.0$	$1.60 - 2.0$	$.9 - 2.0$
1991-95	2.5%	$1.7 - 2.0$	$1.1 - 2.0$	$1.70 - 2.0$	$.9 - 2.0$
1927-30	5.0%	$1.46 - 1.92$	$1.1 - 1.99$	$1.36 - 1.84$	$.9 - 1.99$
1931-35	5.0%	$1.42 - 1.90$	$1.1 - 1.99$	$1.38 - 1.88$	$.9 - 1.99$
1936-40	5.0%	$1.64 - 1.98$	$1.2 - 1.99$	$1.56 - 1.98$	$.9 - 1.99$
1941-45	5.0%	$1.66 - 1.98$	$1.2 - 1.99$	$1.58 - 1.98$	$.9 - 1.99$
1946-50	5.0%	$1.58 - 1.98$	$1.2 - 1.99$	$1.58 - 1.98$	$.9 - 1.99$
1951-55	5.0%	$1.64 - 1.98$	$1.1 - 1.99$	$1.56 - 1.98$	$.9 - 1.99$
1956-60	5.0%	$1.66 - 1.98$	$1.2 - 1.99$	$1.58 - 1.98$	$.9 - 1.99$
1961-65	5.0%	$1.74 - 2.0$	$1.2 - 2.0$	$1.66 - 2.0$	$.9 - 2.0$
1966-70	5.0%	$1.66 - 1.98$	$1.2 - 1.99$	$1.56 - 1.98$	$.9 - 1.99$
1971-75	5.0%	$1.62 - 1.98$	$1.2 - 1.99$	$1.58 - 1.98$	$.9 - 1.99$
1976-80	5.0%	$1.58 - 1.96$	$1.1 - 1.99$	$1.46 - 1.96$	$.9 - 1.99$
1981-85	5.0%	$1.66 - 1.98$	$1.2 - 1.99$	$1.58 - 1.98$	$.9 - 1.99$
1986-90	5.0%	$1.7 - 1.98$	$1.2 - 1.99$	$1.64 - 2.0$	$.9 - 2.0$
1991-95	5.0%	$1.78 - 2.0$	$1.2 - 2.0$	$1.78 - 2.0$	$.9 - 2.0$

Note _ Numbers shown are values of the kurtosis parameter α_s not rejected by the MC GF tests. Columns (1) and (3) pertain to our combined Mardia-type statistic (4.20) - (4.22). Columns (2) and (4) are based on the combined Kilian-Demiroglu JB-type statistic (4.21) - (4.23).

Table 9. Multivariate t and stable distributions:
skewness-based confidence sets for distributional parameters

