The econometrics of mean-variance efficiency tests: a survey*

Enrique Sentana
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<sentana@cemfi.es>

February 2008
Revised: June 2009

Abstract

This paper provides a comprehensive survey of the econometrics of mean-variance efficiency tests. Starting with the classic F test of Gibbons, Ross and Shanken (1989) and its generalised method of moments version, I analyse the effects of the number of assets and portfolio composition on test power. I then discuss asymptotically equivalent tests based on portfolio weights, and study the trade-offs between efficiency and robustness of using parametric and semiparametric likelihood procedures that assume either elliptical innovations or elliptical returns. After reviewing finite sample tests, I conclude with a discussion of mean-variance-skewness efficiency and spanning tests, and other interesting extensions.

Keywords: Elliptical Distributions, Exogeneity, Financial Returns, Generalised Method of Moments, Linear Factor Pricing, Maximum Likelihood, Portfolio choice, Stochastic Discount Factor.

JEL: C12, C13, C16, G11, G12

*This paper was prepared for the *Econometrics Journal* Session on Financial Econometrics that took place during the 2008 Royal Economic Society Annual Conference at the University of Warwick. Some of the material is partly based on a series of Master theses that I supervised at CEMFI between 1994 and 2004, which became joint work with Pedro L. Sánchez-Torres, María Rada, Francisco Peñaranda, Javier Mencía and Dante Amengual, to whom I am grateful. I am also grateful to the last three for their comments on this draft, as well as to Peter Bossaerts, Raymond Kan, Richard Luger, Theo Nijman, Eric Renault, Alessio Sancetta, Jay Shanken, Giovanni Urga and Guofu Zhou for their feedback. The comments of the editor and two anonymous referees have also substantially improved the presentation. Of course, the usual caveat applies. Financial support from the Spanish Ministry of Science and Innovation through grant ECO 2008-00280 is gratefully acknowledged.

1 Introduction

Mean-variance analysis is widely regarded as the cornerstone of modern investment theory. Despite its simplicity, and the fact that more than five and a half decades have elapsed since Markowitz published his seminal work on the theory of portfolio allocation under uncertainty (Markowitz (1952)), it remains the most widely used asset allocation method. There are several reasons for its popularity. First, it provides a very intuitive assessment of the relative merits of alternative portfolios, as their risk and expected return characteristics can be compared in a two-dimensional graph. Second, mean-variance frontiers are spanned by only two funds, a property that simplifies their calculation and interpretation, and that also led to the derivation of the Capital Asset Pricing Model (CAPM) by Sharpe (1964), Lintner (1965) and Mossin (1966). Finally, mean-variance analysis becomes the natural approach if we assume Gaussian or elliptical distributions for asset returns, because in that case it is fully compatible with expected utility maximisation regardless of investor preferences (see e.g. Chamberlain (1983), Owen and Rabinovitch (1983) and Berk (1997); see also Ross (1978) for a related discussion).

A portfolio with excess returns r_{1t} is mean-variance efficient with respect to a given set of N_2 assets with excess returns $\mathbf{r}_{2t}$ if it is not possible to form another portfolio of those assets and r_{1t} with the same expected return as r_{1t} but a lower variance, or more appropriately, with the same variance but a higher expected return. Despite the simplicity of the definition, testing for mean-variance efficiency is of paramount importance in many practical situations, such as mutual fund performance evaluation (see De Roon and Nijman (2001) for a recent survey), gains from portfolio diversification (Errunza, Hogan and Hung (1999)), or tests of linear factor asset pricing models, including the CAPM and APT, which imply that certain portfolio must be mean-variance efficient (see e.g. Campbell, Lo and MacKinlay (1996) or Cochrane (2001) for advanced textbook treatments).

If the first two moments of returns were known, then it would be straightforward to confirm or disprove the mean-variance efficiency of r_{1t} by simply checking whether they lied on the portfolio frontier spanned by $\mathbf{r}_t = (r_{1t}, \mathbf{r}'_{2t})'$. In practice, of course, the mean and variance of portfolio returns are unknown, and the sample mean and standard deviation of r_{1t} will lie inside the estimated mean-variance frontier with probability one.

Therefore, a statistical hypothesis test provides a rather natural decision method in this context because it explicitly takes into account the sampling variability in the estimation of the first two moments of returns. Otherwise, such a variability would be misleading because the inclusion of additional assets systematically leads to the expansion of the sample frontiers irrespective of whether the theoretical frontier is affected, in the same way as the inclusion of additional regressors systematically leads to increments in sample $R^{2\prime}s$ regardless of whether their theoretical regression coefficients are 0.

To emphasise the importance of sampling uncertainty in this context, I have conducted the following simulation experiment. I have assumed that investors have access to a reference asset with excess returns r_{1t} and three additional assets, whose excess returns r_{it}, r_{jt} and r_{kt} are *i.i.d.* with an annual mean of 0%, uncorrelated among themselves and with the original asset, so that the *true* maximum Sharpe ratio (i.e. the ratio of the expected excess return on a portfolio to its standard deviation) does not increase. Then I simulate two years of daily data many times, and compute the original and augmented mean-variance frontiers, as well as the incremental one, which is based on the differences between r_{it}, r_{jt} and r_{kt} and their best tracking portfolios based on r_{1t}. Figure 1 presents part of the ensemble of incremental frontiers, while Figure 2 contains the sampling distribution of the GMM estimator of the incremental Sharpe ratio. As can be seen from both pictures, if one did not take into account sampling uncertainty then one would always conclude that there are clear gains from also investing in r_{it}, r_{jt} and r_{kt} when in reality there are none.

In fact, the sampling uncertainty surrounding expected returns is so large that several authors have forcefully raised some doubts about the usual practice of applying mean-variance investment rules replacing expected returns, variances and covariances by their sampling counterparts. In this sense, there are several solutions that explicitly take into account sampling uncertainty in making portfolio decisions in practice. These include not only Bayesian approaches but also classical ones. For instance, the modifications of the plug-in rule suggested by ter Horst, de Roon and Werker (2006) or Antoine (2008) from a classical perspective, as well as the Bayesian solution proposed by Bawa, Brown and Klein (1979) and others amount to levering up or more likely down the usual mean-variance portfolio rule by effectively changing the risk aversion parameter of the investor. However, the maximum Sharpe ratio attainable remains the same. Hence, an investor who currently applies one of those alternative rules to a vector of N_1 excess returns $\mathbf{r}_{1t}$, say, but who is considering whether or not to diversify her investments into $\mathbf{r}_{2t}$, should still be interested in conducting a mean-variance efficiency test.

The purpose of this paper is to survey mean-variance efficiency tests, with an emphasis on methodology rather than empirical findings, and paying more attention to some recent contributions and their econometric subtleties. In this sense, it complements previous surveys by Shanken (1996), Campbell, Lo and MacKinlay (1997) and Cochrane (2001). In order to accommodate most of the literature, in what follows I shall often work with the vector $\mathbf{r}_{1t}$, so that the null hypothesis should be understood as saying that some portfolio of the N_1 elements in $\mathbf{r}_{1t}$ lies on the efficient part of the mean-variance frontier spanned by $\mathbf{r}_{1t}$ and $\mathbf{r}_{2t}$.[1]

The rest of the paper is organised as follows. I introduce the theoretical set up in section 2, review the original tests in section 3, and analyse the effects of the number of assets and portfolio composition on test power in section 4. Then I discuss asymptotically equivalent tests based on portfolio weights in section 5, and study the trade-offs between efficiency and robustness of using parametric and semiparametric likelihood procedures that assume either elliptical innovations or elliptical returns in section 6. After reviewing finite sample tests in section 7, I conclude with a discussion of mean-variance-skewness efficiency and spanning tests in section 8. Finally, I mention some related topics and suggestions for future work in section 9. Proofs of the few formal results that I present can be found in the original references.

2 Mean-Variance Portfolio Frontiers

Consider a world with one riskless asset, and a finite number N of risky assets. Let R_0 denote the gross return on the safe asset (that is, the total payoff per unit invested, which includes capital gains plus any cash flows received), $\mathbf{R} = (R_1, R_2, \ldots, R_N)'$ the vector of gross returns on the N remaining assets, with vector of means and matrix of variances and covariances $\boldsymbol{\nu}$ and $\boldsymbol{\Sigma}$ respectively, which I assume bounded. Let $p = w_0R_0 + w_1R_1 + \ldots + w_NR_N$ denote the payoffs to a portfolio of the $N+1$ primitive assets with weights given by w_0 and the vector $\mathbf{w} = (w_1, w_2, \ldots, w_N)'$. Importantly, I assume

[1] In this sense, it is important to note that in the case in which r_{1t} contains single asset, the null hypothesis only says that r_{1t} spans the mean-variance frontier, so in principle it could lie on its innefficient part (see GRS).

that there are no transaction costs or other impediments to trade, and in particular, that short-sales are allowed. I also assume that the wealth of any particular investor is such that her individual behaviour does not alter the distribution of returns.

There are at least three characteristics of portfolios in which investors are usually interested: their cost, the expected value of their payoffs, and their variance, given by $C(p) = w_0 + \mathbf{w}'\iota_N$, $E(p) = w_0 R_0 + \mathbf{w}'\boldsymbol{\nu}$ and $V(p) = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$ respectively, where $\boldsymbol{\iota}_N$ is a vector of N ones. Let $\mathcal{P}$ be the set of payoffs from all possible portfolios of the $N+1$ original assets, i.e. the linear span of $(R_0, \mathbf{R}')$, $\langle R_0, \mathbf{R}' \rangle$. Within this set, several subsets deserve special attention. For instance, it is worth considering all unit cost portfolios $\mathcal{R} = \{p \in \mathcal{P} : C(p) = 1\}$, whose payoffs can be directly understood as returns per unit invested; and also all zero cost, or arbitrage portfolios $\mathcal{A} = \{p \in \mathcal{P} : C(p) = 0\}$. In this sense, note that any non-arbitrage portfolio can be transformed into a unit-cost portfolio by simply scaling its weights by its cost. Similarly, if $\mathbf{r} = \mathbf{R} - R_0 \boldsymbol{\iota}_N$ denotes the vector of returns on the N primitive risky assets in excess of the riskless asset, it is clear that $\mathcal{A}$ coincides with the linear span of $\mathbf{r}$, $\langle \mathbf{r} \rangle$. The main advantage of working with excess returns is that their expected values $\boldsymbol{\mu} = \boldsymbol{\nu} - R_0 \boldsymbol{\iota}_N$ directly give us the risk premia of $\mathbf{R}$, without altering their covariance structure. On the other hand, one must distinguish between riskless portfolios, $\mathcal{S} = \{p \in \mathcal{P} : V(p) = 0\}$ and the rest. In what follows, I shall impose restrictions on the elements of $\mathcal{S}$ so that there are no riskless "arbitrage" opportunities. In particular, I shall assume that $\boldsymbol{\Sigma}$ is regular, so that $\mathcal{S}$ is limited to the linear span of R_0, and the law of one price holds (i.e. portfolios with the same payoffs have the same cost). I shall also assume that R_0 is strictly positive (in practice, $R_0 \geq 1$ for nominal returns).

A simple, yet generally incomplete method of describing the choice set of an agent is in terms of the mean and variance of all the portfolios that she can afford. Let us consider initially the case of an agent who has no wealth whatsoever, which means that she can only choose portfolios in $\mathcal{A}$. In this context, frontier arbitrage portfolios, in the usual mean-variance sense, will be those that solve the program $\min V(p)$ subject to the restrictions $C(p) = 0$ and $E(p) = \bar{\mu}$, with $\bar{\mu}$ real. Given that $C(p) = 0$ is equivalent to $p = \mathbf{w}'\mathbf{r}$, I can re-write this problem as $\min_{\mathbf{w}} \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$ subject to $\mathbf{w}'\,\boldsymbol{\mu} = \bar{\mu}$. There are two possibilities: (i) $\boldsymbol{\mu} = \mathbf{0}$, when the frontier can only be defined for $\bar{\mu} = 0$; or (ii) $\boldsymbol{\mu} \neq \mathbf{0}$, in which case the solution for each $\bar{\mu}$ is

$$\mathbf{w}^*(\bar{\mu}) = \bar{\mu}(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$$

As a consequence, the arbitrage portfolio $r_p = (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{-1}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{r}$ generates the whole zero-cost frontier, in what can be called one-fund spanning. Moreover, given that the variance of the frontier portfolios with mean $\bar{\mu}$ will be $\bar{\mu}^2(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{-1}$, in mean-standard deviation space the frontier is a straight line reflected in the origin whose efficient section has slope $\sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}$. Therefore, this slope fully characterises in mean-variance terms the investment opportunity set of an investor with no wealth, as it implicitly measures the trade-off between risk and return that the available assets allow at the aggregate level.

Traditionally, however, the frontier is usually obtained for unit-cost portfolios, and not for arbitrage portfolios. Nevertheless, given that the payoffs of any portfolio in $\mathcal{R}$ can be replicated by means of a unit of the safe asset and a portfolio in $\mathcal{A}$, in mean-standard deviation space, the frontier for $\mathcal{R}$ is simply the frontier for $\mathcal{A}$ shifted upwards in parallel by the amount R_0. And although now we will have two-fund spanning, for a given safe rate, the slope $\sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}$ continues to fully characterise the investment opportunity set of an agent with positive wealth.

An alternative graphical interpretation of the same result would be as follows. The trade-off between risk and return of any unit-cost portfolio in $\mathcal{R}$ is usually measured as the ratio of its risk premium to its standard deviation. More formally, if $R_u \in \mathcal{R}$, then $s(r_u) = \mu_u/\sigma_u$, where $\mu_u = E(r_u)$, $\sigma_u^2 = V(r_u)$, and $r_u = R_u - R_0$. This expression, known as the Sharpe ratio of the portfolio after Sharpe (1966, 1994), remains constant for any portfolio whose mean excess return and standard deviation lie along the ray which, starting at the origin, passes through the point (μ_u, σ_u) because the Sharpe ratio coincides with the slope of this ray. As a result, the steeper (flatter) a ray is (i.e. the closer to the y (x) axis), the higher (lower) the corresponding Sharpe ratio.

Then, since $\mu_p = 1$ and $\sigma_p^2 = (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{-1}$, the slope $s(r_p) = \mu_p/\sigma_p = \sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}$ will give us the Sharpe ratio of

$$R_p(w_{r_p}) = R_0 + w_{r_p} r_p$$

for any $w_{r_p} > 0$, which is the highest attainable. Therefore, in mean excess return-standard deviation space, all $R_p(w_{r_p})$ lie on a positively sloped straight line that starts from the origin. Assuming that $\boldsymbol{\iota}_N'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} > 0$, as the investor moves away from the origin, where she is holding all her wealth in the safe asset, the net total position invested in the riskless asset is steadily decreasing, and eventually becomes zero. Beyond that point, she begins to borrow in the money market to lever up her position in the financial markets.[2] The main point to remember, though, is that a portfolio will span the mean-variance frontier if and only if its square Sharpe ratio is maximum. As we shall see below, this equivalence relationship underlies most mean-variance efficiency tests.

For our purposes, it is useful to relate the maximum Sharpe ratio to the Sharpe ratio of the N underlying assets. Proposition 3 in Sentana (2005) gives the required expression:

Proposition 1 *The Sharpe ratio of the optimal portfolio (in the unconditional mean-variance sense), $s(r_p)$, only depends on the vector of Sharpe ratios of the N underlying assets, $s(\mathbf{r})$, and their correlation matrix, $\boldsymbol{\rho}_{rr} = dg^{-1/2}(\boldsymbol{\Sigma})\boldsymbol{\Sigma} dg^{-1/2}(\boldsymbol{\Sigma})$ through the following quadratic form:*

$$s^2(r_p) = s(\mathbf{r})'\boldsymbol{\rho}_{rr}^{-1}s(\mathbf{r}), \tag{1}$$

where $dg(\boldsymbol{\Sigma})$ is a matrix containing the diagonal elements of $\boldsymbol{\Sigma}$ and zeros elsewhere.

The above expression, which for the case of $N = 2$ adopts the particularly simple form:

$$s^2(r_p) = \frac{1}{1-\rho^2_{r_1 r_2}}\left[s^2(r_1) + s^2(r_2) - 2\rho_{r_1 r_2}s(r_1)s(r_2)\right], \tag{2}$$

where $\rho_{r_1 r_2} = cor(r_1, r_2)$, turns out to be remarkably similar to the formula that relates the R^2 of the multiple regression of r on (a constant and) $\mathbf{x}$ with the correlations of the simple regressions. Specifically,

$$R^2 = \boldsymbol{\rho}'_{xr}\boldsymbol{\rho}_{xx}^{-1}\boldsymbol{\rho}_{xr}. \tag{3}$$

The similarity is not merely coincidental. From the mathematics of the mean-variance frontier, we know that $E(r_j) = cov(r_j, r_p)E(r_p)/V(r_p)$, and therefore, that

[2] The portfolio at which the net position in the riskless asset is exactly 0 is known as the "tangency" portfolio, because when $\boldsymbol{\iota}'_N\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \neq 0$, there is tangency at that particular point between the mean-variance frontier without a riskless asset and the analogous frontier with a riskless asset in expected return - standard deviation space. The exact expression for its weights is:

$$\mathbf{w}^*(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\boldsymbol{\iota}'_N\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}) = (\boldsymbol{\iota}'_N\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}.$$

If $\boldsymbol{\iota}'_N\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} < 0$ (> 0), the expected excess return of this portfolio will be negative (positive), which means that tangency will take place along the inneficient (efficient) section of the mean-variance frontier for excess returns (see e.g. Maller and Turkington (2002)). If $\boldsymbol{\iota}'_N\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = 0$, though, the frontier with a riskless asset coincides with the asymptotes of the frontier without a riskless asset, so strictly speaking no tangency portfolio exists.

$s(r_j) = cor(r_j, r_p)s(r_p)$. In other words, the correlation coefficient between r_j and r_p is $s(r_j)/s(r_p)$, i.e. the ratio of their Sharpe ratios. Hence, the result in Proposition 1 follows from (3) and the fact that the coefficient of determination in the multiple regression of r_p on $\mathbf{r}$ will be 1 because r_p is a linear combination of this vector.

