Identification-robust estimation and testing of the zero-beta CAPM*

Marie-Claude Beaulieu † Université Laval

Jean-Marie Dufour ‡ McGill University

Lynda Khalaf § Carleton University

April 2012

* The authors thank two anonymous referees and the Editor, Bernard Salanié, 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 Montréal (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).

† 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) 656-2926, FAX: (418) 656-2624. email: Marie-Claude.Beaulieu@fas.ulaval.ca

‡ 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 H3A 2T7, Canada. TEL: (1) 514 398 4400 ext. 09156; FAX: (1) 514 398 4800; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

§ 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, K1S 5B6 Canada. Tel (613) 520-2600-8697; FAX: (613)-520-3906. email: Lynda_Khalaf@carleton.ca

ABSTRACT

We propose exact simulation-based procedures for: (i) testing mean-variance efficiency when the zero-beta rate is unknown; (ii) building confidence intervals for the zero-beta rate. On observing that this parameter may be weakly identified, we propose likelihood-ratio-type tests as well as Fieller-type procedures based on a Hotelling-HAC statistic, which are robust to weak identification and allow for non-Gaussian distributions including parametric GARCH structures. The Fieller-Hotelling-HAC procedure also accounts (asymptotically) for general forms of heteroskedasticity and autocorrelation. We propose confidence sets for the zero-beta rate based on "inverting" exact tests for this parameter; for both procedures proposed, these sets can be interpreted as multivariate extensions of the classic Fieller method for inference on ratios. The exact distribution of likelihood-ratio-type statistics for testing efficiency is studied under both the null and the alternative hypotheses. The relevant nuisance parameter structure is established and finite-sample bound procedures are proposed, which extend and improve available Gaussian-specific bounds. Finite-sample distributional invariance results are also demonstrated analytically for the HAC statistic proposed by MacKinlay and Richardson (1991). We study invariance to portfolio repacking for the tests and confidence sets proposed. The statistical properties of the proposed methods are analyzed through a Monte Carlo study and compared with alternative available methods. Empirical results on NYSE returns show that exact confidence sets are very different from asymptotic ones, and allowing for non-Gaussian distributions affects inference results. Simulation and empirical evidence suggests that likelihood-ratio-type statistics - with p-values corrected using the Maximized Monte Carlo test method - are generally preferable to their multivariate Fieller-Hotelling-HAC counterparts from the viewpoints of size control and power.

Key words: capital asset pricing model; CAPM; Black; mean-variance efficiency; non-normality; weak identification; Fieller; multivariate linear regression; uniform linear hypothesis; exact test; Monte Carlo test; bootstrap; nuisance parameters; GARCH; portfolio repacking.
Journal of Economic Literature classification: C3; C12; C33; C15; G1; G12; G14.

Contents

List of Assumptions, Propositions and Theorems **iii**

1. Introduction **1**

2. Model and zero-beta identification **4**
- 2.1. Notation . . . 5
- 2.2. Distributional assumptions . . . 5
- 2.3. Weak identification . . . 6
- 2.4. Standard estimators and test statistics . . . 7
  - 2.4.1. Gaussian-based statistics . . . 7
  - 2.4.2. Fieller-Hotelling-HAC statistics . . . 8

3. Identification-robust Monte Carlo tests for γ **9**
- 3.1. Distribution of $LR(\gamma_0)$ and $\mathscr{J}(\gamma_0)$ under $\mathscr{H}(\gamma_0)$ . . . 9
- 3.2. Monte Carlo test method . . . 10
- 3.3. Monte Carlo tests based on $LR(\gamma_0)$ and $\mathscr{J}(\gamma_0)$ . . . 12

4. Identification-robust confidence sets for γ **12**
- 4.1. Gaussian LR confidence sets . . . 12
- 4.2. Fieller-Hotelling-HAC confidence sets . . . 13

5. Invariance and exact distribution of LR_{B} **14**

6. Exact bound procedures for testing $\mathscr{H}_{\mathrm{B}}$ **15**
- 6.1. Global bound induced by tests of $\mathscr{H}(\gamma_0)$ . . . 15
- 6.2. Tighter bounds . . . 18
- 6.3. Two-stage bound confidence procedures . . . 19

7. Simulation study **20**

8. Empirical analysis **23**

9. Conclusion **26**

A. Appendix: Proofs **26**

List of Tables

1 Size Monte Carlo study 22
2 Power Monte Carlo study, Gaussian and Student designs 23
3 Power Monte Carlo study, Gaussian GARCH design 37
4 QML-based tests of BCAPM 38
5 Confidence sets for intervening parameters 39
6 QML-based point and set estimates for the zero–beta portfolio rate 40
7 Wald-HAC based inference on the zero-beta portfolio rate 41

List of Assumptions, Propositions and Theorems

3.1 **Theorem :** Distribution of the mean-variance CAPM test for a known zero-beta rate . 9
3.2 **Theorem :** Distribution of the Fieller-Hotelling-HAC statistic for γ 9
5.1 **Lemma :** Multivariate scale invariance 15
5.2 **Theorem :** Exact distribution of BCAPM LR tests 15
6.1 **Theorem :** Global bound on the null distribution of the BCAPM test 16
6.2 **Theorem :** Global simulation-based bound on the null distribution of the BCAPM test statistic 17
6.3 **Theorem :** Relation between efficiency tests and zero-beta confidence sets 17
Proof of Theorem **3.1** 26
Proof of Theorem **3.2** 26
Proof of Lemma **5.1** 28
Proof of Theorem **5.2** 29
Proof of Theorem **6.1** 30
Proof of Theorem **6.2** 31
Proof of Theorem **6.3** 31

1. Introduction

One of the most important extensions of the Capital Asset Pricing Model (CAPM) consists in allowing for the absence of a risk-free asset. From a theoretical viewpoint, this can be due to restrictions on borrowing [Black (1972)] or an investor's "riskless" borrowing rate that exceeds the Treasury bill rate [Brennan (1971)]. In this case, portfolio mean-variance efficiency is defined using the expected return in excess of the zero-beta portfolio. The latter is however unobservable which leads to considerable empirical difficulties.

Indeed, there are two basic approaches to estimating and assessing this version of the CAPM (denoted below as BCAPM). The first one uses a "two-pass" approach that may be traced back to Black, Jensen and Scholes (1972) and Fama and MacBeth (1973): *betas* are first estimated from time series regressions for each security, and then the zero-beta rate is estimated by a cross-sectional regression on these *betas*. This raises errors-in-variables problems that affect statistical inference in both finite and large samples.[1] The second approach – which originates in the work of Jensen (1968) – avoids this problem by using a multivariate linear regression as the basic statistical framework.[2] In this paper, we focus on the latter approach and consider two basic problems: (1) testing portfolio efficiency; (2) building a reliable confidence set for the zero-beta rate.

For clarity, let R_{it}, $i = 1, \ldots , n$, be the returns on n securities in period t, and $\tilde{R}_{Mt}$ the return on a market benchmark for $t = 1, \ldots , T$, and consider the n equations ($i = 1, \ldots , n$) associated with the time series regressions of R_{it} on a constant and $\tilde{R}_{Mt}$, where the individual-equation disturbances are heteroskedastic and contemporaneously cross-correlated; let $\Sigma = K'K$ refer to the error scale (or variance/covariance) matrix. If the intercepts from these n equations (the *alphas*) are denoted a_i, and the coefficients on the benchmark regressor (the *betas*) are denoted β_i, $i = 1, \ldots , n$, then the BCAPM equilibrium relations imply the following: there is a scalar γ, the return on the zero-beta portfolio, such that $a_i - \gamma(1 - \beta_i) = 0$, $i = 1, \ldots , n$. Our aim consists in assessing these constraints (denoted below as $\mathscr{H}_B$) as well as estimating γ.

The above cited literature provides analytical formulae for Gaussian likelihood-ratio statistics, the maximum likelihood estimator (MLE) of γ (denoted below as $\hat{\gamma}$), and for a conformable asymptotic variance estimator [denoted $\mathsf{Var}(\hat{\gamma})$]. It is however difficult to find reliable critical points in this context. While Gibbons (1982) used an asymptotic χ^2 critical value for the likelihood-ratio statistic, subsequent authors found this could lead to serious over-rejections, so various finite-sample corrections – such as bounds – have been suggested; see Shanken (1985, 1986, 1996), Stewart (1997), Zhou (1991, 1995), and Velu and Zhou (1999). These corrections depend crucially on normality, which may be inappropriate for financial data [see Fama (1965), Richardson and Smith (1993), Dufour, Khalaf and Beaulieu (2003), and Beaulieu, Dufour and Khalaf (2005, 2007, 2009, 2010)]. Furthermore, evidence on the properties of the confidence interval based on $\mathsf{Var}(\hat{\gamma})$ is unavailable. Despite the simplicity of the above framework, discrepancies between asymptotic and finite-sample

[1] See Litzenberger and Ramaswamy (1979), Banz (1981), Roll (1985), Chen, Roll and Ross (1986), Shanken (1992), Kim (1995), Shanken and Zhou (2007), Lewellen, Nagel and Shanken (2009), Kan, Robotti and Shanken (2012), and Kleibergen (2009).

[2] For other work based on the multivariate regression approach to CAPM analysis, see Gibbons (1982), Jobson and Korkie (1982), Kandel (1984, 1986), Amsler and Schmidt (1985), Shanken (1985, 1986, 1996), Kandel and Stambaugh (1989), Zhou (1991), Shanken (1992), Fama and French (1993), Chou (2000), Fama and French (2004) and Perold (2004).

distributions are not surprising. Indeed, three difficulties deserve notice.
(1) *Dimensionality*: as n increases, the dimension of the scale matrix Σ grows rapidly and available degrees-of-freedom decrease conformably.[3] Even in linear or standard setups where the relevant asymptotic distributions may be free of Σ, this matrix can still affect the distributions in finite samples. Furthermore, positive definite estimates of Σ require a large T relative to n, so portfolios rather than securities are often used in practice.
(2) *Portfolio repacking* [see Kandel and Stambaugh (1989)]: to preserve meaningful pricing relations when portfolios are used, transformations of the return vector $R_t = (R_{1t}, \ldots, R_{nt})'$ into $R_t^* = AR_t$ where A is an $n \times n$ invertible matrix such that $A\iota_n = \iota_n$ and ι_n is an n-dimensional vector of ones, should ideally not affect inference.
(3) *Identification*: as $\beta_i \to 1$, γ becomes weakly identified. Weak identification (WI) strongly affects the distributions of estimators and test statistics, leading to unreliable inference even asymptotically.[4] This should not be taken lightly: reported *betas* are often close to one [see e.g. Fama and MacBeth (1973)]. Further, even if estimated *betas* are not close to one, irregularities associated with WI are not at all precluded [in view of (1) and (2) above]. Indeed, in the regression of R_t^* [from (2)] on a constant and $\tilde{R}_{Mt}$, with intercepts a_i^* and slopes β_i^*, $a_i - \gamma(1-\beta_i) = 0$, $i = 1, \ldots, n$ $\Leftrightarrow a_i^* - \gamma(1-\beta_i^*) = 0$, $i = 1, \ldots, n$, for any γ and A. Portfolio repacking alters *betas* along with scale yet preserves the definition of γ, leading to identification problems as $\beta_i^* \to 1$. So the *betas* and scale parameters play a role in identifying γ.

Our aim in this paper consists in providing inference methods that are robust to dimensionality and identification problems, whose outcomes are invariant to portfolio repacking. We first consider the problem of estimating γ. We show by simulation that available procedures provide poor coverage. So we propose exact confidence sets based on "inverting" exact tests for specific values of γ, *i.e.* the set of values not rejected by these tests. This method is a generalization of the classical procedure proposed by Fieller (1954) to estimate parameter ratios.[5]

To introduce the Fieller-type method in its simplest form with reference to the problem at hand, suppose (for illustrative purposes) that we aim at estimating γ from the univariate regression of the return of the i-th security (R_{it}) on a constant and $\tilde{R}_{Mt}$, so that $\gamma = -a_i/\delta_i$ where $\delta_i = (\beta_i - 1)$. Let $\hat{a}_i$ and $\hat{\delta}_i$ denote the OLS estimates from this regression, with estimated variances and covariance $\mathsf{Var}(\hat{a}_i)$, $\mathsf{Var}(\hat{\delta}_i)$ and $\mathsf{Cov}(\hat{a}_i, \hat{\delta}_i)$. For each possible value γ_0 of the ratio, consider the t-statistic $\mathsf{t}_i(\gamma_0) = (\hat{a}_i + \gamma_0\hat{\delta}_i)/[\mathsf{Var}(\hat{a}_i) + \delta_0^2\mathsf{Var}(\hat{\delta}_i) + 2\delta_0\mathsf{Cov}(\hat{a}_i, \hat{\delta}_i)]^{1/2}$ for testing $\mathscr{H}_i(\gamma_0) : a_i + \gamma_0\delta_i = 0$. Then, we obtain a confidence set with level $1-\alpha$ for γ by finding the set of γ_0 values which are not rejected at level α using $\mathsf{t}_i(\gamma_0)$ and a standard (normal or Student-t) two-tailed critical value $\mathsf{z}_{\alpha/2}$. This means that we collect all γ_0 values such that $|\mathsf{t}_i(\gamma_0)| \le z_{\alpha/2}$ or alternatively such that $(\hat{a}_i + \gamma_0\hat{\delta}_i)^2 \le \mathsf{z}_{\alpha/2}^2(\mathsf{Var}(\hat{a}_i) + \gamma_0^2\mathsf{Var}(\hat{\delta}_i) + 2\gamma_0\mathsf{Cov}(\hat{a}_i, \hat{\delta}_i))$, leading to a second degree inequality in γ_0.

[3] See Shanken (1996), Campbell, Lo and MacKinlay (1997), Dufour and Khalaf (2002), Beaulieu, Dufour and Khalaf (2005, 2007, 2009, 2010), Sentana (2009), and the references therein.

[4] See, *e.g.* Dufour (1997, 2003), Staiger and Stock (1997), Wang and Zivot (1998), Zivot, Startz and Nelson (1998), Dufour and Jasiak (2001), Kleibergen (2002, 2005, 2009), Stock, Wright and Yogo (2002), Moreira (2003), Dufour and Taamouti (2005, 2007) and Andrews, Moreira and Stock (2006).

[5] For the ratio of the means of two normal variables with equal variances, Fieller gave a solution that avoids non-regularities arising from a close-to-zero denominator. Extensions to univariate regressions or to several ratios with equal denominators can be found in Zerbe (1978), Dufour (1997), Bolduc, Khalaf and Yelou (2010).

The resulting confidence set has level $1-\alpha$ irrespective whether δ_i is zero or not. In this paper, we generalize this method to account for the multivariate setup where γ appears, allowing for Gaussian and non-Gaussian error distributions, as well as conditional heteroskedasticity. Empirically, we focus on multivariate Student-t and normal mixture distributions, as well as Gaussian GARCH.

To do so, we consider two statistics [denoted $LR(\gamma_0)$ and $\mathscr{J}(\gamma_0)$] for testing $\mathscr{H}(\gamma_0): a_i+\gamma_0\delta_i = 0,\ i=1,\ \ldots\ ,\ n$. $LR(\gamma_0)$ is the likelihood-ratio statistic for the Gaussian error model, while $\mathscr{J}(\gamma_0)$ is a multivariate Fieller-type statistic based on a generalized Hotelling statistic to which a correction for heteroskedasticity and autocorrelation (HAC) has been applied [as in MacKinlay and Richardson (1991), Ravikumar, Ray and Savin (2000), and Ray and Savin (2008)]. Using any one of these tests, we can build confidence sets by finding the values of γ_0 which are not rejected at level α. This requires a distributional theory for the test statistics. While an F-based cut-off point is available for $LR(\gamma_0)$ in the *i.i.d.* Gaussian case [see Beaulieu, Dufour and Khalaf (2007) and Gibbons, Ross and Shanken (1989)], we show in a simulation study that usual asymptotic critical points perform poorly especially for $\mathscr{J}(\gamma_0)$.

To deal with such difficulties, we apply the maximized Monte Carlo (MMC) test procedure [Dufour (2006)] to obtain finite-sample p-values for $LR(\gamma_0)$ and $\mathscr{J}(\gamma_0)$ in models with non-Gaussian and/or non-*i.i.d.* errors, as follows: a (simulated) p-value function conditional on relevant nuisance parameters is numerically maximized (with respect to these parameters), and the test is significant at level α if the largest p-value is not larger than α.[6] The parametric bootstrap relates to the MMC method, in the sense that the maximization step is replaced by a unique p-value estimation, based on a consistent nuisance parameter estimate. For the GARCH case, such estimates may be unreliable in high-dimensional models; we show that the MMC method avoids this problem, with minimal power costs.

To implement this approach efficiently, it is important to characterize the nuisance parameters in the null distributions of the test statistics. We show that the null distribution of both $LR(\gamma_0)$ and $\mathscr{J}(\gamma_0)$ does not depend on B and Σ, so the only nuisance parameters characterize the form of error distribution: the degree of freedom for Student-t distributions, the mixing probability and scale-ratio parameters for normal mixtures, or GARCH parameters. While related invariance results are available for specific test problems using the multivariate regression likelihood ratio statistic [see Dufour and Khalaf (2002) and the references therein], our results on the HAC statistic $\mathscr{J}(\gamma_0)$ are new to the literature and underscore the usefulness of MacKinlay and Richardson's (1991) statistic.

Because an F-based exact cut-off is available for the Gaussian case, we show that the confidence set which inverts $LR(\gamma_0)$ can be obtained by solving a quadratic inequation. For non-*i.i.d.* or non-Gaussian distributions, we implement a numerical search running the MMC method for each choice for γ_0. Inverting $\mathscr{J}(\gamma_0)$ requires numerical methods even for the Gaussian case. Furthermore, we show that all proposed confidence sets provide relevant information on whether efficiency is supported by the data, a property not shared by standard confidence intervals. Indeed, our confidence sets may turn out to be empty, which occurs when all possible values of γ are rejected.

We next consider testing efficiency in the BCAPM context. We study likelihood-ratio and mul-

[6] This procedure is based on the following fundamental property: when the distribution of a test statistic depends on nuisance parameters, the desired level α is achieved by comparing the largest p-value (over all nuisance parameters consistent with the null hypothesis) with α.

tivariate Fieller-HAC criteria based on minimizing (over γ_0) the above defined $LR(\gamma_0)$ and $\mathscr{J}(\gamma_0)$ statistics. We show that the exact distribution of $\min_{\gamma_0}\{LR(\gamma_0)\}$ depends on a reduced number of nuisance parameters which are functions of both B and Σ. We also generalize Shanken's (1986) exact bound test beyond the Gaussian model, and propose a tighter bound, which involves a numerical search for the tightest cut-off point, based on the MMC method. The MMC based bound is also extended to the $\min_{\gamma_0}\{\mathscr{J}(\gamma_0)\}$ case. This approach, in conjunction with the above defined confidence set based on $\mathscr{J}(\gamma_0)$, provides an interesting alternative to available GMM estimation methods [including the case recently analyzed by Shanken and Zhou (2007)].

We present a simulation study to document the properties of the proposed procedures relative to available alternatives. In particular, we contrast problems arising from small samples with those caused by fundamentally flawed asymptotic approximations. We next examine efficiency of the market portfolio for monthly returns on New York Stock Exchange (NYSE) portfolios, built from the University of Chicago Center for Research in Security Prices (CRSP) 1926-1995 data base. We find more support for efficiency under the non-normal or non-*i.i.d.* hypothesis. Exact confidence sets for γ considerably differ from asymptotic ones, and the Fieller-HAC confidence sets are much wider than the GARCH corrected likelihood-ratio-based ones.

The paper is organized as follows. Section 2 sets the framework and discusses identification of γ. In Section 3, we propose finite-sample tests for specific values of γ, and the corresponding exact confidence set are derived in Section 4. The exact distribution of the likelihood-ratio efficiency test statistic is established in Section 5, and bound procedures are proposed in Section 6. The simulation study is reported in Section 7. Our empirical analysis is presented in Section 8. We conclude in Section 9.