		(1)	(2)	(3)	(4)
Sample	Test level	Student t		Stable ($\beta_s = 0$)	
		ESK_{M}	ESK_{KD}	ESK_{M}	ESK_{KD}
1927-30	2.5%	$3-15$	$1-16$	$1.38-1.98$	$.9-1.99$
1931-35	2.5%	$2-7$	$1-7$	$1.34-1.96$	$.9-1.99$
1936-40	2.5%	$4-25$	$1-31$	$1.46-1.98$	$.9-1.99$
1941-45	2.5%	$4-42$	$1-42$	$1.58-1.98$	$.9-1.99$
1946-50	2.5%	$4-26$	$1-18$	$1.46-1.98$	$.9-1.99$
1951-55	2.5%	$4-37$	$1-13$	$1.54-1.98$	$.9-1.99$
1956-60	2.5%	$5-42$	≥ 1	$1.54-1.98$	$.9-1.99$
1961-65	2.5%	$8-42$	≥ 1	$1.60-2.0$	$.9-2.0$
1966-70	2.5%	$5-42$	≥ 1	$1.54-1.98$	$.9-1.99$
1971-75	2.5%	$4-22$	≥ 1	$1.46-1.98$	$.9-1.99$
1976-80	2.5%	$3-17$	$1-39$	$1.46-1.98$	$.9-1.99$
1981-85	2.5%	$4-29$	≥ 1	$1.58-1.98$	$.9-1.99$
1986-90	2.5%	$4-41$	$1-19$	$1.56-1.98$	$.9-1.99$
1991-95	2.5%	≥ 12	≥ 1	$1.68-2.0$	$.9-2.0$
1927-30	5.0%	$3-13$	$1-10$	$1.48-1.96$	$.9-1.99$
1931-35	5.0%	$3-7$	$1-6$	$1.42-1.94$	$.9-1.99$
1936-40	5.0%	$4-22$	$1-20$	$1.58-1.98$	$.9-1.99$
1941-45	5.0%	$5-36$	≥ 1	$1.6-1.98$	$.9-1.99$
1945-50	5.0%	$4-22$	$1-13$	$1.58-1.98$	$.9-1.99$
1951-55	5.0%	$5-26$	$1-8$	$1.60-1.98$	$.9-1.99$
1956-60	5.0%	$6-42$	≥ 1	$1.64-1.98$	$.9-1.99$
1961-65	5.0%	$14-42$	≥ 1	$1.78-2.0$	$.9-2.0$
1966-70	5.0%	$6-42$	≥ 1	$1.60-1.98$	$.9-1.99$
1971-75	5.0%	$4-19$	≥ 1	$1.60-1.98$	$.9-1.99$
1976-80	5.0%	$4-15$	$1-25$	$1.54-1.98$	$.9-1.99$
1981-85	5.0%	$2-26$	≥ 1	$1.60-1.98$	$.9-1.99$
1986-90	5.0%	$5-41$	$1-12$	$1.66-1.98$	$.9-1.99$
1991-95	5.0%	≥ 35	≥ 1	$1.78-2.0$	$.9-1.99$

Note _ Numbers shown are values of the distributional parameter [κ and α_s respectively] not rejected by the MC-GF tests. Columns (1) and (3) pertain to our extension of the Mardia-type statistic (4.11). Columns (2) and (4) are based on our extension of aggregated individual skewness measures (4.13).

Table 10. Multivariate t and stable distributions:
kurtosis-based confidence sets for distributional parameters

		(1)	(2)	(4)	(5)
Sample	Test level	Student t		Stable ($\beta_s = 0$)	
		EKU_{M}	EKU_{KD}	EKU_{M}	EKU_{KD}
1927-30	2.5%	$3 - 12$	$2 - 28$	$1.38 - 1.96$	$1.1 - 1.99$
1931-35	2.5%	$3 - 7$	$2 - 6$	$1.4 - 1.92$	$1.1 - 1.99$
1936-40	2.5%	$5 - 27$	≥ 2	$1.54 - 1.98$	$1.1 - 1.99$
1941-45	2.5%	≥ 5	≥ 2	$1.54 - 1.98$	$1.1 - 1.99$
1946-50	2.5%	$5 - 34$	≥ 2	$1.56 - 1.98$	$1.1 - 1.99$
1951-55	2.5%	$5 - 29$	$2 - 13$	$1.56 - 1.98$	$1.1 - 1.99$
1956-60	2.5%	≥ 5	≥ 2	$1.60 - 1.98$	$1.1 - 1.99$
1961-65	2.5%	≥ 9	≥ 2	$1.56 - 2.0$	$1.1 - 2.0$
1966-70	2.5%	≥ 5	≥ 2	$1.58 - 1.98$	$1.1 - 1.99$
1971-75	2.5%	$5 - 34$	≥ 2	$1.56 - 1.98$	$1.1 - 1.99$
1976-80	2.5%	$4 - 17$	≥ 2	$1.46 - 1.98$	$1.1 - 1.99$
1981-85	2.5%	$5 - 39$	≥ 2	$1.58 - 1.98$	$1.1 - 1.99$
1986-90	2.5%	$5 - 41$	≥ 2	$1.64 - 2.0$	$1.2 - 1.99$
1991-95	2.5%	≥ 20	≥ 2	$1.76 - 2.0$	$1.1 - 2.0$
1927-30	5.0%	$3 - 11$	$2 - 16$	$1.48 - 1.92$	$1.1 - 1.99$
1931-35	5.0%	$3 - 7$	≥ 2	$1.44 - 1.90$	$1.1 - 1.99$
1936-40	5.0%	$5 - 22$	≥ 2	$1.60 - 1.98$	$1.2 - 1.99$
1941-45	5.0%	$6 - 39$	≥ 2	$1.58 - 1.98$	$1.2 - 1.99$
1945-50	5.0%	$5 - 27$	$2 - 26$	$1.64 - 1.98$	$1.2 - 1.99$
1951-55	5.0%	$5 - 26$	$2 - 7$	$1.64 - 1.98$	$1.1 - 1.99$
1956-60	5.0%	≥ 6	≥ 2	$1.70 - 1.98$	$1.2 - 1.99$
1961-65	5.0%	≥ 11	$2 - 26$	$1.78 - 2.0$	$1.2 - 2.0$
1966-70	5.0%	$6 - 39$	≥ 2	$1.64 - 1.98$	$1.2 - 1.99$
1971-75	5.0%	$5 - 26$	≥ 2	$1.60 - 1.98$	$1.2 - 1.99$
1976-80	5.0%	$4 - 15$	≥ 2	$1.58 - 1.96$	$1.2 - 1.99$
1981-85	5.0%	$6 - 34$	≥ 2	$1.64 - 1.98$	$1.2 - 1.99$
1986-90	5.0%	$7 - 41$	$2 - 41$	$1.70 - 1.98$	$1.2 - 1.99$
1991-95	5.0%	≥ 35	≥ 2	$1.80 - 2.0$	$1.2 - 2.0$

