Tests of Mean-Variance Spanning

RAYMOND KAN and GUOFU ZHOU*

First draft: March, 1997
This version: September, 2001

*Kan is from the University of Toronto, Zhou is from Washington University in St. Louis. We thank Philip Dybvig, Wayne Ferson, Chris Geczy, 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 Social Sciences and Humanities Research Council of Canada.

ABSTRACT

In this paper, we conduct a comprehensive study of tests of mean-variance spanning. We provide both a comparison and a geometrical interpretation of three asymptotic tests (likelihood ratio, Wald, and Lagrange multiplier) of mean-variance spanning under the regression based framework of Huberman and Kandel (1987). For the case of normality, we provide the exact distributions and a comprehensive power analysis of the three tests. For the general case, we provide the GMM version of the spanning tests and evaluate their performance using simulation. In addition, we compare the performance of the spanning tests under the regression approach with those under the stochastic discount factor approach. Our results suggest that the two approaches have similar properties when returns are normally distributed but the regression approach is superior to the stochastic discount factor approach when returns follow a multivariate Student-t distribution.

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 in the literature by Huberman and Kandel (1987, HK hereafter) in which they proposed 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. Subsequent to HK's study, different tests were developed in the finance literature to address the question of mean-variance spanning in different applications. Examples include Ferson, Foerster, and Keim (1993), DeSantis (1993), Bekaert and Urias (1996), De Roon, Nijman, and Werker (2001), and Korkie and Turtle (2001).

In this paper, we aim at providing a complete understanding of various tests of mean-variance spanning. We point out there are two major mistakes in applications of the widely used HK test of spanning. The first mistake is that the test statistic is often incorrectly computed due to a typo in HK's original paper. The second mistake is that the HK test is incorrectly used for the single test asset case (i.e., $N = 1$). In addition to the HK test, we also study other tests of spanning. Statistically, the HK test is a likelihood ratio test. However, unlike the case of testing the CAPM as in Jobson and Korkie (1982) and Gibbons, Ross, and Shanken (1989, GRS hereafter), the likelihood ratio test is in general not the uniformly most powerful invariant test for testing mean-variance spanning. Therefore, it is important not to consider just the likelihood ratio test, but also consider other tests of spanning. Two alternative tests that we consider are the Wald test and the Lagrange multiplier test. We compare their performance with the likelihood ratio test and present geometrical interpretations of all three tests in terms of the *ex post* minimum-variance frontier of the K benchmark assets and that of the entire $N + K$ assets. Under the normality assumption, we present the small sample distribution for all of the three tests and provide a complete analysis of their power under alternative hypotheses. In addition, we also relate the power of these tests to the economic significance of departure from the spanning hypothesis. We find that the power of these three tests does not always align with the economic significance of the difference between the two minimum-variance frontiers. Without the normality assumption, we conduct our analysis using simulation and reach a similar conclusion. As an attempt to overcome the power problem in the traditional tests, we propose a new testing procedure that is potentially more informative than the traditional spanning tests.

Recently, DeSantis (1993), Ferson (1995), and Bekaert and Urias (1996) recast the concept of mean-variance spanning in the stochastic discount factor (SDF) framework and provide various Generalized Method of Moments (GMM) based tests of spanning. As a further development, we provide a comparison of the newer SDF approach with the older regression based approach, in terms of both the size and power. Such a comparison offers insights as to which of these two different approaches is more reliable in practice. We find the GMM spanning tests under the regression approach are superior to the corresponding tests under the SDF approach when returns exhibit conditional heteroskedasticity.

The rest of the paper is organized as follows. The next section discusses the concept of spanning and the regression based approach for tests of spanning. Section II provides a comprehensive power analysis of various regression based spanning tests. Section III discusses how to generalize these tests to the case that the assets returns are not multivariate normally distributed. Section IV introduces the stochastic discount factor approach for tests of spanning and compares their performance with the regression based tests. Section V applies various mean-variance spanning tests to examine if there are benefits of international diversification for a U.S. investor. The final section concludes our findings and the Appendix contains proofs of all propositions.

I. Regression Based Tests of Spanning

A. Mean-Variance Spanning

The concept of mean-variance spanning is simple. We say 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. In the literature, the first set of K risky assets is often called the benchmark assets and the second set of N risky assets is called the test assets. When there exists a risk-free asset and 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, 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 there is no unlimited risk-free lending and borrowing at the same risk-free rate, then investors instead are interested 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, the 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.[1] 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 $E[\epsilon_t R_{1t}'] = \mathrm{O}_{N\times K}$, where 0_N is an N-vector of zeros and $\mathrm{O}_{N\times K}$ 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

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

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

To understand why (4) implies mean-variance spanning, we observe that when (4) holds, then for every test asset, 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

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

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

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.[2]

Denote $\Sigma = V_{22} - V_{21}V_{11}^{-1}V_{12}$ and $Q = [O_{N\times K},\ I_N]$ 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

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

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 assets, 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.[3]

[2] 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.

[3] 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 $Qw_2 = 0_N$.

B. 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:

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

where R 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.[4] 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

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

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

Under the normality assumption, we have

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

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

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

[4] 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.

$H_0 : \Theta = \mathrm{O}_{2\times N}$. Since $\Theta = AB - C$ where

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

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

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

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

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

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

It should be noted that numerically, one does not need to do the constrained estimation 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

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

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 $1/U = (1 + \lambda_1)(1 + \lambda_2)$ and 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)).

Besides the likelihood ratio test, there are two other popular asymptotic tests: the Wald test and the Lagrange multiplier test. From Berndt and Savin (1977), the Wald test is given by

$$W = T(\lambda_1 + \lambda_2) \overset{A}{\sim} \chi^2_{2N}. \tag{24}$$

and the Lagrange multiplier test is given by

$$LM = T \sum_{i=1}^{2} \frac{\lambda_i}{1+\lambda_i} \overset{A}{\sim} \chi^2_{2N}. \tag{25}$$

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.[5] Therefore, using the asymptotic distribution 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 $N = 1$,[6] 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.

C. 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 distribution of the three tests under the null hypothesis.[7] 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[8]

$$\left(\frac{1}{U^{\frac{1}{2}}} - 1\right)\left(\frac{T-K-N}{N}\right) \sim F_{2N,2(T-K-N)}. \tag{26}$$

[5] 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 $T-K+1$.

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

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

[8] 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).

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

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

In this case, the exact distribution of the Wald and Lagrange multiplier tests can be obtained from the F-test in (27) since all three tests are increasing transformations of each other.

For $N \geq 2$, the exact distribution of the Wald test under the null hypothesis is given in Hotelling (1951) and Anderson (1984)

$$\begin{aligned} & P[\lambda_1 + \lambda_2 \leq w] \\ = \; & I_{\frac{w}{2+w}}(N-1, T-K-N) - \\ & \frac{\mathrm{B}\left(\frac{1}{2}, \frac{T-K}{2}\right)}{\mathrm{B}\left(\frac{N}{2}, \frac{T-K-N+1}{2}\right)}(1+w)^{-\left(\frac{T-K-N}{2}\right)} I_{\left(\frac{w}{2+w}\right)^2}\left(\frac{N-1}{2}, \frac{T-K-N}{2}\right), \end{aligned} \tag{28}$$

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 for $N \geq 2$, there is no easy expression available in the literature.[9] Therefore, we provide our own expression here. For $0 \leq v \leq 2$, we have

$$\begin{aligned} & P\left[\frac{\lambda_1}{1+\lambda_1} + \frac{\lambda_2}{1+\lambda_2} \leq v\right] \\ = \; & I_{\frac{v}{2}}(N-1, T-K-N+1) - \frac{\int_{\max[0,v-1]}^{\frac{v^2}{4}} u^{\frac{N-3}{2}}(1-v+u)^{\frac{T-K-N}{2}}\mathrm{d}u}{2\mathrm{B}(N-1, T-K-N+1)}. \end{aligned} \tag{29}$$

Proof of this expression is given in the Appendix.

Under the null hypothesis, the exact distributions of all three tests depend only on N and $T - K$ but are independent of the realizations of R_{1t}. Therefore, under the null hypothesis, the unconditional distributions of the three test statistics are the same as their distributions when unconditional on R_{1t}. In Table I, 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. From Table I, we can 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

[9] 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.

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.

Table I about here

D. 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

$$\hat{G} = \begin{bmatrix} 1 + \hat{a}_1 & \hat{b}_1 \\ \hat{b}_1 & \hat{c}_1 \end{bmatrix}. \tag{30}$$

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

Lemma 1 *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*

$$\hat{H} = \begin{bmatrix} \Delta\hat{a} & \Delta\hat{b} \\ \Delta\hat{b} & \Delta\hat{c} \end{bmatrix}. \tag{31}$$

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

$$U = \frac{1}{|I_2 + \hat{H}\hat{G}^{-1}|} = \frac{|\hat{G}|}{|\hat{G} + \hat{H}|} = \frac{(1 + \hat{a}_1)\hat{c}_1 - \hat{b}_1^2}{(1 + \hat{a})\hat{c} - \hat{b}^2} = \frac{\hat{c}_1 + \hat{d}_1}{\hat{c} + \hat{d}} = \left(\frac{\hat{c}_1}{\hat{c}}\right)\left(\frac{1 + \frac{\hat{d}_1}{\hat{c}_1}}{1 + \frac{\hat{d}}{\hat{c}}}\right), \tag{32}$$

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 (26) can be written as

$$F = \left(\frac{T-K-N}{N}\right)\left(\frac{1}{U^{\frac{1}{2}}} - 1\right) = \left(\frac{T-K-N}{N}\right)\left[\left(\frac{\sqrt{\hat{c}}}{\sqrt{\hat{c}_1}}\right)\left(\frac{\sqrt{1+\frac{\hat{d}}{\hat{c}}}}{\sqrt{1+\frac{\hat{d}_1}{\hat{c}_1}}}\right) - 1\right]. \tag{33}$$

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 (26) can be geometrically represented as[10]

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

Figure 1 about here

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

[10]For $N = 1$, the F-test of (27) can be geometrically represented as

$$F = \left(\frac{T-K-1}{2}\right)\left[\left(\frac{OD}{OC}\right)^2\left(\frac{AH}{BF}\right)^2 - 1\right]. \tag{34}$$

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, we show in the Appendix that the Wald and Lagrange multiplier tests can be represented geometrically as[11]

$$\lambda_1 + \lambda_2 = \frac{\hat{c} - \hat{c}_1}{\hat{c}_1} + \frac{\hat{\theta}^2(\hat{\mu}_{g_1}) - \hat{\theta}_1^2(\hat{\mu}_{g_1})}{1 + \hat{\theta}_1^2(\hat{\mu}_{g_1})} = \left(\frac{OD}{OC}\right)^2 - 1 + \left(\frac{BE}{BF}\right)^2 - 1 \tag{36}$$

and

$$\frac{\lambda_1}{1+\lambda_1} + \frac{\lambda_2}{1+\lambda_2} = \frac{\hat{c} - \hat{c}_1}{\hat{c}} + \frac{\hat{\theta}^2(\hat{\mu}_g) - \hat{\theta}_1^2(\hat{\mu}_g)}{1 + \hat{\theta}^2(\hat{\mu}_g)} = 1 - \left(\frac{OC}{OD}\right)^2 + 1 - \left(\frac{AG}{AH}\right)^2. \tag{37}$$

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 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.

[11]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.

II. Power Analysis of Spanning Tests

A. 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. In fact, 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 (27). 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

$$\left(\frac{1}{U} - 1\right)\left(\frac{T-K-1}{2}\right) \sim F_{2,T-K-1}(T\omega), \tag{38}$$

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[12]

$$\omega = \left[\frac{c - c_1}{\hat{c}_1} + \frac{\theta^2(\hat{\mu}_{g1}) - \theta_1^2(\hat{\mu}_{g1})}{1 + \hat{\theta}_1^2(\hat{\mu}_{g1})}\right], \tag{39}$$

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

[12]The derivation of this expression is similar to that of (36) and therefore not provided.

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 $T-K = 240$.

Figure 2 about here

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.[13] Using monthly returns from 1926/1–1998/12, we estimate $\hat{\mu}_1$ and $\hat{V}_{11}$ and we have $\hat{\mu}_{g_1} = \hat{b}_1/\hat{c}_1 = 0.0079$, $\hat{\sigma}_{g_1} = 1/\sqrt{\hat{c}_1} = 0.049$, and $\hat{\theta}_1(\hat{\mu}_{g_1}) = 0.0875$. 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 4.9%/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 ω for this test asset is 0.1886 (with 0.1838 coming from the first term). Based on Figure 2, this allows us to reject the null hypothesis with a 86% probability for $T-K = 60$, and the probability of rejection goes up to almost one for $T-K = 120$ and

[13]This example was also used by Kandel and Stambaugh (1989).

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 0.0875 to 0.175. 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 0.0228 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 16% for $T-K=60$, 29% for $T-K=120$, and 54% 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 $c-c_1$ 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.

