

Centre interuniversitaire de recherche
en économie quantitative

CIREQ

Cahier 06-2003

*FINITE-SAMPLE DIAGNOSTICS FOR MULTIVARIATE
REGRESSIONS WITH APPLICATIONS TO
LINEAR ASSET PRICING MODELS*

Jean-Marie DUFOUR,
Lynda KHALAF and
Marie-Claude BEAULIEU

Cahier 06-2003

FINITE-SAMPLE DIAGNOSTICS FOR MULTIVARIATE REGRESSIONS WITH APPLICATIONS TO LINEAR ASSET PRICING MODELS

Jean-Marie DUFOUR¹, Lynda KHALAF² and Marie-Claude BEAULIEU³

- ¹ Canada Research Chair Holder (Econometrics), Centre interuniversitaire de recherche en économie quantitative (CIREQ) and Département de sciences économiques, Université de Montréal, and CIRANO
- ² CIREQ, and GREEN, Département d'économie, Université Laval
- ³ Centre interuniversitaire sur le risque, les politiques économiques et l'emploi (CIRPÉE), and Département de finance et assurance, Université Laval

May 2003

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, and the Fonds pour la formation de chercheurs et l'aide à la recherche (FCAR, Government of 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.

CIREQ, Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal (Québec) H3C 3J7
Canada

téléphone : (514) 343-6557
télécopieur : (514) 343-5831
cireq@umontreal.ca
<http://www.cireq.umontreal.ca>

Université 
de Montréal

 McGill

 Concordia
UNIVERSITY

Résumé

Dans cet article, nous proposons plusieurs tests de spécification valides pour des échantillons finis dans le cadre de régressions linéaires multivariées (RLM), avec des applications à des modèles d'évaluation d'actifs. Nous nous concentrons sur les déviations par rapport à l'hypothèse d'erreurs i.i.d. univariée ou multivariée, pour des distributions d'erreurs gaussiennes et non gaussiennes. Les tests univariés étudiés prolongent les procédures exactes existantes en permettant des paramètres non spécifiés dans la distribution des erreurs (e.g., le nombre de degrés de liberté dans le cas de la distribution de Student). Les tests multivariés sont basés sur des résidus standardisés multivariés qui assurent l'invariance par rapport aux coefficients RLM et à ceux de la matrice de covariance des erreurs. Nous considérons des tests contre la dépendance sérielle, contre la présence d'effets GARCH multivariés et des tests de signes contre l'asymétrie. Les procédures proposées sont des versions exactes des tests de Shanken (1990) qui consistent à combiner des tests de spécification univariés. Spécifiquement, nous combinons des tests entre équations en utilisant une approche de test de Monte Carlo (MC), ce qui permet d'éviter des bornes de type Bonferroni. Étant donné que les tests dans un contexte non gaussien ne sont pas pivotaux, nous appliquons une approche de test de Monte Carlo maximisé [Dufour (2002)] où la valeur p simulée pour l'hypothèse testée (qui dépend de paramètres de nuisance) est maximisée (par rapport aux dits paramètres de nuisance) dans le but de contrôler le niveau des tests. Nous appliquons les tests proposés à un modèle d'évaluation d'actifs qui comprend un taux d'intérêt sans risque observable et utilise 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. Nos résultats révèlent que les tests univariés exacts présentent des problèmes de dépendance sérielle, d'asymétrie et d'effets GARCH statistiquement significatifs dans certaines équations. Cependant ces problèmes s'avèrent moins importants, lorsque l'on tient compte de la dépendance entre équations. De plus, les écarts importants par rapport à l'hypothèse i.i.d. sont moins évidents une fois que l'on considère des erreurs non gaussiennes.

Mots clés : modèle d'évaluation d'actifs financiers, CAPM, efficacité moyenne-variance, non-normalité, modèle de régression multivariée, hypothèse uniforme linéaire, test de Monte Carlo, bootstrap, paramètre de nuisance, test de spécification, diagnostics, GARCH, test du ratio des variances

Abstract

In this paper, we propose several finite-sample specification tests for multivariate linear regressions (MLR) with applications to asset pricing models. We focus on departures from the assumption of *i.i.d.* errors assumption, at univariate and multivariate levels, with Gaussian and non-Gaussian (including Student t) errors. The univariate tests studied extend existing exact procedures by allowing for unspecified parameters in the error distributions (e.g., the degrees of freedom in the case of the Student t distribution). The multivariate tests are based on properly *standardized multivariate residuals* to ensure invariance to MLR coefficients and error covariances. We consider tests for serial correlation, tests for multivariate GARCH and sign-type tests against general dependencies and asymmetries. The procedures proposed provide exact versions of those applied in Shanken (1990) which consist in combining univariate specification tests. Specifically, we combine tests across equations using the MC test procedure to avoid Bonferroni-type bounds. Since non-Gaussian based tests are not pivotal, we apply the “maximized MC” (MMC) test method [Dufour (2002)], where the MC p -value for the tested hypothesis (which depends on nuisance parameters) is maximized (with respect to these nuisance parameters) to control the test’s significance level. The tests proposed 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. Our empirical results reveal the following. Whereas univariate exact tests indicate significant serial correlation, asymmetries and GARCH in some equations, such effects are much less prevalent once error cross-equation covariances are accounted for. In addition, significant departures from the *i.i.d.* hypothesis are less evident once we allow for non-Gaussian errors.

Keywords : capital asset pricing model, CAPM, mean-variance efficiency, non-normality, multivariate linear regression, uniform linear hypothesis, exact test, Monte Carlo test, bootstrap, nuisance parameters, specification test, diagnostics, GARCH, variance ratio test

Contents

1. Introduction	1
2. Framework and exact distributional theory	3
2.1. Exact invariance results	4
2.2. Exact Monte Carlo test procedure	5
3. Multivariate specification tests	6
3.1. Generalized diagnostics	6
3.2. Joint tests for GARCH effects	7
3.3. Joint serial dependence tests	8
3.4. Sign tests against asymmetry	9
4. Empirical application	10
4.1. Overview of the results	21
4.2. Detailed discussion of the results	21
5. Conclusion	25
A. Appendix: Monte Carlo goodness-of-fit tests	27

List of Tables

1	Portfolio definitions	10
2	Univariate and Multivariate GARCH Tests	11
3	Table: Univariate and Multivariate Predictability Tests	16

1. Introduction

The multivariate linear regression (MLR) model is certainly one of the most basic and widely used model in finance, econometrics and statistics in general [see Rao (1973, Chapter 8), Anderson (1984, chapters 8 and 13), Kariya (1985), Dufour and Khalaf (2002*d*), and the references therein]. Well known financial applications include: (i) tests of portfolio efficiency in the context of the capital asset pricing model (CAPM) [see for example: Gibbons (1982), Shanken (1986), MacKinlay (1987), Barone-Adesi (1986), Gibbons, Ross and Shanken (1989, GRS), Affleck-Graves and McDonald (1989), Shanken (1990), Zhou (1991), Zhou (1993), Zhou (1995), Fama and French (1993, 1995), Shanken (1996), Campbell, Lo and MacKinlay (1997, Chapter 5), 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 Jobson and Korkie (1982, 1989) and Kan and Zhou (2001)]; and (iii) event studies [see Binder (1985) and Schipper and Thompson (1985)].

A common feature of such models consists in assuming that the disturbances (or the errors) in different equations are correlated across equations, but otherwise constitute independent identically distributed (i.i.d.) random vectors. Of course, violation of the latter assumption can affect the results of various tests and inferences based on the model (such as mean-variance efficiency or spanning tests). This underscores the importance of performing diagnostics on such multivariate models.

As emphasized by Kroner and Ng (1998), the existing literature on multivariate diagnostics is sparse compared to the univariate case. Perhaps because of this, diagnostic tests in empirical MLR-based financial studies are often conducted on an equation by equation basis. Although univariate tests can provide some guidance, contemporaneous correlation of disturbances entails that statistics from individual equations are not independent. As a result, combining test decisions over all equations raises size control problems, so the need for joint testing naturally arises; see Richardson and Smith (1993) and Shanken (1990).

In this context, joint diagnostics are typically based either on asymptotic approximations or on Bonferroni-type bounds. The procedures suggested following the first approach involve test statistics which formally incorporate cross-sectional dependence yet are asymptotically free of nuisance parameters; see, for example, Godfrey (1988), Richardson and Smith (1993) and the recent literature on multivariate GARCH, which may be traced back to Engle and Kroner (1995) and Kroner and Ng (1998). Although this may lead to convenient test procedures, the fact remains that cross-equation correlations can still affect the null distributions of the test statistics in finite samples. In systems with many equations (*e.g.*, many portfolios), the number of correlations can be quite large relative to the sample size, leading to serious degrees-of-freedom losses and size distortions. As a result, asymptotic approximations perform poorly in finite samples; for references and simulation evidence, see Shanken (1996, Section 3.4.2), Campbell et al. (1997, Chapter 5), and Dufour and Khalaf (2002*b*, 2002*d*, 2002*c*). Alternatively, Bonferroni-based bound joint tests require one to divide the significance level of each individual test by the number of tests [see Dufour (1990), Shanken (1990), Dufour and Torrès (1998), Dufour and Khalaf (2002*b*)]. While this can guard against spurious rejections, it can also cause severe power losses if the MLR includes many equations.

Despite the above problems, very few finite sample exact specification tests have been proposed

for MLR models. One exception includes work on testing the independence between the disturbances in different equations [see Harvey and Phillips (1982) and Dufour and Khalaf (2002b)]. But this problem is relatively simple, for the null hypothesis sets the error covariances to zero. The basic difficulty one meets here consists in allowing for an unknown contemporaneous covariance matrix which involves a rapidly growing number of nuisance parameters as the number of equations increases.

In this paper, we consider the problem of testing the specification of MLR models. Of course, the form of a model can be tested against an infinity of alternative formulations or specification errors. Here we focus on the following basic deviations from the MLR specification: (1) detecting the presence of GARCH-type heteroskedasticity; (2) detecting (linear) serial dependence; (3) testing whether the errors follow a symmetric distribution. For these three problems, we propose procedures based on least squares residuals, hence computationally simple. In order to avoid the nuisance parameter problem raised by the unknown error covariance matrix, we apply a multivariate rescaling transformation which eliminates the unknown covariance matrix from the residual distribution. In this way, we get *multivariate standardized residuals* which are *location-scale invariant*, hence do not depend on the (unknown) regression coefficients or the error covariance matrix.

The tests against the presence of GARCH effects include multivariate extensions of the univariate procedures proposed by Engle (1982), Lee (1991) and Lee and King (1993). The tests for linear serial dependence are multivariate versions of the portmanteau Ljung-Box [Ljung and Box (1978)] and variance ratio [Lo and MacKinlay (1988, 1989)] tests. For testing symmetry, we introduce a sign procedure. All these tests are applied to properly standardized residuals. None of the exact procedures is based on a Bonferroni bound, *i.e.* they do not require one to divide the significance level by the number of individual equations.