We can use the partitioned inverse formula to alternatively write expression (1) in the following convenient form

$$s^2(r_p) = s(\mathbf{r}_1)'\boldsymbol{\rho}_{r_1r_1}^{-1}s(\mathbf{r}_1) + s(\mathbf{z}_2)'\boldsymbol{\rho}_{zz}^{-1}s(\mathbf{z}_2) = s^2(r_{p_1}) + \mathbf{a}'\boldsymbol{\Omega}^{-1}\mathbf{a}, \tag{4}$$

where $s(r_{p_1})$ is the Sharpe ratio of the tangency portfolio obtained from $\mathbf{r}_1$ alone, $r_{p_1} = \boldsymbol{\mu}_1'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{r}_1$, the vector $\mathbf{z}_2 = \mathbf{r}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{r}_1$ contains the components of $\mathbf{r}_2$ whose risk has been fully hedged against the risk of $\mathbf{r}_1$, $\mathbf{a} = E(\mathbf{z}_2) = \boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1$, $\boldsymbol{\rho}_{zz} = dg^{-1/2}(\boldsymbol{\Omega})\boldsymbol{\Omega}dg^{-1/2}(\boldsymbol{\Omega})$ and $\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. Given that we can interpret $\mathbf{a}$ as the intercepts in the theoretical least squares projection of $\mathbf{r}_2$ on a constant and $\mathbf{r}_1$, it trivially follows from (4) that $\mathbf{r}_1$ will be mean-variance efficient if and only if $\mathbf{a} = \mathbf{0}$ (see Black, Jensen and Scholes (1972), Jobson and Korkie (1982, 1985), Huberman and Kandel (1987) and Gibbons, Ross and Shanken (1989) (GRS)).

In the bivariate case, (4) reduces to:

$$s^2(r_p) = s^2(r_1) + s^2(z_2),$$

where

$$s(z_2) = \frac{\mu_2 - (\sigma_{12}/\sigma_1^2)\mu_1}{\sqrt{\sigma_2^2 - \sigma_{12}^2/\sigma_1^2}} = \frac{a_2}{\omega_2} = \frac{s(r_2) - \rho_{12}s(r_1)}{\sqrt{1-\rho_{12}^2}}$$

is the Sharpe ratio of $z_2 = r_2 - \sigma_{12}/\sigma_1^2\, r_1$. When r_1 is regarded as a benchmark portfolio, $s(z_2)$ is often known as the information (or appraisal) ratio of r_2.

Corollary 1 in Shanken (1987a) provides the following alternative expression for the maximum Sharpe ratio of $\mathbf{z}_2$ in terms of the Sharpe ratio of r_{p_1} and the correlation between this portfolio and r_p:

$$s(\mathbf{z}_2)'\boldsymbol{\rho}_{zz}^{-1}s(\mathbf{z}_2) = s^2(r_{p_1})\left[\frac{1}{cor^2(r_{p_1}, r_p)} - 1\right].$$

This result exploits the previously mentioned fact that $cor(r_{p_1}, r_p) = s(r_{p_1})/s(r_p)$ (see also Kandel and Stambaugh (1987) and Meloso and Bossaerts (2006)). Intuitively, the incremental Sharpe ratio will reach its minimum value of 0 when $r_{p_1} = r_p$ but it will increase as the correlation between those two portfolios decreases.

3 The original tests

The framework described in the previous section has an implicit time dimension that corresponds to the investment horizon of the agents. To make it econometrically operational for a panel data of excess returns on $N_1 + N_2 = N$ assets over T periods whose length supposedly coincides with the relevant investment horizon, GRS considered the following multivariate, conditionally homoskedastic, linear regression model

$$\mathbf{r}_{2t} = \mathbf{a} + \mathbf{B}\mathbf{r}_{1t} + \mathbf{u}_t = \mathbf{a} + \mathbf{B}\mathbf{r}_{1t} + \mathbf{\Omega}^{1/2}\boldsymbol{\varepsilon}_t^*, \tag{5}$$

where $\mathbf{a}$ is the $N_2 \times 1$ vector of intercepts, $\mathbf{B}$ is a $N_2 \times N_1$ matrix of regression coefficients, $\mathbf{\Omega}^{1/2}$ is an $N_2 \times N_2$ "square root" matrix such that $\mathbf{\Omega}^{1/2}\mathbf{\Omega}^{1/2} = \mathbf{\Omega}$, $\boldsymbol{\varepsilon}_t^*$ is a N_2-dimensional standardised vector martingale difference sequence satisfying $E(\boldsymbol{\varepsilon}_t^*|\mathbf{r}_{1t}, I_{t-1}; \boldsymbol{\gamma}_0, \omega_0) = \mathbf{0}$ and $V(\varepsilon_t^*|\mathbf{r}_{1t}, I_{t-1}; \boldsymbol{\gamma}_0, \omega_0) = \mathbf{I}_{N_2}$, $\boldsymbol{\gamma}' = (\mathbf{a}', \mathbf{b}')$, $\mathbf{b} = vec(\mathbf{B})$, $\boldsymbol{\omega} = vech(\mathbf{\Omega})$, the subscript 0 refers to the true values of the parameters, and I_{t-1} denotes the information set available at $t-1$, which contains at least past values of $\mathbf{r}_{1t}$ and $\mathbf{r}_{2t}$. Crucially, GRS assumed that conditional on $\mathbf{r}_{1t}$ and I_{t-1}, $\boldsymbol{\varepsilon}_t^*$ is independent and identically distributed as a spherical Gaussian random vector, or $\varepsilon_t^*|\mathbf{r}_{1t}, I_{t-1}; \boldsymbol{\gamma}_0, \omega_0 \sim i.i.d.\ N(\mathbf{0}, \mathbf{I}_{N_2})$ for short.

Given the structure of the model, the unrestricted Gaussian ML estimators of $\mathbf{a}$ and $\mathbf{B}$ coincide with the equation by equation OLS estimators in the regression of each element of $\mathbf{r}_{2t}$ on a constant and $\mathbf{r}_{1t}$. Consequently,

$$\begin{aligned}
\hat{\mathbf{a}} &= \hat{\boldsymbol{\mu}}_2 - \hat{\mathbf{B}}\hat{\boldsymbol{\mu}}_1, && (6)\\
\hat{\mathbf{B}} &= \hat{\mathbf{\Sigma}}_{21}\hat{\mathbf{\Sigma}}_{11}^{-1}, && (7)\\
\hat{\mathbf{\Omega}} &= \hat{\mathbf{\Sigma}}_{22} - \hat{\mathbf{\Sigma}}_{21}\hat{\mathbf{\Sigma}}_{11}^{-1}\hat{\mathbf{\Sigma}}_{21}',
\end{aligned}$$

where

$$\begin{aligned}
\hat{\boldsymbol{\mu}} &= \begin{pmatrix} \hat{\boldsymbol{\mu}}_1 \\ \hat{\boldsymbol{\mu}}_2 \end{pmatrix} = \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} \mathbf{r}_{1t} \\ \mathbf{r}_{2t} \end{pmatrix}, \\
\hat{\mathbf{\Gamma}} &= \begin{pmatrix} \hat{\mathbf{\Gamma}}_{11} & \hat{\mathbf{\Gamma}}_{21}' \\ \hat{\mathbf{\Gamma}}_{21} & \hat{\mathbf{\Gamma}}_{22} \end{pmatrix} = \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} \mathbf{r}_{1t}\mathbf{r}_{1t}' & \mathbf{r}_{1t}\mathbf{r}_{2t}' \\ \mathbf{r}_{2t}\mathbf{r}_{1t}' & \mathbf{r}_{2t}\mathbf{r}_{2t}' \end{pmatrix},
\end{aligned}$$

and $\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Gamma}} - \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}'$.

In fact, $\hat{\mathbf{a}}$ and $\hat{\mathbf{B}}$ would continue to be the Gaussian ML estimators if the matrix $\mathbf{\Omega}_0$ were known. In those circumstances, the results in Breusch (1979) would imply that the

Wald (W_T), LR (LR_T) and LM (LM_T) test statistics for the null hypothesis $H_0 : \mathbf{a} = \mathbf{0}$ would all be numerically identical to

$$T \cdot \frac{\hat{\mathbf{a}}'\boldsymbol{\Omega}_0^{-1}\hat{\mathbf{a}}}{1 + \hat{\boldsymbol{\mu}}_1'\hat{\boldsymbol{\Sigma}}_{11}^{-1}\hat{\boldsymbol{\mu}}_1},$$

whose finite sample distribution conditional on the sufficient statistics $\hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\Sigma}}_{11}$ would be that of a non-central χ^2 with N_2 degrees of freedom and non-centrality parameter $T \cdot \mathbf{a}_0'\boldsymbol{\Omega}_0^{-1}\mathbf{a}_0/(1 + \hat{\boldsymbol{\mu}}_1'\hat{\boldsymbol{\Sigma}}_{11}^{-1}\hat{\boldsymbol{\mu}}_1)$.[3] The reason is that the finite sample distribution of $\hat{\mathbf{a}}$, conditional on $\hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\Sigma}}_{11}$, is multivariate normal with mean $\mathbf{a}_0$ and covariance matrix $T^{-1}(1 + \hat{\boldsymbol{\mu}}_1'\hat{\boldsymbol{\Sigma}}_{11}^{-1}\hat{\boldsymbol{\mu}}_1)\boldsymbol{\Omega}_0$.

In practice, of course, $\boldsymbol{\Omega}_0$ is unknown, and has to be estimated along the other parameters. But then, the Wald, LM and LR tests no longer coincide. However, for fixed N_2 and large T all three tests will be asymptotically distributed as the same non-central χ^2 with N_2 degrees of freedom and non-centrality parameter

$$\frac{\tilde{\mathbf{a}}'\boldsymbol{\Omega}^{-1}\tilde{\mathbf{a}}}{1 + \boldsymbol{\mu}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1}$$

under the Pitman sequence of local alternatives $H_{lT} : \mathbf{a} = \tilde{\mathbf{a}}/\sqrt{T}$ (see Newey and MacFadden (1994)). In contrast, they will separately diverge to infinity for fixed alternatives of the form $H_f : \mathbf{a} = \dot{\mathbf{a}}$, which makes them consistent tests. In the case of the Wald test, in particular, we can use Theorem 1 in Geweke (1981) to show that

$$p\lim \frac{1}{T} W_T = \frac{\dot{\mathbf{a}}'\boldsymbol{\Omega}^{-1}\dot{\mathbf{a}}}{1 + \boldsymbol{\mu}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1}$$

coincides with Bahadur's (1960) definition of the approximate slope of the Wald test.[4]

In finite samples, though, the test statistics satisfy the following inequalities

$$W_T \geq LR_T \geq LM_T,$$

which may lead to the conflict among criteria for testing hypotheses pointed out by Berndt and Savin (1977). In effect, the above inequalities reflect the fact that the finite sample distribution of the three tests is not well approximated by their asymptotic distribution, especially when N_2 is moderately large. For that reason, Jobson and Korkie

[3]Consequently, the distribution under the null $H_0 : \mathbf{a} = \mathbf{0}$ is effectively unconditional. In contrast, the unconditional distribution under the alternative is unknown.

[4]Although in general approximate slopes differ from non-centrality parameters for local alternatives, in this case both expressions coincide because the asymptotic variance of $\hat{\mathbf{a}}$ is the same under the null and the alternative.

(1982) proposed a Bartlett (1937) correction that scales the usual LR_T statistic by $1-(N_2+N_1+3)/2T$ to improve the finite sample reliability of its asymptotic distribution.

In this context, the novel contribution of GRS was to exploit results from classic multivariate regression analysis to show that, conditional on the sufficient statistics $\hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\Sigma}}_{11}$, the test statistic

$$F_T = \frac{T-N_2-N_1}{N_2}\frac{\hat{\mathbf{a}}'\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{a}}}{1+\hat{\boldsymbol{\mu}}_1'\hat{\boldsymbol{\Sigma}}_{11}^{-1}\hat{\boldsymbol{\mu}}_1}$$

will be distributed in finite samples as a non-central F with N_2 and $T-N_1-N_2$ degrees of freedom and non-centrality parameter

$$\frac{T\cdot\mathbf{a}_0'\boldsymbol{\Omega}_0^{-1}\mathbf{a}_0}{1+\hat{\boldsymbol{\mu}}_1'\hat{\boldsymbol{\Sigma}}_{11}^{-1}\hat{\boldsymbol{\mu}}_1}.$$

Importantly, for $N_2=1$ this F test coincides with the square of the t-test proposed by Black, Jensen and Scholes (1972). The Wald, LM or LR statistics mentioned before can be written as monotonic transformations of this F test. For instance,

$$F_T = \frac{T-N_2-N_1}{N_2}\left[\exp(LR_T/T)-1\right]$$

GRS also showed that

$$\hat{\mathbf{a}}'\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{a}} = \hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}_1'\hat{\boldsymbol{\Sigma}}_1^{-1}\hat{\boldsymbol{\mu}}_1 = \hat{s}^2(\hat{r}_p)-\hat{s}^2(\hat{r}_{p_1}),$$

where $\hat{s}^2(\hat{r}_p)=\hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$ is the (square) sample Sharpe ratio of the ex-post tangency portfolio that combines $\mathbf{r}_1$ and $\mathbf{r}_2$, while $\hat{s}^2(\hat{r}_{p_1})=\hat{\boldsymbol{\mu}}_1'\hat{\boldsymbol{\Sigma}}_1^{-1}\hat{\boldsymbol{\mu}}_1$ is the (square) sample Sharpe ratio of the ex-post tangency portfolio that uses data on $\mathbf{r}_1$ only.[5] In view of expression (4), an alternative interpretation is that $\hat{\mathbf{a}}'\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{a}}$ is the maximum ex-post square Sharpe ratio obtained by combining $\hat{\mathbf{z}}_2$, which are the components of $\mathbf{r}_2$ that have been fully hedged in sample relative to $\mathbf{r}_1$. The corresponding portfolio, $\hat{\mathbf{a}}'\hat{\boldsymbol{\Omega}}^{-1}(\mathbf{r}_2-\hat{\mathbf{B}}\mathbf{r}_1)=\hat{\mathbf{a}}'\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{z}}_2$, is sometimes known as the (ex post) optimal orthogonal portfolio (see MacKinlay (1995)).

Strictly speaking, GRS considered an incomplete (conditional) model that left unspecified the marginal distribution of $\mathbf{r}_{1t}$. But they would have obtained exactly the same test had they considered the complete (joint) model $\mathbf{r}_t|I_{t-1};\boldsymbol{\rho}\sim i.i.d.\ N[\boldsymbol{\mu}(\boldsymbol{\rho}),\boldsymbol{\Sigma}(\boldsymbol{\rho})]$, where

$$\boldsymbol{\mu}(\boldsymbol{\rho}) = \begin{pmatrix}\boldsymbol{\mu}_1\\ \mathbf{a}+\mathbf{B}\boldsymbol{\mu}_1\end{pmatrix}, \tag{8}$$

$$\boldsymbol{\Sigma}(\boldsymbol{\rho}) = \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{11}\mathbf{B}'\\ \mathbf{B}\boldsymbol{\Omega}_{11} & \mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}'+\boldsymbol{\Omega}\end{pmatrix}, \tag{9}$$

[5]Kandel and Stambaugh (1989) provide an alternative graphical interpretation of the GRS test in *sample* mean-variance space.

and $\boldsymbol{\rho}' = (\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}', \boldsymbol{\mu}_1', \boldsymbol{\sigma}_{11}')$, where $\boldsymbol{\sigma}_{11} = vech(\boldsymbol{\Sigma}_{11})$. The reason is that under this assumption the joint log-likelihood function of $\mathbf{r}_t$ conditional on I_{t-1} can be written as the sum of the conditional log-likelihood function of $\mathbf{r}_{2t}$ given $\mathbf{r}_{1t}$ (and the past), which depends on $\mathbf{a}$, $\mathbf{B}$ and $\boldsymbol{\Omega}$ only, plus the marginal log-likelihood function of $\mathbf{r}_{1t}$ (conditional on the past), which just depends on μ_1 and $\boldsymbol{\Sigma}_{11}$. Given that $\boldsymbol{\theta} = (\mathbf{a}', \mathbf{b}', \boldsymbol{\omega}')'$ and $(\boldsymbol{\mu}_1, \boldsymbol{\sigma}_{11})$ are variation free, we have thus performed a sequential cut of the joint log-likelihood function that makes $\mathbf{r}_{1t}$ weakly exogenous for $(\mathbf{a}, \mathbf{b}, \boldsymbol{\omega})$, which in turn guarantees the efficiency of the GRS procedure (see Engle, Hendry and Richard 1983). In addition, the *i.i.d.* assumption implies that $\mathbf{r}_{1t}$ would in fact be strictly exogenous, which justifies finite sample inferences.

Although the existence of finite sample results is very attractive, particularly when N_2 is moderately large, many empirical studies with financial time series data indicate that the distribution of asset returns is usually rather leptokurtic. For that reason, MacKinlay and Richardson (1991) developed a robust test of mean-variance efficiency by using Hansen's (1982) GMM methodology (see also Harvey and Zhou (1991)). The orthogonality conditions that they considered are

$$\begin{aligned} E\left[\mathbf{m}_R\left(\mathbf{R}_t; \boldsymbol{\gamma}\right)\right] &= \mathbf{0}, \\ \mathbf{m}_R\left(\mathbf{r}_t; \boldsymbol{\gamma}\right) &= \left[\begin{pmatrix} 1 \\ \mathbf{r}_{1t} \end{pmatrix} \otimes \boldsymbol{\varepsilon}_t(\boldsymbol{\gamma})\right], \\ \boldsymbol{\varepsilon}_t(\boldsymbol{\gamma}) &= \mathbf{r}_{2t} - \mathbf{a} - \mathbf{B}\mathbf{r}_{1t}. \end{aligned} \tag{10}$$

The advantage of working within a GMM framework is that under fairly weak regularity conditions inference can be made robust to departures from the assumption of normality, conditional homoskedasticity, serial independence or identity of distribution. But since the above moment conditions exactly identify $\boldsymbol{\gamma}$, the unrestricted GMM estimators coincide with the Gaussian pseudo[6] ML estimators in (6) and (7).[7] An alternative way of reaching the same conclusion is by noticing that the influence function $\mathbf{m}_R\left(\mathbf{R}_t; \boldsymbol{\gamma}\right)$ is a full-rank linear transformation with time-invariant weights of the Gaussian pseudo-score with respect to $\boldsymbol{\gamma}$

$$\mathbf{s}_{\gamma t}(\boldsymbol{\theta}, \mathbf{0}) = \begin{pmatrix} 1 \\ \mathbf{r}_{1t} \end{pmatrix} \otimes \boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\gamma}). \tag{11}$$

[6] In this paper I use "pseudo ML" estimator in the same way as Gourieroux, Monfort and Trognon (1984). In contrast, White (1982) uses the term "quasi ML" for the same concept.