Figure 3 about here

B. 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 (26) 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 $\hat{H}$. In the Appendix, we show that the joint density of λ_1 and λ_2 can be written as

$$\begin{aligned} f(\lambda_1, \lambda_2) &= e^{-\frac{T(\omega_1+\omega_2)}{2}} {}_1F_1\left(\frac{T-K+1}{2}; \frac{N}{2}; \frac{D}{2}, L(I_2+L)^{-1}\right) \times \\ & \frac{N-1}{4\mathrm{B}(N, T-K-N)} \left[\prod_{i=1}^{2} \frac{\lambda_i^{\frac{N-3}{2}}}{(1+\lambda_i)^{\frac{T-K+1}{2}}}\right] (\lambda_1 - \lambda_2), \end{aligned} \tag{40}$$

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

$$f(\lambda_1, \lambda_2) = \frac{N-1}{4\mathrm{B}(N, T-K-N)} \left[\prod_{i=1}^{2} \frac{\lambda_i^{\frac{N-3}{2}}}{(1+\lambda_i)^{\frac{T-K+1}{2}}}\right] (\lambda_1 - \lambda_2). \tag{41}$$

To understand why λ_1 and λ_2 are essential in testing the null hypothesis, note that the null hypothesis $H_0 : \Theta = \mathrm{O}_{2\times N}$ 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 $\Theta = \mathrm{O}_{2\times N}$ 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 (40). However, the resulting expression is numerically very difficult to evaluate under alternative hypotheses because it involves the evaluation of a hypergeometric function with 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-square 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 $T\omega_2$.[14] 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 1 works also for the population counterparts of $\hat{H}$, so we can write

$$H = \begin{bmatrix} \Delta a & \Delta b \\ \Delta b & \Delta c \end{bmatrix} = \begin{bmatrix} a - a_1 & b - b_1 \\ b - b_1 & c - c_1 \end{bmatrix}, \tag{42}$$

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 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.

Figure 4 about here

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

[14] 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.

comparing the power of these three tests, we report in Table II 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 II 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 2 *Define*

$$\mu_z = \arg\min_r \left[\theta^2(r) - \theta_1^2(r)\right] = \frac{\Delta b}{\Delta c}. \tag{43}$$

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*

$$\frac{c - c_1}{\hat{c}_1} = \frac{\theta^2(\mu_z) - \theta_1^2(\mu_z)}{1 + \hat{\theta}_1^2(\mu_z)}. \tag{44}$$

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 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 $\omega_1 \approx \omega_2$.

Table II about here

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.

C. Economic Significance of Alternative Hypotheses

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 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.[15] This procedure, seems new in finance, is potentially more flexible and provides more information than the three standard 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

[15]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.

test $\alpha = 0_N$, similar to the GRS F-test, denote

$$F_1 = \left(\frac{T-K-N}{N}\right)\left(\frac{|\bar{\Sigma}|}{|\hat{\Sigma}|} - 1\right) = \left(\frac{T-K-N}{N}\right)\left(\frac{\hat{a}-\hat{a}_1}{1+\hat{a}_1}\right), \tag{45}$$

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

$$F_2 = \left(\frac{T-K-N+1}{N}\right)\left(\frac{|\tilde{\Sigma}|}{|\bar{\Sigma}|} - 1\right) = \left(\frac{T-K-N+1}{N}\right)\left[\left(\frac{\hat{c}+\hat{d}}{\hat{c}_1+\hat{d}_1}\right)\left(\frac{1+\hat{a}_1}{1+\hat{a}}\right) - 1\right], \tag{46}$$

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 $1-(1-\alpha_1)(1-\alpha_2) = \alpha_1 + \alpha_2 - \alpha_1\alpha_2$.[16] 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 level on economic significance of the 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

[16] 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.

tests reject the spanning hypothesis with probability 0.86 for a test asset that merely reduces the standard deviation of the global minimum-variance portfolio from 4.9% to 4.5%, whereas for a test asset that doubles the slope of the asymptote from 0.0875 to 0.175, the three tests can only reject with probability 0.16. In Table III, 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%.[17] 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 0.00004. 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.

Table III about here

III. Tests of Mean-Variance Spanning Under Nonnormality

A. 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)–(25), 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 I. Nevertheless, those results can still provide a very good approximation for the small sample distribution of the nonnormality case. To illustrate this,

[17]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.

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.[18] In Table IV, 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 IV, 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 derived in Section I even though ϵ_t does not have a multivariate normal distribution.[19]

Table IV about here

B. Conditional Heteroskedasticity

When ϵ_t exhibits conditional heteroskedasticity, the earlier three test statistics, (23)–(25), will no longer be asymptotically χ^2_{2N} distributed under the null hypothesis.[20] 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

$$E[g_t] = E[x_t \otimes \epsilon_t] = 0_{(K+1)N}. \tag{47}$$

[18]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.

[19]For some distributions of ϵ_t, Dufour and Khalaf (2001) 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.

[20]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.

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

$$\bar{g}_T(B) = \frac{1}{T}\sum_{t=1}^{T} x_t \otimes (R_{2t} - B'x_t) \tag{48}$$

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 I. In the Appendix, we show that the GMM version of the Wald test can be written as

$$W_a = T\text{vec}(\hat{\Theta}')' \left[(A_T \otimes I_N)S_T(A_T' \otimes I_N)\right]^{-1} \text{vec}(\hat{\Theta}') \stackrel{A}{\sim} \chi^2_{2N}, \tag{49}$$

where

$$A_T = \begin{bmatrix} 1 + \hat{a}_1 & -\hat{\mu}_1\hat{V}_{11}^{-1} \\ \hat{b}_1 & -1_K'\hat{V}_{11}^{-1} \end{bmatrix}. \tag{50}$$

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 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.

C. 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.[21]

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*

$$\kappa = \frac{E[((R_t-\mu)'V^{-1}(R_t-\mu))^2]}{(N+K)(N+K+2)} - 1. \tag{51}$$

Then the GMM Wald test of spanning is given by

$$W_a^e = T\mathrm{tr}(\hat{H}\hat{G}_a^{-1}) \overset{A}{\sim} \chi^2_{2N}, \tag{52}$$

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

$$\hat{G}_a = \begin{bmatrix} 1+(1+\hat{\kappa})\hat{a}_1 & (1+\hat{\kappa})\hat{b}_1 \\ (1+\hat{\kappa})\hat{b}_1 & (1+\hat{\kappa})\hat{c}_1 \end{bmatrix}, \tag{53}$$

where $\hat{\kappa}$ is a consistent estimate of κ.[22]

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 (49) is valid for all distributions

[21]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.

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

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 $W = T\text{tr}(\hat{H}\hat{G}^{-1})$ is greater than the GMM Wald test W_a^e.[23] 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.[24] 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 V, 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.

Table V about here

As we can see from Table V, 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 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 V also reports the average ratios of W/W_a and W/W_a^e. To understand how what values these average ratios should take, we show in the Appendix that the limit of the expected bias of

[23]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.

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

the regular Wald test under the multivariate Student-t distribution is

$$\lim_{T\to\infty} E\left[\frac{W}{W_a}\right] - 1 = \lim_{T\to\infty} E\left[\frac{W}{W_a^e}\right] - 1 \approx \frac{\kappa}{2} = \frac{1}{\nu - 4} \tag{54}$$

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.[25] 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 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 V 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.

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

[25] The asymptotic variance of $\hat{\alpha}$ is given in (A28) of the Appendix. For the special case of $K = 1$, this expression is given in MacKinlay and Richardson (1991).

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.

Figure 5 about here

IV. SDF Based Tests of Spanning

A. Equivalent Hypotheses of Spanning

DeSantis (1993), Ferson (1995), and Bekaert and Urias (1996, BU hereafter) exploit the duality of the Hansen-Jagannathan (1991) bound and the mean-variance frontier and suggest equivalent hypotheses of spanning that can be tested by using the GMM. Starting with the BU approach, they project a stochastic discount factor m_t with mean c on the returns of $N+K$ assets as

$$m_t = c + (R_t - \mu)'\beta(c) + \varepsilon_t, \tag{55}$$

where c is a constant. Under no arbitrage, we have

$$E[(1_{N+K} + R_t)m_t] = 1_{N+K} \tag{56}$$

and hence $\beta(c)$ is given by

$$\beta(c) = V^{-1}[(1-c)1_{N+K} - c\mu]. \tag{57}$$

In their alternative spanning test, BU choose two distinct values of c, c_1 and c_2, and test

$$H_1: \ Q\beta(c_1) = 0_N \ \text{ and } \ Q\beta(c_2) = 0_N, \tag{58}$$

where $Q = [\mathrm{O}_{N\times K},\ I_N]$. In essence, BU test mean-variance spanning by examining whether the N test assets can help to explain the variance of the stochastic discount factor.

BU prove that H_0 and H_1 are equivalent. We provide an alternative proof here that has a simple geometric interpretation.[26] To understand what $\beta(c_1)$ and $\beta(c_2)$ represent, we note that

$$\beta(c) = (1-c)V^{-1}1_{N+K} - cV^{-1}\mu = (1-c)(1_{N+K}'V^{-1}1_{N+K})w_2 - c(1_{N+K}'V^{-1}\mu)w_1, \tag{59}$$

where w_1 and w_2 are the weights of two frontier portfolios defined in (5) and (6).[27] Therefore, $\beta(c_1)$ and $\beta(c_2)$ are just two different linear combinations of the weights of two frontier portfolios of the $N+K$ assets. With this expression, it is easy to see that H_1 is equivalent to the hypothesis of $Qw_1 = 0_N$ and $Qw_2 = 0_N$, and hence it is also equivalent to $H_0: \alpha = 0_N$ and $\delta = 0_N$.

[26] See also Ferson (1995) for yet another alternative proof and discussion.

[27] Alternatively, one can write $\beta(c) = -cV^{-1}\left[\mu - \left(1 - \frac{1}{c}\right)1_{N+K}\right]$ and we can see that $\beta(c)$ is proportional to the weight of a tangency portfolio, where the y-intercept of the tangent line is $1 - \frac{1}{c}$.

DeSantis (1993) uses a somewhat different approach. He projects the stochastic discount factor on the (gross) returns as

$$m_t = c + (1_{N+K} + R_t)'\gamma(c) + \varepsilon_t. \tag{60}$$

Note that unlike the BU approach, the expected return μ does not appear as a parameter in DeSantis' specification and c is not the mean of the stochastic discount factor m_t. Using (56), we have

$$\gamma(c) = U^{-1}[(1-c)1_{N+K} - c\mu], \tag{61}$$

where

$$U = E[(1_{N+K} + R_t)(1_{N+K} + R_t)'] = V + (1_{N+K} + \mu)(1_{N+K} + \mu)'. \tag{62}$$

Similar to BU's approach, DeSantis (1993) tests mean-variance spanning by choosing two distinct values of c, c_1 and c_2, and test

$$H_2: \; Q\gamma(c_1) = 0_N \;\text{ and }\; Q\gamma(c_2) = 0_N. \tag{63}$$

In the Appendix, we show that like $\beta(c_1)$ and $\beta(c_2)$, $\gamma(c_1)$ and $\gamma(c_2)$ are also two different linear combinations of the weights of two frontier portfolios of the $N+K$ assets. Therefore, H_2 is equivalent to the hypothesis of $Qw_1 = 0_N$ and $Qw_2 = 0_N$, and hence it is also equivalent to $H_0 : \alpha = 0_N$ and $\delta = 0_N$.

B. GMM Tests of Spanning under the SDF Approach

Although the two hypotheses H_1 and H_2 based on the SDF approach are equivalent to the hypothesis H_0 in the regression approach, there are important differences between these two approaches. To test mean-variance spanning, one needs to make ancillary assumptions. Under the regression approach, one assumes α and β are constant over time. For $\beta = V_{21}V_{11}^{-1}$ to be constant over time, one does not need to assume V is constant over time. For example, if (R_t, z_{t-1}) is jointly elliptically distributed where z_{t-1} is the relevant information at time $t-1$, then V is time-varying as a function of z_{t-1} but yet β is still constant over time. Similarly, for $\alpha = \mu_2 - \beta\mu_1$ to be constant over time, one does not need μ to be constant over time. Under the SDF approach, one assumes either $\beta(c)$ or $\gamma(c)$ is constant over time for all c. This implies $V^{-1}1_{N+K}$ and $V^{-1}\mu$ are both constant over time and hence the weights of the frontier portfolios must also be constant over time. This is a stronger ancillary assumption than the constant α and β assumption used by the regression approach. While the regression approach still does not allow for arbitrary time-varying μ and V, it places less restrictive ancillary assumption than the SDF approach.