To overcome multiple-test concerns as well as the fact that the test statistics considered have complicated null distributions which can be extremely difficult to evaluate analytically, we apply the Monte Carlo (MC) test technique [Dwass (1957), Barnard (1963), Dufour and Kiviet (1996), Dufour and Kiviet (1998)], which provides randomized exact versions of the tests considered. The MC simulation-based procedure yields an exact test whenever the joint distribution of the individual test statistics does not depend on unknown nuisance parameters (*i.e.*, the test statistics are jointly *pivotal*) under the null hypothesis and general parametric distributional assumptions. The fact that the relevant analytical distributions are quite complicated is not a problem: all we need is the possibility of simulating the relevant test statistics under the null hypothesis. The combined procedures described also provide exact versions of those applied by Shanken (1990) without the need to use bounds.

Furthermore, our methodology deals, from a finite-sample perspective, with non-normal errors. Formally, this allows one to test for time varying variances or asymmetries, with fat-tailed error distributions such as the Student t . This approach is new, even in the case of univariate tests. Indeed, in this case, the recent literature on simulation based testing [see e.g. Dufour, Khalaf, Bernard and Genest (2001), Dufour and Khalaf (2002a) for univariate residuals based heteroskedasticity and serial correlation tests] allows one to deal with non-Gaussian disturbances, if the error distribution is fully specified. For instance, in the case of the Student t distribution, with an unknown degree-of-freedom parameter, it will typically appear in the null distribution of the diagnostic test statistic.

In the present paper, we extend the procedures proposed in Dufour et al. (2001) and Dufour and Khalaf (2002a) to account for unspecified parameters in the hypothesized error distribution. To control significance level given such difficulties, we apply a “maximized MC” (MMC) test [see Dufour (2002)], where the MC p -value for the tested hypothesis (which depends on the nuisance parameter) is maximized over the relevant nuisance parameter set.¹

The procedures proposed are applied to a typical multivariate market-model based on New York Stock Exchange (NYSE) portfolios, constructed from the University of Chicago Center for Research in Security Prices (CRSP) data base for the period 1926-1995. In Beaulieu, Dufour and Khalaf (2001b), this data set is analyzed in view of testing mean-variance efficiency. Here, we assess the specification using the procedures described above. In particular, we perform finite-sample tests – both univariate and multivariate – of the assumption of i.i.d. errors against the presence of GARCH effects [using joint Engle and Lee-King tests] as well as against linear serial dependence [using joint Ljung-Box and variance ratio tests]. This allows us to compare and contrast evidence from both approaches. Furthermore, the sign procedure is applied to test symmetry.

Our empirical results reveal the following. Whereas univariate exact tests indicate some serial correlation and GARCH effects, the multivariate tests indicate that such effects are less prevalent (at least in the time frames considered) once the cross-equation error covariances are considered. This underscores the merits of formal multivariate testing approaches. More importantly, significant departures from the *i.i.d.* hypothesis are less evident when we allow for non-Gaussian errors. This illustrates the importance of formally accounting for error distributions from a finite-sample perspective.

The paper is organized as follows. Section 2 sets the framework and presents our basic exact distributional results. In Section 3, we present our univariate and multivariate diagnostic criteria. Section 4 reports our empirical analysis. We conclude in Section 5.

2. Framework and exact distributional theory

Many asset pricing models take the form of a system of correlated regression equations, *i.e.*

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

where $Y = [Y_1, \dots, Y_n]$ is a $T \times n$ matrix of observations on n dependent variables, X is an $T \times k$ full-column rank matrix, $B = [B_1, \dots, B_n]$ is a $k \times n$ matrix of unknown coefficients and $U = [U_1, \dots, U_n] = [V_1, \dots, V_T]'$ is a $T \times n$ matrix of random disturbances. For example, a one-factor asset pricing model can be written:

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

¹For further discussions of MMC tests in econometrics, see Dufour and Khalaf (2001, 2002c). When nuisance parameters appear in the null distribution of a test statistic, a test is *exact* at level α if the largest rejection probability over the nuisance parameter space consistent with the null hypothesis is not greater than α [see Lehmann (1986, sections 3.1, 3.5)].

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

$$\begin{aligned} Y &= [r_1, \dots, r_n], \quad r_i = (r_{i1}, \dots, r_{iT})', \quad X = [\nu_T, \tilde{r}_M], \\ \nu_T &= (1, \dots, 1)', \quad \tilde{r}_M = (\tilde{r}_{M1}, \dots, \tilde{r}_{MT})', \quad U_i = (u_{i1}, \dots, u_{iT})'. \end{aligned}$$

Following Dufour and Khalaf (2002d) and Beaulieu et al. (2001b), we restrict the error distribution as follows:

$$V_t = JW_t, \quad t = 1, \dots, T, \quad (2.3)$$

where J is an unknown non-singular lower triangular matrix, the vector $vec(W_1, \dots, W_T)$ has a distribution $\mathcal{F}(\nu)$, which is specified up to an unknown nuisance parameter ν . Let

$$\Sigma = JJ', \quad W = [W_1, \dots, W_T]' = [w_1, \dots, w_n]. \quad (2.4)$$

Assumption (2.3) entails that $W = U(J^{-1})'$ and, when the moments of $\mathcal{F}(\cdot)$ exist, Σ is the covariance of V_t . The least squares estimate of B is

$$\hat{B} = (X'X)^{-1}X'Y \quad (2.5)$$

and the corresponding residual matrix is

$$\hat{U} = [\hat{U}_1, \dots, \hat{U}_n] = Y - X\hat{B} = MY = MU \quad (2.6)$$

where $M = I - X(X'X)^{-1}X'$. Note that the Gaussian-based quasi maximum likelihood estimators for this model are \hat{B} and $\hat{\Sigma} = \hat{U}'\hat{U}/T$.

2.1. Exact invariance results

The test statistics we consider are based on the following *multivariate standardized residual* matrix

$$\tilde{W} = \hat{U} S_{\hat{U}}^{-1} \quad (2.7)$$

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

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

The validity of our proposed test procedures relies on the following decomposition of \tilde{W} .

Theorem 2.1 *Under (2.1), and for all error distributions compatible with (2.3), the standardized residual matrix defined in (2.7) satisfies the identity*

$$\tilde{W} = \hat{U} S_{\hat{U}}^{-1} = \hat{W} S_{\hat{W}}^{-1} \quad (2.8)$$

where $\widehat{W} = MW$ and $S_{\widehat{W}}$ is the Cholesky factor of $\widehat{W}'\widehat{W}$, and thus follows a distribution which does not depend on B and J .

For a proof and further discussion, see Dufour, Khalaf and Beaulieu (2002). The latter result implies that \widehat{W} has a distribution which is completely determined by the distribution of W given X . This also holds for all statistics which depend on the data only through \widehat{W} .

2.2. Exact Monte Carlo test procedure

The above invariance results can be used to obtain Monte Carlo p -values for any test statistic associated with (2.3), so long as the statistic considered, say $S(\widehat{U})$ where \widehat{U} is the residual matrix (2.6), can be written as a function of W and X :

$$S(\widehat{U}) = \overline{S}(W, X) \quad (2.9)$$

where W is defined by (2.3). The Monte Carlo (MC) test procedure goes back to Dwass (1957) and Barnard (1963); extensions to the nuisance parameter dependent case are given in Dufour (2002). The procedures we apply in this paper can be summarized as follows, given a right tailed test statistic of the form (2.9).

Let S_0 denote the observed value of test statistic calculated from the observed data set. For ν given, draw $W^j = [W_1^j, \dots, W_n^j]$, $j = 1, \dots, N$, as in (2.3), and compute

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

Given the rank $\widehat{R}_N(S_0 | \nu)$ of S_0 in the series S_0, S_1, \dots, S_N , let

$$\widehat{p}_N(S_0 | \nu) = \frac{N\widehat{G}_N(S_0 | \nu) + 1}{N + 1}, \quad \widehat{G}_N(S_0 | \nu) = \frac{\widehat{R}_N(S_0 | \nu) - 1}{N - 1}. \quad (2.10)$$

The MC critical region is $\widehat{p}_N(S_0 | \nu) \leq \alpha$, $0 < \alpha < 1$ and $\alpha(N + 1)$ is an integer.

Under the null hypothesis,

$$P[\widehat{p}_N(S_0 | \nu) \leq \alpha] = \alpha \text{ when } \nu = \nu_0 \text{ (}\nu_0 \text{ known)}. \quad (2.11)$$

When ν is an unknown nuisance parameter, we can use the following level-correct critical MC test which we will denote maximized MC (MMC) test defined by the critical region

$$\sup_{\nu \in \Phi_0} [\widehat{p}_N(S_0 | \nu)] \leq \alpha \quad (2.12)$$

where Φ_0 satisfies the null hypothesis under test [see Dufour (2002)].²

²This procedure is based on the following test property [Lehmann (1986)]: when the null distribution of test statistic depends on nuisance parameters [ν in this case], an α -level is guaranteed when the largest p -value (overall values of ν consistent with the null hypothesis) is referred to α . The MMC test method works exactly in this way: the (simulated) p -value function conditional on ν is numerically maximized (with respect to ν); the test is significant if the largest p -value is less than or equal to α .

In Dufour (2002) and Dufour and Kiviet (1996), a modified MMC procedure denoted confidence-set-based MMC (CSMMC) is also proposed. The method involves two stages: (1) an exact confidence set is built for ν , and (2) the MC p -value $[\widehat{p}_N(S_0|\nu)]$ in this case] is maximized over all values of ν in the latter confidence set; see also Beaulieu et al. (2001*b*, 2001*a*). 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 $\alpha_1 = \alpha_2 = \alpha/2$. We also use, following Beaulieu et al. (2001*b*, 2001*a*), a confidence set designed to incorporate information on the goodness-of-fit of the hypothesized error distribution. See Appendix A for a summarized description of the estimation procedure; a detailed presentation is available in Beaulieu et al. (2001*b*).

3. Multivariate specification tests

In this section, we use the above results to derive multivariate specification tests. We first present our general approach to multivariate testing, with applications to GARCH, serial dependence and sign tests. The proposed tests take the form (2.9) so are formally valid for any parametric null hypothesis of the type (2.3). In our empirical application [see Section (4)], we focus on the normal and multivariate $t(\kappa)$ distributions, where κ is the number of degrees of freedom. We also consider (2.3) with possibly unknown parameters.

