ANNALS OF ECONOMICS AND FINANCE **13-1**, 139–187 (2012)

Tests of Mean-Variance Spanning

Raymond Kan

Rotman School of Management, University of Toronto, Canada
E-mail: kan@chass.utoronto.ca

and

Guofu Zhou*

Olin Business School, Washington University, St. Louis, USA
E-mail: zhou@wustl.edu

In this paper, we conduct a comprehensive study of tests for mean-variance spanning. Under the regression framework of Huberman and Kandel (1987), we provide geometric interpretations not only for the popular likelihood ratio test, but also for two new spanning tests based on the Wald and Lagrange multiplier principles. Under normality assumption, we present the exact distributions of the three tests, analyze their power comprehensively. We find that the power is most driven by the difference of the global minimum-variance portfolios of the two minimum-variance frontiers, and it does not always align well with the economic significance. As an alternative, we provide a step-down test to allow better assessment of the power. Under general distributional assumptions, we provide a new spanning test based on the generalized method of moments (GMM), and evaluate its performance along with other GMM tests by simulation.

Key Words: Mean-variance spanning; Spanning tests; Portfolio efficiency.
JEL Classification Numbers: G11,G12,C11.

* We thank an anonymous referee, Stephen Brown, Philip Dybvig, Wayne Ferson, Chris Geczy, Gonzalo Rubio Irigoyen, Bob Korkie, Alexandra MacKay, Shuzhong Shi, Tim Simin, seminar participants at Beijing University, Fields Institute, Indiana University, University of California at Irvine, York University, and participants at the 2000 Northern Finance Meetings, the 2001 American Finance Association Meetings, and the Third Annual Financial Econometrics Conference at Waterloo for helpful discussions and comments. Kan gratefully acknowledges financial support from the National Bank Financial of Canada.

1529-7373/2012

1. INTRODUCTION

In portfolio analysis, one is often interested in finding out whether one set of risky assets can improve the investment opportunity set of another set of risky assets. If an investor chooses portfolios based on mean and variance, then the question becomes whether adding a new set of risky assets can allow the investor to improve the minimum-variance frontier from a given set of risky assets. This question was first addressed by Huberman and Kandel (1987, HK hereafter). They propose a multivariate test of the hypothesis that the minimum-variance frontier of a set of K benchmark assets is the same as the minimum-variance frontier of the K benchmark assets plus a set of N additional test assets. Their study has generated many applications and various extensions. Examples include Ferson, Foerster, and Keim (1993), DeSantis (1993), Bekaert and Urias (1996), De Roon, Nijman, and Werker (2001), Korkie and Turtle (2002), Ahn, Conrad, and Dittmar (2003), Jagannathan, Skoulakis, and Wang (2003), Peñaranda and Sentana (2004), Christiansen, Joensen, and Nielsen (2007), and Chen, Chung, Ho and Hsu (2010).

In this paper, we aim at providing a complete understanding of various tests of mean-variance spanning.[1] First, we provide two new spanning tests based on the Wald and Lagrange multiplier principles. The popular HK spanning test is a likelihood ratio test. Unlike the case of testing the CAP-M as in Jobson and Korkie (1982) and Gibbons, Ross, and Shanken (1989, GRS hereafter), this test is in general not the uniformly most powerful invariant test (as shown below), and hence the new tests are of interest. Second, we provide geometrical interpretations of the three tests in terms of the *ex post* minimum-variance frontier of the K benchmark assets and that of the entire $N+K$ assets, which are useful for better economic understanding of the tests. Third, under the normality assumption, we present the small sample distributions for all of the three tests, and provide a complete analysis of their power under alternative hypotheses. We relate the power of these tests to the economic significance of departure from the spanning hypothesis, and find that the power of the tests does not align well with the economic significance of the difference between the two minimum-variance frontiers. Fourth, as an attempt to overcome the power problem, we propose a new step-down spanning test that is potentially more informative than the earlier three tests. Finally, without the normality assumption, we provide a new spanning test based on the generalized method of mo-

[1] We would like to alert readers two common mistakes in applications of the widely used HK test of spanning. The first is that the test statistic is often incorrectly computed due to a typo in HK's original paper. The second is that the HK test is incorrectly used for the single test asset case (i.e., $N = 1$).

ments (GMM). We evaluate its performance along with other GMM tests by simulation. and reach a similar conclusion to the normality case.

The rest of the paper is organized as follows. The next section discusses the spanning hypothesis and the regression based approach for tests of spanning. Section III provides a comprehensive power analysis of various regression based spanning tests. Section IV discusses how to generalize these tests to the case that the assets returns are not multivariate normally distributed. Section V applies various mean-variance spanning tests to examine whether there are benefits of international diversification for a U.S. investor. The final section concludes.

2. REGRESSION BASED TESTS OF SPANNING

In this section, we introduce the various regression-based spanning tests, and provide both their distributions under the null and their geometric interpretations. The Appendix contains the proofs of all propositions and formulas.

2.1. Mean-Variance Spanning

The concept of mean-variance spanning is simple. Following Huberman and Kandel (1987), we say that a set of K risky assets spans a larger set of $N+K$ risky assets if the minimum-variance frontier of the K assets is identical to the minimum-variance frontier of the K assets plus an additional N assets. The first set is often called the benchmark assets, and the second set the test assets. When there exists a risk-free asset and when unlimited lending and borrowing at the risk-free rate is allowed, then investors who care about the mean and variance of their portfolios will only be interested in the tangency portfolio of the risky assets (i.e., the one that maximizes the Sharpe ratio). In that case, the investors are only concerned with whether the tangency portfolio from using K benchmark risky assets is the same as the one from using all $N+K$ risky assets. However, when a risk-free asset does not exist, or when the risk-free lending and borrowing rates are different, then investors will be interested instead in whether the two minimum-variance frontiers are identical. The answer to this question allows us to address two interesting questions in finance. The first question asks whether, conditional on a given set of $N+K$ assets, an investor can maximize his utility by holding just a smaller set of K assets instead of the complete set. This question is closely related to the concept of K-fund separation and has implications for efficient portfolio management. The second question asks whether an investor, conditional on having a portfolio of K assets, can benefit by investing in a new set of N assets. This latter question addresses the benefits of diversification, and is particularly relevant in the context of international portfolio management when the K

benchmark assets are domestic assets whereas the N test assets are investments in foreign markets.

HK first discuss the question of spanning and formalize it as a statistical test. Define $R_t = [R_{1t}', R_{2t}']'$ as the raw returns on $N + K$ risky assets at time t, where R_{1t} is a K-vector of the returns on the K benchmark assets and R_{2t} is an N-vector of the returns on the N test assets.[2] Define the expected returns on the $N + K$ assets as

$$\mu = E[R_t] \equiv \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \tag{1}$$

and the covariance matrix of the $N + K$ risky assets as

$$V = \mathrm{Var}[R_t] \equiv \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \tag{2}$$

where V is assumed to be nonsingular. By projecting R_{2t} on R_{1t}, we have

$$R_{2t} = \alpha + \beta R_{1t} + \epsilon_t, \tag{3}$$

with $E[\epsilon_t] = 0_N$ and $E[\epsilon_t R_{1t}'] = 0_{N\times K}$, where 0_N is an N-vector of zeros and $0_{N\times K}$ is an N by K matrix of zeros. It is easy to show that α and β are given by $\alpha = \mu_2 - \beta\mu_1$ and $\beta = V_{21}V_{11}^{-1}$. Let $\delta = 1_N - \beta 1_K$ where 1_N is an N-vector of ones. HK provide the necessary and sufficient conditions for spanning in terms of restrictions on α and δ as

$$H_0: \quad \alpha = 0_N, \quad \delta = 0_N. \tag{4}$$

To understand why (4) implies mean-variance spanning, we observe that when (4) holds, then for every test asset (or portfolio of test assets), we can find a portfolio of the K benchmark assets that has the same mean (since $\alpha = 0_N$ and $\beta 1_K = 1_N$) but a lower variance than the test asset (since R_{1t} and ϵ_t are uncorrelated and $\mathrm{Var}[\epsilon_t]$ is positive definite). Hence, the N test assets are dominated by the K benchmark assets.

To facilitate later discussion and to gain a further understanding of what the two conditions $\alpha = 0_N$ and $\delta = 0_N$ represent, we consider two portfolios on the minimum-variance frontier of the $N + K$ assets with their weights given by

$$w_1 = \frac{V^{-1}\mu}{1_{N+K}'V^{-1}\mu}, \tag{5}$$

$$w_2 = \frac{V^{-1}1_{N+K}}{1_{N+K}'V^{-1}1_{N+K}}. \tag{6}$$

[2]Note that we can also define R_t as total returns or excess returns (in excess of risk-free lending rate).

From Merton (1972) and Roll (1977), we know that the first portfolio is the tangency portfolio when the tangent line starts from the origin, and the second portfolio is the global minimum-variance portfolio.[3]

Denote $\Sigma = V_{22} - V_{21}V_{11}^{-1}V_{12}$ and $Q = [0_{N\times K},\ I_N]$ where I_N is an N by N identity matrix. Using the partitioned matrix inverse formula, the weights of the N test assets in these two portfolios can be obtained as

$$Qw_1 = \frac{QV^{-1}\mu}{1'_{N+K}V^{-1}\mu} = \frac{[-\Sigma^{-1}\beta,\ \Sigma^{-1}]\mu}{1'_{N+K}V^{-1}\mu} = \frac{\Sigma^{-1}(\mu_2 - \beta\mu_1)}{1'_{N+K}V^{-1}\mu} = \frac{\Sigma^{-1}\alpha}{1'_{N+K}V^{-1}\mu}, \tag{7}$$

and

$$\begin{aligned} Qw_2 &= \frac{QV^{-1}1_{N+K}}{1'_{N+K}V^{-1}1_{N+K}} = \frac{[-\Sigma^{-1}\beta,\ \Sigma^{-1}]1_{N+K}}{1'_{N+K}V^{-1}1_{N+K}} \\ &= \frac{\Sigma^{-1}(1_N - \beta 1_K)}{1'_{N+K}V^{-1}1_{N+K}} = \frac{\Sigma^{-1}\delta}{1'_{N+K}V^{-1}1_{N+K}}. \end{aligned} \tag{8}$$

From these two expressions, we can see that testing $\alpha = 0_N$ is a test of whether the tangency portfolio has zero weights in the N test asset-s, and testing $\delta = 0_N$ is a test of whether the global minimum-variance portfolio has zero weights in the test assets. When there are two distinct minimum-variance portfolios that have zero weights in the N test assets, then by the two-fund separation theorem, we know that every portfolio on the minimum-variance frontier of the $N + K$ assets will also have zero weights in the N test assets.[4]

2.2. Multivariate Tests of Mean-Variance Spanning

To test (4), additional assumptions are needed. The popular assumption in the literature is to assume α and β are constant over time. Under this assumption, α and β can be estimated by running the following regression

$$R_{2t} = \alpha + \beta R_{1t} + \epsilon_t, \qquad t = 1, 2, \ldots, T, \tag{9}$$

where T is the length of time series. HK's regression based approach is to test (4) in regression (9) by using the likelihood ratio test.

For notational brevity, we use the matrix form of model (9) in what follows:

$$Y = XB + E, \tag{10}$$

[3]In defining w_1, we implicitly assume $1'_{N+K}V^{-1}\mu \neq 0$ (i.e., the expected return of the global minimum-variance portfolio is not equal to zero). If not, we can pick the weight of another frontier portfolio to be w_1.

[4]Instead of testing $H_0 : \alpha = 0_N$ and $\delta = 0_N$, we can generalize the approach of Jobson and Korkie (1983) and Britten-Jones (1999) to test directly $Qw_1 = 0_N$ and $Qw_2 = 0_N$.

where Y is a $T \times N$ matrix of R_{2t}, X is a $T \times (K+1)$ matrix with its typical row as $[1,\ R_{1t}']$, $B = [\,\alpha,\ \beta\,]'$, and E is a $T \times N$ matrix with ϵ_t' as its typical row. As usual, we assume $T \geq N+K+1$ and $X'X$ is nonsingular. For the purpose of obtaining exact distributions of the test statistics, we assume that conditional on R_{1t}, the disturbances ϵ_t are independent and identically distributed as multivariate normal with mean zero and variance Σ.[5] This assumption will be relaxed later in the paper.

The likelihood ratio test of (4) compares the likelihood functions under the null and the alternative. The unconstrained maximum likelihood estimators of B and Σ are the usual ones

$$\hat{B} \equiv [\,\hat{\alpha},\ \hat{\beta}\,]' = (X'X)^{-1}(X'Y), \tag{11}$$

$$\hat{\Sigma} = \frac{1}{T}(Y - X\hat{B})'(Y - X\hat{B}). \tag{12}$$

Under the normality assumption, we have

$$\operatorname{vec}(\hat{B}') \sim N(\operatorname{vec}(B'), (X'X)^{-1} \otimes \Sigma), \tag{13}$$

$$T\hat{\Sigma} \sim W_N(T-K-1, \Sigma), \tag{14}$$

where $W_N(T-K-1,\Sigma)$ is the N-dimensional central Wishart distribution with $T-K-1$ degrees of freedom and covariance matrix Σ. Define $\Theta = [\,\alpha,\ \delta\,]'$, the null hypothesis (4) can be written as $H_0 : \Theta = 0_{2\times N}$. Since $\Theta = AB + C$ with

$$A = \begin{bmatrix} 1 & 0_K' \\ 0 & -1_K' \end{bmatrix}, \tag{15}$$

$$C = \begin{bmatrix} 0_N' \\ 1_N' \end{bmatrix}, \tag{16}$$

the maximum likelihood estimator of Θ is given by $\hat{\Theta} \equiv [\,\hat{\alpha},\ \hat{\delta}\,]' = A\hat{B} + C$. Define

$$\hat{G} = TA(X'X)^{-1}A' = \begin{bmatrix} 1 + \hat{\mu}_1'\hat{V}_{11}^{-1}\hat{\mu}_1 & \hat{\mu}_1'\hat{V}_{11}^{-1}1_K \\ \hat{\mu}_1'\hat{V}_{11}^{-1}1_K & 1_K'\hat{V}_{11}^{-1}1_K \end{bmatrix}, \tag{17}$$

where $\hat{\mu}_1 = \frac{1}{T}\sum_{t=1}^{T} R_{1t}$ and $\hat{V}_{11} = \frac{1}{T}\sum_{t=1}^{T}(R_{1t} - \hat{\mu}_1)(R_{1t} - \hat{\mu}_1)'$, it can be verified that

$$\operatorname{vec}(\hat{\Theta}') \sim N(\operatorname{vec}(\Theta'), (\hat{G}/T) \otimes \Sigma). \tag{18}$$

[5]Note that we do not require R_t to be multivariate normally distributed; the distribution of R_{1t} can be time-varying and arbitrary. We only need to assume that conditional on R_{1t}, R_{2t} is normally distributed.

Let $\tilde{\Sigma}$ be the constrained maximum likelihood estimator of Σ and $U = |\hat{\Sigma}|/|\tilde{\Sigma}|$, the likelihood ratio test of $H_0 : \Theta = 0_{2\times N}$ is given by

$$LR = -T\ln(U) \stackrel{A}{\sim} \chi^2_{2N}. \tag{19}$$

It should be noted that, numerically, one does not need to perform the constrained estimation in order to obtain the likelihood ratio test statistic. From Seber (1984, p.410), we have

$$\tilde{\Sigma} - \hat{\Sigma} = \hat{\Theta}'\hat{G}^{-1}\hat{\Theta}, \tag{20}$$

and hence $1/U$ can be obtained from the unconstrained estimate alone as

$$\begin{aligned}\frac{1}{U} &= \frac{|\tilde{\Sigma}|}{|\hat{\Sigma}|} = |\hat{\Sigma}^{-1}\tilde{\Sigma}| = |\hat{\Sigma}^{-1}(\hat{\Sigma} + \hat{\Theta}'\hat{G}^{-1}\hat{\Theta})| \\ &= |I_N + \hat{\Sigma}^{-1}\hat{\Theta}'\hat{G}^{-1}\hat{\Theta}| = |I_2 + \hat{H}\hat{G}^{-1}|,\end{aligned} \tag{21}$$

where

$$\hat{H} = \hat{\Theta}\hat{\Sigma}^{-1}\hat{\Theta}' = \begin{bmatrix} \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} & \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\delta} \\ \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\delta} & \hat{\delta}'\hat{\Sigma}^{-1}\hat{\delta} \end{bmatrix}. \tag{22}$$

Denoting λ_1 and λ_2 as the two eigenvalues of $\hat{H}\hat{G}^{-1}$, where $\lambda_1 \geq \lambda_2 \geq 0$, we have $1/U = (1+\lambda_1)(1+\lambda_2)$. Then, the likelihood ratio test can then be written as

$$LR = T\sum_{i=1}^{2}\ln(1+\lambda_i). \tag{23}$$

The two eigenvalues of $\hat{H}\hat{G}^{-1}$ are of great importance since all invariant tests of (4) are functions of these two eigenvalues (Theorem 10.2.1 of Muirhead (1982)). In order for us to have a better understanding of what λ_1 and λ_2 represent, we present an economic interpretation of these two eigenvalues in the following lemma.