With these remarks in mind, we now turn to the GMM estimation and tests by using the SDF approach assuming μ and V are constant over time. The sample moment conditions used by BU are

$$\bar{h}_T(\beta(c_1), \beta(c_2)) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^T R_t(c_1 + (R_t - \hat{\mu})'\beta(c_1)) - (1 - c_1)1_{N+K} \\ \frac{1}{T}\sum_{t=1}^T R_t(c_2 + (R_t - \hat{\mu})'\beta(c_2)) - (1 - c_2)1_{N+K} \end{bmatrix}. \tag{64}$$

Note that instead of treating the expected return μ as a parameter, BU use the sample average return $\hat{\mu}$ in constructing these sample moment conditions. The standard approach should treat μ as a parameter and uses the following sample moment conditions

$$\bar{h}^*_T(\beta(c_1), \beta(c_2), \mu) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^T R_t(c_1 + (R_t - \mu)'\beta(c_1)) - (1 - c_1)1_{N+K} \\ \frac{1}{T}\sum_{t=1}^T R_t(c_2 + (R_t - \mu)'\beta(c_2)) - (1 - c_2)1_{N+K} \\ \frac{1}{T}\sum_{t=1}^T (R_t - \mu) \end{bmatrix}. \tag{65}$$

Since the system is exactly identified, it is easy to show that the unconstrained estimates of $\beta(c_1)$ and $\beta(c_2)$ are given by

$$\hat{\beta}(c_1) = \hat{V}^{-1}[(1 - c_1)1_{N+K} - c_1\hat{\mu})], \tag{66}$$

$$\hat{\beta}(c_2) = \hat{V}^{-1}[(1 - c_2)1_{N+K} - c_2\hat{\mu})]. \tag{67}$$

Although the unconstrained estimates of $\beta(c_1)$ and $\beta(c_2)$ are the same regardless of whether we use $\bar{h}_T$ or $\bar{h}^*_T$, using $\bar{h}^*_T$ allows us to come up with the correct asymptotic variance for $\hat{b} = (\hat{\beta}(c_1)',\ \hat{\beta}(c_2)')'$ by incorporating the estimation error of $\hat{\mu}$. However, for the purpose of testing the spanning hypothesis H_1, one can disregard the errors-in-variables adjustment. Let $S = \text{Avar}(\bar{h}^*_T)$ and write

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \tag{68}$$

where S_{11} corresponds to the first two blocks of sample moments of the pricing equation, and S_{22} corresponds to the sample moments for estimating the expected return.[28] The following proposition summarizes the results.

Proposition 3: *The GMM estimator of $(\beta(c_1)',\ \beta(c_2)')'$ in (65) has asymptotic variance*

$$\text{Avar}(\hat{b}) = C^{-1}(S_{11} - F'S_{21} - S_{12}F + F'S_{22}F)C^{-1}, \tag{69}$$

[28] It is important to note that S is a singular matrix with rank $3(N + K) - 1$, as shown in Peñaranda and Sentana (2001). Nevertheless, as we show in the appendix, the fact that S is singular does not affect the asymptotic distribution of the test statistics in Proposition 3.

where $C = I_2 \otimes V$ and $F = [\beta(c_1)\mu',\ \beta(c_2)\mu']$. The GMM Wald test of $H_1 : Q\beta(c_1) = 0_N$, $Q\beta(c_2) = 0_N$ is given by

$$J_1 = T(\hat{b}'Q_2'[Q_2\hat{C}^{-1}(\hat{S}_{11} - \hat{F}'\hat{S}_{21} - \hat{S}_{12}\hat{F} + \hat{F}'\hat{S}_{22}\hat{F})\hat{C}^{-1}Q_2']^{-1}Q_2\hat{b}) \overset{A}{\sim} \chi^2_{2N}, \tag{70}$$

where $Q_2 = I_2 \otimes Q$, and $\hat{C}$, $\hat{F}$, $\hat{S}$ are consistent estimators of C, F, and S, respectively. Under the null hypothesis, J_1 is asymptotically equivalent to

$$J_2 = T(\hat{b}'Q_2'(Q_2\hat{C}^{-1}\hat{S}_{11}\hat{C}^{-1}Q_2')^{-1}Q_2\hat{b}) \overset{A}{\sim} \chi^2_{2N}, \tag{71}$$

the GMM Wald test obtained by BU from (64) without the errors-in-variables adjustment.

The result that we can ignore the errors-in-variables problem for testing the spanning hypothesis is similar to a result in Shanken (1992), where he shows that in the two-pass methodology, one can ignore estimation errors on betas when it comes to testing the beta risk premium is equal to zero. Although both J_1 and J_2 are asymptotically valid under the null hypothesis, their size as well as their power could be quite different in finite samples. We address these issues in the next subsection using simulation. It is easy to show that J_1 and J_2 are numerically independent of the choice of c_1 and c_2. So without loss of generality, we can choose $c_1 = 0$ and $c_2 = 1$. In this case, we have $\hat{\beta}(0) = \hat{V}^{-1}1_{N+K}$ and $\hat{\beta}(1) = -\hat{V}^{-1}\hat{\mu}$, and they are proportional to the weights of the two frontier portfolios that we discussed earlier. Therefore, BU's approach to testing mean-variance spanning is to directly test whether the two frontier portfolios contain zero weights in the N test assets.

Under DeSantis' parameterization, the sample moment conditions are

$$\bar{m}_T(\gamma(c_1), \gamma(c_2)) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T}(1 + R_t)(c_1 + (1 + R_t)'\gamma(c_1)) - 1_{N+K} \\ \frac{1}{T}\sum_{t=1}^{T}(1 + R_t)(c_2 + (1 + R_t)'\gamma(c_2)) - 1_{N+K} \end{bmatrix}. \tag{72}$$

The resulting unconstrained estimates of $\gamma(c_1)$ and $\gamma(c_2)$ are given by

$$\hat{\gamma}(c_1) = \hat{U}^{-1}[1_{N+K} - c_1(1_{N+K} + \hat{\mu})], \tag{73}$$

$$\hat{\gamma}(c_2) = \hat{U}^{-1}[1_{N+K} - c_2(1_{N+K} + \hat{\mu})], \tag{74}$$

where $\hat{U} = \frac{1}{T}\sum_{t=1}^{T}(1 + R_t)(1 + R_t)'$. Note that in DeSantis' specification, μ does not appear as a parameter and therefore we do no need to take care of the errors-in-variables problem as in BU's specification. Under this specification, the GMM Wald test of $H_2:\ Q\gamma(c_1)=0_N,\ Q\gamma(c_2)=0_N$ is given by

$$J_3=T(\hat{g}'Q_2'[(I_2\otimes Q\hat{U}^{-1})\hat{S}_m(I_2\otimes \hat{U}^{-1}Q')]^{-1}Q_2\hat{g})\overset{A}{\sim}\chi^2_{2N}, \tag{75}$$

where $\hat{g}=(\hat{\gamma}(c_1)',\ \hat{\gamma}(c_2)')'$ and $\hat{S}_m$ is a consistent estimate of the asymptotic variance of $\bar{m}_T$. Similar to J_1 and J_2, J_3 is also numerically independent of the choice of c_1 and c_2. Without loss of generality, we can choose $c_1=-(\hat{\mu}'\hat{V}^{-1}1_{N+K}+1_{N+K}'\hat{V}^{-1}1_{N+K})$ and $c_2=1+\hat{\mu}'\hat{V}^{-1}\hat{\mu}+\hat{\mu}'\hat{V}^{-1}1_{N+K}$. With this choice of c_1 and c_2, we can verify that $\hat{\gamma}(c_1)=\hat{V}^{-1}1_{N+K}$ and $\hat{\gamma}(c_2)=-\hat{V}^{-1}\hat{\mu}$, and hence they are also proportional to the weights of the two frontier portfolios that we discussed earlier. Therefore, similar to BU's approach, DeSantis' approach also tests mean-variance spanning by directly testing whether the two frontier portfolios contain zero weights in the N test assets.

In both BU and DeSantis specifications, one can also perform a Lagrange multiplier test or an over-identification test of the spanning hypothesis. However, such tests would require iterations and the specification of an initial weighting matrix. For a general initial weighting matrix, the test results are not independent of the choice of c_1 and c_2.[29] Due to the possible ambiguity of the Lagrange multiplier and over-identification tests, we will not present their results here.

C. Size and Power of SDF Tests of Spanning

In this subsection, we study the size and power of the GMM Wald tests of spanning under the SDF approach, J_1 through J_3, and compare them with the GMM Wald test under the regression approach, W_a. We simulate returns from a multivariate normal distribution with parameters chosen to satisfy the null hypothesis. To make a fair comparison between the SDF based Wald tests and the regression based Wald test, we do not use the information of the normal return distribution in constructing W_a. Instead of using (24) which is valid only under conditional homoskedasticity assumption, the regression based GMM Wald test we use is the W_a in (49). In Table VI, we present the actual probabilities of rejection of J_1 to J_3 and W_a in 100,000 simulations, for different values of K, N, and T, when the rejection decision is based on the 95th percentile of the χ^2_{2N} distribution. As we can see from Table VI, all four GMM Wald tests grossly over-reject the null hypothesis. Although not reported, this over-rejection problem is even more severe when returns

[29] When the initial weighting matrix is an identity matrix, the Lagrange multiplier tests and over-identification tests are indeed independent of the choice of c_1 and c_2, but this is not true for a general initial weighting matrix.

are multivariate Student-t distributed. Therefore, when using the GMM Wald tests of spanning, one must be cautious when using the asymptotic distribution for making the acceptance/rejection decision.

Table VI about here

Although this over-rejection problem of the Wald tests is severe when using their asymptotic distribution, one should not be overly concerned if one can simulate their empirical distributions under the null hypothesis and use them for acceptance/rejection decision. The more important concern is the relative power of these tests. We perform the same simulation experiment as before by generating returns from two benchmark assets as in Figure 3 and a single test asset constructed under different alternative hypotheses. In Figure 6, we generate the returns from a multivariate normal distribution and plot the power functions of J_1 to J_3 and compare them with W_a. As we can see in Figure 6, there are no important differences between the power of the four tests when the returns are multivariate normally distributed, especially when $T = 120$. Therefore, in this case, one could use any of the tests as long as one uses the empirical distribution under the null hypothesis to make the rejection decision. The similarity of the power functions also suggests that the SDF based tests of spanning have good power in rejecting alternative hypotheses that generate even a small reduction in the variance of the global minimum-variance portfolio but little power against alternative hypotheses that generate large difference in the tangency portfolios.

Figure 6 about here

The same conclusion does not hold when returns are not normally distributed. We repeat the same simulation exercise but using returns generated from a multivariate Student-t distribution with five degrees of freedom. The power functions of the four tests are plotted in Figure 7. Here the GMM Wald test under the regression approach, W_a, has the best power in all cases. As for the three SDF based GMM Wald tests, J_2 has better power for alternative hypotheses that have $\alpha \neq 0$, whereas J_3 has better power for alternative hypotheses that have $\delta \neq 0$. While it is not possible to say that the regression based GMM Wald test is better than the SDF based GMM Wald tests for all cases, our simulation evidence seems to prefer the spanning test under the regression approach, especially when returns have a multivariate Student-t distribution.

Figure 7 about here

V. 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 1999. Monthly return data for the S&P 500 index and the 30-year U.S. Treasury bond are obtained from the Center for Research in Security Prices (CRSP). Monthly return data for the equity market of the other seven developed countries are obtained from Morgan Stanley Capital International (MSCI), and they are all converted into U.S. dollar returns.[30]

In Figure 8, 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. Average return and the standard deviation of the other seven developed countries are also indicated in the figure. From Figure 8, we can see that over the 30-year sample period, the Japanese equity market had the highest average return (16.2%/year), whereas the 30-year U.S. Treasury bond had the lowest average return (8.7%/year). Although we observe that all seven foreign equity markets lie within the frontier formed by the U.S. bond and equity, it is possible that the U.S. investor can expand its opportunity set by introducing some foreign equity into his portfolio.

Figure 8 about here

In Table VII, 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 that we suggest in Section II.C. The tests are performed using monthly data over the 30-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

[30]We thank Campbell Harvey for sharing this data set with us.

F-test rejects spanning at the 5% level for Australia, France, Germany, Italy and Japan but not for Canada and the U.K. 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 the U.K., 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 VII, the F_1 tests can only reject $\alpha = 0_N$ for France, Germany, and Japan but the F_2 tests can reject $\delta = 0_N$ for all cases. 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 much weaker evidence that the tangency portfolio can be improved.

Table VII about here

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 VII which based on the normality assumption may not be appropriate. To determine the robustness of the results, we present in Table VIII some asymptotic spanning tests that do not rely on the normality assumption. We report two regression based and two SDF based GMM Wald tests. The regression based Wald tests that we report are W_a^e (which is only valid when returns follow a multivariate elliptical distribution) and W_a. The SDF based GMM Wald tests that we report are J_2 of BU and J_3 of DeSantis.[31] Consistent with results in Table V, 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. As for the SDF based GMM Wald tests, we find that J_3 are all larger than J_2 and when

[31] We do not report J_1 (the BU test with EIV adjustment) because it is numerically very close to J_2.

asymptotic distribution is used, J_3 favors rejection of the spanning hypothesis. This is consistent with the simulation results in Table VI. Keeping in mind that the reported p-values of these tests are only asymptotic, we compare the test results in Table VIII with those in Table VII. We find that once we correct for conditional heteroskedasticity in the Wald tests, the evidence against rejection of spanning in Table VII is further weakened, indicating that there could be over-rejection problems in Table VII due to nonnormality of returns. Nevertheless, the asymptotic tests in Table VIII still can jointly reject spanning for the seven foreign equity indices in almost every case, indicating the rejection in Table VII is robust to conditional heteroskedasticity in the returns.