2. Model and zero-beta identification

Let R_{it}, $i = 1, \ldots, n$, be the returns on n securities in period t, and $\tilde{R}_{Mt}$ the return on a market benchmark $(t = 1, \ldots, T)$. Our analysis of the BCAPM model is based on the following standard multivariate regression setup [Gibbons (1982), Shanken (1986), MacKinlay (1987)]:

$$R_{it} - \tilde{R}_{Mt} = a_i + (\beta_i - 1)\tilde{R}_{Mt} + u_{it}, \quad i = 1, \ldots, n, \; t = 1, \ldots, T, \tag{2.1}$$

where u_{it} is a random disturbance. The testable implication of the BCAPM on (2.1) is the following one: there is a scalar γ, the return on the zero-beta portfolio, such that

$$\mathscr{H}_{B} : a_i + \gamma\delta_i = 0, \quad \delta_i = \beta_i - 1, \quad i = 1, \ldots, n, \quad \text{for some } \gamma \in \varGamma, \tag{2.2}$$

where $\varGamma$ is the set of "admissible" values for γ. Since γ is unknown, $\mathscr{H}_B$ is nonlinear. The latter can be viewed as the union of more restrictive linear hypotheses of the form

$$\mathscr{H}(\gamma_0) : a_i + \gamma_0\delta_i = 0, \; i = 1, \ldots, n, \tag{2.3}$$

where γ_0 is specified. This observation underlies the exact inference approach proposed here.

2.1. Notation

The above model is a special case of the following multivariate regression:

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

where $Y = [Y_1, \ldots, Y_n]$ is $T \times n$, X is $T \times k$ of rank k, $U = [U_1, \ldots, U_n] = [V_1, \ldots, V_T]'$. For (2.1), $Y = [R_1, \ldots, R_n]$, $X = \left[\iota_T, \tilde{R}_{\mathrm{M}}\right]$, $R_i = (R_{i1}, \ldots, R_{iT})'$, $\tilde{R}_{\mathrm{M}} = \left(\tilde{R}_{\mathrm{M}1}, \ldots, \tilde{R}_{\mathrm{M}T}\right)'$, $B = [a, \beta]'$, $a = (a_1, \ldots, a_n)'$, $\beta = (\beta_1, \ldots, \beta_n)'$, and ι_T refers to a T-dimensional vector of ones.

Throughout the paper, we use the following notation. $\mathsf{P}_{(B,K)}$ represents the distribution of Y when the parameters are (B, K), where K represents parameters not included in B (such as parameters of the error distribution). For any matrix $n \times k$ matrix A, $M(A) = I - A(A'A)^{-}A'$ and $\mathrm{vec}(A)$ is the $(nk) \times 1$ vector obtained by stacking the columns of A on top of each other. We also use the following equivalent forms for the model, parameters, and hypotheses considered:

$$\tilde{Y} = Y - \tilde{R}_{\mathrm{M}}\iota_n' = XC + U\,, \tag{2.5}$$

$$\text{where } C = B - \Delta = [a, \beta - \iota_n]', \quad \Delta = [0, \iota_n]'; \tag{2.6}$$

$$\tilde{\mathscr{H}}(\gamma_0) : H(\gamma_0)C = 0\,, \text{ for } \gamma_0 \text{ specified,} \tag{2.7}$$

$$\tilde{\mathscr{H}}(\gamma_0) : \mathsf{R}(\gamma_0)\vartheta = 0\,, \text{ for } \gamma_0 \text{ specified,} \tag{2.8}$$

$$\text{where } H(\gamma_0) = (1, \gamma_0), \quad \mathsf{R}(\gamma_0) = H(\gamma_0) \otimes I_n\,,\ \vartheta = \mathrm{vec}(C')\,,\ \mathsf{R}(\gamma_0)\vartheta = [H(\gamma_0)C]'\,, \tag{2.9}$$

$$H(\gamma)C = 0 \Leftrightarrow \mathsf{R}(\gamma)\vartheta = 0, \text{ for all } \gamma; \tag{2.10}$$

$$\tilde{\mathscr{H}}_{\mathrm{B}} \quad : \quad H(\gamma)C = 0\,, \quad \text{for some } \gamma \in \Gamma\,, \tag{2.11}$$

$$\tilde{\mathscr{H}}_{\mathrm{B}} \quad : \quad \mathsf{R}(\gamma)\vartheta = 0, \quad \text{for some } \gamma \in \Gamma. \tag{2.12}$$

Note that (2.7) - (2.8) represent equivalent representations of $\tilde{\mathscr{H}}(\gamma_0)$, and similarly (2.11) - (2.12) for $\tilde{\mathscr{H}}_{\mathrm{B}}$.

2.2. Distributional assumptions

We further assume that we can condition on $\tilde{R}_{\mathrm{M}}$ and

$$V_t = (u_{1t}, \ldots, u_{nt})' = K'W_t\,, \quad t = 1, \ldots, T\,, \quad W_t = (W_{1t}, \ldots, W_{nt})', \tag{2.13}$$

where K is unknown and nonsingular, $W = [W_1, \ldots, W_T]'$ is independent of X, and the distribution of W is either fully specified or determined up to an unknown distributional shape parameter ν. (2.13) can also be rewritten in matrix form as $U = WK$.

We first present results which require no further regularity assumptions. We use additional restrictions, which entail that the distribution of W belongs to a specific family $\mathscr{H}_W(\mathscr{D}, \nu)$, where $\mathscr{D}$ represents a distribution type and $\nu \in \Omega_{\mathscr{D}}$ a nuisance parameter characterizing the distribution.

In particular, we consider the multivariate normal ($\mathscr{D}_N$), Student-t ($\mathscr{D}_t$) and normal mixture ($\mathscr{D}_m$) distributions:

$$\mathscr{H}_W(\mathscr{D}_N) : W_t \overset{i.i.d}{\sim} \mathrm{N}\big[0, I_n\big], \tag{2.14}$$

$$\mathscr{H}_W(\mathscr{D}_t, \kappa) : W_t = Z_{1t}/(Z_{2t}/\kappa)^{1/2}, \; Z_{1t} \overset{i.i.d}{\sim} \mathrm{N}\big[0, I_n\big], \; Z_{2t} \overset{i.i.d}{\sim} \chi^2(\kappa) , \tag{2.15}$$

$$\mathscr{H}_W(\mathscr{D}_m, \pi, \omega) : W_t = \mathsf{I}_t(\pi)Z_{1t} + [1 - \mathsf{I}_t(\pi)]Z_{3t}, \; Z_{3t} \overset{i.i.d}{\sim} \mathrm{N}\big[0, \omega I_n\big], \; 0 < \pi < 1 , \tag{2.16}$$

where Z_{2t} and Z_{3t} are independent of Z_{1t}, and $\mathsf{I}_t(\pi)$ is an indicator random variable independent of $\big(Z_{1t}, Z_{3t}\big)$ such that $\mathsf{P}[\mathsf{I}_t(\pi) = 0] = 1 - \mathsf{P}[\mathsf{I}_t(\pi) = 1] = \pi$. So, in (2.13), $\nu = \kappa$ under (2.15), and $\nu = (\pi, \omega)$ under (2.16). If $\mathsf{E}(W_t W_t') = I_n$, the covariance matrix of V_t is $\Sigma = K^{'} K$, so Σ is positive definite without further restrictions. If $K^{'}$ is lower triangular, $K^{'}$ and K correspond to the Cholesky factors of Σ.

Time-dependence may be taken into account by an appropriate specification of the distribution of W_t, $t = 1, \ldots , T$. Since time-varying volatility is prevalent in financial data, we consider the parametric GARCH structure:

$$u_{it} = w_{it} h_{it}^{\frac{1}{2}}, \qquad h_{it} = (1 - \phi_{1i} - \phi_{2i})\, \sigma_i^2 + \phi_{1i} w_{i,t-1}^2 + \phi_{2i} h_{i,t-1} , \tag{2.17}$$

where w_{it} are uncorrelated standard normal variables. This process may easily be reparameterized as in (2.13), where K is a diagonal matrix with diagonal terms $(1 - \phi_{1i} - \phi_{2i})^{1/2} \sigma_i$, $i = 1, \ldots , n$, and each W_{it} follows a univariate stationary GARCH process with unit intercept. Conforming with the above notation, we refer to this distributional hypothesis as $\mathscr{H}_W(\mathscr{D}_G, \phi)$, where ϕ is the $2n \times 1$ vector $(\phi_{11}, \ldots , \phi_{1n}, \phi_{21}, \ldots , \phi_{2n})$.[7]

2.3. Weak identification

Even though a_i and β_i are identifiable, γ is defined through a nonlinear transformation that may fail to be well-defined: the ratio $\gamma = a_i/(1 - \beta_i)$ is not defined or, equivalently, the equation $a_i = \gamma(1 - \beta_i)$ does not have a unique solution when $\beta_i = 1$.

In such situations, the distributions of many usual test statistics become non-standard, so the corresponding tests are unreliable and the associated confidence sets invalid. In particular, asymptotic standard errors are unreliable measures of uncertainty, and standard asymptotically justified t-type tests and confidence intervals have sizes that may deviate arbitrarily from their nominal levels; see the literature on weak identification [as reviewed, for example, in Dufour (2003) and Stock et al. (2002)]. Both the finite and large-sample distribution theory of most test statistics can be affected.

While it is straightforward to see that $\beta_i = 1$ corresponds to a discontinuity, the analysis below reveals this is not the whole story. In particular, we study the properties of estimators and test statistics following data transformations of the form $\tilde{Y}_* = \tilde{Y} A$, where A is any nonsingular fixed

[7] Ideally, a multivariate GARCH structure may be considered if T is sufficiently large relative to n; see Bauwens, Laurent and Rombouts (2006) for a recent survey. We adopt (2.17) since our empirical analysis relies on monthly data with 12 portfolios over 5 year subperiods (i.e. $T = 60$ and $n = 12$).

matrix of order n. On comparing (2.1) to its transformed counterpart, we see that irregularities cannot be safely assumed away, even when observed *betas* are not close to one.

2.4. Standard estimators and test statistics

One of the most common inference methods in this context relies on the log-likelihood

$$\ln[L(Y, B, \Sigma)] = -\frac{nT}{2}(2\pi) - \frac{T}{2}\ln(|\Sigma|) - \frac{1}{2}\mathrm{tr}[\Sigma^{-1}(Y - XB)'(Y - XB)]. \tag{2.18}$$

The unrestricted MLE of B and Σ are:

$$\hat{B} = (X'X)^{-1}X'Y = [\hat{a}, \hat{\beta}]', \quad \hat{\Sigma} = \hat{U}'\hat{U}/T,$$

where $\hat{U} = Y - X\hat{B}$, $\hat{a} = (\hat{a}_1, \ldots, \hat{a}_n)'$ and $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_n)'$. If $\hat{C}$ is the MLE of C in (2.5), the corresponding estimate of $\vartheta = \mathrm{vec}(C')$ is

$$\hat{\vartheta} = \mathrm{vec}(\hat{C}') \tag{2.19}$$

where ϑ is defined in (2.6).

2.4.1. Gaussian-based statistics

The likelihood-ratio statistic to test $\mathscr{H}(\gamma_0)$ where $\hat{\Sigma}(\gamma_0)$ is the MLE of Σ under $\mathscr{H}(\gamma_0)$ is:

$$LR(\gamma_0) = T\ln[\Lambda(\gamma_0)], \quad \Lambda(\gamma_0) = |\hat{\Sigma}(\gamma_0)|/|\hat{\Sigma}| = \frac{n}{T-n-1}\mathscr{F}(\gamma_0) + 1, \tag{2.20}$$

$$\hat{\Sigma}(\gamma_0) = \hat{\Sigma} + \frac{1}{T}\hat{B}'H(\gamma_0)'\left[H(\gamma_0)(X'X)^{-1}H(\gamma_0)'\right]^{-1}H(\gamma_0)\hat{B}, \tag{2.21}$$

$$\mathscr{F}(\gamma_0) = \frac{T-n-1}{n}\frac{(\hat{a} + \hat{\delta}\gamma_0)'\hat{\Sigma}^{-1}(\hat{a} + \hat{\delta}\gamma_0)}{1 + \left[(\hat{\mu}_M - \gamma_0)^2/\hat{\sigma}_M^2\right]}, \tag{2.22}$$

$$\hat{\mu}_M = \frac{1}{T}\sum_{t=1}^{T}\tilde{R}_{\mathrm{M}t}, \quad \hat{\sigma}_M^2 = \frac{1}{T}\sum_{t=1}^{T}(\tilde{R}_{\mathrm{M}t} - \hat{\mu}_M)^2, \quad \hat{\delta} = \hat{\beta} - \iota_n. \tag{2.23}$$

$\mathscr{F}(\gamma_0)$ is the Hotelling statistic for testing $\mathscr{H}(\gamma_0)$. Even though this statistic may look like a "Wald-type" statistic, the "covariance matrix" used for $\hat{a} + \hat{\delta}\gamma_0$ – namely, $(1 + [(\hat{\mu}_M - \gamma_0)^2/\hat{\sigma}_M^2])\hat{\Sigma}$ – depends on γ_0 (the tested value of γ). This characteristic makes it similar to Fieller-type statistics in univariate linear regressions [see Dufour (1997)], so $\mathscr{F}(\gamma_0)$ can be described as a "Fieller-Hotelling-HAC-type statistic". We will see below that this feature plays a crucial role in avoiding distributional problems associated with weak identification.

The likelihood-ratio criterion to test $\mathscr{H}_{\mathrm{B}}$ is

$$\begin{aligned} LR_{\mathrm{B}} &= T\ln(\Lambda_{\mathrm{B}}) = \inf\{LR(\gamma_0) : \gamma_0 \in \Gamma\} = LR(\hat{\gamma}), && (2.24)\\ \Lambda_{\mathrm{B}} &= |\hat{\Sigma}_{\mathrm{B}}|/|\hat{\Sigma}|, \quad |\hat{\Sigma}_{\mathrm{B}}| = \inf\{|\hat{\Sigma}(\gamma_0)| : \gamma_0 \in \Gamma\}, && (2.25)\end{aligned}$$

where $\hat{\Sigma}_{\mathrm{B}}$ is the MLE of Σ under $\mathscr{H}_{\mathrm{B}}$ and $\hat{\gamma}$ is the unrestricted MLE of γ; see Shanken (1986). The log-likelihood for (2.5) is

$$\ln\left[\tilde{L}(\tilde{Y}, C, \Sigma)\right] = \ln\left[L(Y - \tilde{R}_{\mathrm{M}}\iota_n', B - \Delta, \Sigma)\right] = \ln[L(Y, B, \Sigma)] \tag{2.26}$$

and the likelihood-ratio statistics for testing $\tilde{\mathscr{H}}(\gamma_0)$ and $\tilde{\mathscr{H}}_{\mathrm{B}}$ coincide with $LR(\gamma_0)$ and LR_{B}. Throughout the paper, we treat LR_{B} and $LR(\gamma_0)$ as quasi likelihood-ratio (QLR) criteria and the associated MLEs as quasi maximum likelihood (QML) estimators. We denote the observed value of these statistics as $LR_{\mathrm{B}}^{(0)}$ and $LR^{(0)}(\gamma_0)$, respectively.

A Wald-type formula for an asymptotic information-matrix-based standard error associated with $\hat{\gamma}$ is provided by Campbell et al. (1997, Chapter 5, equation 5.3.81):

$$\mathsf{Var}(\hat{\gamma}) = \frac{1}{T}\left[1 + \frac{(\hat{\mu}_M - \gamma)^2}{\hat{\sigma}_M^2}\right][(\iota_n - \beta)'\Sigma^{-1}(\iota_n - \beta)]^{-1}. \tag{2.27}$$

Whereas corrections may be derived for the non-Gaussian case [as in Barone-Adesi, Gagliardini and Urga (2004) who study a related asset pricing problem], general results on inference in the presence of identification failure (or weak identification) indicates that a "variance" estimator which does not depend on the tested value γ_0 – which is the case for $\mathsf{Var}(\hat{\gamma})$ – cannot lead to valid pivotal functions, so the associated "asymptotic" confidence sets and tests are fundamentally invalid; see Dufour (1997). Here, for example, it is easy to see that the above formula cannot be valid when $\beta = \iota_n$, and problematic when $\beta \backsimeq \iota_n$.

2.4.2. Fieller-Hotelling-HAC statistics

$\mathscr{F}(\gamma_0)$ may be viewed as a Hotelling-type statistic based on the standardized distance between $\hat{a} + \hat{\delta}\gamma_0$ and zero, which conveys an asymptotic least-squares [Gouriéroux, Monfort and Trognon (1985), Gouriéroux and Monfort (1995, Ch. 9)] and a GMM interpretation of $\hat{\gamma}$. This may be exploited to allow for serial dependence, for example by using a properly corrected weighting matrix, as done for example by MacKinlay and Richardson (1991), Ravikumar et al. (2000), and Ray and Savin (2008). This suggests the following statistic:

$$\mathscr{J}(\gamma_0) = T\,\hat{\vartheta}'\mathsf{R}(\gamma_0)'\left\{\mathsf{R}(\gamma_0)\left[\left(\frac{X'X}{T}\right)^{-1}\otimes I_n\right]\mathsf{S}_T\left[\left(\frac{X'X}{T}\right)^{-1}\otimes I_n\right]\mathsf{R}(\gamma_0)'\right\}^{-1}\mathsf{R}(\gamma_0)\hat{\vartheta}, \tag{2.28}$$

where $\mathsf{R}(\gamma_0)$ is defined in (2.9),

$$\mathsf{S}_T = \Psi_{0,t} + \sum_{j=1}^{q}\left(\frac{q-j}{q}\right)\left[\Psi_{j,T} + \Psi_{j,T}'\right], \quad \Psi_{j,T} = \frac{1}{T}\sum_{t=j+1}^{T}\left(X_t \otimes \hat{U}_t\right)\left(X_{t-j} \otimes \hat{U}_{t-j}\right)',$$

and $\hat{U}_t'$ is the t-th row of $\hat{U}$. A crucial feature of $\mathscr{J}(\gamma_0)$ comes from the fact that the covariance matrix estimator used for $\mathsf{R}(\gamma_0)\hat{\vartheta}$ depends on γ_0 (the tested value of γ). Since it is a HAC-modified version of $\mathscr{F}(\gamma_0)$, $\mathscr{J}(\gamma_0)$ can be described as a "Fieller-Hotelling-HAC" statistic or "$\mathscr{F}(\gamma_0)$-HAC" statistic.

Under $\mathscr{H}(\gamma_0)$, $\mathscr{J}(\gamma_0)$ follows a $\chi^2(n)$ distribution asymptotically. This result does not require any assumption on the identification of γ. A GMM estimator $\tilde{\gamma}$ of γ ,can be obtained by solving the problem

$$\mathscr{J}_{\mathrm{B}} = \inf\{\mathscr{J}(\gamma_0) : \gamma_0 \in \Gamma\} = \mathscr{J}(\tilde{\gamma}). \tag{2.29}$$

We denote the observed value of these statistics as $\mathscr{J}_{\mathrm{B}}^{(0)}$ and $\mathscr{J}^{(0)}(\gamma_0)$, respectively.

3. Identification-robust Monte Carlo tests for γ

We will now derive the exact null distribution of $LR(\gamma_0)$ and $\mathscr{J}(\gamma_0)$ under $\mathscr{H}(\gamma_0)$, where γ_0 is known. We then show how this result can be used to obtain exact Monte Carlo p-values. This will allow us to build confidence sets for γ and yield a way of testing efficiency.

3.1. Distribution of $LR(\gamma_0)$ and $\mathscr{J}(\gamma_0)$ under $\mathscr{H}(\gamma_0)$

Theorems **3.1** and **3.2** show that the null distribution of both $LR(\gamma_0)$ and $\mathscr{J}(\gamma_0)$ given X, is completely determined by X and the distribution of W given X. Proofs are given in the Appendix.