LEMMA 1. *Suppose there exists a risk-free rate r. Let $\hat{\theta}_1(r)$ and $\hat{\theta}(r)$ be the sample Sharpe ratio of the ex post tangency portfolios of the K benchmark asset, and of the $N+K$ assets, respectively. We have*

$$\lambda_1 = \max_r \frac{1+\hat{\theta}^2(r)}{1+\hat{\theta}_1^2(r)} - 1, \qquad \lambda_2 = \min_r \frac{1+\hat{\theta}^2(r)}{1+\hat{\theta}_1^2(r)} - 1. \tag{24}$$

If there were indeed a risk-free rate, it would be natural to measure how close the two frontiers are by comparing the squared sample Sharpe ratios of their tangency portfolios because investors are only interested in the tangency portfolio. However, in the absence of a risk-free rate, it is not entirely clear how we should measure the distance between the two frontiers. This is because the two frontiers can be close over some some region but yet far apart over another region. Lemma 1 suggests that λ_1 measures the maximum difference between the two *ex post* frontiers in terms of squared sample Sharpe ratios (by searching over different values of r), and λ_2 effectively captures the minimum difference between the two frontiers in terms of the squared sample Sharpe ratios.

Besides the likelihood ratio test, econometrically, one can also use the standard the Wald test and Lagrange multiplier tests for almost any hypotheses. As is well known, see. e.g., Berndt and Savin (1977), the Wald test is given by

$$W = T(\lambda_1 + \lambda_2) \overset{A}{\sim} \chi^2_{2N}. \tag{25}$$

and the Lagrange multiplier test is given by

$$LM = T\sum_{i=1}^{2} \frac{\lambda_i}{1+\lambda_i} \overset{A}{\sim} \chi^2_{2N}. \tag{26}$$

Note that although LR, W, and LM all have an asymptotic χ^2_{2N} distribution, Berndt and Savin (1977) and Breusch (1979) show that we must have $W \geq LR \geq LM$ in finite samples.[6] Therefore, using the asymptotic distributions to make an acceptance/rejection decision, the three tests could give conflicting results, with LM favoring acceptance and W favoring rejection.

Note also that unlike the case of testing the mean-variance efficiency of a given portfolio, the three tests are not increasing transformation of each other except for the case of $N = 1$,[7] so they are not equivalent tests in general. It turns out that none of the three tests are uniformly most powerful invariant tests when $N \geq 2$, and which test is more powerful depends on the choice of an alternative hypothesis. Therefore, it is important for us not just to consider the likelihood ratio test but also the other two.

2.3. Small Sample Distributions of Spanning Tests

As demonstrated by GRS and others, asymptotic tests could be grossly misleading in finite samples. In this section, we provide finite sample

[6]The three test statistics can be modified to have better small sample properties. The modified LR statistic is obtained by replacing T by $T-K-(N+1)/2$, the modified W statistic is obtained by replacing T by $T-K-N+1$, and the modified LM statistic is obtained by replacing T by $T-K+1$.

[7]When $N=1$, we have $\lambda_2 = 0$ and hence $LR = T\ln(1+\frac{W}{T})$ and $LM = W/(1+\frac{W}{T})$.

distribution of the three tests under the null hypothesis.[8] Starting with the likelihood ratio test, HK and Jobson and Korkie (1989) show that the exact distribution of the likelihood ratio test under the null hypothesis is given by[9]

$$\left(\frac{1}{U^{\frac{1}{2}}}-1\right)\left(\frac{T-K-N}{N}\right)\sim F_{2N,2(T-K-N)}. \tag{27}$$

Although this F-test has been used to test the spanning hypothesis in the literature for $N=1$, it should be emphasized that this F-test is only valid when $N\geq 2$. When $N=1$, the correct F-test should be

$$\left(\frac{1}{U}-1\right)\left(\frac{T-K-1}{2}\right)\sim F_{2,T-K-1}. \tag{28}$$

In this case, the exact distribution of the Wald and Lagrange multiplier tests can be obtained from the F-test in (28) since all three tests are increasing transformations of each other.

Based on Hotelling (1951) and Anderson (1984), the exact distribution of the Wald test under the null hypothesis is, when $N\geq 2$,

$$\begin{aligned}&P[\lambda_1+\lambda_2\leq w]\\ =\ &I_{\frac{w}{2+w}}(N-1,T-K-N)-\\ &\frac{\mathrm{B}\left(\frac{1}{2},\frac{T-K}{2}\right)}{\mathrm{B}\left(\frac{N}{2},\frac{T-K-N+1}{2}\right)}(1+w)^{-\left(\frac{T-K-N}{2}\right)}I_{\left(\frac{w}{2+w}\right)^2}\left(\frac{N-1}{2},\frac{T-K-N}{2}\right)\end{aligned} \tag{29}$$

where $\mathrm{B}(\cdot,\cdot)$ is the beta function, and $I_x(\cdot,\cdot)$ is the incomplete beta function.

For the exact distribution of the Lagrange multiplier test when $N\geq 2$, there are no simple expressions available in the literature.[10] The simplest expression we have obtained is, for $0\leq v\leq 2$,

$$\begin{aligned}&P\left[\frac{\lambda_1}{1+\lambda_1}+\frac{\lambda_2}{1+\lambda_2}\leq v\right]\\ =\ &I_{\frac{v}{2}}(N-1,T-K-N+1)-\frac{\int_{\max[0,v-1]}^{\frac{v^2}{4}}u^{\frac{N-3}{2}}(1-v+u)^{\frac{T-K-N}{2}}\mathrm{d}u}{2\mathrm{B}(N-1,T-K-N+1)}\end{aligned} \tag{30}$$

[8] The small sample version of the likelihood ratio, the Wald and the Lagrange multiplier tests are known as the Wilks' U, the Lawley-Hotelling trace, and the Pillai trace, respectively, in the multivariate statistics literature.

[9] HK's expression of the F-test contains a typo. Instead of $U^{\frac{1}{2}}$, it was misprinted as U. This mistake was unfortunately carried over, to our knowledge, to all later studies such as Bekaert and Urias (1996) and Errunza, Hogan, and Hung (1999), with the exception of Jobson and Korkie (1989).

[10] Existing expressions in Mikhail (1965) and Pillai and Jayachandran (1967) require summing up a large number of terms and only work for the special case that both N and $T-K$ are odd numbers.

Unlike that for the Wald test, this formula requires the numerical computation of an integral, which can be done using a suitable computer program package.

Under the null hypothesis, the exact distributions of all the three tests depend only on N and $T-K$, and are independent of the realizations of R_{1t}. Therefore, under the null hypothesis, the unconditional distributions of the test statistics are the same as their distributions when conditional on R_{1t}. In Table 1, we provide the actual probabilities of rejection of the three tests under the null hypothesis when the rejection is based on the 95% percentile of their asymptotic χ^2_{2N} distribution. We see that the actual probabilities of rejection can differ quite substantially from the asymptotic p-value of 5%, especially when N and K are large relative to T. For example, when $N = 25$, even when T is as high as 240, the probabilities of rejection can still be two to four times the size of the test for the Wald and the likelihood ratio tests. Therefore, using asymptotic distributions could lead to a severe over-rejection problem for the Wald and the likelihood ratio tests. For the Lagrange multiplier test, the actual probabilities of rejection are actually quite close to the size of the test, except when T is very small. If one wishes to use an asymptotic spanning test, the Lagrange multiplier test appears to be preferable to the other two in terms of the size of the test.

2.4. The Geometry of Spanning Tests

While it is important to have finite sample distributions of the three tests, it is equally important to develop a measure that allows one to examine the economic significance of departures from the null hypothesis. Fortunately, all three tests have nice geometrical interpretations. To prepare for our presentation of the geometry of the three test statistics, we introduce three constants $\hat{a} = \hat{\mu}'\hat{V}^{-1}\hat{\mu}$, $\hat{b} = \hat{\mu}'\hat{V}^{-1}1_{N+K}$, $\hat{c} = 1'_{N+K}\hat{V}^{-1}1_{N+K}$, where $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t$ and $\hat{V} = \frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})'$. It is well known that these three constants determine the location of the *ex post* minimum-variance frontier of the $N + K$ assets. Similarly, the corresponding three constants for the mean-variance efficiency set of just the K benchmark assets are $\hat{a}_1 = \hat{\mu}_1'\hat{V}_{11}^{-1}\hat{\mu}_1$, $\hat{b}_1 = \hat{\mu}_1'\hat{V}_{11}^{-1}1_K$, $\hat{c}_1 = 1'_K\hat{V}_{11}^{-1}1_K$. Using these constants, we can write

$$\hat{G} = \begin{bmatrix} 1 + \hat{a}_1 & \hat{b}_1 \\ \hat{b}_1 & \hat{c}_1 \end{bmatrix}. \tag{31}$$

The following lemma relates the matrix $\hat{H}$ to these two sets of efficiency constants.

TABLE 1.

Sizes of Three Asymptotic Tests of Spanning Under Normality

			Actual Probabilities of Rejection		
K	N	T	W	LR	LM
2	2	60	0.078	0.063	0.048
		120	0.063	0.056	0.049
		240	0.056	0.053	0.050
	10	60	0.249	0.125	0.037
		120	0.126	0.080	0.044
		240	0.082	0.063	0.047
	25	60	0.879	0.500	0.015
		120	0.422	0.185	0.033
		240	0.183	0.099	0.042
5	2	60	0.094	0.076	0.059
		120	0.069	0.062	0.054
		240	0.059	0.056	0.052
	10	60	0.315	0.172	0.058
		120	0.146	0.095	0.054
		240	0.089	0.069	0.052
	25	60	0.932	0.638	0.038
		120	0.479	0.229	0.047
		240	0.203	0.113	0.049
10	2	60	0.126	0.105	0.084
		120	0.081	0.073	0.064
		240	0.064	0.060	0.057
	10	60	0.446	0.279	0.118
		120	0.186	0.126	0.075
		240	0.103	0.081	0.061
	25	60	0.981	0.838	0.146
		120	0.579	0.315	0.082
		240	0.238	0.138	0.063

The table presents the actual probabilities of rejection of three asymptotic tests of spanning (Wald (W), likelihood ratio (LR), and Lagrange multiplier (LM)), under the null hypothesis for different values of number of benchmark assets (K), test assets (N), and time series observations (T). The asymptotic p-values of all three tests are set at 5% based on the asymptotic distribution of χ^2_{2N} and the actual p-values reported in the table are based on their finite sample distributions under normality assumption.

Lemma 2. *Let $\Delta\hat{a} = \hat{a} - \hat{a}_1$, $\Delta\hat{b} = \hat{b} - \hat{b}_1$, and $\Delta\hat{c} = \hat{c} - \hat{c}_1$, we have*

$$\hat{H} = \begin{bmatrix} \Delta\hat{a} & \Delta\hat{b} \\ \Delta\hat{b} & \Delta\hat{c} \end{bmatrix}. \tag{32}$$

Since $\hat{H}$ summarizes the marginal contribution of the N test assets to the efficient set of the K benchmark assets, Jobson and Korkie (1989) call this matrix the "marginal information matrix." With this lemma, we have

$$\begin{aligned} U &= \frac{1}{|I_2 + \hat{H}\hat{G}^{-1}|} = \frac{|\hat{G}|}{|\hat{G}+\hat{H}|} = \frac{(1+\hat{a}_1)\hat{c}_1 - \hat{b}_1^2}{(1+\hat{a})\hat{c} - \hat{b}^2} \\ &= \frac{\hat{c}_1 + \hat{d}_1}{\hat{c}+\hat{d}} = \left(\frac{\hat{c}_1}{\hat{c}}\right)\left(\frac{1+\frac{\hat{d}_1}{\hat{c}_1}}{1+\frac{\hat{d}}{\hat{c}}}\right), \end{aligned} \tag{33}$$

where $\hat{d} = \hat{a}\hat{c} - \hat{b}^2$ and $\hat{d}_1 = \hat{a}_1\hat{c}_1 - \hat{b}_1^2$. Therefore, the F-test of (27) can be written as

$$\begin{aligned} F &= \left(\frac{T-K-N}{N}\right)\left(\frac{1}{U^{\frac{1}{2}}} - 1\right) \\ &= \left(\frac{T-K-N}{N}\right)\left[\left(\frac{\sqrt{\hat{c}}}{\sqrt{\hat{c}_1}}\right)\left(\frac{\sqrt{1+\frac{\hat{d}}{\hat{c}}}}{\sqrt{1+\frac{\hat{d}_1}{\hat{c}_1}}}\right) - 1\right]. \end{aligned} \tag{34}$$

In Figure 1, we plot the *ex post* minimum-variance frontier of the K benchmark assets as well as the frontier for all $N+K$ assets in the $(\hat{\sigma}, \hat{\mu})$ space. Denote g_1 the *ex post* global minimum-variance portfolio of the K assets and g the *ex post* global minimum-variance portfolio of all $N+K$ assets. It is well known that the standard deviation of g_1 and g are $1/\sqrt{\hat{c}_1}$ and $1/\sqrt{\hat{c}}$, respectively. Therefore, the first ratio $\sqrt{\hat{c}}/\sqrt{\hat{c}_1}$ is simply the ratio of the standard deviation of g_1 to that of g, and this ratio is always greater than or equal to one. To obtain an interpretation of the second ratio $\sqrt{1+\frac{\hat{d}}{\hat{c}}}\Big/\sqrt{1+\frac{\hat{d}_1}{\hat{c}_1}}$, we note that the absolute value of the slopes of the asymptotes to the efficient set hyperbolae of the K benchmark assets and of all $N+K$ assets are $\sqrt{\hat{d}_1/\hat{c}_1}$ and $\sqrt{\hat{d}/\hat{c}}$, respectively. Therefore, $\sqrt{1+\frac{\hat{d}_1}{\hat{c}_1}}$ is the length of the asymptote to the hyperbola of the K benchmark assets from $\hat{\sigma}=0$ to $\hat{\sigma}=1$, and $\sqrt{1+\frac{\hat{d}}{\hat{c}}}$ is the corresponding length of the asymptote to the hyperbola of the $N+K$ assets. Since the *ex post* frontier of the $N+K$ assets dominates the *ex post* frontier of the K benchmark assets, the ratio $\sqrt{1+\frac{\hat{d}}{\hat{c}}}\Big/\sqrt{1+\frac{\hat{d}_1}{\hat{c}_1}}$ must be greater than or equal to one. In Figure 1, we can see that for $N>1$, the F-test of (27) can be geometrically represented as[11]

$$F = \left(\frac{T-K-N}{N}\right)\left[\left(\frac{OD}{OC}\right)\left(\frac{AH}{BF}\right)-1\right]. \qquad (36)$$

Under the null hypothesis, the two minimum-variance frontiers are *ex ante* identical, so the two ratios $\sqrt{\hat{c}}/\sqrt{\hat{c}_1}$ and $\sqrt{1+\frac{\hat{d}}{\hat{c}}}\Big/\sqrt{1+\frac{\hat{d}_1}{\hat{c}_1}}$ should be close to one and the F-statistic should be close to zero. When either g_1 is far enough from g or the slopes of the asymptotes to the two hyperbolae are very different, we get a large F-statistic and we will reject the null hypothesis of spanning.