Table VIII about here

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.

VI. 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. In addition, we provide a linkage between the traditional regression based spanning tests with those that are based on the newer SDF approach. The finite sample properties and the power studies of these two types of tests are compared and we find evidence that the spanning tests under the traditional regression approach can be superior to the ones under the SDF approach, especially when returns follow a multivariate Student-t distribution.

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

Proof of (29): 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

$$\Gamma(k)\Gamma\left(k-\frac{1}{2}\right) = \frac{\sqrt{\pi}\Gamma(2k-1)}{2^{2k-2}} \tag{A1}$$

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

$$f(\xi_1,\xi_2) = \frac{n+1}{2\mathrm{B}(2m+2,2n+3)}\left[\prod_{i=1}^{2}\xi_i^m(1-\xi_i)^n\right](\xi_1-\xi_2) \quad \text{for} \quad 1 \ge \xi_1 \ge \xi_2 \ge 0, \tag{A2}$$

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

$$f(a_1,a_2) = \frac{n+1}{2\mathrm{B}(2m+2,2n+3)}a_2^m(1-a_1+a_2)^n. \tag{A3}$$

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

$$\begin{aligned}
& P[a_1 \le v] \\
= \; & \frac{n+1}{2\mathrm{B}(2m+2,2n+3)}\int_0^{\frac{v^2}{4}}\int_{2\sqrt{a_2}}^{\min[v,1+a_2]} a_2^m(1-a_1+a_2)^n \mathrm{d}a_1\mathrm{d}a_2 \\
= \; & \frac{n+1}{2\mathrm{B}(2m+2,2n+3)}\int_0^{\frac{v^2}{4}} a_2^m\left[\left[\frac{(1-a_1+a_2)^{n+1}}{n+1}\right]\Bigg|_{\min[v,1+a_2]}^{2\sqrt{a_2}}\right]\mathrm{d}a_2 \\
= \; & \frac{1}{2\mathrm{B}(2m+2,2n+3)}\left[\int_0^{\frac{v^2}{4}} a_2^m(1-\sqrt{a_2})^{2n+2}\mathrm{d}a_2 - \int_{\max[0,v-1]}^{\frac{v^2}{4}} a_2^m(1-v+a_2)^{n+1}\mathrm{d}a_2\right] \\
= \; & I_{\frac{v}{2}}(2m+2,2n+3) - \frac{1}{2\mathrm{B}(2m+2,2n+3)}\int_{\max[0,v-1]}^{\frac{v^2}{4}} a_2^m(1-v+a_2)^{n+1}\mathrm{d}a_2.
\end{aligned} \tag{A4}$$

This completes the proof. *Q.E.D.*

Proof of Lemma 1: 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 formula, it is easy to verify that

$$\hat{V}^{-1} = \begin{bmatrix} \hat{V}_{11}^{-1} + \hat{\beta}'\hat{\Sigma}^{-1}\hat{\beta} & -\hat{\beta}'\hat{\Sigma}^{-1} \\ -\hat{\Sigma}^{-1}\hat{\beta} & \hat{\Sigma}^{-1} \end{bmatrix} = \begin{bmatrix} \hat{V}_{11}^{-1} & \mathrm{O}_{K\times N} \\ \mathrm{O}_{N\times K} & \mathrm{O}_{N\times N} \end{bmatrix} + \begin{bmatrix} -\hat{\beta}' \\ I_N \end{bmatrix}\hat{\Sigma}^{-1}[-\hat{\beta} \;\; I_N]. \tag{A5}$$

Therefore,

$$
\begin{aligned}
& \begin{bmatrix} \hat{a} & \hat{b} \\ \hat{b} & \hat{c} \end{bmatrix} \\
= & \begin{bmatrix} \hat{\mu}' \\ 1'_{N+K} \end{bmatrix} \hat{V}^{-1}[\hat{\mu}\ 1_{N+K}] \\
= & \begin{bmatrix} \hat{\mu}' \\ 1'_{N+K} \end{bmatrix} \begin{bmatrix} \hat{V}_{11}^{-1} & \mathrm{O}_{K\times N} \\ \mathrm{O}_{N\times K} & \mathrm{O}_{N\times N} \end{bmatrix} [\hat{\mu}\ 1_{N+K}] + \begin{bmatrix} \hat{\mu}' \\ 1'_{N+K} \end{bmatrix} \begin{bmatrix} -\hat{\beta}' \\ I_N \end{bmatrix} \hat{\Sigma}^{-1}[-\hat{\beta}\ I_N][\hat{\mu}\ 1_{N+K}] \\
= & \begin{bmatrix} \hat{\mu}_1' \\ 1_K' \end{bmatrix} \hat{V}_{11}^{-1}[\hat{\mu}_1\ 1_K] + \begin{bmatrix} (\hat{\mu}_2 - \hat{\beta}\hat{\mu}_1)' \\ (1_N - \hat{\beta}1_K)' \end{bmatrix} \hat{\Sigma}^{-1}[\hat{\mu}_2 - \hat{\beta}\hat{\mu}_1\ \ 1_N - \hat{\beta}1_K] \\
= & \begin{bmatrix} \hat{a}_1 & \hat{b}_1 \\ \hat{b}_1 & \hat{c}_1 \end{bmatrix} + \hat{H}.
\end{aligned}
\tag{A6}
$$

This completes the proof. *Q.E.D.*

Proof of (36) and (37): Since λ_1 and λ_2 are the two eigenvalues of $\hat{H}\hat{G}^{-1}$, they are the solutions to the following equation

$$|\hat{H}\hat{G}^{-1} - \lambda I_2| = 0, \tag{A7}$$

or equivalently the solutions to

$$|\hat{H} - \lambda\hat{G}| = \begin{vmatrix} \Delta\hat{a} - \lambda(1+\hat{a}_1) & \Delta\hat{b} - \lambda\hat{b}_1 \\ \Delta\hat{b} - \lambda\hat{b}_1 & \Delta\hat{c} - \lambda\hat{c}_1 \end{vmatrix} = 0. \tag{A8}$$

Simplifying, we have

$$(\hat{c}_1 + \hat{d}_1)\lambda^2 - \left[\Delta\hat{a}\hat{c}_1 - 2\Delta\hat{b}\hat{b}_1 + \Delta\hat{c}(1+\hat{a}_1)\right]\lambda + \left[\Delta\hat{a}\Delta\hat{c} - (\Delta\hat{b})^2\right] = 0. \tag{A9}$$

It is easy to see that

$$
\begin{aligned}
\lambda_1 + \lambda_2 &= \frac{\Delta\hat{a}\hat{c}_1 - 2\Delta\hat{b}\hat{b}_1 + \Delta\hat{c}(1+\hat{a}_1)}{\hat{c}_1 + \hat{d}_1} \\
&= \frac{\Delta\hat{a} - 2\Delta\hat{b}\hat{\mu}_{g_1} + \Delta\hat{c}\hat{\mu}_{g_1}^2}{1 + \frac{\hat{d}_1}{\hat{c}_1}} + \frac{\Delta\hat{c}\left(\frac{1+\hat{a}_1}{\hat{c}_1} - \hat{\mu}_{g_1}^2\right)}{1 + \frac{\hat{d}_1}{\hat{c}_1}} \\
&= \frac{\hat{\theta}^2(\hat{\mu}_{g_1}) - \hat{\theta}_1^2(\hat{\mu}_{g_1})}{1 + \hat{\theta}_1^2(\hat{\mu}_{g_1})} + \frac{\Delta\hat{c}}{\hat{c}_1},
\end{aligned}
\tag{A10}
$$

where the last equality follows from the fact that

$$\hat{\theta}^2(r) - \hat{\theta}_1^2(r) = (\hat{a} - 2\hat{b}r + \hat{c}r^2) - (\hat{a}_1 - 2\hat{b}_1 r + \hat{c}_1 r^2) = \Delta\hat{a} - 2\Delta\hat{b}r + \Delta\hat{c}r^2. \tag{A11}$$

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

$$|\hat{H} - \xi(\hat{H}+\hat{G})| = \begin{vmatrix} \Delta\hat{a} - \xi(1+\hat{a}) & \Delta\hat{b} - \xi\hat{b} \\ \Delta\hat{b} - \xi\hat{b} & \Delta\hat{c} - \xi\hat{c} \end{vmatrix} = 0. \tag{A12}$$

Comparing (A8) with (A12), 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 (A10). This completes the proof. *Q.E.D.*

Proof of (40): 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

$$\begin{aligned} f(\lambda_1,\lambda_2) &= e^{-\mathrm{tr}(\Omega)/2}\,{}_1F_1\left(\frac{T-K+1}{2};\frac{N}{2};\frac{\Omega}{2},L(I_2+L)^{-1}\right)\times \\ &\qquad \frac{N-1}{4\mathrm{B}(N,T-K-N)}\left[\prod_{i=1}^{2}\frac{\lambda_i^{\frac{N-3}{2}}}{(1+\lambda_i)^{\frac{T-K+1}{2}}}\right](\lambda_1-\lambda_2), \end{aligned} \tag{A13}$$

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

$$\Omega = T\hat{G}^{-\frac{1}{2}}\Theta\Sigma^{-1}\Theta'\hat{G}^{-\frac{1}{2}}. \tag{A14}$$

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. *Q.E.D.*

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 (40). 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 $L_{11} \sim \sqrt{\chi^2_{T-K-N+1}}$ $L_{22} \sim \sqrt{\chi^2_{T-K-N}}$, and $L_{12} \sim \mathrm{N}(0,1)$. 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 $\mathrm{vec}(Z) \sim \mathrm{N}(\mathrm{vec}(D^{\frac{1}{2}}), I_4)$, 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. *Q.E.D.*

Proof of Lemma 2: By replacing $\Delta\hat{a}$, $\Delta\hat{b}$, $\Delta\hat{c}$ by Δa, Δb, and Δc, we have from (A10)

$$\omega_1 + \omega_2 = \frac{\Delta c}{\hat{c}_1} + \frac{\theta^2(\hat{\mu}_{g_1}) - \theta_1^2(\hat{\mu}_{g_1})}{1 + \hat{\theta}_1^2(\hat{\mu}_{g_1})}. \tag{A15}$$

Similarly, with the same replacement, we have from (A9)

$$\omega_1\omega_2 = \frac{\Delta a \Delta c - (\Delta b)^2}{\hat{c}_1 + \hat{d}_1} = \left(\frac{\Delta c}{\hat{c}_1}\right)\left(\frac{\theta^2(\mu_z) - \theta_1^2(\mu_z)}{1 + \hat{\theta}_1^2(\hat{\mu}_{g_1})}\right), \tag{A16}$$

where the last equality follows from the fact that

$$\theta^2(\mu_z) - \theta_1^2(\mu_z) = \Delta a - 2\Delta b\left(\frac{\Delta b}{\Delta c}\right) + \Delta c\left(\frac{\Delta b}{\Delta c}\right)^2 = \Delta a - \frac{(\Delta b)^2}{\Delta c}. \tag{A17}$$

(i) Since under the alternative hypothesis, we have $\omega_1 > 0$. Therefore, from (A16), 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

$$\begin{aligned}
(\omega_1 + \omega_2)^2 &= \left[\frac{\Delta c}{\hat{c}_1} + \frac{\theta^2(\hat{\mu}_{g_1}) - \theta_1^2(\hat{\mu}_{g_1})}{1 + \hat{\theta}_1^2(\hat{\mu}_{g_1})}\right]^2 \\
&\geq 4\left(\frac{\Delta c}{\hat{c}_1}\right)\left(\frac{\theta^2(\hat{\mu}_{g_1}) - \theta_1^2(\hat{\mu}_{g_1})}{1 + \hat{\theta}_1^2(\hat{\mu}_{g_1})}\right) \\
&\geq 4\left(\frac{\Delta c}{\hat{c}_1}\right)\left(\frac{\theta^2(\mu_z) - \theta_1^2(\mu_z)}{1 + \hat{\theta}_1^2(\hat{\mu}_{g_1})}\right) \\
&= 4\omega_1\omega_2.
\end{aligned} \tag{A18}$$

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

$$\frac{\Delta c}{\hat{c}_1} = \frac{\theta^2(\hat{\mu}_{g_1}) - \theta_1^2(\hat{\mu}_{g_1})}{1 + \hat{\theta}_1^2(\hat{\mu}_{g_1})} \tag{A19}$$

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

Proof of the distribution of (45) and (46): 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),

$$\frac{|\bar{\Sigma}|}{|\tilde{\Sigma}|} \sim U_{N,1,T-K} \tag{A20}$$

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

$$F_2 = \left(\frac{T-K-N+1}{N}\right)\left(\frac{|\tilde{\Sigma}|}{|\bar{\Sigma}|} - 1\right) \sim F_{N,T-K-N+1}. \tag{A21}$$