[7] The obvious GMM estimator of $\boldsymbol{\omega}$ is given by $\hat{\boldsymbol{\Omega}}$, which is the sample analogue to the residual covariance matrix.

Not surprisingly, GMM asymptotic theory yields the same answer as standard Gaussian PML results for multivariate regression models:

Proposition 2 *Under appropriate regularity conditions*

$$\sqrt{T}(\hat{\boldsymbol{\gamma}}_{GMM} - \boldsymbol{\gamma}_0) \to N\left[\mathbf{0}, \mathcal{C}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\phi}_0)\right], \tag{12}$$

where

$$\mathcal{C}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\phi}) = \mathcal{A}^{-1}_{\gamma\gamma}(\boldsymbol{\phi})\mathcal{B}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\phi})\mathcal{A}^{-1}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\phi}),$$
$$\mathcal{A}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\phi}) = -E\left[\mathbf{h}_{\boldsymbol{\gamma}\boldsymbol{\gamma}t}(\boldsymbol{\theta}, \mathbf{0})|\phi\right] = E\left[\mathcal{A}_{\boldsymbol{\gamma}\boldsymbol{\gamma}t}(\boldsymbol{\phi})|\phi\right],$$
$$\mathcal{A}_{\boldsymbol{\gamma}\boldsymbol{\gamma}t}(\boldsymbol{\phi}) = -E[\mathbf{h}_{\boldsymbol{\gamma}\boldsymbol{\gamma}t}(\boldsymbol{\theta};\mathbf{0})|\,\mathbf{r}_{1t}, I_{t-1};\boldsymbol{\phi}] = \begin{pmatrix} 1 & \mathbf{r}_{1t} \\ \mathbf{r}_{1t} & \mathbf{r}_{1t}\mathbf{r}'_{1t} \end{pmatrix} \otimes \boldsymbol{\Omega}^{-1},$$
$$\mathcal{B}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\phi}) = \lim_{T\to\infty} V\left[\left.\frac{\sqrt{T}}{T}\bar{\mathbf{s}}_{\boldsymbol{\gamma}T}(\boldsymbol{\theta}, \mathbf{0})\right|\boldsymbol{\phi}\right],$$

where $\mathbf{h}_{\boldsymbol{\gamma}\boldsymbol{\gamma}t}(\boldsymbol{\theta};\mathbf{0})$ is the block of the component of the Gaussian Hessian matrix corresponding to $\boldsymbol{\gamma}$ attributable to the t^{th} observation, $\bar{\mathbf{s}}_{\boldsymbol{\gamma}T}(\boldsymbol{\theta}, \mathbf{0})$ is the sample mean of the Gaussian scores, and $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta})'$ the $2N_2 + N_2(N_2+1)/2 + q$ parameters of the model, which include some q additional parameters $\boldsymbol{\eta}$ that determine the shape of the distribution of $\boldsymbol{\varepsilon}^*_t$ conditional on $\mathbf{r}_{1t}$ and I_{t-1}.

From here, it is straightforward to obtain robust, efficient versions of the Wald and LM tests, which will continue to be asymptotically equivalent to each other under the null and sequences of local alternatives (see Property 18.2 in Gouriéroux and Monfort (1995)). However, the LR test will not be asymptotically valid unless $\boldsymbol{\varepsilon}_t(\boldsymbol{\gamma}_0)$ is *i.i.d.* conditional on $\mathbf{r}_{1t}$ and I_{t-1}. But it is possible to define a LR analogue as the difference in the GMM criterion functions under the null and the alternative. This "distance metric" test will have an asymptotic χ^2 distribution only if the GMM weighting matrix is optimally chosen, in which case it will be asymptotically equivalent to the optimal GMM versions of the W_T and LM_T tests under the null and sequences of local alternatives (see e.g. Theorem 9.2 in Newey and MacFadden (1994)).

Importantly, the optimal distance metric test will coincide with the usual overidentification test since the moment conditions (10) exactly identify $\boldsymbol{\gamma}$ under the alternative. In addition, given that the influence functions (10) are linear in the parameters $\boldsymbol{\gamma}$, the results in Newey and West (1987) imply that regardless of whether we use the Wald, Lagrange multiplier or Distance Metric tests, there will be two numerical distinct test statistics only: those that use the optimal GMM weighting matrix computed under the null, and those based on the optimal weighting matrix computed under the alternative.

4 The effects of the number of assets and portfolio composition on test power

Although at first sight this section may only seem interesting for theoretically inclined econometricians, arguably it is also relevant for applied researchers because in practice the substantive conclusions about the mean-variance efficiency of a candidate portfolio can be rather sensitive to the way in which tests are implemented. Let us start by considering a very simple practical situation. As we mentioned in the previous section, Black, Jensen and Scholes (1972) proposed the use of the t ratio of a_i in the regression of r_2 on a constant and $\mathbf{r}_1$ to test the mean-variance efficiency of $\mathbf{r}_1$. However, when $\mathbf{r}_2$ contains more than one element, it seems natural to follow GRS and conduct a joint test of $H_0 : \mathbf{a} = \mathbf{0}$ in order to increase the probability of rejecting the null hypothesis when $\mathbf{r}_1$ is not mean-variance efficient. Somewhat surprisingly, the answer is not so straightforward. For simplicity, let us initially assume that there are only two assets in $\mathbf{r}_2$, r_i and r_j, say. According to (4), the incremental Sharpe ratio that one can attain by combining $\mathbf{r}_{1t}$, r_{it} and r_{jt} is given by $\mathbf{a}'\mathbf{\Omega}^{-1}\mathbf{a}$, which is the maximum (square) Sharpe ratio that one can achieve by combining the components of r_i and r_j that are fully hedged with respect to $\mathbf{r}_1$, $z_i = a_i + \varepsilon_i$ and $z_j = a_j + \varepsilon_j$. But if apply (4) to z_i and z_j we get

$$\mathbf{a}'\mathbf{\Omega}^{-1}\mathbf{a} = \frac{a_i^2}{\omega_i^2} + \frac{[s(z_j) - \rho_{z_i z_j} s(z_i)]^2}{\sqrt{1 - \rho_{z_i z_j}^2}}$$

where $\rho_{z_i z_j}$ is the correlation between z_i and z_j. An alternative way to interpret this expression is to think of the second summand as the (square) Sharpe ratio of $u_j = z_j - (\omega_{ij}/\omega_j^2)z_i$, which is the component of r_j that is fully hedged with respect to both $\mathbf{r}_{1t}$ and r_{it}.[8] Therefore, when we add r_j to r_i for the purpose of testing the mean-variance efficiency of $\mathbf{r}_1$ we must consider three effects:

1) The increase in the so-called non-centrality parameter of the test statistic, which is proportional to $s^2(u_j)$ and *ceteris paribus* increases power.

[8] It is important to remember that as the correlation between z_i and z_j increases, the law of one price guarantees that $s^2(u_j) = 0$ in the limit of $\rho_{z_i z_j}^2 = 1$.

2) The increase in the number of degrees of freedom of the numerator, which *ceteris paribus* decreases power.

3) The decrease in the number of degrees of freedom of the denominator resulting from the fact that there are additional parameters to be estimated, which *ceteris paribus* decreases power too, although not by much if T is reasonably large.

The net effect is studied in detail by Rada and Sentana (1997). For a given value of $\hat{s}^2(\mathbf{r}_1)$ and different values of T, these authors obtain *isopower* lines, defined as the locus of points in $s^2(z_i), s^2(u_j)$ space for which the power of the univariate test is exactly the same as the power of the bivariate test. GRS also present some evidence on the effects of increasing the number of assets on power under the assumption that the innovations are cross-sectionally homoskedastic and equicorrelated, so that

$$\mathbf{\Omega} = \omega[(1-\rho)\mathbf{I}_{N_2} + \rho \boldsymbol{\iota}_{N_2}\boldsymbol{\iota}'_{N_2}], \tag{13}$$

where ω and ρ are two scalars. Given that the F test estimates a fully unrestricted $\mathbf{\Omega}$, it is not surprising that their results suggest that one should not use a large N_2 (see also MacKinlay (1987)). In fact, the F test can no longer be computed if $N_2 \geq T - N_1$.[9]

The answer to the previous practical question leads to another practical question: If we want to increase the chances of rejecting the null hypothesis when $\mathbf{r}_{1t}$ is not mean-variance efficient, should we group r_{it} and r_{jt} into a portfolio and carry out a single individual t test, or should we consider them separately? Rada and Sentana (1997) study this question in a multivariate context. For simplicity, I will only discuss the situation in which $\mathbf{\Omega}$ is assumed to be a known diagonal matrix, in which case one could work with the vector of re-scaled excess returns $\mathbf{r}_2^* = dg^{-1/2}(\mathbf{\Omega})\mathbf{r}_2$, which are such that

$$\mathbf{r}_2^* = \mathbf{a}^* + \mathbf{B}^*\mathbf{r}_1 + \boldsymbol{\varepsilon}^*,$$

where $\mathbf{a}^* = dg^{-1/2}(\mathbf{\Omega})\mathbf{a}$, $\mathbf{B}^* = dg^{-1/2}(\mathbf{\Omega})\mathbf{B}$ and $V(\boldsymbol{\varepsilon}^*|\mathbf{r}_1) = \mathbf{I}_{N_2}$. Note that the i^{th} element of $\mathbf{a}^*$, $a_i^* = a_i/\omega_i$, coincides with the "information ratio" of r_i introduced at the end of section 2, since it reflects the Sharpe ratio of $z_i = r_i - \boldsymbol{\sigma}'_{i1}\mathbf{\Sigma}_{11}^{-1}\mathbf{r}_1$, which is the component of r_i that cannot be hedged against $\mathbf{r}_1$. In this simplified context, Rada and Sentana

[9] Affleck-Graves and McDonald (1990) proposed a maximum entropy statistic that ensures the non-singularity of the estimated residual covariance matrix $\mathbf{\Omega}$ even if $N_2 > T$. Unfortunately, the finite sample distribution of their test statistic is generally unknown even under normality, and can only be assessed by simulation. In addition, it is not clear either what is limiting behaviour will be when both N_2 and T go to infinity at the same rate.

(1997) express the non-centrality parameter of the joint Wald test of $H_0 : \mathbf{a}^* = \mathbf{0}$ as the sum of the non-centrality parameters of a Wald test whose null is that all information ratios are equal ($H_0 : \mathbf{a}^* = a^* \boldsymbol{\iota}_{N_2}$) and a Wald test whose null is that the average information ratio is 0 ($H_0 : a^* = 0$). Their result is based on a standard analysis of variance argument applied to the ML estimator of $\mathbf{a}^*$. Specifically, they exploit the fact that

$$\sum_{i=1}^{N_2} \hat{a}_i^{*2} = N_2(\hat{a}^{*2} + \hat{\delta}), \tag{14}$$

where

$$\hat{a}^* = N_2^{-1} \sum_{i=1}^{N_2} \hat{a}_i^*, \quad \hat{\delta} = N_2^{-1} \sum_{i=1}^{N_2} (\hat{a}_i^* - \hat{a}^*)^2.$$

It is then easy to see that under their maintained distributional assumptions, $\hat{a}^{*2}$ is proportional to a non-central chi-square with one degree of freedom, while $\hat{\delta}$ is proportional to an independent non-central chi-square with $N_2 - 1$ degrees of freedom. Not surprisingly, Rada and Sentana (1997) show that the contribution of each of those two components to the power of the test depend exclusively on the relative values of the cross-sectional mean of the information ratios $a^* = N_2^{-1} \sum_{i=1}^{N_2} a_i^*$, and their cross-sectional variance $\delta = N_2^{-1} \sum_{i=1}^{N_2} (a_i^* - a^*)^2$. In particular, if there were no cross-sectional variability in the information ratios because $\delta = 0$, then one should simply apply the Black, Jensen and Scholes (1972) test to the equally weighted portfolio of $\mathbf{r}_2$. In contrast, if a^* were 0, such a test would have no power to reject the mean-variance efficiency of $\mathbf{r}_1$ regardless of how big δ could be.

Finally, Rada and Sentana (1997) extend their analysis to the case in which one forms L equally weighted portfolios of M different assets from the N_2 elements of $\mathbf{r}_2^*$, where $M = N_2/L$. In that case, an analysis of variance decomposes the test into three components: a test that the overall mean of the information ratios is zero, as in the previous case, a test that the between group variance in information ratios is 0, and finally a test that their within groups variance is 0. More specifically, if we denote by $\hat{a}_l^*$ the average value of $\hat{a}_i^*$ for those assets that belong to the l^{th} group, so that $\hat{a}^* = L^{-1} \sum_{l=1}^{L} \hat{a}_l^*$, then we will have that

$$\hat{\delta} = \frac{1}{L} \sum_{l=1}^{L} (\hat{a}_l^* - \hat{a})^2 + \frac{1}{L} \sum_{l=1}^{L} \frac{L}{N_2} \sum_{j=1}^{M} (\hat{a}_i^* - \hat{a}_l^*)^2. \tag{15}$$

Note that the first summand is proportional to a non-central chi-square with $L - 1$ degrees of freedom, while the second one is proportional to an independent non-central chi-square with $N_2 - L$ degrees of freedom. In this context, Rada and Sentana (1997) provide isopower lines in the space of within group and between group variances. Their analysis suggests that randomly chosen portfolios will have very little power over and above a test that the overall mean is zero, since the between groups variance is likely to be close to 0 for large M. In contrast, if we could form portfolios that reduce the within group variance in information ratios but increase their between group variance then we would have substantially more power in the portfolio tests than in the test that considers the individual assets. The above results provide a formal justification for the usual practice of grouping returns according to the ranked values of certain observable characteristics that are likely to yield disperse information ratios, such as size or book to value, as opposed to grouping them by industry, which is likely to produce very similar information ratios. Nevertheless, it is important to realise that such procedures may introduce some data snooping size distortions, as illustrated by Lo and MacKinlay (1990).

Another fact that is worth remembering in this context is that the maximum Sharpe ratio attainable for any particular N_2 will be bounded from above by the limiting maximum Sharpe ratio, s_∞, which is also bounded if we rule out arbitrage opportunities as $N_2 \to \infty$ (see Ross (1976) and Chamberlain (1983)). This is important because an increasing number of assets cannot result in an unbounded Sharpe ratio, and consequently, an unbounded non-centrality parameter, as explained by MacKinlay (1987, 1995). In other words, $N_2(a^{*2} + \delta)$ must remain bounded as N_2 goes to infinity, which requires that $(a^{*2} + \delta) = O(N_2^{-1})$.

To see the effects of this restriction, let us obtain the asymptotic distribution of the mean-variance efficiency test when $N_2 \to \infty$ in the case in which $\mathbf{\Omega}$ is diagonal but unknown and the distribution of returns is $i.i.d.$ multivariate normal. Conditional on $\hat{s}^2(\mathbf{r}_1)$, the squared t-ratio of the intercept of the i^{th} asset

$$\tilde{t}_i^{*2} = \frac{T - N_1 - 1}{[1 + \hat{s}^2(r_{p_1})]} \cdot \frac{\hat{a}_i^2}{\hat{\omega}_{ii}}$$

will be distributed independently of the t-ratios of the intercepts of the other assets as a non-central F distribution with 1 and $T - N_1 - 1$ degrees of freedom and non-centrality parameter $Ta_i^{*2}[1 + \hat{s}^2(r_{p_1})]^{-1}$. Hence, its mean will be

$$\pi_i = \frac{T - N_1 - 1}{T - N_1 - 3} \left[1 + \frac{T}{[1 + \hat{s}^2(r_{p_1})]} a_i^{*2} \right] \tag{16}$$

and its variance

$$\lambda_i^2 = \frac{2(T-N_1-1)^2}{(T-N_1-3)^2(T-N_1-5)} \left\{ \begin{array}{c} \left[1+\frac{T}{[1+\hat{s}^2(r_{p_1})]}a_i^{*2}\right]^2 \\ +(T-N_1-3)\left[1+\frac{2T}{[1+\hat{s}^2(r_{p_1})]}a_i^{*2}\right] \end{array} \right\}.$$

Given that the mean-variance efficiency test that exploits the diagonality of $\mathbf{\Omega}$ will be proportional to $\sum_{i=1}^{N_2} \tilde{t}_i^{*2}$, we can use the Linderberg-Feller central limit theorem for independent but heterogeneously distributed random variables[10] to obtain the asymptotic distribution of the joint test for fixed T but large N_2, which under the null will be given by

$$\frac{\sqrt{N_2}}{N_2}\sum_{i=1}^{N_2}\left(\tilde{t}_i^{*2} - \frac{T-N_1-1}{T-N_1-3}\right) \to N(0,2).$$

In contrast, the mean under the alternative will be proportional to $a^{*2}+\delta$ in view of (16). But since we saw before that $a^{*2}+\delta = O(N_2^{-1})$ in order to rule out limiting arbitrage opportunities, one cannot even allow for local alternatives of the form $(\bar{a}^{*2}+\bar{\delta})/\sqrt{N_2}$, and therefore the mean-variance efficiency test is likely to have negligible asymptotic power in those circumstances.[11]

Affleck-Graves and McDonald (1990) suggest to use the statistic $\sum_{i=1}^{N_2} \tilde{t}_i^{*2}$ even when $\mathbf{\Omega}$ is not diagonal. Part of their motivation is that in this way there is no longer any need to form portfolios for the purposes of avoiding a singular estimated covariance matrix. The problem is that the distribution of such a statistic is non-standard if $\mathbf{\Omega}$ is not diagonal, although in samples in which N_2 is small but T is large, we could use Imhof's (1961) results (see also Farebrother (1990)) to approximate the distribution of the statistic $\sum_{i=1}^{N_2} \tilde{t}_i^{*2}$, replacing the matrix $\mathbf{\Omega}$ by its unrestricted sample counterpart $\hat{\mathbf{\Omega}}$ in computing the weights of the associated quadratic form in normal variables. Alternatively, we could impose structure on the cross-sectional distribution of the asset returns. Bossaerts and Hillion (1995) take a first step in this direction and derive the

[10] As is well known, this central limit theorem says that

$$\frac{\sum_{i=1}^{N_2} \tilde{t}_i^{*2} - \sum_{i=1}^{N_2} \pi_i}{\sqrt{\sum_{i=1}^{N_2} \lambda_i^2}} \to N(0,1)$$

as long as the Lindeberg condition is satisfied, which we are implicitly assuming. This condition guarantees that the individual variances λ_i^2 are small compared to their sum, in the sense that for given ϵ and for all sufficiently large N_2, $\lambda_i^2/\sum_{j=1}^{N_2}\lambda_j^2 < \epsilon$ for $i = 1, ..., N_2$ (see Feller 1971, p. 256).

