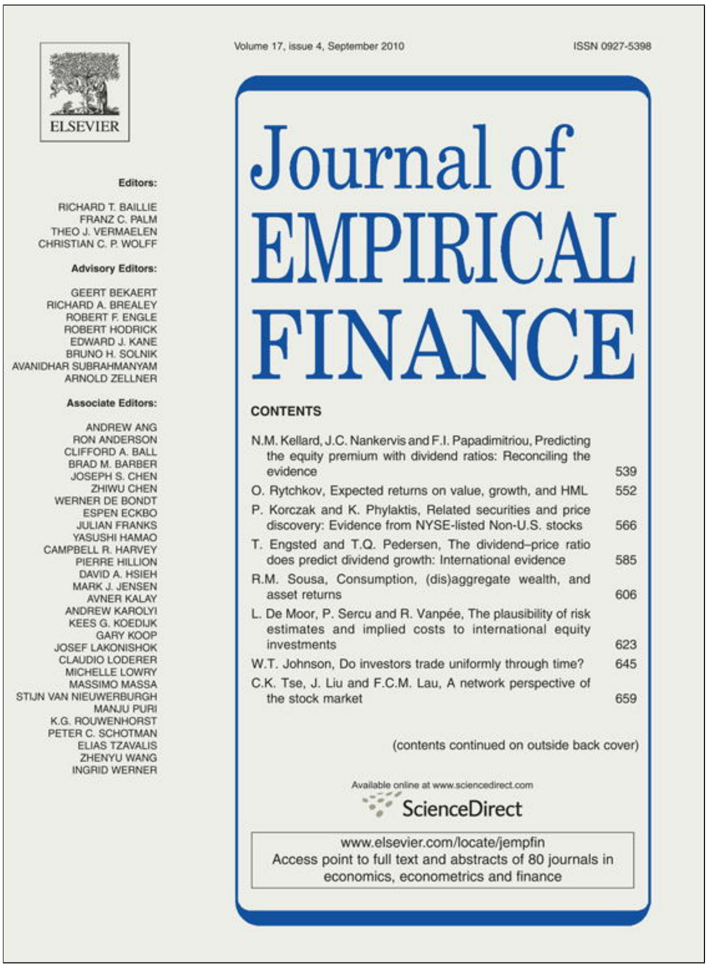

Volume 17, issue 4, September 2010 ISSN 0927-5398

Journal of EMPIRICAL FINANCE

CONTENTS

(contents continued on outside back cover)

Journal of Empirical Finance 17 (2010) 763–782

Contents lists available at ScienceDirect

Journal of Empirical Finance

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

Asset-pricing anomalies and spanning: Multivariate and multifactor tests with heavy-tailed distributions

Marie-Claude Beaulieu[a,1], Jean-Marie Dufour[b,*,2], Lynda Khalaf[c,3]

[a] Université Laval, Canada
[b] McGill University, Canada
[c] Carleton University, Canada

ARTICLE INFO

Available online 12 March 2010

Keywords:
Asset pricing models
Multifactor model
Mean-variance spanning
Non-normality
Multivariate linear regression
Exact test
Monte Carlo test
Bootstrap
Specification test
Diagnostics
GARCH
Variance ratio test.

ABSTRACT

In this paper we propose a multivariate regression based assessment of the multifactor model first developed by Fama and French (1993). We study mean-variance efficiency and spanning, as well as factor relevance. In particular, we assess the relative contribution of the factors in accounting for asset pricing anomalies. Our tests are motivated by a finite-sample distributional theory, invariant to portfolio repackaging, and achieve size control exactly conditioning on observed factors, in normal and non-normal contexts. We focus on the multivariate normal and Student-**t** distributions, in which case we rely on the simulation procedure proposed and applied in Beaulieu et al. (2007). We also assess, from a finite-sample and multivariate test perspective, the specification and fit of the model and error distributions considered. In its most general form, the model considered includes six factors: the market portfolio, size, the ratio of book equity to market equity as well as term structure variables (a term premium and a default premium) and momentum. Portfolio returns (coming from assets traded at NYSE, AMEX and NASDAQ) from Fama and French's data base are analyzed on monthly frequencies from 1961–2000.

Our results show the following. (1) Normality in model residuals becomes more dependable as a working hypothesis, over short time spans, when the book to market equity and size factors or when the momentum factor are accounted for. (2) Allowing for heavy tailed distributions empirically accommodates some stylized asset pricing anomalies. (3) Loadings on the term structure variables and the momentum factor seem (jointly, across portfolios) statistically insignificant at usual levels in many sub-periods. (4) Mean-variance efficiency is rejected in fewer subperiods allowing for non-normal errors in multi-factor settings; the book to market equity and size factors contribute importantly in reinforcing efficiency. (5) Enlarging the set of assets [from a one factor to a six factor model] does not reinforce the mean-variance spanning hypothesis, which is globally rejected at usual levels.

* Corresponding author. Tel.: +1 514 398 8879; fax: +1 514 398 4938.
E-mail addresses: Marie-Claude.Beaulieu@fas.ulaval.ca (M.-C. Beaulieu), jean-marie.dufour@mcgill.ca (J.-M. Dufour), Lynda_Khalaf@carleton.ca (L. Khalaf).
URL: http://www.jeanmariedufour.com (J.-M. Dufour).
[1] Chaire RBC en innovations financières, Centre interuniversitaire sur le risque, les politiques économiques et l'emploi (CIRPÉE), Département de finance et assurance, Pavillon Palasis-Prince, local 3620-A, Université Laval, Québec, Canada G1K 7P4. Tel.: +418 656 2131 2926; fax: +418 656 2624.
[2] William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec Canada, H3A 2T7.
[3] Groupe de recherche en économie de l'énergie, de l'environnement et des ressources naturelles (GREEN) Université Laval, Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Economics Department, Carleton University. Mailing address: Economics Department, Carleton University, Loeb Building 1125 Colonel By Drive, Ottawa, Ontario, Canada K1S 5B6. Tel.:+613 520 2600-8697; fax: +613 520 3906.

0927-5398/$ – see front matter © 2010 Published by Elsevier B.V.
doi:10.1016/j.jempfin.2010.03.001

1. Introduction

While the traditional capital asset pricing model (CAPM) developed by Sharpe (1964) and Lintner (1965) is still widely used to estimate the cost of capital for firms and evaluate the performance of managed portfolios – for reviews and references see Campbell et al. (1997 Chapters 5 and 6), Shanken (1996), Fama and French (2004), Perold (2004) and Sentana (2009) – the model proposed by Fama and French (1992, 1993) reveals that multifactor pricing models can be used for pricing anomalies in the CAPM. In fact, Fama and French (2004) argue that the average return anomalies of the CAPM suggest the following: if asset pricing is rational, a multifactor version of Merton (1973)'s intertemporal CAPM (ICAPM) or Ross (1976)'s arbitrage pricing theory (APT) can provide a better description of average returns (see *e.g.* Campbell and Vuolteenaho, 2004 two-beta model). While the excess market return of the CAPM is a relevant risk in many multifactor alternatives, like the ICAPM and most of its applications, the empirical evidence of a positive relation between beta and expected returns is not sufficient to favor CAPM over multifactor models given the documented empirical failures of the model.

The CAPM anomalies include Banz (1981)'s results that size adds to the explanation of average returns provided by the classic beta. The author documents that when stocks are sorted on market capitalization, average returns on small stocks are higher than predicted by the CAPM. This is a source of concern given that size is not the only anomaly of the CAPM. Variables which (unlike size) are not correlated with beta (such as earnings/price (E/P), cashflow/price, book equity to market equity ratio (BE/ME), and past sales growth) add to the explanation of average return provided by beta (Basu, 1977; Chan et al., 1991; Fama and French, 1992; Fama and French, 1993 and Lakonishok et al., 1994). Basu (1977) shows evidence that when common stocks are sorted on earnings–price ratios, future returns on high E/P stocks are higher than predicted by the CAPM. Bhandari (1977)'s results reveal that high debt–equity ratios of returns are too high relative to beta prediction of the CAPM. Stattman (1980) and Rosenberg et al. (1985) document that stocks with high book-to-market equity ratios have higher average returns than predicted by their beta. Furthermore, there is empirical evidence that stocks that have outperformed (underperformed) in the past may continue to do so (Jegadeesh and Titman, 1993, 2001). These authors attempt to explain the profitability of momentum strategies (portfolios constructed from past winners excluding or shorting past losers) using simple risk measures. Their results show that the intercept of the CAPM and of the three-factor model of Fama and French (1993) for winners and losers are both significantly different from zero. This is confirmed by Grundy and Martin (2001) who show, after adjusting for Fama and French three factors, the stability of profitability for momentum strategies; see also Carhart (1997).

In this context, we use the set of variables proposed by Fama and French (1993) to assess a multifactor model with finite sample methods that can depart from the normality assumption. The factors include the classic market portfolio, size, book to market as well as term structure variables and an additional factor capturing one year momentum anomaly. Furthermore, following the work of Jobson and Korkie (1982), Huberman and Kandel (1987), Korkie and Turtle (2001), Kan and Zhou (2008) and Mencia and Sentana (2009), we test whether enlarging the set of risky assets through additional factors allows an investor to improve the minimum variance frontier of the s benchmark assets given p new risky assets. In that respect our contribution is twofold, for we provide a finite sample test for *spanning* (see the survey of De Roon and Nijman, 2001) with an enlarged number of benchmarks allowing for non-normality. We also assess factor relevance from a non-normal finite sample and joint test perspective; such tests are, to the best of our knowledge, unavailable for the Fama-French three-six factor models. Their usefulness is definite, given the emerging literature on redundant factors (see *e.g.* Kan and Zhang, 1999a, b; Beaulieu et al., 2008b) and recent critiques on tight factor structures (Lewellen and Nagel, 2006); see also Lewellen et al. (2009), Beaulieu et al. (2008b) and Shanken and Weinstein (2006) on the importance of joint (across-portfolios) tests.

Since the work of Jensen (1968) and Gibbons (1982), empirical tests on various asset pricing models are often conducted within a multivariate linear regression (**MLR**) framework, which we also retain in this paper. In this framework, to test the hypothesis that given factors are sufficient to explain expected returns, one estimates the time series regression for a set of assets (or portfolios), and then jointly tests intercepts against zero. In forming portfolios in a certain way, one might expose shortcomings of the CAPM prediction that any factor other than the market is not necessary to explain expected returns. To determine whether the intercepts in a MLR setting are zero, early tests shared the same asymptotic properties, while it was not clear which had the best finite-sample properties. Indeed, several studies in econometrics and finance have shown that standard asymptotic theory provides a poor approximation to the finite-sample distribution of MLR-based tests, even with fairly large samples; see Campbell et al. (1997 Chapters 5 and 6), Dufour and Khalaf (2002), Beaulieu et al. (2005, 2007, 2009) and the references therein. In particular, test size distortions grow quickly as the number of portfolios (equations) increases. This is an important issue. The fact that we have long time series does not allow us to ignore the finite sample properties of the tests we use given that most financial series are not temporally stable over long horizons. Furthermore, in many cases, given the length of the time series and the number of equations in the MLR system, the number of degrees of freedom left to run the tests are limited and again, one has to be concerned about the finite sample properties of the chosen test. As a result, the conclusions of financial MLR-based empirical studies can be strongly affected and may lead to spurious rejections.

In this context, as emphasized by Shanken (1996), Campbell et al. (1997) and Sentana (2009), applying finite-sample statistical methods is particularly important. Consequently, a number of authors have proposed tests based on a finite-sample distributional theory in order to assess mean-variance efficiency. Early works may be traced back to Jobson and Korkie (1982), Shanken (1986), MacKinlay (1987) and Gibbons et al. (1989, henceforth GRS). In particular, the zero-intercept test from GRS is still viewed as a fundamental contribution, despite its reliance on normality. GRS' test exploits results from the statistical literature on multivariate analysis-of-variance, in particular on Hotelling's T^2 statistic which may be transformed into a statistic with an exact Fisher distribution assuming normality. More importantly, from the asset pricing perspective, GRS also emphasize the test economic interpretation in addressing whether the market proxy is the tangency portfolio in the set of portfolios that can be constructed by combining the market portfolio with the specific assets used as dependent variables in the time series regressions. Beaulieu et al. (2007) provide a simulation-based extension of GRS' procedures for mean-variance testing in non-normal contexts; for reviews and further references, see e.g. Stewart (1997), Amengual and Sentana (2010), Gungor and Luger (2009) and Sentana (2009).

In view of its fundamental importance, mean-variance efficiency is one of the first and very few MLR-based asset pricing models that have been approached from an exact perspective. Nevertheless, a few authors have recognized that hypotheses dealing with the joint significance of the coefficients of *two* regression coefficients across equations can also be tested exactly applying related F tests. Examples include inter-temporal asset pricing tests in Shanken (1990, see footnote 18). Furthermore, as discussed in Shanken (1996), econometric tests of *spanning* fall within this class. Indeed, spanning tests (see Jobson and Korkie, 1989, Kan and Zhou, 2008) may be written in terms of a model of the GRS form. The hypothesis is however more restrictive than GRS', in the sense that over and above the restriction on the intercepts, the betas for each regression are required to sum to one. The results in the present paper extend available exact tests of these important financial problems beyond the Gaussian context.