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. *Q.E.D.*

Proof of (49): From Hansen (1982), the asymptotic variance of $\mathrm{vec}(\hat{B}')$ is given by $(D_0' S_0^{-1} D_0)^{-1}$, where

$$D_0 = E\left[\frac{\partial \bar{g}_T(B)}{\partial \mathrm{vec}(B')'}\right] = E[x_t x_t'] \otimes I_N. \tag{A22}$$

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

$$\begin{aligned}(A \otimes I_N)(D_0' S_0^{-1} D_0)^{-1}(A' \otimes I_N) &= (A \otimes I_N) D_0^{-1} S_0 D_0^{-1} (A' \otimes I_N) \\ &= (A(E[x_t x_t'])^{-1} \otimes I_N) S_0 ((E[x_t x_t'])^{-1} A' \otimes I_N). \end{aligned} \tag{A23}$$

Using the partitioned matrix inverse formula, we have

$$\begin{aligned} A(E[x_t x_t'])^{-1} &= \begin{bmatrix} 1 & 0_K' \\ 0 & -1_K' \end{bmatrix} \begin{bmatrix} 1 & \mu_1' \\ \mu_1 & V_{11} + \mu_1 \mu_1' \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 0_K' \\ 0 & -1_K' \end{bmatrix} \begin{bmatrix} 1 + \mu_1' V_{11}^{-1} \mu_1 & -\mu_1' V_{11}^{-1} \\ -V_{11}^{-1}\mu_1 & V_{11}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} 1 + \mu_1' V_{11}^{-1} \mu_1 & -\mu_1 V_{11}^{-1} \\ 1_K' V_{11}^{-1} \mu_1 & -1_K' V_{11}^{-1} \end{bmatrix}. \end{aligned} \tag{A24}$$

Replacing S_0 and $A(E[x_t x_t'])^{-1}$ by their consistent estimates S_T and A_T, we obtain (49). This completes the proof. *Q.E.D.*

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

$$E[R_{1t} R_{1t}' \otimes \epsilon_t \epsilon_t'] = \mu_1 \mu_1' \otimes \Sigma + (1+\kappa) V_{11} \otimes \Sigma = (V_{11} + \mu_1 \mu_1') \otimes \Sigma + \kappa V_{11} \otimes \Sigma, \tag{A25}$$

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

$$S_0 = E[x_t x_t'] \otimes \Sigma + \begin{bmatrix} 0 & 0_K' \\ 0_K & \kappa V_{11} \end{bmatrix} \otimes \Sigma. \tag{A26}$$

Using this expression and (A22), the asymptotic variance of $\mathrm{vec}(\hat{B}')$ is given by

$$(D_0')^{-1} S_0 D_0^{-1} = \begin{bmatrix} 1 + a_1 & -\mu_1' V_{11}^{-1} \\ -V_{11}^{-1} \mu_1 & V_{11}^{-1} \end{bmatrix} \otimes \Sigma + \kappa \begin{bmatrix} a_1 & -\mu_1' V_{11}^{-1} \\ -V_{11}^{-1}\mu_1 & V_{11}^{-1} \end{bmatrix} \otimes \Sigma. \tag{A27}$$

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

$$(A \otimes I_N)(D_0^{-1})' S_0 D_0^{-1} (A' \otimes I_N) = \begin{bmatrix} 1 + (1+\kappa) a_1 & (1+\kappa) b_1 \\ (1+\kappa) b_1 & (1+\kappa) c_1 \end{bmatrix} \otimes \Sigma. \tag{A28}$$

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

$$W_a = T\mathrm{vec}(\hat{\Theta}')'(\hat{G}_a^{-1} \otimes \hat{\Sigma}^{-1})\mathrm{vec}(\hat{\Theta}') = T\mathrm{vec}(\hat{\Theta}')'\mathrm{vec}(\hat{\Sigma}^{-1}\hat{\Theta}'\hat{G}_a^{-1}) = T\mathrm{tr}(\hat{H}\hat{G}_a^{-1}), \tag{A29}$$

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

Proof of (54): 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 (A10),

$$W = \mathrm{tr}(\hat{H}\hat{G}^{-1}) = \frac{\hat{\theta}^2(\hat{\mu}_g) - \hat{\theta}_1^2(\hat{\mu}_g)}{1 + \hat{\theta}_1^2(\hat{\mu}_g)} + \frac{\Delta\hat{c}}{\hat{c}_1}. \tag{A30}$$

Using a similar proof, we have

$$W_a^e = \mathrm{tr}(\hat{H}\hat{G}_a^{-1}) = \frac{\hat{\theta}^2(\hat{\mu}_g) - \hat{\theta}_1^2(\hat{\mu}_g)}{1 + (1+\hat{\kappa})\hat{\theta}_1^2(\hat{\mu}_g)} + \frac{\Delta\hat{c}}{\hat{c}_1(1+\hat{\kappa})} \equiv X_1 + X_2. \tag{A31}$$

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

$$\mathrm{tr}(\hat{H}\hat{G}^{-1}) \approx X_1 + (1+\kappa)X_2, \tag{A32}$$

and hence

$$\lim_{T\to\infty} \frac{W}{W_a^e} - 1 \approx \frac{X_1 + (1+\kappa)X_2}{X_1 + X_2} - 1 = \kappa\left(\frac{X_2}{X_1 + X_2}\right). \tag{A33}$$

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

$$\lim_{T\to\infty} E\left[\frac{W}{W_a^e}\right] - 1 \approx \frac{\kappa}{2}. \tag{A34}$$

This completes the proof. *Q.E.D.*

Proof that $\gamma(c)$ in (61) is a linear combination of w_1 and w_2: Note that we can write U^{-1} as

$$\begin{aligned} U^{-1} &= V^{-1} - \frac{V^{-1}(1_{N+K} + \mu)(1_{N+K} + \mu)'V^{-1}}{1 + (1_{N+K} + \mu)'V^{-1}(1_{N+K} + \mu)} \\ &= V^{-1} - (d_1 w_1 + d_2 w_2)(1_{N+K} + \mu)'V^{-1}, \end{aligned} \tag{A35}$$

where w_1 and w_2 are defined in (5) and (6) and

$$d_1 = \frac{1_{N+K}'V^{-1}\mu}{1 + (1_{N+K} + \mu)'V^{-1}(1_{N+K} + \mu)}, \tag{A36}$$

$$d_2 = \frac{1_{N+K}'V^{-1}1_{N+K}}{1 + (1_{N+K} + \mu)'V^{-1}(1_{N+K} + \mu)}. \tag{A37}$$

Using this identity, we have

$$\begin{aligned}\gamma(c) &= U^{-1}[(1-c)1_{N+K} - c\mu] \\ &= V^{-1}[(1-c)1_{N+K} - c\mu] - (d_1 w_1 + d_2 w_2)(1_{N+K} + \mu)' V^{-1}[(1-c)1_{N+K} - c\mu] \\ &= (1-c)(1'_{N+K} V^{-1} 1_{N+K}) w_2 - c(1'_{N+K} V^{-1} \mu) w_1 \\ &\quad - (1_{N+K} + \mu)' V^{-1}[(1-c)1_{N+K} - c\mu](d_1 w_1 + d_2 w_2), \end{aligned} \tag{A38}$$

which is a linear combination of w_1 and w_2. This completes the proof. *Q.E.D.*

Proof of Proposition 3: Define

$$D_0 = E\left[\frac{\partial \bar{h}_T^*}{\partial \beta(c_1)'}, \frac{\partial \bar{h}_T^*}{\partial \beta(c_2)'}, \frac{\partial \bar{h}_T^*}{\partial \mu'}\right] = \begin{bmatrix} C & -F' \\ \mathrm{O}_{(N+K)\times 2(N+K)} & -I_{N+K} \end{bmatrix}, \tag{A39}$$

we have the asymptotic variance of $(\hat{b},\ \hat{\mu})$ as $(D_0' W D_0)^{-1}(D_0' W S W D_0)(D_0' W D_0)^{-1} = D_0^{-1} S (D_0')^{-1}$ for any positive definite weighting matrix W. Using

$$D_0^{-1} = \begin{bmatrix} C^{-1} & -C^{-1}F' \\ \mathrm{O}_{(N+K)\times 2(N+K)} & -I_{N+K} \end{bmatrix}, \tag{A40}$$

we can express $D_0^{-1} S (D_0')^{-1}$ as

$$\begin{bmatrix} C^{-1}(S_{11} - F'S_{21} - S_{12}F + F'S_{22}F)C^{-1} & -C^{-1}(S_{12} - F'S_{22}) \\ -(S_{21} - S_{22}F)C^{-1} & S_{22} \end{bmatrix}, \tag{A41}$$

and the asymptotic variance of $\hat{b}$ is given by its first submatrix.[32] Since the asymptotic variance of $Q_2\hat{b}$ is

$$Q_2 C^{-1}(S_{11} - F'S_{21} - S_{12}F + F'S_{22}F)C^{-1}Q_2', \tag{A42}$$

the GMM Wald test of H_1 follows by replacing the asymptotic variance of $Q_2\hat{b}$ with its consistent estimate. Under the null hypothesis, we have from (7) that $QV^{-1}\mu = 0_N$ and hence

$$Q_2 C^{-1} F' = \begin{bmatrix} QV^{-1}\mu\beta(c_1)' \\ QV^{-1}\mu\beta(c_2)' \end{bmatrix} = \mathrm{O}_{2N\times(N+K)}. \tag{A43}$$

Therefore, the asymptotic variance of $Q_2\hat{b}$ can be simplified to $Q_2 C^{-1} S_{11} C^{-1} Q_2'$ under the null hypothesis and J_2 is asymptotically equivalent to J_1. This completes the proof. *Q.E.D.*

[32] Note that although S is a singular matrix, Avar$[\hat{b}]$ and Avar$[Q_2\hat{b}]$ are nonsingular and hence the test statistic J_1 is nevertheless valid.

References

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

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

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

Blattberg, Robert C., and Nicholas 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, Mark, 1999, The sampling error in estimates of mean-variance efficient portfolio weights, *Journal of Finance* 54, 655–671.

De Roon, Frans A., Theo E. Nijman, and Bas 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, Giorgio, 1993, Volatility bounds for stochastic discount factors: tests and implications from international financial markets, Ph.D. dissertation, Department of Economics, University of Chicago.

Dufour, Jean-Marie, and Lynda Khalaf, 2001, Simulation based finite and large sample tests in multivariate regressions, *Journal of Econometrics* forthcoming.

Errunza, Vihang, Ked Hogan, and Mao-Wei Hung, 1999, Can the gains from international diversification be achieved without trading abroad? *Journal of Finance* 54, 2075–2107.

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

Ferson, Wayne, 1995, Theory and empirical testing of asset pricing models, *The Finance Handbook* (Robert A. Jarrow, William T. Ziemba, and Vojislav Maksimovich, ed.), North Holland Publishers, Amsterdam.

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

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

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

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

Hansen, Lars Peter, and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

Hotelling, Harold, 1951, A generalized T test and measure of multivariate dispersion, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Jerzy Neyman, ed.), University of California, Los Angeles and Berkeley, 23–41.

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

Jobson, J. D., and Bob Korkie, 1982, Potential performance and tests of portfolio efficiency, *Journal of Financial Economics* 10, 433–466.

Jobson, J. D., and Bob 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 Bob 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, Shmuel, and Robert F. Stambaugh, 1992, A mean-variance framework for tests of asset pricing models, *Review of Financial Studies* 2, 125–156.

Korkie, Bob, and Harry J. Turtle, 2001, An analysis of self-financing portfolios in a mean-variance framework, Working Paper, University of Alberta.

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

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

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

Mathai, A. M., Serge B. Provost, and Takesi Hayakawa, 1995, *Bilinear forms and zonal polynomials* (Springer-Verlag, New York).

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

Merton, Robert 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, Robb J., 1982, *Aspects of multivariate statistical theory* (Wiley, New York).

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

Peñaranda, Francisco, and Enrique Sentana, 2001, Spanning tests in return and stochastic discount mean-varaince 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. Sreedharan, and Kanta Jayachandran, 1967, Power comparison of tests of two multivariate hypotheses based on four criteria, *Biometrika* 54, 195–210.

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

Roll, Richard, 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, Takashi, and Takuya Toyama, 1996, On the estimation of kurtosis parameter in elliptical distributions, *Journal of the Japan Statistical Society* 26, 59–68.

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

Zhou, Guofu, 1993, Asset pricing test under alternative distributions, *Journal of Finance* 48, 1925–1942.