Note _ Numbers shown are values of the distributional parameter (κ and α_s respectively) not rejected by the MC GF tests. Columns (1) and (3) pertain to our extension of the Mardia-type statistic (4.12). Columns (2) and (4) are based on our extension of aggregated individual kurtosis measures (4.14).

parameters not rejected by the GF tests considered. From an empirical perspective, the most relevant result from these tables is the following: Mardia-type confidence sets are the tightest. Specifically, smaller values of κ and α_s (which signal more extreme kurtosis) are more easily rejected with Mardia-type tests, although, in a few cases, larger values of κ and α_s (which imply tails approaching the normal) are more easily rejected with JB-type tests. Following our conclusions regarding normality tests, we see that combining both type of test statistics may yield more powerful procedures; this is easily achieved in a MC tests framework. The min-p-value KD yields lower p-values and (tighter confidence sets) than its omnibus counterpart. This is particularly noticeable in the stable distribution case where $\widetilde{JB}$ suggests that $.9 \leq \alpha_s \leq 1$ [which signals severely extreme kurtosis] is compatible with our data, whereas all other tests reject $\alpha_s \leq 1$.[6]

7. Conclusion

In this paper, we have proposed a class of exact procedures for testing goodness-of-fit of the error distribution in MLR models. The test statistics are based on multivariate skewness and kurtosis measures computed on appropriately standardized multivariate residuals, so their null distributions do not depend on the unknown error covariance matrix (or the regression coefficients). To deal with the fact that the statistics may have analytically intractable null distributions, the tests are implemented using simple, double and triple Monte Carlo test methods. Special cases considered include testing multivariate normal, Student t, normal mixtures and stable error models. In the Gaussian case, the procedures proposed include finite-sample versions of standard multivariate skewness and kurtosis tests for multivariate normality, as well as new ways of combining skewness and kurtosis measures for that purpose. For non-Gaussian distribution families involving nuisance parameters, the problem of building confidence sets (through GF test "inversion") for the nuisance parameters and the error distribution was also considered.

We have also demonstrated the usefulness of the proposed GF tests with a size and power study which suggest guidelines for empirical work. In particular, it is evident that asymptotic theory is highly unreliable; in contrast, the MC tests are straightforward to use and achieve perfect size control. Furthermore, whereas empirical researchers in econometrics seem to favor JB-type criteria (perhaps because the available underlying theory allows for regressors), our MC versions of the Mardia-type tests emerge as a better choice.