It has long been recognized that financial returns do not exhibit normality (see *e.g.* Fama, 1965; Baillie and Bollerslev, 1989 and Beaulieu et al., 2005, 2007, 2009), though normality is not necessary for various asset pricing models (see *e.g.* Ingersoll, 1987; Berk, 1997; Beaulieu et al., 2005 and the references therein). Empirically however, return non-normality may not necessarily lead to non-normal factor model residuals, particularly in multifactor settings. In this paper, we study (using the finite sample methods from Dufour et al., 2003) whether commonly considered factors have a statistical "normalizing" effect on short horizons. Formally, we study whether the addition of specific factors over and above a market proxy may render multinormality acceptable as a working hypothesis. To the best of our knowledge, this analysis is also unavailable from a finite sample perspective, for the basic Fama-French multifactor pricing model, or for models with momentum.

Considering the importance of the distributional assumption for the quality of inference, one has to choose how to address that problem when the normality hypothesis is convincingly non-tenable. Various methods are currently adopted for this purpose in asset pricing. Normality is typically assumed away via large-sample procedures (such as GMM) based on the central limit theorem, in which case resampling techniques (such as the non-parametric bootstrap) are increasingly considered motivated by asymptotic refinements that promise, at least in principle, better finite-sample performance; for some examples in the context of beta pricing models including the CAPM, see Affleck-Graves and McDonald (1989), Fama and French (1993), MacKinlay and Richardson (1991), Jagannathan and Wang (1996), Ferson and Harvey (1999), Groenwold and Fraser (2001), Campbell and Vuolteenaho (2004) and Mencia and Sentana (2009). In parallel, and partly because of related finite sample issues, Bayesian methods (see e.g. Kandel et al., 1995; MacKinlay and Pastor, 2000 and the discussion in Sentana, 2009) have gained popularity. A more specialized approach restricts attention to procedures for elliptical distributions: examples include the alternative kernel-based semiparametric asymptotic procedure proposed by Hodgson et al. (2002), Vorkink (2003) and Hodgson and Vorkink (2003).[4] Within the elliptical family, focus on parametric procedures fitting non-normal distributions (such as the multivariate t and multivariate mixtures) constitute another modelling alternative; see *e.g.* Zhou (1993), Fiorentini et al. (2003), Geczy (2001), Kan and Zhou (2008), Amengual and Sentana (2010) and Mencia and Sentana (2009). Such approaches, that are also for the most part asymptotic, have the fundamental advantage of removing indeterminacies associated with very weak distributional assumptions and thus promise finite sample improvements (relative to *e.g.* moment-based central limit theory underlying *e.g.* GMM).

In this paper, we pursue this parametric line of work, with focus on the finite sample motivated procedures proposed by Beaulieu et al. (2007) and Dufour et al. (2003, 2010) .[5] These procedures involve testing both the usual cross-equation parameter restrictions entailed by a multifactor asset pricing model as well as various specification tests, including tests of normality and distributional goodness-of-fit, and tests against the presence of ARCH effects and serial dependence, which take into account the multivariate nature of the model. All tests applied involve the restricted and unrestricted least-squares [**LS**] residual matrices from a MLR model that describes returns, and are finite-sample exact in normal and non-normal settings, conditional [as with all available exact procedures in this literature] on observed factors. Under the usual weak exogeneity assumptions which validate regression based asset pricing tests, our methods remain asymptotically size correct and provide much better finite-sample fit than usual asymptotic tests. Normal and non-normal test p-values are obtained via the Monte Carlo [**MC**] test method (described below; see Dufour, 2006).

Perhaps most importantly, all tests we use (on the factor loadings and on the specification and fit) are invariant to portfolio repackaging (see Kandel and Stambaugh, 1989). Specifically, although we formally account for the model *betas* and for the residual return variances and covariances via their LS estimates, our test procedures are invariant to the unknown values of these parameters. In other words, the null distributions we obtain do not depend on the loadings nor on the error variance-covariance matrix (Σ). These properties may be seen via the following motivating example.

In the LS regression of a $T\times n$ matrix Y of portfolio returns on a $T\times k$ matrix X of factors (including an intercept), the LS residual matrix $\hat{U}$ satisfies $\hat{U}=(I-X(X'X)^{-1}X')U$ where U refers to the error term. This means that $\hat{U}$ depends on the distribution of U [and thus on Σ], though regression coefficients are "projected out". Invariance to factor loadings obtains in our context using related projections, for constrained and unconstrained residuals. For the constrained case, this result is less trivial since it does not automatically follow from a general linearity assumption (see Dufour and Khalaf, 2002). Furthermore, since $\hat{U}$ depends on U, the fact that Σ is, nevertheless, "evacuated out" from the null distribution of all test statistics we use, is a nontrivial and quite

[4] Note that Hodgson et al. (2002) report size problems on high dimensional systems and restrict their analysis to systems with a small number of portfolios. Furthermore Vorkink (2003) specifies that the procedure was applied on a portfolio-by-portfolio basis, not on a multivariate estimation system.

[5] Our method is parametric, yet is not restricted to the elliptical family; see (Beaulieu et al., 2005) for an application with Stable distributions.

consequential outcome, particularly for models with many portfolios.[6] To the best of our knowledge, this property has not been established as yet, for non-Gaussian multifactor spanning tests.

The above outcomes stem from the so-called location-scale or rotation invariance principle [invariance to multiplication by an invertible matrix], which for asset pricing practice, entails the following. Suppose the return vector (r_t, $t=1, ...,T$) is transformed into $r_t^* = Ar_t$ where A is an $n \times n$ invertible matrix such that $A\iota_n = \iota_n$, which would correspond to conducting inference on another set of portfolios. Our tests will not be affected by such a mapping, *i.e.* will yield the same decision whether derived from r_t or r_t^* . This important property, which is not straightforward in non-normal contexts, has regained attention lately given recent critiques (see Lewellen and Nagel, 2006 and Lewellen et al., 2009) on using portfolios sorted according to the Fama and French (1993) factors.

From the empirical perspective, this paper has five main contributions. The tests proposed are applied to an asset pricing model which, in its most general form, includes six factors: the market portfolio, size, the ratio of book equity to market equity as well as term structure variables (a term premium and a default premium) and momentum. Portfolio returns (coming from assets traded at NYSE, AMEX and NASDAQ) are all extracted from Fama and French's data base on Kenneth French web site. We produce results for monthly returns of 25 value weighted portfolios from 1961–2000, under the multivariate normal and the multivariate Student-**t** distribution.[7] First, we assess mean-variance efficiency in a multifactor context. Second, we test the hypothesis that the frontier spanned by factor portfolios is the same as the frontier spanned by twenty-five stock portfolios. Third, we assess the joint (across these portfolios) significance of the factors, and analyze the implications on this issue on efficiency and spanning. Fourth, we analyze the impact of factors on residual normality. Fifth, we document the fit of the Student-**t** distribution, controlling for the factors. Our analysis thus sheds light on pricing anomalies along these five related directions. Our results may be interpreted with some generality as allowed by the invariance to portfolio repackaging property. When commitment to given factors for portfolio sorts and model tests becomes difficult (following critiques such as those raised by e.g. Lewellen and Nagel, 2006 and Lewellen et al., 2009), test methods that are invariant-to-repackaging provide sound model assessment practices.

Our key results can be summarized as follows. (1) In general, normality in model residuals becomes more dependable as a working hypothesis, over short time spans, when the book to market equity and size factors or when the momentum factor are accounted for. In contrast, the term structure variables do not play an important “normalizing” role. (2) Overall, the Student-**t** distribution for residuals passes our tests almost everywhere. In view of the above cited literature which exploits the Student-**t** hypothesis, our result is useful and suggests that allowing for heavy tailed distributions may empirically accommodate some stylized asset pricing anomalies. The latter conclusion must however be weighed against further evidence from our factor relevance, spanning and efficiency tests. (3) On these tests, χ^2 based p-values [denoted p_∞ in Section 3] are, in far too many cases, much smaller than their simulated counterparts [denoted p_N and p_T for the normal and the Student-**t** case respectively], leading to noticeable test decision changes. In general [except for tests on the joint significance of the market, book to market equity and size factors, where all p-values are small], p_N is smaller than p_T, yet the discrepancy between p_∞ and p_N is much more marked and more consequential than that between p_N and p_T. (4) Our tests reveal that the term structure variables and the momentum factor seem irrelevant [loadings on these factors seem statistically insignificant at usual levels] in many sub-periods. Whereas some recent evidence (see e.g. Shanken and Weinstein, 2006 in cross-sectional tests) may corroborate our findings on the term structure factors, our results on momentum are noteworthy, particularly since to the best of our knowledge, this paper provides the first study on factor relevance with finite sample motivated methods. (5) Mean-variance efficiency is rejected in fewer subperiods allowing for non-normal errors in multi-factor settings; nevertheless, the zero-intercept hypothesis is not uniformly upheld even in the six factor model. The “normalizing” effect of some factors does not necessarily translate into better model fit, since abnormal excess returns (measured via MLR intercepts) remain, in some subperiods, significant at conventional levels. (6) Enlarging importantly the set of assets (going from a one factor to a six factor asset pricing model) does not lead to fewer rejections of the mean-variance spanning hypothesis. Spanning is not rejected at 5% [based on p_N and p_T although $p_\infty < 5\%$] in a few subperiods for a model including the market proxy in addition to the term structure variables or the momentum factor; this result must be interpreted with caution, since the relevance of these factors was questionable, based on the joint tests we performed.

The paper is organized as follows. Section 2 describes our econometric methods. In particular, our model and underlying hypotheses are defined in Section 2.1; our inference methods are discussed in Sections 2.2 and 2.3, whereas companion model diagnostics are briefly summarized in Section 2.4. In Section 3 we present the empirical analyses. We conclude in Section 4. Technical appendices provide comprehensive formulas, algorithms and further references that document our econometric methods.

2. The econometric methodology

We consider the following MLR framework to describe portfolio returns:

$$r_{it} = a_i + \sum_{j=1}^{s} b_{ij}\tilde{r}_{jt} + u_{it}, t = 1, ..., T, i = 1, ..., n, \tag{2.1}$$

where $r_{it} = R_{it} - R_t^F$, R_{it}, $i=1, \ldots, n$, are returns on n (25 in our case) portfolios for period t, $t=1, ..., T$, R_t^F is the riskless rate of return, $\tilde{r}_{jt} = \tilde{R}_{jt} - R_t^F$, $\tilde{R}_{jt}$, $j=1, ..., s$ are returns on s benchmark factors, and u_{it} is a random disturbance. In the most general form of our

[6] The pioneering works of *e.g.* (Shanken, 1986), (MacKinlay, 1987), (Gibbons et al., 1989), (Jobson & Korkie, 1989) and (Zhou, 1993) which exploited similar invariance results have been quite influential for regression based testing.

[7] Note that we ran the same analysis over equally-weighted portfolios and that results remained similar overall.

multifactor model, the factors considered include: the return on the market portfolio (R_t^M), the average return on three small portfolios minus the average return on three large portfolios (SMB), the average return on two value portfolios minus the average returns on two growth portfolios (HML), the difference between the monthly long-term government bond return and the one-month Treasury bill rate measured at the end of the previous month (TERM), the difference between the return on a market portfolio of long-term corporate bonds and the long-term government bond return (DEF), and the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios (MOM).

2.1. Spanning, mean-variance efficiency, and factor relevance

We focus on the spanning hypothesis, which in this framework implies that the intercepts are zero *and* the betas for each portfolio sum to one (see Huberman and Kandel, 1987; Jobson and Korkie, 1989; De Roon and Nijman, 2001; Penaranda and Sentana, 2008, Kan and Zhou, 2008), or formally:

$$a_i = 0, \sum_{j=1}^{s} b_{ij} = 1, \quad i = 1, \ldots, n. \tag{2.2}$$

We also study the efficiency hypothesis:

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

and assess the significance of each factor, *i.e.* we test whether the coefficients of each factor ($\tilde{r}_{jt}$) are jointly equal to zero; formally, we test (in turn) for $j = 1, \ldots, s$, whether

$$b_{ij} = 0, i = 1, \ldots, n. \tag{2.4}$$

Standard exact methods for testing these hypotheses rely on the assumption that the disturbance vectors are *i.i.d.* Gaussian, namely:[8]

$$U_1, \ldots, U_T \text{ are i.i.d.} N[0, \Sigma], \quad U_t = (u_{1t}, \ldots, u_{nt})', t = 1, \ldots, T. \tag{2.5}$$

Since security returns tend to follow more heavy-tailed distributions than the normal, this may be unduly restrictive. In this paper, and following Beaulieu et al. (2007), we use a more comprehensive statistical theory which yields finite-sample tests assuming that

$$U_t = JW_t, t = 1, \ldots, T, \tag{2.6}$$

where J is an unknown, non-singular matrix and the joint distribution of $W_t, t = 1, \ldots, T,$ is either known, or is a fully specified function of a parameter κ.

The parameter κ may be known (in which case the joint distribution of $W_1, \ldots, W_T$ is free of nuisance parameters), or may be unknown and needs to be estimated from the data (or in other words, κ intervenes as a nuisance parameter in the test problem). Assumption (2.5) is thus a special case of Eq. (2.6) where $W_1, \ldots, W_T$ are *i.i.d.* $N[0, I_n]$. We also focus on the special case where $W_1, \ldots, W_T$ are *i.i.d.* such that

$$W_t = Z_{1t} / (Z_{2t}/\kappa)^{1/2}, \tag{2.7}$$

where $Z_{1t} \sim N[0, I_n]$ and $Z_{2t} \sim \chi^2(\kappa)$ and is independent of Z_{1t}, that is W_t follows a Student-**t** distribution with degrees-of-freedom parameter κ.[9]

The idea here is to disentangle the following two components from the distribution of error terms: (i) the variate W_t which may be interpreted as a *normalized disturbance* term, so that the distribution of $W_t, t = 1, \ldots, T,$ gives the fundamental data generating process [DGP]; and (ii) the matrix J which sets "the scale", which we define as

$$\Sigma = JJ'$$

[8] On exact asset pricing tests, see (MacKinlay, 1987), (Jobson and Korkie, 1989), (Gibbons et al., 1989), (Shanken, 1996), ((Campbell et al., 1997), Chapters 5 and 6), (Beaulieu et al., 2007) and the references therein. For further discussion of these tests in the statistics and econometrics literature, the reader may consult ((Anderson, 1984), chapters 8 and 13), (Berndt & Savin, 1977), (Dufour & Khalaf, 2002), (Kariya, 1985), (Kariya & Kim, 1997), ((Rao, 1973), chapter 8) and (Stewart, 1997).