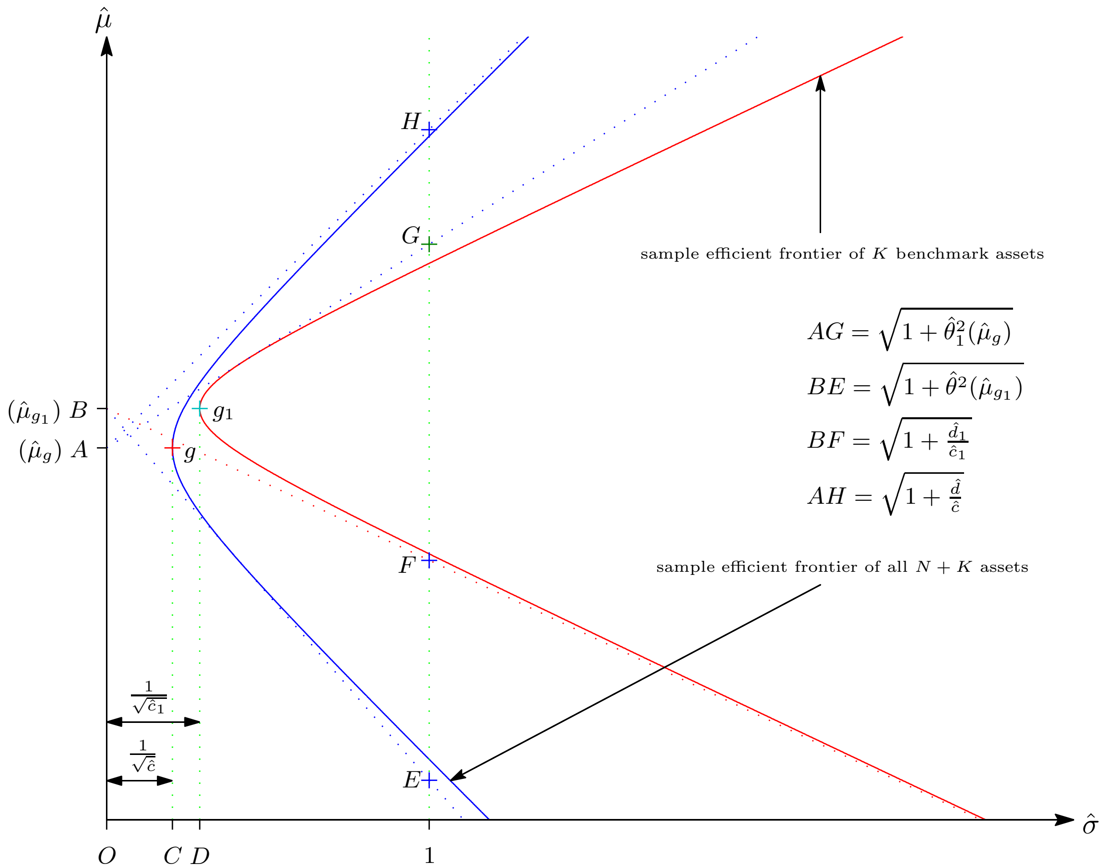

Figure 1
The Geometry of Mean-Variance Spanning Tests
The figure plots the *ex post* minimum-variance frontier hyperbola of K benchmark assets and that of all $N+K$ assets on the $(\hat{\sigma}, \hat{\mu})$ space. The constants that determine the hyperbola of K benchmark assets are $\hat{a}_1 = \hat{\mu}_1'\hat{V}_{11}\hat{\mu}_1$, $\hat{b}_1 = \hat{\mu}_1'\hat{V}_{11}1_K$, $\hat{c}_1 = 1_K'\hat{V}_{11}1_K$, 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 $\hat{\mu}_{g_1}$.

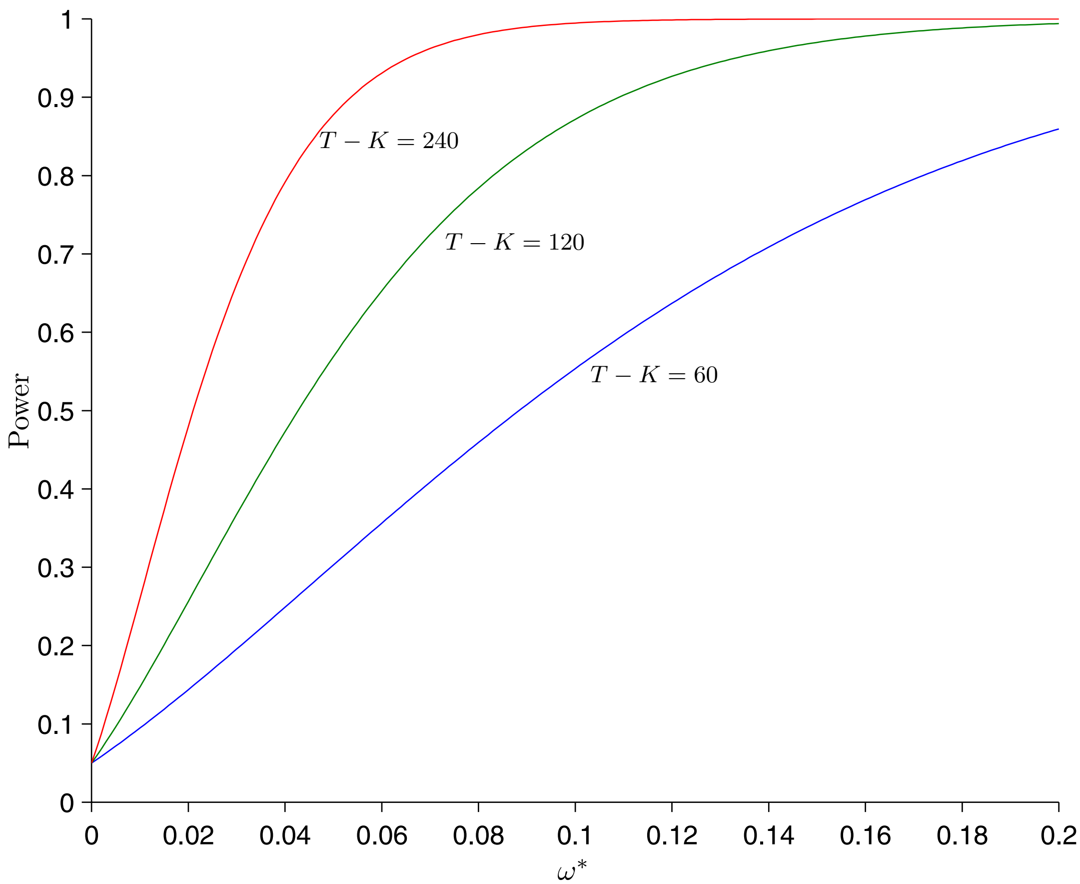

Figure 2
Power Function of Mean-Variance Spanning Test with Single Test Asset
The figure plots the probability of rejecting the null hypothesis of mean-variance spanning as a function of ω^* for three different values of $T - K$ (the number of time series observations minus the number of benchmark assets), 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.

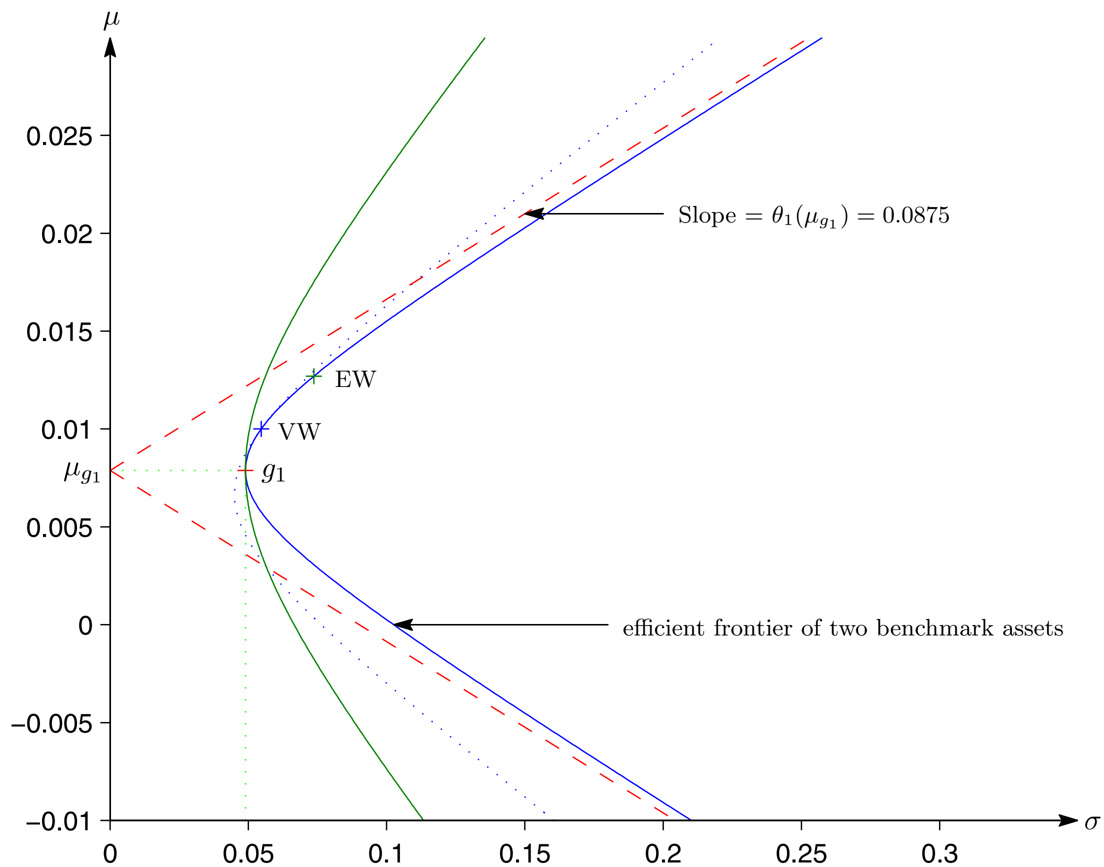

Figure 3
Minimum-Variance Frontier of Two Benchmark Assets
The figure plots the minimum-variance frontier hyperbola of two benchmark assets in the (σ, μ) space. The two benchmark assets are the value-weighted (VW) and equally weighted (EW) portfolios of the NYSE. g_1 is the global minimum-variance portfolio 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 1926/1–1998/12. 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 4.9%/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 0.0875 to 0.175.

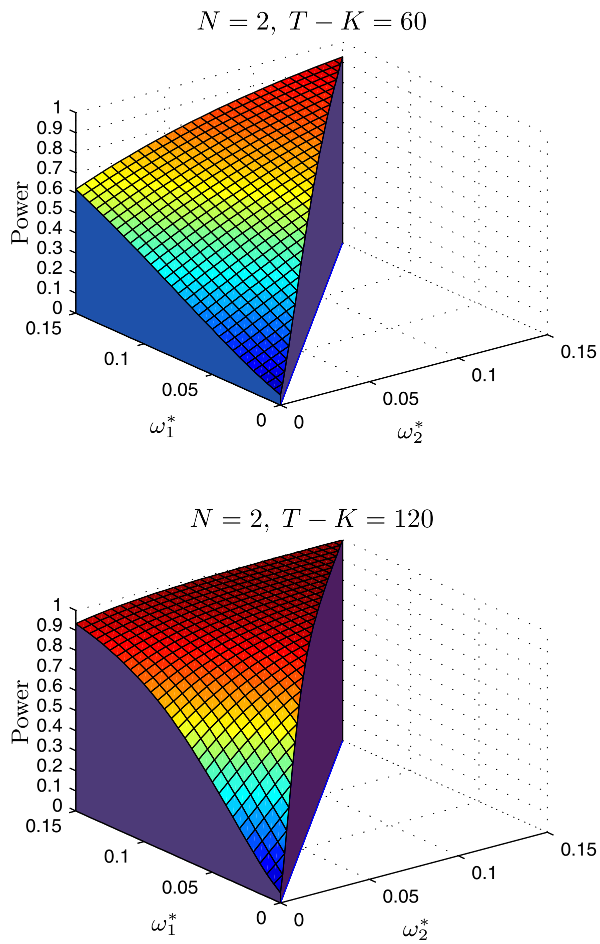

Figure 4
Power Function of Likelihood Ratio Test
The figure plots the probability of rejecting the null hypothesis of mean-variance spanning as a function of ω_1^* and ω_2^* using the likelihood ratio test when the size of the test is 5%, where $(T-K-1)\omega_1^*$ and $(T-K-1)\omega_2^*$ are the eigenvalues 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.