3.1. Generalized diagnostics

Assumptions routinely checked in multivariate regressions are in many ways similar to those relevant in a univariate regressions, which include: (i) autocorrelated disturbances, (ii) heteroskedastic disturbances, (iii) lack-of-fit of the error distribution, (iv) parameter constancy. If one pursues a univariate approach, standard tests for these hypotheses may be applied to the residuals of each equation of (2.1), which we write

$$Y_i = XB_i + U_i \quad (3.1)$$

following the notation of Section 2. In this context, it is natural to consider the univariate counterparts of (2.3)-(2.4):

$$U_i = \sigma w_i, \quad w_i \sim \mathcal{F}(\nu), \quad (3.2)$$

where the distribution $\mathcal{F}(\nu)$ of the vector w_i is specified up to an unknown nuisance parameter ν . In this case, the argument of Section (2.2) implies that any statistic of the form

$$S_i(\widehat{U}_i) = \overline{S}_i(w_i, X), \quad (3.3)$$

where \widehat{U}_i is the residual vector and w_i satisfies (3.2), yields an exact MC univariate test conditional on ν . In Dufour, Farhat, Gardiol and Khalaf (1998), Dufour and Khalaf (2001), Dufour et al. (2001) and Dufour and Khalaf (2002*a*), we show that standard residual-based test statistics for heteroskedasticity, non-normality, autocorrelation and structural breaks are of the form (3.3), for any parametric error distribution (Gaussian or non-Gaussian) specified up to an unknown scale parameter (σ in our notation); this corresponds to (3.2) with known ν . Furthermore, in Dufour

et al. (2001), we argue that the latter property underlies the validity of diagnostic tests based on regression residuals (rather than observables).³ This paper presents further extensions to the case of unknown ν , applying the MMC or the CSMMC procedure. We note that these extensions are new to econometrics even from a univariate perspective.

Extensions to multivariate setups can be pursued imposing e.g. vector-autoregressive based procedures; see, for example, Engle and Kroner (1995), Deschamps (1996). However, it is clear that such parametrization may raise further identification and nuisance parameter problems. Alternatively, procedures which combine univariate standard tests [see e.g. Shanken (1990)] have been appealing to practitioners, because of parsimony and identification ease. Here we follow the latter approach, with specific modifications to ensure exactness.

Let \tilde{w}_i denote the columns of the matrix \tilde{W} defined by (2.7). Obtain standardized versions of the univariate tests denoted $S_i(\tilde{U}_i)$, $i = 1, \dots, n$, replacing \tilde{U}_i by \tilde{w}_i in the formula for these statistics. These modified statistics, denoted, $\hat{S}_i(\tilde{w}_i)$, $i = 1, \dots, n$, satisfy the conditions of Theorem 2.1 by construction. Hence, under (2.3), their *joint* distribution does not depend on the error covariances. To obtain combined inference across equations, we consider the following combined statistic:

$$S_{\min}(\tilde{W}) = 1 - \min_{1 \leq i \leq n} [p(\hat{S}_i(\tilde{w}_i))] \quad (3.4)$$

where $p(\hat{S}_i(\tilde{w}_i))$ refers to p -values; these may be obtained applying a MC test method, or using asymptotic null distributions [to cut execution time]. The underlying intuition here is to reject the null hypothesis if at least one of the individual (standardized) tests is significant; for further reference on tests combined in this way, see Dufour and Khalaf (2002b), Dufour et al. (2001) and Dufour and Khalaf (2002a). By construction, the latter statistic(3.4) takes the form (2.9). Then we can apply the MMC or CSMMC test procedure to the combined statistic imposing (2.3).

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 MC test technique. Indeed, the property underlying exactness is joint pivotality, which is achieved by using properly standardized residuals. In what follows, we give the formulae for the tests considered in our empirical analysis.

3.2. Joint tests for GARCH effects

We consider procedures based on the standard GARCH tests [e.g. see Engle (1982), Lee (1991) or Lee and King (1993)]. The Engle test statistic for equation i , which will be denoted E_i is given by TR_i^2 , where T is the sample size, R_i^2 is the coefficient of determination in the regression of the equation's squared OLS residuals \hat{u}_{it}^2 on a constant and $\hat{u}_{(t-j),i}^2$ ($j = 1, \dots, q$). Under standard regularity conditions, the asymptotic null distribution of this statistic is $\chi^2(q)$. Lee and King (1993)

³To assess the validity of residual based homoskedasticity tests, Godfrey (1996, section 2) defined the notion of "robustness to estimation effects". Following Godfrey (1996, section 2), a test is robust to estimation effects if the asymptotic distribution is the same irrespective of whether disturbances or residuals are used to construct the test statistic. Our approach to residual based diagnostic tests departs from Godfrey's asymptotic framework.

proposed an alternative GARCH test which exploits the one-sided nature of H_A . The test statistic is

$$LK_i = \frac{\left\{ (T-q) \sum_{t=q+1}^T [(\hat{u}_{it}^2/\hat{\sigma}_i^2 - 1)] \sum_{j=1}^q \hat{u}_{i,t-j}^2 \right\} / \left\{ \sum_{t=q+1}^T (\hat{u}_{it}^2/\hat{\sigma}_i^2 - 1)^2 \right\}^{1/2}}{\left\{ (T-q) \sum_{t=q+1}^T \left(\sum_{j=1}^q \hat{u}_{i,t-j}^2 \right)^2 - \left(\sum_{t=q+1}^T \left(\sum_{j=1}^q \hat{u}_{i,t-j}^2 \right) \right)^2 \right\}^{1/2}} \quad (3.5)$$

where $\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^T \hat{u}_{it}^2$, and its asymptotic null distribution is standard normal. Let \widetilde{W}_{it} denote the elements of the matrix \widetilde{W} defined by (2.7). As shown in the previous section, we obtain standardized versions of the univariate Engle and Lee-King test, denoted \widetilde{E}_i and \widetilde{LK}_i , replacing \hat{u}_{it} by \widetilde{W}_{it} in the formula for these statistics. We next construct the following combined statistics:

$$\widetilde{E} = 1 - \min_{1 \leq i \leq n} [p(\widetilde{E}_i)] \quad (3.6)$$

$$\widetilde{LK} = 1 - \min_{1 \leq i \leq n} [p(\widetilde{LK}_i)] \quad (3.7)$$

where $p(\widetilde{E}_i)$ and $p(\widetilde{LK}_i)$ refer to p -values; in the empirical application considered next, we obtained these p -values using asymptotic null distributions. The MMC or CSMMC technique can be used to obtain p -values for these tests, as described above.

3.3. Joint serial dependence tests

We focus on two popular criteria which combine empirical residual autocorrelations: (i) the Ljung-Box portmanteau statistic [Ljung and Box (1978)],

$$LB_i = T(T+2) \sum_{j=1}^J \frac{\widehat{\rho}_{ij}^2}{T-j}, \quad (3.8)$$

and (ii) the variance ratio statistic [Lo and MacKinlay (1988), Lo and MacKinlay (1989)],

$$VR_i = 1 + 2 \sum_{j=1}^J \left(1 - \frac{j}{l}\right) \widehat{\rho}_{ij} \quad (3.9)$$

where

$$\widehat{\rho}_{ij} = \frac{\sum_{t=j+1}^T \widehat{u}_{it} \widehat{u}_{i,t-j}}{\sum_{t=1}^T \widehat{u}_{it}^2}, \quad j = 1, \dots, J. \quad (3.10)$$

We propose multivariate extensions following the combination procedure of Section 3.1. First we obtain *orthogonalized* versions of the univariate variance ratio and Ljung-Box statistics, denoted \widetilde{LB}_i and \widetilde{VR}_i , replacing \widehat{u}_{it} by \widetilde{W}_{it} [the elements of \widetilde{W} , the standardized residuals matrix (2.7)] in the formula for these statistics; with this modification, the statistics *joint* null (under (2.3)) does not

depend on the error cross-covariances. Second, we use the combined criteria

$$\widetilde{VR} = 1 - \min_{1 \leq i \leq n} [p(\widetilde{VR}_i)] \quad (3.11)$$

$$\widetilde{LB} = 1 - \min_{1 \leq i \leq n} [p(\widetilde{LB}_i)] \quad (3.12)$$

where $p(\widetilde{VR}_i)$ and $p(\widetilde{LB}_i)$ are individual p -values associated with \widetilde{LB}_i and \widetilde{VR}_i ; these are derived applying approximate null distributions:

$$(VR_i - 1) \stackrel{asy}{\sim} N[0, 2(2J - 1)(J - 1)/3J], \quad (3.13)$$

$$LB(J) \stackrel{T \rightarrow \infty}{\sim} \chi^2(J). \quad (3.14)$$

Once again, to obtain an exact combined p -value under (2.3), we can apply the MMC (or the CSMMC) test procedures to the combined statistic.

3.4. Sign tests against asymmetry

As emphasized in Campbell et al. (1997, Chapter 2), tests based on signs can be useful for assessing deviations from the *i.i.d.* errors hypothesis. However, the usual distributional theory underlying these tests breaks down, in the presence of regression residuals, even in a univariate perspective. Nevertheless, due to the flexibility of the MC testing method, we show in this section how to extend such procedures to the context of (3.2). The tests as proposed are not non-parametric; yet they retain their intuitive appeal. We consider here a sign procedure which can be sensitive to possible asymmetry of then distribution.

Let PP_i refer to the proportion of positive residuals in equation i . To test (3.2) with $\nu = \nu_0$, we propose the following statistic:

$$sg_i = |PP_i - \overline{PP}_i(\nu_0)| \quad (3.15)$$

where $\overline{PP}_i(\nu_0)$ is a simulation-based estimate of the proportion of positive residuals under (3.2). The proportion of positive residuals is unchanged if the residuals are multiplied by any positive constant; hence sg_i is scale invariant and can be simulated [in a single equation framework] under (3.2) as follows. Given ν_0 , draw N_0 realizations of w_i consistent with (3.2); for each draw, construct the simulated residuals as Mw_i which yield N_0 realizations of the sign statistic. The empirical average of the latter simulated series gives $\overline{PP}_i(\nu_0)$. In the same manner, following Section 2.2, a MC p -value for each statistic can be obtained, conditional on $\overline{PP}_i(\nu_0)$.

To extend this approach to the multivariate framework, we proceed as described in Section 3.1. Formally, we obtain standardized versions of the univariate sign tests, denoted \widetilde{sg}_i , replacing \widehat{u}_{it} by \widetilde{W}_{it} in their formula. Second, we use the combined criteria

$$\widetilde{sg} = 1 - \min_{1 \leq i \leq n} [p(\widetilde{sg}_i)] \quad (3.16)$$

where $p(\widetilde{sg}_i)$ are individual p -values associated with \widetilde{sg}_i ; these may be derived applying the MC test technique, drawing residuals [in the form MW] as in (2.3). Finally, exact MMC (or CSMMC)

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, Gibbons and Litzenberger (1989).

p -values are obtained for the joint criterion (3.16).

4. Empirical application

Our empirical analysis focuses on the asset pricing model (2.2) 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 et al. (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.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. 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 reported in Tables 2–3. We report, in addition to the asymptotic p -values [denoted \hat{p}_∞] when available, three MC p -values: (i) the Gaussian based MC p -value [denoted \hat{p}_g], (ii) the Student t based MMC p -value (maximized over all relevant degrees-of-freedom $\kappa \geq 2$) [denoted \hat{p}_a], and (ii) the Student t based CSMMC [denoted \hat{p}_i] (where the maximization is restricted to the degrees-of-freedom not rejected by a prior 2.5% goodness-of-fit tests); the associated confidence sets are reported in column 1 of Table 2. The results can be summarized as follows.