Theorem 3.1 DISTRIBUTION OF THE MEAN-VARIANCE CAPM TEST FOR A KNOWN ZERO-BETA RATE. *Under* (2.1), (2.13) *and* $\mathscr{H}(\gamma_0)$, $LR(\gamma_0)$ *is distributed like*

$$\overline{LR}(\gamma_0, W) = T\,\ln(|W'\bar{M}(\gamma_0)W| / |W'MW|) \tag{3.1}$$

where $\bar{M}(\gamma_0) = M(X) + X(X'X)^{-1}H(\gamma_0)'[H(\gamma_0)(X'X)^{-1}H(\gamma_0)']^{-1}H(\gamma_0)(X'X)^{-1}X'$.

In the *i.i.d.* Gaussian case (2.14), we have:

$$[(T-1-n)/n][\Lambda(\gamma_0) - 1] \sim F(n, T-1-n); \tag{3.2}$$

see Dufour and Khalaf (2002). This result was used by Gibbons et al. (1989) in studying efficiency with an observable risk-free rate. Indeed, testing $\mathscr{H}(\gamma_0)$ is equivalent to testing whether the intercepts are jointly zero in a market model with returns in excess of γ_0.

Theorem 3.2 DISTRIBUTION OF THE FIELLER-HOTELLING-HAC STATISTIC FOR γ. *Under* (2.1), (2.13) *and* $\mathscr{H}(\gamma_0)$, *the statistic* $\mathscr{J}(\gamma_0)$ *defined in* (2.28) *is distributed like*

$$\bar{\mathscr{J}}(\gamma_0, W) = TH(\gamma_0)(X'X)^{-1}X'W\bar{\mathscr{Q}}(\gamma_0, W)^{-1}W'X(X'X)^{-1}H(\gamma_0)', \tag{3.3}$$

where

$$\bar{\mathscr{Q}}(\gamma_0, W) = \left[H(\gamma_0)\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right]\hat{\mathsf{S}}_T\left[\left(\frac{X'X}{T}\right)^{-1} H(\gamma_0)' \otimes I_n\right],$$

$$\hat{\mathsf{S}}_T = \hat{\Psi}_{0,T} + \sum_{j=1}^{q}\left(\frac{q-j}{q}\right)\left[\hat{\Psi}_{j,T} + \hat{\Psi}'_{j,T}\right], \quad \hat{\Psi}_{j,T} = \frac{1}{T}\sum_{t=j+1}^{T}\left(X_t \otimes \hat{W}_t\right)\left(X_{t-j} \otimes \hat{W}_{t-j}\right)',$$

and $\hat{W}_t'$ *is the t-th row of* $\hat{W} = M(X)W$.

For non-Gaussian distributions compatible with (2.13) [including the GARCH case (2.17)], Theorem **3.1** shows that the exact distribution of both statistics, although non-standard, may easily be simulated once X, the distribution of W and γ_0 [given by $\mathscr{H}(\gamma_0)$] are set. This invariance property entails that the Monte Carlo (MC) test method can be easily applied, provided the distribution of W can be simulated (given a finite number of parameters); see Dufour (2006). By eliminating a potentially large number of nuisance parameters, this invariance is also relevant for applying asymptotic approximations. Of course, tests based on $\mathscr{J}(\gamma_0)$ are asymptotically valid in the usual way under standard asymptotic assumptions, such as those required by the HAC corrections considered by MacKinlay and Richardson (1991), Ravikumar et al. (2000), and Ray and Savin (2008).

3.2. Monte Carlo test method

This section explains the MC and MMC methods for implementing hypothesis tests based on a test statistic S. We also set associated notation, given the following assumptions on S.

In the context of model (2.4), consider a null hypothesis H_* which consists of (possibly nonlinear) restrictions on B, and an associated test statistic S. We reject H_* when $S \geq c$. Under H_*, S is distributed like a function $\bar{S}(\eta, W)$ of W and a vector of parameters $\eta \in \Xi$ (as well as the known matrix X), where Ξ describes the restricted parameter space (typically of dimension lower than the number of parameters in B and K). Finally, given (2.13), we can simulate W, once ν has been specified [e.g., as suggested by one of the assumptions (2.14) - (2.16)].

For notation simplicity, the dependence upon X is implicit through the definition of $\bar{S}$. Examples of $\bar{S}(.)$ include (3.1) and (3.3) in which case η corresponds to γ_0 which is set by the null hypothesis. The MC method assesses the rank of the observed value of S denoted $\left[\text{denoted } S^{(0)}\right]$, relative to a finite number N of simulated statistics [denoted $S^{(1)}, \ldots, S^{(N)}$] drawn under the null hypothesis. Given the above general assumptions, we can apply the following algorithm.

A1 Given: (i) a value of ν, (ii) N draws $W^{(1)}, \ldots, W^{(N)}$ from the distribution of W, (iii) a value of η, and (iv) the test function $\bar{S}(\eta, W)$, compute $\bar{S}(\eta, W)$ leading to the vector

$$\bar{S}_N(\eta, \nu) = \left[\bar{S}\big(\eta, W^{(1)}\big), \ldots, \bar{S}\big(\eta, W^{(N)}\big)\right]'. \tag{3.4}$$

A2 Build the *MC* p-value function

$$p_N[S^{(0)}|\bar{S}_N(\eta, \nu)] = \frac{NG_N\left[S^{(0)}; \bar{S}_N(\eta, \nu)\right] + 1}{N+1}, \tag{3.5}$$

$$G_N\left[S^{(0)}; \bar{S}_N(\eta, \nu)\right] = \frac{1}{N}\sum_{j=1}^{N} I_{[0,\infty)}\left[\bar{S}(W^{(j)}, \eta) - S^{(0)}\right], \tag{3.6}$$

where $I_A[x] = 1$, if $x \in A$, and $I_A[x] = 0$, if $x \notin A$. Set $\hat{p}_N(S; \eta, \nu) = p_N[S^{(0)}|\bar{S}_N(\eta, \nu)]$.

A3 If ν and η are set by H_*, the test which rejects H_* when

$$\hat{p}_N(S;\eta,\nu) \leq \alpha \tag{3.7}$$

is a test with *level* α for H_*; if furthermore the distribution of S is continuous under H_* and $\alpha(N+1)$ is an integer [as assumed in the cases considered in this paper], this test has *size* α: the probability of rejection under the null hypothesis is equal to α, for finite T and N. If ν or η is not set by the null hypothesis, then maximize $p_N[S^{(0)}\big|\bar{S}_N(\eta,\nu)]$ over all the (ν,η) values compatible with H_*, and reject the latter if the maximal p-value is less than or equal to α. For any $A \subseteq \Xi$ and $E \subseteq \Omega_{\mathscr{D}}$, let

$$\hat{p}_N(S;\eta,E) = \sup\{\hat{p}_N(S;\eta,\nu) : \nu \in E\}, \tag{3.8}$$

$$\hat{p}_N(S;A,\nu) = \sup\{\hat{p}_N(S;\eta,\nu) : \eta \in A\}, \tag{3.9}$$

$$\hat{p}_N(S;A,E) = \sup\{\hat{p}_N(S;\eta,\nu) : \eta \in A,\ \nu \in E\}, \tag{3.10}$$

where, by convention, $\hat{p}_N(S;A,\cdot) = 0$ if A is empty, and $\hat{p}_N(S;\cdot,E) = 0$ if E is empty. Then the probability of rejection under H_* based on either: (i) $\hat{p}_N(S;\eta,\Omega_{\mathscr{D}})$ when η is set by H_*, (ii) $\hat{p}_N(S;\Xi,\nu)$ when ν is set by H_*, or (iii) $\hat{p}_N(S;\Xi,\Omega_{\mathscr{D}})$ when both ν and η are not set by H_*, is itself not larger than α for finite T and N; see Dufour (2006).

A test for which the probability of rejection under the null hypothesis is not larger than α is sometimes described as a *conservative* test. We prefer to rely on the more precise and traditional distinction between a *level-correct* test and a *size-correct* test. A test is *level-correct for* α (or *exact at level* α) if the probability of rejection under all distributions compatible with the null hypothesis is *less than or equal* to α (for finite T). In contrast, a test is *size*-correct for α (or has *exact size* α) if the *supremum* of the probability of rejection under all distributions compatible with the null hypothesis is *equal* to α (for finite T). It is clear that a size-correct test is level-correct, but the converse may not hold. Both these definitions account for the common situation where the rejection probability under the null hypothesis may depend on nuisance parameters, and thus vary over the null hypothesis. We use the term *exact* when underlying probability statements hold for finite T and N, whether null rejection probabilities are equal to or not larger than the hypothesized cut-off. We follow the same notation in the case of a confidence set. A confidence set with exact size $1-\alpha$ implies that the infimum of the set's coverage probabilities compatible with the considered model is equal to $1-\alpha$, for finite T and N. In contrast, a confidence set with exact level α implies that the set's coverage probability is greater than or equal to $1-\alpha$, for finite T and N. This definition implies that inverting a size-correct test leads to a size-correct confidence set, while inverting a level-correct test leads to a level-correct confidence set.

When a Monte Carlo p-value is computed given a hypothesized distribution, and where in this case, the underlying test statistic converges to any distribution (say $\bar{F}$) the associated Monte Carlo test will remain asymptotically valid under any set of weaker assumptions for which the statistic still converges to the same limiting distribution $\bar{F}$. The implications for the present paper are as follows: our tests though exact with the considered distributions on W remain asymptotically valid in a semi-parametric context under the usual set of assumptions typically required for the CAPM to hold.

3.3. Monte Carlo tests based on $LR(\gamma_0)$ and $\mathscr{J}(\gamma_0)$

With respect to the algorithm in the previous section, in the case of $LR(\gamma_0)$, $S^{(0)} \equiv LR^{(0)}(\gamma_0)$, $\eta \equiv \gamma_0$, and using (3.1), $\bar{S}(\eta, W^{(i)}) = \overline{LR}(\gamma_0, W^{(i)})$. With $\bar{S}_N(\eta, \nu) = \overline{LR}_N(\gamma_0, \nu) = \left[\overline{LR}(\gamma_0, W^{(1)}), \ldots, \overline{LR}(\gamma_0, W^{(N)})\right]'$, and using the p-value function $p_N[.]$ from (3.5), let

$$\hat{p}_N(LR; \gamma_0, \nu) \equiv p_N\left[LR^{(0)}(\gamma_0)\middle|\overline{LR}_N(\gamma_0, \nu)\right]. \tag{3.11}$$

As a result of Theorem **3.1**, we have, under $\mathscr{H}(\gamma_0)$ in conjunction with $\mathscr{H}_W(\mathscr{D}, \nu)$ and given the notation from (3.8)-(3.10):

$$\mathsf{P}\left[\hat{p}_N(LR; \gamma_0, \nu_0) \leq \alpha\right] = \alpha\,, \text{ when } \nu = \nu_0, \tag{3.12}$$
$$\mathsf{P}\left[\hat{p}_N(LR; \gamma_0, \Omega_{\mathscr{D}}) \leq \alpha\right] \leq \alpha\,, \text{ when } \nu \text{ may be unknown.} \tag{3.13}$$

Similarly, in the case of $\mathscr{J}(\gamma_0)$, $S^{(0)} \equiv \mathscr{J}^{(0)}(\gamma_0)$, $\eta \equiv \gamma_0$, and using (3.3), $\bar{S}(\eta, W^{(i)}) = \bar{\mathscr{J}}(\gamma_0, W^{(i)})$. With $\bar{S}_N(\eta, \nu) = (\gamma_0, \nu) = \left[\bar{\mathscr{J}}(\gamma_0, W^{(1)}), \ldots, \bar{\mathscr{J}}(\gamma_0, W^{(N)})\right]'$, let

$$\hat{p}_N(\mathscr{J}; \gamma_0, \nu) \equiv p_N\left[\mathscr{J}^{(0)}(\gamma_0)\middle|\bar{\mathscr{J}}_N(\gamma_0, \nu)\right]. \tag{3.14}$$

As a result of Theorem **3.2**, we have, under $\mathscr{H}(\gamma_0)$ in conjunction with $\mathscr{H}_W(\mathscr{D}, \nu)$:

$$\mathsf{P}\left[\hat{p}_N(\mathscr{J}; \gamma_0, \nu_0) \leq \alpha\right] = \alpha\,, \text{ when } \nu = \nu_0, \tag{3.15}$$
$$\mathsf{P}\left[\hat{p}_N(\mathscr{J}; \gamma_0, \Omega_{\mathscr{D}}) \leq \alpha\right] \leq \alpha\,, \text{ when } \nu \text{ may be unknown.} \tag{3.16}$$

4. Identification-robust confidence sets for γ

Under $\mathscr{H}_{\mathrm{B}}$, the ratios $a_i/(1-\beta_i)$, 1, … , n, are equal. This definition of γ leads to the classical problem of inference on ratios from Fieller (1954). The problem here is clearly more complex, so to extend Fieller's arguments, we use the above defined tests of $\mathscr{H}(\gamma_0)$.

4.1. Gaussian LR confidence sets

Consider the Gaussian model given by (2.1), (2.13) and (2.14). In this case, under $\mathscr{H}_0(\gamma_0)$, $\mathscr{F}(\gamma_0)$ follows a Fisher distribution $F(n, T-n-1)$; see (3.2). Let F_α denote the cut-off point for a test with level α based on the $F(n, T-n-1)$ distribution. Then

$$CF_\gamma(\alpha) = \{\gamma_0 \in \Gamma : \mathscr{F}(\gamma_0) \leq F_\alpha\} \tag{4.1}$$

has level $1-\alpha$ for γ, *i.e.* the probability that γ be covered by $CF_\gamma(\alpha)$ is not smaller than $1-\alpha$. Indeed, $\mathsf{P}[\gamma \in CF_\gamma(\alpha)] = 1-\alpha$, that is $CF_\gamma(\alpha)$ is size-correct. On noting that $\mathscr{F}(\gamma_0) \leq F_\alpha$ can be rewritten as

$$M_F(\gamma_0) - \frac{nF_\alpha}{T-n-1} N_F(\gamma_0) \le 0\,, \tag{4.2}$$

$$M_F(\gamma_0) = \big(\hat{a}+\hat{\delta}\gamma_0\big)'\hat{\Sigma}^{-1}\big(\hat{a}+\hat{\delta}\gamma_0\big) = \big(\hat{\delta}'\hat{\Sigma}^{-1}\hat{\delta}\big)\gamma_0^2 + \big(2\hat{\delta}'\hat{\Sigma}^{-1}\hat{a}\big)\gamma_0 + \hat{a}'\hat{\Sigma}^{-1}\hat{a}\,, \tag{4.3}$$

$$N_F(\gamma_0) = 1 + \frac{(\hat{\mu}_M - \gamma_0)^2}{\hat{\sigma}_M^2} = \frac{1}{\hat{\sigma}_M^2}\gamma_0^2 - \frac{2\hat{\mu}_m}{\hat{\sigma}_M^2}\gamma_0 + 1 + \frac{\hat{\mu}_M^2}{\hat{\sigma}_M^2}\,, \tag{4.4}$$

we see, after a few manipulations, that $CF_\gamma(\alpha)$ reduces to a simple quadratic inequation:

$$CF_\gamma(\alpha) = \{\gamma_0 \in \Gamma : A\gamma_0^2 + B\gamma_0 + C \le 0\}\,, \tag{4.5}$$

$$A = \hat{\delta}'\hat{\Sigma}^{-1}\hat{\delta} - \left(\frac{nF_\alpha}{T-n-1}\right)\frac{1}{\hat{\sigma}_M^2}\,, \quad B = 2\left[\hat{\delta}'\hat{\Sigma}^{-1}\hat{a} + \left(\frac{nF_\alpha}{T-n-1}\right)\frac{\hat{\mu}_M}{\hat{\sigma}_M^2}\right], \tag{4.6}$$

$$C = \hat{a}'\hat{\Sigma}^{-1}\hat{a} - \left(\frac{nF_\alpha}{T-n-1}\right)\left[1 + \frac{\hat{\mu}_M^2}{\hat{\sigma}_M^2}\right]. \tag{4.7}$$

For $\Gamma = \mathbb{R}$, the resulting confidence set can take several forms depending on the roots of the polynomial $A\gamma_0^2 + B\gamma_0 + C$: (a) a closed interval; (b) the union of two unbounded intervals; (c) the entire real line; (d) an empty set.[8] Case (a) corresponds to a situation where γ is well identified, while (b) and (c) correspond to unbounded confidence sets and indicate (partial or complete) non-identification.

The possibility of getting an empty confidence set may appear surprising. But, on hindsight, this is quite natural: it means that no value of γ_0 allows $\mathscr{H}(\gamma_0)$ to be acceptable. Since $\mathscr{H}_\mathrm{B}$ states there exists a real scalar γ such that $a_i = (1-\beta_i)\gamma,\ i = 1,\ \ldots\ ,\ n$, this can be interpreted as a rejection of $\mathscr{H}_\mathrm{B}$. Further, under $\mathscr{H}_\mathrm{B}$, the probability that $CF_\gamma(\alpha)$ covers the true value γ is $1-\alpha$, and an empty set obviously does not cover γ. Consequently, the probability that $CF_\gamma(\alpha)$ be empty $[CF_\gamma(\alpha) = \emptyset]$ cannot be greater than α under $\mathscr{H}_\mathrm{B}$: $\mathsf{P}[CF_\gamma(\alpha) = \emptyset] \le \alpha$. The event $CF_\gamma(\alpha) = \emptyset$ is an exact critical region at level α for $\mathscr{H}_\mathrm{B}$ under normality (although its size may be smaller than α). When the confidence set $C_\gamma^{LR}(\alpha;\,\nu)$ is not empty, the identity $LR(\hat{\gamma}) = \inf\{LR(\gamma_0) : \gamma_0 \in \Gamma\}$ entails that $\hat{\gamma}$ must belong to $C_\gamma^{LR}(\alpha;\,\nu)$, for the critical value is the same for all γ_0.

4.2. Fieller-Hotelling-HAC confidence sets

The quadratic confidence set described above relies heavily on the fact that the same critical point F_α can be used to test all values of γ_0. This occurs under the *i.i.d.* Gaussian distributional assumption, but not necessarily otherwise. Although the quadratic confidence set will remain "asymptotically valid" as long as $\mathscr{F}(\gamma_0)$ converges to a $\chi^2(n)$ distribution, this cannot provide an exact confidence set. Consider the MC p-value $\hat{p}_N(LR;\gamma_0,\,\nu)$ function associated with this statistic, as defined in (3.11). Since the critical region $\hat{p}_N(LR;\gamma_0,\,\nu_0) \le \alpha$ has size α for testing $\gamma = \gamma_0$ when ν_0 is known,

[8]For further discussion, see Dufour and Jasiak (2001), Zivot et al. (1998), Dufour and Taamouti (2005), Kleibergen (2009), and Mikusheva (2009).

the set of γ_0 values for which $\hat{p}_N(\gamma_0, \nu_0)$ exceeds α, *i.e.*

$$C_\gamma^{LR}(\alpha; \nu) = \left\{\gamma_0 \in \Gamma : \hat{p}_N(LR; \gamma_0, \nu_0) > \alpha\right\}, \tag{4.8}$$

is a confidence set with size $1-\alpha$ for γ. Similarly, when ν is not specified, the test $\hat{p}_N(LR; \gamma_0, \Omega_{\mathscr{D}}) \leq \alpha$ yields:

$$C_\gamma^{LR}(\alpha; \mathscr{D}) = \left\{\gamma_0 \in \Gamma : \hat{p}_N(LR; \gamma_0, \Omega_{\mathscr{D}}) > \alpha\right\}, \tag{4.9}$$

whose level is $1-\alpha$. $C_\gamma(\alpha; \nu)$ or $C_\gamma(\alpha; \mathscr{D})$ must be drawn by numerical methods. Our empirical analysis reported below, relies on nested grid searches, over γ_0 and κ, for the Student-t case (2.15), and over γ_0 and (π, ω) for the normal-mixture case (2.16); for the GARCH case (2.17), we conduct a grid search on γ_0 where for each candidate value, we run the simulated annealing optimization algorithm to calculate the maximal p-value from (4.9) over the $2n$ nuisance parameters in ϕ.