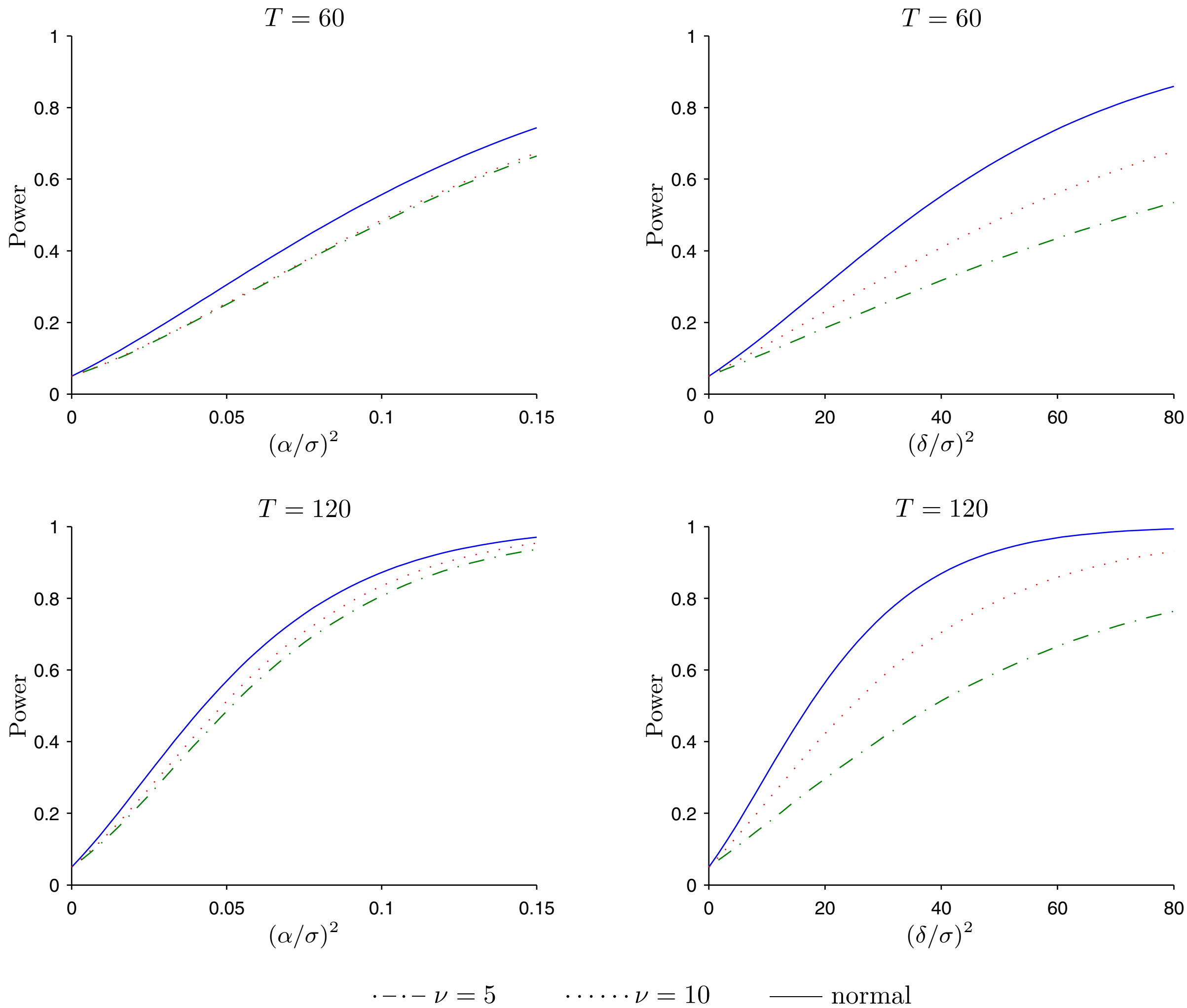

Figure 5
Power Function of GMM Wald Test Under Multivariate Student-t Returns
The figure plots the probability of rejecting the null hypothesis of mean-variance spanning for two different types of alternative hypotheses using the GMM Wald test. The plots on the left hand side are for alternative hypotheses 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.

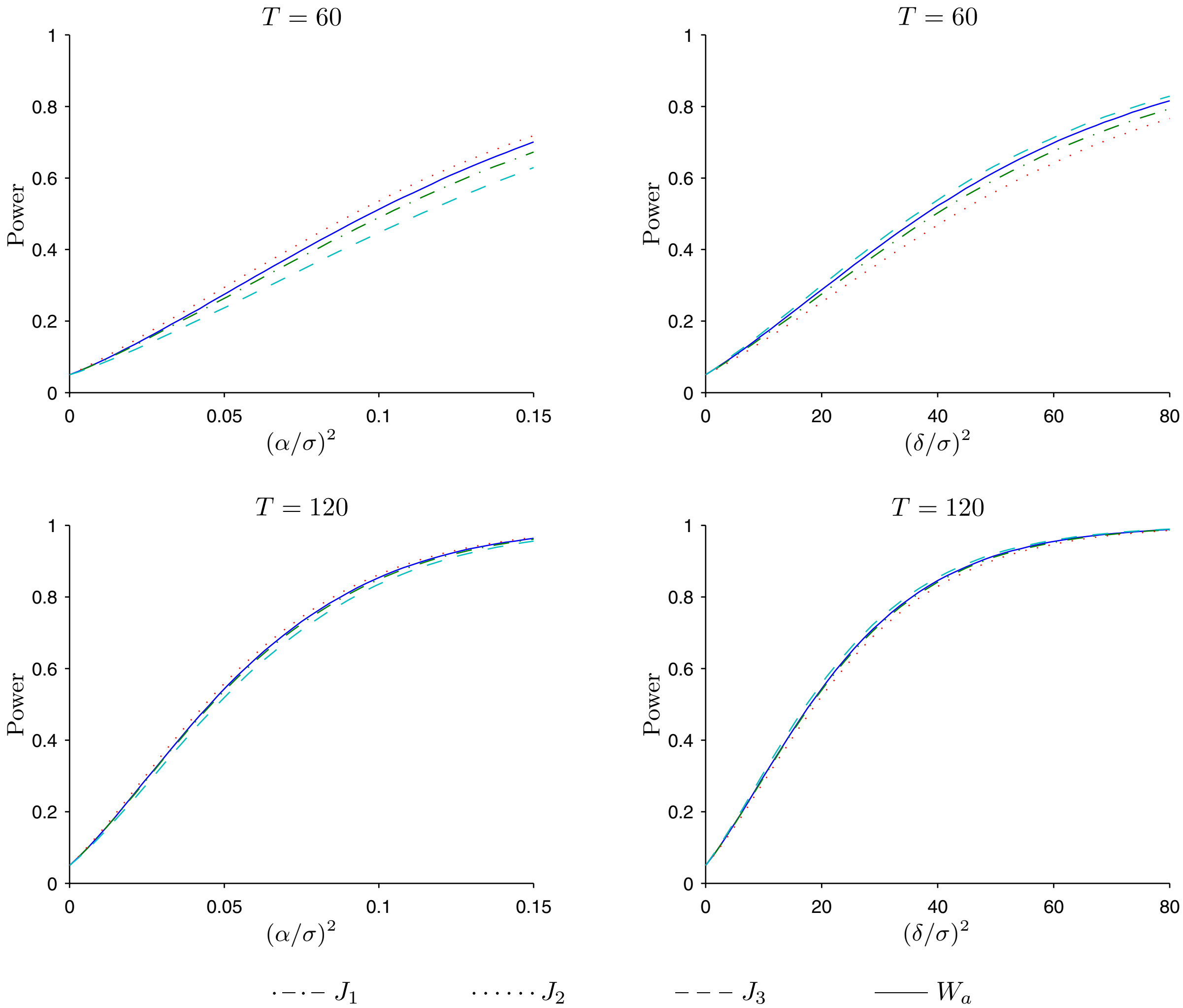

Figure 6
Power Function of GMM Wald Tests Under Multivariate Normal Returns
The figure plots the probability of rejecting the null hypothesis of mean-variance spanning for two different types of alternative hypotheses using four different GMM Wald tests (J_1 to J_3 are SDF based, W_a is regression based). 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 normal distribution and the probability of rejection in 100,000 simulations is plotted.

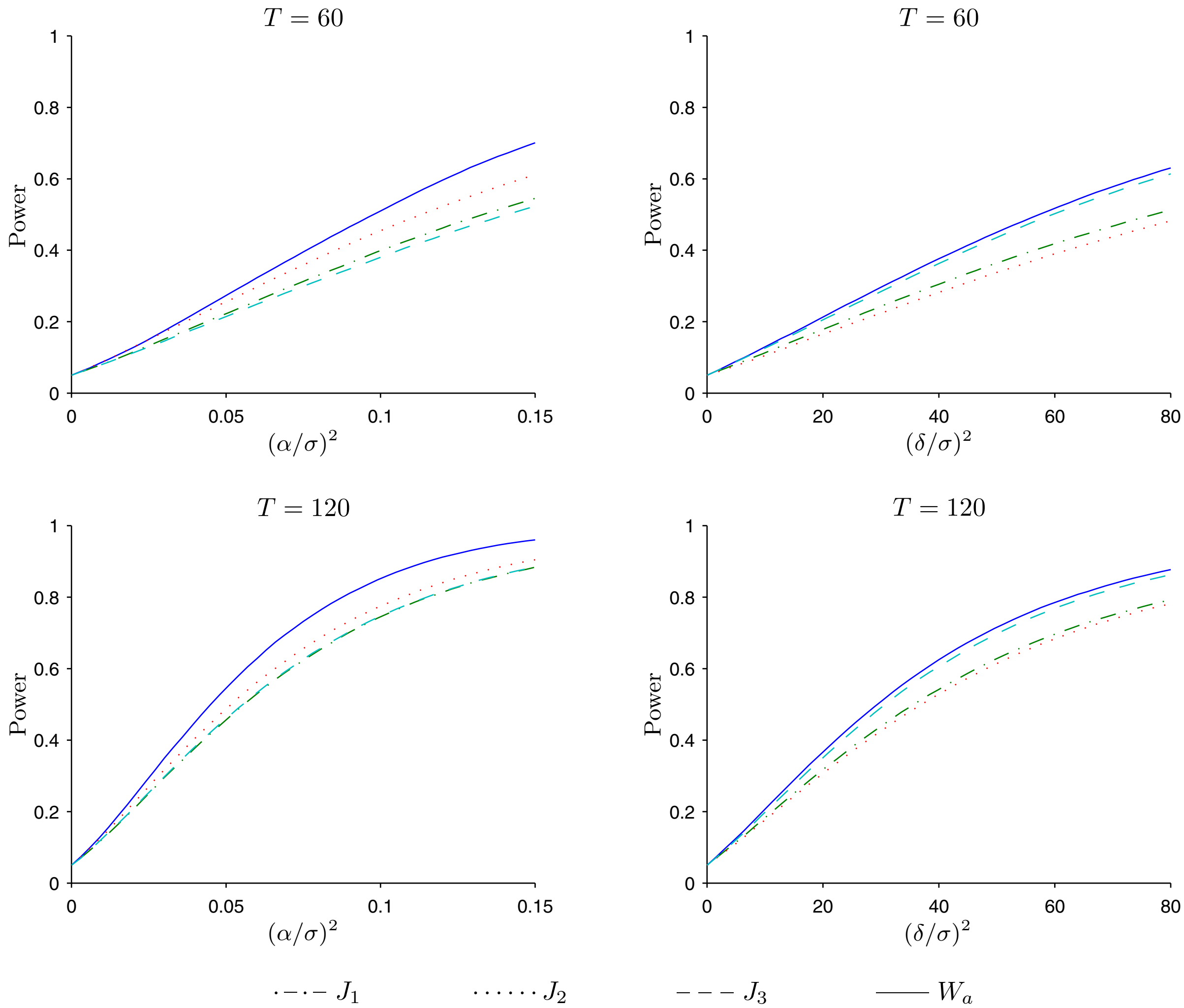

Figure 7
Power Function of GMM Wald Tests Under Multivariate Student-*t* Returns
The figure plots the probability of rejecting the null hypothesis of mean-variance spanning for two different types of alternative hypotheses using four different GMM Wald tests (J_1 to J_3 are SDF based, W_a is regression based). 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 degrees of freedom and the probability of rejection in 100,000 simulations is plotted.