⁴Note that as in Breeden et al. (1989), firms with SIC code 39 (Miscellaneous manufacturing industries) are excluded from the dataset for portfolio formation.

Table 2. Univariate and Multivariate GARCH Tests

			Engle				Lee-King			
			\widehat{p}_i	\widehat{p}_a	p_∞	\widehat{p}_g	\widehat{p}_i	\widehat{p}_a	p_∞	\widehat{p}_g
1927-1930	Confidence Set $3 \leq \kappa \leq 12$	1	.25	.31	.428	.25	.11	.14	.065	.15
		2	.85	.86	.879	.78	.89	.89	.998	.84
		3	.95	.94	.946	.90	.54	.55	.709	.52
		4	.10	.10	.164	.02	.48	.45	.526	.48
		5	.11	.14	.259	.05	.17	.11	.046	.16
		6	.83	.83	.847	.71	.96	.96	1.00	.93
		7	.79	.79	.827	.69	.15	.17	.113	.16
		8	.70	.75	.810	.65	.17	.19	.121	.16
		9	.70	.74	.797	.65	.45	.49	.574	.45
		10	.47	.48	.604	.46	.13	.15	.090	.16
		11	.01	.06	.007	.01	.06	.08	.023	.08
		12	.10	.11	.217	.04	.04	.07	.013	.07
Joint			.04	.11	n.a.	.01	.29	.38	n.a.	.34
1931-1935	Confidence Set $3 \leq \kappa \leq 8$	1	.43	.49	.547	.45	.09	.12	.138	.09
		2	.81	.93	.910	.92	.66	.68	.839	.58
		3	.90	.97	.951	.92	.07	.06	.057	.08
		4	.13	.22	.268	.18	.12	.16	.190	.12
		5	.32	.41	.447	.31	.07	.09	.088	.08
		6	.76	.87	.849	.82	.34	.38	.547	.33
		7	.21	.31	.334	.26	.18	.26	.338	.24
		8	.06	.09	.104	.09	.16	.21	.256	.17
		9	.87	.96	.939	.94	.20	.29	.382	.24
		10	.94	1.0	.987	.99	.19	.26	.352	.24
		11	.05	.06	.054	.04	.09	.12	.135	.09
		12	.10	.19	.251	.16	.73	.74	.892	.67
Joint			.06	.23	n.a.	.04	.61	.78	n.a.	.79

Note _ Numbers shown are MC p -values. p_∞ refers to the test's asymptotic p -value and \widehat{p}_g is the Gaussian based MC p -value. \widehat{p}_i is CSMMC p -value imposing student $t(\kappa)$ errors (the method for constructing the underlying confidence set for the degrees-of-freedom κ is presented in Appendix A). \widehat{p}_a is the MMC p -value overall degrees-of-freedom. The individual equation test statistic are: Engle's TR^2 and Lee-King's statistic (3.5) defined in Section 3.2; the joint tests are defined in (3.6)-(3.7). n.a. means not available (because asymptotic critical values have not been derived).

Table 2(b). Univariate and Multivariate GARCH Tests (continued)

			Engle				Lee-King			
			\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g	\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g
1936–1940	Confidence Set $4 \leq \kappa \leq 29$	1	.96	.96	.937	.94	.60	.66	.796	.56
		2	1.0	1.0	.992	1.0	.87	.87	.994	.84
		3	.48	.52	.550	.43	.81	.81	.976	.78
		4	.49	.62	.660	.50	.24	.31	.440	.24
		5	.48	.52	.563	.50	.12	.17	.232	.13
		6	.96	1.0	.986	1.0	.29	.37	.490	.33
		7	.18	.23	.281	.24	.86	.89	.997	.85
		8	.99	1.0	1.00	1.0	.64	.66	.806	.60
		9	.15	.29	.317	.17	.56	.55	.742	.54
		10	.99	1.0	.999	1.0	.74	.73	.918	.68
		11	.12	.26	.280	.17	.72	.74	.901	.68
		12	.72	.85	.886	.79	.20	.25	.357	.21
Joint			.16	.33	n.a.	.08	.62	.67	n.a.	.61
1941–1945	Confidence Set $\kappa \geq 5$	1	.32	.62	.656	.55	.98	.99	1.00	.95
		2	.13	.14	.203	.09	.56	.54	.723	.52
		3	.79	.81	.851	.78	.54	.57	.748	.49
		4	.11	.10	.128	.05	.87	.85	.993	.84
		5	.11	.10	.137	.05	.12	.11	.137	.10
		6	.44	.38	.433	.38	.87	.85	.994	.84
		7	.21	.40	.457	.15	.16	.15	.187	.12
		8	.39	.35	.416	.35	.31	.31	.400	.26
		9	.11	.13	.171	.06	.21	.21	.280	.14
		10	.52	.53	.561	.45	.13	.13	.166	.10
		11	1.0	1.0	.989	.99	.31	.31	.400	.26
		12	.73	.65	.673	.65	.74	.80	.959	.67
Joint			.90	.90	n.a.	.87	1.0	1.0	n.a.	.99
1946–1950	Confidence Set $4 \leq \kappa \leq 31$	1	.26	.27	.320	.23	.16	.14	.187	.13
		2	.66	.68	.729	.64	.59	.64	.791	.53
		3	.72	.73	.789	.66	.63	.65	.794	.58
		4	.05	.08	.086	.03	.10	.10	.116	.08
		5	.52	.54	.598	.49	.52	.28	.398	.58
		6	.95	.97	.963	.91	.33	.40	.506	.29
		7	.11	.17	.255	.08	.88	.72	.868	.85
		8	.61	.61	.679	.60	.10	.10	.116	.08
		9	.92	.85	.864	.90	.55	.67	.834	.50
		10	.15	.12	.161	.09	.89	.92	.999	.84
		11	.28	.18	.282	.20	.81	.76	.943	.77
		12	.57	.71	.766	.57	.71	.75	.924	.67
Joint			.03	.08	n.a.	.01	.81	.83	n.a.	.82

Table 2(c). Univariate and Multivariate GARCH Tests (continued)

			Engle				Lee-King			
			\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g	\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g
1951-1956	Confidence Set $5 \leq \kappa \leq 34$	1	1.0	1.0	.992	1.0	.83	.83	.987	.77
		2	.34	.31	.386	.32	.64	.66	.810	.59
		3	.09	.09	.149	.06	.05	.05	.054	.05
		4	.99	1.0	.990	.99	.08	.10	.108	.06
		5	.42	.47	.510	.42	.77	.76	.933	.69
		6	.25	.27	.350	.23	.11	.12	.128	.09
		7	.03	.05	.024	.01	.02	.03	.017	.02
		8	.29	.29	.367	.26	.11	.12	.127	.10
		9	.27	.55	.597	.24	.07	.06	.058	.06
		10	.96	.91	.900	.92	.77	.75	.904	.71
		11	1.0	1.0	.982	.99	.29	.26	.356	.26
		12	.75	.71	.750	.70	.52	.52	.677	.46
Joint			.22	.32	n.a.	.16	.28	.38	n.a.	.37
1956-1960	Confidence Set $\kappa \geq 5$	1	1.0	1.0	.993	1.0	.66	.70	.839	.60
		2	.49	.48	.547	.42	.75	.77	.949	.71
		3	.53	.56	.626	.51	.86	.87	.995	.83
		4	.57	.61	.640	.52	.63	.66	.820	.57
		5	.36	.30	.396	.31	.39	.51	.682	.33
		6	.55	.53	.633	.49	.30	.29	.460	.25
		7	.44	.46	.531	.39	.03	.03	.014	.02
		8	.58	.58	.667	.53	.03	.03	.015	.02
		9	.72	.79	.807	.68	.18	.18	.267	.15
		10	.33	.33	.424	.28	.90	.91	.999	.89
		11	.21	.24	.320	.18	.08	.10	.080	.07
		12	.99	.99	.977	.98	.65	.70	.846	.59
Joint			1.0	1.0	n.a.	1.0	.67	.67	n.a.	.66
1961-1965	Confidence Set $\kappa \geq 7$	1	.64	.65	.693	.57	1.0	1.0	1.00	1.0
		2	.57	.59	.639	.47	.83	.84	.983	.78
		3	.99	.98	.973	.96	.83	.84	.983	.77
		4	.27	.30	.350	.25	.71	.73	.892	.66
		5	.88	.88	.869	.91	.83	.82	.977	.76
		6	.71	.71	.768	.65	.93	.94	1.00	.89
		7	.88	.86	.884	.91	.68	.67	.828	.60
		8	.12	.10	.147	.07	.76	.79	.963	.73
		9	.30	.32	.374	.27	.84	.84	.986	.79
		10	.90	.85	.895	.92	.90	.91	.999	.87
		11	.15	.14	.185	.10	.49	.50	.684	.43
		12	.36	.37	.425	.28	1.0	1.0	1.00	.97
Joint			.47	.50	n.a.	.36	.25	.25	n.a.	.22

Table 2(d). Univariate and Multivariate GARCH Tests (continued)

			Engle				Lee-King			
			\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g	\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g
1966-1970	Confidence Set $\kappa \geq 5$	1	.56	.57	.619	.49	.09	.09	.106	.09
		2	.83	.83	.853	.79	.84	.83	.987	.79
		3	.88	.86	.870	.85	.56	.57	.749	.52
		4	.51	.53	.586	.46	.09	.09	.118	.09
		5	.01	.05	.005	.01	.15	.15	.182	.11
		6	.45	.45	.491	.37	.95	.96	1.00	.90
		7	.28	.29	.339	.23	.86	.86	.995	.83
		8	.30	.30	.355	.27	.91	.91	.999	.86
		9	.25	.23	.277	.20	.65	.68	.827	.58
		10	.04	.07	.056	.02	.51	.52	.701	.47
		11	.61	.57	.606	.51	.10	.12	.129	.09
		12	.05	.08	.092	.04	.68	.71	.850	.62
Joint			.08	.23	n.a.	.05	.30	.30	n.a.	.29
1971-1975	Confidence Set $4 \leq \kappa \leq 28$	1	.06	.09	.100	.06	.10	.11	.164	.10
		2	.25	.25	.271	.21	.05	.05	.046	.04
		3	1.0	1.0	.999	1.0	.28	.28	.381	.22
		4	.42	.42	.464	.39	.02	.02	.007	.01
		5	.92	.92	.915	.89	.14	.16	.229	.11
		6	.97	.98	.954	.95	.21	.22	.291	.15
		7	.94	.95	.940	.91	.21	.22	.302	.18
		8	.50	.50	.563	.46	.82	.83	.980	.78
		9	.35	.38	.416	.33	.62	.63	.774	.58
		10	.12	.15	.184	.11	.05	.05	.044	.04
		11	1.0	.99	.964	.99	.37	.40	.520	.34
		12	.61	.63	.673	.53	.11	.11	.173	.10
Joint			.36	.48	n.a.	.32	.26	.28	n.a.	.24
1976-1980	Confidence Set $4 \leq \kappa \leq 17$	1	.25	.25	.286	.20	.03	.04	.014	.03
		2	.27	.33	.393	.26	.93	.94	1.00	.89
		3	.27	.27	.320	.23	.06	.06	.067	.06
		4	.15	.16	.210	.14	.05	.05	.029	.03
		5	.01	.02	.002	.01	.01	.01	.000	.01
		6	.04	.07	.045	.01	.06	.07	.072	.06
		7	.04	.09	.066	.01	.01	.01	.000	.01
		8	.82	.85	.867	.82	.13	.16	.206	.12
		9	.42	.42	.496	.40	.13	.16	.209	.10
		10	.15	.15	.196	.13	.59	.59	.753	.52
		11	.30	.36	.420	.26	.90	.90	.998	.87
		12	.18	.22	.257	.17	.27	.31	.420	.23
Joint			.03	.11	n.a.	.01	.03	.03	n.a.	.01