[11] Rada and Sentana (1997) also combine the decompositions of $\sum_{i=1}^{N} \hat{a}_i^{*2}$ in (14) and (15) with this asymptotic approximation to obtain the asymptotic distribution of the components of the mean-variance efficiency test attributable to the overall mean of the information ratios, their between groups variance and the within groups one.

asymptotic distribution of $\sum_{i=1}^{N_2}(r_{it} - \sum_{j=1}^{N_1} \tilde{b}_{ij} r_{jt})$ for large N_2 but fixed T, where $\tilde{b}_{ij}$ is the restricted OLS estimator of b_{ij} that imposes the null hypothesis $a_i = 0$, under the assumptions that (i) the conditional distribution of $\boldsymbol{\varepsilon}_t$ given $\mathbf{r}_{1t}$ is exchangeable (see e.g. Kingman (1978)), which among other things requires that $\boldsymbol{\Omega}$ can be written as in (13), and (ii) $\boldsymbol{\Omega}$ has an approximate zero factor structure as N_2 grows (see Chamberlain and Rothschild (1983)), which requires that $\rho = O(1/N)$ so that the largest eigenvalue of $\boldsymbol{\Omega}$ in (13) is bounded. Bossaerts and Hillion (1995) show that their test, which is effectively focusing on $H_0 : a^* = 0$, is not consistent for fixed T if we rule out limiting arbitrage opportunities, but at least has non-trivial power against admissible alternatives of the form $\mathbf{a} = (\bar{a}^*/\sqrt{N_2})\boldsymbol{\iota}_{N_2}$. As expected, though, their test becomes consistent as $T \to \infty$. However, the application of mean-variance efficiency tests in situations in which N_2/T is not negligible would require not only a different asymptotic theory in which the object of interest is the cross-sectional limit of $\mathbf{a}'\boldsymbol{\Omega}^{-1}\mathbf{a}$, but also the imposition of plausible restrictions on the matrix $\boldsymbol{\Omega}$, with exact or approximate factor structures being the most natural candidates.

5 Asymptotically equivalent tests

Both Jobson and Korkie (1983) and Britten-Jones (1999) suggested to test the mean-variance efficiency of a given portfolio by regressing 1 on $\mathbf{r}_t$. The rationale is that the coefficients of such a projection, $\boldsymbol{\Gamma}^{-1}\boldsymbol{\mu}$, are proportional to the weights of the tangency portfolio, $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, by virtue of the Woodbury formula. In a GMM framework, the moment conditions and parametric restrictions of their proposed test are

$$E(\mathbf{r}_t\mathbf{r}_t'\boldsymbol{\phi}^+ - \mathbf{r}_t) = E[\mathbf{m}_U(\mathbf{r}_t; \boldsymbol{\phi}^+)] = \mathbf{0}, \tag{17}$$

and $H_0 : \boldsymbol{\phi}_2^+ = \mathbf{0}$, respectively. This test is essentially identical to the GMM test of the moment conditions

$$E[\mathbf{r}_t(\varkappa + \psi_1'\mathbf{r}_{1t})] = \mathbf{0}$$

studied by Cochrane (2001) as a test of linear factor pricing models, since in the case of excess returns the choice of $\varkappa$ is arbitrary. Intuitively, Cochrane's moment conditions can be understood as simply saying that under the null there is a stochastic discount factor (SDF) generated from $(1, \mathbf{r}_{1t})$ alone that prices correctly all N assets under consideration.

Peñaranda and Sentana (2004) provide a third interpretation of (17) by using the fact that the arbitrage (i.e. zero-cost) mean variance frontier (AMVF) can be written as

$$r^{MV}(\mu) = \mu \left(\frac{1 + \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}} \right) p^{+},$$

where p^{+} is the (uncentred) mean representing portfolio for arbitrage portfolios, i.e. the arbitrage portfolio that satisfies:

$$E(\mathbf{r}p^{+}) = \boldsymbol{\mu}. \tag{18}$$

Specifically, they show that the test of $H_0 : \boldsymbol{\phi}_2^{+} = \mathbf{0}$ based on (17) can be understood as checking that $\mathcal{A}_{N_1} = \langle \mathbf{r}_1 \rangle$ and $\mathcal{A}_N = \langle \mathbf{r} \rangle$ share the same mean representing portfolio (see also Sentana (2005)).

In this context, we can once more apply the trinity of asymptotic GMM tests, which will again have a limiting chi-square distribution with N_2 degrees of freedom under the null. But since the moment conditions defining $\boldsymbol{\phi}^{*}$ and $\boldsymbol{\phi}^{+}$ are exactly identified, the distance metric test will coincide with the overidentifying restrictions test. In addition, all the tests can be made numerically identical by using a common estimator of the asymptotic covariance matrix of $\sqrt{T}\bar{\mathbf{m}}_{UT}(\boldsymbol{\phi}^{0})$, because both the moment conditions and the restrictions to test are linear in the parameters (see Newey and West (1987)).

Peñaranda and Sentana (2004) also consider an alternative approach based on the centred mean representing portfolio, $Cov(\mathbf{r}, p^{++}) = \mu$, which leads to the moment conditions

$$E\left[\begin{array}{c} \mathbf{r}_t - \boldsymbol{\mu} \\ (\mathbf{r}_t - \boldsymbol{\mu})(\mathbf{r}_t - \boldsymbol{\mu})' \boldsymbol{\varphi}^{+} - \mathbf{r}_t \end{array} \right] = E\left[\begin{array}{c} \mathbf{m}_M(\mathbf{r}_t; \boldsymbol{\mu}) \\ \mathbf{m}_C(\mathbf{r}_t; \boldsymbol{\varphi}^{+}, \boldsymbol{\mu}) \end{array} \right] = E[\mathbf{m}_E(\mathbf{r}_t; \boldsymbol{\varphi}^{+}, \boldsymbol{\mu})] = \mathbf{0}, \tag{19}$$

to test $H_0 : \boldsymbol{\varphi}_2^{+} = \mathbf{0}$. The advantage of working with centred moments is that $\boldsymbol{\varphi}^{+} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, which means that their test can also be regarded as a test based on the most frequent presentation of the weights of the tangency portfolio. In this sense, $\boldsymbol{\varphi}_2^{+} = \mathbf{0}$ means that the tangency portfolio does not involve any asset in $\mathbf{r}_{2t}$. In addition, their test is entirely analogous to the one considered by De Santis (1995) and Bekaert and Urias (1996). Although these authors were interested in assessing the gains to US investors from internationally diversifying their portfolios, they exploited the duality between return mean-variance frontiers and Hansen and Jagannathan (1991) frontiers by basing their tests on the SDF moment conditions

$$E\{\mathbf{r}_t[c + (\mathbf{r}_{1t} - \boldsymbol{\mu}_1)'\boldsymbol{\beta}_1]\} = \mathbf{0},$$

in which the choice of c is arbitrary. In this context, sequential GMM can be successfully applied to (19), and it retains the computational advantage of linearity in $\boldsymbol{\varphi}^{+}$ (see Ogaki (1993)). In addition, since $E[\mathbf{m}_M(\mathbf{r}_t;\boldsymbol{\mu})] = \mathbf{0}$ exactly identifies the nuisance parameter $\boldsymbol{\mu}$, Peñaranda and Sentana (2004) show that SGMM entails no asymptotic efficiency loss.

Therefore, we have three different ways to test for the mean variance efficiency of $\mathbf{r}_{1t}$: centred and uncentred representing portfolios (or portfolio weights), and the GRS regression version. The equivalence between their respective parametric restriction can be easily proved by showing that $\mathbf{a}$ is a full-rank linear transformation of $\boldsymbol{\phi}_2^{+}$, which in turn is proportional to $\boldsymbol{\varphi}_2^{+}$. However, the fact that the restrictions to test are equivalent does not necessarily imply that the corresponding GMM-based test statistics will be equivalent too. This is particularly true in the case of the regression version of the test, in which the number of moments and parameters involved is different, although the number of degrees of freedom is the same.

It turns out, however, that those three families of mean-variance efficiency tests are asymptotically equivalent under the null and sequences of local alternatives, as shown by Peñaranda and Sentana (2004). Therefore, there is no basis to prefer one test to the other from this perspective because all three statistics converge to exactly the same random variable. In this respect, note that this equivalence result is valid as long as the asymptotic distributions of the different tests are standard, which happens under fairly weak assumptions on the distribution of asset returns.

However, such an equivalence is lost under fixed alternatives. But by strengthening the distributional assumptions, Peñaranda and Sentana (2004) prove that if $\mathbf{r}_t$ are independently and identically distributed as an elliptical random vector with mean $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$, and bounded fourth moments, then the approximate slope of the Wald version of the regression test is at least as large as the approximate slope of the Wald version of the centred RP test.

In contrast, it is fairly easy to find parametric configurations for which the approximate slope of the uncentred RP test is either bigger or smaller than the approximate slope of the GMM version of the GRS test. In particular, Peñaranda and Sentana (2004) prove that the uncentred RP test is more powerful than the regression test under normality regardless of the parameter values. Although these results are fairly specific, they can rationalise Monte Carlo results obtained under commonly made assumptions, since the elliptical distributions nest both the multivariate normal and Student t.

Finally, it is worth mentioning that the moment condition (17) and (19), as well the ones used by MacKinlay and Richardson (1991) (see (10)) are exactly identified under the alternative, so that weigthing matrix is asymptotically irrelevant for the unrestricted estimators. Under the null, though, those systems of moment conditions are overidentified, so we may need an initial estimate of the optimal weighting matrix based on a consistent estimator of the parameters. Although the choice of preliminary estimator does not affect the asymptotic distribution of two-step GMM estimators up to $O_p(T^{-1/2})$ terms, there is some Monte Carlo evidence suggesting that their finite sample properties can be negatively affected by an arbitrary choice of initial weighting matrix such as the identity (see e.g. Kan and Zhou (2001)).

For that reason, Peñaranda and Sentana (2004) provide the following useful expressions for first-step, consistent restricted estimators, which are optimal under the assumption that $\mathbf{r}_t$ is independently and identically distributed as an elliptical random vector with mean μ, covariance matrix $\mathbf{\Sigma}$, and bounded coefficient of multivariate excess kurtosis κ (see Mardia (1970)):[12]

Proposition 3 *1. The linear combinations of the moment conditions in (17) that provide the most efficient estimators of $\boldsymbol{\phi}_1^+$ under $H_0 : \boldsymbol{\phi}_2^+ = \mathbf{0}$ will be given by*

$$E(\mathbf{r}_{1t}\mathbf{r}_{1t}'\boldsymbol{\phi}_1^+ - \mathbf{r}_{1t}) = \mathbf{0},$$

so that $\bar{\boldsymbol{\phi}}_1^+ = \hat{\mathbf{\Gamma}}_{11}^{-1}\hat{\boldsymbol{\mu}}_1$.

2. The linear combinations of the moment conditions (19) that provide the most efficient estimators of $\boldsymbol{\varphi}_1^+$ under $H_0 : \boldsymbol{\varphi}_2^+ = \mathbf{0}$ will be given by

$$E\left[\begin{array}{c} \mathbf{r}_{1t} - \boldsymbol{\mu}_1 \\ (\mathbf{r}_{1t} - \boldsymbol{\mu}_1)(\mathbf{r}_{1t} - \boldsymbol{\mu}_1)'\boldsymbol{\varphi}_1^+ - \mathbf{r}_{1t} \end{array}\right] = \mathbf{0},$$

so that $\bar{\boldsymbol{\mu}}_{1T} = \hat{\boldsymbol{\mu}}_1$ and $\bar{\boldsymbol{\varphi}}_1^+ = \hat{\mathbf{\Sigma}}_{11}^{-1}\hat{\boldsymbol{\mu}}_1$, and

3. The linear combinations of the moment conditions (10) that provide the most efficient estimators of $\mathbf{b}$ under $H_0 : \mathbf{a} = \mathbf{0}$ will be given by

$$E[(\mathbf{r}_{1t} + \kappa\boldsymbol{\mu}_1) \otimes (\mathbf{r}_{2t} - \mathbf{B}\mathbf{r}_{1t})] = \mathbf{0}.$$

[12]See Renault (1997) for a result analogous to part 3 in the special case in which the payoffs of the arbitrage portfolios are i.i.d. Gaussian.

In this respect, note that since $\mathbf{\Gamma}^{-1}\boldsymbol{\mu} = (1 + \boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu})^{-1}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}$, $\bar{\boldsymbol{\phi}}_1^+$ and $\bar{\boldsymbol{\varphi}}_1^+$ will be proportional to each other, and the same applies to $\hat{\boldsymbol{\phi}}^+$ and $\hat{\boldsymbol{\varphi}}^+$. However, since the factor of proportionality depends on the data, the Wald tests of $H_0 : \boldsymbol{\phi}_2^+ = \mathbf{0}$ and $H_0 : \boldsymbol{\varphi}_2^+ = \mathbf{0}$ cannot be made numerically identical.

6 More efficient tests

6.1 Tests based on the distribution of $\mathbf{r}_{2t}$ conditional on $\mathbf{r}_{1t}$

The GMM tests discussed in previous sections provide asymptotically valid inferences under fairly weak assumptions on the distribution of returns. However, this robustness may come at the cost of a power loss. In this sense, Hodgson, Linton, and Vorkink (2002; hereinafter HLV) developed a semiparametric estimation and testing methodology that enabled them to obtain optimal mean-variance efficiency tests under the assumption that the distribution of $\mathbf{r}_{2t}$ conditional on $\mathbf{r}_{1t}$ (and their past) is elliptically symmetric. Specifically, HLV showed that their proposed estimators of $\mathbf{a}$ and $\mathbf{b}$ are adaptive under the aforementioned assumptions of linear conditional mean and constant conditional variance, which means that they are as efficient as infeasible maximum likelihood estimators that use the correct parametric elliptical density with full knowledge of its shape parameters. The main advantage of elliptical distributions in this context is that they generalise the multivariate normal distribution, but at the same time they retain its analytical tractability irrespective of the number of assets.

Before discussing their test, though, it is pedagogically convenient to introduce a parametric version, which will be based on the assumption that conditional on $\mathbf{r}_{1t}$ and I_{t-1}, $\boldsymbol{\varepsilon}_t^*$ is independent and identically distributed as a spherical random vector with a well defined density, or $\boldsymbol{\varepsilon}_t^*|r_{Mt}, I_{t-1}; \boldsymbol{\gamma}_0, \boldsymbol{\omega}_0, \boldsymbol{\eta}_0 \sim i.i.d.\ s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta}_0)$ for short, where $\boldsymbol{\eta}$ is the $q \times 1$ vector of shape parameters that determine the distribution of $\varsigma_t = \boldsymbol{\varepsilon}_t^{*\prime}\boldsymbol{\varepsilon}_t^*$. Apart from the normal distribution, another popular and more empirically realistic example is a standardised multivariate t with ν_0 degrees of freedom, or $i.i.d.\ t(\mathbf{0}, \mathbf{I}_N, \nu_0)$ for short. As is well known, the multivariate Student t approaches the multivariate normal as $\nu_0 \to \infty$, but has generally fatter tails. Zhou (1993) and Amengual and Sentana (2009) consider two other illustrative examples: a Kotz distribution and a discrete scale mixture of normals.

Let $\boldsymbol{\phi} = (\boldsymbol{\gamma}', \boldsymbol{\omega}', \boldsymbol{\eta})' \equiv (\boldsymbol{\theta}', \boldsymbol{\eta})'$ denote the $2N_2 + N_2(N_2+1)/2 + q$ parameters of interest, which we assume variation free. The log-likelihood function of a sample of size T based on a particular parametric spherical assumption will take the form $L_T(\boldsymbol{\phi}) = \sum_{t=1}^{T} l_t(\boldsymbol{\phi})$, with $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + c(\boldsymbol{\eta}) + g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$, where $d_t(\boldsymbol{\theta}) = -\frac{1}{2}\ln|\boldsymbol{\Omega}|$ corresponds to the Jacobian, $c(\boldsymbol{\eta})$ to the constant of integration of the assumed density, and $g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$ to its kernel, where $\varsigma_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\theta}) = \boldsymbol{\Omega}^{-1/2}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \boldsymbol{r}_{2t} - \mathbf{a} - \mathbf{B}\mathbf{r}_{1t}$.[13]

Let $\mathbf{s}_t(\boldsymbol{\phi})$ denote the score function $\partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$, and partition it into three blocks, $\mathbf{s}_{\boldsymbol{\gamma}t}(\boldsymbol{\phi})$, $\mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\phi})$, and $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\gamma}$, $\boldsymbol{\omega}$ and $\boldsymbol{\eta}$, respectively. A straightforward application of expression (2) in Fiorentini and Sentana (2007) implies that

$$\mathbf{s}_{\boldsymbol{\gamma}t}(\boldsymbol{\phi}) = \begin{pmatrix} 1 \\ \mathbf{r}_{1t} \end{pmatrix} \otimes \delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}), \tag{20}$$

where

$$\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}] = -2\partial g[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]/\partial\varsigma,$$

which reduces to 1 under Gaussianity (cf. (11)).