We have no closed-form description of the structure of $C_\gamma^{LR}(\alpha; \nu)$ or $C_\gamma^{LR}(\alpha; \mathscr{D})$. While these can be bounded intervals (this is shown numerically in Section 8), $C_\gamma^{LR}(\alpha; \nu)$ or $C_\gamma^{LR}(\alpha; \mathscr{D})$ must be unbounded with a high probability if γ is not identifiable or weakly identified [see Dufour (1997)]. An empty confidence set is also possible and provides evidence that $\mathscr{H}_{\mathrm{B}}$ is not compatible with the data. The event $C_\gamma^{LR}(\alpha; \nu) = \emptyset$ [or $C_\gamma^{LR}(\alpha; \mathscr{D}) = \emptyset$] is a test with level α for $\mathscr{H}_{\mathrm{B}}$ under (2.13).

The Fieller-Hotelling-HAC confidence set we obtain for the GARCH case is exact, because the cut-off point we use when inverting $\mathscr{F}(\gamma_0)$ is adjusted for the parametric form (2.17) via the maximized p-value from (4.9). Inverting $\mathscr{J}(\gamma_0)$ in (2.28) may however be more appropriate. We thus define using (3.15)-(3.16)

$$C_\gamma^{\mathscr{J}}(\alpha; \nu_0) = \left\{\gamma_0 \in \Gamma : \hat{p}_N(\mathscr{J}; \gamma_0, \nu_0) > \alpha\right\}, \tag{4.10}$$

which gives a robust confidence set with size $1-\alpha$ for γ. Similarly, when ν is not specified, the test $\hat{p}_N(LR; \gamma_0, \Omega_{\mathscr{D}}) \leq \alpha$ yields

$$C_\gamma^{\mathscr{J}}(\alpha; \mathscr{D}) = \left\{\gamma_0 \in \Gamma : \hat{p}_N(\mathscr{J}; \gamma_0, \Omega_{\mathscr{D}}) > \alpha\right\}, \tag{4.11}$$

a robust confidence set with level $1-\alpha$. Again, this must be implemented by numerical methods. Inverting an asymptotic test based on $\mathscr{J}(\gamma_0)$ can also be considered, relaxing the GARCH restriction. For example, a grid search can be conducted on γ_0 where for each candidate value, $\mathscr{J}(\gamma_0)$ is referred to the $\chi^2(n)$ distribution; this would circumvent the identification problem asymptotically [as argued e.g. in Stock and Wright (2000)], yet in finite samples, the $\chi^2(n)$ approximation may perform poorly. Indeed, our simulation results reported below illustrate the severity of this problem. Consequently, we use the MMC method for each candidate γ_0: we maximize over the model parameters as well as over ϕ.

5. Invariance and exact distribution of LR_{B}

In this section, we study the exact distribution of the statistic LR_{B}, under both the null hypothesis and the corresponding unrestricted multivariate regression alternative model. Our analysis also provides further information (beyond Theorem **3.1**) on distribution of $LR(\gamma_0)$. We track and control for the joint role *betas* and scale parameters play in identifying γ.

Lemma 5.1 MULTIVARIATE SCALE INVARIANCE. *The likelihood-ratio statistics $LR(\gamma_0)$ and LR_{B} defined in* (2.24) *and* (2.20) *are invariant to replacing $\tilde{Y}$ by $\tilde{Y}_* = \tilde{Y}A$, where A is an arbitrary nonsingular $n \times n$ matrix.*

Such transformations can be viewed as the following affine transformations on Y :

$$Y_* = YA + \tilde{R}_{\mathrm{M}}\iota_n'(I_n - A). \tag{5.1}$$

Theorem 5.2 EXACT DISTRIBUTION OF BCAPM LR TESTS. *Under* (2.1) *and* (2.13), *the distributions of $LR(\gamma_0)$ and LR_{B} depend on (B, K) only through $\bar{B} = (B - \Delta)K^{-1}$, and*

$$LR(\gamma_0) = T\,\ln\left(|\hat{W}(\gamma_0)'\hat{W}(\gamma_0)|/|\hat{W}'\hat{W}|\right), \quad LR_{\mathrm{B}} = \inf\{LR(\gamma_0) : \gamma_0 \in \Gamma\}, \tag{5.2}$$

where $\Delta = [0,\, \iota_n]'$, $\hat{W} = M(X)W$, $\bar{M}(\gamma_0)$ is defined as in (3.1) *and*

$$\hat{W}(\gamma_0) = \bar{M}(\gamma_0)(X\bar{B} + W) = \bar{M}(\gamma_0)\{\iota_T[a + \gamma_0(\beta - \iota_n)]'K^{-1} + W\}. \tag{5.3}$$

If, furthermore, the null hypothesis $\mathscr{H}_{\mathrm{B}}$ holds, then

$$\hat{W}(\gamma_0) = (\gamma_0 - \gamma)\bar{M}(\gamma_0)\iota_T(\beta - \iota_n)'K^{-1} + \bar{M}(\gamma_0)W \tag{5.4}$$

and the distribution of LR_{B} depends on (B, K) only through γ and $(\beta - \iota_n)'K^{-1}$; in the Gaussian case (2.14), *this distribution involves only one nuisance parameter.*

Even though B and K may involve up to $2n+\,n^2$ different nuisance parameters [or $2n + n(n+1)/2$ parameters, if K is triangular], the latter theorem shows that the number of free parameters in the distributions of $LR(\gamma_0)$ and LR_{B} does not exceed $2n$; when $\mathscr{H}_{\mathrm{B}}$ holds, the number of free parameters is at most $n+1$. Further, under $\mathscr{H}(\gamma_0)$ [using (5.4)] $\bar{B}$ is evacuated, entailing Theorem **3.1**. Theorem **5.2** also provides the power function.

6. Exact bound procedures for testing $\mathscr{H}_{\mathrm{B}}$

In this section, we propose tests for $\mathscr{H}_{\mathrm{B}}$ in the presence of nuisance parameters induced by nonlinearity and non-Gaussian error distributions. Because our rationale leads to Shanken's (1986) bound for the Gaussian case, we study first global bounds based on tests of $\mathscr{H}(\gamma_0)$. Second, we describe more general but computationally more expensive methods based on the technique of MMC tests to obtain tighter bounds.

6.1. Global bound induced by tests of $\mathscr{H}(\gamma_0)$

The results of Section 3 on testing $\gamma = \gamma_0$ can be used to derive a global bound on the distribution of the statistic LR_{B}. This is done in the following theorem.

Theorem 6.1 GLOBAL BOUND ON THE NULL DISTRIBUTION OF THE BCAPM TEST. *Under the assumptions* (2.1), (2.13) *and* $\mathscr{H}_{\mathrm{B}}$, *we have, for any given* $\nu \in \Omega_{\mathscr{D}}$,

$$\mathsf{P}\big[LR_{\mathrm{B}} \geq x\big] \leq \sup_{\gamma_0 \in \Gamma} \mathsf{P}\big[\overline{LR}(\gamma_0, W) \geq x\big], \forall x, \tag{6.1}$$

where $\overline{LR}(\gamma_0, W)$ *is defined in* (3.1). *Further, in the Gaussian case* (2.14), *we have:*

$$\mathsf{P}\big[(T-1-n)(\Lambda_{\mathrm{B}}-1)/n \geq x\big] \leq \mathsf{P}[F(n, T-1-n) \geq x], \ \forall x. \tag{6.2}$$

To relate this result to available bounds, observe that (6.1) and (6.2) easily extend to the following multi-beta setups: for $i = 1, \ldots, n$, $t = 1, \ldots, T$,

$$R_{it} = a_i + \sum_{j=1}^{s} \beta_{ij} \tilde{R}_{jt} + u_{it}, \qquad \mathscr{H}_{\mathrm{B}}: a_i = \gamma\Big(1 - \sum_{j=1}^{s} \beta_{ij}\Big), \tag{6.3}$$

where $\tilde{R}_{jt}$, $j = 1, \ldots, s$, are returns on s benchmarks. In this case, the bounding distribution of LR_{B} obtains as in Theorem **6.1** where $X = [\iota_T, \tilde{R}_1, \ldots, \tilde{R}_s]$, $\tilde{R}_j = (\tilde{R}_{j1}, \ldots, \tilde{R}_{jT})'$, $j = 1, \ldots, s$, and H is the k-dimensional row vector $(1, \gamma_0, \ldots, \gamma_0)$. In the Gaussian case, $\mathsf{P}\big[\overline{LR}(\gamma_0, W) \geq x\big]$ does not depend on γ_0, and the bounding distribution under normality is $F(n, T-s-n)$. Shanken (1986) suggests the statistic

$$\hat{Q} = \min_{\gamma} \left\{ \frac{T\big[\hat{a} - \gamma(\iota_n - \hat{\beta}\iota_s)\big]'\big[\big(T/(T-2)\big)\hat{\Sigma}\big]^{-1}\big[\hat{a} - \gamma(\iota_n - \hat{\beta}\iota_s)\big]}{1 + (\bar{R}_{\mathrm{M}} - \gamma\iota_s)'\hat{\Delta}_{\mathrm{M}}^{-1}(\bar{R}_{\mathrm{M}} - \gamma\iota_s)} \right\} \tag{6.4}$$

where $\hat{a}$ is an n-dimensional vector which includes the (unconstrained) intercept estimates, $\hat{\beta}$ is an $n \times s$ matrix whose rows include the unconstrained OLS estimates of $(\beta_{i1}, \ldots, \beta_{is})$, $i = 1, \ldots, n$, $\bar{R}_{\mathrm{M}}$ and $\hat{\Delta}_{\mathrm{M}}$ include respectively the time-series means and sample covariance matrix corresponding to the right-hand-side total portfolio returns. Further, the minimum in (6.4) occurs at the constrained MLE $\hat{\gamma}$ of γ, and

$$LR_{\mathrm{B}} = T\ln(1 + \hat{Q}/(T-s-1)). \tag{6.5}$$

For normal errors, $(T-s-n)\hat{Q}/[n(T-s-1)]$ can be bounded by the $F(n, T-n-s)$ distribution. The latter obtains from Gibbons et al.'s (1989) joint test of zero intercepts, where returns are expressed in excess of a known γ.

Independently, Stewart (1997) shows [using Dufour (1989)] under normal errors, that $(T - s - n)\big[\big(|\hat{\Sigma}_{\mathrm{B}}|/|\hat{\Sigma}|\big) - 1\big]/n$ can be bounded by the $F(n, T-n-s)$ distribution. Now, from (2.24) and (6.5), we see that Shanken and Stewart's bounds are equivalent, and both results obtain from Theorem **6.1** in the special case of normal errors.

When disturbances are non-Gaussian, Theorem **6.1** entails that the bounding distribution can easily be simulated. Using the notation we introduced in subsection 3.2 [definitions of the p-value function $p_N[.]$ from (3.5) and the maximized p-values (3.8)-(3.10)] the following algorithm may be used.

B1 Given a value of ν, generate N *i.i.d.* draws from the distribution of W then, for any given γ_0, apply the (bounding) function $\overline{LR}(\gamma_0, W)$ from (3.1) to each draw. With reference to Section 3.2, this implies using $\overline{LR}(\gamma_0, W)$ for $\bar{S}(\eta, W)$, which will yield a vector $\overline{LR}_N(\gamma_0, \nu)$ of N simulated values of the bounding test statistic.

B2 Compute the MC p-value [denoted the *bound MC* (BMC) p-value]

$$\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \gamma_0, \nu) \equiv p_N[LR_{\mathrm{B}}^{(0)} | \overline{LR}_N(\gamma_0, \nu)] . \tag{6.6}$$

In contrast with the Gaussian case, $\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \gamma_0, \nu)$ may depend on γ_0; nevertheless, for any γ_0,

$$LR_{\mathrm{B}} \leq LR(\gamma_0) \Rightarrow \hat{p}_N(LR; \gamma_0, \nu) \leq \hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \gamma_0, \nu). \tag{6.7}$$

B3 Critical regions that provably satisfy the level constraint can be obtained by maximizing $\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \gamma_0, \nu)$ over the relevant nuisance parameters, leading to the maximized p-values $\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \Gamma, \nu)$, $\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \gamma_0, \Omega_{\mathscr{D}})$ and $\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \Gamma, \Omega_{\mathscr{D}})$.

Theorem 6.2 GLOBAL SIMULATION-BASED BOUND ON THE NULL DISTRIBUTION OF THE BCAPM TEST STATISTIC. *Under* (2.1), (2.13) *and* $\mathscr{H}_{\mathrm{B}}$*, we have, using the notation in* (3.8)-(3.10)*:*

$$\mathsf{P}\big[\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \Gamma, \nu) \leq \alpha\big] \leq \alpha , \quad \mathsf{P}\big[\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \Gamma, \Omega_{\mathscr{D}}) \leq \alpha\big] \leq \alpha , \tag{6.8}$$

where ν *represents the true distributional shape of* W.

These bound tests are closely related to the confidence set-based test proposed in Section 4: the null hypothesis is rejected when the confidence set for γ is empty, *i.e.* if no value of γ_0 can be deemed acceptable (at level α), either with ν specified or ν taken as a nuisance parameter. This may be seen on comparing (4.9) with the probabilities in Theorem **6.2**. Since $LR_{\mathrm{B}} = \inf\{LR(\gamma_0) : \gamma_0 \in \Gamma\}$, this suggests a relatively easy way of showing that $C_{\gamma}^{LR}(\alpha; \nu)$ or $C_{\gamma}^{LR}(\alpha; \mathscr{D})$ is not empty, through the specific p-value $\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \hat{\gamma}, \nu)$ obtained by taking $\gamma_0 = \hat{\gamma}$ in (6.6). We shall call $\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \hat{\gamma}, \nu)$ the *QML-BMC* p-value.

Theorem 6.3 RELATION BETWEEN EFFICIENCY TESTS AND ZERO-BETA CONFIDENCE SETS. *Under* (2.1), (2.13) *and* $\mathscr{H}_{\mathrm{B}}$*, let* $\hat{\gamma}$ *be the QML estimator of* γ *in* (2.25)*. Then, using the notation from* (3.8)-(3.10)*:*

$$\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \hat{\gamma}, \nu) > \alpha \Rightarrow \hat{p}_N(LR_{\mathrm{B}}; \Gamma, \nu) > \alpha \Rightarrow C_{\gamma}^{LR}(\alpha; \nu) \neq \emptyset , \ \forall \nu \in \Omega_{\mathscr{D}} ,$$
$$\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \hat{\gamma}, \Omega_{\mathscr{D}}) > \alpha \Rightarrow \hat{p}_N(LR_{\mathrm{B}}; \Gamma_0, \Omega_{\mathscr{D}0}) > \alpha \Rightarrow C_{\gamma}^{LR}(\alpha; \mathscr{D}) \neq \emptyset ,$$

where $C_{\gamma}^{LR}(\alpha, \nu)$ *and* $C_{\gamma}^{LR}(\alpha; \mathscr{D})$ *are the sets defined in* (4.8) *and* (4.9).

For the Gaussian case, Zhou (1991) and Velu and Zhou (1999) proposed a potentially tighter bound applicable to statistics which can be written as ratios of independent Wishart variables and does not seem to extend easily to other classes of distributions. In the next section, we propose an approach which yields similarly tighter bounds for non-Gaussian distributions as well. Finally, the HAC statistic $\mathscr{J}_{\mathrm{B}}$ may be used to obtain alternative identification-robust bound tests following the same rationale. The correspondence between such tests and empty confidence sets entailed by test inversion also follows from similar arguments. Finite-sample MMC level corrections are recommended, given the simulation results in Section 7.

6.2. Tighter bounds

Another approach to testing $\mathscr{H}_{\mathrm{B}}$ with the statistic LR_{B} consists in directly assessing its dependence on nuisance parameters and adjusting the test accordingly through the MMC method. Let $\theta = \psi(B, K)$ represent the parameter vector upon which the distribution of LR_{B} actually depends, and Ω_{B} the set of admissible values for θ under $\mathscr{H}_{\mathrm{B}}$. The dimension of θ may be lower than the number of parameters in B and K. Define the function $\overline{LR}_{\mathrm{B}}(\theta, W) = \overline{LR}_{\mathrm{B}}\big(\psi(B, K), W\big)$ which assigns to each value of (B, K) and the noise matrix W the following outcome: using θ and a draw from the distribution of W (which may depend on ν), generate a sample from (2.1)-(2.2), and compute LR_{B} [as defined in (2.24)] from this sample. This suggests we can use $\overline{LR}_{\mathrm{B}}(\theta, W)$ for $\bar{S}(\eta, W)$. Maintaining further notation from Section 3.2 [again, recall in particular the definitions of the p-value function $p_N[.]$ from (3.5) and the maximized p-values (3.8)-(3.10)] the following algorithm may be used.

- C1 Generate N independent replications of W and apply $\overline{LR}_{\mathrm{B}}(\theta, W)$ to each draw, for any value of θ, leading to a vector $\overline{LR}_{\mathrm{B}N}(\theta, \nu)$ of N simulated statistics.
- C2 Compute the MC p-value
$$\hat{p}_N^{\mathrm{B}}(LR_{\mathrm{B}}; \theta, \nu) = p_N\big[LR_{\mathrm{B}}^{(0)} \big| \overline{LR}_{\mathrm{B}N}(\theta, \nu)\big].$$
- C3 Critical regions that provably satisfy the level constraint can be obtained by maximizing $\hat{p}_N^{\mathrm{B}}(LR_{\mathrm{B}}; \theta, \nu)$ over relevant nuisance parameters leading to $\hat{p}_N^{\mathrm{B}}(LR_{\mathrm{B}}; \Omega_{\mathrm{B}}, \nu)$, $\hat{p}_N^{\mathrm{B}}(LR_{\mathrm{B}}; \theta, \Omega_{\mathscr{D}})$ and $\hat{p}_N^{\mathrm{B}}(LR_{\mathrm{B}}; \Omega_{\mathrm{B}}, \Omega_{\mathscr{D}})$.

Theorem **6.3** guarantees that $\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \Gamma, \nu) \leq \alpha \Rightarrow \hat{p}_N^{\mathrm{B}}(LR_{\mathrm{B}}; \Omega_{\mathrm{B}}, \nu) \leq \alpha$ for any given ν. So it may be useful to check the global bound for significance before turning to the MMC one. Furthermore, it is not always necessary to run the numerical maximization underlying MMC to convergence: if $\hat{p}_N^{\mathrm{B}}(LR_{\mathrm{B}}; \theta, \nu) > \alpha$ given any relevant θ (or ν), then a non-rejection is confirmed. We suggest to use the QML estimate $\hat{\theta}$ of θ as start-up value, because this provides *parametric bootstrap-type* [or a *local MC* (LMC)] p-values:

$$p_N^b(LR_{\mathrm{B}}; \nu) = \hat{p}_N^{\mathrm{B}}(LR_{\mathrm{B}}; \hat{\theta}, \nu), \quad p_N^b(LR_{\mathrm{B}}; \Omega_{\mathscr{D}}) = \hat{p}_N^{\mathrm{B}}(LR_{\mathrm{B}}; \hat{\theta}, \Omega_{\mathscr{D}}). \tag{6.9}$$

Then $p_N^b(LR_{\mathrm{B}}; \nu) > \alpha$ entails $\hat{p}_N^{\mathrm{B}}(LR_{\mathrm{B}}; \Omega_{\mathrm{B}}, \nu) > \alpha$, and $p_N^b(LR_{\mathrm{B}}; \Omega_{\mathscr{D}}) > \alpha$ entails $\hat{p}_N^{\mathrm{B}}(LR_{\mathrm{B}}; \Omega_{\mathrm{B}}, \Omega_{\mathscr{D}}) > \alpha$.