[9] If the χ^2 term underlying the simulable form we use for the multivariate Student-t distribution is replaced by a *gamma* distributed variate, non-integer values for κ can be accommodated. Our theoretical framework does not require non-integer values [which is rarely the case with usual asymptotic tests based on usual asymptotic theory]; this point is worth noting.

so that $det(\Sigma)\neq 0$. This decomposition allows to clearly identify all unknowns which intervene in the null distribution of the tests we propose.

2.2. Non-Gaussian tests

One of the most commonly used statistics (see Dufour and Khalaf, 2002 and the references therein) to test hypotheses (2.2)–(2.4) is the Gaussian Likelihood Ratio [LR] criterion, which derives in this model from the constrained and unconstrained estimates of Σ. The latter, denoted $\hat{\Sigma}$, coincides with the least-squares [LS] estimator, and all constrained estimates, for the problems at hand, have closed form non-iterative LS based solutions; refer to Appendix A, Eq. (A.3) for a formal definition. In this paper, we use this statistic although we relax normality, following the method applied in Beaulieu et al. (2007) which allows one to obtain valid p-values for this statistic under Assumption (2.6). The method and underlying theory is presented, for completion, in Appendices A and B. For the sake of clarity, we summarize, discuss and motivate the main steps involved in what follows.

Our methodology is based on the following fundamental result [formally stated in Eqs. (A.3) and (A.4)]: under the null hypotheses considered and Assumption (2.6), the LR test statistic is a *pivotal function* of [*i.e.* depends exclusively on] realizations of the factors r_{jt}, $j=1, ..., s$, and the normalized disturbances W_t, $t=1, \ldots, T$. In other words, the null distribution of LR does not depend on the factor loadings nor on the scale matrix Σ. On recalling the dimension of Σ, the fact that it is evacuated out is quite useful, particularly for models where n is large relative to T. From an asset pricing perspective, this property results from invariance to portfolio repackaging (see Kandel and Stambaugh, 1989), *i.e.* to transformations of the returns vector $r_t=(r_{1t}, ..., r_{nt})'$ into $r_t^*=Ar_t$ where A is an $n\times n$ invertible matrix such that $A\iota_n=\iota_n$. Our results imply that the LR tests we run yield the same decision [as may be seen from the algorithm below] whether derived from r_t or from r_t^*.

Beside its theoretical appeal, the above result provides the basis to derive simulation based p-values for the LR statistic, conditional on the sample realizations of $\tilde{r}_{jt}$, and given draws from W_t, $t=1, \ldots, T$, conforming with Eq. (2.6). Indeed, for each such draw, a simulated value for LR can be calculated using the *pivotal expression* Eq. (A.4). There is no need to reconstruct an artificial sample as is usual in the bootstrap literature, which avoids reliance on estimates of the factor loadings and of Σ. To see this, consider a regular parametric bootstrap method which would in our case be run as follows: (i) draw a sample path from the normalized disturbances; (ii) use the *constrained* [*i.e.* imposing Eq. (2.2), (2.3) or (2.4)] estimates of the factor loadings and of Σ to reconstruct, given the sample realizations of $\tilde{r}_{jt}$, bootstrap simulated returns conforming with regression (2.1); and (iii) derive a simulated value for LR using the latter simulated samples. Our invariance results simply imply that one would obtain exactly the same numerical value for the simulated LR if following step (i), the simulated W_t are directly *plugged* into the function (A.4).

Given N simulated values of LR [obtained given N draws from the normalized disturbances], an empirical p-value [denoted $\hat{p}_N$] can be calculated from the *rank* of the observed relative to the simulated criteria, or formally, from the number of simulated statistics that exceed the value calculated from the observed sample; refer to formula (A.5) in Appendix A.

Our discussion has so far left out the specifics underlying step (i) in the above procedure. Yet the method validity depends crucially on this step. In particular, if no unknown parameters intervene in simulating the normalized disturbances, then it can be shown (see Dufour, 2006) that the size of the test so obtained (conditional on the realizations of the factors) is correct for any finite sample size T and any finite number of Monte Carlo replications N.[10] This occurs in particular under the i.i.d. Student-**t** hypothesis with known κ. Here we treat κ as an unknown parameter, so we proceed as follows. First, we derive a confidence set for κ [denoted $C(\kappa)$] of level α_1, conformable with our maintained model and assumptions. Each value of κ in the latter set would yield a MC p-value for the LR statistic, denoted $\hat{p}_N(\kappa)$ [refer to Eq. (A.5) in Appendix A]. To guard against spurious rejections, at level α, for all values of κ consistent with our data, we base our decision on the largest such p-value: formally, we reject the null hypothesis if

$$\sup_{\kappa\in C(\kappa)} \hat{p}_N(\kappa)\leq(\alpha-\alpha_1).$$

The resulting procedure has level α for any DGP compatible with the null hypotheses under consideration.

The confidence set for κ we derive is size correct under the same assumptions which validate the above described MC LR test with known κ. This set is built by inverting an exact multivariate moment-based goodness-of-fit test for a Student-**t** distribution on the innovations, with known κ. Concretely, this implies collecting the values of κ that are not rejected [here at level α_1] by the test in question. The test statistic we use is defined in Eq. (B.8) in Appendix B and is denoted $CSK_{\mathrm{M}}(\kappa)$.[11] We also run alternative tests for multivariate normality [discussed in Appendix B] all of which satisfy, as does CSK_{M}, the same location-scale invariance property we obtained for the above described LR test. Indeed (see Dufour et al., 2003), all these statistics depend on the data exclusively via the standardized residuals matrix [denoted $\tilde{W}$ and formally defined in Eq. (B.4)], that is, the matrix of OLS residuals from regression (2.1) orthogonalized via the Cholesky factor of $\hat{\Sigma}$. Most importantly, under the model assumptions, the standardized residuals can be rewritten as a sole function of the factors $\tilde{r}_{jt}$ $j=1,...,s$ and of the standardized disturbances, which once again

[10] Note the division by $N+1$ in formula (A.5). Type I error control obtains, as shown in (Dufour, 2006), because under the null hypothesis, the N simulated statistics and the observed LR are exchangeable. Distributional theory on the *rank* of any observation [here, of the observed LR] within a series of exchangeable random variables [here the $N+1$ observed and simulated LRs] is used to prove size control. A series of random variables indexed $Z_1, Z_2, ..., Z_n$ are exchangeable when for any finite permutation of the indices [here, 1, ..., n], the permuted series have the same joint distribution as the original one. An *i.i.d.* series is exchangeable, the condition is however weaker than *i.i.d.*

[11] *CSK* stands for "combined Skewness and Kurtosis", conditioning on κ is emphasized, and the subscript M refers to (Mardia, 1970) since the moments we use are modifications of Mardia's (1970) measures.

establishes their invariance to the factor loadings and to Σ (thus to portfolio repackaging) leading to a straightforward application of Monte Carlo test methods.

2.3. The conditioning setting

The above discussed test procedures provide exact (for finite T and N) type I error control assuming we can condition on the full realization of the factors $\tilde{r}_{jt}$, $t=1,...,T$, $j=1,...,s$, i.e. assuming we can consider these factors as fixed for statistical analysis. In other words, if a MC p-value is calculated as described above, conditioning on the full-sample realization of the factors, and is referred to a level α, then the associated rejection probability (for finite T and N) under the null hypothesis will not exceed α. This holds true for all available finite sample tests in this literature.

The fact remains that the LR statistic we use is asymptotically pivotal under fairly general regularity conditions which include the weak exogeneity assumption, in which case (as emphasized in e.g. MacKinlay and Richardson, 1991 and Sentana, 2009) LR coincides with GMM estimation. So in particular, if errors are i.i.d. and first and second moments exist [which covers the distributional assumption we focus on in Section 3, namely Eq. (2.7) with $\kappa \geq 2$], and assuming we can condition on the contemporaneous (and past) realizations of the factors $\tilde{r}_{j,t-s}$, $s=0,...,t-1$, $j=1,...,s$, the LR statistic admits a χ^2 asymptotic null distribution with degrees of freedom that depend exclusively on the number of constraints tested [here, n for Eqs. (2.3) and (2.4), and $2n$ under Eq. (2.2)]; see e.g. Gouriéroux et al. (1995), Anderson (1984, chapters 8 and13), Berndt and Savin (1977), Kariya (1985), Rao (1973, chapter 8)). Clearly, the null distribution of the LR statistic will converge to this same distribution if the factors are further restricted and assumed fixed.

This property, namely the fact that the statistic's null distribution is the same under weak and strong exogeneity, ensures the asymptotic validity of our MC test procedure given a weak exogeneity assumption that fits more realistically within the asset pricing framework we consider here. Indeed, as shown in Coudin and Dufour (2009) and Dufour and Kiviet (2004) in a generic setting, when a MC p-value is computed given the full realization of some conditioning variable (say X), and where in this case, the underlying test statistic converges to any distribution (say $\bar{F}$) that does not depend on the realization of X, the associated MC test will remain asymptotically valid under any set of weaker assumptions for which the statistic still converges (not necessarily at the same rate) to the same limiting distribution $\bar{F}$. For the tests discussed in Section 2.2, and assuming, within the setting of Section 2.1, the usual linear regression based weak exogeneity hypotheses, the results of Coudin and Dufour (2009) and Dufour and Kiviet (2004) imply the following; if a MC p-value is calculated as described above, conditioning on the full realizations of the factors, and is referred to a level α, then the limiting (for $T\rightarrow\infty$ and finite N) rejection probability (under the null hypothesis) of the associated MC test will not exceed α. So to sum up, our MC tests though exact with fixed regressors remain asymptotically valid under the usual set of assumptions typically required in regular regression settings (such as regressions with lagged endogenous variables). The associated tests will nevertheless perform much better in finite samples than their usual χ^2 counterparts.[12]

The generic conditions from Coudin and Dufour (2009) and Dufour and Kiviet (2004) suggest that our MC tests may in fact be valid given even milder assumptions on the joint distributions of the factors $\tilde{r}_{jt}$, $j=1,...,s$, and the normalized disturbances W_t, $t=1,...,T$. In particular, weaker conditions on the existence of moments and on convergence rates may be formulated to justify inferences for more heavy tailed distributions (as considered in Beaulieu et al., 2005). Such an analysis (which would clearly constitute a worthy theoretical extension to the MLR testing problems analyzed in e.g. Dufour and Khalaf, 2002 and Dufour et al., 2003, 2010) is beyond the scope of the present paper, because our empirical analysis focuses on Assumption (2.7) that fits within the usual regular regression framework.

In principle, conditional heteroskedasticity of the GARCH form, or arising from joint ellipticity assumptions on the returns and factors (see Zhou, 1993; Mencia and Sentana, 2009; Amengual and Sentana, 2010, the survey in Sentana, 2009 and Beaulieu et al., 2007) may invalidate the size of our tests. However, Assumption (2.6) is, in principle, not restricted to the intertemporal i.i.d. setting. Indeed, dynamic structures including GARCH processes may be fit via appropriate specifications for the distribution of W_t. A MC p-value which would control for these specifications can be computed where nuisance parameters can be dealt with by maximization (following the treatment of κ in the present paper). To see this, recall that in the Student-**t** case, we do obtain a valid non-normal test using the Gaussian LR statistic; similarly, one can still use the same statistic allowing for conditional heteroskedasticity, provided its p-value is corrected appropriately.

For instance, in the context of non-linear asset pricing models, Beaulieu et al. (2008a,b) consider Gaussian-GARCH processes for W_t, $t=1,...,T$ and apply a maximized MC test based on the Gaussian LR statistic, where the p-value is maximized over the values of the GARCH parameter. Interestingly, a simulation study reveals that such a correction has a negligible effect on inference in the sense that the size and the power of the uncorrected MC LR test do not deviate much from the size and power of the corrected test, even when GARCH is present in the DGP. For a related test problem, Beaulieu et al. (2007) also show (given a specific data set) that correcting inference to account for heteroskedasticity linked to conditioning on market returns does not affect the test decision.

Of course, it can be argued that the i.i.d. LR statistic (even if it is possible to correct its p-value) may not be the most efficient one in the presence of serial dependence. Indeed, following the work pioneered in this literature by MacKinlay and Richardson (1991), Wald-HAC [heteroskedasticity and autocorrelation consistent] statistics may also be considered. It is however well known that such statistics suffer from the curse of dimensionality as much as (and perhaps even more than) their i.i.d. counterparts; for recent

[12] For a review of the simulation-based tests literature which documents this fact, see *e.g.* (Dufour, 2006), (Dufour and Khalaf, 2001) and the references therein.

simulation evidence in asset pricing contexts, see Ray and Savin (2008), Beaulieu et al. (2008a), and Gungor and Luger (2009). Despite the semi-parametric motivation of HAC statistics, one may thus consider to correct their size given parametric special cases [e.g. GARCH] compatible with the asset pricing model under test. For such a strategy, Beaulieu et al. (2008a) show that expected efficiency gains may not materialize into effective power gains; in contrast, an i.i.d. MC LR test (following the method we apply in the present paper, for a different although related asset pricing problem) seems to perform markedly well in finite samples, even in GARCH contexts.