Table 2(e). Univariate and Multivariate GARCH Tests (continued)

			Engle				Lee-King			
			\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g	\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g
1981-1985	Confidence Set $5 \leq \kappa \leq 33$	1	.70	.72	.751	.67	.96	.96	1.00	.93
		2	.41	.41	.465	.34	.42	.42	.577	.39
		3	.80	.78	.794	.76	.99	1.0	1.00	.95
		4	.15	.16	.176	.10	.72	.74	.902	.65
		5	.11	.08	.102	.07	.19	.21	.262	.13
		6	.03	.06	.016	.01	.21	.21	.270	.16
		7	.01	.02	.002	.01	.57	.57	.728	.47
		8	.11	.11	.156	.08	.25	.23	.332	.20
		9	.62	.68	.692	.61	.08	.09	.081	.07
		10	.94	.94	.914	.90	.37	.41	.552	.34
		11	.65	.61	.652	.63	.95	.96	1.00	.93
		12	.12	.18	.203	.12	.85	.87	.990	.84
		Joint			.08	.21	n.a.	.06	.12	.14
1986-1990	Confidence Set $5 \leq \kappa \leq 41$	1	.02	.06	.021	.01	.50	.56	.673	.50
		2	.44	.41	.557	.35	.12	.12	.091	.09
		3	.46	.47	.598	.40	.76	.81	.930	.67
		4	.50	.50	.654	.45	.31	.31	.355	.28
		5	.59	.60	.727	.57	.97	.97	1.00	.96
		6	.86	.85	.879	.82	.58	.64	.746	.57
		7	.37	.41	.558	.32	.94	.94	.996	.89
		8	.62	.65	.760	.59	.88	.88	.981	.80
		9	.59	.60	.719	.58	.96	.96	.999	.95
		10	.14	.16	.277	.12	.72	.79	.895	.63
		11	.98	.99	.981	.97	.43	.43	.508	.38
		12	.06	.12	.152	.04	.81	.87	.969	.77
		Joint			.14	.23	n.a.	.06	.47	.47
1991-1995	Confidence Set $\kappa \geq 15$	1	.71	.71	.759	.64	.31	.31	.370	.27
		2	.60	.61	.667	.54	.03	.03	.020	.02
		3	.76	.75	.803	.73	.96	.96	1.00	.95
		4	.71	.68	.730	.61	.03	.03	.020	.02
		5	.99	.99	.974	.98	.74	.75	.892	.70
		6	.40	.44	.494	.36	.85	.85	.988	.80
		7	.36	.39	.419	.27	.02	.02	.003	.01
		8	.50	.53	.584	.44	.68	.68	.834	.64
		9	.79	.81	.848	.76	.09	.09	.106	.06
		10	.94	.94	.933	.88	.69	.70	.794	.60
		11	.52	.53	.598	.48	.17	.17	.203	.12
		12	.20	.20	.233	.16	.38	.38	.494	.34
		Joint			.86	.88	n.a.	.76	.31	.31

Table 3. Univariate and Multivariate Predictability Tests

		Sign Test			Ljung-Box				Variance ratio			
		\hat{p}_i	\hat{p}_a	\hat{p}_g	\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g	\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g
1927-1930	1	.159	.291	.028	.49	.52	.511	.49	.76	.74	.852	.72
	2	.139	.295	.043	.47	.49	.478	.49	.61	.63	.777	.51
	3	.804	.752	.576	.66	.68	.690	.58	.37	.36	.646	.34
	4	.743	.787	.736	.84	.89	.840	.84	.70	.68	.806	.63
	5	.999	.999	.982	.69	.62	.608	.61	.58	.43	.680	.48
	6	.294	.437	.080	.96	.25	.188	.17	.11	.13	.539	.09
	7	.743	.811	.661	.18	.18	.145	.13	.38	.79	.664	.35
	8	.515	.702	.397	.78	.78	.780	.79	.82	.75	.845	.80
	9	.272	.422	.099	.35	.34	.306	.31	.02	.02	.341	.02
	10	.460	.570	.212	.59	.61	.594	.55	.02	.02	.211	.01
	11	.264	.434	.070	.04	.04	.010	.04	.04	.05	.466	.03
	12	.994	.995	.982	.01	.02	.001	.01	1.0	.88	.911	.96
Joint		.548	.728	.199	.05	.07	n.a.	.04	.02	.02	n.a.	.02
1931-1935	1	.036	.180	.004	.30	.35	.322	.31	.64	.67	.800	.61
	2	.987	.993	.968	.55	.61	.579	.58	.81	.82	.880	.77
	3	.492	.673	.274	.25	.29	.253	.26	.57	.54	.714	.56
	4	.930	.990	.839	.59	.72	.705	.61	.53	.59	.758	.45
	5	.229	.465	.073	.06	.07	.039	.06	.91	1.0	.998	.88
	6	.222	.384	.055	.71	.71	.693	.69	.90	.90	.928	.86
	7	.317	.493	.139	.11	.11	.077	.09	.59	.71	.827	.58
	8	.990	.990	.966	.39	.37	.339	.37	.21	.21	.564	.23
	9	.175	.322	.037	.45	.49	.458	.45	.23	.17	.539	.23
	10	.791	.874	.779	.20	.21	.177	.20	.56	.60	.764	.50
	11	.484	.659	.309	.73	.75	.713	.73	.78	.75	.848	.74
	12	.807	.840	.688	.05	.05	.009	.04	.46	.39	.660	.43
Joint		.569	.874	.048	.44	.52	n.a.	.46	.08	.09	n.a.	.06

Note _ Numbers shown are MC p -values. p_∞ refers to the test's asymptotic p -value and \hat{p}_g is the Gaussian based MC p -value. \hat{p}_i is the CSMC p -value imposing student t errors; the underlying confidence sets (see Appendix A) are reported in Table 2. \hat{p}_a is the MMC p -value overall degrees-of-freedom. The individual equation test statistics are: the sign test (3.15) defined in Section 3.4; the Variance ratios (3.9) and the Ljung-Box criteria (3.8); the joint tests are defined in (3.16)-(3.11)-(3.12).

Table 3(b). Univariate and Multivariate Predictability Tests (continued)

		Sign Test			Ljung-Box				Variance ratio			
		\hat{p}_i	\hat{p}_a	\hat{p}_g	\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g	\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g
1936-1940	1	.689	.798	.532	.90	.90	.848	.90	.99	1.0	.997	.96
	2	.982	.997	.910	.87	.85	.759	.83	.44	.41	.679	.40
	3	.396	.693	.302	.18	.16	.114	.18	.59	.68	.828	.52
	4	.538	.670	.368	.93	.97	.918	.94	.61	.65	.809	.57
	5	.154	.451	.076	.54	.57	.543	.52	.86	.99	.974	.84
	6	.120	.391	.072	.66	.70	.666	.69	.91	.94	.947	.90
	7	.254	.477	.165	.51	.54	.528	.50	.86	.86	.927	.85
	8	.239	.486	.150	.11	.11	.065	.09	.38	.39	.674	.36
	9	.624	.791	.626	.24	.32	.283	.19	.19	.29	.612	.17
	10	.817	.838	.670	.90	.93	.868	.91	.11	.12	.512	.10
	11	.008	.169	.006	.05	.06	.017	.05	.05	.07	.436	.05
	12	.844	.892	.741	.79	.85	.762	.77	.56	.52	.745	.50
Joint		.652	.722	.535	.28	.28	n.a.	.25	.92	.21	n.a.	.92
1941-1945	1	.684	.772	.530	.90	.92	.854	.90	.28	.28	.613	.23
	2	.519	.731	.449	.64	.68	.653	.59	.35	.39	.655	.31
	3	.270	.585	.196	.95	.95	.892	.95	.57	.58	.744	.55
	4	.126	.377	.052	.59	.58	.563	.56	.70	.10	.830	.64
	5	.283	.619	.177	.92	.92	.862	.90	.31	.33	.628	.29
	6	.090	.380	.051	.27	.27	.255	.26	.21	.23	.588	.19
	7	.094	.337	.065	.08	.09	.049	.07	.66	.72	.832	.61
	8	.396	.530	.237	.17	.14	.088	.13	.40	.36	.645	.35
	9	.784	.920	.809	.35	.32	.311	.30	.04	.05	.356	.04
	10	.878	.944	.896	.83	.85	.801	.79	.61	.65	.773	.60
	11	.288	.580	.279	.64	.65	.583	.59	.93	.99	.971	.89
	12	.396	.683	.434	.43	.41	.381	.35	.10	.12	.508	.07
Joint		.303	.604	.268	.79	.79	n.a.	.74	.20	.21	n.a.	.17
1946-1950	1	.337	.556	.171	.91	.91	.872	.86	.76	.77	.842	.71
	2	.540	.737	.476	.09	.08	.030	.07	.85	.87	.890	.81
	3	.637	.785	.512	.29	.35	.324	.26	.02	.04	.345	.02
	4	.693	.782	.524	1.0	1.0	.989	1.0	.41	.26	.593	.39
	5	.830	.876	.733	.57	.39	.316	.53	.17	.13	.517	.14
	6	.160	.433	.103	.72	.73	.687	.71	.73	.79	.851	.66
	7	.656	.803	.519	.18	.16	.104	.16	.55	.77	.840	.54
	8	.022	.174	.001	.14	.16	.106	.12	.77	.80	.874	.72
	9	.041	.215	.020	.43	.43	.374	.41	.32	.34	.324	.28
	10	.257	.458	.156	.05	.07	.024	.04	.32	.33	.627	.27
	11	.240	.487	.165	.95	.96	.920	.90	.97	.91	.930	.93
	12	.511	.661	.406	1.0	1.0	.991	.98	.48	.91	.925	.44
Joint		.353	.581	.152	.28	.30	n.a.	.30	.08	.08	n.a.	.05

Table 3(c). Univariate and Multivariate Predictability Tests (continued)