Following the same reasoning, a parametric MMC test imposing (2.17) may be applied to the HAC statistics $\mathscr{J}(\gamma_0)$ and $\mathscr{J}_{\mathrm{B}}$, as an attempt to correct their size for the GARCH alternative of interest. We investigate the size-corrected power associated with these statistics in Section 7.

Weak identification of γ implies that the data is weakly informative on the parameter and thus on the related specification. This means that the data may not provide sufficient information on whether $\mathscr{H}_{\mathrm{B}}$ should be refuted or not. When γ is weakly identified, tests for $\mathscr{H}_{\mathrm{B}}$ will suggest accepting the underlying pricing restrictions. We thus recommend to interpret tests for $\mathscr{H}_{\mathrm{B}}$ in conjunction with associated confidence sets which will be unbounded when γ is weakly identified and will provide a much more complete statistical analysis.

6.3. Two-stage bound confidence procedures

To deal with the fact that the distribution of W may involve an unknown parameter $\nu \in \Omega_{\mathscr{D}}$, we suggested above to maximize the relevant p-values over $\Omega_{\mathscr{D}}$. We next consider restricting the maximization over ν to a set which is empirically relevant, as in Beaulieu et al. (2007). This leads to two basic steps: (i) an exact confidence set with level $1-\alpha_1$ is built for ν, and (ii) the MC p-values (presented above) are maximized over all values of ν in the latter confidence set and are referred to level α_2, so that the global test level is $\alpha = \alpha_1 + \alpha_2$. In our empirical application, we used $\alpha/2$. Let $\mathscr{C}_\nu(\alpha_1) = \mathscr{C}_\nu(\alpha_1; Y)$ be a confidence set with level $1-\alpha_1$ for ν. Then, under $\mathscr{H}(\gamma_0)$, we have $\mathsf{P}\left[\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \gamma_0, \mathscr{C}_\nu(\alpha_1)) \leq \alpha_2\right] \leq \alpha_1 + \alpha_2$ while, under $\mathscr{H}_{\mathrm{B}}$:

$$\mathsf{P}\left[\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}}; \Gamma, \mathscr{C}_\nu(\alpha_1)) \leq \alpha_2\right] \leq \alpha_1 + \alpha_2, \qquad \mathsf{P}\left[\hat{p}_N^{\mathrm{B}}(LR_{\mathrm{B}}; \Omega_{\mathrm{B}}, \mathscr{C}_\nu(\alpha_1)) \leq \alpha_2\right] \leq \alpha_1 + \alpha_2 . \tag{6.10}$$

Note also that for $\hat{p}_N^{\mathrm{B}}(LR_{\mathrm{B}}; \Omega_{\mathrm{B}}, \mathscr{C}_\nu(\alpha_1)) \leq \alpha_2$ *not to hold*, the following condition is sufficient:

$$\hat{p}_N^{\mathrm{B}}(LR_{\mathrm{B}}; \hat{\theta}, \mathscr{C}_\nu(\alpha_1)) > \alpha_2. \tag{6.11}$$

To build a confidence sets for ν, we invert a test (of level α_1) for the specification underlying (2.13) where $\nu = \nu_0$ for known ν_0; this avoids the need to use regularity assumptions on ν. The test we invert is the three-stage MC goodness-of-fit test introduced in Dufour et al. (2003), which uses Mardia's (1970) well known multivariate skewness and kurtosis measures [SK and KU below]:

$$\mathrm{CSK}(\nu_0) = 1 - \min\left\{\hat{p}_N(\mathrm{ESK}(\nu_0); \nu_0), \hat{p}_N(\mathrm{EKU}(\nu_0); \nu_0)\right\} \tag{6.12}$$

$$\mathrm{ESK}(\nu_0) = \left|\mathrm{SK} - \overline{\mathrm{SK}}(\nu_0)\right|, \quad \mathrm{EKU}(\nu_0) = \left|\mathrm{KU} - \overline{\mathrm{KU}}(\nu_0)\right| \tag{6.13}$$

$$\mathrm{SK} = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{i=1}^{T}\hat{d}_{ii}^{3}, \quad \mathrm{KU} = \frac{1}{T}\sum_{t=1}^{T}\hat{d}_{tt}^{2}, \tag{6.14}$$

where $\hat{d}_{it}$ are the elements of the matrix $\hat{U}(\hat{U}'\hat{U}/T)^{-1}\hat{U}'$, $\overline{\mathrm{SK}}(\nu_0)$ and $\overline{\mathrm{KU}}(\nu_0)$ are simulation-based estimates of the expected SK and KU given (2.13) and $\hat{p}[\mathrm{ESK}(\nu_0)]$ and $\hat{p}[\mathrm{EKU}(\nu_0)]$ are p-values, obtained by MC methods under (2.13). The test's three stages [summarized below] are motivated by the following [see Dufour et al. (2003) for proofs and more detailed algorithms].

Let $\bar{d}_{it}(\nu)$ refer to the elements of the matrix $\hat{W}(\hat{W}'\hat{W}/T)^{-1}\hat{W}'$ with $\hat{W} = M(X)W$ and recall that $\hat{U} = M(X)U = M(X)WK = \hat{W}K$. Then under assumption (2.13) [including in particular the distributions from (2.14)-(2.16)], SK and KU depend on the data only via $\bar{d}_{it}(\nu)$ and thus may easily be simulated given draws from the distribution of W. The three stages of the test use this result, as follows.

D1 Stage I, to obtain $\overline{\mathrm{SK}}(\nu_0)$ and $\overline{\mathrm{KU}}(\nu_0)$: for ν_0 given, draw N_0 samples from the hypothesized distribution of W; then, apply $\bar{d}_{it}(.)$ to each draw, replacing $\hat{d}_{it}$ in (6.14) by $\bar{d}_{it}(\nu)$, compute the corresponding measures of skewness and kurtosis, and take the average. The remaining two stages condition on these estimates. Refer to $\overline{\mathrm{SK}}(\nu_0)$ and $\overline{\mathrm{KU}}(\nu_0)$ so obtained as the *reference simulated moments* [**RSM**].

D2 Stage II, to derive $\mathrm{CSK}(\nu_0)$: first obtain $\hat{p}_N\left(\mathrm{ESK}(\nu_0);\nu_0\right)$ and $\hat{p}_N\left(\mathrm{EKU}(\nu_0);\nu_0\right)$, using the observed statistics and N_1 [where $\alpha(N_1+1)$ is an integer] draws from the hypothesized distribution of W (independently from stage I), as in the algorithm A1-A3 from Section 3.2 and using the same RSM for the observed and the simulated values of SK and KU.[9] Refer to the N simulated SK and KU from step A1 as the reference simulated series [**RSS**]. Applying (6.12) gives the observed value of the test statistic, $\mathrm{CSK}(\nu_0)^{(0)}$.

D3. Stage III, to derive a MC p-values for $\mathrm{CSK}(\nu_0)$: independently of the previous RSM and BSS, generate N [where $\alpha(N+1)$ is an integer] additional *i.i.d.* realizations of W. Applying $\bar{d}_{it}(.)$ to each draw, and replacing $\hat{d}_{it}$ in (6.14) by $\bar{d}_{it}(\nu)$ compute the corresponding measures of excess skewness and kurtosis [using the RSM]. On comparing each of these to the RSS, compute MC p-values: in other words, repeat step D2 replacing the observed statistics by those last simulated. These p-values provide a vector $\overline{\mathrm{CSK}}_N(\nu_0)$ of N replications of the combined statistic, leading in turn to $\hat{p}_N(\mathrm{CSK}(\nu_0);\,\nu_0) = p_N\left[\mathrm{CSK}(\nu_0)^{(0)}\middle|\overline{\mathrm{CSK}}_N(\nu_0)\right]$.

The confidence set for ν corresponds to the values of ν_0 which are not rejected at level α_1, using the latter p-value. For the GARCH case, pre-estimating the $2n\times 1$ vector ϕ is infeasible with 5 or even 10 year sub-samples of monthly data. Nevertheless, the single stage MMC is valid despite this limitation. Interestingly, the simulation study we report next suggests that power costs are unimportant even with relatively small samples.

7. Simulation study

We now present a small simulation study to assess the performance of the proposed methods. The design is calibrated to match our empirical analysis (see Section 8) which relies on monthly returns of 12 portfolios of NYSE firms over 1927-1995. We consider model (2.1) where $\tilde{R}_{\mathrm{M}t}$, $t = 1,\ldots,T$, are the returns on the market portfolio from the aforementioned data over the last 5 and 10 year subperiods, as well as the whole sample. We thus take $n = 12$ and $T = 60, 120$ and 828. The coefficients of (2.1) including γ are set to their QML estimates (restricted under $\mathscr{H}_{\mathrm{B}}$ over the conformable sample period). From the QML regression, we also retain the estimated error covariance matrix, to generate model shocks; formally, we compute the corresponding empirical Cholesky factor (denoted $\hat{K}$) and use it for K in (2.13). Test sizes with $K = I_{12}$ are also analyzed to illustrate the effects of portfolio repacking.

We consider normal and Student t-errors (with $\kappa = 8$, in accordance with the kurtosis observed in the empirical application), so the random vectors W_t, $t = 1,\ldots,T$, in (2.13) are generated following (2.14) and (2.15) respectively. The MC tests are applied imposing and ignoring information

[9] With respect to the notation in A1-A3, disregard η since the statistics considered here are invariant to B and K.

on κ, which allows us to document the cost of estimating this parameter. When κ is considered unknown, MMC p-values are calculated over the interval $4 \leq \kappa \leq 13$ to keep execution time manageable (a wider range is allowed for the empirical application in Section 8). We also consider the case of GARCH errors (2.17), with $\phi_{1i} = \phi_1$ and $\phi_{2i} = \phi_2$, $i = 1, \ldots, n$ (the coefficients are the same across equations). This restriction is motivated by execution time, but it is relaxed in Section 8. We use the diagonal elements of $\hat{K}\hat{K}'$ to scale the intercept, yet we also consider the case where $\sigma_i^2 = 1$, $i = 1, \ldots, n$. Samples are simulated with $(\phi_1, \phi_2) = (.15, .80)$. These parameters are treated, in turn, as known and as unknown quantities. In view of the low dimension of the nuisance parameter space in this case, when (ϕ_1, ϕ_2) is treated as unknown, p-value maximization is achieved through a coarse grid search (for the purpose of this simulation). The p-value function does not appear to be very sensitive to the value of (ϕ_1, ϕ_2), and the results presented below indicate this is sufficient for controlling test level in the relevant cases. A more thorough optimization is however used in Section 8.

The results of the simulation are summarized in tables 1 - 3. These tables report empirical rejection rates for various tests of $\mathscr{H}(\gamma_0)$ with nominal size 5%. These rejection rates determine the coverage properties of confidence sets derived from the tests. Since we focus on estimating γ, $\mathscr{H}_{\mathrm{B}}$ is imposed for both the size and power studies. We compare the following tests: (1) a Wald-type test which rejects $\gamma = \gamma_0$ when γ_0 falls outside the Wald-type confidence interval $[\hat{\gamma} - 1.96 \times \mathrm{AsySE}(\hat{\gamma}), \hat{\gamma} + 1.96 \times \mathrm{AsySE}(\hat{\gamma})]$, using the QML estimator $\hat{\gamma}$, an asymptotic standard error [$\mathrm{AsySE}(\hat{\gamma})$, based on (2.27)], and a normal limiting distribution; (2) the MC and MMC tests based on the QLR test statistic $LR(\gamma_0)$ defined in Theorem **3.1**, with MC p-values for *i.i.d.* normal or Student-t errors (with known or unknown κ), Gaussian GARCH with known or unknown (ϕ_1, ϕ_2), as well as $\phi_1 = \phi_2 = 0$ (*i.e.*, ignoring the GARCH dependence even when it is present in the simulated process); (3) tests based on the HAC Wald-type statistic $\mathscr{J}(\gamma_0)$ in (2.28), using a $\chi^2(n)$ critical value, MC with known (ϕ_1, ϕ_2), and MMC where (ϕ_1, ϕ_2) is taken as unknown.

In the size study (Table 1), γ_0 is calibrated to its QML counterpart from the data set [$\gamma_0 = -0.000089$ for $T = 60$, $\gamma_0 = .004960$ for $T = 120$, $\gamma_0 = .005957$ for $T = 828$]. In the power study (tables 2 - 3), we focus on the $\hat{K}$ case; samples are drawn with γ set to its QML estimate, and γ_0 is set to the latter value + $step \times \widehat{\sigma}_i^{\min}$, where $\widehat{\sigma}_i^{\min} = [\min\{\widehat{\sigma}_i^2\}]^{1/2}$, and $\widehat{\sigma}_i^2$ are the diagonal terms of $\hat{K}\hat{K}'$ (with various $step$ values). $N = 99$ is used for MC tests ($N = 999$ is used in the empirical application). In each experiment, the number of simulations is 1000. We use 12 lags for the HAC correction.

Our results can be summarized as follows. The asymptotic *i.i.d.* or robust procedures are very unreliable from the viewpoint of controlling level. Whereas we observe empirical frequencies of type I errors over 70% and sometimes 90% with $T = 60$, we still see empirical rejections near 55% with $T = 828$. The results also show that the empirical size of the HAC-based tests is not affected by K, though a formal proof of its invariance is not available. This observation is however compatible with the fact that its size improves with larger samples: while the level of the Wald-type test shows no improvement (around 55%) even with $T = 828$ and normal errors, the size of the $\mathscr{J}(\gamma_0)$ statistic drops from 95% with $T = 60$ to 12% with $T = 828$. The likelihood-ratio and robust MC and MMC tests achieve level control; in the GARCH case, the MC likelihood-ratio test has the correct size even when GARCH dependence is not accounted for.

Table 1. Tests on zero-beta rate: empirical size

$n = 12$			$T = 60$		$T = 120$		$T = 828$	
			K		K		K	
W	Statistic	p-value	I_{12}	$\hat{K}$	I_{12}	$\hat{K}$	I_{12}	$\hat{K}$
Normal	Wald-type	N[0,1]	.709	.196	.633	.096	.578	.050
	$LR(\gamma_0)$	MC	.057	.057	.048	.048	.041	.041
Student-t	Wald-type	N[0,1]	.714	.218	.645	.106	.587	.055
	$LR(\gamma_0)$	MC: κ known	.053	.053	.046	.046	.043	.043
		MMC: κ unknown	.043	.043	.035	.035	.031	.031
GARCH	Wald-type	N[0,1]	.676	.200	.628	.086	.579	.047
	$LR(\gamma_0)$	MC: $\phi_1 = \phi_2 = 0$	.059	.059	.048	.048	.046	.046
		MC: ϕ_1, ϕ_2 known	.064	.064	.043	.043	.050	.028
		MMC: ϕ_1, ϕ_2 unknown	.054	.054	.032	.032	.028	.050
	$\mathscr{J}(\gamma_0)$	$\chi^2(12)$	.954	.954	.686	.686	.127	.127
		MC: ϕ_1, ϕ_2 known	.049	.049	.045	.045	.049	.040
		MMC: ϕ_1, ϕ_2 unknown	.040	.040	.034	.034	.040	.049

Note – The table reports the empirical rejection rates of various tests for $\mathscr{H}(\gamma_0)$ with nominal level 5%. The values of γ_0 tested are: $\gamma_0 = -0.000089$ for $T = 60$, $\gamma_0 = .004960$ for $T = 120$, $\gamma_0 = .005957$ for $T = 828$. The design is calibrated to match our empirical analysis (see Section 8). The tests compared are the following. (1) A Wald-type test which rejects $\gamma = \gamma_0$ when γ_0 falls outside the Wald-type confidence interval $[\hat{\gamma} - 1.96 \times \text{AsySE}(\hat{\gamma}), \hat{\gamma} + 1.96 \times \text{AsySE}(\hat{\gamma})]$, using the QML estimator $\hat{\gamma}$ with asymptotic standard error $[\text{AsySE}(\hat{\gamma})]$ based on (2.27), and a normal limiting distribution. (2) MC and MMC tests based on $LR(\gamma_0)$ in (2.20), with MC p-values for *i.i.d.* normal and Student-t errors (with known or unknown κ), Gaussian GARCH with known or unknown (ϕ_1, ϕ_2), as well as $\phi_1 = \phi_2 = 0$ (*i.e.*, ignoring the GARCH dependence even when it is present in the simulated process). (3) Tests based on the HAC Wald-type statistic $\mathscr{J}(\gamma_0)$ in (2.28), using a $\chi^2(n)$ critical value, MC with known (ϕ_1, ϕ_2), and MMC where (ϕ_1, ϕ_2) is taken as unknown. In the *i.i.d.* cases, the errors are generated using (2.13) with K set to either I_{12} or $\hat{K}$, which corresponds to the Cholesky factor of the least-squares error covariance estimate from the empirical data used for the simulation design. In the GARCH case, samples are generated with conditional variance as in (2.17) using $\hat{K}$ or I_{12} for K.

Table 2. Tests on zero-beta rate: empirical power
Gaussian and Student designs

$n = 12$			$T = 60$		$T = 120$		$T = 828$	
W	Statistic	p-value	Step	Power	Step	Power	Step	Power
Normal	$LR(\gamma_0)$	MC: $\phi_1 = \phi_2 = 0$	.50	.151	.50	.226	.20	.129
			.75	.315	.75	.529	.30	.313
			1.0	.544	1.0	.835	.50	.814
			2.0	.981	1.5	.999	.75	.998
Student-t	$LR(\gamma_0)$	MC: κ known	.50	.134	.50	.181	.20	.109
		MMC: κ unknown		.126		.158		.080
		MC: κ known	.75	.264	.75	.428	.30	.237
		MMC: κ unknown		.239		.384		.182
		MC: κ known	1.0	.494	1.0	.709	.50	.660
		MMC: κ unknown		.440		.673		.605
		MC: κ known	2.0	.939	1.5	.997	.75	.966
		MMC: κ unknown		.925		.997		.960

Note – The table reports the empirical rejection rates of various tests for $\mathscr{H}(\gamma_0)$ with nominal level 5%. The values of γ_0 tested are: $\gamma_0 = -0.000089$ for $T = 60$, $\gamma_0 = .004960$ for $T = 120$, $\gamma_0 = .005957$ for $T = 828$. The sampling design conforms with the size study, for the $\hat{K}$ case. Samples are drawn with γ calibrated to its QML counterpart from the 1991-95 subsample; values for γ_0 are set to the latter value + $step \times \widehat{\sigma}_i^{\min}$, where $\widehat{\sigma}_i^{\min} = [\min\{\widehat{\sigma}_i^2\}]^{1/2}$, and $\widehat{\sigma}_i^2$ are the diagonal terms of $\hat{K}\hat{K}'$. See Table 1 for further details on the design and tests applied.

In view of the poor size performance of the asymptotic tests, the power study focuses on procedures whose level appears to be under control. Overall, the MMC correction is not too costly from the power viewpoint, with both Student-t and GARCH errors. In the latter case, the likelihood-ratio-type test uncorrected for GARCH effects outperforms all the other tests. When GARCH corrections are performed via MMC, the likelihood-ratio-type test performs generally better than the $\mathscr{J}(\gamma_0)$ test.

8. Empirical analysis

In this section, we assess $\mathscr{H}_{\mathrm{B}}$ as defined in (2.2) in the context of (2.1) under the distributional assumptions (2.15)-(2.16), as well as the Gaussian GARCH in (2.17). We use real monthly returns over the period going from January 1926 to December 1995, obtained from CRSP. The data studied involve 12 portfolios of NYSE firms grouped by standard two-digit industrial classification (SIC). The sectors studied include: (1) petroleum; (2) finance and real estate; (3) consumer durables; (4) basic industries; (5) food and tobacco; (6) construction; (7) capital goods; (8) transportation; (9) utilities; (10) textile and trade; (11) services; (12) leisure; for details on the SIC codes, see Beaulieu et al. (2007). For each month, the industry portfolios include the firms for which the return, the price per common share and the number of shares outstanding are recorded by CRSP. Furthermore, portfolios are value-weighted in each month. We measure the market return by value-weighted NYSE returns, and the real risk-free rate by the one month Treasury bill rate net of inflation, both available from CRSP. All MC tests were applied with $N = 999$ and MMC p-values are obtained using the simulated annealing algorithm.