Given correct specification, the results in Crowder (1976) imply that the score vector $\mathbf{s}_t(\boldsymbol{\phi})$ evaluated at the true parameter values has the martingale difference property. His results also imply that, under suitable regularity conditions, which typically require that both $\mathbf{r}_{1t}$ and $vech(\mathbf{r}_{1t}\mathbf{r}_{1t}')$ are strictly stationary process with absolutely summable autocovariances, the asymptotic distribution of the feasible ML estimator will be given by the following expression

$$\sqrt{T}\left(\hat{\boldsymbol{\phi}}_{ML} - \boldsymbol{\phi}_0\right) \longrightarrow N\left[\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\phi}_0)\right]$$

where $\mathcal{I}(\boldsymbol{\phi}_0) = E[\mathcal{I}_t(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0]$,

$$\mathcal{I}_t(\boldsymbol{\phi}) = V\left[\mathbf{s}_t(\boldsymbol{\phi})|r_{Mt}, I_{t-1}; \boldsymbol{\phi}\right] = -E\left[\mathbf{h}_t(\boldsymbol{\phi})|r_{Mt}, I_{t-1}; \boldsymbol{\phi}\right],$$

and $\mathbf{h}_t(\boldsymbol{\phi})$ denotes the Hessian function $\partial\mathbf{s}_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}' = \partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$. On this basis, Amengual and Sentana (2009) prove the following result:

Proposition 4 *If $\boldsymbol{\varepsilon}_t^*|r_{Mt}, I_{t-1}; \boldsymbol{\phi}_0$ in (5) is i.i.d. $s(\mathbf{0}, \mathbf{I}_{N_2}, \boldsymbol{\eta}_0)$ with density $\exp[c(\boldsymbol{\eta}) + g(\varsigma_t, \boldsymbol{\eta})]$ such that $\mathrm{M}_{ll}(\boldsymbol{\eta}_0) < \infty$, and both $\mathbf{r}_{1t}$ and $vech(\mathbf{r}_{1t}\mathbf{r}_{1t}')$ are strictly stationary processes with absolutely summable autocovariances, then*

$$\sqrt{T}(\hat{\mathbf{a}}_{ML} - \mathbf{a}_0) \rightarrow N[\mathbf{0}, \mathcal{I}^{\mathbf{aa}}(\boldsymbol{\phi}_0)], \tag{21}$$

[13] Fiorentini, Sentana and Calzolari (2003) provide expressions for $c(\eta)$ and $g_t\left[\varsigma_t(\boldsymbol{\theta}), \eta\right]$ in the multivariate t case, which under normality collapse to $-(N_2/2)\log\pi$ and $-\frac{1}{2}\varsigma_t(\boldsymbol{\theta})$, respectively.

where

$$\begin{aligned}\mathcal{I}^{\mathbf{aa}}(\boldsymbol{\phi}) &= [\mathcal{I}_{\mathbf{aa}}(\boldsymbol{\phi})-\mathcal{I}_{\mathbf{ab}}(\boldsymbol{\phi})\mathcal{I}_{\mathbf{bb}}^{-1}(\boldsymbol{\phi})\mathcal{I}_{\mathbf{ab}}'(\boldsymbol{\phi})]^{-1}=\frac{1}{\mathrm{M}_{ll}(\boldsymbol{\eta})}[1+s^2(r_{p_1})]\boldsymbol{\Omega},\\ \mathrm{M}_{ll}(\boldsymbol{\eta}) &= E\left\{\delta^2[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\frac{\varsigma_t(\boldsymbol{\theta})}{N}\middle|\boldsymbol{\phi}\right\}=E\left\{\frac{2\partial\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\varsigma}\frac{\varsigma_t(\boldsymbol{\theta})}{N}+\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\middle|\boldsymbol{\phi}\right\},\end{aligned}$$

$\boldsymbol{\mu}_1=E(\mathbf{r}_{1t}|\boldsymbol{\phi})$ *and* $\boldsymbol{\Sigma}_{11}=V(\mathbf{r}_{1t}|\boldsymbol{\phi})$*, so that* $s(r_{p_1})=\sqrt{\boldsymbol{\mu}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1}$ *is the maximum Sharpe ratio attainable with the reference portfolios.*

Importantly, expression (21) is valid regardless of whether or not the shape parameters $\boldsymbol{\eta}$ are fixed to their true values $\boldsymbol{\eta}_0$, as in an infeasible ML estimator, $\hat{\mathbf{a}}_{IML}$ say, or jointly estimated with $\boldsymbol{\theta}$, as in an unrestricted one, $\hat{\mathbf{a}}_{UML}$ say. The reason is that the scores corresponding to the mean parameters, $\mathbf{s}_{\boldsymbol{\gamma}t}(\boldsymbol{\phi}_0)$, and the scores corresponding to variance and shape parameters, $\mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\phi}_0)$ and $s_{\boldsymbol{\eta}t}(\boldsymbol{\phi}_0)$, respectively, are asymptotically uncorrelated under the sphericity assumption. The usual asymptotic efficiency properties of maximum likelihood estimators and associated test procedures imply that mean-variance efficiency tests based on this elliptical assumption will be more efficient than those based on the assumption of normality. Specifically, it is easy to see that

$$\mathcal{C}_{\boldsymbol{\alpha\alpha}}(\boldsymbol{\phi}_0)=[1+s^2(r_{p_1})]\boldsymbol{\Omega}_0, \tag{22}$$

which does not depend on the specific distribution for the innovations that we are considering, regardless of whether or not the conditional distribution of $\boldsymbol{\varepsilon}_t^*$ is spherical, as long as it is *i.i.d.* Since $\mathrm{M}_{ll}(\boldsymbol{\eta})\geq 1$, with equality if and only if $\boldsymbol{\varepsilon}_t^*$ is normal, it is clear that the parametric procedure is more efficient than the GMM one.

However, unless one is careful, the elliptically symmetric parametric approach may provide misleading inference if the relevant conditional distribution does not coincide with the assumed one, even if both are elliptical. Nevertheless, Amengual and Sentana (2009) show that the parametric pseudo ML estimator of $\boldsymbol{\gamma}$ that makes the wrong distributional assumption remains consistent in that case. In contrast, the ML estimator of $\boldsymbol{\Omega}$ is only consistent up to scale, in the sense that if reparametrise $\boldsymbol{\Omega}$ as $\tau\boldsymbol{\Upsilon}(\boldsymbol{\upsilon})$, where $\boldsymbol{\upsilon}$ are $N_2(N_2+1)/2-1$ parameters that ensure that $|\boldsymbol{\Upsilon}(\boldsymbol{\upsilon})|=1\ \forall\boldsymbol{\upsilon}$, $\boldsymbol{\upsilon}$ will be consistently estimated but τ will not. They illustrate their results when the pseudo log-likelihood function is based on the multivariate t, in which case the correct asymptotic distribution for the pseudo t-based ML estimator of $\mathbf{a}$ is given by the following expression:

Proposition 5 *If $\boldsymbol{\varepsilon}_t^*|r_{Mt}, I_{t-1}; \boldsymbol{\varphi}_0$ is i.i.d. $s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\varrho}_0)$ but not t with $\kappa_0 > 0$, where $\boldsymbol{\varphi}_0 = (\boldsymbol{\gamma}_0, \boldsymbol{\upsilon}_0, \tau_0, \boldsymbol{\varrho}_0)$, then:*

$$\sqrt{T}\,(\hat{\mathbf{a}}_{UML} - \mathbf{a}_0) \rightarrow N\left[\mathbf{0}, \frac{\mathrm{M}_{ll}^{O}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)}{\lambda_\infty\left[\mathrm{M}_{ll}^{H}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)\right]^2} \cdot \mathcal{C}_{\mathbf{aa}}(\boldsymbol{\varphi}_0)\right], \tag{23}$$

where

$$\begin{aligned}
\mathrm{M}_{ll}^{O}(\boldsymbol{\phi}; \boldsymbol{\varphi}) &= E\left\{\delta^2[\varsigma_t(\boldsymbol{\vartheta}), \eta] \cdot [\varsigma_t(\boldsymbol{\vartheta})/N] \middle|\, \boldsymbol{\varphi}\right\}, \\
\mathrm{M}_{ll}^{H}(\boldsymbol{\phi}; \boldsymbol{\varphi}) &= E\left\{2\partial\delta[\varsigma_t(\boldsymbol{\vartheta}), \eta]/\partial\varsigma \cdot [\varsigma_t(\boldsymbol{\vartheta})/N] + \delta[\varsigma_t(\boldsymbol{\theta}), \eta] \middle|\, \boldsymbol{\varphi}\right\},
\end{aligned}$$

$\lambda_\infty = \tau_0/\tau_\infty$, and τ_∞ is the pseudo-true value of τ.

The analysis of a restricted t-based PML estimator which fixes η to some value $\bar{\eta}$, is entirely analogous, except for the fact that the pseudo-true value of τ becomes $\tau_\infty(\bar{\eta})$, as opposed to $\tau_\infty = \tau_\infty(\eta_\infty)$.[14]

A natural question in this context is a comparison of the efficiency of the t-based pseudo ML estimator and the GMM estimator when the distribution is elliptical but not t. Amengual and Sentana (2009) answer this question by assuming that the conditional distribution is either normal, Kotz, or the two-component scale mixture of normals previously discussed, for which they obtain analytical expressions for the inefficiency ratio $\mathrm{M}_{ll}^{O}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)/\{\lambda_\infty[\mathrm{M}_{ll}^{H}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)]^2\}$. Trivially, they find that if the true conditional distribution is Gaussian, then the restricted ML estimator that makes the erroneous assumption that it is a Student t with $\bar{\eta}^{-1}$ degrees of freedom is inefficient relative to the GMM estimator, the more so the larger the value of $\bar{\eta}$. Nevertheless, this inefficiency becomes smaller and less sensitive to $\bar{\eta}$ as the number of assets increases. But of course $\eta_\infty = 0$ in this case, which suggests that estimating η is clearly beneficial under misspecification. They also find that the restricted t-based PML estimator seems to be strictly more efficient than the GMM one when the true conditional distribution is leptokurtic. And again, they find that as N_2 increases the restricted t-based PML estimator tends to achieve the full efficiency of the ML estimator for any $\bar{\eta} > 0$.

As we mentioned before, HLV proposed a semiparametric estimator of multivariate linear regression models that updates $\hat{\boldsymbol{\theta}}_{GMM}$ (or any other root-T consistent estimator)

[14] When the true distribution is either mesokortic ($\kappa = 0$) or platikurtic ($\kappa < 0$), Amengual and Sentana (2009) show that the t-based pseudo ML estimators will be asymptotically equivalent to the GMM estimators.

by means of a single scoring iteration without line searches. The crucial ingredient of their method is the so-called elliptically symmetric semiparametric efficient score (see Proposition 7 in Fiorentini and Sentana (2007)):

$$\mathring{\mathbf{s}}_{\theta t}(\boldsymbol{\phi}_0)=\mathbf{s}_{\theta t}(\boldsymbol{\phi}_0)-\mathbf{W}_s(\boldsymbol{\phi}_0)\left\{\left[\delta[\varsigma_t(\boldsymbol{\theta}_0),\boldsymbol{\eta}_0]\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N}-1\right]-\frac{2}{(N+2)\kappa_0+2}\left[\frac{\varsigma_t(\boldsymbol{\theta}_0)}{N}-1\right]\right\},$$

where $\mathbf{W}'_s(\boldsymbol{\phi})=[\mathbf{0},\mathbf{0},\frac{1}{2}vec'(\boldsymbol{\Omega}^{-1})\mathbf{D}_{N_2}]$ and $\mathbf{D}_{N_2}$ the duplication matrix of order N_2 (see Magnus and Neudecker (1988)). In fact, the special structure of $\mathbf{W}_s(\boldsymbol{\phi})$ implies that we can update the GMM estimator of $\boldsymbol{\gamma}$ by means of the following simple BHHH correction:

$$\left[\sum_{t=1}^{T}\mathbf{s}_{\gamma t}(\boldsymbol{\phi}_0)\mathbf{s}'_{\gamma t}(\boldsymbol{\phi}_0)\right]^{-1}\sum_{t=1}^{T}\mathbf{s}_{\gamma t}(\boldsymbol{\phi}_0), \tag{24}$$

which does not require the computation of $\mathring{\mathbf{s}}_{\omega t}(\boldsymbol{\phi}_0)$. In practice, of course, $\mathbf{s}_{\gamma t}(\boldsymbol{\phi}_0)$ has to be replaced by a semiparametric estimate obtained from the joint density of $\boldsymbol{\varepsilon}^*_t$. However, the elliptical symmetry assumption allows one to obtain such an estimate from a nonparametric estimate of the univariate density of ς_t, $h\left(\varsigma_t;\boldsymbol{\eta}\right)$, avoiding in this way the curse of dimensionality (see HLV and appendix B1 in Fiorentini and Sentana (2007) for details).

Proposition 7 in Fiorentini and Sentana (2007) shows that the elliptically symmetric semiparametric efficiency bound will satisfy $\mathring{\mathcal{S}}_{\gamma\gamma}(\boldsymbol{\phi}_0)=\mathcal{I}_{\gamma\gamma}(\boldsymbol{\phi}_0)$ in view of the structure of $\mathbf{W}_s(\boldsymbol{\phi}_0)$. This result confirms that the HLV estimator of $\boldsymbol{\gamma}$ is adaptive.[15]

Unfortunately, the HLV approach may also lead to erroneous inferences if the true conditional distribution is asymmetric, and the same is true of the parametric procedure. Amengual and Sentana (2009) illustrate the problem for the case in which $\boldsymbol{\varepsilon}^*_t$ is distributed as an *i.i.d.* multivariate asymmetric t (see Mencía and Sentana (2009b)). In that context, they show that the unrestricted t-based PMLE of $\mathbf{a}$ will be inconsistent. In contrast, $\mathbf{B}$ will be consistently estimated precisely because the estimator of $\mathbf{a}$ will fully mop up the bias in the mean. Unfortunately, mean-variance efficiency tests are based on $\mathbf{a}$, not $\mathbf{B}$.

For analogous reasons, the HLV estimator of $\mathbf{a}$ also becomes inconsistent under asymmetry. Intuitively, the problem is that it will not be true any more that the N_2-dimensional density of $\boldsymbol{\varepsilon}^*_t$ could be written as a function of $\varsigma_t=\boldsymbol{\varepsilon}^{*\prime}_t\boldsymbol{\varepsilon}^*_t$ alone. Therefore, a

[15]HLV also consider alternative estimators that iterate the semiparametric adjustment (24) until it becomes negligible. However, since they have the same first-order asymptotic distribution, we shall not discuss them separately.

semiparametric estimator of $\mathbf{s}_{\gamma t}(\boldsymbol{\phi}_0)$ that combines the elliptical symmetry assumption with a non-parametric specification for $\delta[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}]$ will be contaminated by the skewness of the data. In contrast, the GMM estimator always yields a consistent estimator of $\mathbf{a}$, on the basis of which we can develop a GMM-based Wald test with the correct asymptotic size because (12) remains valid under asymmetry.

Another problem that the semiparametric procedures could have is that their finite sample performance may not be well approximated by the first-order asymptotic theory that justifies them. In this respect, the Monte Carlo evidence presented in Amengual and Sentana (2009) suggests that HLV-based joint and individual tests have systematically the largest size distortions. In contrast, GMM tests have finite sample sizes that are close to the asymptotic levels. As for the tests that use the unrestricted t-based PML estimator, they find that both the robust and non-robust versions are well behaved.

6.2 Tests based on the joint distribution of $\mathbf{r}_{1t}$ and $\mathbf{r}_{2t}$

In this section we explicitly study the framework analysed by MacKinlay and Richardson (1991) and Kan and Zhou (2006), who considered a *joint* distribution of excess returns for the N assets in $\mathbf{r}_t$. Such an assumption is particularly relevant in this context because in the presence of a safe asset a sufficient condition for mean-variance analysis applied to $\mathbf{r}_t$ to be compatible with expected utility maximisation is that the *joint* distribution of $\mathbf{r}_t$ is elliptical (see e.g. Chamberlain (1983), Owen and Rabinovitch (1983) and Berk (1997)). As we mentioned before, when the joint distribution of $\mathbf{r}_t$ is *i.i.d.* Gaussian, the distribution of $\mathbf{r}_{2t}$ conditional on $\mathbf{r}_{1t}$ must also be normal, with a mean $\mathbf{a}+\mathbf{B}\mathbf{r}_{1t}$ that is a linear function of $\mathbf{r}_{1t}$, and a covariance matrix $\boldsymbol{\Omega}$ that does not depend on $\mathbf{r}_{1t}$. However, while the linearity of the conditional mean will be preserved when $\mathbf{r}_t$ is elliptically distributed but non-Gaussian, the conditional covariance matrix will no longer be independent of $\mathbf{r}_{1t}$. For instance, if we assume that $\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\rho})[\mathbf{r}_t - \boldsymbol{\mu}(\boldsymbol{\rho})] \sim$ *i.i.d.* $t(\mathbf{0}, \mathbf{I}_N, \nu)$, where $\boldsymbol{\mu}(\boldsymbol{\rho})$ and $\boldsymbol{\Sigma}(\boldsymbol{\rho})$ are defined in (8) and (9), then

$$\begin{aligned} E\left[\mathbf{r}_{2t}|\mathbf{r}_{1t}; \boldsymbol{\rho}, \nu\right] &= \mathbf{a} + \mathbf{B}\mathbf{r}_{1t}, \\ V\left[\mathbf{r}_{2t}|\mathbf{r}_{1t}; \boldsymbol{\rho}, \nu\right] &= \left(\frac{\nu-2}{\nu+N_1-2}\right)\left[1 + \frac{1}{(\nu-2)}\left(\mathbf{r}_{1t}-\boldsymbol{\mu}_1\right)'\boldsymbol{\Sigma}_{11}^{-1}\left(\mathbf{r}_{1t}-\boldsymbol{\mu}_1\right)\right]\boldsymbol{\Omega}, \end{aligned}$$

which means that model (5) will be misspecified due to contemporaneous, conditionally heteroskedastic innovations. In other words, the variances and covariances of the regression residuals will be a function of the regressor. In addition, note that we can no longer operate the sequential cut of the joint log-likelihood function discussed in section 3, which invalidates the exogeneity of $\mathbf{r}_{1t}$.

As MacKinlay and Richardson (1991) pointed out, the GMM estimator of $\boldsymbol{\gamma}$ remains consistent in this case. In fact, Kan and Zhou (2001) and Amengual and Sentana (2009) show that if $\mathbf{r}_t$ is independently and identically distributed as an elliptical random vector with mean $\boldsymbol{\mu}(\boldsymbol{\rho})$, covariance matrix $\boldsymbol{\Sigma}(\rho)$, and bounded fourth moments, then

$$V(\hat{\mathbf{a}}_{GMM}) = \left[1 + s^2(r_{p_1})\,(1+\kappa_0)\right]\boldsymbol{\Omega}_0 \tag{25}$$

In this sense, note that the only difference with respect to (12) is that the maximum (square) Sharpe ratio of the reference portfolios $s^2(r_{p_1})$ is multiplied by the factor $(1+\kappa_0)$. In practice, we could estimate $V(\hat{\mathbf{a}}_{GMM})$ by using heteroskedastic robust standard errors a la White (1980).