Finally, the tests proposed were applied to an asset pricing model using monthly returns on New York Stock Exchange (NYSE) portfolios over five-year subperiods from 1926-1995. The results confirm through exact test that multivariate normality is rejected in all subperiods. Further, on inverting the GF tests for Student t and stable error distributions, we found heavy (though non-extreme) kurtosis. The reader may consult Beaulieu, Dufour and Khalaf (2001*b*, 2001*a*) for mean-variance efficiency tests which exploit these results.

[6]One may argue that tests based on empirical moments are not best suited for such alternatives since the true moments of the associated stable distributions do not exist; yet our tests as conceived somewhat circumvent this difficulty, because the simulation-based RBM are not necessarily estimates of moments: these may be viewed as estimates of an expected measure of central tendency and scale compatible with the hypothesized distribution.

A. Appendix: Monte Carlo tests

The Monte Carlo (MC) test procedure goes back to Dwass (1957) and Barnard (1963). Extensions to the nuisance-parameter-dependent case are from Dufour (2002). Here we summarize the underlying methodology (given a right tailed test), as it applies to the test statistics we consider in this paper. Let us first consider the pivotal statistics case, i.e. the case where the statistic considered, say $S = S(Y, X)$ can be written as a pivotal function of W [in (2.2)], formally $S(Y, X) = \overline{S}\,(W, X)$,where Y and X are as in (2.1), W is defined by (2.2), and is fully specified.

1. Let $S^{(0)}$ denote the test statistic calculated from the observed data set.
2. Generate N replications $S^{(1)}, \ldots, S^{(N)}$ of the test statistic S in such a way that $S^{(0)}, S^{(1)}, \ldots, S^{(N)}$ be exchangeable.
3. Given the series of simulated statistics $S^{(1)}, \ldots, S^{(N)}$, compute $\hat{p}_N[S] \equiv p_N(S^{(0)}; S)$, where
$$p_N(x; S) \equiv \frac{NG_N(x; S) + 1}{N + 1}, \tag{A.1}$$
$$G_N(x; S) \equiv \frac{1}{N}\sum_{i=1}^{N} s\big(S^{(i)} - x\big), \tag{A.2}$$
where $s(x) = 1$ if $x \geq 0$, and $s(x) = 0$ if $x < 0$. In other words, $p_N(S^{(0)}; S) = [NG_N(S^{(0)}; S) + 1]/(N + 1)$ where $NG_N(S^{(0)}; S)$ is the number of simulated values which are greater than or equal to $S^{(0)}$. When $S^{(0)}, S^{(1)}, \ldots, S^{(N)}$ are all distinct [an event with probability one when the when the vector $\big(S^{(0)}, S^{(1)}, \ldots, S^{(N)}\big)'$ has absolutely continuous distribution], $\hat{R}_N(S^{(0)}) = N + 1 - NG_N(S^{(0)}; S)$ is the rank of $S^{(0)}$ in the series $S^{(0)}, S^{(1)}, \ldots, S^{(N)}$.
4. The MC critical region is
$$p_N(S^{(0)}; S) \leq \alpha, \quad 0 < \alpha < 1. \tag{A.3}$$
If $\alpha(N + 1)$ is an integer, then, under the null hypothesis (provided the distribution of S is continuous),
$$\mathsf{P}\big[p_N(S^{(0)}; S) \leq \alpha\big] = \alpha; \tag{A.4}$$
see Dufour (2002).

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.

Baringhaus, L. and Henze, N. (1992), 'Limit distributions for Mardia's measure of multivariate skewness', *The Annals of Statistics* **20**, 1889–1902.

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*a*), 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.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2001*b*), Testing the CAPM with possibly non-Gaussian errors: An exact simulation-based approach, Technical report, Département d'économique, Université Laval, and CIREQ, Université de Montréal.