Our QML-based BCAPM test results are summarized in Table 4. Non-Gaussian p-values are the largest MC p-values over the error distribution parameters [respectively: κ and (π, ω) for (2.15)-(2.16)] within the specified 97.5% confidence sets; the latter are reported in Table 5. In the GARCH case (2.17), p-values are the largest MC p-values over all (ϕ_{1i}, ϕ_{2i}). Given a 5% level test, the benchmark is .05 for p_∞, normal LMC, MMC and GARCH p-values, while the Student-t and normal mixture p-values should be compared to .025 (to ensure that the test has level .05). Non-rejections by LMC MC p-values are conclusive (though rejections are not); rejections based on the conservative bound reported under the heading BND are conclusive under normality; non-rejections based on the QML bound in the non-Gaussian case (reported under the heading BND) signal that the confidence set for γ is not empty; however, since the MMC p-value is based on the tightest bound, this evidence does not necessarily imply non-rejection of $\mathscr{H}_{\mathrm{B}}$.

The empirical results presented in Table 4 show that the asymptotic test and the Gaussian-based bound test yield the same decision at level 5%, although the former p-values are much lower. The non-normal p-values exceed the Gaussian-based p-value, enough to change the test decision. For instance, at the 5% significance level, we find seven rejections of the null hypothesis for the asymptotic $\chi^2(11)$ test, seven for the MC p-values under normality and with normal GARCH, and five (relying on the MMC p-value) under the Student-t and normal mixture distributions.

Focusing on Student-t and normal mixture distributions with parameters not rejected by proper GF tests, we see that mean-variance efficiency test results can change relative to the available F-based bound. The power advantages of the MMC procedure are illustrated by the results of the 1966-70 subperiod where the QML p-value exceeds 2.5% for the Student-t and normal mixture distributions, whereas the MMC p-value signals a rejection.

The confidence sets for distributional parameters are reported in Table 5. In the mixture case, the confidence region is summarized as follows for presentation ease: we give the confidence set for ω corresponding to five different values of π.

In Table 6, we present: (i) the average real market return as well as the average real risk-free rate over each subperiod, (ii) the QML estimate of γ (denoted $\hat{\gamma}$) and 95% confidence sets for this parameter, using respectively the asymptotic standard errors (2.27) (under the heading Wald-type), and the likelihood-ratio-type tests with *i.i.d.* Gaussian, $t(\kappa)$ and normal mixture (π, ω) errors, plus Gaussian GARCH errors (lower panel).[10] The values of γ in the Fieller-type confidence set are not rejected by the test defined in Theorem **3.1** to test $\mathscr{H}(\gamma_0)$. Rejection decisions are based on the largest MC p-values over all κ and (π, ω) respectively; we did not restrict maximization to the confidence set for these parameters here. As expected in view of the $\mathscr{H}_{\mathrm{B}}$ test results, the exact confidence sets are empty for several subperiods. The usefulness of the asymptotic confidence

[10]Note that some values of $\hat{\gamma}$ are high. Nonetheless, comparing the average real market return for those subperiods with our estimate of γ reveal that these high occurrences of γ are consistent with subperiods during which the estimated zero-beta rate is higher than the market portfolio return. This is an illustration of finding, ex post, a linear relationship between risk and return with a negative slope. Furthermore, rerunning our analysis using 10-year subperiods leads to γ estimates below the benchmark average return.

intervals is obviously questionable here. Other results which deserve notice are the empty sets for 1956-60 subperiod; these sets correspond to the case where the efficiency bound test is significant (at 5%).[11]

To illustrate the differences between the asymptotic confidence set and ours, we next check whether the average real risk-free rate is contained in the confidence sets. For many subperiods, like 1966-70, the evidence produced by the asymptotic and MC Fieller-type confidence intervals is similar. There are nonetheless cases where the set estimates do not lead to the same decision. For instance, for 1941-45 and 1971-75, the average risk-free rate is not included in the asymptotic confidence interval, while it is covered by our MMC confidence sets. These are cases where, using the asymptotic confidence interval, the hypothesis $\gamma = r_f$ is rejected, whereas exact confidence sets indicate it should not be rejected. Conversely, in 1986-90, the asymptotic confidence interval includes the average risk-free rate, whereas our confidence sets are empty.

In Table 7, we report the $\mathscr{J}(\gamma_0)$ counterparts of the above QML-based tests (columns 2 and 3) as well as point and set estimates of γ (columns 1 and 4). Column (2) reports the values of our proposed J-test-type minimum $\mathscr{J}(\gamma_0)$ statistic. In column (3), MMC refers to the maximal MC p-value [over all (ϕ_{1i}, ϕ_{2i})] for this statistic, assuming the GARCH specification (2.17), and the level is 5%; alternatively, an asymptotic $\chi^2(12)$ critical value (21.03 for a 5% level) can be used. In column (1), we report the GMM-type point estimate (denoted $\tilde{\gamma}$); the associated set estimate which inverts the $\mathscr{J}(\gamma_0)$ MC Gaussian GARCH based test is reported in column (4).

We first note that, on using the asymptotic critical value, a $\mathscr{J}(\gamma_0)$ test would reject the model in all subperiods at level 5%. In contrast, the GARCH-MMC p-value is less than 5% only in the 1986-90 subperiod. In view of our simulation results from Section 7, these results illustrate the serious implications of asymptotic test size distortions. Recall that the likelihood-ratio-based MC and MMC (Gaussian and non-Gaussian, with and without GARCH) tests reject the model at the 5% level in at least three other sub-periods: 1946-50, 1950-55, 1960-65. This reflects the test relative power, as illustrated in Section 7. Turning to the estimates of γ, we note that the $\mathscr{J}(\gamma_0)$ based MMC confidence sets are substantially wider than the likelihood-ratio-based counterparts, only one confidence set is empty (in the 1986-1990 subperiods, in which case the model would be rejected), and the set is unbounded in the 1990-95 subperiod. Had we relied on the asymptotic $\chi^2(12)$ cut-off to invert the $\mathscr{J}(\gamma_0)$ test, all confidence sets would be empty. Again, these observations line up with our simulation results.

The above procedures applied to the full data yields empty confidence sets using the exact GARCH corrected likelihood-ratio and $\mathscr{J}(\gamma_0)$ criteria; the confidence interval using (2.27) is $[.0007, .0088]$. Since our subperiod analysis suggests that γ is temporally unstable, one must be careful in interpreting such results. On using a Bonferroni argument (that accounts for time-varying parameters) based on the minimum (over subperiods) GARCH-corrected p-value which is $.003 < .05/12$, the model can be safely rejected at level 5%, over the full sample.

[11] This can be checked by referring to Table 4: although the reported maximal p-values in this table are performed over the confidence set for κ and (π, ω), we have checked that the global maximal p-value leads to the same decision here.

9. Conclusion

This paper proposes exact mean-variance efficiency tests when the zero-beta (or risk-free) rate is not observable, which raises identification difficulties. Proposed methods are robust to this problem as well as to portfolio repacking, and allow for heavy-tailed return distributions. In particular, a useful invariance result for MacKinlay and Richardson's (1991) HAC statistic is shown analytically. We also derive exact confidence sets for the zero-beta rate γ. While available Wald-type intervals are unreliable and lead to substantially different inference concerning γ, our confidence sets are valid in finite samples without assuming identification, and are empty by construction if efficiency is rejected.

We report a simulation study which illustrates the properties of our proposed procedures. Our results allow to disentangle small-sample problems from asymptotic failures arising from weak identification. We also examine efficiency of the market portfolio for monthly returns on NYSE CRSP portfolios. We find that efficiency is less rejected with non-normal assumptions. Exact confidence sets for γ differ importantly from asymptotic ones, and likelihood-ratio-based confidence sets are tighter than their Wald counterparts. All confidence sets nevertheless suggest that γ is not stable over time.

These results provide the motivation to extend our method to more general factor models, as discussed by Campbell et al. (1997, Chapter 6) and Shanken and Zhou (2007). These models raise the same statistical issues as the BCAPM, except that their definitional parameter is non-scalar. In this case, Fieller-type methods are clearly more challenging and raise worthy theoretical and empirical research questions.

A. Appendix: Proofs

PROOF OF THEOREM **3.1** Under (2.13) and $\mathscr{H}(\gamma_0)$, we have: $T\hat{\Sigma} = \hat{U}'\hat{U} = K^{'}W'M(X)WK$, $T\hat{\Sigma}(\gamma_0) = K^{'}W'\bar{M}(\gamma_0)WK$. Then, under $\mathscr{H}(\gamma_0)$,

$$\Lambda(\gamma_0) = \frac{|\hat{\Sigma}(\gamma_0)|}{|\hat{\Sigma}|} = \frac{|K^{'}W'\bar{M}(\gamma_0)WK|}{|K^{'}W'M(X)WK|} = \frac{|K^{'}|\,|W'\bar{M}(\gamma_0)W|\,|K|}{|K^{'}|\,|W'M(X)W|\,|K|} = \frac{|W'\bar{M}(\gamma_0)W|}{|W'M(X)W|},$$

hence $\mathsf{P}[LR(\gamma_0) \geq x] = \mathsf{P}[T\ln(|W'\bar{M}(\gamma_0)W|\,/\,|W'M(X)W|) \geq x]$, $\forall x$. □

PROOF OF THEOREM **3.2** For this model,

$$\hat{U} = M(X)U = M(X)WK = \hat{W}K, \quad \hat{W} = M(X)W\,.$$

If we denote by $\hat{U}_t^{'}$ and $\hat{W}_t^{'}$ the t-th rows of $\hat{U}$ and $\hat{W}$, respectively, this means that $\hat{U}_t = K'\hat{W}_t$, for all t. Then,

$$\begin{aligned}\left(X_t \otimes \hat{U}_t\right)\left(X_{t-j} \otimes \hat{U}_{t-j}\right)' &= \left(X_t \otimes K'\hat{W}_t\right)\left(X_{t-j} \otimes K'\hat{W}_{t-j}\right)' = \left(X_t \otimes K'\hat{W}_t\right)\left(X_{t-j}' \otimes \hat{W}_{t-j}^{'}K\right)\\ &= X_tX_{t-j}' \otimes K'\hat{W}_t\hat{W}_{t-j}^{'}K = \left(I_k \otimes K'\right)\left(X_t \otimes \hat{W}_t\right)\left(X_{t-j} \otimes \hat{W}_{t-j}\right)'\left(I_k \otimes K\right)\end{aligned}$$

for $j \geq 0, \ldots$. Setting

$$\Psi_{j,T} = \frac{1}{T}\sum_{t=j+1}^{T}\left(X_t \otimes \hat{U}_t\right)\left(X_{t-j} \otimes \hat{U}_{t-j}\right)', \; \mathsf{S}_T = \Psi_{0,t} + \sum_{j=1}^{q}\left(\frac{q-j}{q}\right)\left[\Psi_{j,T} + \Psi_{j,T}'\right],$$

we see that

$$\Psi_{j,T} = \left(I_k \otimes K'\right)\hat{\Psi}_{j,T}\left(I_k \otimes K\right), \quad \hat{\Psi}_{j,T} = \frac{1}{T}\sum_{t=j+1}^{T}\left(X_t \otimes \hat{W}_t\right)\left(X_{t-j} \otimes \hat{W}_{t-j}\right)',$$

$$\Psi_{j,T} + \Psi_{j,T}' = \left(I_k \otimes K'\right)\left[\hat{\Psi}_{j,T} + \hat{\Psi}_{j,T}'\right]\left(I_k \otimes K\right),$$

$$\mathsf{S}_T = \left(I_k \otimes K'\right)\hat{\Psi}_{0,T}\left(I_k \otimes K\right) + \sum_{j=1}^{q}\left(\frac{q-j}{q}\right)\left(I_k \otimes K'\right)\left[\hat{\Psi}_{j,T} + \hat{\Psi}_{j,T}'\right]\left(I_k \otimes K\right) = \left(I_k \otimes K'\right)\hat{\mathsf{S}}_T\left(I_k \otimes K\right),$$

$$\hat{\mathsf{S}}_T = \hat{\Psi}_{0,T} + \sum_{j=1}^{q}\left(\frac{q-j}{q}\right)\left[\hat{\Psi}_{j,T} + \hat{\Psi}_{j,T}'\right].$$

On replacing S_T with $(I_k \otimes K')\,\hat{\mathsf{S}}_T\,(I_k \otimes K)$, we get:

$$\left[\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right]\mathsf{S}_T\left[\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right] = \left[\left(\frac{X'X}{T}\right)^{-1} \otimes K'\right]\hat{\mathsf{S}}_T\left[\left(\frac{X'X}{T}\right)^{-1} \otimes K\right].$$

Furthermore, since $\mathsf{R}(\gamma_0) = (1,\gamma_0) \otimes I_n = H(\gamma_0) \otimes I_n$ and $H(\gamma_0)'$ is a k-dimensional vector in $\tilde{\mathscr{H}}(\gamma_0) : \mathsf{R}(\gamma_0)\vartheta = C'H(\gamma_0)' = 0$, we see that:

$$\mathsf{R}(\gamma_0)\hat{\vartheta} = \hat{C}'H(\gamma_0)',$$

$$\mathsf{R}(\gamma_0)\left[\left(\frac{X'X}{T}\right)^{-1} \otimes K'\right] = \left(H(\gamma_0) \otimes I_n\right)\left[\left(\frac{X'X}{T}\right)^{-1} \otimes K'\right] = H(\gamma_0)\left(\frac{X'X}{T}\right)^{-1} \otimes K',$$

so $\mathscr{J}(\gamma_0)$ can be rewritten as

$$\begin{aligned}
\mathscr{J}(\gamma_0) &= T\,\hat{\vartheta}'\mathsf{R}(\gamma_0)'\left\{\mathsf{R}(\gamma_0)\left[\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right]\mathsf{S}_T\left[\left(\frac{X'X}{T}\right)^{-1} \otimes I_n\right]\mathsf{R}(\gamma_0)'\right\}^{-1}\mathsf{R}(\gamma_0)\hat{\vartheta} \\
&= T\,\hat{\vartheta}'\mathsf{R}(\gamma_0)'\left\{\mathsf{R}(\gamma_0)\left[\left(\frac{X'X}{T}\right)^{-1} \otimes K'\right]\hat{\mathsf{S}}_T\left[\left(\frac{X'X}{T}\right)^{-1} \otimes K\right]\mathsf{R}(\gamma_0)'\right\}^{-1}\mathsf{R}(\gamma_0)\hat{\vartheta} \\
&= T\,H(\gamma_0)\hat{C}\mathscr{Q}(\gamma_0)^{-1}\hat{C}'H(\gamma_0)' = T\,H(\gamma_0)\hat{C}K^{-1}K\mathscr{Q}(\gamma_0)^{-1}K'\left(K'\right)^{-1}\hat{C}'H(\gamma_0)' \\
&= T\,H(\gamma_0)\hat{C}K^{-1}\bar{\mathscr{Q}}(\gamma_0,W)^{-1}\left(K'\right)^{-1}\hat{C}'H(\gamma_0)'
\end{aligned}$$

where

$$\mathscr{Q}(\gamma_0)=\left[H(\gamma_0)\left(\frac{X'X}{T}\right)^{-1}\otimes K'\right]\hat{\mathsf{S}}_T\left[\left(\frac{X'X}{T}\right)^{-1}H(\gamma_0)'\otimes K\right],\ \bar{\mathscr{Q}}(\gamma_0,W)=\left(K'\right)^{-1}\mathscr{Q}(\gamma_0)K^{-1}.$$

Under the null hypothesis $H(\gamma_0)C=0$, we have:

$$H(\gamma_0)\hat{C}=H(\gamma_0)(X'X)^{-1}X'\left[XC+U\right]=H(\gamma_0)C+H(\gamma_0)(X'X)^{-1}X'U=H(\gamma_0)(X'X)^{-1}X'WK,$$

$$\begin{aligned}\bar{\mathscr{Q}}(\gamma_0,W) &= \left(K'\right)^{-1}\left[H(\gamma_0)\left(\frac{X'X}{T}\right)^{-1}\otimes K'\right]\hat{\mathsf{S}}_T\left[\left(\frac{X'X}{T}\right)^{-1}H(\gamma_0)'\otimes K\right]K^{-1}\\ &= \left[H(\gamma_0)\left(\frac{X'X}{T}\right)^{-1}\otimes\left(K'\right)^{-1}K'\right]\hat{\mathsf{S}}_T\left[\left(\frac{X'X}{T}\right)^{-1}H(\gamma_0)'\otimes KK^{-1}\right]\\ &= \left[H(\gamma_0)\left(\frac{X'X}{T}\right)^{-1}\otimes I_n\right]\hat{\mathsf{S}}_T\left[\left(\frac{X'X}{T}\right)^{-1}H(\gamma_0)'\otimes I_n\right]\end{aligned}$$

where we take into account the fact that $H(\gamma_0)'$ is a column vector, hence

$$\begin{aligned}\mathscr{J}(\gamma_0) &= T\,H(\gamma_0)(X'X)^{-1}X'WKK^{-1}\bar{\mathscr{Q}}(\gamma_0,W)^{-1}\left(K'\right)^{-1}K'W'X(X'X)^{-1}H(\gamma_0)'\\ &= T\,H(\gamma_0)(X'X)^{-1}X'W\bar{\mathscr{Q}}(\gamma_0,W)^{-1}W'X(X'X)^{-1}H(\gamma_0)'.\end{aligned}$$

□

PROOF OF LEMMA **5.1** The Gaussian log-likelihood function for model (2.5) is

$$\ln[\tilde{L}(\tilde{Y},C,\Sigma)]=-\frac{T}{2}[n(2\pi)+\ln(|\Sigma|)]-\frac{1}{2}\mathrm{tr}[\Sigma^{-1}(\tilde{Y}-XC)'(\tilde{Y}-XC)]=\ln[L(Y,B,\Sigma)].$$

Setting $\tilde{\Sigma}(C)\equiv\frac{1}{T}(\tilde{Y}-XC)'(\tilde{Y}-XC)$, for any given value of C, $\ln[\tilde{L}(\tilde{Y},C,\Sigma)]$ is maximized by taking $\Sigma=\tilde{\Sigma}(C)$ yielding the concentrated log-likelihood

$$\ln[\tilde{L}(\tilde{Y},C,\Sigma)_c=-\frac{nT}{2}[(2\pi)+1]-\frac{T}{2}\ln(|\tilde{\Sigma}(C)|).\tag{A.1}$$

The Gaussian MLE of C thus minimizes $|\tilde{\Sigma}(C)|$ with respect to C. Let us denote by $\hat{C}(Y)$ the unrestricted MLE of C so obtained, and by $\hat{C}(Y;\gamma_0)$ and $\hat{C}_{\mathrm{B}}(Y)$ the restricted estimators subject to $\tilde{\mathscr{H}}(\gamma_0)$ and $\tilde{\mathscr{H}}_{\mathrm{B}}$ respectively. Suppose that $\tilde{Y}$ is replaced by $\tilde{Y}_*=\tilde{Y}A$ where A is a nonsingular $n\times n$ matrix. We need to show that $LR_*(\gamma_0)=LR(\gamma_0)$ and $LR_{\mathrm{B}*}=LR_{\mathrm{B}}$, where $LR_*(\gamma_0)$ and $LR_{\mathrm{B}*}$ represent the corresponding test statistics based on the transformed data. Following this transformation,