At the other extreme of the efficiency range, we can use Proposition 6 in Amengual and Sentana (2009) to show that

$$V(\hat{\mathbf{a}}_{JML}) = \left[\frac{1}{\mathrm{M}_{ll}(\eta_0)} + \frac{1}{\mathrm{M}_{ss}(\boldsymbol{\eta}_0)}s^2(r_{p_1})\right]\boldsymbol{\Omega}, \tag{26}$$

where $\hat{\mathbf{a}}_{JML}$ denotes the joint ML estimator that makes the correct assumption that $\boldsymbol{\epsilon}_t^*(\boldsymbol{\rho}) = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\rho})[\mathbf{r}_t - \boldsymbol{\mu}(\boldsymbol{\rho})] \sim i.i.d.\ s(\mathbf{0}, \mathbf{I}_N, \boldsymbol{\eta})$, and both $\mathrm{M}_{ll}(\boldsymbol{\eta})$ and

$$\mathrm{M}_{ss}(\boldsymbol{\eta}) = \frac{N}{N+2}\left[1 + V\left\{\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]\frac{\varsigma_t}{N}\middle|\boldsymbol{\phi}\right\}\right] = E\left\{\frac{2\partial\delta[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]}{\partial\varsigma}\frac{\varsigma_t^2(\boldsymbol{\theta})}{N(N+2)}\middle|\boldsymbol{\phi}\right\} + 1$$

correspond to the N-dimensional joint distribution of $\mathbf{r}_t$. This estimator has been proposed by Kan and Zhou (2006) for the case of the multivariate t.

Amengual and Sentana (2009) also prove the consistency of the t-based estimators of $\boldsymbol{\gamma}$ which make the erroneous assumption that $V[\mathbf{r}_{2t}|\mathbf{r}_{1t}] = \tau\boldsymbol{\Upsilon}(\boldsymbol{\upsilon})$, where $\tau = |\boldsymbol{\Omega}|^{1/N_2}$ and $\boldsymbol{\Upsilon}(\boldsymbol{\upsilon}) = \boldsymbol{\Omega}/|\boldsymbol{\Omega}|^{1/N_2}$, and provide expressions for the conditional variance of the score and expected Hessian matrix under such misspecification. Specifically, they show that a sandwich formula analogous to the one in (23) can still be applied to obtain the asymptotic variance of the unrestricted ML estimator. They also quantify the efficiency of the GMM and conditional ML estimator relative to the full information ML estimator when $\mathbf{r}_t$ is distributed as a multivariate t. Their results indicate that the restricted t-based PML estimator of $\boldsymbol{\gamma}$ is more efficient than the GMM estimator for all values of $\bar{\eta}$, the more so the larger N_2 is. Furthermore, the unrestricted t-based PML estimator that also estimates η gets close to achieving the full efficiency of the joint ML estimator, especially for large N_2.

In principle, their results will continue to hold if we replace the t-based ML estimator by any other estimator based on a specific *i.i.d.* elliptical distribution for $\mathbf{r}_{2t}|\mathbf{r}_{1t}, I_{t-1}$. But since the HLV estimator is asymptotically equivalent to a parametric estimator that uses a flexible elliptical distribution as we increase the number of parameters, their results suggest that the HLV estimator of $\boldsymbol{\gamma}$ will continue to be consistent. In fact, an argument analogous to the one made by Hodgson (2000) in a closely related univariate context would imply that the HLV estimator is as efficient as the parametric estimator that used the true *unconditional* distribution of the innovations $\boldsymbol{\varepsilon}_t = \mathbf{r}_{2t} - \mathbf{a}_0 - \mathbf{B}_0\mathbf{r}_{1t}$. Nevertheless, inferences about $\mathbf{a}$ and $\mathbf{B}$ would have to be adjusted to reflect the contemporaneous conditional heteroskedasticity of $\boldsymbol{\varepsilon}_t$, which is not straightforward.

7 Finite sample tests

As we discussed in section 3, one of the nicest features of the GRS test is that it allows us to make exact finite sample inferences conditional on the observations of $\mathbf{r}_{1t}$ for $t = 1, \ldots, T$ under the assumption of conditional normality and homoskedasticity. But since their distributional assumption turns out to be empirically implausible, several studies have analysed the finite sample properties of their tests in more realistic circumstances. In particular, Affleck-Graves and McDonald (1989) found that while the nominal size and power of the GRS test can be seriously misleading if the non-normalities are severe, they are reasonably robust to minor departures from normality (see also MacKinlay (1987), and Zhou (1993), who shows that the finite sample results differ depending on whether the non-normality affects the conditional distribution of $\mathbf{r}_{2t}$ given $\mathbf{r}_{1t}$, or the joint distribution of $\mathbf{r}_{1t}$ and $\mathbf{r}_{2t}$, which is not surprising in view of the discussion in the previous section).

Given that elliptical distributions are natural alternatives to multivariate normality in this context, Zhou (1993) proposed simulation-based p-values for the GRS statistic for a few *fully specified* elliptical distributions, including multivariate t, Kotz and discrete scale mixtures of normals (see also Harvey and Zhou (1991)). Similarly, Gezcy (2001) suggested an adjustment to the F version of the GRS test that has approximately the correct size under the same distributional assumptions.

More recently, Beaulieu, Dufour and Khalaf (2007a) have developed a method to obtain the exact distribution of the Gaussian-based Wald, LR, LM and F versions of the mean-variance efficiency tests described at the beginning of section 3 when the innovations are *i.i.d.* but not necessarily Gaussian or elliptical. For the sake of clarity, let us discuss first the case in which the distribution of the innovations is fully specified, including the nuisance parameters $\boldsymbol{\eta}$. Their approach relies on the fact that in classical multivariate regression models such as (5) the numerical values of the LR, W and LM test of $\mathbf{a} = \mathbf{0}$ depend exclusively on the realisations of the regressors $\mathbf{r}_{1t}$ and innovations $\boldsymbol{\varepsilon}_t^*$ over the full sample $t = 1, \ldots, T$. Consequently, tests of linear hypothesis on the regression coefficients $\mathbf{a}$ are pivotal with respect to the parameters $\mathbf{b}$ and $\boldsymbol{\omega}$ for any finite T. On this basis, one can simulate to any desired degree of accuracy the finite sample distribution of the trinity of classical tests conditional on the full sample realisation of $\mathbf{r}_{1t}$ by generating artificial sample paths of the standardised disturbances $\boldsymbol{\varepsilon}_t^*$ according to some specific *i.i.d.* distribution, such a multivariate t with some fixed degrees of freedom ν_0.[16]

Interestingly, their procedure could also be trivially applied to the Wald, LM and DM versions of the MacKinlay and Richardson (1991) test, as long as one exploits the *i.i.d.* assumption in computing the efficient GMM weighting matrix according to expression (22).

To handle the more realistic situation in which the distribution of the innovations depends on some unknown parameters $\boldsymbol{\eta}$, Beaulieu, Dufour and Khalaf (2007a) exploit the fact that the sample values of the multivariate skewness and kurtosis measures underlying Mardia's (1970) multivariate normality tests are also pivotal with respect to $\mathbf{b}$ and $\boldsymbol{\omega}$ conditional on the full sample realisation of $\mathbf{r}_{1t}$ (see Zhou (1993) and Dufour, Khalaf and Beaulieu (2003)). On this basis, they manage to construct an exact $1 - \alpha_1$ confidence set for the nuisance parameters by "inverting" a simulated moment-based distributional goodness of fit test that they construct by comparing the aforementioned skewness and kurtosis components with their finite sample expectations computed by simulation under the assumed *i.i.d.* distribution for the innovations.[17] Then, they repeat the procedure

[16] In fact, if one is only interested in finding the exact p-value for a given value of the LR statistic say, as opposed to the exact critical values at some pre-specificed level α, the Beaulie, Dufour and Khlaf (2007a) procedure provides the answer with a finite number of simulations.

[17] That is, their $1 - \alpha_1$ confidence level set for $\boldsymbol{\eta}$ is made up by all the values of this parameter for which their distribution goodness of fit test has an exact Monte Carlo p-value less than or equal to α_1.

described in the previous paragraph at a confidence level α_2 for all values of η in the $1-\alpha_1$ confidence set, and report the maximum p-value. Somewhat remarkably, they show that the resulting maximised Monte Carlo p-value has exact level $\alpha_1+\alpha_2$, in the sense that the probability of rejecting the null hypothesis of mean-variance efficiency is not greater than $\alpha_1+\alpha_2$ for any data generating process compatible with the null (see Lehmann (1986, chap. 3)).

Like in the original GRS test, the sampling framework of their tests is one in which the full sample path of the excess returns on the candidate portfolio $\mathbf{r}_{1t}$ is "fixed in repeated samples". Except in the *i.i.d.* normal case, though, it is not clear whether the null distribution of the Beaulieu, Dufour and Khalaf (2007a) tests is in fact independent in finite samples from the values of the regressors.

Despite the fact that it may seem a contradiction in terms, it is interesting to analyse the asymptotic behaviour of their finite sample procedures in order to relate them to the analysis in section 6. Although the exact confidence set for $\boldsymbol{\eta}$ that they construct should become more and more concentrated around the true value $\boldsymbol{\eta}_0$ as $T\to\infty$, let us consider for simplicity the case in which a researcher specifies that the distribution of the innovations is *i.i.d.* t with ν_0 degrees of freedom. Given that the multivariate regression Wald test numerically coincides with a GMM version that exploits the *i.i.d.* assumption in computing the efficient GMM weighting matrix, the asymptotic size and power properties of the Beaulieu, Dufour and Khalaf (2007a) procedure are identical to the asymptotic size and power properties of the GMM tests discussed in section 6.1 as long as the distribution of the innovations is *i.i.d.*, regardless of whether or not they really follow a t with ν_0 degrees of freedom. However, their test will have asymptotically the wrong size if the conditional distribution of the innovations is not *i.i.d.*, and the same is obviously true in finite samples. As we saw in section 6.2, a potentially relevant example would be one in which the joint distribution of $\mathbf{r}_{1t}$ and $\mathbf{r}_{2t}$ were elliptical.

Obviously, standard simulation techniques, such as bootstrap and subsampling methods, can in principle be applied to any of the tests that we have previously discussed, although once again it would important to distinguish the situation in which $\mathbf{r}_{1t}$ is treated as if it were "fixed in repeated samples" from the more realistic situation in which the relevant sampling framework involves all assets in $\mathbf{r}_t$.

In this sense, it is worth remembering that the same exogeneity considerations apply to Bayesian testing methods, such as the ones considered by Shanken (1987b), Harvey and Zhou (1990), Kandel, McCulloch and Stambaugh (1995) or Cremers (2006), which can also be regarded as finite sample methods.

8 Mean-variance-skewness efficiency and spanning tests

Despite its popularity, mean-variance analysis also suffers from important limitations. Specifically, it neglects the effect of higher order moments on asset allocation. In particular, it ignores the third central moment of returns, which as a measure of skewness is undoubtedly a crucial ingredient in analysing derivative assets, games of chance and insurance contracts. In this sense, Patton (2004) uses a bivariate copula model to show the empirical importance of asymmetries in asset allocation. Further empirical evidence has been provided by Harvey et al. (2002) and Jondeau and Rockinger (2006). From the theoretical point of view, Athayde and Flores (2004) derive several useful properties of mean-variance-skewness frontiers, and obtain their shape for some examples by simulation techniques. Similarly, Briec, Kerstens and Jokung (2007) propose an optimisation algorithm that, starting from a specific portfolio, obtains the mean-variance-skewness efficient portfolio along a given direction that reflects investors' relative preferences for those three moments.

From an econometric point of view, it is important to distinguish between testing the mean-variance-skewness efficiency of a particular portfolio, and testing spanning of the mean-variance-skewness frontier.

Let us start with the first test. Using a variational argument, Kraus and Litzenberger (1976) showed that the risk premia of any portfolio could be expressed as a linear combination of its covariance and co-skewness with any mean-variance-skewness efficient portfolio (see also Barone-Adessi (1985), Ingersoll (1987) and Lim (1989)). Specifically, they showed that[18]

$$\mu_i = \tau_r \sigma_{i1} + \tau_s \phi_{i11} \quad \forall i, \tag{27}$$

[18]Strictly speaking, Kraus and Litzenberger (1976) derived a "beta" version of (27), in which σ_{i1} is divided by σ_{11} and ϕ_{i11} by ϕ_{111}, with the appropriate adjustments to τ_r and τ_s An advantage of the formulation in (27) relative to the original one is that it does not require the reference portfolio to be asymmetric.

where

$$\begin{aligned}\sigma_{ij} &= cov(r_i, r_j),\\ \phi_{ijk} &= E[(r_i-\mu_i)(r_j-\mu_j)(r_k-\mu_k)],\end{aligned}$$

and the coefficients τ_r and τ_s are common across assets. These restrictions were cast in a GMM framework by Sánchez-Torres and Sentana (1998) as follows:

$$\begin{aligned}E(r_{1t}-\tau_r\sigma_{11}-\tau_s\phi_{111}) &= 0\\ E[(r_{1t}-\tau_r\sigma_{11}-\tau_s\phi_{111})^2-\sigma_{11}] &= 0\\ E[(r_{1t}-\tau_r\sigma_{11}-\tau_s\phi_{111})^3-\phi_{111}] &= 0\\ E(r_{it}-\tau_r\sigma_{i1}-\tau_s\phi_{i11}) &= 0\\ E[(r_{it}-\tau_r\sigma_{i1}-\tau_s\phi_{i11})(r_{1t}-\tau_r\sigma_{11}-\tau_s\phi_{111})-\sigma_{i1}] &= 0\\ E[(r_{it}-\tau_r\sigma_{i1}-\tau_s\phi_{i11})(r_{1t}-\tau_r\sigma_{11}-\tau_s\phi_{111})^2-\phi_{i11}] &= 0\end{aligned}$$

Note that for each asset except the reference portfolio there are three restrictions but only two parameters, while for the reference portfolio there are four parameters but only three restrictions. All in all, there are $3(N_2+1)$ moment restrictions on $\mathbf{r}$ with $2(N_2+1)+2$ parameters $(\tau_r, \tau_s, \sigma_{i1}, \phi_{i11})$. Therefore, the corresponding overidentification test has N_2-1 degrees of freedom under the null hypothesis of mean-variance-skewness efficiency of r_1, the loss of one degree of freedom relative to the MacKinlay and Richardson (1991) test being due to the addition of the parameter τ_s. As in the case of mean-variance frontiers, the overidentifying test can be made robust to departures from the assumption of normality, conditional homoskedasticity, serial independence or identity of distribution.

Given that (27) would also arise from an asset pricing model in which the SDF were proportional to

$$1-\tau_r(r_{1t}-\mu_1)-\tau_s[r_{1t}^2-(\mu_1^2+\sigma_{11})], \tag{28}$$

we could always interpret a test of $H_0:\tau_s=0$ as a test that (co-)skewness with r_{1t} is not priced.[19] This interpretation also suggests that an alternative test of the mean-variance-skewness efficiency of r_{1t} could be obtained from the SDF-type restrictions:

$$E[r_{it}\{1-\tau_r(r_{1t}-\mu_1)-\tau_s[r_{1t}^2-(\mu_1^2+\sigma_{11})]\}]=0\ \forall i.$$

[19] Chabi-Yo, Leisen and Renault (2007) extend the infinitesimal risk analysis of Samuelson (1970) to provide a justification for a SDF specification such as (28). They also provide an alternative representation of the SDF in terms of r_{1t} and a skewness-representing portfolio, which is the least squares projection of r_{1t}^2 on a constant and $\mathbf{r}_t$.

An econometric problem that arises in this set-up is that σ_{i1} and ϕ_{i11} are highly cross-sectionally collinear in practice (see Barone-Adessi, Gagliardini and Urga (2004)), which makes the separate identification of τ_r and τ_s problematic (see Kan and Zhang (1999a,b) or Kleibergen (2007) for related discussions in more general contexts).

Given the well-known relationship between beta pricing and SDF pricing, Barone-Adessi, Gagliardini and Urga (2004) proposed a "quadratic" regression version of the above problem. Specifically, they showed that if the SDF is a linear combination of r_{1t} and $(R_{1t}^2 - R_{0t})$, then the intercept of the following multivariate regression

$$\mathbf{r}_{2t} = \boldsymbol{\alpha} + \boldsymbol{\beta} r_{1t} + \boldsymbol{\gamma}(R_{1t}^2 - R_{0t}) + \mathbf{v}_t$$

must satisfy the restriction

$$\boldsymbol{\alpha} = \tau_g \boldsymbol{\gamma}, \tag{29}$$

where τ_g is a scalar parameter (see also Barone-Adesi (1985)). However, it is necessary to bear in mind that unless r_{1t} is symmetric, γ_i will not be exactly proportional to the co-skewness of asset i with r_1 even if one makes the additional assumptions that $E(v_{it}|r_{1t}, I_{t-1})$ is 0 and both R_{0t} and $V(v_{it}|r_{1t}, I_{t-1})$ are constant because

$$\phi_{i11} = cov(r_{it}, r_{1t}^2) = \gamma_i V(r_{1t}^2) + \beta_i cov(r_{1t}, r_{1t}^2).$$

As a result, one has to be careful in testing whether co-skewness with r_{1t} is priced (see also Chabi-Yo, Leisen and Renault (2007)). Nevertheless, Barone-Adessi, Gagliardini and Urga (2004) argue that the difference between γ_i and $\phi_{i11}/V(r_{1t}^2)$ is likely to be fairly small in practice when r_{1t} is a well diversified portfolio, since the distribution of such portfolios is strongly leptokurtic but only mildly asymmetric, if at all.[20] More recently, Beaulieu, Dufour and Khalaf (2008) have explained how to obtain by simulation the finite sample size of the Wald and LR test of the non-linear restriction (29) under the assumption that the distribution of $\boldsymbol{\varepsilon}_t$ conditional on I_{t-1} and the past, present and future of r_{1t} is $i.i.d.(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\rho})$.[21]

Notice, though, that like in the case of the mean-variance frontier without a riskless asset, the fact that a portfolio is mean-variance-skewness efficient does not imply that

[20] Sánchez-Torres and Sentana (1998) proposed a moment test of the restriction $E(r_{1t} - \mu_1)^3 = 0$ to assess the asymmetry of the distribution of r_{1t}. The advantage of their test relative to the skewness component of the usual Jarque-Bera (1981) test is that it can be made robust to non-normality, heteroskedasticity and serial correlation (see also Bai and Ng (2005) and Bontempts and Meddahi (2005) for closely related approaches).

[21] In addition, they explicitly consider the more general case in which a riskless asset is not available.

any particular agent would be interested in investing in it. An obvious example is the usual mean-variance tangency portfolio. The properties of the mean-variance frontier imply that such a portfolio will trivially satisfy (27) with $\tau_s = 0$. However, only those agents who do not care about skewness will choose it.