For the Wald and the Lagrange multiplier tests, mean-variance spanning is tested by examining different parts of the two minimum-variance frontiers. To obtain a geometrical interpretation of these two test statistics, we define $\hat{\theta}_1(r)$ and $\hat{\theta}(r)$ as the slope of the tangent lines to the sample frontier of the K benchmark assets and of all $N+K$ assets, respectively, when the tangent lines have a y-intercept of r. Also denote $\hat{\mu}_{g_1} = \hat{b}_1/\hat{c}_1$ and $\hat{\mu}_g = \hat{b}/\hat{c}$ as the sample mean of the *ex post* global minimum-variance portfolio of the K benchmark assets and of all $N+K$ assets, respectively. Using these definitions, the Wald and Lagrange multiplier tests can be represented geometrically as[12]

$$\lambda_1+\lambda_2 = \frac{\hat{c}-\hat{c}_1}{\hat{c}_1} + \frac{\hat{\theta}^2(\hat{\mu}_{g_1})-\hat{\theta}_1^2(\hat{\mu}_{g_1})}{1+\hat{\theta}_1^2(\hat{\mu}_{g_1})} = \left(\frac{OD}{OC}\right)^2 - 1 + \left(\frac{BE}{BF}\right)^2 - 1 \quad (37)$$

and

$$\frac{\lambda_1}{1+\lambda_1}+\frac{\lambda_2}{1+\lambda_2} = \frac{\hat{c}-\hat{c}_1}{\hat{c}} + \frac{\hat{\theta}^2(\hat{\mu}_g)-\hat{\theta}_1^2(\hat{\mu}_g)}{1+\hat{\theta}^2(\hat{\mu}_g)} = 1-\left(\frac{OC}{OD}\right)^2 + 1 - \left(\frac{AG}{AH}\right)^2. \qquad (38)$$

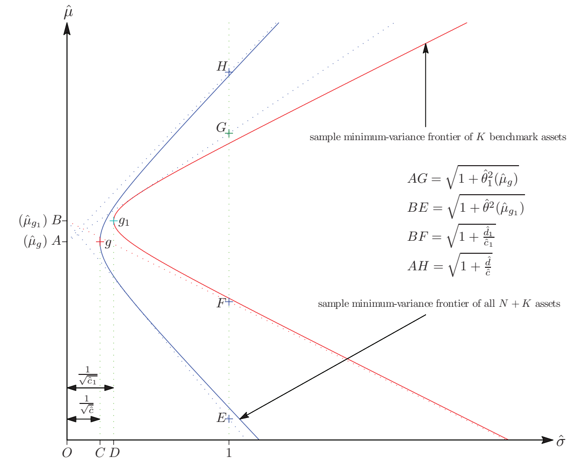
From these two expressions and Figure 1, we can see that both the Wald and the Lagrange multiplier test statistics are each the sum of two quantities. The first quantity measures how close the two *ex post* global minimum-variance portfolios g_1 and g are, and the second quantity measures how close together the two tangency portfolios are. However, there is a subtle

[11] For $N=1$, the F-test of (28) can be geometrically represented as

$$F = \left(\frac{T-K-1}{2}\right)\left[\left(\frac{OD}{OC}\right)^2\left(\frac{AH}{BF}\right)^2-1\right]. \qquad (35)$$

[12] Note that $\hat{\theta}_1^2(\hat{\mu}_{g_1}) = \hat{d}_1/\hat{c}_1$ and $\hat{\theta}^2(\hat{\mu}_g) = \hat{d}/\hat{c}$ and they are just the square of the slope of the asymptote to the efficient set hyperbolae of the K benchmark assets and of all $N+K$ assets, respectively.

FIG. 1. The Geometry of Mean-Variance Spanning Tests



The figure plots the *ex post* minimum-variance frontier hyperbola of K benchmark assets and that of all $N+K$ assets on the $(\hat{\sigma}, \hat{\mu})$ space. The constants that determine the hyperbola of K benchmark assets are $\hat{a}_1 = \hat{\mu}_1' \hat{V}_{11} \hat{\mu}_1$, $\hat{b}_1 = \hat{\mu}_1' \hat{V}_{11} 1_K$, $\hat{c}_1 = 1_K' \hat{V}_{11} 1_K$, and $\hat{d}_1 = \hat{a}_1 \hat{c}_1 - \hat{b}_1^2$, where $\hat{\mu}_1$ and $\hat{V}_{11}$ are maximum likelihood estimates of the expected return and covariance matrix of the K benchmark assets. The constants that determine the hyperbola of all $N+K$ assets are $\hat{a} = \hat{\mu}' \hat{V} \hat{\mu}$, $\hat{b} = \hat{\mu}' \hat{V} 1_{N+K}$, $\hat{c} = 1_{N+K}' \hat{V} 1_{N+K}$, and $\hat{d} = \hat{a}\hat{c} - \hat{b}^2$, where $\hat{\mu}$ and $\hat{V}$ are maximum likelihood estimates of the expected return and covariance matrix of all $N+K$ assets. Portfolios g_1 and g are the *ex post* global minimum-variance portfolios of the two frontiers. The dotted line going through BF is one of the asymptotes to the hyperbola of K benchmark assets. It has slope $-\sqrt{\frac{\hat{d}_1}{\hat{c}_1}}$ and the distance BF is $\sqrt{1+\frac{\hat{d}_1}{\hat{c}_1}}$. The dotted line going through AH is one of the asymptotes to the hyperbola of all $N+K$ assets. It has slope $\sqrt{\frac{\hat{d}}{\hat{c}}}$ and the distance AH is $\sqrt{1+\frac{\hat{d}}{\hat{c}}}$. The distance AG is $\sqrt{1+\hat{\theta}_1^2(\hat{\mu}_g)}$ where $\hat{\theta}_1(\hat{\mu}_g)$ is the slope of the tangent line to the frontier of the K benchmark assets when the y-intercept of the tangent line is $\hat{\mu}_g$. The distance BE is $\sqrt{1+\hat{\theta}^2(\hat{\mu}_{g_1})}$ where $\hat{\theta}(\hat{\mu}_{g_1})$ is the slope of the tangent line to the frontier of all $N+K$ assets when the y-intercept of the tangent line is $\hat{\mu}_{g_1}$.

difference between the two test statistics. For the Wald test, g_1 is the reference point and the test measures how close the sample frontier of the

$N + K$ assets is to g_1 in terms of the increase in the variance of going from g to g_1, and in terms of the improvement of the square of the slope of the tangent line from introducing N additional test assets, with $\hat{\mu}_{g_1}$ as the y-intercept of the tangent line. For the Lagrange multiplier test, g is the reference point and the test measures how close the sample frontier of the K assets is to g in terms of the reduction in the variance of going from g_1 to g, and in terms of the reduction of the square of the slope of the tangent line when using only K benchmark assets instead of all the assets, with $\hat{\mu}_g$ as the y-intercept of the tangent line. Such a difference is due to the Wald test being derived under the unrestricted model but the Lagrange multiplier test being derived under the restricted model.

3. POWER ANALYSIS OF SPANNING TESTS

3.1. Single Test Asset

In the mean-variance spanning literature, there are many applications and studies of HK's likelihood ratio test. However, not much has been done on the power of this test. GRS consider the lack of power analysis as a drawback of HK test of spanning. Since the likelihood ratio test is not in general the uniformly most powerful invariant test, it is important for us to understand the power of all three tests.

We should first emphasize that although in finite samples we have the inequality $W \geq LR \geq LM$, this inequality by no means implies the Wald test is more powerful than the other two. This is because the inequality holds even when the null hypothesis is true. Hence, the inequality simply suggests that the tests have different sizes when we use their asymptotic χ^2_{2N} distribution. In evaluating the power of these three tests, it is important for us to ensure that all of them have the correct size under the null hypothesis. Therefore, the acceptance/rejection decisions of the three tests must be based on their exact distributions but not on their asymptotic χ^2_{2N} distribution. It also deserves emphasis that the distributions of the three tests under the alternative are conditional on $\hat{G}$, i.e., conditional on the realizations of the *ex post* frontier of K benchmark assets. Thus, similar to GRS, we study the power functions of the three tests conditional on a given value of $\hat{G}$, not the unconditional power function.

When there is only one test asset (i.e., $N = 1$), all three tests are increasing transformations of the F-test in (28). For this special case, the power analysis is relatively simple to perform because it can be shown that this F-test has the following noncentral F-distribution under the alternative hypothesis

$$\left(\frac{1}{U} - 1\right)\left(\frac{T - K - 1}{2}\right) \sim F_{2,T-K-1}(T\omega), \tag{39}$$

where $T\omega$ is the noncentrality parameter and $\omega = (\Theta'\hat{G}^{-1}\Theta)/\sigma^2$, with σ^2 representing the variance of the residual of the test asset. Geometrically, ω can be represented as[13]

$$\omega = \left[\frac{c - c_1}{\hat{c}_1} + \frac{\theta^2(\hat{\mu}_{g_1}) - \theta_1^2(\hat{\mu}_{g_1})}{1 + \hat{\theta}_1^2(\hat{\mu}_{g_1})}\right], \tag{40}$$

where $c_1 = 1_K' V_{11}^{-1} 1_K$ and $c = 1_{N+K}' V^{-1} 1_{N+K}$ are the population counterparts of the efficient set constants $\hat{c}_1$ and $\hat{c}$, and $\theta_1(\hat{\mu}_{g_1})$ and $\theta(\hat{\mu}_{g_1})$ are the slope of the tangent lines to the *ex ante* frontiers of the K benchmark assets, and of all $N + K$ assets, respectively, with the y-intercept of the tangent lines as $\hat{\mu}_{g_1}$.

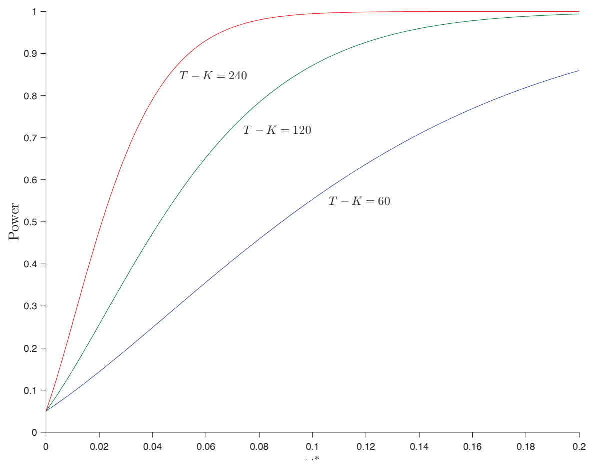
In Figure 2, we present the power of the F-test as a function of $\omega^* = T\omega/(T - K - 1)$ for $T - K = 60$, 120, and 240, when the size of the test is 5%. It shows that the power function of the F-test is an increasing function of $T - K$ and ω^* and this allows us to determine what level of ω^* that we need to reject the null hypothesis with a certain probability. For example, if we wish the F-test to have at least a 50% probability of rejecting the spanning null hypothesis, then we need ω^* to be greater than 0.089 for $T - K = 60$, 0.043 for $T - K = 120$, and 0.022 for $T - K = 240$.

Note that ω is the sum of two terms. The first term measures how close the *ex ante* global minimum-variance portfolios of the two frontiers are in terms of the reciprocal of their variances. The second term measures how close the *ex ante* tangency portfolios of the two frontiers are in terms of the square of the slope of their tangent lines.

In determining the power of the test, the distance between the two global minimum-variance portfolios is in practice a lot more important than the distance between the two tangency portfolios. We provide an example to illustrate this. Consider the case of two benchmark assets (i.e., $K = 2$), chosen as the equally weighted and value-weighted market portfolio of the NYSE.[14] Using monthly returns from 1926/1–2006/12, we estimate $\hat{\mu}_1$ and $\hat{V}_{11}$ and we have $\hat{\mu}_{g_1} = \hat{b}_1/\hat{c}_1 = 0.0074$, $\hat{\sigma}_{g_1} = 1/\sqrt{\hat{c}_1} = 0.048$, and $\hat{\theta}_1(\hat{\mu}_{g_1}) = 0.0998$. We plot the *ex post* minimum-variance frontier of these two benchmark assets in Figure 3. Suppose we take this frontier as the *ex ante* frontier of the two benchmark assets and consider the power of the F-test for two different cases. In the first case, we consider a test asset that slightly reduces the standard deviation of the global minimum-variance portfolio from 4.8%/month to 4.5%/month. This case is represented by the dotted frontier in Figure 3. Although geometrically this asset does not improve the opportunity set of the two benchmark assets by much, the

[13]The derivation of this expression is similar to that of (37) and therefore not provided.
[14]This example was also used by Kandel and Stambaugh (1989).

FIG. 2. Power Function of Mean-Variance Spanning Test with Single Test Asset



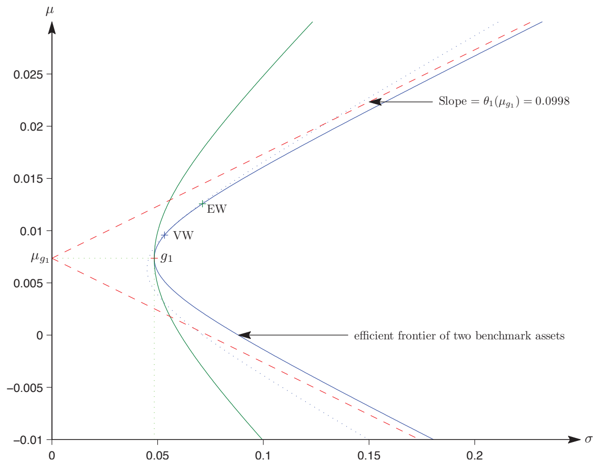
The figure plots the probability of rejecting the null hypothesis of mean-variance spanning as a function of ω^* for three different values of $T - K$ (the number of time series observations minus the number of benchmark assets), when there is only one test asset and the size of the test is 5%. The spanning test is an F-test, which has a central F-distribution with 2 and $T - K - 1$ degrees of freedom under the null hypothesis, and has a noncentral F-distribution with 2 and $T - K - 1$ degrees of freedom with noncentrality parameter $(T - K - 1)\omega^*$ under the alternatives.

ω for this test asset is 0.1610 (with 0.1574 coming from the first term). Based on Figure 2, this allows us to reject the null hypothesis with a 79% probability for $T - K = 60$, and the probability of rejection goes up to almost one for $T - K = 120$ and 240. In the second case, we consider a test asset that does not reduce the variance of the global minimum-variance portfolio but doubles the slope of the asymptote of the frontier from 0.0998 to 0.1996. This case is represented by the outer solid frontier in Figure 3. While economically this test asset represents a great improvement in the opportunity set, its ω is only 0.0299 and the F-test does not have much power to reject the null hypothesis. From Figure 2, the probability of rejecting the null hypothesis is only 20% for $T - K = 60$, 37% for $T - K = 120$, and 66% for $T - K = 240$.

It is easy to explain why the F-test has strong power rejecting the spanning hypothesis for a test asset that can improve the variance of the global minimum-variance portfolio but little power for a test asset that can only improve the tangency portfolio. This is because the sampling error of the former is in practice much less than that of the latter. The first term of ω

involves $c - c_1 = 1'_{N+K}V^{-1}1_{N+K} - 1'_K V_{11}^{-1} 1_K$ which is determined by V but not μ. Since estimates of V are in general a lot more accurate than estimates of μ (see Merton (1980)), even a small difference in c and c_1 can be detected and hence the test has strong power to reject the null hypothesis when $c \neq c_1$. However, the second term of ω involves $\theta^2(\hat{\mu}_{g_1}) - \theta_1^2(\hat{\mu}_{g_1})$, which is difficult to estimate accurately as it is determined by both μ and V. Therefore, even when we observe a large difference in the sample measure $\hat{\theta}^2(\hat{\mu}_{g_1}) - \hat{\theta}_1^2(\hat{\mu}_{g_1})$, it is possible that such a difference is due to sampling errors rather than due to a genuine difference. As a result, the spanning test has little power against alternatives that only display differences in the tangency portfolio but not in the global minimum-variance portfolio.

FIG. 3. Minimum-Variance Frontier of Two Benchmark Assets



The figure plots the minimum-variance frontier hyperbola of two benchmark assets in the (σ, μ) space. The two benchmark assets are the value-weighted (VW) and equally weighted (EW) portfolios of the NYSE. g_1 is the global minimum-variance portfolio and the two dashed lines are the asymptotes to the efficient set parabola. The frontier of the two benchmark assets is estimated using monthly data from the period 1926/1–2006/12. The figure also presents two additional frontiers for the case that a test asset is added to the two benchmark assets. The dotted frontier is for a test asset that improves the standard deviation of the global minimum-variance portfolio from 4.8%/month to 4.5%/month. The outer solid frontier is for a test asset that does not improve the global minimum-variance portfolio but doubles the slope of the asymptote from 0.0998 to 0.1996.

3.2. Multiple Test Assets

The calculation for the power of the spanning tests is extremely difficult when $N > 1$. For example, even though the F-test in (27) has a central F-distribution under the null, it does not have a noncentral F-distribution under the alternatives. To study the power of the three tests for $N > 1$, we need to understand the distribution of the two eigenvalues, λ_1 and λ_2, of the matrix $\hat{H}\hat{G}^{-1}$ under the alternatives. In this subsection, we provide first the exact distribution of λ_1 and λ_2 under the alternative hypothesis, then a simulation approach for computing the power in small samples, and finally examples illustrating the power under various alternatives.

Denote $\omega_1 \geq \omega_2 \geq 0$ the two eigenvalues of $H\hat{G}^{-1}$ where $H = \Theta\Sigma^{-1}\Theta'$ is the population counterpart of $\hat{H}$. The joint density of λ_1 and λ_2 can be written as

$$f(\lambda_1, \lambda_2) = e^{-\frac{T(\omega_1+\omega_2)}{2}} {}_1F_1\left(\frac{T-K+1}{2}; \frac{N}{2}; \frac{D}{2}, L(I_2+L)^{-1}\right) \times \frac{N-1}{4\mathrm{B}(N, T-K-N)} \left[\prod_{i=1}^{2} \frac{\lambda_i^{\frac{N-3}{2}}}{(1+\lambda_i)^{\frac{T-K+1}{2}}}\right] (\lambda_1 - \lambda_2), \quad (41)$$

for $\lambda_1 \geq \lambda_2 \geq 0$, where $L = \mathrm{Diag}(\lambda_1, \lambda_2)$, ${}_1F_1$ is the hypergeometric function with two matrix arguments, and $D = \mathrm{Diag}(T\omega_1, T\omega_2)$. Under the null hypothesis, the joint density function of λ_1 and λ_2 simplifies to

$$f(\lambda_1, \lambda_2) = \frac{N-1}{4\mathrm{B}(N, T-K-N)} \left[\prod_{i=1}^{2} \frac{\lambda_i^{\frac{N-3}{2}}}{(1+\lambda_i)^{\frac{T-K+1}{2}}}\right] (\lambda_1 - \lambda_2). \quad (42)$$

To understand why λ_1 and λ_2 are essential in testing the null hypothesis, note that the null hypothesis $H_0 : \Theta = 0_{2\times N}$ can be equivalently written as $H_0 : \omega_1 = \omega_2 = 0$. This is because $H\hat{G}^{-1}$ is a zero matrix if and only if H is a zero matrix, and this is true if and only if $\Theta = 0_{2\times N}$ since Σ is nonsingular. Therefore, tests of H_0 can be constructed using the sample counterparts of ω_1 and ω_2, i.e., λ_1 and λ_2. In theory, distributions of all functions of λ_1 and λ_2 can be obtained from their joint density function (41). However, the resulting expression is numerically very difficult to evaluate under alternative hypotheses because it involves the evaluation of a hypergeometric function with two matrix arguments. Instead of using the exact density function of λ_1 and λ_2, the following proposition helps us to obtain the small sample distribution of functions of λ_1 and λ_2 by simulation.