$|\tilde{\Sigma}(C)|$ becomes:

$$\begin{aligned}|\tilde{\Sigma}_*(C_*)| &= \Big|\frac{1}{T}(\tilde{Y}_* - XC_*)'(\tilde{Y}_* - XC_*)\Big| = \Big|\frac{1}{T}A'(\tilde{Y} - XC_*A^{-1})'(\tilde{Y} - XC_*A^{-1})A\Big| \\ &= |A'A|\,\Big|\frac{1}{T}(\tilde{Y} - XC)'(\tilde{Y} - XC)\Big| = |A'A|\,|\tilde{\Sigma}(C)| \end{aligned} \tag{A.2}$$

where $C = C_*A^{-1}$. Then $|\tilde{\Sigma}(C_*)|$ is minimized by $\hat{C}_*(Y_*) = \hat{C}(Y)A$ and $|\tilde{\Sigma}_*\big(\hat{C}_*(Y_*)\big)| = |A'A||\tilde{\Sigma}\big(\hat{C}(Y)\big)$. On observing that $H(\gamma_0)C = 0 \Longleftrightarrow H(\gamma_0)CA = 0 \Longleftrightarrow H(\gamma_0)C_* = 0$ for any γ_0, the restricted estimators of C under $\tilde{\mathscr{H}}(\gamma_0)$ and $\mathscr{H}_{\mathrm{B}}$ are transformed in the same way: $\hat{C}_*(Y_*;\gamma_0) = \hat{C}(Y;\gamma_0)A$ and $\hat{C}_{*\mathrm{B}}(Y_*) = \hat{C}_{\mathrm{B}}(Y)A$. This entails that $|\tilde{\Sigma}_*\big(\hat{C}_*(Y_*;\gamma_0)\big)| = |A'A|\,|\tilde{\Sigma}\big(\hat{C}(Y;\gamma_0)\big)|$ and $|\tilde{\Sigma}_*\big(\hat{C}_{*\mathrm{B}}(Y_*)\big)| = |A'A|\,|\tilde{\Sigma}\big(\hat{C}_{\mathrm{B}}(Y)\big)|$, so that

$$\tilde{\Lambda}_*(\gamma_0) = \frac{|\tilde{\Sigma}_*\big(\hat{C}_*(Y_*;\gamma_0)\big)|}{|\tilde{\Sigma}_*\big(\hat{C}_*(Y_*)\big)|} = \frac{|\tilde{\Sigma}\big(\hat{C}(Y;\gamma_0)\big)|}{|\tilde{\Sigma}\big(\hat{C}(Y)\big)|} = \tilde{\Lambda}(\gamma_0)\,, \tag{A.3}$$

$$\tilde{\Lambda}_{\mathrm{B}*} = \frac{|\tilde{\Sigma}_*\big(\hat{C}_{*\mathrm{B}}(Y_*)\big)|}{|\tilde{\Sigma}_*\big(\hat{C}_*(Y_*)\big)|} = \frac{|\tilde{\Sigma}\big(\hat{C}_{\mathrm{B}}(Y)\big)|}{|\tilde{\Sigma}\big(\hat{C}(Y)\big)|} = \tilde{\Lambda}_{\mathrm{B}}\,. \tag{A.4}$$

Finally, in view of (2.20) and (2.26), we have $LR_*(\gamma_0) = T\ln[\tilde{\Lambda}_*(\gamma_0)] = T\ln[\tilde{\Lambda}(\gamma_0)] = LR(\gamma_0)$ and $LR_{\mathrm{B}*} = T\ln(\tilde{\Lambda}_{\mathrm{B}*}) = T\ln(\tilde{\Lambda}) = LR_{\mathrm{B}}$. □

PROOF OF THEOREM **5.2** Consider a transformation of the form $\tilde{Y}_* = \tilde{Y}K^{-1}$ or, equivalently, $Y_* = YK^{-1} + \tilde{R}_{\mathrm{M}}\iota_n'(I - K^{-1})$. Using (2.1) and (2.13), we then have:

$$\begin{aligned} Y_* &= (XB + WK)K^{-1} + \tilde{R}_{\mathrm{M}}\iota_n'(I - K^{-1}) = XBK^{-1} + \tilde{R}_{\mathrm{M}}\iota_n'(I - K^{-1}) + W \\ &= (\iota_T a' + \tilde{R}_{\mathrm{M}}\beta')K^{-1} + \tilde{R}_{\mathrm{M}}\iota_n'(I - K^{-1}) + W \\ &= \tilde{R}_{\mathrm{M}}\iota_n' + [\iota_T a' + \tilde{R}_{\mathrm{M}}(\beta - \iota_n)']K^{-1} + W \\ &= \tilde{R}_{\mathrm{M}}\iota_n' + X(B - \Delta)K^{-1} + W = \tilde{R}_{\mathrm{M}}\iota_n' + X\bar{B} + W \end{aligned} \tag{A.5}$$

where $\bar{B} = (B - \Delta)K^{-1}$ and $\Delta = [0,\, \iota_n]'$. Using Lemma **5.1**, $LR(\gamma_0)$ and LR_{B} can be viewed as functions of Y_*, and depend on (B, K) only through $\bar{B} = (B - \Delta)K^{-1}$. Under $\mathscr{H}_{\mathrm{B}}$, the nuisance parameter only involves γ and $(\beta - \iota_n)'K^{-1}$. Now the distribution of $LR(\gamma_0)$ and LR_{B} can be explicitly characterized by using (A.3) - (A.4) and observing that

$$\begin{aligned} \tilde{\Lambda}(\gamma_0) &= \frac{|\tilde{\Sigma}_*\big(\hat{C}_*(Y_*;\gamma_0)\big)|}{|\tilde{\Sigma}_*\big(\hat{C}_*(Y_*)\big)|} = \frac{|\hat{W}(\gamma_0)'\hat{W}(\gamma_0)|}{|\hat{W}'\hat{W}|}\,, \\ \tilde{\Lambda}_{\mathrm{B}} &= \frac{|\tilde{\Sigma}_*\big(\hat{C}_{*\mathrm{B}}(Y_*)\big)|}{|\tilde{\Sigma}_*\big(\hat{C}_*(Y_*)\big)|} = \frac{\inf\{|\tilde{\Sigma}_*\big(\hat{C}_*(Y_*;\gamma_0)\big)| : \gamma_0 \in \Gamma\}}{|\tilde{\Sigma}_*\big(\hat{C}_*(Y_*)\big)|} = \inf\{\tilde{\Lambda}(\gamma_0) : \gamma_0 \in \Gamma\}\,, \end{aligned}$$

where $\hat{W}(\gamma_0) = \bar{M}(\gamma_0)(Y_* - \tilde{R}_{\mathrm{M}}\iota_n') = \bar{M}(\gamma_0)(X\bar{B} + W) = \bar{M}(\gamma_0)\{[\iota_T a' + \tilde{R}_{\mathrm{M}}(\beta - \iota_n)']K^{-1} + W\} = \bar{M}(\gamma_0)\{\big[\iota_T\big(a' + \gamma_0(\beta - \iota_n)'\big) + (\tilde{R}_{\mathrm{M}} - \gamma_0\iota_T)(\beta - \iota_n)'\big]K^{-1} + W\} = \bar{M}(\gamma_0)\{\iota_T\big(a + \gamma_0(\beta - \iota_n)'\big)K^{-1} + W\}$ and $\hat{W} = M(X)W$. Under $\mathscr{H}_{\mathrm{B}}$ where $a = -\gamma(\beta - \iota_n)$, $\hat{W}(\gamma_0) = (\gamma_0 - \gamma)\bar{M}(\gamma_0)\iota_T(\beta -$

$\iota_n)'K^{-1} + \bar{M}(\gamma_0)W$. The theorem then follows on observing that $LR(\gamma_0) = T\,\ln[\tilde{\Lambda}(\gamma_0)]$ and $LR_{\mathrm{B}} = T\,\ln(\tilde{\Lambda}_{\mathrm{B}})$. Further information can be drawn from the singular value decomposition of $\bar{B}$. Let r be the rank of $\bar{B}$. Since $\bar{B}$ is a $2 \times n$ matrix, we have $0 \leq r \leq 2$ and we can write:

$$\bar{B} = PDQ', \quad D = [\bar{D}, 0], \quad \bar{D} = \mathrm{diag}\left(\lambda_1^{1/2}, \lambda_2^{1/2}\right), \tag{A.6}$$

where D is a $2 \times n$ matrix, λ_1 and λ_2 are the two largest eigenvalues of $\bar{B}'\bar{B}$ (where $\lambda_1 \geq \lambda_2 \geq 0$), $Q = [Q_1, Q_2]$ is an orthogonal $n \times n$ matrix whose columns are eigenvectors of $\bar{B}'\bar{B}$, Q_1 is a $2 \times r$ matrix which contains eigenvectors associated with the non-zero eigenvalues of $\bar{B}'\bar{B}$, $P = [P_1, P_2]$ is a 2×2 orthogonal matrix such that $P_1 = \bar{B}Q_1 D_1^{-1}$ and D_1 is a diagonal matrix which contains the non-zero eigenvalues of $\bar{B}'\bar{B}$, setting $P = P_1$ and $D_1 = \bar{D}$ if $r = 2$, and $P = P_2$ if $r = 0$; see Harville (1997, Theorem 21.12.1). Using Lemma **5.1** and Theorem **5.2**, $LR(\gamma_0)$ and LR_{B} may then be reexpressed as:

$$LR(\gamma_0) = T\,\ln\left(|\tilde{W}(\gamma_0)'\tilde{W}(\gamma_0)|/|\tilde{W}'\tilde{W}|\right), \quad LR_{\mathrm{B}} = \inf\{LR(\gamma_0) : \gamma_0 \in \Gamma\}, \tag{A.7}$$

$$\tilde{W} = \hat{W}Q = M(X)\bar{W}, \quad \bar{W} = WQ, \quad \tilde{W}(\gamma_0) = \hat{W}(\gamma_0)Q = \bar{M}(\gamma_0)(XPD + \bar{W}), \tag{A.8}$$

$PD = [P\bar{D},\, 0]$ and $P\bar{D}$ has at most 3 free coefficients (P is orthogonal). Under $\mathscr{H}_{\mathrm{B}}$,

$$\tilde{W}(\gamma_0) = \bar{M}(\gamma_0)\iota_T\left[(\gamma_0 - \gamma)\left(\varphi'\varphi\right)^{1/2}\bar{\varphi}'\right] + \bar{M}(\gamma_0)\bar{W},$$
$$\varphi = Q'\left(K^{-1}\right)'(\beta - \iota_n), \quad \bar{\varphi} = \varphi/\left(\varphi'\varphi\right)^{1/2}.$$

Define $\Phi = \left[\bar{\varphi}, \bar{\Phi}\right]$ as an orthogonal matrix such that $\Phi'\Phi = \Phi\Phi' = I_n$, so

$$\Phi'\Phi = \begin{bmatrix} \bar{\varphi}'\bar{\varphi} & \bar{\varphi}'\bar{\Phi} \\ \bar{\Phi}'\bar{\varphi} & \bar{\Phi}'\bar{\Phi} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & I_{n-1} \end{bmatrix}, \qquad \bar{\varphi}'\Phi = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}. \tag{A.9}$$

Then as in (A.7), $LR(\gamma_0)$ and LR_{B} may again be expressed under $\mathscr{H}_{\mathrm{B}}$ as:

$$LR(\gamma_0) = T\,\ln\left(|\tilde{W}_B(\gamma_0)'\tilde{W}_B(\gamma_0)|/|\tilde{W}_B'\tilde{W}_B|\right), \quad LR_{\mathrm{B}} = \inf\{LR(\gamma_0) : \gamma_0 \in \Gamma\}, \tag{A.10}$$

$$\tilde{W}_B = \tilde{W}\bar{\Phi} = M(X)\bar{W}_B, \quad \bar{W}_B = \bar{W}\Phi, \tag{A.11}$$

$$\tilde{W}_B(\gamma_0) = \tilde{W}(\gamma_0)\bar{\Phi} = \bar{M}(\gamma_0)\iota_T\varphi_B' + \bar{M}(\gamma_0)\bar{W}_B, \tag{A.12}$$

where $\varphi_B' = (\gamma_0 - \gamma)\left(\varphi'\varphi\right)^{1/2}\bar{\varphi}'\Phi = (\gamma_0 - \gamma)\left(\varphi'\varphi\right)^{1/2}\begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$ which involves at most one free coefficient. When W is non-Gaussian, the distributions of $LR(\gamma_0)$ and LR_{B} may be influenced by $\bar{B}$ through Q in $\bar{W}$. Under the Gaussian assumption (2.14), the rows of $\bar{W}$ are *i.i.d.* $\mathrm{N}(0, I_n)$, so that $LR(\gamma_0)$ and LR_{B} follow distributions which depend on (B, K) only through $P\bar{D}$. Under $\mathscr{H}_{\mathrm{B}}$, since the rows of $\bar{W}_B$ are *i.i.d.* $\mathrm{N}(0, I_n)$, this distribution involves only one nuisance parameter, in accordance with the result from Zhou (1991, Theorem 1), derived through a different method. □

PROOF OF THEOREM **6.1** $\mathscr{H}_{\mathrm{B}} = \cup_{\gamma_0}\mathscr{H}(\gamma_0)$. Since $LR_{\mathrm{B}} = \inf\{LR(\gamma_0) : \gamma_0 \in \Gamma\}$, we have $LR_{\mathrm{B}} \leq LR(\gamma_0)$, for any γ_0, hence $\mathsf{P}[LR_{\mathrm{B}} \geq x] \leq \mathsf{P}_{(B,K)}\left[LR(\gamma_0) \geq x\right]$, $\forall x$, for each γ_0 and for any (B, K)

compatible with $\mathscr{H}(\gamma_0)$. Furthermore, under $\mathscr{H}_{\mathrm{B}}$, there is a value of γ_0 such that the distribution of $LR(\gamma_0)$ is given by Theorem **3.1**, which entails (6.1). The result for the Gaussian special case then follows upon using (3.2). $\square$

PROOF OF THEOREM **6.2** The result follows from (6.7), (3.12), and the inequalities $\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}};\gamma,\nu) \leq \hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}};\Gamma,\nu)$ and $\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}};\gamma,\nu) \leq \hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}};\gamma,\Omega_{\mathscr{D}}) \leq \hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}};\Gamma,\Omega_{\mathscr{D}})$. $\square$

PROOF OF THEOREM **6.3** When ν is specified, by (6.6), (2.25) and (3.11), we have: $\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}};\hat{\gamma},\nu) \equiv p_N[LR_{\mathrm{B}}^{(0)}\big|\overline{LR}_N(\hat{\gamma},\nu)] = p_N[LR^{(0)}(\hat{\gamma})\big|\overline{LR}_N(\hat{\gamma},\nu)] = \hat{p}_N(LR;\hat{\gamma},\nu)$, hence $\sup\{\hat{p}_N(LR;\gamma_0,\nu) : \gamma_0 \in \Gamma\} \leq \alpha \Rightarrow \hat{p}_N(LR;\hat{\gamma},\nu) \leq \alpha \Rightarrow \hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}};\hat{\gamma},\nu) \leq \alpha$; on noting that $\sup\{\hat{p}_N(LR;\gamma_0,\nu) : \gamma_0 \in \Gamma\} \leq \alpha$ means that $C_\gamma^{LR}(\alpha,\nu)$ is empty, $\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}};\hat{\gamma},\nu) > \alpha \Rightarrow \sup\{\hat{p}_N(LR;\gamma_0,\nu) : \gamma_0 \in \Gamma\} > \alpha \Rightarrow C_\gamma^{LR}(\alpha,\nu) \neq \emptyset$. For ν unknown,

$$\begin{aligned}
\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}};\hat{\gamma},\Omega_{\mathscr{D}}) &= \sup\{\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}};\hat{\gamma},\nu_0) : \nu_0 \in \Omega_{\mathscr{D}}\} = \sup\{p_N[LR_{\mathrm{B}}^{(0)}\big|\overline{LR}_N(\hat{\gamma},\nu_0)] : \nu_0 \in \Omega_{\mathscr{D}}\}, \\
&= \sup\{p_N[LR^{(0)}(\hat{\gamma})\big|\overline{LR}_N(\hat{\gamma},\nu_0)] : \nu_0 \in \Omega_{\mathscr{D}}\} = \sup\{\hat{p}_N(LR;\hat{\gamma},\nu) : \nu_0 \in \Omega_{\mathscr{D}}\},
\end{aligned}$$

hence $\sup\{\hat{p}_N(LR;\gamma_0,\nu_0) : \gamma_0 \in \Gamma, \nu_0 \in \Omega_{\mathscr{D}}\} \leq \alpha \Rightarrow \sup\{\hat{p}_N(LR;\hat{\gamma},\nu_0) : \nu_0 \in \Omega_{\mathscr{D}}\} \leq \alpha \Rightarrow \hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}};\hat{\gamma},\Omega_{\mathscr{D}}) \leq \alpha$ and $\hat{p}_N^{\mathrm{U}}(LR_{\mathrm{B}};\hat{\gamma},\Omega_{\mathscr{D}}) > \alpha \Rightarrow \sup\{\hat{p}_N(LR;\gamma_0,\nu_0) : \gamma_0 \in \Gamma,\ \nu_0 \in \Omega_{\mathscr{D}}\} > \alpha \Rightarrow C_\gamma^{LR}(\alpha;\mathscr{D}) \neq \emptyset$. $\square$

References

Amsler, C. E. and Schmidt, P. (1985), 'A Monte Carlo investigation of the accuracy of multivariate CAPM tests', *Journal of Financial Economics* **14**, 359–375.

Andrews, D. W. K., Moreira, M. J. and Stock, J. H. (2006), 'Optimal two-sided invariant similar tests for instrumental variables regression', *Econometrica* **74**, 715–752.

Banz, R. W. (1981), 'The relationship between return and market value of common stocks', *Journal of Financial Economics* **9**, 3–18.

Barone-Adesi, G., Gagliardini, P. and Urga, G. (2004), 'Testing asset pricing models with coskewness', *Journal of Business and Economic Statistics* **22**, 474–485.

Bauwens, L., Laurent, S. and Rombouts, J. V. K. (2006), 'Multivariate GARCH models: a survey', *Journal of Applied Econometrics* **21**, 79–109.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2005), Exact multivariate tests of asset pricing models with stable asymmetric distributions, *in* M. Breton and H. Ben Ameur, eds, 'Numerical Methods in Finance', Kluwer, Dordrecht, The Netherlands.

Beaulieu, M.-C., Dufour, J.-M. and 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. and Khalaf, L. (2009), 'Finite sample multivariate tests of asset pricing models with coskewness', *Computational Statistics and Data Analysis* **53**, 2008–2021.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2010), 'Asset-pricing anomalies and spanning: Multivariate multifactor tests with heavy tailed distributions', *Journal of Empirical Finance* **17**, 763–782.

Black, F. (1972), 'Capital market equilibrium with restricted borrowing', *Journal of Business* **45**, 444–454.

Black, F., Jensen, M. C. and Scholes, M. (1972), The capital asset pricing model: Some empirical tests, *in* M. C. Jensen, ed., 'Studies in the Theory of Capital Markets', Praeger.

Bolduc, D., Khalaf, L. and Yelou, C. (2010), 'Identification robust confidence sets methods for inference on parameter ratios with application to discrete choice models', *Journal of Econometrics* **157**, 317–327.

Brennan, M. (1971), 'Capital market equilibrium with divergent borrowing and lending rates', *Journal of Financial Economics* **9**, 3–18.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton University Press, New Jersey.

Chen, N.-F., Roll, R. and Ross, S. A. (1986), 'Economic forces and the stock market', *The Journal of Business* **59**, 383–403.

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

Dufour, J.-M. (1989), 'Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions', *Econometrica* **57**, 335–355.

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

Dufour, J.-M. (2003), 'Identification, weak instruments and statistical inference in econometrics', *Canadian Journal of Economics* **36**(4), 767–808.

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* **133**, 443–478.

Dufour, J.-M. and Jasiak, J. (2001), 'Finite sample limited information inference methods for structural equations and models with generated regressors', *International Economic Review* **42**, 815–843.

Dufour, J.-M. and 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. and 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* **65**, 891–906.

Dufour, J.-M. and Taamouti, M. (2005), 'Projection-based statistical inference in linear structural models with possibly weak instruments', *Econometrica* **73**, 1351–1365.