2.4. Diagnostic tests

In view of the above discussion, the specification of our model is carefully evaluated using the finite-sample residual-based diagnostics proposed in Dufour et al. (2003) that are specifically designed for MLR models. These tests assess the hypothesis of i.i.d. disturbances, against linear serial dependence (using multivariate extensions of the variance-ratio tests from Lo and MacKinlay, 1988, 1989) and against the presence of GARCH-type effects (using multivariate versions of tests suggested by Engle, 1982 and Lee and King, 1993). The tests are described in Appendix C, and may be summarized as follows. A diagnostic statistic for each equation in Eq. (2.1) is derived using the same standardized residuals [specifically, the rows of the matrix $\widetilde{W}$ defined in Eq. (B.4)] considered for the above goodness-of-fit tests. This ensures that the joint distribution of all criteria (over all equations) is a pivotal function (free from the loadings and from Σ) of the factors $\tilde{r}_{jt}$ and the normalized disturbances, and is thus invariant to portfolio repackaging. As with all above defined tests which satisfy this property, we obtain Monte Carlo p-values for each individual criteria, and retain the smallest p-value over all equations as a combined test statistic. The Monte Carlo method is then applied again to the combined statistic (as in double bootstrap or a double Monte Carlo test; see Dufour et al., 2003). These tests are performed under the different error distributions considered, namely the multivariate normal and the Student-**t**, in which case we rely on the maximized Monte Carlo p-value over all values of κ in the same α_1-level confidence set $C(\kappa)$ we use for the LR spanning, efficiency and factor significance tests. All diagnostics are valid in finite and large samples following the assumptions discussed above.

3. Empirical analysis

3.1. Data

In our empirical analysis of a multifactor asset pricing model, we use Fama and French's data base. We produce results for monthly returns of 25 value weighted portfolios from 1961 to 2000. The portfolios which are constructed at the end of June, are the intersections of five portfolios formed on size (market equity) and five portfolios formed on the ratio of book equity to market equity. The size breakpoints for year s are the New York Stock Exchange (NYSE) market equity quintiles at the end of June of year s. The ratio of book equity to market equity for June of year s is the book equity for the last fiscal year end in $s-1$ divided by market equity for December of year $s-1$. The ratio of book equity to market equity are NYSE quintiles. The portfolios for July of year s to June of year $s+1$ include all NYSE, AMEX, NASDAQ stocks for which market equity data is available for December of year $s-1$ and June of year s, and (positive) book equity data for $s-1$. The benchmark factors are: 1) the excess return on the market, defined as the value-weighted return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) minus the one-month Treasury bill rate (from Ibbotson Associates), 2) SMB (small minus big) defined as the average return on three small portfolios minus the average return on three big portfolios, 3) HML (high minus low) defined as the average return on two value portfolios minus the average return on two growth portfolios, 4) TERM the difference between the monthly long term government bond return (from Ibbotson Associates) and the one-month Treasury bill rate measured at the end of the previous month, 5) DEF the difference between the return on a market portfolio of long term corporate bonds (the Composite portfolio on the corporate bond module of Ibbotson Associates) and the long term government bond return and finally 6), MOM, the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios. Fama and French benchmark factors, SMB and HML, are constructed from six size/book-to-market benchmark portfolios that do not include hold ranges and do not incur transaction costs. The portfolios for these factors are rebalanced quarterly using two independent sorts, on size (market equity, ME) and book-to-market (the ratio of book equity to market equity, BE/ME). The size breakpoint (which determines the buy range for the small and big portfolios) is the median NYSE market equity. The BE/ME breakpoints (which determine the buy range for the growth, neutral, and value portfolios) are the 30th and 70th NYSE percentiles. For the construction of the MOM factor, six value-weighted portfolios formed on size and prior (2–12) returns are used. The portfolios, which are formed monthly, are the intersections of two portfolios formed on size (market equity, ME) and three portfolios formed on prior (2–12) return. The monthly size breakpoint is the median NYSE market equity. The monthly prior (2–12) return breakpoints are the 30th and 70th NYSE percentiles.

3.2. Multifactor models assessment

Our results are presented in Tables 1–8. These tables regroup normality tests on the residuals of different model specifications that include the classic one beta asset pricing model as well as a more general multifactor model with six factors. We also present, in each case, finite-sample correct multivariate assessments of the joint significance of: (i) the vector of coefficients on each factor (factor loadings) across the 25 portfolios, and (ii) the vector of intercepts. These tests serve three different although related purposes. First, we can assess which (and how many) factors are necessary for residuals to respect the normality assumption.

Table 1
Model with MARKET.

A	Normality Tests								Student
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	SK_M	KU_M	CSK_M	MSK	SK_{KD}	KU_{KD}	JB	CSK_{KD}	$C(\kappa)$
61–65	.015	.004	.001	.015	.115	.252	.184	.185	8–34
66–70	.060	.054	.076	.047	.010	.026	.018	.012	10–34
71–75	.004	.005	.001	.004	.474	.815	.631	.759	8–34
76–80	.001	.001	.001	.001	.002	.001	.001	.001	6–28
81–85	.001	.001	.001	.001	.361	.239	.379	.273	6–34
86–90	.568	.663	.687	.301	.040	.479	.064	.161	15–34
91–95	.123	.056	.079	.089	.296	.025	.049	.038	10–34
96–00	.001	.001	.001	.001	.001	.001	.001	.001	3–9
61–00	.001	.001	.001	.001	.001	.001	.001	.001	5–8

B	Intercept				Market			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	LR	p_∞	p_N	p_T	LR	p_∞	p_N	p_T
61–65	40.267	.027	.244	.257	445.529	.000	.001	.001
66–70	31.036	.188	.610	.622	401.248	.000	.001	.001
71–75	34.102	.106	.487	.503	431.621	.000	.001	.001
76–80	44.113	.010	.134	.142	352.947	.000	.001	.001
81–85	77.425	.000	.001	.001	389.349	.000	.001	.001
86–90	70.434	.000	.001	.003	441.244	.000	.001	.001
91–95	71.682	.000	.002	.003	374.018	.000	.001	.001
96–00	50.531	.002	.055	.057	323.194	.000	.001	.001
61–00	76.367	.000	.001	.001	2251.87	.000	.001	.001

Notes — Factor included: MARKET. In Panel A, numbers in columns (1)–(8) represent p-values for multinormality tests: the underlying statistics are defined in Eqs. (B.1)–(B.7). In column (9) we report the values of κ that are not rejected by the joint test GF test associated with Eq. (B.8) under multivariate Student-**t** errors; see Appendix B. In Panel B, Columns (1) and (5) presents the quasi-LR statistic defined in Eq. (A.3) associated with the hypothesis under consideration; columns {(2) and (6)}, {(3) and (7)} and {(4) and (8)} are the associated p-values using, respectively, the asymptotic $\chi^2(n)$ distribution, the pivotal statistics based MC test method imposing multivariate normal regression errors, and an MMC confidence set based method imposing multivariate $t(\kappa)$ errors which yields the largest MC p-value for all κ within the specified confidence set. The latter corresponds the values reported in column (9) of Panel A. See Apendix A for a description of the MMC test.

Secondly, tests on the intercepts reflect each model ability to respect (allowing for potential non-normality) the mean variance efficiency property of a well-specified asset pricing model. Thirdly, tests on factor loadings assess the weight (again, controlling for potential non-normality) of the different factors within an asset pricing framework. Taken collectively, our results identify what (and how many) factors help in pricing the anomalies described in the introduction, which are present in asset return models. All tests we conduct are motivated by finite sample arguments and are multivariate, that is, pertain jointly to the full set of the 25 portfolios controlling for their correlation structure.

In each table, we present results for value-weighted portfolios.[13] In interpreting all tables, recall that our tests for MC p-values under the Student distribution are joint tests for nuisance parameters consistent with the data and the mean-variance efficiency hypothesis. We have attributed a level of 2.5% to the construction of the confidence set [reported in the column with heading $C(\kappa)$ in each table]. So to establish a fair comparison with the MC p-values under the normality assumption or the asymptotic p-values, we must refer the p-values for the efficiency tests under the Student distribution to 2.5%.[14] All MC tests where applied with 999 replications.

In Panel A of each table, we show results from the different multivariate normality tests. These tests allow us to evaluate whether observed residuals exhibit non-Gaussian behavior through excess skewness and kurtosis: the multivariate statistics SK_M and SK_{KD} provide joint [overall portfolios] evidence on the presence of excess skewness; the statistics KU_M and KU_{KD} assess excess kurtosis, whereas CSK_M, MSK, JB and CSK_{KD} test for both excess skewness and kurtosis.

In Panel B of each table, under the heading "Intercept", we present mean-variance efficiency test statistics [in the column titled LR], the corresponding asymptotic p-values obtained from the asymptotic $\chi^2(n)$ distribution (p_∞), the exact Gaussian-based MC p-values (p_N) and the maximized MC p-values based on the Student-**t** error model (p_T). The associated confidence set for the number of degrees of freedom κ appears in the column titled $C(\kappa)$. These results allow one to compare rejection decisions across different distributional assumptions on the returns of the 25 portfolios. Results on the joint factor significance tests appear in columns with titles indicating the factor under test, where the definition of the p-values p_∞, p_N and p_T conform to the efficiency test case. The same confidence set for κ is used for all tests in each model.

[13] Note that we ran the same tests over equally-weighted portfolios. Results were not significantly affected.
[14] Unless indicated otherwise, the term "rejection" and "significance" in the following discussion will refer to a 5% level.

Table 2
Model with MARKET and MOM.

A	Normality Tests								Student
	SK_M	KU_M	CSK_M	MSK	SK_{KD}	KU_{KD}	JB	CSK_{KD}	$C(\kappa)$
61–65	.107	.035	.045	.071	.289	.227	.198	.303	10–34
66–70	.291	.256	.334	.150	.149	.054	.004	.011	11–34
71–75	.130	.146	.162	.082	.663	.770	.784	.818	11–34
76–80	.003	.002	.005	.003	.062	.091	.382	.107	7–34
81–85	.001	.001	.001	.001	.269	.266	.332	.384	6–34
86–90	.585	.669	.708	.708	.497	.322	.187	.278	21–34
91–95	.936	.553	.655	.533	.051	.095	.520	.086	17–34
96–00	.001	.001	.001	.001	.008	.005	.003	.010	5–34
61–00	.001	.001	.001	.001	.001	.001	.001	.001	5–9

B	Intercept				Market	MOM			
	LR	p_∞	p_N	p_T	LR^*	LR	p_∞	p_N	p_T
61–65	32.75	.138	.529	.542	443.67	43.18	.013	.182	.199
66–70	30.51	.206	.647	.660	402.80	70.90	.000	.002	.002
71–75	37.97	.047	.353	.358	413.66	75.82	.000	.001	.001
76–80	37.33	.054	.370	.383	331.75	95.99	.000	.001	.001
81–85	78.73	.000	.002	.002	389.83	45.81	.007	.116	.129
86–90	66.92	.000	.007	.009	438.00	56.48	.000	.029	.031
91–95	74.70	.000	.001	.003	381.62	60.69	.000	.013	.018
96–00	49.38	.003	.086	.090	320.49	93.45	.000	.001	.001
61–00	84.97	.000	.011	.001	2260.9	136.7	.000	.001	.001

Notes — Factors included are: MARKET, and MOM. See notes to Table 1. The * symbol implies that $p_\infty = 0$ and $p_N = p_T = .001$ for all subsamples and over the whole sample.

As usual in this literature, we estimate and test the model over intervals of 5 years, as well as over the whole sample: temporal instability of regression coefficients is well known in finance. It led researchers to choose fairly short time periods for tests of asset pricing. The justification for this approach is that, although the coefficients may vary, their short run variation is negligible to the extent that the tested pricing relation is unlikely to be rejected due to the temporal variation of the coefficients. Note that Lewellen and Nagel (2006) recognize that fact when they design a new test of the conditional CAPM using direct estimates of conditional alphas and betas from short window regressions, which allow them to avoid to specify conditioning information.[15] Temporal instabilities are clearly illustrated in all tables below. Indeed, our tests conducted over the full sample reject efficiency and spanning at conventional levels, and all factors [except the term structure variables in the six factor model] are jointly relevant. Normality is soundly rejected, the Student-**t** distribution fits well, the confidence set for κ is tight and reflects heavy tails. However, diagnostic tests signal departures from the i.i.d. hypotheses, which calls for caution in interpreting such evidence.[16] Furthermore, on examining evidence across the subperiods [discussed in details below], we see that test results vary importantly over time, a point worth emphasizing.

In Table 1, for most subperiods, normality is rejected with the single beta market model. These results are interesting since, although it is well accepted in the finance literature that continuously compounded returns are skewed and leptokurtic, empirical evidence (see e.g. Affleck-Graves and McDonalds 1989; Richardson and Smith, 1993 or Fiorentini et al., 2003) of non-normality is weaker for monthly data. Our results for the market model residuals indicate stronger evidence against normality, and confirm the analysis in Beaulieu et al. (2007) that is based on industry portfolios.

In general, our results show that asymptotic *p*-values are often spuriously significant. For instance, in Panel B of Table 1, at the 5% level of confidence, we find six rejections [efficiency is rejected in all subperiods except from 1966–75] of the efficiency hypothesis using the asymptotic $\chi^2(25)$ test, and only three [in the 1981–95 subperiods] using the MC *p*-values under normality and under the Student-**t** distribution. In these cases, even if the market factor is strongly significant, the mean-variance efficiency hypothesis is equally strongly rejected: indeed, p_∞, p_N and p_T for tests on the intercept and on the market factors, are all very small in the 1981–95 subperiods.