Therefore, from an investors' point of view it may be more interesting to consider mean-variance-skewness spanning tests. The problem with those tests is that in general the mean-variance-skewness frontier is not generated by any finite number of assets. Nevertheless, Mencía and Sentana (2009a) make mean-variance-skewness analysis fully operational by working with a rather flexible family of multivariate asymmetric distributions, known as location-scale mixtures of normals ($LSMN$), which nest as particular cases several important elliptically symmetric distributions, such as the Gaussian or the Student t, and also some well known asymmetric distributions like the Generalised Hyperbolic (GH) introduced by Barndorff-Nielsen (1977). The GH distribution in turn nests many other well known and empirically relevant special cases, such as symmetric and asymmetric versions of the Hyperbolic (Chen, Hardle and Jeong (2008)), Normal Gamma (Madan and Milne (1991)), Normal Inverse Gaussian (Aas, Dimakos and Haff (2005)) or Multivariate Laplace (Cajigas and Urga (2007)). In addition, $LSMN$ nest other interesting examples, such as finite mixtures of normals, which have been shown to be a flexible and empirically plausible device to introduce non-Gaussian features in high dimensional multivariate distributions (see e.g. Kon (1984)), but which at the same time remain analytically tractable.

Formally, a random vector $\mathbf{r}$ of dimension N follows a $LSMN$ if it can be generated as:

$$\mathbf{r} = \boldsymbol{\upsilon} + \xi^{-1}\boldsymbol{\Upsilon}\boldsymbol{\delta} + \xi^{-1/2}\boldsymbol{\Upsilon}^{1/2}\boldsymbol{\varepsilon}^o, \tag{30}$$

where $\boldsymbol{\upsilon}$ and $\boldsymbol{\delta}$ are N-dimensional vectors, $\boldsymbol{\Upsilon}$ is a positive definite matrix of order N, $\boldsymbol{\varepsilon}^o \sim N(\mathbf{0}, \mathbf{I}_N)$, and ξ is an independent positive mixing variable whose distribution function depends on a vector of q shape parameters $\boldsymbol{\varrho}$. Since $\mathbf{r}$ given ξ is Gaussian with conditional mean $\boldsymbol{\upsilon} + \boldsymbol{\Upsilon}\boldsymbol{\delta}\xi^{-1}$ and covariance matrix $\boldsymbol{\Upsilon}\xi^{-1}$, it is clear that $\boldsymbol{\upsilon}$ and $\boldsymbol{\Upsilon}$ play the roles of location vector and dispersion matrix, respectively. The parameters $\boldsymbol{\varrho}$ allow for flexible tail modelling, while the vector $\boldsymbol{\delta}$ introduces skewness in this distribution. For ease of interpretation, Mencía and Sentana (2009) re-write the data generation process for returns as

$$\mathbf{r} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\varepsilon}^{*}, \tag{31}$$

where $\boldsymbol{\varepsilon}^{*}$ is a standardised $LSMN$ vector that is obtained from (30) by choosing $\boldsymbol{\upsilon}$ and $\boldsymbol{\Upsilon}$ appropriately. In addition, they choose

$$\boldsymbol{\delta} = \boldsymbol{\Sigma}^{-1/2}\mathbf{d} \tag{32}$$

in order to make the distribution of $\mathbf{r}$ independent of the particular factorisation of $\boldsymbol{\Sigma}$ in (31).

In terms of portfolio allocation, Mencía and Sentana (2009a) show that if the distribution of asset returns can be expressed as a $LSMN$, then the distribution of any portfolio that combines those assets will be uniquely characterised by its mean, variance and skewness parameter $\mathbf{w}'\boldsymbol{\Sigma}\mathbf{d}$. This implies that, from an investor's point of view, the relative attractiveness of any two portfolios can always be explained in terms of those three quantities because all higher-order moments depend on the lower ones and the common tail parameters $\boldsymbol{\varrho}$. Hence, one only needs to characterise the investment opportunity set in terms of these moments to fully describe the investor's available strategies.

Furthermore, Mencía and Sentana (2009a) show that the efficient part of this frontier can be spanned by three funds: the fund that together with the safe asset generates the usual mean-variance frontier, whose weights are proportional to $\boldsymbol{\varphi}^{+} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, plus an additional fund whose weights are given by the vector $\mathbf{d}$ in (32). This second vector can be interpreted as an asymmetry-variance efficient portfolio because one can maximise efficiency for a given standard deviation by considering portfolios with weights proportional to $\mathbf{d}$. Consequently, any portfolio in the efficient part of the mean-variance-skewness frontier will be of the type $w_r\boldsymbol{\varphi}^{+} + w_s\mathbf{d}$, where w_r and w_s are two scalars.[22]

On this basis, Mencía and Sentana (2009a) develop a mean-variance-skewness spanning test that jointly assesses whether $\boldsymbol{\varphi}_2^{+} = \mathbf{0}$ and $\mathbf{d}_2 = \mathbf{0}$. Given that they work within a fully parametric framework, their test is based on the asymptotic distribution of the ML estimator of the parameters of the $LSMN$ model. In this regard, they provide

[22]There are other asymmetric distributions that satisfy this property. Specifically, Simaan (1983) studies portfolio allocation when excess returns are the sum of an elliptical random vector and an independent scalar asymmetric variable times a constant vector. Similalry, Mencía and Sentana (2009b) consider a multivariate Hermite expansion of a multivariate normal vector in which asymmetry is a common feature.

analytical expressions for the score by means of the EM algorithm, and explain how to reliably evaluate the information matrix.[23]

9 Conclusions

This paper provides a survey of the econometrics of mean-variance efficiency tests. Starting with the classic F test of Gibbons, Ross and Shanken (1989) and its generalised method of moments version, I analyse the effects of the number of assets and portfolio composition on test power. I then discuss asymptotically equivalent tests based on portfolio weights, and study the trade-offs between efficiency and robustness of using parametric and semiparametric likelihood procedures that assume either elliptical innovations or elliptical returns. After reviewing finite sample tests, I conclude with a discussion of mean-variance-skewness efficiency and spanning tests.

A unifying theme of this survey is that empirical researchers must decide how much a priori knowledge about the degree of inefficiency of the candidate portfolio, its exogeneity, the pattern of the residual covariance matrix or the conditional distribution of asset returns they want to use in order to obtain tests that are either more powerful or have more reliable finite sample distributions. As usual, if they make the wrong a priori assumptions they may inadvertently introduce potential biases in their conclusions. In this sense, it is important that they are aware of and understand those biases, so that they can robustify their inferences. However, it does not necessarily follow that they should systematically rely on "asymptotically robust" procedures whose main justification is based on first-order limiting results if they provide a poor approximation in finite samples.

In any case, there are many important issues that I have unfortunately not considered in the interest of space. In particular, I have not looked at mean-variance efficiency tests when a riskless asset is not available (as in e.g. Gibbons (1982), Kandel (1986), Shanken (1985, 1986), Zhou (1991), Velu and Zhou (1999) and more recently Beaulieu, Dufour and Khalaf (2007b)), in which case the regression should be run in terms of returns instead of excess returns, and the null hypothesis should become $H_0 : \alpha_i = \varpi(1 - \sum_{j=1}^{N_1} b_{ij})\ \forall i$, where ϖ is a scalar parameter representing the expected return of the so-called zero-beta

[23]In principle, one could exploit the non-elliptical nature of the distribution of returns for the only purpose of obtaining more efficient parameter estimates of the mean vector and covariance matrix of returns, as in section 6. As we have just seen, though, mean-variance analysis is generally suboptimal for asymmetric return distributions.

portfolio. As we mentioned before, in those circumstances it is important to distinguish between mean-variance efficiency tests on the one hand, and spanning tests on the other (see Huberman and Kandel (1987), and De Roon and Nijman (2001) for a recent survey), in which the null hypothesis involves restrictions on both intercepts and slopes of the multivariate regression model (5) (see Peñaranda and Sentana (2008a) for a comparison of alternative GMM procedures).

Moreover, I have ignored the effects of transaction costs and short sale constraints on testing for mean-variance analysis, which are discussed in detail by De Roon, Nijman and Werker (2000). Short sale and additivity constraints are particularly relevant in style analysis, which is often used in practice (see Sharpe (1992) for a definition and De Roon, Nijman and ter Horst (2004) for a discussion of the econometric issues).

I have also disregarded the effects of using proxies of the true benchmark portfolios $\mathbf{r}_{1t}$, which is particularly relevant in asset pricing applications in view of the so-called Roll (1977) critique (see Shanken (1987a) and Kandel and Stambaugh (1987)).

There is also an extensive body of literature that looks at the two-pass procedures of Fama and McBeth (1973), which continue to attract substantial attention from practitioners (see Shanken (1992), Shanken and Zhou (2006) and Lewellen, Nagel and Shanken (2007)), and also Cochrane (2001, p. 247) for a re-interpretation of their procedure in cross-sectional and pooled regression contexts in which the estimated regression coefficients $\hat{\mathbf{B}}$ are held constant over the full sample period).

Similarly, there is a growing literature that discusses portfolio selection and its pricing implications taking into account either fourth order moments of the distribution of returns through expansions of general expected utility von Neumann-Morgenstern preferences (see e.g. Dittmar (2002), Jondeau and Rockinger (2006), Guidolin and Timmermann (2008) and Chabi-Yo, Ghysels and Renault (2008)), or a specific parametric class of utility functions (see Gourieroux and Monfort (2005)). Relatedly, Jurczenko, Maillet and Merlin (2006) extend the dual approach in Briec, Kerstens and Jokung (2007) to obtain the portfolio frontier for fourth order moments.

Finally, a very important issue that I have ignored is the fact that nowadays it is widely accepted that asset returns are predictable, if not in mean at least in variance, and that investors can exploit this fact to their advantage by using conditional distributions as opposed to unconditional ones in deciding their portfolio strategies.[24] For instance, an investor can not only choose a passive "buy and hold" portfolio strategy whose weights are fixed over time, but also define a dynamic trading strategy as a function of the volatility level of the stock market, as measured by the VIX, say. Frontiers for such active strategies were introduced by Hansen and Richard (1987), and have been recently revisited by Ferson and Siegel (2001), Abhyankar, Basu and Stremme (2007) and Peñaranda and Sentana (2008b). Hansen and Richard (1987) carefully distinguish between conditional mean variance frontiers, which refer to *conditional* moments of *active* strategies, from unconditional mean-variance frontiers, which bound the first two *unconditional* moments of all conceivable actively managed portfolios. In turn, these unconditional frontiers should not be confused with unconditional mean-variance frontiers for passive portfolios, where by passive we mean portfolios whose weights do not depend on the information available at the time of trading.[25]

In line with most of the existing literature on mean-variance efficiency tests, though, the information that is available at the time of trading has played no explicit role in this paper. In this strict sense, therefore, one could regard the procedures that I have surveyed as tests of passive mean-variance efficiency, although the underlying assets could be portfolios managed according to some specific dynamic strategy. At first sight, it may seem irrelevant to study passive strategies in the presence of conditioning information. However, following Hansen and Richard (1987) and many others, empirical work on unconditional mean-variance frontiers typically relies on passive strategies of managed portfolios such as $\mathbf{r}_t \otimes \mathbf{x}_{t-1}$, where $\mathbf{x}_{t-1}$ is a vector of predictor variables known at time $t-1$, as a way of approximating the complexity of active strategies without running the risk of misspecifying the conditional distribution of asset returns (see chapter 8 in Cochrane (2001) for a justification).

Still, other authors prefer to impose functional form restrictions on the conditional distribution of $\mathbf{r}_t$ given $\mathbf{x}_{t-1}$. In some cases, those restrictions amount to assuming

[24]See Cochrane (2001) for a summary of the empirical evidence on mean predictability, and Sentana (2005) for a recent example of the link between regression forecasts and optimal portfolios.

[25]Peñaranda and Sentana (2008b) also discuss *extended* mean variance frontiers, which correspond to actively managed portfolios whose cost is one on average, but not necessarily one for every possible value of the variables in the information set. In addition, there is a nontrivial connection between mean-variance preferences and return frontiers when investors rely on active strategies. Peñaranda (2008) studies such a connection, showing that different mean-variance preferences lead to different interpretations of the results of portfolio efficiency tests.

that the conditional analogue to the multivariate regression slope and intercepts in (5) linearly depend on $\mathbf{x}_{t-1}$ while $\mathbf{\Omega}$ remains constant (see Beaulieu, Dufour and Khalaf (2007a) or Morales (2009) for recent examples). Alternatively, the conditional regression coefficients and residual covariance matrix may be kept constant, but the conditional means, variances and covariances of $\mathbf{r}_{1t}$ are allowed to change over time (as in Gourieroux, Monfort and Renault (1991)). A third possibility is to assume that the conditional mean of $\mathbf{r}_t$ is linear in $\mathbf{x}_{t-1}$ but the corresponding conditional covariance is constant (see e.g. Ferson and Siegel (2009)).

Such parametric restrictions typically imply that some of the procedures surveyed in the previous sections can be easily adapted. For instance, Beaulieu, Dufour and Khalaf (2007a) test conditional mean-variance efficiency by checking that the coefficients of $\mathbf{x}_{t-1}$ in the regression of $\mathbf{r}_{2t}$ on $\mathbf{x}_{t-1}$ and $\mathbf{r}_{1t} \otimes \mathbf{x}_{t-1}$ are simultaneously 0. Similarly, Property 17 in Gourieroux, Monfort and Renault (1991) implies that under their assumptions $\mathbf{a}'\mathbf{\Omega}^{-1}\mathbf{a}$ also reflects the time-invariant incremental Sharpe ratio that separates the conditional mean-variance frontier generated from $\mathbf{r}_{1t}$ alone from the one generated from both $\mathbf{r}_{1t}$ and $\mathbf{r}_{2t}$, even though the unconditional means of the corresponding maximum conditional Sharpe ratios do not coincide with the maximum Sharpe ratios of constant-weight portfolios discussed in section 2. As a result, a test of $H_0 : \mathbf{a} = \mathbf{0}$ is relevant for both conditional and passive mean-variance frontiers. Finally, Ferson and Siegel (2009) compare the maximum Sharpe ratio of the unconditional Hansen and Richard (1987) frontier for arbitrage portfolios constructed from $\mathbf{r}_{1t}$ alone, with the one generated from both $\mathbf{r}_{1t}$ and $\mathbf{r}_{2t}$. To do so, they exploit a result in Ferson and Siegel (2001) which indicates that the arbitrage portfolio with maximum unconditional Sharpe ratio is given by $[\mu(\mathbf{x}_{t-1})\boldsymbol{\mu}'(\mathbf{x}_{t-1}) + \mathbf{\Sigma}(\mathbf{x}_{t-1})]^{-1}\boldsymbol{\mu}'(\mathbf{x}_{t-1})\mathbf{r}_t$, where $\boldsymbol{\mu}(\mathbf{x}_{t-1})$ and $\mathbf{\Sigma}(\mathbf{x}_{t-1})$ are the mean vector and covariance matrix of the distribution of $\mathbf{r}_t$ given $\mathbf{x}_{t-1}$.

In principle, the procedures described in the earlier sections could also be modified to test for conditional mean variance efficiency for a specific value that the conditioning variables may take at the time of trading. Non-parametric procedures can be developed by localising either with respect to state (as in Wang (2002, 2003) and Kayahan and Stengos (2007)) or with respect to time (as in Lewellen and Nagel (2006); see also Fan, Fan and Jiang (2007) for a combined approach).

All these issues constitute interesting avenues for further research.

References

Aas, K., X.K. Dimakos and I.H. Haff (2005): "Risk estimation using the multivariate normal inverse gaussian distribution", *Journal of Risk* 8, 39-60.

Abhyankar, A., D. Basu, and A. Stremme (2007): Portfolio Efficiency and Discount Factor Bounds with Conditioning Information: An Empirical Study, *Journal of Banking & Finance* 31, 419-437.

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

Affleck-Graves, J., and B. McDonald (1990): "Multivariate tests of asset pricing: the comparative power of alternative statistics", *Journal of Financial and Quantitative Analysis* 25, 163-185.

Amengual, D. and E. Sentana (2009): "A comparison of mean-variance efficiency tests", *Journal of Econometrics*, forthcoming.

Antoine, B. (2008): "Portfolio selection with estimation risk: a test-based approach", mimeo, Simon Fraser University.

Athayde, G. M. de and R.G. Flôres (2004): "Finding a maximum skewness portfolio- a general solution to three-moments portfolio choice", *Journal of Economic Dynamics and Control* 28, 1335-1352.

Bahadur, R. (1960). "Stochastic comparison of tests", *Annals of Mathematical Statistics* 31, 276-295.

Bai, J. and S. Ng (2005): "Tests for skewness, kurtosis, and normality for time series data", *Journal of Business and Economic Statistics* 23, 49-60.

Bartlett, M. (1937): "Properties of sufficiency and statistical tests", *Proc. Roy. Soc.* Ser. A, 160, 268-282.

Barndorff-Nielsen, O.E. (1977): "Exponentially decreasing distributions for the logarithm of particle size", *Proc. R. Soc.* 353, 401-419.

Barone-Adesi, G. (1985): "Arbitrage equilibrium with skewed asset returns", *Journal of Financial and Quantitative Analysis* 20, 299-313.

Barone-Adesi, G., Gagliardini, P. and Urga, G. (2004): "Testing asset pricing models with co-skewness", *Journal of Business and Economic Statistics* 22, 474-485.

Bawa, V.S., Brown, S.J., and Klein, R.W. (1979): *Estimation Risk and Optimal Portfolio*

Choice, North-Holland.

Beaulieu, M.C., J.M. Dufour, and L. Khalaf (2007a): Testing mean-variance efficiency in CAPM with possibly non-gaussian errors: an exact simulation-based approach", *Journal of Business and Economic Statistics* 25, 398-410.

Beaulieu, M.C., J.M. Dufour, and L. Khalaf (2007b): "Testing Black's CAPM with possibly non-Gaussian errors: an exact identification-robust simulation-based approach," mimeo, CIRANO and CIREQ, University of Montréal.

Beaulieu, M.C., J.M. Dufour, and L. Khalaf (2008): "Finite-sample multivariate tests of asset pricing models with coskewness", mimeo, McGill University.

Bekaert, B., and M.S. Urias (1996): "Diversification, integration and emerging market closed-end funds", *Journal of Finance* 51, 835-869.