PROPOSITION 1. *λ_1 and λ_2 have the same distribution as the eigenvalues of AB^{-1} where $A \sim W_2(N, I_2, D)$ and $B \sim W_2(T-K-N+1, I_2)$, independent of A.*

With this proposition, we can simulate the exact sampling distribution of any functions of λ_1 and λ_2 as long as we can generate two random matrices A and B from the noncentral and central Wishart distributions, respectively. In the proof of Proposition 1 (in the Appendix), we give details on how to do so by drawing a few observations from the chi-squared and the standard normal distributions.

Before getting into the specific results, we first make some general observations on the power of the three tests. It can be shown that the power is a monotonically increasing function in $T\omega_1$ and $T\omega_2$.[15] This implies that, as expected, the power is an increasing functions of T. The more interesting question is how the power is determined for a fixed T. For such an analysis, we need to understand what the two eigenvalues of $H\hat{G}^{-1}$, ω_1 and ω_2, represent. The proof of Lemma 2 works also for the population counterparts of $\hat{H}$, so we can write

$$H = \begin{bmatrix} \Delta a & \Delta b \\ \Delta b & \Delta c \end{bmatrix} = \begin{bmatrix} a - a_1 & b - b_1 \\ b - b_1 & c - c_1 \end{bmatrix}, \qquad (43)$$

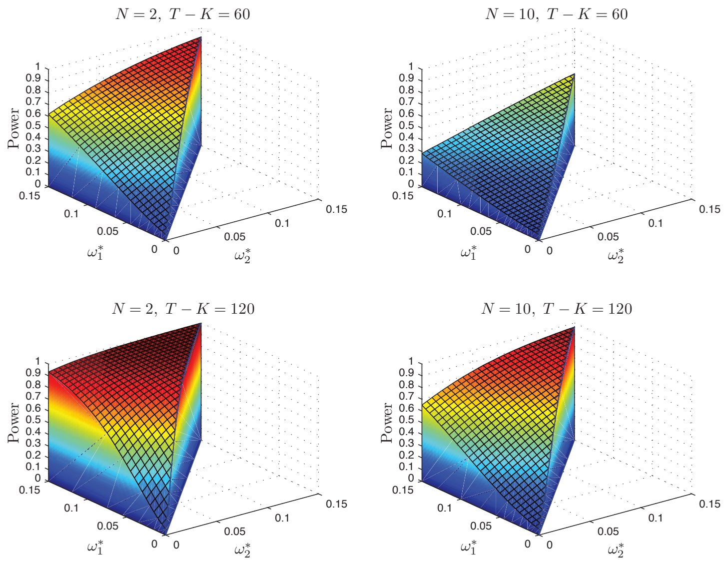
where $a = \mu' V^{-1} \mu$, $b = \mu' V^{-1} 1_{N+K}$, $c = 1'_{N+K} V^{-1} 1_{N+K}$, $a_1 = \mu_1' V_{11}^{-1} \mu_1$, $b_1 = \mu_1' V_{11}^{-1} 1_K$, and $c_1 = 1_K' V_{11}^{-1} 1_K$ are the population counterparts of the efficient set constants. Therefore, H is a measure of how far apart the *ex ante* minimum-variance frontier of K benchmark assets is from the *ex ante* minimum-variance frontier of all $N + K$ assets. Conditional on a given value of $\hat{G}$, the further apart the two frontiers, the bigger the H, the bigger the ω_1 and ω_2, and the more powerful the three tests. However, for a given value of H, the power also depends on $\hat{G}$, which is a measure of the *ex post* frontier of K benchmark assets. The better is the *ex post* frontier of K benchmark assets, the bigger the $\hat{G}$, and the less powerful the three tests. This is expected because if $\hat{G}$ is large, we can see from (18) that the estimates of α and δ will be imprecise and hence it is difficult to reject the null hypothesis even though it is not true.

In Figure 4, we present the power of the likelihood ratio test as a function of $\omega_1^* = T\omega_1/(T - K - 1)$ and $\omega_2^* = T\omega_2/(T - K - 1)$ for $N = 2$ and 10, and for $T - K = 60$ and 120, when the size of the test is 5%. Figure 4 shows that for fixed ω_1^* and ω_2^*, the power of the likelihood ratio test is an increasing function of $T - K$ and a decreasing function of N. The fact that the power of the test is a decreasing function of N does not imply we should use fewer test assets to test the spanning hypothesis. It only

[15] It is possible for the Lagrange multiplier test that its power function is not monotonically increasing in $T\omega_1$ and $T\omega_2$ when the sample size is very small. (See Perlman (1974) for a discussion of this.) However, for the usual sample sizes and significance levels that we consider, this problem will not arise.

suggests that if the additional test assets do not increase ω_1 and ω_2 (i.e., the additional test assets do not improve the frontier), then increasing the number of test assets will reduce the power of the test. However, if the additional test assets can improve the frontier, then it is possible that the power of the test can be increased by using more test assets.

FIG. 4. Power Function of Likelihood Ratio Test



The figure plots the probability of rejecting the null hypothesis of mean-variance spanning as a function of ω_1^* and ω_2^* using the likelihood ratio test when the size of the test is 5%, where $(T-K-1)\omega_1^*$ and $(T-K-1)\omega_2^*$ are the eigenvalues of the noncentrality matrix $TH\hat{G}^{-1}$. The four plots are for two different values of N (number of test assets) and two different values of $T-K$ (number of time series observations minus number of benchmark assets). The likelihood ratio test is an F-test, which has a central F-distribution with $2N$ and $2(T-K-N)$ degrees of freedom under the null hypothesis.

The plots for the power function of the Wald and the Lagrange multiplier tests are very similar to those of the likelihood ratio test, so we do not report them separately. For the purpose of comparing the power of these three tests, we report in Table 2 the probability of rejection of the three tests for $N = 10$ and $T - K = 60$ under different values of ω_1^* and ω_2^*. Although the difference in the power between the three tests is not large, a pattern emerges. When $\omega_2 \approx 0$, the Wald test is the most powerful among the three. However, when $\omega_1 \approx \omega_2$, the Lagrange multiplier test is more powerful than the other two. There are only a few cases where the likelihood ratio test is the most powerful one. The pattern that we observe in Table 2 holds for other values of N and $T-K$. Therefore, which test is more powerful depends on the relative magnitude of ω_1 and ω_2. The following lemma presents two extreme cases that help to identify alternative hypotheses with $\omega_2 \approx 0$ or $\omega_1 \approx \omega_2$.

LEMMA 3. *Define*

$$\mu_z = \arg\min_r \left[\theta^2(r) - \theta_1^2(r)\right] = \frac{\Delta b}{\Delta c}. \tag{44}$$

Under alternative hypotheses, we have (i) $\omega_2 = 0$ *if and only if* $c = c_1$ *or* $\theta^2(\mu_z) = \theta_1^2(\mu_z)$, *(ii)* $\omega_1 = \omega_2$ *if and only if*

$$\frac{c - c_1}{\hat{c}_1} = \frac{\theta^2(\mu_z) - \theta_1^2(\mu_z)}{1 + \hat{\theta}_1^2(\mu_z)}. \tag{45}$$

The first part of the lemma suggests that when there is a point at which the two *ex ante* minimum-variance frontiers are very close, then we have $\omega_2 \approx 0$. The second part of the lemma suggests that if the percentage reduction of the inverse of the variance of the global minimum-variance portfolio is roughly the same as the percentage increase in one plus the square of the slope of the tangent line (when the y-intercept of the tangent line is μ_z), then we will have $\omega_1 \approx \omega_2$.

As discussed earlier in the single test asset case, the effect of a small improvement of the standard deviation of the global minimum-variance portfolio is more important than the effect of a large increase in the slope of the tangent lines. Therefore, if we believe that the test assets could allow us to reduce the standard deviation of the global minimum-variance portfolio by even a small amount under the alternative hypothesis, then we should expect ω_1 to dominate ω_2 and the Wald test should be slightly more powerful than the other two tests.

4. A STEP-DOWN TEST

For reasonable alternative hypotheses, as shown earlier, the distance between the standard deviations of the two global minimum-variance portfolios is the primary determinant of the power of the three spanning tests whereas the distance between the two tangency portfolios is relatively unimportant. This is expected because the test of spanning is a joint test of $\alpha = 0_N$ and $\delta = 0_N$ and it weighs the estimates $\hat{\alpha}$ and $\hat{\delta}$ according to their statistical accuracy. Since δ does not involve μ (recall that δ is proportional to the weights of the N test assets in the global minimum-variance

TABLE 2.

Comparison of Power of Three Tests of Spanning Under Normality

Likelihood Ratio Test

	$\omega_2^* = 0.0$	$\omega_2^* = 0.3$	$\omega_2^* = 0.6$	$\omega_2^* = 0.9$	$\omega_2^* = 1.2$	$\omega_2^* = 1.5$
$\omega_1^* = 0.0$	0.0500					
$\omega_1^* = 0.3$	0.0823	0.1251				
$\omega_1^* = 0.6$	0.1226	0.1752	0.2338			
$\omega_1^* = 0.9$	0.1724	0.2307	0.2952	0.3612		
$\omega_1^* = 1.2$	0.2260	**0.2913**	0.3596	0.4257	0.4913	
$\omega_1^* = 1.5$	0.2834	0.3533	0.4228	0.4897	0.5533	0.6127

Wald Test

	$\omega_2^* = 0.0$	$\omega_2^* = 0.3$	$\omega_2^* = 0.6$	$\omega_2^* = 0.9$	$\omega_2^* = 1.2$	$\omega_2^* = 1.5$
$\omega_1^* = 0.0$	0.0500					
$\omega_1^* = 0.3$	**0.0825**	0.1243				
$\omega_1^* = 0.6$	**0.1241**	0.1735	0.2292			
$\omega_1^* = 0.9$	**0.1739**	0.2289	0.2901	0.3546		
$\omega_1^* = 1.2$	**0.2299**	0.2905	0.3547	0.4193	0.4834	
$\omega_1^* = 1.5$	**0.2902**	**0.3538**	0.4195	0.4829	0.5450	0.6042

Lagrange Multiplier Test

	$\omega_2^* = 0.0$	$\omega_2^* = 0.3$	$\omega_2^* = 0.6$	$\omega_2^* = 0.9$	$\omega_2^* = 1.2$	$\omega_2^* = 1.5$
$\omega_1^* = 0.0$	0.0500					
$\omega_1^* = 0.3$	0.0820	**0.1260**				
$\omega_1^* = 0.6$	0.1216	**0.1754**	**0.2362**			
$\omega_1^* = 0.9$	0.1685	**0.2314**	**0.2981**	**0.3650**		
$\omega_1^* = 1.2$	0.2199	0.2902	**0.3617**	**0.4296**	**0.4962**	
$\omega_1^* = 1.5$	0.2731	0.3496	**0.4234**	**0.4930**	**0.5589**	**0.6195**

The table presents the probabilities of rejection of Wald, likelihood ratio, and Lagrange multiplier tests of spanning in 100,000 simulations under the alternative hypotheses when the number of test assets (N) is equal to 10 and the number of time series observations less the number of benchmark assets ($T - K$) is equal to 60. The size of the tests is set at 5% and the alternative hypotheses are summarized by two measures ω_1^* and ω_2^*, where $(T - K - 1)\omega_1^*$ and $(T - K - 1)\omega_2^*$ are the eigenvalues of the noncentrality matrix $T H \hat{G}^{-1}$. Numbers that are boldfaced indicate the test has the highest power among the three tests.

portfolio of all $N + K$ assets), it can be estimated a lot more accurately than α. Therefore, tests of spanning inevitably place heavy weights on $\hat{\delta}$

and little weights on $\hat{\alpha}$. Although this practice is natural from a statistical point of view, it does not take into account the economic significance of the departure from the spanning hypothesis. A small difference in the global minimum-variance portfolios, while statistically significant, is not necessarily economically important. On the other hand, a big difference in the tangency portfolios can be of great economic importance, but this importance is difficult to detect statistically.

The fact that statistical significance does not always correspond to economic significance for the three spanning tests suggests that researchers need to be cautious in interpreting the p-values of these tests. A low p-value does not always imply that there is an economically significant difference between the two frontiers, and a high p-value does not always imply that the test assets do not add much to the benchmark assets. To mitigate this problem, we suggest researchers should examine the two components of the spanning hypothesis ($\alpha = 0_N$ and $\delta = 0_N$) individually instead of jointly. Such a practice could allow us to better assess the statistical evidence against the spanning hypothesis.

To be more specific, we suggest the following step-down procedure to test the spanning hypothesis.[16] This procedure is potentially more flexible and provides more information than the three tests discussed earlier.

The step-down procedure is a sequential test. We first test $\alpha = 0_N$, and then test $\delta = 0_N$ but conditional on the constraint $\alpha = 0_N$. To test $\alpha = 0_N$, similar to the GRS F-test, denote

$$F_1 = \left(\frac{T-K-N}{N}\right)\left(\frac{|\bar{\Sigma}|}{|\hat{\Sigma}|} - 1\right) = \left(\frac{T-K-N}{N}\right)\left(\frac{\hat{a}-\hat{a}_1}{1+\hat{a}_1}\right), \tag{46}$$

where $\hat{\Sigma}$ is the unconstrained estimate of Σ and $\bar{\Sigma}$ is the constrained estimate of Σ by imposing only the constraint of $\alpha = 0_N$. Under the null hypothesis, F_1 has a central F-distribution with N and $T-K-N$ degrees of freedom. Now to test $\delta = 0_N$ conditional $\alpha = 0_N$, we use the following F-test

$$\begin{aligned} F_2 &= \left(\frac{T-K-N+1}{N}\right)\left(\frac{|\tilde{\Sigma}|}{|\bar{\Sigma}|} - 1\right) \\ &= \left(\frac{T-K-N+1}{N}\right)\left[\left(\frac{\hat{c}+\hat{d}}{\hat{c}_1+\hat{d}_1}\right)\left(\frac{1+\hat{a}_1}{1+\hat{a}}\right) - 1\right], \end{aligned} \tag{47}$$

[16]See Section 8.4.5 of Anderson (1984) for a discussion of the step-down procedure. It should be noted that the step-down procedure there applies to each of the test assets but not to each component of the hypothesis as in our case.

where $\tilde{\Sigma}$ is the constrained estimate of Σ by imposing both the constraints of $\alpha = 0_N$ and $\delta = 0_N$. In the Appendix, we show that under the null hypothesis, F_2 has a central F-distribution with N and $T - K - N + 1$ degrees of freedom, and it is independent of F_1.

Suppose the level of significance of the first test is α_1 and that of the second test is α_2. Under the step-down procedure, we will accept the spanning hypothesis if we accept both tests. Therefore, the significance level of this step-down test is $1 - (1 - \alpha_1)(1 - \alpha_2) = \alpha_1 + \alpha_2 - \alpha_1\alpha_2$.[17] There are two benefits of using this step-down test. The first is that we can get an idea of what is causing the rejection. If the rejection is due to the first test, we know it is because the two tangency portfolios are statistically very different. If the rejection is due to the second test, we know the two global minimum-variance portfolios are statistically very different. The second benefit is flexibility in allocating different significance levels to the two tests based on their relative economic significance. For example, knowing that it does not take a big difference in the two global minimum-variance portfolio to reject $\delta = 0_N$ at the traditional significance level of 5%, we may like to set α_2 to a smaller number so that it takes a bigger difference in the two global minimum-variance portfolios for us to reject this hypothesis. Contrary to the three traditional tests that permit the statistical accuracy of $\hat{\alpha}$ and $\hat{\delta}$ to determine the relative importance of the two components of the hypothesis, the step-down procedure could allow us to adjust the significance levels based on the economic significance of the two components. Such a choice could result in a power function that is more sensible than those of the traditional tests.