		Sign Test			Ljung-Box				Variance ratio			
		\hat{p}_i	\hat{p}_a	\hat{p}_g	\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g	\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g
1951-1955	1	.298	.564	.155	.58	.59	.551	.56	.90	.83	.877	.86
	2	.390	.597	.320	.62	.69	.679	.58	.03	.04	.318	.03
	3	.840	.900	.781	.41	.41	.364	.37	.56	.58	.761	.55
	4	.540	.688	.365	.99	.98	.966	.97	.79	.92	.934	.76
	5	.274	.593	.192	1.0	1.0	.997	.99	.87	1.0	.998	.85
	6	.688	.778	.519	.01	.01	.000	.01	.25	.24	.575	.20
	7	.094	.372	.054	.42	.41	.367	.40	.88	.88	.920	.85
	8	.532	.688	.376	.18	.18	.139	.13	.25	.26	.592	.21
	9	.161	.424	.101	.92	1.0	.809	.86	1.0	.85	.986	1.0
	10	.477	.729	.393	.70	.67	.620	.67	.82	1.0	.995	.78
	11	.383	.654	.301	.28	.26	.206	.26	.39	.46	.705	.36
	12	.441	.615	.287	.68	.74	.696	.66	.93	.85	.896	.89
Joint		.313	.398	.159	.98	.98	n.a.	.94	.17	.17	n.a.	.14
1956-1960	1	.944	.949	.929	.60	.60	.568	.59	.05	.05	.362	.04
	2	.294	.546	.209	.14	.14	.547	.11	.03	.02	.225	.02
	3	.534	.708	.450	.88	.88	.837	.85	.05	.05	.410	.05
	4	.787	.867	.732	.89	.87	.835	.86	.34	.34	.653	.30
	5	.405	.655	.261	.04	.03	.006	.04	.01	.01	.049	.01
	6	.277	.534	.217	.47	.47	.453	.42	.32	.31	.639	.29
	7	.388	.598	.274	.40	.40	.388	.36	.82	.83	.892	.75
	8	.445	.617	.287	.30	.31	.272	.25	.56	.58	.761	.55
	9	.984	.979	.953	.53	.47	.448	.46	.93	.89	.911	.92
	10	.616	.801	.539	.29	.19	.168	.22	.02	.02	.137	.01
	11	.963	.963	.924	.69	.66	.627	.67	.20	.21	.579	.15
	12	.959	.959	.978	.69	.68	.631	.66	.71	.65	.796	.65
Joint		.811	.791	.697	.91	.91	n.a.	.85	.11	.11	n.a.	.09
1961-1965	1	.641	.775	.517	.79	.79	.720	.71	.86	.86	.911	.79
	2	.393	.655	.337	.64	.64	.614	.60	.86	.86	.912	.78
	3	.774	.876	.724	.67	.67	.643	.65	.10	.10	.492	.10
	4	.093	.369	.044	.87	.84	.799	.85	.74	.73	.830	.67
	5	.460	.725	.397	.49	.47	.451	.47	.06	.06	.403	.05
	6	.546	.753	.494	.58	.59	.543	.54	.72	.71	.821	.67
	7	.506	.735	.400	.60	.58	.505	.56	.12	.12	.505	.10
	8	.581	.769	.581	.88	.90	.834	.87	.72	.72	.829	.67
	9	.964	.977	.857	.36	.35	.293	.32	.13	.12	.518	.10
	10	.984	.984	.939	.44	.41	.398	.39	.35	.35	.645	.28
	11	.034	.236	.006	.46	.47	.424	.43	.06	.06	.350	.04
	12	.961	.925	.899	.54	.57	.615	.52	.27	.29	.615	.24
Joint		.654	.920	.556	.97	.97	n.a.	.94	.68	.68	n.a.	.62

Table 3(d). Univariate and Multivariate Predictability Tests (continued)

		Sign Test			Ljung-Box				Variance ratio			
		\hat{p}_i	\hat{p}_a	\hat{p}_g	\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g	\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g
1966-1970	1	.668	.789	.491	.13	.16	.088	.10	1.0	1.0	.998	.99
	2	.680	.784	.581	.33	.33	.282	.30	.37	.37	.660	.31
	3	.648	.841	.523	.99	.99	.971	.97	.68	.68	.806	.62
	4	.951	.937	.867	.75	.75	.685	.74	.59	.60	.773	.54
	5	.396	.632	.276	.74	.74	.629	.68	.42	.41	.684	.40
	6	.375	.645	.335	.26	.28	.218	.22	.20	.17	.568	.16
	7	.828	.828	.633	.61	.60	.529	.56	.66	.67	.800	.60
	8	.437	.598	.311	.86	.87	.811	.84	.76	.81	.868	.72
	9	.200	.492	.150	.87	.87	.821	.84	.85	.86	.903	.82
	10	.379	.589	.294	.87	.87	.814	.84	.88	.94	.931	.86
	11	.161	.402	.081	.75	.75	.661	.74	.57	.56	.754	.51
	12	.312	.562	.190	.83	.84	.768	.78	.49	.51	.720	.45
Joint		.705	.728	.484	.66	.66	n.a.	.55	.91	.91	n.a.	.84
1971-1975	1	.827	.841	.744	.16	.16	.124	.13	.55	.55	.762	.49
	2	.323	.562	.238	.21	.24	.191	.18	.43	.42	.690	.36
	3	.661	.810	.522	.72	.73	.679	.72	.97	.96	.966	.92
	4	.120	.377	.043	.18	.18	.152	.17	.15	.16	.538	.13
	5	.126	.333	.038	.23	.27	.220	.22	.16	.16	.544	.14
	6	.996	.998	.908	.23	.27	.225	.23	.53	.53	.750	.45
	7	.196	.420	.107	.46	.50	.436	.42	.52	.52	.738	.45
	8	.993	.993	.873	.17	.19	.158	.17	.88	.89	.931	.80
	9	.453	.588	.338	.84	.88	.828	.79	.88	.87	.918	.78
	10	.275	.524	.232	.46	.50	.450	.42	.99	1.0	.997	.94
	11	.967	.967	.998	.87	.88	.822	.82	.84	.82	.894	.74
	12	.156	.448	.112	.13	.13	.096	.08	.85	.87	.903	.74
Joint		.191	.597	.039	.40	.39	n.a.	.37	.28	.28	n.a.	.24
1976-1980	1	.694	.792	.521	.51	.56	.503	.48	.81	.78	.871	.75
	2	.865	.891	.812	.94	.93	.887	.91	.46	.48	.721	.43
	3	.846	.884	.722	.33	.31	.310	.26	.95	.98	.972	.97
	4	.689	.764	.585	.86	.86	.829	.83	1.0	1.0	.991	.99
	5	.367	.601	.245	.30	.32	.321	.26	.65	.63	.794	.57
	6	.986	.986	.879	.14	.14	.115	.13	.17	.17	.541	.16
	7	.985	.999	.986	.99	1.0	.973	.98	.71	.73	.833	.64
	8	.643	.839	.666	.08	.08	.040	.07	.31	.30	.620	.26
	9	.514	.663	.460	.18	.20	.177	.17	.35	.27	.609	.29
	10	.131	.346	.052	.30	.31	.312	.26	.25	.26	.604	.24
	11	.181	.448	.064	1.0	1.0	.999	1.0	.99	1.0	.988	.98
	12	.807	.885	.713	.21	.94	.895	.90	.94	.96	.382	.04
Joint		.637	.743	.535	1.0	1.0	n.a.	.97	.24	.24	n.a.	.20

Table 3(e). Univariate and Multivariate Predictability Tests (continued)

		Sign Test			Ljung-Box				Variance ratio			
		\hat{p}_i	\hat{p}_a	\hat{p}_g	\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g	\hat{p}_i	\hat{p}_a	p_∞	\hat{p}_g
1981-1985	1	.142	.405	.057	.53	.53	.474	.47	.53	.55	.734	.47
	2	.191	.501	.167	.22	.22	.165	.20	.35	.35	.628	.30
	3	.770	.882	.688	.33	.33	.271	.30	.96	.98	.970	.92
	4	.987	.990	.843	.91	.90	.846	.88	.52	.51	.703	.44
	5	.677	.783	.520	.32	.35	.297	.30	.69	.67	.801	.62
	6	.342	.660	.345	.76	.75	.704	.74	.93	.94	.940	.90
	7	.107	.344	.056	.23	.23	.174	.20	.97	.97	.959	.91
	8	.984	.984	.878	.49	.53	.474	.46	.29	.29	.602	.28
	9	.290	.567	.255	.91	.94	.883	.88	.13	.16	.535	.12
	10	.388	.602	.278	.14	.15	.100	.11	.65	.67	.798	.60
	11	.378	.662	.298	.76	.77	.732	.76	.41	.42	.668	.35
	12	.511	.690	.479	.26	.24	.194	.22	.53	.55	.732	.47
Joint		.853	.922	.508	.04	.04	n.a.	.02	.29	.29	n.a.	.24
1986-1990	1	.917	.964	.989	.57	.58	.542	.52	.62	.63	.803	.56
	2	.483	.732	.444	.92	.96	.901	.93	.27	.30	.637	.22
	3	.546	.721	.438	.80	.78	.747	.75	.06	.07	.482	.05
	4	.824	.914	.711	.13	.13	.080	.09	.05	.05	.459	.03
	5	.489	.693	.407	1.0	1.0	.978	.98	.92	.93	.931	.85
	6	.211	.559	.125	.62	.64	.567	.54	.89	.87	.906	.81
	7	.623	.808	.620	.97	.96	.927	.94	.51	.50	.727	.46
	8	.425	.647	.329	.28	.28	.230	.25	.92	.94	.944	.87
	9	.781	.889	.658	.52	.53	.499	.49	.72	.63	.805	.69
	10	.953	.953	.913	.37	.34	.292	.34	.12	.12	.531	.06
	11	.799	.857	.681	.90	.91	.860	.85	.32	.39	.683	.26
	12	.025	.251	.011	.99	1.0	.973	.98	.76	.77	.870	.72
Joint		.734	.734	.430	.59	.58	n.a.	.52	.48	.48	n.a.	.42
1991-1995	1	.379	.653	.238	.20	.20	.156	.16	.18	.18	.554	.15
	2	.279	.540	.243	.38	.38	.366	.33	.56	.58	.773	.50
	3	.966	.966	.911	.57	.55	.523	.55	.27	.29	.620	.24
	4	.811	.919	.657	.37	.37	.362	.33	.18	.20	.576	.16
	5	.518	.748	.379	.46	.46	.456	.39	.47	.49	.729	.47
	6	.081	.369	.054	1.0	1.0	.990	1.0	.51	.51	.742	.48
	7	.488	.675	.445	.09	.10	.039	.08	.01	.02	.065	.01
	8	.442	.690	.448	.98	.99	.945	.97	.95	.99	.979	.94
	9	.936	.993	.907	.60	.62	.581	.57	.05	.06	.331	.04
	10	.260	.563	.254	.24	.25	.200	.23	.96	.99	.973	.95
	11	.505	.716	.422	.45	.43	.423	.36	.64	.62	.800	.59
	12	.376	.611	.297	.33	.32	.313	.30	.98	.99	.978	.99
Joint		.962	.969	.956	.91	.91	n.a.	.88	.12	.13	n.a.	.08

4.1. Overview of the results

First, focusing on the univariate tests, we see that the asymptotic p -values may lead to under-rejections, particularly in the case of the variance ratio test. In general, the MC p -values are lower than the asymptotic ones for GARCH and variance ratio tests. The converse seems to hold for the Ljung-Box tests; note that a small scale simulation study we performed in Dufour and Khalaf (2002a) suggested this test might over-reject in regression contexts.