Dufour, J.-M. and Taamouti, M. (2007), 'Further results on projection-based inference in IV regressions with weak, collinear or missing instruments', *Journal of Econometrics* **139**, 133–153.

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

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

Fama, E. F. and French, K. R. (2004), 'The capital asset pricing model: Theory and evidence', *Journal of Economic Perspectives* **18**, 25–46.

Fama, E. F. and MacBeth, J. (1973), 'Risk, return, and equilibrium: Empirical tests', *Journal of Political Economy* **71**, 607–636.

Fieller, E. C. (1954), 'Some problems in interval estimation', *Journal of the Royal Statistical Society, Series B* **16(2)**, 175–185.

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

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

Gouriéroux, C. and Monfort, A. (1995), *Statistics and Econometric Models, Volumes One and Two*, Cambridge University Press, Cambridge, U.K.

Gouriéroux, C., Monfort, A. and Trognon, A. (1985), 'Moindres carré asymptotiques', *Annales de l'INSEE* **58**, 91–122.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag, New York.

Jensen, M. (1968), 'The performance of mutual funds in the period 1945-64', *Journal of Finance* **23**, 389–416.

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

Kan, R., Robotti, C. and Shanken, J. (2012), ‘Pricing model performance and the two-pass cross-sectional regression methodology’, *Jounal of Finance* **forthcoming**.

Kandel, S. (1984), ‘The likelihood ratio test statistics of mean-variance efficiency without a riskless asset’, *Journal of Financial Economics* **13**, 575–592.

Kandel, S. (1986), ‘The geometry of the likelihood estimator of the zero beta return’, *Journal of Finance* **31**, 339–346.

Kandel, S. and Stambaugh, R. F. (1989), ‘A mean-variance framework for tests of asset pricing models’, *The Review of Financial Studies* **2**, 125–156.

Kim, D. (1995), ‘The errors-in-variables problem in the cross-section of expected stock returns’, *Journal of Finance* **50**, 1605–1634.

Kleibergen, F. (2002), ‘Pivotal statistics for testing structural parameters in instrumental variables regression’, *Econometrica* **70**, 1781–1803.

Kleibergen, F. (2005), ‘Testing parameters in GMM without assuming that they are identified’, *Econometrica* **73**, 1103–1123.

Kleibergen, F. (2009), ‘Tests of risk premia in linear factor models’, *Journal of Econometrics* **149**, 149–173.

Lewellen, J., Nagel, S. and Shanken, J. (2009), ‘A skeptical appraisal of asset-pricing tests’, *Journal of Financial Economics* **92**, 175–194.

Litzenberger, G. N. and Ramaswamy, K. (1979), ‘The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence’, *Journal of Financial Economics* **7**, 163–195.

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

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

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

Mikusheva, A. (2009), Robust confidence sets in the presence of weak instruments, Technical report, MIT Department of Economics, Cambridge, Massachusetts.

Moreira, M. J. (2003), ‘A conditional likelihood ratio test for structural models’, *Econometrica* **71**(4), 1027–1048.

Perold, A. F. (2004), 'The capital asset pricing model', *Journal of Economic Perspectives* **18**, 3–24.

Ravikumar, B., Ray, S. and Savin, E. (2000), 'Robust Wald tests in SUR systems with adding-up restrictions', *Econometrica* **68**, 715–720.

Ray, S. and 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* **23**, 91–109.

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

Roll, R. (1985), 'A note on the geometry of Shanken's CSR T2 test for mean/variance efficiency', *Journal of Financial Economics* **14**, 349–357.

Sentana, E. (2009), 'The econometrics of mean-variance efficiency tests: A survey', *Econometrics Journal* **12**, C65–C101.

Shanken, J. (1985), 'Multivariate tests of the zero-beta CAPM', *Journal of Financial Economics* **14**, 325–348.

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

Shanken, J. (1992), 'On the estimation of beta-pricing models', *Review of Financial Studies* **5**, 1–33.

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

Shanken, J. and Zhou, G. (2007), 'Estimating and testing beta pricing models: Alternative methods and their performance in simulations', *Journal of Financial Economics* **84**, 40–86.

Staiger, D. and Stock, J. H. (1997), 'Instrumental variables regression with weak instruments', *Econometrica* **65**(3), 557–586.

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

Stock, J. H. and Wright, J. H. (2000), 'GMM with weak identification', *Econometrica* **68**, 1097–1126.

Stock, J. H., Wright, J. H. and Yogo, M. (2002), 'A survey of weak instruments and weak identification in generalized method of moments', *Journal of Business and Economic Statistics* **20**(4), 518–529.

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

Wang, J. and Zivot, E. (1998), ‘Inference on structural parameters in instrumental variables regression with weak instruments’, *Econometrica* **66**(6), 1389–1404.

Zerbe, G. O. (1978), ‘On Fieller’s theorem and the general linear model’, *The American Statistician* **32(3)**, 103–105.

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

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

Zivot, E., Startz, R. and Nelson, C. R. (1998), ‘Valid confidence intervals and inference in the presence of weak instruments’, *International Economic Review* **39**, 1119–1144.

Table 3. Tests on zero-beta rate: empirical power
Gaussian GARCH design

$n = 12$		$T = 60$		$T = 120$		$T = 828$	
Statistic	p-value	Step	Power	Step	Power	Step	Power
$LR(\gamma_0)$	MC: $\phi_1 = \phi_2 = 0$	.50	.112	.50	.203	.20	.195
	MC: ϕ_1, ϕ_2 known		.113		.204		.198
	MMC: ϕ_1, ϕ_2 unknown		.106		.170		.168
$\mathscr{J}(\gamma_0)$	MMC: ϕ_1, ϕ_2 known		.088		.155		.208
	MMC: ϕ_1, ϕ_2 unknown		.078		.133		.183
$LR(\gamma_0)$	MC: $\phi_1 = \phi_2 = 0$	.75	.247	.75	.449	.30	.465
	MC: ϕ_1, ϕ_2 known		.248		.452		.471
	MMC: ϕ_1, ϕ_2 unknown		.213		.411		.425
$\mathscr{J}(\gamma_0)$	MMC: ϕ_1, ϕ_2 known		.177		.316		.442
	MMC: ϕ_1, ϕ_2 unknown		.158		.276		.411
$LR(\gamma_0)$	MC: $\phi_1 = \phi_2 = 0$	1.0	.447	1.0	.753	.50	.945
	MC: ϕ_1, ϕ_2 known		.441		.753		.950
	MMC: ϕ_1, ϕ_2 unknown		.395		.709		.937
$\mathscr{J}(\gamma_0)$	MC: ϕ_1, ϕ_2 known		.300		.552		.934
	MMC: ϕ_1, ϕ_2 unknown		.269		.505		.920
$LR(\gamma_0)$	MC, $\phi_1 = \phi_2 = 0$	2.0	.913	1.5	.973	.75	1.0
	MC: ϕ_1, ϕ_2 known		.915		.970		1.0
	MMC: ϕ_1, ϕ_2 unknown		.892		.962		1.0
$\mathscr{J}(\gamma_0)$	MMC: ϕ_1, ϕ_2 known		.719		.856		1.0
	MMC: ϕ_1, ϕ_2 unknown		.664		.931		1.0

Note – The values of γ_0 tested are: $\gamma_0 = -0.000089$ for $T = 60$, $\gamma_0 = .004960$ for $T = 120$, $\gamma_0 = .005957$ for $T = 828$. Numbers reported are empirical rejection rates for various tests of $\mathscr{H}(\gamma_0)$ with nominal size 5%. The sampling design conforms with the size study, for the $\hat{K}$ case; errors are generated with conditional variance as in (2.17) using $\hat{K}$. See Table 1 for a complete description of the designs and tests applied. Samples are drawn with γ calibrated to its QML counterpart from the 1991-95 subsample; values for γ_0 are set to the latter value + $step \times \widehat{\sigma}_i^{\min}$ (for various $step$ values) where $\widehat{\sigma}_i^{\min} = [\min\{\widehat{\sigma}_i^2\}]^{1/2}$, and $\widehat{\sigma}_i^2$ are the diagonal terms of $\hat{K}\hat{K}'$.

Table 4. QML-based tests of BCAPM

Sample	LR_B	p_∞	Normal			GARCH
			LMC	MMC	BND	BND
1927 – 30	16.10	.137	.269	.308	.366	.340
1931 – 35	14.09	.228	.344	.381	.432	.451
1936 – 40	15.36	.167	.257	.284	.345	.355
1941 – 45	18.62	.068	.148	.163	.203	.213
1946 – 50	32.69	**.001**	.005	**.006**	**.007**	**.006**
1951 – 55	37.04	**.000**	.003	**.004**	**.004**	**.003**
1956 – 60	26.10	**.006**	.027	**.031**	**.042**	**.039**
1961 – 65	29.21	**.002**	.011	**.016**	**.020**	**.015**
1966 – 70	27.45	**.004**	.016	**.018**	**.026**	**.029**
1971 – 75	16.81	.113	.213	.224	.292	.294
1976 – 80	25.76	**.007**	.027	**.031**	**.040**	**.042**
1981 – 85	14.98	.183	.316	.335	.387	.404
1986 – 90	35.41	**.000**	.003	**.004**	**.004**	**.005**
1991 – 95	16.41	.127	.219	.253	.310	.320

	Student-*t*			Normal mixture		
	LMC	MMC	BND	LMC	MMC	BND
1927 – 30	.272	.316	.360	.279	.313	.381
1931 – 35	.359	.399	.468	.342	.387	.452
1936 – 40	.282	.308	.372	.265	.302	.357
1941 – 45	.147	.169	.210	.150	.165	.211
1946 – 50	.007	**.007**	**.010**	.007	**.007**	**.008**
1951 – 55	.003	**.005**	**.005**	.003	**.003**	**.003**
1956 – 60	.030	.040	.052	.028	.035	.045
1961 – 65	.013	**.017**	**.023**	.014	**.021**	**.024**
1966 – 70	.020	**.025**	.032	.018	**.023**	.028
1971 – 75	.217	.248	.300	.206	.238	.292
1976 – 80	.026	.035	.039	.026	.034	.042
1981 – 85	.323	.399	.405	.318	.339	.406
1986 – 90	.004	**.005**	**.005**	.004	**.004**	**.005**
1991 – 95	.226	.263	.325	.226	.261	.319

Note – LR_B is the statistic in (2.24). Remaining numbers are associated p-values. p_∞ is based on $\chi^2(n-1)$. All other non-Gaussian p-values are the largest MC p-values over the shape parameter ν within the specified confidence sets [$\nu = \kappa$ or $\nu = (\pi, \omega)$; refer to Table 5]. LMC is the bootstrap p-value in (6.11) and MMC is the maximal p-value in (6.10) (refer to Section 6.2). BND is the bound (6.2) for the Gaussian case and the *QML-BMC* bound from Theorem **6.3** otherwise; the GARCH BND is the largest *QML-BMC* over ϕ_{1i}, ϕ_{2i} [from (2.17)]. Returns for the months of January and for October 1987 are excluded. Given a 5% level, the cut-off is .05 for p_∞, the normal LMC, MMC and the GARCH p-values; for the Student-*t* and mixtures, the cut-off is .025. p-values which lead to significant tests with this benchmark are in bold.

Table 5. Confidence sets for intervening parameters

	Mixture (π, ω), confidence set for ω					$t(\kappa)$
	(1)	(2)	(3)	(4)	(5)	(6)
Sample	$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.3$	$\pi = 0.4$	$\pi = 0.5$	κ
$1927-30$	≥ 1.8	$1.6-2.8$	$1.6-2.5$	$1.6-2.5$	$1.6-2.6$	$3-12$
$1931-35$	$2.1-10.0$	$1.9-3.0$	$1.9-2.7$	$1.9-2.7$	$2.1-3.0$	$3-8$
$1966-40$	$1.5-3.5$	$1.5-2.3$	$1.4-2.1$	$1.4-2.0$	$1.4-2.1$	$4-25$
$1941-45$	$1.3-3.5$	$1.3-2.1$	$1.3-1.9$	$1.3-1.8$	$1.3-1.9$	≥ 5
$1946-50$	$1.4-3.5$	$1.3-2.2$	$1.3-2.0$	$1.3-1.9$	$1.3-1.9$	$5-37$
$1951-55$	$1.4-3.5$	$1.4-2.2$	$1.3-2.0$	$1.3-1.9$	$1.3-2.0$	$5-34$
$1956-60$	$1.3-2.8$	$1.2-2.0$	$1.2-1.9$	$1.2-1.8$	$1.2-1.8$	≥ 5
$1961-65$	$1.0-2.2$	$1.0-1.6$	$1.0-1.5$	$1.0-1.5$	$1.0-1.5$	≥ 7
$1966-70$	$1.3-3.0$	$1.3-2.0$	$1.3-1.9$	$1.3-1.8$	$1.2-1.9$	≥ 5
$1971-75$	$1.5-3.5$	$1.5-2.2$	$1.4-2.0$	$1.4-1.9$	$1.4-2.0$	$4-24$
$1976-80$	$1.6-4.0$	$1.5-2.5$	$1.5-2.2$	$1.5-2.2$	$1.5-2.3$	$4-19$
$1981-85$	$1.4-3.5$	$1.4-2.1$	$1.3-2.0$	$1.3-1.9$	$1.4-2.0$	$5-33$
$1986-90$	$1.1-3.0$	$1.1-2.0$	$1.1-1.8$	$1.0-1.7$	$1.1-1.8$	≥ 5
$1991-95$	$1.0-1.9$	$1.0-1.5$	$1.0-1.4$	$1.0-1.3$	$1.0-1.3$	≥ 19

Note – Numbers in columns (1)-(5) represent a confidence set for the parameters (π, ω) [respectively, the probability of mixing and the ratio of scales] of the multivariate mixtures-of-normal error distribution. Column (6) presents the confidence set for κ, the degrees-of-freedom parameter of the multivariate Student-t error distribution. See Section 6 for details on the construction of these confidence sets: the values of (π, ω) or κ (respectively) in this set are not rejected by the *CSK* test (6.12) [see Dufour et al. (2003)] under multivariate mixtures or Student-t errors (respectively). Note that the maximum of the p-value occurs in the closed interval for ω. Returns for the month of January and October 1987 are excluded from the data set.

Table 6. QML-based point and set estimates for the zero-beta portfolio rate

Sample	$\bar{R}_M$	$\bar{r}_f$	$\hat{\gamma}$	Wald-type
1927 – 30	.0045	.0045	.0047	[−.0037 , .0130]
1931 – 35	.0103	.0025	−.0130	[−.0301 , .0039]
1926 – 40	.0031	−.0006	−.0069	[−.0192 , .0055]
1941 – 45	.0097	−.0042	.0117	[.0037 , .0198]
1946 – 50	.0021	−.0051	−.0219	[−.0189 , −.0070]
1951 – 55	.0145	.0001	.0024	[−.0015 , .0064]
1956 – 60	.0086	.0002	.0156	[.0109 , .0202]
1961 – 65	.0080	.0014	.0571	[.0398 , .0744]
1966 – 70	.0008	.0004	.0169	[.0096 , .0242]
1971 – 75	−.0061	−.0010	.0150	[.0030 , .0270]
1976 – 80	.0056	−.0012	−.0096	[−.0169 , −.0024]
1981 – 85	.0081	.0037	.0197	[.0125 , .0268]
1986 – 90	.0088	.0020	.0053	[−.0024 , .0131]
1991 – 95	.0104	.0011	.0010	[−.0130 , .0062]
	95% Confidence set, Fieller-type			
Sample	Normal errors	Student t errors	Mixture errors	GARCH
1927 – 30	[−.0133 , .0227]	[−.0143 , .0229]	[−.0141 , .0227]	[−.0125, .020]
1931 – 35	[−.0509 , .0225]	[−.0520 , .0225]	[−.0157 , .0227]	[−.0517, .0217]
1926 – 40	[−.0341 , .0187]	[−.0350 , .0190]	[−.0349 , .0817]	[−.0300, .0175]
1941 – 45	[−.0045 , .0275]	[−.0048 , .0287]	[−.0045 , .0283]	[−.0025, .0275]
1946 – 50	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
1951 – 55	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
1956 – 60	$\emptyset$	[.0149 , .0161]	$\emptyset$	$\emptyset$
1961 – 65	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
1966 – 70	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
1971 – 75	[−.0069 , .0454]	[−.0081 , .0488]	[−.0069 , .0531]	[−.0050, .0450]
1976 – 80	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
1981 – 85	[.0059 , .0371]	[.0051 , .0376]	[.0051 , .0387]	[.0075, .0350]
1986 – 90	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
1991 – 95	[−.0285 , .0147]	[−.0303 , .0154]	[−.0325 , .0147]	[−.0275, .0125]

Note – $\bar{R}_M$ is the average real market portfolio return for each subperiod, $\bar{r}_f$ is the real average risk-free rate for each subperiod; $\hat{\gamma}$ is the QML estimate of γ; the remaining columns report 95% confidence sets for this parameter, using, respectively, the asymptotic standard errors (2.27) [$\hat{\gamma}\pm 1.96\times$ AsySE$(\hat{\gamma})$], the inverted test based on $LR(\gamma_0)$ from Theorem **3.1**, and the MC Gaussian p-value, the MMC p-value imposing multivariate $t(\kappa)$ errors and mixture-of-normals (π, ω) errors, and the MMC GARCH p-value. See Section 4 for details on the construction of these confidence sets. Non-Gaussian p-values are the largest MC p-values over the shape parameters κ or (π, ω). The GARCH p-value is the largest MC p-value over ϕ_{1i}, ϕ_{2i} [from (2.17)]. Returns for the months of January and October 1987 are excluded from the data set.

Table 7. $\mathscr{J}(\gamma_0)$ based inference on the zero-beta portfolio rate

	(1)	(2)	(3)	(4)
sample	$\tilde{\gamma} = \underset{\gamma_0}{\operatorname{argmin}} \mathscr{J}(\gamma_0)$	$\underset{\gamma_0}{\min} \mathscr{J}(\gamma_0)$	BND	95% Confidence set, MMC
1927 – 30	.0090	71.29	.650	$[-.0195, .0235]$
1931 – 35	$-.0045$	71.06	.541	$[-.0240, .0250]$
1926 – 40	$-.0045$	54.52	.620	$[-.0355, .0550]$
1941 – 45	.0415	163.26	.143	$[-.0455, .0670]$
1946 – 50	.0000	133.76	.121	$[-.0105, .0075]$
1951 – 55	.0075	104.93	.250	$[.0000, .0120]$
1956 – 60	.0195	110.18	.280	$[-.0385, .0415]$
1961 – 65	.0370	149.61	.142	$[-.0295, -.0150] \cup [.0250, .0670]$
1966 – 70	.0090	168.54	.081	$[.0045, .0135]$
1971 – 75	.0060	61.06	.623	$[-.0180, .0067]$
1976 – 80	.0060	172.09	.061	$[-.0225, .0135]$
1981 – 85	.0195	121.41	.201	$[.0105, .0385]$
1986 – 90	.0030	184.38	**.030**	$\emptyset$
1991 – 95	.0100	53.60	.841	$\{\leq .0075\} \cup \{\geq .0310\}$

Note – $\mathscr{J}(\gamma_0)$ is the HAC statistic in (2.28). $\tilde{\gamma}$ is the minimum distance estimator from (2.29). Column (3) provides a bound MC p-value simulated at $\tilde{\gamma}$ and maximized over ϕ_{1i}, ϕ_{2i} [from (2.17)]. Column (4) provides the confidence set for γ which inverts the inverted test based on $\mathscr{J}(\gamma_0)$ and the MMC GARCH p-value; again, this is the largest MC p-value over ϕ_{1i}, ϕ_{2i} [from (2.17)]. Returns for the months of January and October 1987 are excluded from the data set. Given a 5% level, the cut-off of the BND p-value is .05; p-values which lead to significant tests with this benchmark are in bold. Note that the confidence set which inverts $\mathscr{J}(\gamma_0)$ based on the asymptotic $\chi^2(12)$ cut-off is empty for all sub-periods.