To illustrate the step-down procedure, we return to our earlier example of two benchmark assets in Figure 3. For $T - K = 60$ and a level of significance of 5%, we show that the three traditional tests reject the spanning hypothesis with probability 0.79 for a test asset that merely reduces the standard deviation of the global minimum-variance portfolio from 4.8% to 4.5%, whereas for a test asset that doubles the slope of the asymptote from 0.0998 to 0.1996, the three tests can only reject with probability 0.20. In Table 3, we provide the power function of the step-down test for these two cases, using different values of α_1 and α_2 while keeping the significance level of the test at 5%.[18] For different values of α_1 and α_2, the step-down test has different power in rejecting the spanning hypothesis. However, in order for the step-down test to be more powerful in rejecting the test asset that doubles the slope of the asymptote, we need to set α_2 to be less than

[17] Alternatively, one can reverse the order by first testing $\delta = 0_N$ and then testing $\alpha = 0_N$ conditional on $\delta = 0_N$. In choosing the ordering of the tests, the natural choice is to test the more important component first.

[18] Under the alternative hypotheses, F_1 and F_2 are not independent. Details on the computation of the power of the step-down test are available upon request.

0.0001. Note that if we wish to accomplish roughly the same power as the traditional tests, all we need to do is to set $\alpha_1 = \alpha_2 = 0.02532$. While choosing the appropriate α_1 and α_2 is not a trivial task, it is far better to be able to have control over them than to leave them determined by statistical considerations alone.

5. TESTS OF MEAN-VARIANCE SPANNING UNDER NONNORMALITY

5.1. Conditional Homoskedasticity

Exact small sample tests are always preferred if they are available. The normality assumption is made so far to derive the small sample distributions. These results also serve as useful benchmarks for the general nonnormality case. In this section, we present the spanning tests under the assumption that the disturbance ϵ_t in (9) is nonnormal. There are two cases of nonnormality to consider. The first case is when ϵ_t is nonnormal but it is still independently and identically distributed when conditional on R_{1t}. The second case is when the variance of ϵ_t can be time-varying as a function of R_{1t}, i.e., the disturbance ϵ_t exhibits conditional heteroskedasticity.

For the first case that ϵ_t is conditionally homoskedastic, the three tests, (23)–(26), are still asymptotically χ^2_{2N} distributed under the null hypothesis, but their finite sample distributions will not be the same as the ones presented in Section II. Nevertheless, those results can still provide a very good approximation for the small sample distribution of the nonnormality case. To illustrate this, we simulate the returns on the test assets under the null hypothesis but with ϵ_t independently drawn from a multivariate Student-t distribution with five degrees of freedom.[19] In Table 4, we present the actual probabilities of rejection of the three tests in 100,000 simulations, for different values of K, N, and T, when the rejection decision is based on the 95th percentile of the exact distribution under the normality case. As we can see from Table 4, even when ϵ_t departs significantly from normality, the small sample distribution derived for the normality case still works amazingly well. Our findings are very similar to those of MacKinlay (1985) and Zhou (1993), in which they find that when ϵ_t is conditionally homoskedastic, nonnormality of ϵ_t has little impact on the finite sample distribution of the GRS test even for T as small as 60. Therefore, if one believes conditional homoskedasticity is a good working assumption, one should not hesitate to use the small sample version of the three tests de-

[19] Due to the invariance property, it can be shown that the joint distribution of λ_1 and λ_2 does not depend on Σ when ϵ_t has a multivariate elliptical distribution. Details are available upon request.

TABLE 3.

Power of Step-Down Test of Spanning Under Normality

Significance Levels		Probability of Rejection	
		$\Delta a = 0.0299$	$\Delta a, \Delta b = 0$
α_1	α_2	$\Delta b, \Delta c = 0$	$\Delta c = 67.16$
0.00000	0.05000	0.05117	0.87457
0.02532	0.02532	0.19930	0.80914
0.04040	0.01000	0.23996	0.70256
0.04905	0.00100	0.25889	0.42230
0.04914	0.00090	0.25908	0.41071
0.04924	0.00080	0.25927	0.39798
0.04933	0.00070	0.25946	0.38385
0.04943	0.00060	0.25966	0.36794
0.04952	0.00050	0.25985	0.34971
0.04962	0.00040	0.26004	0.32829
0.04971	0.00030	0.26023	0.30217
0.04981	0.00020	0.26041	0.26823
0.04990	0.00010	0.26060	0.21800
0.04995	0.00005	0.26070	0.17710
0.04996	0.00004	0.26071	0.16578
0.04997	0.00003	0.26073	0.15240
0.04998	0.00002	0.26075	0.13574
0.04999	0.00001	0.26077	0.11254
0.05000	0.00000	0.26068	0.05000

The table presents the probabilities of rejection of step-down test for two different alternatives, conditional on the frontier of two benchmark assets is given in Figure 3. The first alternative ($\Delta a = 0.0299$) is a test asset that doubles the slope of the asymptote to the efficient hyperbola of the two benchmark assets. The second alternative ($\Delta c = 67.16$) is a test asset that reduces the standard deviation of the global minimum-variance portfolio of the two benchmark assets from 4.8%/month to 4.5%/month. The step-down test is a sequential test. The first test is an F-test on $\alpha = 0_N$ and the second test is an F-test of $\delta = 0_N$ conditional on the restriction of $\alpha = 0_N$. The null hypothesis of spanning is only accepted if we accept both tests. α_1 and α_2 are the significance levels for the first and the second F-test, respectively. The number of time series observations is 62.

rived in Section II even though ϵ_t does not have a multivariate normal distribution.[20]

[20]For some distributions of ϵ_t, Dufour and Khalaf (2002) provide a simulation based method to construct finite sample tests in multivariate regressions. Their methodology can be used to obtain exact tests of spanning under multivariate elliptical errors.

TABLE 4.

Sizes of Small Sample Tests of Spanning Under Nonnormality of Residuals

			Actual Probabilities of Rejection		
K	N	T	W	LR	LM
2	2	60	0.048	0.048	0.048
		120	0.049	0.050	0.050
		240	0.051	0.051	0.051
	10	60	0.047	0.047	0.047
		120	0.046	0.046	0.046
		240	0.047	0.049	0.050
	25	60	0.046	0.047	0.047
		120	0.046	0.046	0.046
		240	0.047	0.048	0.048
5	2	60	0.049	0.048	0.048
		120	0.051	0.051	0.051
		240	0.051	0.051	0.051
	10	60	0.047	0.047	0.047
		120	0.048	0.048	0.048
		240	0.049	0.049	0.048
	25	60	0.046	0.046	0.047
		120	0.046	0.046	0.046
		240	0.048	0.048	0.048
10	2	60	0.050	0.049	0.049
		120	0.049	0.049	0.049
		240	0.051	0.051	0.051
	10	60	0.048	0.048	0.048
		120	0.049	0.049	0.049
		240	0.049	0.049	0.049
	25	60	0.048	0.048	0.048
		120	0.047	0.047	0.047
		240	0.047	0.047	0.047

The table presents the probabilities of rejection of Wald (W), likelihood ratio (LR), and Lagrange multiplier (LM) tests of spanning under the null hypothesis when the residuals follow a multivariate Student-t distribution with five degrees of freedom. The rejection decision is based on 95th percentile of their exact distributions under normality and the results for different values of the number of benchmark assets (K), test assets (N), and time series observations (T) are based on 100,000 simulations.

5.2. Conditional Heteroskedasticity

When ϵ_t exhibits conditional heteroskedasticity, the earlier three test statistics, (23)–(26), will no longer be asymptotically χ^2_{2N} distributed under the null hypothesis.[21] In this case, Hansen's (1982) GMM is the common viable alternative that relies on the moment conditions of the model. In this subsection, we present the GMM tests of spanning under the regression approach. This is the approach used by Ferson, Foerster, and Keim (1993).

Define $x_t = [1,\ R'_{1t}]'$, $\epsilon_t = R_{2t} - B'x_t$, the moment conditions used by the GMM estimation of B are

$$E[g_t] = E[x_t \otimes \epsilon_t] = 0_{(K+1)N}. \tag{48}$$

We assume R_t is stationary with finite fourth moments. The sample moments are given by

$$\bar{g}_T(B) = \frac{1}{T}\sum_{t=1}^{T} x_t \otimes (R_{2t} - B'x_t) \tag{49}$$

and the GMM estimate of B is obtained by minimizing $\bar{g}_T(B)'S_T^{-1}\bar{g}_T(B)$ where S_T is a consistent estimate of $S_0 = E[g_t g_t']$, assuming serial uncorrelatedness of g_t. Since the system is exactly identified, the unconstrained estimate $\hat{B}$, and hence $\hat{\Theta}$, does not depend on S_T and remains the same as their OLS estimates in Section II. The GMM version of the Wald test can be written as

$$W_a = T\mathrm{vec}(\hat{\Theta}')' \left[(A_T \otimes I_N)S_T(A_T' \otimes I_N)\right]^{-1} \mathrm{vec}(\hat{\Theta}') \overset{A}{\sim} \chi^2_{2N}, \tag{50}$$

where

$$A_T = \begin{bmatrix} 1+\hat{a}_1 & -\hat{\mu}_1\hat{V}_{11}^{-1} \\ \hat{b}_1 & -1_K'\hat{V}_{11}^{-1} \end{bmatrix}. \tag{51}$$

Since both the model and the constraints are linear, Newey and West (1987) show that the GMM version of the likelihood ratio test and the Lagrange multiplier test have exactly the same form as the Wald test, even though one needs the constrained estimate of B to calculate the likelihood ratio and Lagrange multiplier tests. Therefore, all three tests are numerically identical if they use the same S_T. In practice, different estimates of S_T are often used for the Wald test and the Lagrange multiplier test. For the case of the Wald test, S_T is computed using the unconstrained estimate of B whereas for the Lagrange multiplier test, S_T is usually computed using the

[21] It can be shown that under the null hypothesis, the asymptotic distribution of the three test statistics is a linear combination of $2N$ independent χ^2_1 random variables.

constrained estimate of B. Since the constrained estimate of B depends on the choice of S_T, a two-stage or an iterative approach is often used for performing the Lagrange multiplier test. Despite using different S_T, the two tests are still asymptotically equivalent under the null hypothesis. For the rest of this section, we focus on the GMM Wald test because its analysis does not require a specification of the initial weighting matrix and the number of iterations.

5.3. A Specific Example: Multivariate Elliptical Distribution

To study the potential impact of conditional heteroskedasticity on tests of spanning, we look at the case that the returns have a multivariate elliptical distribution. Under this class of distributions, the conditional variance of ϵ_t is in general not a constant, but a function of R_{1t}, unless the returns are multivariate normally distributed. The use of the multivariate elliptical distribution to model returns can be motivated both empirically and theoretically. Empirically, Mandelbrot (1963) and Fama (1965) find that normality is not a good description for stock returns because stock returns tend to have excess kurtosis compared with the normal distribution. This finding has been supported by many later studies, including Blatteberg and Gonedes (1974), Richardson and Smith (1993) and Zhou (1993). Since many members in the elliptical distribution like the multivariate Student-t distribution can have excess kurtosis, one could better capture the fat-tail feature of stock returns by assuming that the returns follow a multivariate elliptical distribution. Theoretically, we can justify the choice of multivariate elliptical distribution because it is the largest class of distributions for which mean-variance analysis is consistent with expected utility maximization.

For our purpose, the choice of multivariate elliptical distribution is appealing because the GMM Wald test has a simple analytical expression in this case. This analytical expression allows for simple analysis of the GMM Wald tests under conditional heteroskedasticity. The following proposition summarizes the results.[22]

PROPOSITION 2. *Suppose R_t is independently and identically distributed as a non-degenerate multivariate elliptical distribution with finite fourth moments. Define its kurtosis parameter as*

$$\kappa = \frac{E[((R_t - \mu)'V^{-1}(R_t - \mu))^2]}{(N+K)(N+K+2)} - 1. \tag{52}$$

[22]We thank Chris Geczy for suggesting the use of kurtosis parameter in this proposition. See Geczy (1999) for a similar conditional heteroskedasticity adjustment for tests of mean-variance efficiency under elliptical distribution.

Then the GMM Wald test of spanning is given by

$$W_a^e = T\mathrm{tr}(\hat{H}\hat{G}_a^{-1}) \overset{A}{\sim} \chi^2_{2N}, \tag{53}$$

where $\hat{H}$ *defined in (22) and*

$$\hat{G}_a = \begin{bmatrix} 1+(1+\hat{\kappa})\hat{a}_1 & (1+\hat{\kappa})\hat{b}_1 \\ (1+\hat{\kappa})\hat{b}_1 & (1+\hat{\kappa})\hat{c}_1 \end{bmatrix}, \tag{54}$$

where $\hat{\kappa}$ *is a consistent estimate of* κ*.*[23]

We use the notation W_a^e here to indicate that this GMM Wald test is only valid when R_t has a multivariate elliptical distribution, whereas the GMM Wald test W_a in (50) is valid for all distributions of R_t. Note that when returns exhibit excess kurtosis, $\hat{G}_a - \hat{G}$ is a positive definite matrix, so the regular Wald test $W = T\mathrm{tr}(\hat{H}\hat{G}^{-1})$ is greater than the GMM Wald test W_a^e.[24] Since $\hat{G}_a - \hat{G}$ does not go to zero asymptotically when $\kappa > 0$, using the regular Wald test W will lead to over-rejection problem when returns follow a multivariate elliptical distribution with excess kurtosis. In the following, we study a popular member of the multivariate elliptical distribution: the multivariate Student-t distribution.[25] To assess the impact of the multivariate Student-t distribution on tests of spanning, we perform a simulation experiment using the same two benchmark assets given in Figure 3. For different choices of N, we simulate returns of the benchmark assets and the test assets jointly from a multivariate Student-t distribution with mean and variance satisfying the null hypothesis. In Table 5, we present the actual size of the regular Wald test W and the two GMM Wald tests W_a and W_a^e, when the significance level of the tests is 5%. The results are presented for two different values of degrees of freedom for the multivariate Student-t distribution, $\nu = 5$ and 10.

As we can see from Table 5, the regular Wald tests reject far too often. The over-rejection problem is severe when N is large and when the degrees of freedom are small. In addition, the over-rejection problem does not go away as T increases. For the GMM Wald test under the elliptical distribution (W_a^e), it works reasonably well except when N is large and T is small, and its probability of rejection gets closer to the size of the

[23] In our empirical work, we use the biased-adjusted estimate of the kurtosis parameter developed by Seo and Toyama (1996).

[24] It can be shown that $-2/(N+K+2) < \kappa < \infty$ for multivariate elliptical distribution with finite fourth moments. Therefore, $\hat{G}_a$ cannot be too much smaller than $\hat{G}$ when the total number of assets $(N+K)$ is large, but $\hat{G}_a$ can be much bigger than $\hat{G}$ when the return distribution has fat tails.

[25] For multivariate Student t-distribution with ν degrees of freedom, we have $\kappa = 2/(\nu-4)$.

TABLE 5.

Sizes of Spanning Tests Under Multivariate Student-t Returns

		Actual Probabilities of Rejection			Average	Average
N	T	W	W_a	W_a^e	W/W_a	W/W_a^e
			Degrees of Freedom = 5			
2	60	0.195	0.166	0.091	1.141	1.474
	120	0.197	0.113	0.078	1.305	1.564
	240	0.204	0.084	0.070	1.452	1.648
10	60	0.555	0.832	0.231	0.685	1.424
	120	0.469	0.536	0.112	0.962	1.519
	240	0.459	0.309	0.073	1.191	1.609
25	60	0.979	1.000	0.844	0.138	1.386
	120	0.851	0.995	0.346	0.569	1.480
	240	0.756	0.870	0.137	0.892	1.570
			Degrees of Freedom = 10			
2	60	0.116	0.134	0.090	1.003	1.136
	120	0.101	0.090	0.071	1.070	1.148
	240	0.095	0.070	0.063	1.113	1.156
10	60	0.373	0.747	0.243	0.677	1.142
	120	0.239	0.399	0.121	0.878	1.155
	240	0.183	0.201	0.079	0.998	1.162
25	60	0.942	1.000	0.871	0.157	1.136
	120	0.636	0.982	0.402	0.583	1.151
	240	0.406	0.724	0.172	0.826	1.160

The table presents the probabilities of rejection of using regular Wald test (W) and two GMM Wald tests (W_a and W_a^e) of spanning under the null hypothesis when the returns follow a multivariate Student-t distribution with five and with ten degrees of freedom. The number of benchmark assets is two and they are chosen to have the same characteristics as the value-weighted and equally weighted market portfolios of the NYSE. The rejection decisions of the Wald tests are based on 95th percentile of χ^2_{2N}. The table also presents the average ratios of the regular Wald tests to the GMM Wald tests. Results for different values of number of test assets (N) and time series observations (T) are based on 100,000 simulations.

test as T increases. However, for the general GMM Wald test (W_a), it does not work well at all except when N is very small. In many cases, it over-rejects even more than the regular Wald test. Such over-rejection is due to the fact that W_a requires the estimation of a large S_0 matrix using S_T, which is imprecise when N is relatively large to T. While W_a is asymptotically equivalent to W_a^e under elliptical distribution, the poor finite sample performance of W_a suggests that it is an ineffective way to correct for conditional heteroskedasticity when N is large.