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

Binder, J. J. (1985), 'Measuring the effects of regulation with stock price data', *Rand Journal of Economics* **16**, 167–183.

Bontemps, C. and Meddahi, N. (2002), Testing normality: A GMM approach, Technical Report 2002s-63, CIRANO, Montréal, Canada.

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

Campbell, J. 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.

Chou, P.-H. (2000), 'Alternative tests of the zero-beta CAPM', *The Journal of Financial Research* **23**, 469–493.

D'Agostino, R. B. and Stephens, M. A., eds (1986), *Goodness-of-Fit Techniques*, Marcel Dekker, New York.

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*), 'Simulation based finite and large sample tests in multivariate regressions', *Journal of Econometrics* **111**(2), 303–322.

Dufour, J.-M. and Khalaf, L. (2003), Finite sample tests in seemingly unrelated regressions, *in* D. E. A. Giles, ed., 'Computer-Aided Econometrics', Marcel Dekker, New York, chapter 2, pp. 11–35.

Dufour, J.-M., Khalaf, L., Bernard, J.-T. and Genest, I. (2003), 'Simulation-based finite-sample tests for heteroskedasticity and ARCH effects', *Journal of Econometrics* **forthcoming**.

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 Kiviet, J. F. (1998), 'Exact inference methods for first-order autoregressive distributed lag models', *Econometrica* **66**, 79–104.

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

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.

Fama, E. F. and French, K. R. (1995), 'Size and book-to-market factors in earnings and returns', *The Journal of Finance* **50**, 131–155.

Fiorentini, G., Sentana, E. and Calzolari, G. (2003), 'Maximum likelihodd estimation and inference in multivariate conditionally heteroskedastic dynamic models with Student t innovations', *Journal of Business and Economic Statistics* **21**(4), 532–546.

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.

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.

Henze, N. (1994), 'On Mardia's kurtosis test for multivariate normality', *Communications in Statistics, Part A - Theory and Methods* **23**, 1031–1045.

Holgersson, H. E. T. and Shukur, G. (2001), 'Some aspects of non-normality tests in systems of regression equations', *Communications in Statistics, Simulation and Computation* **30**(2), 291–310.

Horswell, R. L. and Looney, S. W. (1992), 'A comparison of tests for multivariate normality that are based on measures of multivariate skewness and kurtosis', *Journal of Statistical Computation and Simulation* **42**, 21–38.

Horswell, R. L. and Looney, S. W. (1993), 'Diagnostic limitations of skewness coefficients in assessing departures from univariate and multivariate normality', *Communications in Statistics, Part B - Simulation and Computation* **22**, 437–459.

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

Jarque, C. M. and Bera, A. K. (1987), 'A test for normality of observations and regression residuals', *International Statistical Review* **55**, 163–172.

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.

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.

Kilian, L. and Demiroglu, U. (2000), 'Residual-based tests for normality in autoregressions: Asymptotic theory and simulation evidence', *Journal of Business and Economic Statistics* **18**, 40–50.

Lütkepohl, H. and Theilen, B. (1991), 'Measures of multivariate skewness and kurtosis for tests of nonnormality', *Statistical Papers* **32**, 179–193.

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

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

Mardia, K. V. (1974), 'Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies', *Sankhyā A* **36**, 115–128.

Mardia, K. V. (1980), Tests of univariate and multivariate normality, *in* P. R. Krishnaiah, ed., 'Handbook of Statistics 1: Analysis of Variance', North-Holland, Amsterdam, pp. 279–320.

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

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.

Schipper, K. and Thompson, R. (1985), 'The impact of merger-related regulations using exact distributions of test statistics', *Journal of Accounting Research* **23**, 408–415.

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.

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

Thode, Jr., H. C. (2002), *Testing for Normality*, number 164 *in* 'Statistics: Textbooks and Monographs', Marcel Dekker, New York.

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

Weron, R. (1996), 'On the Chambers-Mallows-Stuck method for simulating skewed stable random variables', *Statistics and Probability Letters* **28**, 165–171.

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.