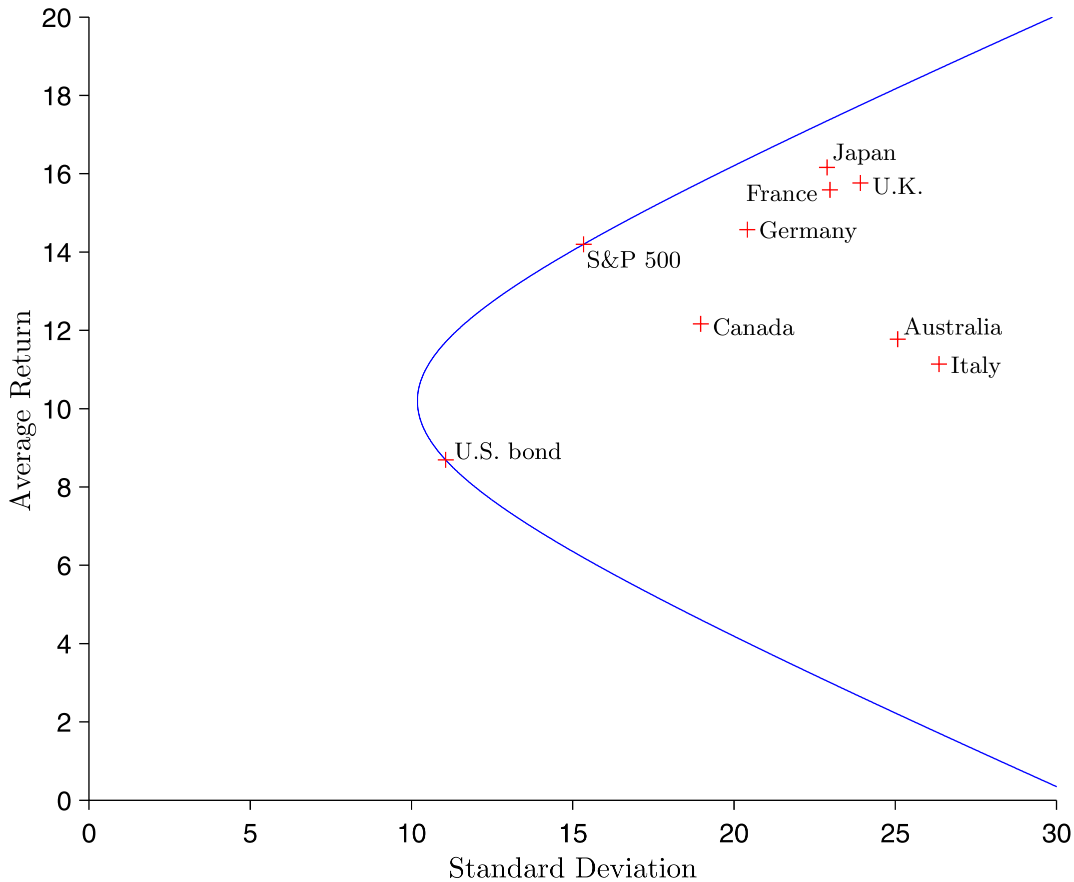

Figure 8
Average Return and Standard Deviation of U.S. and Foreign Investments
The figure plots the average return and sample standard deviation (in annualized percentage) of S&P 500 index, 30-year U.S. Treasury bond, and seven foreign equity indices, computed using monthly data over the period January 1970 to December 1999. The figure also plots the opportunity set from combining the S&P 500 index and the 30-year U.S. Treasury bond.

Table I
Sizes of Three Asymptotic Tests of Spanning Under Normality

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.

			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
	5	60	0.123	0.080	0.044
		120	0.081	0.063	0.047
		240	0.064	0.056	0.049
	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
	5	60	0.155	0.104	0.060
		120	0.092	0.073	0.055
		240	0.069	0.060	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
	5	60	0.222	0.159	0.100
		120	0.114	0.091	0.070
		240	0.077	0.068	0.059
	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

Table II

Comparison of Power of Three Tests of Spanning Under Normality

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 $TH\hat{G}^{-1}$. Numbers that are boldfaced indicate the test has the highest power among the three tests.

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**

Table III
Power of Step-Down Test of Spanning Under Normality

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 ($\Delta a = 0.0229$) is a test asset that doubles the slope of the asymptote to the efficient hyperbola of the two benchmark assets. The second alternative ($\Delta c = 76.69$) is a test asset that reduces the standard deviation of the global minimum-variance portfolio of the two benchmark assets from 4.9%/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.

Significance Levels		Probability of Rejection	
α_1	α_2	$\Delta a = 0.0229$ $\Delta b, \Delta c = 0$	$\Delta a, \Delta b = 0$ $\Delta c = 76.69$
0.00000	0.05000	0.05133	0.91981
0.02532	0.02532	0.16149	0.87008
0.04040	0.01000	0.19416	0.78207
0.04905	0.00100	0.20955	0.51506
0.04914	0.00090	0.20971	0.50289
0.04924	0.00080	0.20986	0.48942
0.04933	0.00070	0.21002	0.47432
0.04943	0.00060	0.21018	0.45715
0.04952	0.00050	0.21033	0.43722
0.04962	0.00040	0.21049	0.41348
0.04971	0.00030	0.21064	0.38398
0.04981	0.00020	0.21080	0.34476
0.04990	0.00010	0.21095	0.28458
0.04995	0.00005	0.21103	0.23337
0.04996	0.00004	0.21104	0.21878
0.04997	0.00003	0.21106	0.20128
0.04998	0.00002	0.21107	0.17904
0.04999	0.00001	0.21109	0.14711
0.05000	0.00000	0.21100	0.05000

Table IV

Sizes of Small Sample Tests of Spanning Under Nonnormality of Residuals

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.

			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
	5	60	0.047	0.048	0.048
		120	0.049	0.049	0.049
		240	0.050	0.050	0.050
	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
	5	60	0.047	0.047	0.048
		120	0.049	0.049	0.049
		240	0.050	0.050	0.050
	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
	5	60	0.048	0.048	0.048
		120	0.049	0.049	0.050
		240	0.050	0.051	0.050
	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

Table V
Sizes of Spanning Tests Under Multivariate Student-t Returns

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.

		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
5	60	0.332	0.409	0.121	0.942	1.450
	120	0.313	0.240	0.079	1.155	1.547
	240	0.318	0.150	0.064	1.338	1.635
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
5	60	0.194	0.319	0.127	0.872	1.144
	120	0.148	0.172	0.083	0.993	1.155
	240	0.130	0.107	0.066	1.066	1.162
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

Table VI

Sizes of GMM Wald Tests of Spanning Under Multivariate Normal Returns

The table presents the probabilities of rejection of three SDF based (J_1 to J_3) and a regression based GMM Wald (W_a) tests of spanning under the null hypothesis with the returns follow a multivariate normal distribution. The rejection decision is based on 95th percentile of their asymptotic distributions of χ^2_{2N}, and the results for different values of number of benchmark assets (K), test assets (N), and time series observations (T) are based on 100,000 simulations.

			Actual Probabilities of Rejection			
K	N	T	J_1	J_2	J_3	W_a
2	2	60	0.064	0.086	0.097	0.114
		120	0.054	0.065	0.071	0.078
		240	0.050	0.055	0.058	0.063
	5	60	0.120	0.189	0.218	0.260
		120	0.072	0.103	0.114	0.137
		240	0.056	0.070	0.075	0.087
	10	60	0.366	0.543	0.596	0.658
		120	0.135	0.224	0.254	0.308
		240	0.075	0.110	0.121	0.150
	25	60	1.000	1.000	1.000	1.000
		120	0.751	0.908	0.929	0.952
		240	0.252	0.439	0.482	0.573
5	2	60	0.075	0.099	0.110	0.132
		120	0.060	0.072	0.075	0.087
		240	0.053	0.059	0.061	0.067
	5	60	0.147	0.221	0.250	0.305
		120	0.083	0.116	0.125	0.153
		240	0.061	0.076	0.079	0.093
	10	60	0.428	0.598	0.648	0.722
		120	0.157	0.251	0.276	0.342
		240	0.085	0.122	0.131	0.164
	25	60	1.000	1.000	1.000	1.000
		120	0.785	0.924	0.941	0.964
		240	0.274	0.464	0.496	0.602
10	2	60	0.103	0.130	0.145	0.172
		120	0.069	0.082	0.087	0.099
		240	0.058	0.064	0.066	0.072
	5	60	0.206	0.285	0.325	0.395
		120	0.102	0.137	0.151	0.183
		240	0.070	0.085	0.089	0.104
	10	60	0.539	0.686	0.747	0.821
		120	0.196	0.298	0.327	0.402
		240	0.096	0.137	0.147	0.183
	25	60	1.000	1.000	1.000	1.000
		120	0.842	0.947	0.960	0.980
		240	0.314	0.505	0.537	0.643

Table VII

Mean-Variance Spanning Tests on Seven Foreign Equity Indices Under Normality

The table presents two sets of mean-variance spanning tests on seven foreign 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 foreign equity index as well as jointly on all seven foreign 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.

					Step-Down Test			
Country	$\hat{\alpha}$	$\hat{\delta}$	F-test	p-value	F_1	p-value	F_2	p-value
			Entire Period: 1970/1—1999/12					
Australia	0.0025	0.551	12.334	0.000	0.531	0.811	24.169	0.000
Canada	−0.0002	0.156	2.982	0.052	0.008	1.000	5.971	0.000
France	0.0051	0.331	5.064	0.007	2.385	0.021	7.713	0.000
Germany	0.0054	0.375	7.674	0.001	3.281	0.002	11.990	0.000
Italy	0.0042	0.567	9.154	0.000	1.051	0.395	17.255	0.000
Japan	0.0085	0.571	13.055	0.000	5.818	0.000	20.024	0.000
U.K.	0.0037	0.188	1.757	0.174	1.258	0.270	2.254	0.030
All			3.783	0.000	1.080	0.376	6.648	0.000
			First Subperiod: 1970/1—1984/12					
Australia	0.0007	0.361	2.538	0.082	0.019	1.000	5.086	0.000
Canada	0.0017	0.043	0.187	0.829	0.288	0.958	0.087	0.999
France	0.0033	0.513	4.983	0.008	0.443	0.874	9.554	0.000
Germany	0.0036	0.504	7.962	0.000	0.835	0.560	15.103	0.000
Italy	−0.0002	0.757	8.999	0.000	0.002	1.000	18.099	0.000
Japan	0.0116	0.563	10.199	0.000	7.504	0.000	12.440	0.000
U.K.	0.0037	0.137	0.414	0.662	0.431	0.882	0.397	0.903
All			2.993	0.000	1.235	0.286	4.900	0.000
			Second Subperiod: 1985/1—1999/12					
Australia	0.0052	0.760	12.263	0.000	1.186	0.313	23.316	0.000
Canada	−0.0018	0.268	6.963	0.001	0.492	0.840	13.472	0.000
France	0.0063	0.148	1.284	0.279	2.196	0.037	0.370	0.919
Germany	0.0063	0.238	1.655	0.194	1.876	0.076	1.427	0.197
Italy	0.0078	0.381	2.205	0.113	1.735	0.104	2.665	0.012
Japan	0.0051	0.563	4.594	0.011	0.776	0.608	8.423	0.000
U.K.	0.0041	0.245	2.387	0.095	1.351	0.229	3.417	0.002
All			2.560	0.002	0.790	0.597	4.481	0.000

Table VIII
Asymptotic Mean-Variance Spanning Tests on Seven Foreign Equity Indices

The table presents four mean-variance spanning tests on seven foreign equity indices, using the S&P 500 index and the 30-year U.S. Treasury bond as benchmark assets. The first 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. The other two tests, J_2 and J_3, are SDF based GMM Wald tests. J_2 is the version used by Bekaert and Urias (1996), and J_3 is the version used by DeSantis (1993). The four tests are performed on each foreign equity index as well as jointly on all seven foreign equity indices. All four tests 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 presented for the entire sample period as well as for its two subperiods.

	Regression Based				SDF Based			
Country	W_a^e	p-value	W_a	p-value	J_2	p-value	J_3	p-value
				Entire Period: 1970/1—1999/12				
Australia	11.172	0.004	20.742	0.000	17.566	0.000	20.872	0.000
Canada	3.925	0.140	4.428	0.109	4.344	0.114	4.613	0.100
France	7.303	0.026	9.836	0.007	9.490	0.009	11.051	0.004
Germany	13.773	0.001	14.583	0.001	13.551	0.001	14.983	0.001
Italy	15.459	0.000	13.687	0.001	11.568	0.003	15.104	0.001
Japan	23.294	0.000	23.077	0.000	20.285	0.000	25.939	0.000
U.K.	1.321	0.517	3.260	0.196	2.559	0.278	2.617	0.270
All	26.743	0.021	48.329	0.000	39.984	0.000	61.513	0.000
				First Subperiod: 1970/1—1984/12				
Australia	2.949	0.229	4.146	0.126	3.805	0.149	4.151	0.125
Canada	0.329	0.848	0.316	0.854	0.314	0.855	0.315	0.854
France	5.383	0.068	9.782	0.008	8.262	0.016	10.570	0.005
Germany	11.244	0.004	13.315	0.001	11.895	0.003	15.088	0.001
Italy	11.857	0.003	22.865	0.000	17.026	0.000	20.902	0.000
Japan	15.709	0.000	18.247	0.000	17.988	0.000	24.679	0.000
U.K.	0.402	0.818	0.974	0.615	0.880	0.644	0.890	0.641
All	26.227	0.024	45.627	0.000	35.093	0.001	57.247	0.000
				Second Subperiod: 1985/1—1999/12				
Australia	7.176	0.028	23.723	0.000	18.304	0.000	22.425	0.000
Canada	7.645	0.022	12.091	0.002	11.804	0.003	12.829	0.002
France	2.515	0.284	3.182	0.204	3.282	0.194	3.417	0.181
Germany	3.149	0.207	3.679	0.159	3.558	0.169	3.700	0.157
Italy	4.415	0.110	3.289	0.193	3.148	0.207	3.543	0.170
Japan	9.566	0.008	8.128	0.017	6.917	0.031	8.740	0.013
U.K.	3.413	0.181	5.031	0.081	4.684	0.096	4.914	0.086
All	21.359	0.093	36.764	0.001	28.701	0.011	38.369	0.000