The next tables show the multivariate normality evidence for asset pricing models that include more than one factor. In particular, results from Tables 1–4 reveal that the number of multivariate rejections decreases importantly when momentum is added to the analysis. For example, whereas for the market model, at least one of the normality test statistic has a *p*-value smaller than 5% in all subperiods, when momentum is added, all *p*-values exceed 5% in 3 subperiods [1971–75, 86–90, 91–95]. The same observation holds when SMB and HML are added (with and without momentum) to the market model.

[15] We ran the analysis with October 1987 and January returns over 5 and 10 year subperiods as well as without those observations over 10 year subperiods. Our results are not significantly affected by such modifications.
[16] (Beaulieu et al., 2008a) report that GARCH effects may be less consequential on our methodology than what may be expected, for hypotheses on regression coefficients.

Table 3
Model with MARKET, SMB, HML.

A	Normality Tests								Student
	SK_M	KU_M	CSK_M	MSK	SK_{KD}	KU_{KD}	JB	CSK_{KD}	$C(\kappa)$
61–65	.009	.002	.001	.009	.309	.536	.424	.417	8–34
66–70	.438	.256	.321	.247	.001	.006	.001	.001	13–34
71–75	.183	.220	.288	.114	.796	.794	.916	.878	10–34
76–80	.544	.206	.245	298	.027	.231	.059	.087	13–34
81–85	.004	.005	.005	.004	.388	.324	.483	.324	8–34
86–90	.450	.772	.573	.804	.383	.374	.543	.438	23–34
91–95	.937	.929	.992	.559	.658	.369	.526	.537	16–34
96–00	.001	.001	.001	.001	.014	.001	.001	.001	4–15
All	.001	.001	.001	.001	.001	.001	.001	.001	6–10

B	Intercept				Market	SMB	HML
	LR	p_∞	p_N	p_T	LR^*	LR^*	LR^*
61–65	33.303	.124	.577	.586	441.194	292.065	232.189
66–70	27.611	.326	.793	.809	387.158	322.297	237.899
71–75	36.000	.072	.456	.469	411.597	355.454	311.138
76–80	35.637	.077	.472	.484	339.791	293.914	256.774
81–85	55.410	.000	.043	.046	353.048	315.527	232.669
86–90	65.775	.000	.007	.009	426.706	284.057	258.001
91–95	73.834	.000	.005	.004	368.244	339.164	289.868
96–00	49.725	.002	.092	.093	294.907	351.904	278.362
61–00	54.932	.001	.001	.001	2161.42	2294.67	1759.02

Notes — Factors included are: MARKET, and SMB and HML. See notes to Table 1. The * symbol implies that $p_\infty = 0$ and $p_N = p_T = .001$ for all subsamples and over the whole sample.

From Table 2, we find that inference on the significance of the intercept is comparable with what was found in a one-factor model. While momentum coefficients are statistically significant under asymptotic p-values, their significance is reduced by over 40% with MC p-values. This leads us to think that the role of momentum in the asset pricing literature might be over-rated.

Table 3 shows the analysis for an asset pricing model that include market, SMB and HML. In that case multivariate normality is rejected in only three sub-periods over eight while the multivariate evidence on the significance of the intercept is strongly reduced compared with the two previous models. Consistent with the analysis of Fama and French (1993) even in finite samples, the coefficients on all three factors appear to be strongly significant. We next include momentum to the previous three-factor

Table 4
Model with MARKET, SMB, HML and MOM.

A	Normality Tests								Student
	SK_M	KU_M	CSK_M	MSK	SK_{KD}	KU_{KD}	JB	CSK_{KD}	$C(\kappa)$
61–65	.088	.019	.026	.052	.807	.816	.708	.874	7–34
66–70	.919	.848	.958	.527	.137	.031	.001	.001	13–34
71–75	.467	.500	.610	.241	.743	.794	.735	.881	11–34
76–80	.211	.422	.283	.909	.600	.278	.086	.134	26–34
81–85	.002	.004	.001	.002	.557	.399	.249	.344	6–34
86–90	.230	.457	.287	.896	.449	.452	.436	.577	31–34
91–95	.270	.504	.322	.879	.308	.443	.678	.472	26–34
96–00	.001	.001	.001	.001	.020	.019	.070	.037	5–34
61–00	.001	.001	.001	.001	.001	.001	.001	.001	6–10

B	Intercept				Market	SMB	HML	MOM			
	LR	p_∞	p_N	p_T	LR^*	LR^*	LR^*	LR	p_∞	p_N	p_T
61–65	23.50	.548	.906	.912	440.65	294.10	233.75	44.71	.009	.199	.217
66–70	26.74	.369	.825	.843	389.10	322.17	237.29	70.44	.000	.004	.007
71–75	36.56	.064	.486	.489	403.20	329.55	310.96	49.01	.003	.109	.118
76–80	30.64	.201	.716	.718	331.46	292.47	251.50	94.64	.000	.001	.001
81–85	55.04	.000	.056	.067	350.50	315.92	231.26	44.25	.010	.202	.211
86–90	63.38	.000	.018	.018	426.87	282.17	253.09	50.58	.002	.088	.094
91–95	75.03	.000	.003	.004	376.26	338.73	289.86	60.25	.000	.019	.024
96–00	41.74	.019	.284	.288	294.16	349.32	270.47	68.46	.000	.007	.011
61–00	45.23	.008	.015	.020	2167.6	2294.7	1751.0	90.80	.000	.001	.001

Notes — Factors included are: MARKET, SMB, HML and MOM. See notes to Table 1. The * symbol implies that $p_\infty = 0$ and $p_N = p_T = .001$ for all subsamples and over the whole sample.

Table 5
Model with TERM and DEF.

A	Normality Tests								Student
	SK_M	KU_M	CSK_M	MSK	SK_{KD}	KU_{KD}	JB	CSK_{KD}	$C(\kappa)$
61–65	.102	.092	.106	.077	.086	.543	.137	.284	9–34
66–70	.011	.010	.014	.011	.054	.039	.063	.032	8–34
71–75	.068	.041	.066	.054	.935	.294	.436	.746	10–34
76–80	.001	.001	.001	.001	.001	.001	.001	.001	5–30
81–85	.011	.002	.001	.010	.331	.679	.470	.557	8–34
86–90	.062	.077	.102	.048	.001	.001	.001	.001	10–34
91–95	.268	.077	.098	.159	.539	.145	.220	.275	10–34
96–00	.001	.001	.001	.001	.001	.001	.001	.001	4–10
61–00	.001	.001	.001	.001	.001	.001	.001	.001	5–9

B	Intercept				TERM				DEF			
	LR	p_∞	p_N	p_T	LR	p_∞	p_N	p_T	LR	p_∞	p_N	p_T
61–65	43.88	.011	.166	.169	59.28	.000	.017	.024	38.47	.042	.335	.357
66–70	28.80	.273	.723	.736	40.74	.024	.264	.263	60.03	.000	.018	.024
71–75	33.89	.110	.528	.532	39.50	.033	.273	.280	55.34	.000	.038	.041
76–80	46.48	.006	.118	.119	36.21	.068	.400	.418	42.85	.014	.177	.215
81–85	77.00	.000	.001	.001	34.63	.095	.474	.496	33.83	.112	.528	.536
86–90	72.82	.000	.001	.002	22.51	.606	.905	.903	20.97	.694	.921	.931
91–95	74.68	.000	.003	.004	44.86	.009	.139	.152	44.66	.009	.155	.161
96–00	55.99	.000	.031	.030	40.34	.027	.274	.294	54.04	.000	.032	.048
61–00	82.79	.000	.001	.001	45.69	.007	.001	.001	66.97	.000	.001	.001

Notes — Factors included are: TERM and DEF. See notes to Table 1.

Table 6
Model with MARKET, TERM, DEF and MOM

A	Normality Tests								Student
	SK_M	KU_M	CSK_M	MSK	SK_{KD}	KU_{KD}	JB	CSK_{KD}	$C(\kappa)$
61–65	.931	.570	.697	.468	.211	.199	.223	.331	13–34
66–70	.183	.132	.156	.101	.100	.051	.033	.063	9–34
71–75	.359	.330	.417	.192	.691	.742	.682	.834	11–34
76–80	.068	.072	.092	.047	.054	.046	.104	.097	8–34
81–85	.003	.002	.002	.003	.556	.509	.424	.579	6–34
86–90	.427	.385	.494	.795	.767	.604	.388	.512	22–34
91–95	.684	.558	.652	.665	.152	.279	.712	.239	17–34
96–00	.001	.001	.001	.001	.023	.012	.004	.010	5–29
61–00	.001	.001	.001	.001	.001	.001	.001	.001	5–9

B	Intercept				Market	MOM			
	LR	p_∞	p_N	p_T	LR^*	LR	p_∞	p_N	p_T
61–65	33.19	.126	.579	.587	445.00	44.13	.010	.205	.225
66–70	36.51	.064	.460	.467	394.68	81.85	.000	.001	.002
71–75	46.24	.006	.174	.179	397.99	78.68	.000	.001	.003
76–80	39.03	.037	.389	.397	318.99	89.53	.000	.001	.001
81–85	83.93	.000	.002	.002	365.52	48.68	.003	.107	.122
86–90	68.07	.000	.010	.011	439.36	57.75	.000	.037	.042
91–95	73.57	.000	.003	.004	369.43	54.93	.000	.048	.060
96–00	58.00	.000	.031	.035	312.60	107.2	.000	.001	.001
61–00	85.08	.000	.001	.001	2199.9	134.2	.000	.001	.001

B	TERM				DEF			
	LR	p_∞	p_N	p_T	LR	p_∞	p_N	p_T
61–65	56.577	.000	.038	.052	40.510	.026	.324	.335
66–70	46.255	.006	.173	.181	57.960	.000	.033	.040
71–75	41.263	.021	.315	.322	48.259	.003	.137	.147
76–80	38.820	.038	.379	.397	45.184	.008	.188	.209
81–85	41.948	.018	.284	.295	34.828	.091	.495	.512
86–90	23.104	.571	.908	.917	21.594	.659	.942	.947
91–95	38.658	.040	.368	.376	35.732	.076	.489	.500
96–00	39.267	.035	.338	.373	55.944	.000	.047	.058
61–00	47.849	.004	.005	.008	47.696	.004	.003	.007

Notes — Factors included are: MARKET, TERM, DEF and MOM. See notes to Table 1.

Table 7
Five factor model.

A	Normality Tests								Student
	SK_M	KU_M	CSK_M	MSK	SK_{KD}	KU_{KD}	JB	CSK_{KD}	$C(\kappa)$
61–65	.283	.188	.221	.166	.575	.311	.449	.471	9–34
66–70	.221	.121	.169	.134	.001	.014	.001	.004	10–34
71–75	.469	.319	.459	.249	.783	.746	.896	.840	10–34
76–80	.348	.724	.407	.854	.005	.028	.006	.010	18–34
81–85	.010	.011	.016	.009	.747	.169	.275	.329	8–34
86–90	.359	.645	.427	.851	.531	.294	.412	.447	20–34
91–95	.406	.324	.412	.236	.725	.247	.382	.423	10–34
96–00	.001	.001	.001	.001	.032	.006	.005	.005	5–30
61–00	.001	.001	.001	.001	.001	.001	.001	.001	6–10

B	Intercept				Market	SMB	HML
	LR	p_∞	p_N	p_T	LR^*	LR^*	LR^*
61–65	34.411	.099	.598	.611	444.388	292.878	231.659
66–70	27.881	.313	.827	.838	375.134	320.849	227.236
71–75	45.538	.007	.203	.214	386.482	353.060	311.848
76–80	37.164	.056	.490	.497	334.963	294.361	250.293
81–85	57.304	.000	.048	.049	334.761	322.791	232.503
86–90	69.252	.000	.007	.008	427.308	281.928	258.728
91–95	80.742	.000	.004	.003	350.356	328.026	286.169
96–00	52.854	.001	.093	.092	284.229	354.063	276.945
61–00	55.721	.000	.001	.001	2071.50	2277.46	1774.74

B	TERM				DEF			
	LR	p_∞	p_N	p_T	LR	p_∞	p_N	p_T
61–65	54.8895	.005	.049	.064	41.1912	.219	.330	.344
66–70	50.0995	.021	.114	.126	41.8486	.187	.301	.310
71–75	43.7912	.114	.234	.244	42.7294	.150	.271	.293
76–80	41.1074	.224	.332	.354	45.6543	.070	.199	.220
81–85	38.4013	.423	.431	.450	30.0095	.224	.756	.761
86–90	26.6033	.376	.845	.843	23.1021	.571	.908	.916
91–95	34.2781	.102	.569	.585	36.4776	.064	.470	.492
96–00	37.9797	.465	.438	.464	46.2263	.006	.195	.216
61–00	45.6903	.070	.006	.014	66.9671	.001	.001	.001

Notes — Factors included are: MARKET, SMB, HML, TERM, and DEF. See notes to Table 1. The * symbol implies that $p_\infty = 0$ and $p_N = p_T = .001$ for all subsamples and over the whole sample.

model; results are presented in Table 4. With respect to multivariate normality, the inference is comparable to that of the three factor model. In that case, momentum does not add very much to the analysis. The same is true of the inference on the intercept. Finally, results on the significance of all factors is positive except for momentum in which case the coefficient appears significant in all sub-periods under asymptotic p-values while the significance is reduced to only four sub-periods out of nine under finite sample correct inference.