Table 5 also reports the average ratios of W/W_a and W/W_a^e. To understand what values these average ratios should take, we note that the limit of the expected bias of the regular Wald test under the multivariate Student-t distribution is

$$\lim_{T\to\infty} E\left[\frac{W}{W_a}\right] - 1 = \lim_{T\to\infty} E\left[\frac{W}{W_a^e}\right] - 1 \approx \frac{\kappa}{2} = \frac{1}{\nu-4}, \qquad (55)$$

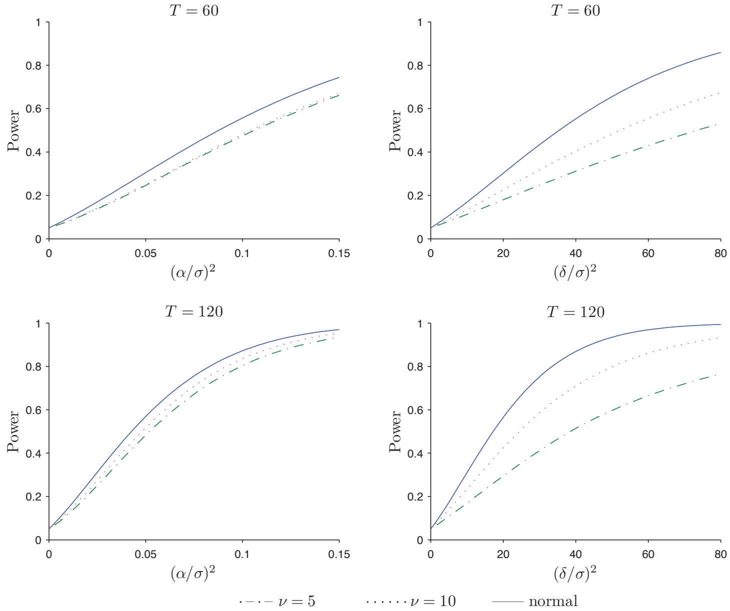
when the square of the slope of the asymptote to the sample frontier of the K benchmark assets, $\hat{\theta}_1^2(\hat{\mu}_g)$, is small compared with one (which is usually the case for monthly data). Therefore, when $\nu = 5$, the limit of the expected bias is about 100%, and when $\nu = 10$, the limit of the expected bias is about 16.7%. The magnitude of this bias is much greater than the one reported by MacKinlay and Richardson (1991) for test of mean-variance efficiency of a given portfolio. They find that when $\nu = 5$, the bias of the regular Wald test is less than 35% even when the squared Sharpe ratio of the benchmark portfolio is very large, and is negligible when the squared Sharpe ratio is small. To resolve this difference, we note that the test of mean-variance efficiency of a given portfolio is a test of $\alpha = 0_N$. The asymptotic variance of $\hat{\alpha}$ with and without the conditional heteroskedasticity adjustment are $\left[1 + \left(\frac{\nu-2}{\nu-4}\right) a_1\right]\Sigma$ and $(1+a_1)\Sigma$, respectively.[26] When the squared Sharpe ratio of the benchmark portfolio, a_1, is small compared with one, $1 + a_1$ is very close to $1 + \left(\frac{\nu-2}{\nu-4}\right) a_1$, and hence the impact of the conditional heteroskedasticity adjustment on test of $\alpha = 0_N$ is minimal. For the case of test of spanning, it is a joint test of $\alpha = 0_N$ and $\delta = 0_N$. The asymptotic variance of $\hat{\delta}$ with and without the conditional heteroskedasticity adjustment are $\left(\frac{\nu-2}{\nu-4}\right) c_1\Sigma$ and $c_1\Sigma$, respectively, and the ratio of the two is always equal to $(\nu-2)/(\nu-4)$. Hence, when ν is small, the asymptotic bias of W could still be very large even when the asymptotic variance of $\hat{\alpha}$ is almost unaffected. Therefore, conditional heteroskedasticity has potentially much bigger impact on tests of spanning than on tests of mean-variance efficiency of a given portfolio, and it is advisable not to ignore such adjustment for tests of spanning. In finite samples, Table 5 shows that for $\nu = 5$, W_a^e is only about 60% but not 100% larger than W, even when $T = 240$. For $\nu = 10$, the average ratio of W/W_a^e is roughly 1.16 and it is very close to the limit of 1.167. As for the average ratios of W/W_a, they are far away from its limit and often less than one. This again suggests that we should be cautious in using W_a to adjust for conditional heteroskedasticity when N is large.

[26] The asymptotic variance of $\hat{\alpha}$ is given in (A.32) of the Appendix. For the special case of $K = 1$, this expression is given in MacKinlay and Richardson (1991).

Besides its impact on the size of the regular Wald test, multivariate Student-t distribution also has significant impact on the power of the spanning test. This is because when returns follow a multivariate Student-t distribution, the asymptotic variances of $\hat{\alpha}$ and $\hat{\delta}$ are higher than the normality case. As a result, departures from the null hypothesis become more difficult to detect. Nevertheless, the power reduction is not uniform across all alternative hypotheses. For test assets that improve the tangency portfolio (i.e., $\alpha \neq 0_N$), we do not expect a significant change in power because the asymptotic variances of $\hat{\alpha}$ under multivariate Student-t and multivariate normality are almost identical. However, for test assets that improve the variance of the global minimum-variance portfolio (i.e., $\delta \neq 0_N$), we expect there can be a substantial loss in power when returns follow a multivariate Student-t distribution. This is because the asymptotic variance of $\hat{\delta}$ under multivariate Student-t returns is much higher than in the case of multivariate normal returns, especially when the degrees of freedom is small.

In Figure 5, we plot the power function of W_a^e under multivariate Student-t returns for these two types of alternative hypotheses. We use the same two benchmark assets as in Figure 3 and a single test asset constructed under different alternative hypotheses. Since we do not have the analytical expression for the power function of W_a^e under multivariate Student-t returns, the power functions are obtained by simulation. In addition, the power functions are size-adjusted so that W_a^e has the correct size under the null hypothesis. The two plots on the left hand side are for the power function of a test asset that has $\alpha \neq 0$. For both $T = 60$ and 120, we can see from Figure 5 that the power function for a test asset that has nonzero α does not change much by going from multivariate normal returns to multivariate Student-t returns. However, for a test asset that has $\delta \neq 0$, the two plots on the right hand side of Figure 5 show that there is a substantial decline in the power of W_a^e when returns follow a multivariate Student-t distribution, as compared with the case of multivariate normal. While there is a substantial reduction in the probability for W_a^e to reject nonzero δ when the returns follow a multivariate Student-t distribution with a low degrees of freedom, we still find that small difference in the global minimum-variance portfolio is easier to detect than large difference in the tangency portfolio. Therefore, just like the regular Wald test in the normality case, we cannot easily interpret the statistical significance in the GMM Wald test W_a^e. To better understand the source of rejection, we can construct a GMM version of the step-down test similar to the one for the case of normality. For the sake of brevity, we do not present the GMM step-down test here but details are available upon request.

FIG. 5. Power Function of GMM Wald Test Under Multivariate Student-t Returns



The figure plots the probability of rejecting the null hypothesis of mean-variance spanning for two different types of alternative hypotheses using the GMM Wald test. The plots on the left hand side are for alternative hypotheses with nonzero α, where $(\alpha/\sigma)^2$ is the improvement of the square of the slope of the tangent line with a y-intercept equals to zero. The plots on the right hand side are for alternative hypotheses with nonzero δ, where $(\delta/\sigma)^2$ is the improvement of the reciprocal of the variance of the global minimum variance portfolio. T is the length of time series observations used in the GMM Wald test. The significance level of the test is 5% and the rejection decision is based on the empirical distribution obtained from 100,000 simulations under the null hypothesis. For each one of the alternative hypotheses, returns on two benchmark assets and one test asset are generated using a multivariate Student-t distribution with five or ten degrees of freedom and the probability of rejection in 100,000 simulations is plotted. The figure also plots the power function for the case of multivariate normal returns for comparison.

6. AN APPLICATION

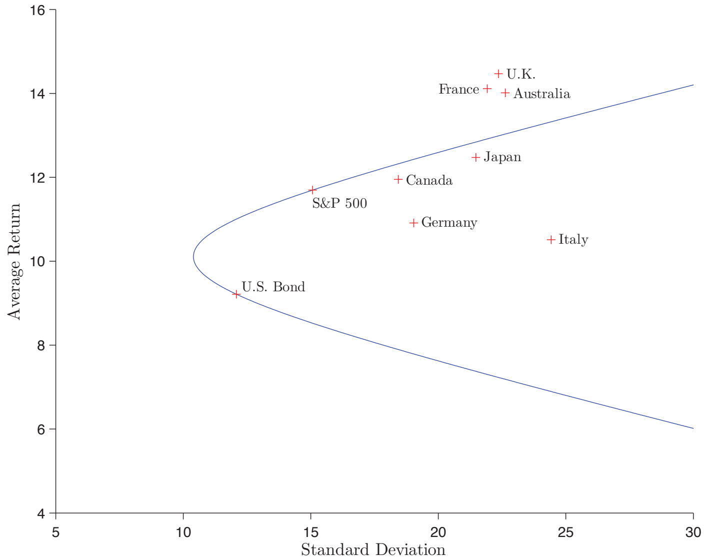
In this section, we apply various spanning tests to investigate if there are benefits for international diversification for a US investor who has an existing investment opportunity set that consists of the S&P 500 index and the

30-year U.S. Treasury bond. We assume the investor is considering investing in the equity markets of seven developed countries: Australia, Canada, France, Germany, Italy, Japan, and U.K. To address the question whether there are benefits for international diversification for this U.S. investor, we rely on monthly data over the period January 1970 to December 2007. Monthly data for all the return series are obtained the Global Financial Data, and they are all converted into U.S. dollar returns.

In Figure 6, we plot the *ex post* opportunity set available to the U.S. investor from combining the S&P 500 index and the 30-year U.S. Treasury bond. The sample return and standard deviation of the other seven developed countries are also indicated in the figure. From Figure 6, we can see that over the 38-year sample period, the U.K. equity market had the highest average return (14.5%/year), whereas the 30-year U.S. Treasury bond had the lowest average return (9.2%/year). Although we observe that some international equity markets (France, U.K. and Australia) lie outside the frontier formed by the U.S. bond and equity, it is possible that this occurs because of sampling errors, and a U.S. investor may not be able to expand his opportunity set reliably by introducing some foreign equity into his portfolio.

In Table 6, we report two mean-variance spanning tests on each of the seven foreign equity indices as well as a joint test on all seven indices. The first test is the corrected HK F-test and the second test is the step-down test. The tests are performed using monthly data over the 38-year sample period and its two subperiods. Both tests are exact under normality assumption on the residuals. Results from the entire period show that the traditional F-test rejects spanning at the 5% level for all the countries except for Canada. The joint test also rejects spanning for all seven countries. While we can reject spanning using the traditional F-test, it is not entirely clear how to interpret the results. For example, since we can reject spanning for Australia but not for Canada, does it mean the former is a better investment than the latter for the U.S. investor? Without knowing where the rejection comes from, one cannot easily answer this question. The step-down test can help in this case. There are two components in the step-down test, F_1 and F_2. F_1 is a test of $\alpha = 0_N$ whereas F_2 is a test of $\delta = 0_N$ conditional on $\alpha = 0_N$. From Table 6, the F_1 tests can only reject $\alpha = 0_N$ at the 5% level for Australia and Japan but the F_2 tests can reject $\delta = 0_N$ for all cases except for Canada. In addition, the joint test cannot reject $\alpha = 0_N$ for all seven countries but the evidence against $\delta = 0_N$ is overwhelming. By separating the sources of the rejection, we can conclude that there is strong evidence that the global minimum-variance portfolio can be improved by the seven foreign equity indices, but there is weaker evidence that the tangency portfolio can be improved.

FIG. 6. Average Return and Standard Deviation of U.S. and International Investments



The figure plots the average return and sample standard deviation (in annualized percentage) of S&P 500 index, 30-year U.S. Treasury bond, and seven international equity indices, computed using monthly data over the period January 1970 to December 2007. The figure also plots the opportunity set from combining the S&P 500 index and the 30-year U.S. Treasury bond.

The subperiod results are not very stable. Although we can jointly reject spanning for the seven equity indices in each subperiod, the evidence again is limited to rejection of $\delta = 0_N$ but not to rejection of $\alpha = 0_N$. Overall, the first subperiod offers more rejections of the spanning hypothesis than the second subperiod. One could interpret this as evidence that the global equity markets are becoming more integrated in the second subperiod, hence reducing the benefits of international diversification.

Given that returns exhibit conditional heteroskedasticity and fat-tails, the spanning tests in Table 6 which based on the normality assumption may not be appropriate. To determine the robustness of the results, we present in Table 7 some asymptotic spanning tests that do not rely on the normality assumption. We report two regression based Wald W_a^e (which is only valid when returns follow a multivariate elliptical distribution) and W_a. Consistent with results in Table 5, we find that for the regression based Wald tests, W_a^e are mostly smaller than W_a, possibly due to W_a is inflated in small sample. Keeping in mind that the reported p-values of these tests

TABLE 6.
Mean-Variance Spanning Tests on Seven International Equity Indices Under Normality

Country	$\hat{\alpha}$	$\hat{\delta}$	F-test	p-value	Step-Down Test F_1	p-value	F_2	p-value
Entire Period: 1970/1—2007/12								
Australia	0.00597	0.468	14.479	0.000	4.734	0.030	24.026	0.000
Canada	0.00130	0.116	2.132	0.120	0.549	0.459	3.718	0.054
France	0.00492	0.305	6.700	0.001	3.361	0.067	9.987	0.002
Germany	0.00349	0.439	17.083	0.000	2.222	0.137	31.858	0.000
Italy	0.00416	0.536	12.999	0.000	1.610	0.205	24.355	0.000
Japan	0.00619	0.561	18.376	0.000	4.560	0.033	31.943	0.000
U.K.	0.00461	0.244	4.534	0.011	2.973	0.085	6.070	0.014
All			5.161	0.000	1.262	0.267	9.314	0.000
First Subperiod: 1970/1—1988/12								
Australia	0.00668	0.545	9.113	0.000	2.174	0.142	15.970	0.000
Canada	0.00128	0.113	1.206	0.301	0.245	0.621	2.174	0.142
France	0.00639	0.320	3.389	0.035	1.939	0.165	4.819	0.029
Germany	0.00362	0.511	14.798	0.000	1.199	0.275	28.372	0.000
Italy	0.00468	0.595	8.420	0.000	0.839	0.361	16.011	0.000
Japan	0.01555	0.580	19.600	0.000	17.449	0.000	20.275	0.000
U.K.	0.00658	0.220	1.831	0.163	1.877	0.172	1.778	0.184
All			4.575	0.000	2.580	0.014	6.718	0.000
Second Subperiod: 1989/1—2007/12								
Australia	0.00455	0.329	4.273	0.015	2.215	0.138	6.297	0.013
Canada	0.00136	0.120	0.879	0.417	0.314	0.576	1.448	0.230
France	0.00332	0.271	3.504	0.032	1.459	0.228	5.538	0.019
Germany	0.00284	0.299	3.106	0.047	0.788	0.376	5.429	0.021
Italy	0.00319	0.420	3.745	0.025	0.607	0.437	6.894	0.009
Japan	−0.00343	0.508	6.527	0.002	0.602	0.439	12.473	0.001
U.K.	0.00280	0.285	5.463	0.005	1.480	0.225	9.426	0.002
All			1.738	0.046	0.795	0.593	2.722	0.010

The table presents two sets of mean-variance spanning tests on seven international equity indices, using the S&P 500 index and the 30-year U.S. Treasury bond as benchmark assets. The first test is an F-test of $H_0 : \alpha = 0_N$ and $\delta = 0_N$. The second test is a step down test where F_1 is an F-test of $\alpha = 0_N$, and F_2 is an F-test of $\delta = 0_N$ conditional on $\alpha = 0_N$. The two tests are performed on each international equity index as well as jointly on all seven international equity indices. The reported p-values are exact under the normality assumption on the residuals. The results are presented for the entire sample period as well as for its two subperiods.

are only asymptotic, we compare the test results in Table 7 with those in Table 6. We find that once we correct for conditional heteroskedasticity in the Wald tests, the evidence against rejection of spanning in Table 6 is further weakened, indicating that there could be over-rejection problems in Table 6 due to nonnormality of returns. Nevertheless, the asymptotic tests in Table 7 still can jointly reject spanning for the seven foreign equity indices in almost every case, indicating the rejection in Table 6 is robust to conditional heteroskedasticity in the returns.

In summary, we find that an U.S. investor with an existing opportunity set of the S&P 500 index and the 30-year U.S. Treasury bond can expand his opportunity set by investing in the equity indices of the seven developed countries. However, the improvement is only statistically significant at the global minimum-variance part of the frontier, but not at the part that is close to the tangency portfolio. To the extent that the U.S. investor is not interested in holding the global minimum-variance portfolio, there is no compelling evidence that international diversification can benefit this U.S. investor.