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

Berndt, E.R. and N.E. Savin. (1977): "Conflict among criteria for testing hypotheses in the multivariate linear regression model", *Econometrica* 45, 1263-1278.

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

Bontemps, C. and N. Meddahi (2005): "Testing normality: a GMM approach", *Journal of Econometrics* 124, 149-186.

Bossaerts, P. and P. Hillion (1995): "Testing the mean-variance efficiency of well-diversified portfolios in large cross-sections", *Annales d'Economie et de Statistique* 40, 93-124.

Britten-Jones, M. (1999): "The sampling error in estimates of mean-variance efficient portfolio weights", *Journal of Finance* 54, 655-671.

Breusch, T.S. (1979): "Conflict among criteria for testing hypotheses: extensions and comments", *Econometrica* 47, 203-207.

Briec , W., K. Kerstens and O. Jokung (2007): "Mean-variance-skewness portfolio performance gauging: a general shortage function and dual approach", *Management Science* 53, 135–149.

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

Cajigas, J. and G. Urga, G. (2007): "Dynamic conditional correlation models with asymmetric multivariate Laplace innovations", Cass Business School Centre for Econometric Analysis Working Paper 08-2007.

Chabi-Yo, F., E. Ghysels and E. Renault (2008). "On portfolio separation theorems with heterogeneous beliefs and attitudes towards risk", mimeo, University of North Carolina.

Chabi-Yo, F., D. Leisen and E. Renault (2007). "Implications of asymmetry risk for portfolio analysis and asset pricing", Bank of Canada Working Paper 2007-47.

Chamberlain, G. (1983). "A characterization of the distributions that imply mean-variance utility functions", *Journal of Economic Theory* 29, 185-201.

Chamberlain, G. (1983): "Funds, factors, and diversification in arbitrage pricing models", *Econometrica* 51, 1305-1323.

Chamberlain, G. and M. Rothschild (1983): "Arbitrage, factor structure, and mean-variance analysis on large asset markets", *Econometrica* 51, 1281-1304.

Chen, Y., W. Härdle and S. Jeong (2008): "Nonparametric risk management with generalised hyperbolic distributions", *Journal of the American Statistical Association* 103, 910-923.

Cochrane, J. (2001). *Asset Pricing*, Princeton: Princeton University Press.

Cremers, K.J.M. (2006): "Multifactor efficiency and Bayesian inference", *Journal of Business* 79, 2951-2998.

Crowder, M.J. (1976). "Maximum likelihood estimation for dependent observations", *Journal of the Royal Statistical Society B* 38, 45-53.

De Roon, F.A. and T.E. Nijman (2001). "Testing for mean-variance spanning: a survey", *Journal of Empirical Finance* 8-2, 111-156.

De Roon, F.A., T.E. Nijman and J.R. ter Horst (2004). "Evaluating style analysis", *Journal of Empirical Finance* 11-1, 29-53.

De Roon, F.A., T.E. Nijman and B.J.M. Werker (2001). "Testing for mean-variance spanning with short sales constraints and transaction costs: the case of emerging markets", *Journal of Finance* 56, 723-744.

De Santis, G. (1995): "Volatility bounds for stochastic discount factors: tests and implications from international financial markets", mimeo, University of Southern California.

Dittmar, R.F. (2002). "Non-linear pricing kernels, kurtosis preference and evidence from the cross section of equity returns", *Journal of Finance* 57, 368-403.

Dufour, J.-M., L. Khalaf 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.

Engle, R. F., D. F. Hendry and J.F. Richard (1983): "Exogeneity", *Econometrica* 51, 277–304.

Errunza, V., K. Hogan and M. Hung (1999). "Can the gains from international diversification be achieved without trading abroad", *Journal of Finance* 54, 2075-2107.

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. and MacBeth, J. (1973): "Risk, return and equilibrium: Empirical tests", *Journal of Political Economy* 81, 607-636

Fan, J., Fan, Y. and Jiang, J. (2007): "Dynamic integration of time- and state-domain methods for volatility estimation", *Journal of the American Statistical Association* 102, 618-631.

Farebrother, R. (1990): "The distribution of a quadratic form in normal variables (Algorithm AS 256.3)", *Applied Statistics* 39, 294–309.

Feller, W. (1971): *An introduction to probability theory and its applications*, Vol. 2, 3rd ed. Wiley.

Ferson, W.A. and A. F. Siegel (2009): "The efficient use of conditioning information in portfolios", *Journal of Finance* 56, 967-982.

Ferson, W.A. and A. F. Siegel (2009): "Testing portfolio efficiency with conditioning information" *Review of Financial Studies*, forthcoming.

Fiorentini, G., E. Sentana, and G. Calzolari (2003). "Maximum likelihood estimation and inference on multivariate conditionally heteroscedastic dynamic regression models with Student t innovations", *Journal of Business and Economic Statistics* 24, 532-546.

Fiorentini, G. and E. Sentana (2007). "On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models", CEMFI Working Paper 0713.

Geczy, C.C. (2001): "Some generalized tests of mean-variance effciency and multifactor model performance", mimeo, Wharton School.

Geweke, J. (1981). "The approximate slopes of econometric tests", *Econometrica* 49, 1427-1442.

Gibbons, M. (1982): "Multivariate tests of financial models: a new approach", *Journal* *of Financial Economics* 10, 3-27.

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

Gouriéroux, C. and Monfort, A. (1995): *Statistics and Econometric models*, vols. I and II, Cambridge.

Gouriéroux, C. and A. Monfort (2005): "The econometrics of efficient portfolios", *Journal of Empirical Finance* 12, 1-41.

Gouriéroux, C., A. Monfort and E. Renault (1991): "A general framework for factor models", mimeo, INSEE.

Gouriéroux, C., A. Monfort and A. Trognon (1984): "Pseudo maximum likelihood methods: theory", *Econometrica* 52, 681-700.

Groenwold, N. and P. Fraser (2001): "Tests of asset-pricing models: How important is the *i.i.d.* normal assumption", *Journal of Empirical Finance* 8, 427–449.

Guidolin, M. and A. Timmermann (2008): "International asset allocation under regime switching, skew and kurtosis preferences", *Review of Financial Studies* 21, 889-935.

Hansen, L.P. (1982). "Large sample properties of generalized method of moments estimators", *Econometrica* 50, 1029–1054.

Hansen, L.P., and R. Jagannathan (1991): "Implications of security market data for models of dynamic economies", *Journal of Political Economy* 99, 225-262.

Hansen, L.P. and S.F. Richard (1987): "The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models", *Econometrica* 55, 587-613.

Harvey, C. R., J.C. Liechty, M.W. Liechty and P. Müller (2002): "Portfolio selection with higher moments", Duke University Working Paper.

Harvey, C.R. and G. Zhou (1990): "Bayesian inference in asset pricing tests", *Journal of Financial Economics* 26, 221-254.

Harvey, C.R. and G. Zhou (1991): "International asset pricing with alternative distributional specifications", *Journal of Empirical Finance* 1, 107–131.

Hodgson, D.J. (2000). "Unconditional pseudo-maximum likelihood and adaptive estimation in the presence of conditional heterogeneity of unknown form", *Econometric Reviews* 19, 175-206.

Hodgson, D., O. Linton and K. Vorkink (2002). "Testing the capital asset pricing model efficiently under elliptical symmetry: a semiparametric approach", *Journal of Applied Econometrics* 17, 617-639.

Horst, J.R. ter, F.A. de Roon and B.J.M. Werker (2006): "An alternative approach to estimation risk", in L. Renneboog (ed.) *Advances in Corporate Finance and Asset Pricing*, Elsevier.

Huberman, G. and S. Kandel (1987). "Mean-variance spanning", *Journal of Finance* 42, 873-888.

Imhof, J.P. (1961): "Computing the distribution of quadratic forms in normal variables", *Biometrika* 48, 419-426.

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

Jarque, C. M. and A. Bera (1980): "Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals", *Economics Letters* 6, 255-259.

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

Jobson, J.D. and B. Korkie (1983): "Statistical inference in two-parameter portfolio theory with multiple regression software", *Journal of Financial and Quantitative Analysis* 18, 189-197.

Jobson, J.D. and B. Korkie (1985): "Some tests of linear asset pricing with multivariate normality", *Canadian Journal of Administrative Science* 2, 114–138.

Jondeau, E. and M. Rockinger (2006): "Optimal portfolio allocation under higher moments", *European Financial Management* 12, 29-55.

Jurczenko, E., B. Maillet, and P. Merlin (2006): "Hedge funds portfolio selection with higher order moments: a non-parametric mean-variance-skewness-kurtosis efficient frontier", in E. Jurczenko and B. Maillet (eds.), *Multi-moment asset allocation and pricing models*, pp. 51–66. John Wiley & Sons.

Kan, R. and C. Zhang (1999a): "Two-pass tests of asset pricing models with useless factors", *Journal of Finance* 54, 204-235.

Kan, R. and C. Zhang (1999b): "GMM tests of stochastic discount factor models with useless factors", *Journal of Financial Economics* 54, 103-127.

Kan, R. and G. Zhou (2001): "Tests of mean-variance spanning", mimeo, J.M. Olin School of Business, Washington University in St. Louis.

Kan, R. and G. Zhou (2006). "Modeling non-normality using multivariate t: implications for asset pricing", mimeo J.M. Olin School of Business, Washington University in St.Louis.

Kandel, S. (1984): "The likelihood ratio test statistic of mean-variance efficiency without a riskless asset", *Journal of Financial Economics* 13, 575-592.

Kandel, S., and R. Stambaugh (1987): "On correlations and inferences about mean-variance efficiency", *Journal of Financial Economics* 18, 61-90.

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

Kandel, S., R. McCulloch and R. Stambaugh (1995): "Bayesian inference and portfolio inference", *Review of Financial Studies* 8, 1-53.

Kayahan, B. and T. Stengos (2007): "Testing the capital asset pricing model with local maximum likelihood methods", *Mathematical and Computer Modelling* 46, 138-150.

Kingman, J. F. C. (1978): "Uses of exchangeability", *Annals of Probability* 6, 183–197.

Kleibergen, F. (2007): "Test of risk premia in linear factor models", mimeo, Brown University.

Kon, S. J. (1984): "Models of stock returns-a comparison", *Journal of Finance* 39, 147-165.

Kotz, S. (1975). "Multivariate distributions at a cross-road", in G. P. Patil, S. Kotz and J.K. Ord (eds.) *Statistical Distributions in Scientific Work*, vol. I, 247-270, Reidel.

Kraus, A., R. H. Litzenberger (1976): "Skewness preference and the valuation of risky assets", *Journal of Finance* 31 1085–1100.

Lehmann, E. L. (1986): *Testing Statistical Hypotheses* (2nd ed.), Wiley.

Lewellen, J. and S. Nagel (2006): "The conditional CAPM does not explain asset-pricing anomalies", *Journal of Financial Economics* 82, 289-314.

Lim, K.G. (1989): "A new test of the three-moment capital asset pricing model", *Journal of Financial and Quantitative Analysis* 24, 205-216.

Lintner, J. (1965): "The valuation of risky assets and the selection of risky investments in portfolio selection and capital budgets", *Review of Economic and Statistics* 47, 13-37.

Lo, A.W. and A.C. MacKinlay (1990): "Data-snooping biases in tests of financial asset pricing models", *Review of Financial Studies* 3, 431-467.

Lo, A.W. and A.C. MacKinlay (1990): "Data-snooping biases in tests of financial asset pricing models", *Review of Financial Studies* 3, 431-467.

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

MacKinlay, A.C. (1995): "Multifactor models do not explain deviations from the Capital Asset Pricing Model", *Journal of Financial Economics* 38, 3-28.

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

Madan, D. B. and F. Milne (1991): "Option pricing with VG martingale components", *Mathematical Finance* 1, 39-55.

Magnus, J.R. and H. Neudecker (1988). *Matrix Differential Calculus with Applications on Statistics and Econometrics*, Wiley.

Maller, R.A. and Turkington, D.A. (2002): "New light on the portfolio allocation problem", *Mathematical Methods of Operations Research* 56, 501-511.

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

Markowitz, H. (1952): "Portfolio selection", *Journal of Finance* 8, 77-91.

Meloso, D. and Bossaerts, P. (2006): "Portfolio correlation and the power of portfolio efficiency tests", mimeo, Bocconi University.

Mencía, J. and E. Sentana (2009a): "Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation", *Journal of Econometrics*, forthcoming.

Mencía, J. and E. Sentana (2009b): "Distributional tests in multivariate dynamic models with normal and Student t innovations", mimeo, CEMFI.

Morales, L. (2009): "Mean-variance efficiency tests with conditioning information: a comparison", CEMFI Master Thesis 0902.

Mossin, J. (1966): "Equilibrium in a capital asset market", *Econometrica* 35, 768-783.

Newey, W. K. and D. L. McFadden (1994). "Large sample estimation and hypothesis testing", in R. F. Engle and (eds.), *Handbook of Econometrics*, vol. IV, 2111-2245, Elsevier.

Newey, W.K. and K.D. West (1987): "Hypothesis testing with efficient method of moments estimation", *International Economic Review* 28, 777-787.

Ogaki, M. (1993): "Generalized method of moments: econometric applications", in Maddala, G.S. , C.R. Rao and H.D. Vinod, *Handbook of Statistics* vol. 11, 455-488, Elsevier.

Owen, J. and R. Rabinovitch (1983). "On the class of elliptical distributions and their applications to the theory of portfolio choice", *Journal of Finance* 38, 745-752.

Patton, A. J. (2004): "On the out-of-sample importance of skewness and asymmetric dependence for asset allocation", *Journal of Financial Econometrics* 2, 130-168.

Peñaranda, F. (2008): "Understanding portfolio efficiency with conditioning information", FMG Discussion Paper 626.

Peñaranda, F. and E. Sentana (2004): "Tangency tests in return and stochastic discount factor mean-variance frontiers: a unifying approach", mimeo, CEMFI.

Peñaranda, F. and E. Sentana (2008a). "Spanning tests in return and stochastic discount factor mean-variance frontiers: a unifying approach", mimeo CEMFI.

Peñaranda, F. and E. Sentana (2008b). "Duality in mean-variance frontiers with conditioning information ", mimeo CEMFI.

Rada, M. and E. Sentana (1997): "The power of mean-variance efficiency tests: portfolio aggregation considerations", mimeo, CEMFI.

Renault, E. (1997): "Econométrie de la Finance: la Méthode des Moments Généralisés", in Y. Simon (ed.), *Encyclopédie des Marchés Financiers*, 330-407, Economica, Paris.

Roll, R.A. (1977): "A critique of the asset pricing theory's tests, Part I: On past and potential testability of the theory", *Journal of Financial Economics* 4, 129-176.

Ross, S.A. (1976): "The arbitrage theory of capital asset pricing", *Journal of Economic Theory* 13, 341-360.

Ross, S.A. (1978): "Mutual fund separation in financial theory - The separating distributions", *Journal of Economic Theory* 17, 254-286.

Samuelson, P. (1970): "The fundamental approximation theorem of portfolio anaysis in terms of mean, variances and higher moments", *Review of Economic Studies* 36, 537-541.

Sánchez-Torres, P.L. and Sentana, E. (1998): "Mean-variance-skewness analysis: an application to risk premia in the Spanish stock market" with P.L. Sánchez Torres, *Investigaciones Económicas* 22, 5-17.

Sentana, E. (2005): Least squares predictions and mean-variance analysis, *Journal of Financial Econometrics* 3, 56-78.

Shanken, J. (1985): "Multivariate tests of the zero-beta CAPM", *Journal of Financial Economics* 14, 327-348.

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

Shanken, J. (1987a): "Multivariate proxies and asset pricing relations: Living with Roll's critique", *Journal of Financial Economics* 18, 91-110.

Shanken, J. (1987b): "A Bayesian approach to testing portfolio efficiency", *Journal of Financial Economics* 19, 195-215.

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, C.R. Rao *Handbook of Statistics 14: Statistical Methods in Finance*, 693-711, Elsevier.

Shanken, J. and G. Zhou (2006): "Estimating and testing beta pricing models: alternative methods and their performance in simulations", *Journal of Financial Economics*, forthcoming.

Sharpe, W.F. (1964): "Capital Asset Prices: A theory of capital market equilibrium under conditions of risk", *Journal of Finance* 19, 425-442.

Sharpe, W.F. (1966): "Mutual fund performance", *Journal of Business* 39, 119-138.

Sharpe, W.F. (1992): "Asset allocation: management style and performance measurement", *Journal of Portfolio Management* Winter 7-19.

Sharpe, W.F. (1994): "The Sharpe ratio", *Journal of Portfolio Management* 21, Fall 49-58.

Simaan, Y. (1993): "Portfolio selection and asset pricing - Three parameter framework", *Management Science* 39, 568-577.

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

Wang, K.Q. (2002): "Nonparametric tests of conditional mean-variance efficiency of a benchmark portfolio", Journal of Empirical Finance 9,133-169.

Wang, K.Q. (2003): "Asset pricing with conditioning information: a new test", *Journal of Finance* 58, 161–196

White, H. (1980). "A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity", *Econometrica*, 45, 817-838.

White, H. (1982): "Maximum likelihood estimation of misspecified models", *Econometrica* 50, 1-25.

Zhou, G. (1991): "Small sample tests of portfolio efficiency", *Journal of Financial Eco-*

nomics 30, 165–191.

Zhou, G. (1993): "Asset-pricing tests under alternative distributions", *Journal of Finance* 48, 1927–1942.

Figure 1: Incremental mean-variance frontiers

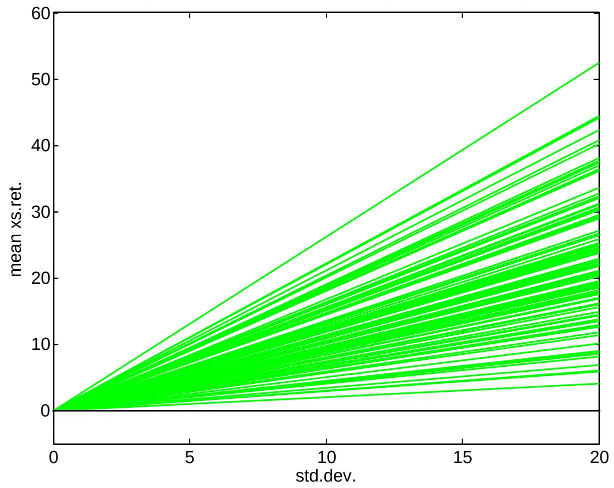

Figure 2: GMM estimator of incremental Sharpe ratio