The next models include TERM and DEF given that Fama and French (1993) show that these variables are important in the time series variations of stock returns; results are shown in Tables 5–8. On comparing the results from Tables 1 and 5, we see that in a two-factor model, TERM and DEF on their own have better abilities to retain the multivariate normality hypothesis of residuals than market alone. The rejection rate of the intercept with those two factors are similar to those of the market model. Although on an asymptotic basis those factors might appear important, the finite sample inference shows that these results are spurious in most sub-periods. This can be attributed to the fact that TERM and DEF are slowly evolving variables over time. They appear significant under asymptotic inference given that the asymptotic critical point is unreliable and overrejects too easily. Results [not tabulated for space considerations] of a three factor model with market, TERM and DEF, confirm these findings. In particular, based on P_N and P_T, tests on the intercept, and on each factor in turn [including market] are not significant at the 5% level except in the three subperiods from 1981–1995. The fact that even the market factor seems irrelevant in 5 out of 8 sub-periods is worth noting; this result was however not observed in all other models. We suspect [in view of the results discussed below on models with more factors] that statistical redundancy of the term structure variables may lead to identification problems that may have escaped notice to date, because of reliance on spurious χ^2 based tests.

The next model discussed in Table 6 adds market and momentum to TERM and DEF. With respect to multivariate normality, the null hypothesis is rejected in only three subperiods. Yet the inference on the intercept is almost unchanged compared to the previous model or the market model. In terms of factor relevance, market remains strongly significant, while momentum appeared to be significant in all subperiods asymptotically, its significance is reduced by half [that is, momentum is significant at 5% in only four subperiods] based on the Student-**t** MC p-values. The role of TERM and DEF is greatly reduced since the coefficients appeared significant under asymptotic p-values, none are significant at usual levels given the MC Student-**t** p-values.

Table 8
Model with all factors.

A	Normality Tests								Student
	SK_M	KU_M	CSK_M	MSK	SK_{KD}	KU_{KD}	JB	CSK_{KD}	$C(\kappa)$
61–65	.366	.601	.847	.522	.366	.601	.847	.522	10–34
66–70	.093	.019	.001	.001	.093	.019	.001	.001	11–34
71–75	.774	.851	.809	.909	.774	.851	.809	.909	11–34
76–80	.428	.204	.058	.098	.428	.204	.058	.098	–
81–85	.548	.489	.399	.558	.548	.489	.399	.558	6–34
86–90	.652	.639	.551	.699	.652	.639	.551	.699	30–34
91–95	.196	.315	.651	.299	.196	.315	.651	.299	13–34
96–00	.025	.018	.050	.038	.025	.018	.050	.038	6–34
61–00	.001	.001	.001	.001	.001	.001	.001	.001	6–11

B	TERM				DEF			
	LR	p_∞	p_N	p_T	LR	p_∞	p_N	p_T
61–65	54.815	.005	.073	.092	41.363	.210	.354	.368
66–70	49.173	.027	.180	.194	47.137	.047	.207	.214
71–75	44.208	.103	.284	.292	43.679	.118	.296	.314
76–80	38.508	.412	.456	–	43.381	.127	.307	–
81–85	44.822	.088	.274	.279	32.730	.138	.638	.647
86–90	28.095	.303	.824	.829	25.062	.458	.894	.895
91–95	29.911	.227	.772	.779	33.364	.122	.634	.644
96–00	45.273	.078	.231	.258	55.910	.004	.065	.075
61–00	38.698	.040	.052	.057	37.968	.046	.064	.077

B	Intercept				Market	SMB	HML	MOM			
	LR	p_∞	p_N	p_T	$LR^\dagger$	$LR^\dagger$	$LR^\dagger$	LR	p_∞	p_N	p_T
61–65	24.09	.514	.916	.918	442.82	294.42	232.52	44.93	.009	.244	.262
66–70	31.99	.158	.688	.713	374.81	320.98	228.81	86.35	.000	.001	.002
71–75	46.99	.005	.204	.210	382.81	326.30	311.70	50.04	.002	.136	.148
76–80	33.43	.120	.681	–	317.82	293.35	250.02	92.21	.000	.001	–
81–85	60.36	.000	.043	.048	322.77	322.64	228.38	43.09	.014	.290	.313
86–90	66.11	.000	.019	.021	425.76	278.53	254.76	51.31	.002	.117	.129
91–95	79.15	.000	.001	.004	355.95	327.82	286.04	54.42	.000	.082	.094
96–00	46.58	.006	.213	.226	278.78	353.61	261.12	72.72	.000	.006	.009
61–00	45.69	.007	.014	.017	2081.2	2277.3	1735.2	92.30	.000	.001	.001

Notes — Factors included are: MARKET, SMB, HML, TERM, DEF and MOM. The – symbol indicates an empty confidence set for κ i.e. the Student-**t** family is rejected, in which case p_T is not computed. The † symbol implies that $p_\infty = 0$ and $p_N = .001$ for all subsamples and over the whole sample, and $p_T = .001$ for all samples except in the 1976–80 subperiod, in which case the Student-**t** distribution is rejected. See notes to Table 1.

In Table 7 we present results for a model including market, SMB, HML, TERM and DEF as factors. Results are comparable to those on Table 6 with respect to multivariate normality while the model fares marginally better with rejection rates of the intercept. The role of market, SMB and HML remains important while the coefficients on TERM and DEF do not appear significant under finite-sample p-values even under Gaussianity; this holds for all subperiods and over the full sample. Despite our cautionary interpretation of full sample results, we can safely conclude that irrelevance of TERM and DEF emerges as a structurally stable conclusion in the five factor model.

The last model (refer to Table 8) includes all factors. In that case the hypothesis of multivariate normality cannot be rejected except in two subperiods [1966–70, and 1996–00]. Note that the Student-**t** distribution is rejected in the 1976–80 subperiod. To assess whether this result may be linked to outliers, we replicate the analysis excluding January; the confidence set for κ is not empty in this case: indeed $C(\kappa) = \{19 - 27\}$.[17] Furthermore, for all test rejections based on normal MC p-values are also confirmed by the Student-**t** p-value. Multivariate joint tests on the intercept show rejections in three subperiods given normality, and only two sub-periods under Student-**t** MC p-values while we see five rejections under the asymptotic p-values. Market, SMB and HML are again strongly significant while momentum is significant in three sub-periods out of eight under MC p-values. Again TERM and DEF appear to be irrelevantly present in the model.

Chen et al. (1986) use TERM and a variable like DEF to help explain the cross-section of average returns on NYSE stocks. They use the Fama and MacBeth (1973) cross-section regression approach; the cross-section of average stock returns is explained with the cross-section of slopes from time series regressions of returns on TERM, a default factor, and other factors. In their tests, the default factor is the most powerful factor in average stock returns, and TERM sometimes has power. Fama and French (1993)

[17] Except for this case, results excluding the January months and October 1987 are qualitatively similar to the full data analysis in each subperiod and for the full sample.

Table 9
Multivariate diagnostics.

	E		LK		VR	
	Normal	Student	Normal	Student	Normal	Student
	(1)	(2)	(3)	(4)	(5)	(6)
61–65	.861	.864	.984	.990	.991	.997
66–70	.893	.909	.382	.399	.191	.199
71–75	.704	.727	.011	.016	.206	.212
76–80	.024	.205	.627	.626	.294	.293
81–85	.509	.606	.717	.701	.319	.333
86–90	.105	.309	.664	.648	.140	.159
91–95	.219	.440	.373	.380	.773	.777
96–00	.006	.106	.018	.023	.497	.516
61–00	.001	.045	.001	.001	.002	.005

Notes — Factors included are: MARKET, SMB, HML, TERM, DEF and MOM. Numbers shown are p-values associated with the combined tests E, LK, and VR, defined by (C.1). E and LK are multivariate versions of Engle's and Lee and King's GARCH tests and VR is a multivariate version of Lo and MacKinlay's variance ratio tests; see Appendix C. In columns (1)–(3)–(5), the p-values are MC pivotal statistics based; p-values in columns (2)–(4)–(6) are MMC confidence set based. The relevant 2.5% confidence set for the nuisance parameters is reported in the column titled $C(\kappa)$ in Panel A of Table 8.

confirm that the tracks of TERM and DEF show up clearly in the time series variation of stock returns. In contrast to the cross-section regressions of Chen, Roll and Ross, however, their time series regressions say that the average premium on DEF and TERM risks are too small to explain much variation in the cross-section of average stock returns. Shanken and Weinstein (2006) make a similar point. Our evidence confirms that DEF and TERM explain less of the time series variation in stock returns than suggested in cross section analyses given previous available tests on the significance of those factors.

Finally, Table 9 presents the results of our multivariate exact diagnostic checks for departures from the i.i.d. assumption, namely our proposed multivariate versions of the Engle, Lee-King and variance ratio tests; we use 12 month lags. The results show very few rejections of the null hypothesis both at the 1% and 5% level of significance over five-year sub-periods while in all but one case, the multivariate diagnostics reject the null hypotheses over the whole sample. This implies that, in our statistical framework and for the shorter time spans analyzed, i.i.d. errors provide an acceptable working assumption while it is not the case for our forty year sample.

Table 10
Spanning tests.

Factors	Market				Market, MOM				Market SMB HML
	LR	p_∞	p_N	p_T	LR	p_∞	p_N	p_T	LR*
61–65	109.474	.000	.004	.005	81.857	.003	.168	.168	335.911
66–70	96.862	.000	.027	.026	99.368	.000	.022	.022	332.017
71–75	108.686	.000	.007	.006	95.548	.000	.046	.042	404.828
76–80	140.874	.000	.001	.001	178.162	.000	.001	.001	364.380
81–85	143.142	.000	.001	.001	123.245	.000	.001	.001	327.869
86–90	149.951	.000	.001	.001	127.295	.000	.002	.002	328.523
91–95	150.340	.000	.001	.001	143.090	.000	.001	.001	388.300
96–00	120.179	.000	.001	.001	141.544	.000	.001	.001	324.563
61–00	309.834	.000	.001	.001	272.374	.000	.001	.001	2121.561

	Market, TERM, DEF				Market, SMB, HML,MOM	5 factors	All factors
	LR	p_∞	p_N	p_T	*LR**	*LR**	*LR**
61–65	84.932	.001	.157	.169	310.588	237.723	223.481
66–70	82.113	.003	.205	.214	315.055	320.557	304.345
71–75	72.128	.022	.435	.449	337.001	384.977	337.631
76–80	116.558	.000	.004	.011	346.831	360.763	335.510
81–85	138.640	.000	.001	.001	318.858	340.799	335.932
86–90	116.326	.000	.001	.005	300.253	311.373	277.831
91–95	114.313	.000	.003	.006	382.178	376.665	370.050
96–00	87.026	.001	.122	.133	308.165	311.470	303.726
61–00	145.560	.000	.001	.001	1969.007	2011.970	1889.005

Notes — The * symbol implies that $p_\infty = 0$ and $p_N = p_T = .001$ for all subsamples and over the whole sample. LR refer to statistic used to assess the spanning hypothesis (2.2); p_∞, p_N and p_T are the associated p-values using, respectively, the asymptotic $\chi^2(2n)$ distribution, the pivotal statistics based MC test method imposing multivariate normal regression errors, and an MMC confidence set based method imposing multivariate $t(\kappa)$ errors which yields the largest MC p-value for all κ within the specified confidence set. The latter (for each model) are reported in the columns titled $C(\kappa)$ in Tables 1–8. See Appendix A for a description of the MMC test.

3.3. Spanning tests

Table 10 present spanning test results for asymptotic and MC p-values under normal and Student-**t** distributions. Results show that mean-variance spanning is preserved in almost no case. That is in a one-factor model, the mean-variance spanning hypothesis is strongly rejected independently of the p-value we consider. The same is true of all multifactor models except for the one that includes DEF and TERM in addition to the market factor where the mean-variance spanning hypothesis cannot be rejected in four sub-periods out of eight. What is interesting to note in that case is that decisions based on asymptotic p-values lead to rejection of the null in all sub-periods. The non-rejections only occur with finite sample correct p-values. Whether we assume normality or Student-**t** distributions does not affect the decision.

We note however that for the latter model, and for the same subperiods where the spanning hypothesis could not be rejected at usual levels, the model intercept and each factor including the market proxy were also insignificant at usual levels [when tested in turn for joint significance overall portfolios]. On observing that the market proxy was strongly significant [i.e. our joint test had very small p-values] for all other models considered for these same sub-periods, one must refrain from a casual conclusion in favour of spanning with this model. Indeed, our results for this sub-model when interpreted collectively, signal the presence of identification problems. These may be attributed to the fact that TERM and DEF evolve quite slowly over time, so a clear distinction, between their effect and that of the constant regressor is rendered difficult in relatively small samples. Observe that reliance on asymptotics masks this problem, since spuriously small χ^2-based p-values contribute to blurring this evidence.

To conclude, a brief discussion of the above results may prove useful beyond the specific model analyzed. The statistics associated with MLR coefficients and the residual diagnostics we performed do behave differently, in so far as dependence on κ is concerned. For the efficiency, spanning and factor relevance tests, the above reported results use an $\alpha_1 = 2.5\%$ level for $C(\kappa)$. Though we analyze results for an overall 5% level [for conventional rather than methodological reasons], the tables above report the maximal p-values in question for every test, leaving the reader with more complete information. Note that changing α_1 has practically little impact on the maximal p-values particularly when the p-values are small. In the diagnostic tests case, p-values are more sensitive to κ; we thus consider the alternative MMC procedure proposed in Dufour et al. (2010), which consists in maximizing the diagnostic test p-value for any κ [in a one-stage maximized Monte Carlo (MMC) procedure, not restricted to a confidence set], and referring the maximal p-value so obtained to a 5% [rather than a 2.5%] cut-off.[18] For the tests reported in Table 9, applying the latter MMC test does not alter our main conclusion. It is also worth noting that the simulation study in Dufour et al. (2010) reveals that all the Monte Carlo tests we apply here perform well with samples of 60 observations on as many as 40 equations. Nevertheless, our diagnostics presume that regression coefficients are stable over the test period; our parametric approach also assumes a constant degrees-of-freedom parameter for the hypothesized Student-**t** distribution. These assumptions may be disputed over the long run which calls for caution in interpreting our full-sample test outcomes, relative to the sub-period analysis. On balance, our results suggest that: (i) a multivariate regression with Fama-French-Carhart factors and Student-**t** errors seems acceptable as a statistical working framework within subperiods, yet (ii) the pricing relation and financial implications arising from this model are temporally unstable over long time spans.