7. CONCLUSIONS

In this paper, we conduct a comprehensive study of various tests of mean-variance spanning. We provide geometrical interpretations and exact distributions for three popular test statistics based on the regression model. We also provide a power analysis of these tests that offers economic insights for understanding the empirical performance of these tests. In realistic situations, spanning tests have very good power for assets that could improve the variance of the global minimum-variance portfolio, but they have very little power against assets that could only improve the tangency portfolio. To mitigate this problem, we suggest a step-down test of spanning that allows us to extract more information from the data as well as gives us the flexibility to adjust the size of the test by weighting the two components of the spanning hypothesis based on their relative economic importance.

As an application, we apply the spanning tests to study benefits of international diversification for a U.S. investor. We find that there is strong evidence that equity indices in seven developed countries are not spanned by the S&P 500 index and the 30-year U.S. Treasury bond. However, the data cannot offer conclusive evidence that there are benefits for international diversification, except for those who are interested in investing in the part of the frontier that is close to the global minimum-variance portfolio.

APPENDIX A

Proof of Lemma 1 and Lemma 2: We first prove Lemma 2. Denote $\hat{\beta} = \hat{V}_{21}\hat{V}_{11}^{-1}$ and $\hat{\Sigma} = \hat{V}_{22} - \hat{V}_{21}\hat{V}_{11}^{-1}\hat{V}_{12}$. Using the partitioned matrix inverse

TABLE 7.

Asymptotic Mean-Variance Spanning Tests on Seven International Equity Indices

Country	W_a^e	p-value	W_a	p-value
Entire Period: 1970/1—2007/12				
Australia	10.964	0.004	21.573	0.000
Canada	2.227	0.328	3.397	0.183
France	8.469	0.014	10.373	0.006
Germany	19.573	0.000	28.422	0.000
Italy	18.215	0.000	17.584	0.000
Japan	25.428	0.000	30.432	0.000
U.K.	3.090	0.213	8.641	0.013
All	33.718	0.002	57.608	0.000
First Subperiod: 1970/1—1988/12				
Australia	6.859	0.032	15.149	0.001
Canada	1.288	0.525	1.925	0.382
France	4.646	0.098	6.055	0.048
Germany	22.291	0.000	25.391	0.000
Italy	12.648	0.002	11.872	0.003
Japan	34.102	0.000	31.163	0.000
U.K.	1.829	0.401	4.577	0.101
All	37.741	0.001	56.848	0.000
Second Subperiod: 1989/1—2007/12				
Australia	7.080	0.029	7.801	0.020
Canada	1.114	0.573	1.614	0.446
France	5.303	0.071	6.320	0.042
Germany	3.773	0.152	5.378	0.068
Italy	6.405	0.041	6.095	0.047
Japan	10.644	0.005	14.761	0.001
U.K.	8.620	0.013	11.124	0.004
All	20.094	0.127	27.294	0.018

The table presents four mean-variance spanning tests on seven international equity indices, using the S&P 500 index and the 30-year U.S. Treasury bond as benchmark asset-s. The two tests, W_a^e and W_a, are regression based GMM Wald tests. W_a is valid under general distribution whereas W_a^e is only valid when returns follow a multivariate elliptical distribution. Both tests are performed on each internation-al equity index as well as jointly on all seven international equity indices, and they both have an asymptotic χ^2_{2N} distribution, where N is the number of test assets, and the reported p-values are asymptotic ones. The results are p-resented for the entire sample period as well as for its two subperiods.

formula, it is easy to verify that

$$\begin{aligned}\hat{V}^{-1} &= \begin{bmatrix} \hat{V}_{11}^{-1} + \hat{\beta}'\hat{\Sigma}^{-1}\hat{\beta} & -\hat{\beta}'\hat{\Sigma}^{-1} \\ -\hat{\Sigma}^{-1}\hat{\beta} & \hat{\Sigma}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \hat{V}_{11}^{-1} & 0_{K\times N} \\ 0_{N\times K} & O_{N\times N} \end{bmatrix} + \begin{bmatrix} -\hat{\beta}' \\ I_N \end{bmatrix} \hat{\Sigma}^{-1}[-\hat{\beta} \;\; I_N]. \end{aligned} \tag{A.1}$$

Therefore,

$$\begin{aligned} &\begin{bmatrix} \hat{a} & \hat{b} \\ \hat{b} & \hat{c} \end{bmatrix} \\ &= \begin{bmatrix} \hat{\mu}' \\ 1'_{N+K} \end{bmatrix} \hat{V}^{-1}[\hat{\mu} \;\; 1_{N+K}] \\ &= \begin{bmatrix} \hat{\mu}' \\ 1'_{N+K} \end{bmatrix} \begin{bmatrix} \hat{V}_{11}^{-1} & 0_{K\times N} \\ 0_{N\times K} & 0_{N\times N} \end{bmatrix} [\hat{\mu} \;\; 1_{N+K}] + \begin{bmatrix} \hat{\mu}' \\ 1'_{N+K} \end{bmatrix} \begin{bmatrix} -\hat{\beta}' \\ I_N \end{bmatrix} \hat{\Sigma}^{-1}[-\hat{\beta} \;\; I_N][\hat{\mu} \;\; 1_{N+K}] \\ &= \begin{bmatrix} \hat{\mu}_1' \\ 1'_K \end{bmatrix} \hat{V}_{11}^{-1}[\hat{\mu}_1 \;\; 1_K] + \begin{bmatrix} (\hat{\mu}_2 - \hat{\beta}\hat{\mu}_1)' \\ (1_N - \hat{\beta}1_K)' \end{bmatrix} \hat{\Sigma}^{-1}[\hat{\mu}_2 - \hat{\beta}\hat{\mu}_1 \;\; 1_N - \hat{\beta}1_K] \\ &= \begin{bmatrix} \hat{a}_1 & \hat{b}_1 \\ \hat{b}_1 & \hat{c}_1 \end{bmatrix} + \hat{H}. \end{aligned} \tag{A.2}$$

This completes the proof of Lemma 2.

For the proof of Lemma 1, we write

$$1 + \hat{\theta}^2(r) = 1 + \hat{a} - 2\hat{b}r + \hat{c}r^2 = [1, \; -r] \begin{bmatrix} 1+\hat{a} & \hat{b} \\ \hat{b} & \hat{c} \end{bmatrix} \begin{bmatrix} 1 \\ -r \end{bmatrix}, \tag{A.3}$$

and similarly

$$1 + \hat{\theta}_1^2(r) = 1 + \hat{a}_1 - 2\hat{b}_1 r + \hat{c}_1 r^2 = [1, \; -r] \begin{bmatrix} 1+\hat{a}_1 & \hat{b}_1 \\ \hat{b}_1 & \hat{c}_1 \end{bmatrix} \begin{bmatrix} 1 \\ -r \end{bmatrix}. \tag{A.4}$$

Therefore, we can write

$$\frac{1+\hat{\theta}^2(r)}{1+\hat{\theta}_1^2(r)} - 1 = \frac{[1, \; -r] \begin{bmatrix} \Delta\hat{a} & \Delta\hat{b} \\ \Delta\hat{b} & \Delta\hat{c} \end{bmatrix} \begin{bmatrix} 1 \\ -r \end{bmatrix}}{[1, \; -r] \begin{bmatrix} 1+\hat{a}_1 & \hat{b}_1 \\ \hat{b}_1 & \hat{c}_1 \end{bmatrix} \begin{bmatrix} 1 \\ -r \end{bmatrix}}, \tag{A.5}$$

and it is just a ratio of two quadratic forms in $[1, \; -r]'$. The maximum and minimum of this ratio of two quadratic forms are given by the two eigenvalues of

$$\begin{bmatrix} \Delta\hat{a} & \Delta\hat{b} \\ \Delta\hat{b} & \Delta\hat{c} \end{bmatrix} \begin{bmatrix} 1+\hat{a}_1 & \hat{b}_1 \\ \hat{b}_1 & \hat{c}_1 \end{bmatrix}^{-1} = \hat{H}\hat{G}^{-1}, \tag{A.6}$$

which are λ_1 and λ_2, respectively. This completes the proof of Lemma 1.
Proof of (30): Let $\xi_1 = \lambda_1/(1+\lambda_1)$ and $\xi_2 = \lambda_2/(1+\lambda_2)$. From Anderson (1984, p.529) and using the duplication formula

$$\Gamma(k)\Gamma\left(k-\frac{1}{2}\right) = \frac{\sqrt{\pi}\Gamma(2k-1)}{2^{2k-2}} \tag{A.7}$$

when $2k$ is an integer, we can write the joint density function of ξ_1 and ξ_2 under the null hypothesis as

$$f(\xi_1,\xi_2) = \frac{n+1}{2\mathrm{B}(2m+2,2n+3)}\left[\prod_{i=1}^{2}\xi_i^m(1-\xi_i)^n\right](\xi_1-\xi_2) \quad \text{for } 1\ge\xi_1\ge\xi_2\ge 0, \tag{A.8}$$

where $m=(N-3)/2$ and $n=(T-K-N-2)/2$.

Using a transformation $a_1=\xi_1+\xi_2$ and $a_2=\xi_1\xi_2$, we have the joint density function of a_1 and a_2 as

$$f(a_1,a_2) = \frac{n+1}{2\mathrm{B}(2m+2,2n+3)}a_2^m(1-a_1+a_2)^n. \tag{A.9}$$

Since $a_1=(\xi_1+\xi_2)\ge 2\sqrt{\xi_1\xi_2}=2\sqrt{a_2}$ and $1-a_1+a_2=(1-\xi_1)(1-\xi_2)\ge 0$, the probability for $\xi_1+\xi_2\le v$ is equal to

$$\begin{aligned}
& P[a_1\le v] \\
={}& \frac{n+1}{2\mathrm{B}(2m+2,2n+3)}\int_0^{\frac{v^2}{4}}\int_{2\sqrt{a_2}}^{\min[v,1+a_2]} a_2^m(1-a_1+a_2)^n \mathrm{d}a_1\mathrm{d}a_2 \\
={}& \frac{n+1}{2\mathrm{B}(2m+2,2n+3)}\int_0^{\frac{v^2}{4}} a_2^m\left[\frac{(1-a_1+a_2)^{n+1}}{n+1}\right]\Bigg|_{\min[v,1+a_2]}^{2\sqrt{a_2}}\mathrm{d}a_2 \\
={}& \frac{1}{2\mathrm{B}(2m+2,2n+3)}\left[\int_0^{\frac{v^2}{4}} a_2^m(1-\sqrt{a_2})^{2n+2}\mathrm{d}a_2 - \int_{\max[0,v-1]}^{\frac{v^2}{4}} a_2^m(1-v+a_2)^{n+1}\mathrm{d}a_2\right] \\
={}& I_{\frac{v}{2}}(2m+2,2n+3) - \frac{1}{2\mathrm{B}(2m+2,2n+3)}\int_{\max[0,v-1]}^{\frac{v^2}{4}} a_2^m(1-v+a_2)^{n+1}\mathrm{d}a_2. \qquad \text{(A.10)}
\end{aligned}$$

This completes the proof.
Proof of (37) and (38): Since λ_1 and λ_2 are the two eigenvalues of $\hat{H}\hat{G}^{-1}$, they are the solutions to the following equation

$$|\hat{H}\hat{G}^{-1}-\lambda I_2| = 0, \tag{A.11}$$

or equivalently the solutions to

$$|\hat{H} - \lambda\hat{G}| = \begin{vmatrix} \Delta\hat{a} - \lambda(1+\hat{a}_1) & \Delta\hat{b} - \lambda\hat{b}_1 \\ \Delta\hat{b} - \lambda\hat{b}_1 & \Delta\hat{c} - \lambda\hat{c}_1 \end{vmatrix} = 0. \quad \text{(A.12)}$$

Simplifying, we have

$$(\hat{c}_1 + \hat{d}_1)\lambda^2 - \left[\Delta\hat{a}\hat{c}_1 - 2\Delta\hat{b}\hat{b}_1 + \Delta\hat{c}(1+\hat{a}_1)\right]\lambda + \left[\Delta\hat{a}\Delta\hat{c} - (\Delta\hat{b})^2\right] = 0. \quad \text{(A.13)}$$

It is easy to see that

$$\begin{aligned} \lambda_1 + \lambda_2 &= \frac{\Delta\hat{a}\hat{c}_1 - 2\Delta\hat{b}\hat{b}_1 + \Delta\hat{c}(1+\hat{a}_1)}{\hat{c}_1 + \hat{d}_1} \\ &= \frac{\Delta\hat{a} - 2\Delta\hat{b}\hat{\mu}_{g_1} + \Delta\hat{c}\hat{\mu}_{g_1}^2}{1 + \frac{\hat{d}_1}{\hat{c}_1}} + \frac{\Delta\hat{c}\left(\frac{1+\hat{a}_1}{\hat{c}_1} - \hat{\mu}_{g_1}^2\right)}{1 + \frac{\hat{d}_1}{\hat{c}_1}} \\ &= \frac{\hat{\theta}^2(\hat{\mu}_{g_1}) - \hat{\theta}_1^2(\hat{\mu}_{g_1})}{1 + \hat{\theta}_1^2(\hat{\mu}_{g_1})} + \frac{\Delta\hat{c}}{\hat{c}_1}, \end{aligned} \quad \text{(A.14)}$$

where the last equality follows from the fact that

$$\hat{\theta}^2(r) - \hat{\theta}_1^2(r) = (\hat{a} - 2\hat{b}r + \hat{c}r^2) - (\hat{a}_1 - 2\hat{b}_1 r + \hat{c}_1 r^2) = \Delta\hat{a} - 2\Delta\hat{b}r + \Delta\hat{c}r^2. \quad \text{(A.15)}$$

For the Lagrange multiplier test, we define $\xi_i = \lambda_i/(1+\lambda_i)$ and we have ξ_1 and ξ_2 as the two eigenvalues of $\hat{H}(\hat{H}+\hat{G})^{-1}$, which are the solutions to the following equation

$$|\hat{H} - \xi(\hat{H}+\hat{G})| = \begin{vmatrix} \Delta\hat{a} - \xi(1+\hat{a}) & \Delta\hat{b} - \xi\hat{b} \\ \Delta\hat{b} - \xi\hat{b} & \Delta\hat{c} - \xi\hat{c} \end{vmatrix} = 0. \quad \text{(A.16)}$$

Comparing (A.12) with (A.16), the only difference is $\hat{a}_1$, $\hat{b}_1$, $\hat{c}_1$ are replaced by $\hat{a}$, $\hat{b}$, and $\hat{c}$. Therefore, by making the corresponding substitutions, $\xi_1+\xi_2$ takes the same form as (A.14). This completes the proof.

Proof of (41): Following Muirhead (1982), it is easy to show that $Y_1^* = \sqrt{T}\hat{G}^{-\frac{1}{2}}\hat{\Theta}$ and $\hat{\Sigma}$ are independent of each other. Furthermore, the eigenvalues of $Y_1^*(T\hat{\Sigma})^{-1}Y_1^{*\prime} = \hat{G}^{-\frac{1}{2}}\hat{\Theta}\hat{\Sigma}^{-1}\hat{\Theta}'\hat{G}^{-\frac{1}{2}}$ are the same as the eigenvalues of $\hat{\Theta}\hat{\Sigma}^{-1}\hat{\Theta}'\hat{G}^{-1} = \hat{H}\hat{G}^{-1}$, so from Theorem 10.4.5 of Muirhead (1982), we have the joint density function of the two eigenvalues of $\hat{H}\hat{G}^{-1}$ as

$$\begin{aligned} f(\lambda_1,\lambda_2) &= e^{-\mathrm{tr}(\Omega)/2}{}_1F_1\left(\frac{T-K+1}{2};\frac{N}{2};\frac{\Omega}{2},L(I_2+L)^{-1}\right)\times \\ &\quad \frac{N-1}{4\mathrm{B}(N,T-K-N)}\left[\prod_{i=1}^{2}\frac{\lambda_i^{\frac{N-3}{2}}}{(1+\lambda_i)^{\frac{T-K+1}{2}}}\right](\lambda_1-\lambda_2), \end{aligned} \quad \text{(A.17)}$$

for $\lambda_1 \geq \lambda_2 \geq 0$, where $L = \text{Diag}(\lambda_1, \lambda_2)$, ${}_1F_1$ is the hypergeometric function with two matrix arguments, and

$$\Omega = T\hat{G}^{-\frac{1}{2}}\Theta\Sigma^{-1}\Theta'\hat{G}^{-\frac{1}{2}}. \tag{A.18}$$

It is well known that the hypergeometric function only depends on the eigenvalues of Ω, which is the same as the eigenvalues of $TH\hat{G}^{-1}$. Therefore, the joint density function of λ_1 and λ_2 depends only on the eigenvalues of $TH\hat{G}^{-1}$ and we can replace Ω with D. This completes the proof.