Second, our results illustrate the importance of considering multivariate tests. In many cases, whereas univariate tests detect significant serial correlations and/or GARCH effects, the joint tests do not detect significant departures from the null hypothesis of interest. Interestingly however, in some cases, multivariate tests are more powerful in rejecting the joint null. This occurs for example in the case of Engle's test, in the 1946-50 sample, and for the Ljung-Box case, in the 1981-85 sub-sample. The smallest univariate p -value is 15%, yet the joint test p -value is only 4%. Recall that the univariate tests do not account for error covariances, which justifies such conflicts.

Thirdly, the confidence set based MMC approach is useful here. Indeed, in many cases, if degrees-of-freedom which are not compatible with the data are allowed, GARCH effects may end up undetected. For example, refer to the joint Engle-type test in 1927-30, 1946-50, 1976-80; this also occurs more noticeably with univariate tests.

Finally, we observe that almost invariably, the tests considered have led to conflicting decisions regarding the maintained hypothesis. For example, although in principle, Lee-King's test is supposed to be superior to the Engle test (because it is designed to account for positivity of variance), we observe that in many cases, the Engle test is significant whereas its Lee-King counterpart is not. The same holds true for the Ljung-Box and Variance Ratio tests. Of course, a formal simulation study is required to assess the relative power of competing tests, yet the conflicts we observed with this well known data set is worth noting, particularly when using exact p -values, for conflicts in test decisions are often attributed to finite sample distortions.

Overall, the multivariate diagnostics do not detect serious deviations from the *i.i.d.* assumption.⁵ The assumption of *i.i.d.* errors provides an acceptable working framework with this data set. This result supports the common practice of considering 5 years sub-groups to perform asset pricing tests (e.g. efficiency, spanning, etc.) tests which require *i.i.d.* errors.

4.2. Detailed discussion of the results

We next discuss the test results in details, across all subperiods. We use α to represent the level of the tests.

- 1927-30 subperiod

Gaussian-based Engle tests detect significant GARCH in equations 4, 5, 11 and 12 [for $\alpha \geq 2\%$ and 5%, 1% and 4% respectively]; the associated joint test is significant for $\alpha \geq 1\%$. Note that the asymptotic Engle test is significant ($p_\infty = .007$) only for equation 11. With Student t errors, the

⁵Recall that in the case of the confidence set based MMC \hat{p}_i , the cut-off level is 2.5% (if an overall level of 5% is desired) since 2.5% was used to construct the underlying confidence set.

MC p -values for Engle's tests exceed 10% for equations 4, 5 and 12; in the case of equation 11, GARCH-t effects are still significant for levels exceeding 3.5% ($\hat{p}_i = .01$); the associated joint test remains marginally significant for $\alpha \geq 6.5\%$ ($\hat{p}_i = .04$). Regarding Lee-King's test, although the asymptotic p -values are less than 5% in the case of equations 5, 11 and 12, all MC p -values exceed 5% on univariate and joint levels.

The MC variance ratio tests are significant at 5% in Gaussian and non-Gaussian cases, for equations 9, 10 and 11 and on a joint level. The MC Gaussian and non-Gaussian Ljung-Box tests are significant at 5% for equations 11 and 12; however, the associated joint test is significant for level exceeding 4% under Gaussian errors and levels exceeding 7% for non-Gaussian errors. Sign tests are significant ($\hat{p}_g = .028$ and $.043$ respectively) only for equations 1 and 2.

- 1931-35 subperiod

Gaussian-based Engle tests detect significant GARCH in equations 11 and on a joint test basis [for $\alpha \geq 4\%$]; the associated asymptotic test is not significant at 5% ($p_\infty = .054$). With Student t errors, GARCH-t effects are still significant in this equation, for levels exceeding 6% ($\hat{p}_a = .06$); the Student t joint test is marginally significant for $\alpha \geq 8.5\%$ ($\hat{p}_i = .06$). Lee-King's tests are not significant overall equations in this subperiod.

While all variance ratio tests are not significant at conventional levels, the Ljung-Box MC test is significant at 5% in Gaussian and non-Gaussian context; with normal errors, the sign test is also significant at 5% for equation 9 and marginally ($\hat{p}_g = .055$) for equation 6. However, the corresponding joint p -values exceed 44%. The sign test is significant for equation 1, for $\alpha \geq .004$ assuming normality and $\alpha \geq 6.1\%$ assuming t -errors; the associated joint test is significant (for $\alpha \geq 4.8\%$) only under normality.

- 1936-40 subperiod

While all univariate p -values exceed 17%, the joint Gaussian-based Engle test has a p -value of 8%. Otherwise, no significant departure from the *i.i.d.* hypotheses are detected in this subperiod. Turning to Table 3, we see that the variance ratio, Ljung-Box and sign tests are all significant at 5% in the case of equation 11, under normality; in this case, with t -errors, the sign test remains significant (for $\alpha \geq 3.3\%$) while the variance ratio and Ljung-Box are significant for $\alpha \geq 7\%$ and 6% respectively. All joint tests have p -values larger than 20%.

- 1941-45 subperiod

While all asymptotic univariate p -values exceed 20%, Gaussian-based Engle tests detect significant GARCH in equations 4 and 5 [for $\alpha \geq 5\%$] and marginally for equation 9 [for $\alpha \geq 6\%$]; however, the associated joint p -value is not significant ($\hat{p}_g = .87$) and the same holds for the associated GARCH-t p -values which exceed 10%. Lee-King's tests are not significant overall equations in this subperiod.

MC p -values for the sign and Ljung-Box test all exceed 5%; Gaussian and t -based variance ratio tests are significant at 5% in equation 9 [for $\alpha \geq 4\%$]. However, all associated joint p -values exceeds 21%.

- 1946-50 subperiod

While all asymptotic univariate p -values exceed 8.6%, Gaussian-based Engle tests detect significant GARCH in equations 4 [for $\alpha \geq 3\%$]; the associated joint p -value is highly significant ($\hat{p}_g = .01$). In the case of GARCH-t p -values, the MC Engle tests detect significant GARCH for levels exceeding 7.5% ($\hat{p}_i = .05$); the associated joint test is still significant for $\alpha \geq 5.5\%$. Lee-King's tests are also not significant in this subperiod.

>From Table 3, we see that the MC variance ratios are significant at 5% in the case of equation 3, in normal and non-normal cases; on a joint basis, the normal based MC variance ratio test is significant at 5% whereas its t-based counterpart is significant at 8%. Regarding the MC Ljung-Box test, we see that it is significant in the case of equation 12, at $\alpha \geq 4\%$ under normality and $\alpha \geq 7\%$ under Student t errors; the test is however not significant on a joint level. Examining the sign test under normal errors, we find significant departures from the *i.i.d.* hypothesis in equations 8 and 9 (for $\alpha \geq 0.1\%$ and 2% respectively); under t-errors, the test remains significant in these equations (for $\alpha \geq 5.7\%$ and 6.6% respectively). The joint sign tests are not significant at conventional levels.

- 1951-56 subperiod

Gaussian-based Engle tests detect significant GARCH in equations 7 ($\hat{p}_g = .01$) and marginally for equation 3 [for $\alpha \geq 6\%$]; the associated joint test is however not significant, and the asymptotic Engle test is significant ($p_\infty = .024$) only for equation 7. With Student t errors, the MC p -values for Engle's tests exceed 9% for equations 3; in the case of equation 11, GARCH-t effects are still significant for levels exceeding 5% ($\hat{p}_a = .05$); the associated joint test is not significant for levels. Regarding Lee-King's test, we find significant Gaussian and t-based p -values in both equations 3 and 7 (in equation 3, $\hat{p}_g = .05$ and $\hat{p}_a = .05$; in equation 7, $\hat{p}_g = .02$ and $\hat{p}_a = .03$). However, all joint Lee-King tests have p -values not smaller than 28%.

With respect to tests predictability, the normal and non-normal variance ratio and Ljung-Box MC tests are significant at 5% for equations 2 and 6 respectively. However, all joint tests and all sign tests are not significant at usual levels.

- 1956-60 subperiod

While all Engle tests are not significant at conventional levels, Lee-King's test detect significant GARCH in equations 7 and 8 with Gaussian and Student t errors (in both equations $\hat{p}_g = .02$ and $\hat{p}_a = .03$). The asymptotic p -values also yield the same evidence. However, none of the joint p -values is significant (all exceed 60%).

In this subperiod, the normal and non-normal variance ratio MC tests are significant at 5% for equations 1, 2, 3, 5 and 10. Despite such univariate evidence, the associated joint tests have p -values which exceed 9%. It is also worth noting that the asymptotic Variance Ratio test is not significant for equations 1, 2, 3 and 10. In the case of the Ljung-Box test, we find significant (at 5%) MC p -values under normal and non-normal errors only for equation 10; we note that the joint MC Ljung-Box p -values all exceed 85% and none of the sign tests is significant at usual levels.

- 1961-65 subperiod

While all univariate p -values exceed 14%, the Gaussian-based Engle test has a p -value of 7% for equation 8. Otherwise, no significant departures from the *i.i.d.* hypotheses are detected in this subperiod.

Turning to the Ljung-Box tests, we observe no significant p -values at standard levels. In contrast, the MC normal and non-normal sign test is significant (at 5%) for equation 11; the tests is significant at 5% given normal errors in equation 4, but the Student t p -values are not significant; the same holds true for the joint sign tests. Similarly, none of the Ljung-Box tests is significant. The variance ratio MC tests are significant, in equation 4 and 11, for $\alpha \geq 5\%$ and 4% respectively, under normality, and 6% under the Student t hypothesis; the associated joint p -values exceed 62%.

- 1966-70 subperiod

Gaussian-based Engle tests detect significant GARCH in equations 5, 10 and 12 [for $\alpha \geq 1\%$, 2%, and 4% respectively]; the associated joint test is significant for $\alpha \geq 5\%$. Note that the asymptotic Engle test is significant at 5% only for equation 5. With Student t errors, in the case of equation 5, GARCH effects are still significant for levels exceeding 3.5% ($\hat{p}_i = .01$); for equations 10 and 12 GARCH-t effects are marginally significant for $\alpha \geq 6.5\%$ and 7.5% respectively; the associated joint test is not marginally significant ($\hat{p}_i = .08$ and $\hat{p}_\alpha = .23$). All Lee-King tests have p -values not smaller than 9%. Similarly, none of the predictability tests is significant at usual levels.

- 1971-75 subperiod

Gaussian-based Lee-King tests signal significant GARCH in equations 2, 4 and 10 [for $\alpha \geq 4\%$, 1%, and 4% respectively]; the associated joint test is however not significant. With Student t errors, GARCH effects are still significant at 5% in all three cases. However, the joint p -values are not significant. Engle's tests are not significant except for the Gaussian based case in equation 1, with a p -value of .06%.