4. Conclusion

This paper conducts within a MLR non-normal framework a comprehensive assessment of the multifactor model first proposed by Fama and French (1993). We study mean-variance efficiency and spanning, as well as factor relevance in accounting for stylized anomalies. Our tests are finite sample motivated and are invariant to portfolio repackaging. Our results, using monthly portfolio returns from Fama and French's data from 1961–2000 reveal the following. (1) The book to market equity and size factors and/or the momentum factor have a “normalizing” effect on model residuals. (2) Allowing for heavy tailed distributions may empirically accommodate some stylized asset pricing anomalies. (3) The term structure variables seem statistically redundant; the momentum factor also suffers from weak relevance problems; aside from a “normalizing” role on model residuals, momentum appears statistically insignificant at usual levels in many sub-periods. (4) Mean-variance efficiency is rejected in fewer subperiods allowing for non-normal errors in multi-factor settings; the book to market equity and size factors contribute importantly in reinforcing efficiency. Nevertheless, the zero-intercept hypothesis does not pass our tests uniformly even in the six factor model. (5) The addition of factors fails to prop up the mean-variance spanning hypothesis.

Our results on factor relevance call for caution in interpreting general spanning and efficiency test non-rejections. Indeed, despite the simplicity and the linear structure of the underlying test framework, adding non-relevant factors can blur evidence, leading to identification difficulties as would occur with e.g. multicollinearity. This observation lines up with conventional wisdom; our results however illustrate that reliance on usual χ^2 tests may have delayed awareness on problems stemming from the term structure and possibly the momentum factor. Such problems may be further avoided if in addition to testing for statistical significance, one constructs confidence sets for key model parameters. Unfortunately, most available procedures for confidence set inference in MLR rely on the same asymptotics underlying the χ^2 tests that were shown to be highly unreliable. In Beaulieu et al. (2008a) we provide new analytical tools that may be used for this purpose in linear and some non-linear multifactor models. We also expect that more pronounced problems may result from redundant factors in non-linear models that call for instrumental variables methods. Pursuing work in this direction is a worthy research objective.

[18] Dufour et al. (2010) show that the one step MMC test in question does not uniformly dominate the two-step procedure we apply here.

Acknowledgments

The authors thank Stéphane Chrétien, Christian Gouriéroux, Raymond Kan, Blake LeBaron, Guofu Zhou, three anonymous referees and the Editor, Fanz Palm, for several useful comments. This work was supported by the William Dow Chair in Political Economy (McGill University), the Chaire RBC en innovations financières (Université Laval), the Canada Research Chair Program, the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de finance mathématique de Montré al (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], 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 and NATECH (Government of Québec).

Appendix A. Finite sample tests on regression coefficients

To simplify presentation, let us rewrite Eq. (2.1) as follows

$$Y = XB + U \tag{A.1}$$

where $Y = [r_1,...,r_n]$ is the $T \times n$ matrix of portfolio returns $r_i = (r_{i1},...,r_{iT})'$, $i = 1,...,n$, X is the $T \times k$ matrix of factors $X = [l_T, \tilde{r}_1,...,\tilde{r}_s]$, $r_j = (\tilde{r}_{j1},...,\tilde{r}_{jT})'$, $j = 1,...,s$ where $k = s+1$, l_T is a T-dimensional vector of ones, and $U = [U_1,...,U_T]'$ is the $T \times n$ matrix which includes u_{it}. In this context, Eq. (2.6) may be rewritten as $U = WJ'$, where $W = [W_1,...,W_T]'$, and Eqs. (2.2)–(2.4) can all be reformulated as

$$HB = C \tag{A.2}$$

where: (i) for Eq. (2.2), H and C are (respectively) the $2 \times k$ and $2 \times n$ matrices

$$H = \begin{bmatrix} 1\,0 \ldots 0 \\ 0\,1 \ldots 1 \end{bmatrix}, C = \begin{bmatrix} 0\,0 \ldots 0 \\ 1\,1 \ldots 1 \end{bmatrix};$$

(ii) for Eq. (2.3), C is $1 \times n$ vector of zeros and H is the $1 \times k$ vector (1,0,...,0); and (iii) for Eq. (2.4), C is an $1 \times n$ vector of zeros and H is the $1 \times k$ selection vector with one at the position $j+1$ and zeros elsewhere. The constrained and unconstrained estimates of Σ and B are $\hat{\Sigma}(H,C) = (\hat{U}(H,C))'\hat{U}(H,C)/T$, $\hat{U}(H,C) = Y - X\hat{B}(H,C)$, $\hat{B}(H,C) = \hat{B} + (X'X)^{-1}H'[H(X'X)^{-1}H']^{-1}[C - H\hat{B}]$, $\hat{\sum} = \hat{U}'\hat{U}/T$, $\hat{U} = Y - X\hat{B}$, $\hat{B} = (X'X)^{-1}X'Y$. Furthermore (see Dufour and Khalaf, 2002), the LR statistic

$$LR(H,C) = Tln(\Lambda(H,C)), \;\; \Lambda(H,C) = |\hat{\Sigma}(H,C)| / |\hat{\Sigma}| \tag{A.3}$$

is distributed, under Eqs. (2.1), (A.2) and (2.6) like

$$\overline{LR}(H) \equiv T\ln(|W'M_HW| / |W'MW|), \tag{A.4}$$

where $M = I - X(X'X)^{-1}X'$ and $M_H = M + X(X'X)^{-1}H'[H(X'X)^{-1}H']^{-1}H(X'X)^{-1}X'$. Under the normality hyphothesis (2.5), and $T - s - n > 0$, $(\rho\tau - 2\lambda)\,(\Lambda_{(H,C)} - 1)/hc \sim F(hn, \rho\tau - 2\lambda)$ where $h = rank(H)$, $\rho = (T - k(n-h+1))/2$, $\lambda = (hn-2)/4$, $\tau = [(h^2n^2 - 4)/(h^2 + n^2 - 5)]^{1/2}$, if $h^2 + n^2 - 5 > 0$, otherwise $\tau = 1$.This result can be used to obtain an exact Monte Carlo p-value (see Dufour, 2006), as follows. Conditional on κ, generate, imposing Eq. (2.6), N *i.i.d.* draws from the distribution of $W_1,...,W_T$; these yield N simulated values of $LR(H,C)$ applying Eq. (A.4). Let $\hat{G}_N(\kappa)$ denote the number of simulated criteria $\geq$ to the observed value of $LR(H,C)$. Then, the MC p-value is

$$\hat{p}_N(\kappa) = \left(N\hat{G}_N(\kappa) + 1\right) / (N + 1). \tag{A.5}$$

To deal with unknown κ, we apply the two stage MMC procedure from Beaulieu et al. (2007): (1) we build an exact confidence set [denoted $C(\kappa)$] for κ, of level $(1 - \alpha_1)$; and (2) we maximize the p-value $\hat{p}_N(\kappa)$ over the values of κ in $C(\kappa)$; then we refer the latter maximal p-value [denoted P_T] to α_2 where $\alpha = \alpha_1 + \alpha_2$. To obtain $C(\kappa)$, we invert a goodness-of-fit test (presented below), for the null hypothesis Eq. (2.6) with known κ.

Appendix B. Goodness-of-fit tests

Consider the skewness and kurtosis based normality test statistics (see Mardia, 1970; Kilian and Demiroglu, 2000 and Dufour et al., 2003)

$$SK_M = \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\hat{d}_{st}^3, \quad SK_{KD} = (sk_1,...,sk_n)'(sk_1,...,sk_n), \tag{B.1}$$

$$KU_{KD} = (eku_1,...,eku_n)'(eku_1,...,eku_n), \quad KU_M = \frac{1}{T}\sum_{t=1}^{T}\hat{d}_{tt}^2, \tag{B.2}$$

$$MSK = \frac{T}{6}SK_M + \frac{T[KU_M - n(n+2)]^2}{8n(n+2)}, \quad JB = \frac{T}{6}SK_{KD} + \frac{T}{24}KU_{KD}, \tag{B.3}$$

where $\hat{d}_{st}$ are the elements of $\hat{U} = \hat{U}(\hat{U}'\hat{U}/T)^{-1}\hat{U}'$, $eku_i = ku_i - 3$, $sk_i = \left(T^{-1}\sum_{t=1}^{T}\tilde{W}_{ti}^3\right) / \left(\left(T^{-1}\sum_{t=1}^{T}\tilde{W}_{ti}^2\right)^{3/2}\right)$, $ku_i = \left(T^{-1}\sum_{t=1}^{T}\tilde{W}_{ti}^4\right) / \left(\left(T^{-1}\sum_{t=1}^{T}\hat{W}_{ti}^2\right)^2\right)$, $i = 1,..., n$, $\hat{W}_{ti}$ are the elements of the standardized residuals matrix

$$\tilde{W} = \hat{U}S_{\hat{U}}^{-1}, S_{\hat{U}} = \text{the Cholesky factor of } \hat{U}'\hat{U}. \tag{B.4}$$

Under Eq. (2.5), the null distributions of these statistics do not depend on B and Σ, so MC p-values can be obtained as in Eq. (A.5).

Consider next the following modified criteria (see Dufour et al., 2003) to assess the fit of Eq. (2.6) including the normal and multivariate Student-**t** special cases Eqs. (2.5) and (2.7):

$$ESK_M(\kappa) = |SK_M - \overline{SK}_M(\kappa)|, \ EKU_M(\kappa) = |KU_M - \overline{KU}_M(\kappa)|, \tag{B.5}$$

$$ESK_{KD}(\kappa) = (esk_1(\kappa),...,esk_n(\kappa))'(esk_1(\kappa),...,esk_n(\kappa)), \tag{B.6}$$

$$EKU_{KD}(\kappa) = (eku_1(\kappa),...,eku_n(\kappa))'(eku_1(\kappa),...,eku_n(\kappa)), \tag{B.7}$$

where $esk_i(\kappa) = \left(sk_i(\kappa) - \mu_{sk_i(\kappa)}\right) / \sigma_{sk_i(\kappa)}$, $eku_i(\kappa) = (ku_i(\kappa) - \mu_{ku_i(\kappa)}) / \sigma_{ku_i(\kappa)}$, $i=,...n$, $\overline{SK}_M(\kappa)$ and $\overline{KU}_M(\kappa)$ are simulation-based estimates of the mean of SK_M and KU_M given Eq. (2.6), $\mu_{sk_i(\kappa)}$ and $\sigma_{sk_i(\kappa)}$ are simulation-based estimates of the mean and standard deviation of $sk_i(\kappa)$ under Eq. (2.6), ku_i ($ku_i(\kappa)$ and $\sigma_{ku_i(\kappa)}$ are simulation-based estimates of the mean and the standard deviation of $ku_i(\kappa)$ given Eq. (2.6). Under Eq. (2.6) with known κ, the null distributions of these statistics do not depend on B and $\sum$, so MC p-values [denoted $\hat{p}_N[ESK_M(\kappa_0)]$, $\hat{p}_N[EKU_M(\kappa_0)]$, $\hat{p}_N[ESKU_{KD}(\kappa_0)]$ and is $\hat{p}_N[EKU_{KD}(\kappa_0)]$ can be obtained as in Eq. (A.5). We also use the combined statistics

$$CSK_M(\kappa) = 1 - \min\{\hat{p}_N[ESK_M(\kappa)], \hat{p}_N[EKU_M(\kappa)]\}, \tag{B.8}$$

$$CSK_{KD}(\kappa) = 1 - \min\{\hat{p}_N[ESK_{KD}(\kappa)], \hat{p}_N[EKU_{KD}(\kappa)]\}, \tag{B.9}$$

and apply the MC test technique once again to obtain their p-values. In our empirical analysis, we apply all above test procedures to assess multivariate normality. In the Student-**t** case, we invert CSK_M (κ) given the results from Dufour et al. (2003) which document its superiority (in terms of power, in this case).

Appendix C. Multivariate diagnostic checks

We apply the multivariate diagnostic statistics from Dufour et al. (2010):

$$E = 1 - \min_{1 \le i \le n}[p(E_i)], LK = 1 - \min_{1 \le i \le n}[p(LK_i)], VR = 1 - \min_{1 \le i \le n}[p(VR_i)], \tag{C.1}$$

where, for equation i, E_i and LK_i are the GARCH test statistics from Engle (1982) and Lee and King (1993), VR_i is the variance ratio statistic from Lo and MacKinlay (1988, 1989), all statistics are computed from standardized residuals [namely $\tilde{W}_{it}$, the elements of

$\tilde{W}$ in Eq. (B.4)] and $p(E_i)$, $p(LK_i)$ and $p(VR_i)$ refer to p-values.[19] E_i is given by T× (the coefficient of determination in the regression of $\tilde{W}_{it}^2$ on a constant and $\tilde{W}_{i,(t-j)}^2$, $j=1,...,q$), and

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

$$VR_i = 1+2\sum_{j=1}^{K}\left(1-\frac{j}{K}\right)\hat{\rho}_{ij},\ \hat{\rho}_{ij} = \frac{\sum_{t=j+1}^{T}\tilde{W}_{it}\tilde{W}_{i,t-j}}{\sum_{t=1}^{T}\tilde{W}_{ti}^2},\ \hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\tilde{W}_{it}^2.$$

Under Eq. (2.6) with known κ, the null distributions of these statistics (see Dufour et al., 2010) do not depend on B and Σ, so MC p-values can be obtained as in Eq. (A.5). To deal with an unknown κ, we apply an MMC test procedure following the same technique we proposed for tests of Eq. (A.2). Specifically, we use the same confidence set for κ, of level $(1-\alpha_1)$; we then maximize the p-value function associated with E, LK and VR over all values of κ in the latter confidence set; we then refer the latter maximal p-value to α_2 where $\alpha=\alpha_1+\alpha_2$.