Proof of Proposition 1: Using Theorem 10.4.2 of Muirhead (1982), we can find out the density function of the two eigenvalues of AB^{-1} is exactly the same as (41). To generate B, we use the Bartlett's decomposition of central Wishart distribution (see Muirhead (1982), Theorem 3.2.14). Define L a lower triangular 2 by 2 matrix with $L_{11} \sim \sqrt{\chi^2_{T-K-N+1}}$, $L_{22} \sim \sqrt{\chi^2_{T-K-N}}$, $L_{21} \sim N(0, 1)$, and they are independent of each other. Then $B = LL' \sim W_2(T - K - N + 1, I_2)$. To generate A, we generate a central Wishart $S \sim W_2(N - 2, I_2)$ using the same procedure and a 2 by 2 matrix Z where $\text{vec}(Z) \sim N(\text{vec}(D^{\frac{1}{2}}), I_4)$, then we have $Z'Z \sim W_2(2, I_2, D)$ and $A = S + Z'Z \sim W_2(N, I_2, D)$. This completes the proof.

Proof of Lemma 3: By replacing $\Delta\hat{a}$, $\Delta\hat{b}$, $\Delta\hat{c}$ by Δa, Δb, and Δc, we have from (A.14)

$$\omega_1 + \omega_2 = \frac{\Delta c}{\hat{c}_1} + \frac{\theta^2(\hat{\mu}_{g_1}) - \theta_1^2(\hat{\mu}_{g_1})}{1 + \hat{\theta}_1^2(\hat{\mu}_{g_1})}. \tag{A.19}$$

Similarly, with the same replacement, we have from (A.13)

$$\omega_1\omega_2 = \frac{\Delta a \Delta c - (\Delta b)^2}{\hat{c}_1 + \hat{d}_1} = \left(\frac{\Delta c}{\hat{c}_1}\right)\left(\frac{\theta^2(\mu_z) - \theta_1^2(\mu_z)}{1 + \hat{\theta}_1^2(\hat{\mu}_{g_1})}\right), \tag{A.20}$$

where the last equality follows from the fact that

$$\theta^2(\mu_z) - \theta_1^2(\mu_z) = \Delta a - 2\Delta b\left(\frac{\Delta b}{\Delta c}\right) + \Delta c\left(\frac{\Delta b}{\Delta c}\right)^2 = \Delta a - \frac{(\Delta b)^2}{\Delta c}. \tag{A.21}$$

(i) Since under the alternative hypothesis, we have $\omega_1 > 0$. Therefore, from (A.20), we can see that $\omega_2 = 0$ if and only if $\Delta c = 0$ or $\theta^2(\mu_z) - \theta_1^2(\mu_z)$.
(ii) Using the inequality $(a + b)^2 \geq 4ab$ for a and b nonnegative and the

definition of μ_z, we have

$$\begin{aligned}
(\omega_1+\omega_2)^2 &= \left[\frac{\Delta c}{\hat{c}_1}+\frac{\theta^2(\hat{\mu}_{g_1})-\theta_1^2(\hat{\mu}_{g_1})}{1+\hat{\theta}_1^2(\hat{\mu}_{g_1})}\right]^2 \\
&\geq 4\left(\frac{\Delta c}{\hat{c}_1}\right)\left(\frac{\theta^2(\hat{\mu}_{g_1})-\theta_1^2(\hat{\mu}_{g_1})}{1+\hat{\theta}_1^2(\hat{\mu}_{g_1})}\right) \\
&\geq 4\left(\frac{\Delta c}{\hat{c}_1}\right)\left(\frac{\theta^2(\mu_z)-\theta_1^2(\mu_z)}{1+\hat{\theta}_1^2(\hat{\mu}_{g_1})}\right) \\
&= 4\omega_1\omega_2. \qquad (A.22)
\end{aligned}$$

For $\omega_1=\omega_2>0$, we need the two inequalities to be equalities. This is true if and only if

$$\frac{\Delta c}{\hat{c}_1}=\frac{\theta^2(\hat{\mu}_{g_1})-\theta_1^2(\hat{\mu}_{g_1})}{1+\hat{\theta}_1^2(\hat{\mu}_{g_1})} \qquad (A.23)$$

and $\hat{\mu}_{g_1}=\mu_z$. Combining these two conditions, we prove the lemma.

Proof of the distribution of (46) and (47): The proof that under the null hypothesis, F_1 has a central F-distribution with N and $T-K-N$ degrees of freedom follows directly from Theorem 8.4.5 of Anderson (1984). For F_2, we have from Seber (1984, pp.412–413),

$$\frac{|\bar{\Sigma}|}{|\tilde{\Sigma}|}\sim U_{N,1,T-K} \qquad (A.24)$$

under the null hypothesis, and hence from 2.42 of Seber (1984), we have

$$F_2=\left(\frac{T-K-N+1}{N}\right)\left(\frac{|\tilde{\Sigma}|}{|\bar{\Sigma}|}-1\right)\sim F_{N,T-K-N+1}. \qquad (A.25)$$

The independence of F_1 and F_2 under the null hypothesis follows from Corollary 10.5.4 of Muirhead (1982). This completes the proof.

Proof of (50): From Hansen (1982), the asymptotic variance of $\mathrm{vec}(\hat{B}')$ is given by $(D_0'S_0^{-1}D_0)^{-1}$, where

$$D_0=E\left[\frac{\partial\bar{g}_T(B)}{\partial\mathrm{vec}(B')'}\right]=-E[x_tx_t']\otimes I_N. \qquad (A.26)$$

Since $\hat{\Theta}=A\hat{B}-C$, the asymptotic variance of $\mathrm{vec}(\hat{\Theta}')$ is given by

$$\begin{aligned}
&(A\otimes I_N)(D_0'S_0^{-1}D_0)^{-1}(A'\otimes I_N) \\
=\ &(A\otimes I_N)D_0^{-1}S_0D_0^{-1}(A'\otimes I_N) \\
=\ &(A(E[x_tx_t'])^{-1}\otimes I_N)S_0((E[x_tx_t'])^{-1}A'\otimes I_N). \qquad (A.27)
\end{aligned}$$

Using the partitioned matrix inverse formula, we have

$$\begin{aligned} A(E[x_t x_t'])^{-1} &= \begin{bmatrix} 1 & 0_K' \\ 0 & -1_K' \end{bmatrix} \begin{bmatrix} 1 & \mu_1' \\ \mu_1 & V_{11}+\mu_1\mu_1' \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 0_K' \\ 0 & -1_K' \end{bmatrix} \begin{bmatrix} 1+\mu_1'V_{11}^{-1}\mu_1 & -\mu_1'V_{11}^{-1} \\ -V_{11}^{-1}\mu_1 & V_{11}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} 1+\mu_1'V_{11}^{-1}\mu_1 & -\mu_1 V_{11}^{-1} \\ 1_K'V_{11}^{-1}\mu_1 & -1_K'V_{11}^{-1} \end{bmatrix}. \end{aligned} \tag{A.28}$$

Replacing S_0 and $A(E[x_t x_t'])^{-1}$ by their consistent estimates S_T and A_T, we obtain (50). This completes the proof.

Proof of Proposition 2: When R_t follows a multivariate elliptical distribution, we have

$$E[R_{1t}R_{1t}'\otimes\epsilon_t\epsilon_t'] = \mu_1\mu_1'\otimes\Sigma+(1+\kappa)V_{11}\otimes\Sigma = (V_{11}+\mu_1\mu_1')\otimes\Sigma+\kappa V_{11}\otimes\Sigma, \tag{A.29}$$

using Corollary 3.2.1 and 3.2.2 in Mathai, Provost, and Hayakawa (1995). It follows that

$$S_0 = E[x_t x_t']\otimes\Sigma + \begin{bmatrix} 0 & 0_K' \\ 0_K & \kappa V_{11} \end{bmatrix}\otimes\Sigma. \tag{A.30}$$

Using this expression and (A.26), the asymptotic variance of $\mathrm{vec}(\hat{B}')$ is given by

$$(D_0')^{-1}S_0D_0^{-1} = \begin{bmatrix} 1+a_1 & -\mu_1'V_{11}^{-1} \\ -V_{11}^{-1}\mu_1 & V_{11}^{-1} \end{bmatrix}\otimes\Sigma+\kappa\begin{bmatrix} a_1 & -\mu_1'V_{11}^{-1} \\ -V_{11}^{-1}\mu_1 & V_{11}^{-1} \end{bmatrix}\otimes\Sigma. \tag{A.31}$$

Note that the first term is the asymptotic variance of $\mathrm{vec}(\hat{B}')$ under the conditional homoskedasticity assumption, and the second term is the adjustment matrix due to the conditional heteroskedasticity. The asymptotic variance of $\mathrm{vec}(\hat{\Theta}')$ is then given by

$$(A\otimes I_N)(D_0^{-1})'S_0D_0^{-1}(A'\otimes I_N) = \begin{bmatrix} 1+(1+\kappa)a_1 & (1+\kappa)b_1 \\ (1+\kappa)b_1 & (1+\kappa)c_1 \end{bmatrix}\otimes\Sigma. \tag{A.32}$$

By replacing a_1, b_1, c_1, κ, Σ by their consistent estimates $\hat{a}_1$, $\hat{b}_1$, $\hat{c}_1$, $\hat{\kappa}$ and $\hat{\Sigma}$, the consistent estimate of the asymptotic variance of $\mathrm{vec}(\hat{\Theta}')$ is $\hat{G}_a\otimes\hat{\Sigma}$. Therefore, the GMM Wald test is

$$\begin{aligned} W_a &= T\mathrm{vec}(\hat{\Theta}')'(\hat{G}_a^{-1}\otimes\hat{\Sigma}^{-1})\mathrm{vec}(\hat{\Theta}') \\ &= T\mathrm{vec}(\hat{\Theta}')'\mathrm{vec}(\hat{\Sigma}^{-1}\hat{\Theta}'\hat{G}_a^{-1}) = T\mathrm{tr}(\hat{H}\hat{G}_a^{-1}), \end{aligned} \tag{A.33}$$

where the last equality follows from the identity $\mathrm{tr}(AB) = \mathrm{vec}(A')'\mathrm{vec}(B)$. This completes the proof.

Proof of (55): Since W_a is asymptotically equivalent to W_a^e, the limit of $E[W/W_a]$ is the same as the limit of $E[W/W_a^e]$. For W, we have from (A.14),

$$W = \mathrm{tr}(\hat{H}\hat{G}^{-1}) = \frac{\hat{\theta}^2(\hat{\mu}_g) - \hat{\theta}_1^2(\hat{\mu}_g)}{1 + \hat{\theta}_1^2(\hat{\mu}_g)} + \frac{\Delta\hat{c}}{\hat{c}_1}. \tag{A.34}$$

Using a similar proof, we have

$$W_a^e = \mathrm{tr}(\hat{H}\hat{G}_a^{-1}) = \frac{\hat{\theta}^2(\hat{\mu}_g) - \hat{\theta}_1^2(\hat{\mu}_g)}{1 + (1+\hat{\kappa})\hat{\theta}_1^2(\hat{\mu}_g)} + \frac{\Delta\hat{c}}{\hat{c}_1(1+\hat{\kappa})} \equiv X_1 + X_2. \tag{A.35}$$

Under the null hypothesis, the two terms X_1 and X_2 are asymptotically independent of each other and distributed as χ_N^2. When $\hat{\theta}_1^2(\hat{\mu}_g)$ is small compared with one, we have

$$\mathrm{tr}(\hat{H}\hat{G}^{-1}) \approx X_1 + (1+\kappa)X_2, \tag{A.36}$$

and hence

$$\lim_{T\to\infty} \frac{W}{W_a^e} - 1 \approx \frac{X_1 + (1+\kappa)X_2}{X_1 + X_2} - 1 = \kappa\left(\frac{X_2}{X_1 + X_2}\right). \tag{A.37}$$

Asymptotically, $X_2/(X_1 + X_2)$ has a beta distribution and its expected value is 1/2. Therefore, we have

$$\lim_{T\to\infty} E\left[\frac{W}{W_a^e}\right] - 1 \approx \frac{\kappa}{2}. \tag{A.38}$$

This completes the proof.

REFERENCES

Ahn, D., J. Conrad, and R.F. Dittmar, 2003. Risk adjustment and trading strategies. *Review of Financial Studies* **16** , 459–485.

Anderson, T. W, 1984. *An introduction of multivariate statistical analysis*, 2nd edition. Wiley, New York.

Bekaert, G. and M. S. Urias, 1996. Diversification, integration and emerging market closed-end funds. *Journal of Finance* **51**, 835–869.

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

Blattberg, R. C. and N. J. Gonedes, 1974. A comparison of the stable and Student distributions as statistical models of stock prices. *Journal of Business* **47**, 244–280.

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

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

Chen, W., H. Chung, K. Ho, and T. Hsu, 2010. Portfolio optimization models andmean Cvariance spanning tests, in *Handbook of Quantitative Finance and Risk Management*, eds. by Lee et al, Springe, pp 165–184.

Chong, T. K., 1988. An Inquiry into the Appropriateness and Sensitivity of Recent Multivariate Tests of The CAPM, Ph.D. dissertation. New York University.

Christiansen, C., J. S. Joensen, and H. S. Nielsen, 2007. The risk-return trade-off in human capital investment. *Labour Economics* **14**, 971–986.

De Roon, F. A. and T. E. Nijman, 2001. Testing for mean-variance spanning: A survey. *Journal of Empirical Finance* **8**, 111–155.

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**, 721–742.

DeSantis, G., 1993. Volatility bounds for stochastic discount factors: tests and implications from international financial markets. Ph.D. dissertation, Department of Economics, University of Chicago.

Dufour, J. and L. Khalaf, 2002. Simulation based finite and large sample tests in multivariate regressions. *Journal of Econometrics* **111**, 303–322.

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., 1965.The behavior of stock market prices. *Journal of Business* **38**, 34–105.

Ferson, W., S. R.Foerster, and D. B. Keim, 1993. General tests of latent variable models and mean-variance spanning. *Journal of Finance* **48**, 131–156.

Geczy, C., 1999. Some generalized tests of mean-variance efficiency and multifactor model performance. Working Paper, University of Pennsylvania.

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

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

Hotelling, H., 1951. A generalized T test and measure of multivariate dispersion, in: J. Neyman, (Eds.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California, Los Angeles and Berkeley, pp. 23–41.

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

Jagannathan, R., G. Skoulakis, and Z. Wang, 2003. Analysis of large cross sections of security returns, in: Y. Aït-Sahalia, and L. Hansen, (Eds.), *Handbook of Financial Econometrics*, forthcoming.

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, 1989. A performance interpretation of multivariate tests of asset set intersection, spanning and mean-variance efficiency. *Journal of Financial and Quantitative Analysis* **24**, 185–204.

Kandel, S.and R. F. Stambaugh, 1992. A mean-variance framework for tests of asset pricing models. *Review of Financial Studies* **2**, 125–156.

Korkie, B.and H. J. Turtle, 2002. A mean-variance analysis of self-financing portfolios. *Management Science* **48**, 427–443.

Mandelbrot, B., 1963. The variation of certain speculative prices. *Journal of Business* **36**, 394–419.

MacKinlay, A. C., 1985. Analysis of multivariate financial tests. Ph.D. dissertation, Graduate School of Business, University of Chicago.

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

Mathai, A. M., S. B. Provost, and T. Hayakawa, 1995. *Bilinear forms and zonal polynomials.* Springer-Verlag, New York.

Merton, R. C., 1972. An analytic derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis* **7**, 1851–1872.

Merton, R. C., 1980. On estimating the expected return on the market. *Journal of Financial Economics* **8**, 323–361.

Mikhail, N. N., 1965. A comparison of tests of the Wilks-Lawley hypothesis in multivariate analysis. *Biometrika* **52**, 149–156.

Muirhead, R. J., 1982. *Aspects of multivariate statistical theory.* Wiley, New York.

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

Peñaranda, F.and E. Sentana, 2004. Spanning tests in return and stochastic discount mean-variance frontiers: a unifying approach. Working paper, CEMFI.

Perlman, M. D., 1974. On the monotonicity of the power function of tests based on traces of multivariate beta matrices. *Journal of Multivariate Analysis* **4**, 22–30.

Pillai, K. C. S.and K. Jayachandran, 1967. Power comparison of tests of two multivariate hypotheses based on four criteria. *Biometrika* **54**, 195–210.

Richardson, M. P., and T. Smith, 1993. A test for multivariate normality in stock returns. *Journal of Business* **66**, 295–321.

Roll, R., 1977. A critique of the asset pricing theory's test; Part I: on past and potential testability of theory. *Journal of Financial Economics* **4**, 129–176.

Seber, G. A. F., 1984. *Multivariate Observations.* Wiley, New York.

Seo, T.and T. Toyama, 1996. On the estimation of kurtosis parameter in elliptical distributions. *Journal of the Japan Statistical Society* **26**, 59–68.

Zhou, G., 1983. Asset pricing test under alternative distributions. *Journal of Finance* **48**, 1925–1942.