The serial correlation tests are all not significant at usual levels. The normal-errors based sign test is significant (at 5%) jointly and for equations 4 and 5. However, its t -based counterparts are not significant.

- 1976-80 subperiod

Gaussian and Student t based Lee-King tests are significant at 5% in equations 1, 4, 5 and 7; the associated joint tests all have p -values $\leq 3\%$. Engle's tests provide the same evidence given Gaussian errors for equations 5, 6 and 7 and on joint basis. The Student t based Engle test for these equations is significant for levels exceeding 2%, 6.5% and 6.5% respectively; the joint Engle Student t based CSMMC p -value is 3%.

Except for equation 12 under normality, the variance ratio tests are not significant. All MC Ljung-Box and sign tests exceed 5% in this subperiod.

- 1981-85 subperiod

While all Lee-King tests are not significant at conventional levels, Engle's test is significant in equations 6 and 7 with Gaussian and Student t errors ($\hat{p}_g = .01$ in both equations; in equation 7, $\hat{p}_a = .02$ while in equation 6, Engle's sup-MC test is significant for $\alpha \geq 5.5\%$). However, joint p -values are not significant at 5%.

In this subperiod, we note an interesting result regarding the Ljung-Box test: whereas all univariate tests are not significant, the joint MC normal and non-normal tests are significant at 5%. Recall that the joint tests are based on orthogonalized residuals, which corrects for cross-correlations. All other tests (univariate or multivariate) are not significant on a joint level.

- 1986-90 subperiod

While all Lee-King tests are not significant at conventional levels, Engle's test is significant in equations 1 and 12 with Gaussian errors ($\hat{p}_g = .01$ and $.04$ respectively). However, there is evidence for GARCH errors only in equation 1 ($\hat{p}_a = .02$). Furthermore, joint p -values are not significant at 5%. Similarly, none of the Ljung-Box tests is significant. The variance ratio MC tests are significant, in equation 3 and 4, for $\alpha \geq 5\%$ and 3% respectively, under normality, and for $\alpha \geq 6\%$ and 5% under the Student t hypothesis; the associated joint p -values exceed 52%. As for the sign test, we find significant MC normal and non-normal p -values (at 5%), in equation 12; however, this no longer holds on a joint test level.

- 1991-95 subperiod

In this subperiod, Engle tests are not significant jointly and individually Lee-King's test is significant in equations 2 and 7 with Gaussian and Student t errors ($\hat{p}_g = .02$ and $.01$ respectively; in equation 7, $\hat{p}_a = .02$ while in equation 1, Lee-King's sup-MC t -based test is significant for $\alpha \geq 5.5\%$). However, all joint p -values exceed 23%.

Turning to Table 3, we see that Sign tests and Ljung-Box tests are not significant jointly and individually The MC variance ratio test is significant in equations 7 and 9 with Gaussian and Student t errors ($\hat{p}_g = .01$ and $.04$ respectively; in equation 7, $\hat{p}_a = .02$ while in equation 9, the variance ratio sup-MC t -based test is significant for $\alpha \geq 6\%$). However, all joint p -values exceed 8%.

5. Conclusion

Previous research typically assess MLR-based asset pricing statistical models using tests based on individual equations. Due to error cross-correlations, statistics from individual equations which are not independent, which raises simultaneous test problems. In this paper, we consider a diagnostic test procedure that accounts for cross-equation correlations exactly, in possibly non-normal contexts. We focus on departures from the *i.i.d.* errors hypothesis, on univariate and multivariate levels. Our univariate tests extend existing exact procedures by allowing for unspecified parameters in the error distributions. Our multivariate tests are invariant to MLR coefficients and error covariances; dependence on further unknown parameters in the error distribution is circumvented applying MC test techniques. We consider, tests for serial correlation, tests for multivariate GARCH and sign-type tests against general dependencies and asymmetries. The tests proposed 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. Our tests reveal noticeable conflicts in the univariate versus multivariate procedures; this highlights the importance of formally controlling for contemporaneous correlation of disturbances.

A. Appendix: Monte Carlo goodness-of-fit tests

This appendix presents the set estimation method we propose to obtain a confidence set denoted $\mathcal{C}(Y)$ for the nuisance parameter ν associated with assumption (2.3). The set is obtained by "inverting" a GF test, of level α_1 , for the error distribution underlying (2.3). In other words, the confidence set corresponds to the set of ν_0 values that are not rejected by a GF test which assesses (2.3) imposing $\nu = \nu_0$. We use the multivariate skewness and kurtosis criteria introduced in Beaulieu et al. (2001b).⁶

$$\text{ESK}(\nu_0) = |\text{SK} - \overline{\text{SK}}(\nu_0)|, \quad (\text{A.1})$$

$$\text{EKU}(\nu_0) = |\text{KU} - \overline{\text{KU}}(\nu_0)|, \quad (\text{A.2})$$

where

$$\text{SK} = \frac{1}{T^2} \sum_{t=1}^T \sum_{i=1}^T \hat{d}_{ii}^3, \quad (\text{A.3})$$

$$\text{KU} = \frac{1}{T} \sum_{t=1}^T \hat{d}_{tt}^2, \quad (\text{A.4})$$

\hat{d}_{it} are the elements of the matrix $\hat{U}(\hat{U}'\hat{U})^{-1}\hat{U}'$ and $\overline{\text{SK}}(\nu_0)$ and $\overline{\text{KU}}(\nu_0)$ are simulation-based estimates of the expected SK and KU given (2.3). Conditional on $\overline{\text{SK}}(\nu_0)$ and $\overline{\text{KU}}(\nu_0)$, these tests satisfy the conditions of Theorem 2.1. Thus the MC test technique may be applied to obtain their corresponding exact p -values, $\hat{p}(\text{ESK}_0 | \nu_0)$, $\hat{p}(\text{EKU}_0 | \nu_0)$. To obtain a joint test based on these two statistics without relying on Bonferroni rules, we propose the following joint test statistic:

$$\text{CSK} = 1 - \min \{ \hat{p}(\text{ESK}_0 | \nu_0), \hat{p}(\text{EKU}_0 | \nu_0) \}. \quad (\text{A.5})$$

The MC test technique may once again be applied to obtain a size correct p -value for the combined test. For further details on the procedure, see Dufour et al. (2002).

⁶See also Dufour et al. (2002), Mardia (1970) and Zhou (1993).

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.
- Barone-Adesi, G. (1986), 'Arbitrage equilibrium with skewed asset returns', *Journal of Financial and Quantitative Analysis* **20**(3), 299–313.
- Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2001a), 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. (2001b), 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.
- Binder, J. J. (1985), 'Measuring the effects of regulation with stock price data', *Rand Journal of Economics* **16**, 167–183.
- 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.
- Chou, P.-H. (2000), 'Alternative tests of the zero-beta CAPM', *The Journal of Financial Research* **23**, 469–493.
- Deschamps, P. (1996), 'Monte Carlo methodology for LM and LR autocorrelation tests in multivariate regressions', *Annales d'Économie et de Statistique* **43**, 150–169.
- Dufour, J.-M. (1990), 'Exact tests and confidence sets in linear regressions with autocorrelated errors', *Econometrica* **58**, 475–494.
- 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. (2002a), Exact simulation based multiple hypothesis tests, Technical report, CIRANO and CIREQ, Université de Montréal, and Département d’économique, Université Laval.
- Dufour, J.-M. and Khalaf, L. (2002b), ‘Exact tests for contemporaneous correlation of disturbances in seemingly unrelated regressions’, *Journal of Econometrics* **106**(1), 143–170.
- Dufour, J.-M. and Khalaf, L. (2002c), 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. (2002d), ‘Simulation based finite and large sample tests in multivariate regressions’, *Journal of Econometrics* **111**(2), 303–322.
- Dufour, J.-M., Khalaf, L. and Beaulieu, M.-C. (2002), Exact skewness-kurtosis tests for multivariate normality and goodness-of-fit in multivariate regressions with application to asset pricing models, Technical report, CIRANO and CIREQ, Université de Montréal, and Département d’économique, Université Laval.
- Dufour, J.-M., Khalaf, L., Bernard, J.-T. and Genest, I. (2001), Simulation-based finite-sample tests for heteroskedasticity and ARCH effects, Technical report, C.R.D.E., Université de Montréal, and GREEN, Université Laval. 48 pages.
- Dufour, J.-M. and Kiviet, J. F. (1996), ‘Exact tests for structural change in first-order dynamic models’, *Journal of Econometrics* **70**, 39–68.
- Dufour, J.-M. and Kiviet, J. F. (1998), ‘Exact inference methods for first-order autoregressive distributed lag models’, *Econometrica* **66**, 79–104.
- 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.
- Engle, R. F. and Kroner, K. F. (1995), ‘Multivariate simultaneous generalized ARCH’, *Econometric Theory* **11**, 122–150.
- 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.
- 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.
- Godfrey, L. G. (1988), *Misspecification Tests in Econometrics : The Lagrange Multiplier Principle and Other Approaches*, Cambridge University Press, Cambridge, UK.
- Godfrey, L. G. (1996), 'Some results on the Glejser and Koenker tests of heteroscedasticity', *Journal of Econometrics* **72**, 275–299.
- 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.
- Harvey, A. C. and Phillips, G. D. A. (1982), 'Testing for contemporaneous correlation of disturbances in systems of regression equations', *Bulletin of Economic Research* **34(2)**, 79–81.
- 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.
- Kroner, K. F. and Ng, V. P. (1998), 'Modeling Asymmetric Comovements of Asset Returns', *The Review of Financial Studies* **11**, 817–844.
- Lee, J. H. H. (1991), 'A Lagrange multiplier test for GARCH models', *Economics Letters* **37**, 265–271.
- Lee, J. H. and King, M. L. (1993), 'A locally most mean powerful based score test for arch and garch regression disturbances', *Journal of Business and Economic Statistics* **11**, 17–27. Correction 12 (1994), 139.
- Lehmann, E. L. (1986), *Testing Statistical Hypotheses, 2nd edition*, John Wiley & Sons, New York.

- Ljung, G. M. and Box, G. E. P. (1978), 'On a measure of lack of fit in time series models', *Biometrika* **65**, 297–303.
- Lo, A. and MacKinlay, C. (1988), 'Stock prices do not follow random walks: Evidence from a simple specification test', *Review of Financial Studies* **1**, 41–66.
- Lo, A. and MacKinlay, C. (1989), 'The size and power of the variance ratio test in finite samples: A Monte Carlo investigation', *Journal of Econometrics* **40**, 203–238.
- MacKinlay, A. C. (1987), 'On multivariate tests of the Capital Asset Pricing Model', *Journal of Financial Economics* **18**, 341–372.
- 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.
- 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. Madala 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.
- Velu, R. and Zhou, G. (1999), 'Testing multi-beta asset pricing models', *Journal of Empirical Finance* **6**, 219–241.
- 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.