References

Affleck-Graves, J., McDonald, B., 1989. Nonnormalities and tests of asset pricing theories. Journal of Finance 44, 889–908.
Amengual, D., Sentana, E., 2010. A comparison of mean-variance efficiency tests. Journal of Econometrics 154, 16–34.
Anderson, T.W., 1984. An Introduction to Multivariate Statistical Analysissecond edn. John Wiley & Sons, New York.
Baillie, R., Bollerslev, T., 1989. The message in daily exchange rates: a conditional variance tale. Journal of Business and Economic Statistics 7, 297–305.
Banz, R.W., 1981. Anomalies in relationships between securities' yields and yield surrogates. Journal of Financial Economics 6, 103–126.
Basu, S., 1977. Investments performance of common stocks in relation to their price-earnings ratios: a test of the efficient market hypothesis. Journal of Finance 12, 129–156.
Beaulieu, M.-C., Dufour, J.-M., Khalaf, L., 2005. Exact multivariate tests of asset pricing models with stable asymmetric distributions. In: Breton, M., Ben Ameur, H. (Eds.), Numerical Methods in Finance. Kluwer, Dordrecht, The Netherlands.
Beaulieu, M.-C., Dufour, J.-M., Khalaf, L., 2007. Multivariate tests of mean-variance efficiency with possibly non-Gaussian errors: an exact simulation-based approach. Journal of Business and Economic Statistics 25, 398–410.
Beaulieu, M.-C., Dufour, J.-M., Khalaf, L., 2008a. Finite-sample identification-robust inference for unobservable zero-beta rates and portfolio efficiency with non-Gaussian distributions. Technical report. Mc Gill University, Université Laval and Carleton University.
Beaulieu, M.-C., Dufour, J.-M., Khalaf, L., 2008b. Identification robust inference in multivariate reduced rank regression and factor models. Technical report. Mc Gill University, Université Laval and Carleton University.
Beaulieu, M.-C., Dufour, J.-M., Khalaf, L., 2009. Finite sample multivariate tests of asset pricing models with coskewness. Computational Statistics and Data Analysis 53, 2008–2021.
Berk, J.B., 1997. Necessary conditions for the CAPM. Journal of Economic Theory 73, 245–257.
Berndt, E.R., Savin, N.E., 1977. Conflict among criteria for testing hypotheses in the multivariate linear regression model. Econometrica 45, 1263–1277.
Bhandari, L.C., 1977. Debt/equity ratio and expected common stocks returns: empirical evidence. Journal of Finance 9, 3–18.
Campbell, J.Y., Lo, A.W., MacKinlay, A.C., 1997. The Econometrics of Financial Markets. Princeton University Press, New Jersey.
Campbell, J.Y., Vuolteenaho, T., 2004. Bad beta, good beta. American Economics Review 94, 1249–1275.
Carhart, M.M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.
Chan, L.K., H.Y., Lakonishok, J., 1991. Fundamentals and stock returns in Japan. The Journal of Finance 45, 1739–1765.
Chen, N.-F., Roll, R., Ross, S.A., 1986. Economic forces and the stock market. The Journal of Business 59, 383–403.
Coudin, E., Dufour, J.-M., 2009. Finite-sample distribution-free inference in linear median regressions under heteroskedasticity and nonlinear dependence of unknown form. Econometrics Journal S1, S19–S49 (10th anniversary special edition).
De Roon, F.A., Nijman, T.E., 2001. Testing for mean-variance spanning: a survey. Journal of Empirical Finance 8, 111–155.
Dufour, J.-M., 2006. Monte Carlo tests with nuisance parameters: a general approach to finite-sample inference and nonstandard asymptotics in econometrics. Journal of Econometrics 443–478.
Dufour, J.-M., Khalaf, L., 2001. Monte Carlo test methods in econometrics. In: Baltagi, B. (Ed.), Basil Blackwell, Oxford, U.K, pp. 494–519. chapter 23.
Dufour, J.-M., Khalaf, L., 2002. Simulation based finite and large sample tests in multivariate regressions. Journal of Econometrics 111 (2), 303–322.
Dufour, J.-M., Khalaf, L., Beaulieu, M.-C., 2003. Exact skewness–kurtosis tests for multivariate normality and goodness-of-fit in multivariate regressions with application to asset pricing models. Oxford Bulletin of Economics and Statistics, pp. 891–906.
Dufour, J.-M., Khalaf, L., Beaulieu, M.-C., 2010. 'Multivariate residual-based finite-sample tests for serial dependence and GARCH with applications to asset pricing models'. Journal of Applied Econometrics 25 (2), 263–285.
Dufour, J.-M., Kiviet, J.F., 2004. Generic asymptotic validity of Monte Carlo tests. Technical report. CIRANO and CIREQ, Université de Montréal, and Tinbergen Institute, University of Amsterdam.
Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50 (4), 987–1008.
Fama, E.F., 1965. The behaviour of stock prices. Journal of Business 60, 401–424.
Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427–465.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.
Fama, E.F., French, K.R., 2004. The capital asset pricing model: theory and evidence. Journal of Economic Perspectives 18, 25–46.
Fama, E.F., MacBeth, J., 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy 71, 607–636.
Ferson, W.E., Harvey, C.R., 1999. Conditioning variables and the cross section of stock returns. The Journal of Finance 54, 1325–1360.
Fiorentini, G., Sentana, E., Calzolari, G., 2003. Maximum likelihodd estimation and inference in multivariate conditionally heteroskedastic dynamic models with Student t innovations. Journal of Business and Economic Statistics 4, 532–546.

[19] These may be obtained applying a MC test method, or using asymptotic null distributions, respectively, $\chi^2(q)$, the standard normal and $N[1, 2(2K-1)(K-1)/(3K)]$.

Geczy, C.C., 2001. Some generalized tests for mean-variance efficiency and multifactor model performance. Technical report. University of Pennsylvania, Philadelphia, Pennsylvania.
Gibbons, M.R., 1982. Multivariate tests of financial models: a new approach. Journal of Financial Economics 3–27.
Gibbons, M.R., Ross, S.A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.
Gouriéroux, C., Monfort, A., Renault, E., 1995. Inference in factor models. In: Maddala, G.S., Phillips, P.C.B., Srinivasan, T.N. (Eds.), Advances in Econometrics and Quantitative Economics. Blackwell, Oxford, U.K, pp. 311–353. chapter 13.
Groenwold, N., Fraser, P., 2001. Tests of asset-pricing models: how important is the iid-normal assumption. Journal of Empirical Finance 8, 427–449.
Grundy, B.D., Martin, J.S., 2001. Understanding the nature of the risks and the sources of the rewards to momentum investing. Review of Financial Studies 14, 29–78.
Gungor, S., Luger, R., 2009. Exact distribution-free tests of mean-variance efficiency. Journal of Empirical Finance 16, 816–829.
Hodgson, D.J., Linton, O., Vorkink, K., 2002. Testing the capital asset pricing model efficiently under elliptical symmetry: a semiparametric approach. Journal of Applied Econometrics 17, 617–639.
Hodgson, D.J., Vorkink, K., 2003. Efficient estimation of conditional asset pricing models. Journal of Business and Economic Statistics 21, 269–283.
Huberman, G., Kandel, S., 1987. Mean-variance spanning. Journal of Finance 42, 873–888.
Ingersoll, J., 1987. Theory of Financial Decision Making. Rowman & Littlefield, NJ.
Jagannathan, R., Wang, Z., 1996. The conditioning CAPM and the cross section of expected returns. The Journal of Finance 51, 3–53.
Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65–91.
Jegadeesh, N., Titman, S., 2001. Profitability of momentum strategies: an evaluation of alternative explanations. Journal of Finance 56, 699–720.
Jensen, M., 1968. The performance of mutual funds in the period 1945–64. Journal of Finance 23, 389–416.
Jobson, J.D., Korkie, B., 1982. Potential performance and tests of portfolio efficiency. Journal of Financial Economics 10, 433–466.
Jobson, J.D., Korkie, B., 1989. A performance interpretation of multivariate tests of asset set intersection, spanning, and mean-variance efficiency. Journal of Financial and Quantitative Analysis 24, 185–204.
Kan, R., Zhang, C., 1999a. GMM tests of stochastic discount factor models with useless factors. Journal of Financial Economics 54 (1), 103–127.
Kan, R., Zhang, C., 1999b. Two-pass tests of asset pricing models with useless factors. Journal of Finance 54, 204–235.
Kan, R., Zhou, G., 2008. Tests of mean-variance spanning. Technical report. Rotman School of Management, University of Toronto, Toronto, Canada.
Kandel, S., McCulloch, R., Stambaugh, R.F., 1995. Bayesian inference and portfolio efficiency. The Review of Financial Studies 8, 1–53.
Kandel, S., Stambaugh, R.F., 1989. A mean-variance framework for tests of asset pricing models. The Review of Financial Studies 2, 125–156.
Kariya, T., 1985. Testing in the Multivariate General Linear Model, number 22 in 'Economic Research Series, The Institute of Economic Research, Hitotsubashi University, Japan'. Kinokuniya Company Ltd., Tokyo.
Kariya, T., Kim, P., 1997. Finite sample robustness of tests: an overview. In: Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics 15: Robust Inference. Amsterdam, North-Holland, pp. 645–660. chapter 22.
Kilian, L., Demiroglu, U., 2000. Residual-based tests for normality in autoregressions: asymptotic theory and simulation evidence. Journal of Business and Economic Statistics 40–50.
Korkie, R.M., Turtle, H.J., 2001. What's a portfolio manager worth? A new style performance measure. Journal of Portfolio Management 28, 65–69.
Lakonishok, J., Shleifer, A., Vishny, R.W., 1994. Contrarian investment, extrapolation, and risk. Journal of Finance 49, 1541–1578.
Lee, J.H., 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 (1994), 17–27 (Correction 12 (1994), 139).
Lewellen, J., Nagel, S., Shanken, J., 2010. 'A skeptical appraisal of asset-pricing tests', Journal of Financial Economics. Journal of Financial Economics 96, 175–194.
Lewellen, J.W., Nagel, S., 2006. The conditional CAPM does not explain asset-pricing anomalies. Journal of Financial Economics 82, 289–314.
Lintner, J., 1965. The valuation of risk assets and selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.
Lo, A., 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., MacKinlay, C., 1989. The size and power of the variance ratio test in finite samples: a Monte Carlo investigation. Journal of Econometrics 40, 203–238.
MacKinlay, A.C., 1987. On multivariate tests of the Capital Asset Pricing Model. Journal of Financial Economics 18, 341–372.
MacKinlay, A.C., Pastor, L., 2000. Asset pricing models: implications for expected returns and portfolio selection. Review of Financial Studies 883–916.
MacKinlay, A.C., Richardson, M.P., 1991. Using generalized method of moments to test mean-variance efficiency. The Journal of Finance 46, 511–527.
Mardia, K.V., 1970. Measures of multivariate skewness and kurtosis with applications. Biometrika 519–530.
Mencia, J., Sentana, E., 2009. Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation. Journal of Econometrics 153, 105–121.
Merton, R.C., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867–887.
Penaranda, F., Sentana, E., 2008. Spanning tests in returns and stochastic discount factor mean-variance frontiers. A unifying approach, Technical report. Spain, Madrid. CEMFI Working Paper 0410, revised June 2008.
Perold, A.F., 2004. The capital asset pricing model. Journal of Economic Perspectives 18, 3–24.
Rao, C.R., 1973. Linear Statistical Inference and its Applicationssecond edn. John Wiley & Sons, New York.
Ray, S., Savin, E., 2008. The performance of heteroskedasticity and autocorrelation robust tests: a Monte Carlo study with an application to the three-factor Fama-French asset-pricing model. Journal of Applied Econometrics 91–109.
Richardson, M., Smith, T., 1993. A test for multivariate normality in stock returns. Journal of Business 66, 295–321.
Rosenberg, B., Reid, K., Lanstein, R., 1985. Persuasive evidence of market inefficiency. Journal of Portfolio Management 11, 9–17.
Ross, S.A., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360.
Sentana, E., 2009. The econometrics of mean-variance efficiency tests: a survey. Econometrics Journal C65–C101.
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 99–120.
Shanken, J., 1996. Statistical methods in tests of portfolio efficiency: a synthesis. In: Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics 14: Statistical Methods in Finance. North-Holland, Amsterdam, pp. 693–711.
Shanken, J., Weinstein, M.I., 2006. Economic forces and the stock market revisited. Journal of Empirical Finance 13, 129–144.
Sharpe, W.F., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442.
Stattman, D., 1980. Book values and stock returns. The Chicago MBA: A Journal of Selected Papers, pp. 25–45.
Stewart, K.G., 1997. Exact testing in multivariate regression. Econometric Reviews 16, 321–352.
Vorkink, K., 2003. Return distributions and improved tests of asset pricing models. The Review of Financial Studies 16 (3), 845–874.
Zhou, G., 1993. Asset-pricing tests under alternative distributions. The Journal of Finance 1927–1942.