

Tests of Mean-Variance Spanning

RAYMOND KAN and GUOFU ZHOU*

First draft: March, 1997

This version: January, 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, Indiana University, University of California at Irvine, York University, and participants at the 2000 Northern Finance Meetings and the 2001 American Finance Meetings 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), and De Roon, Nijman, and Werker (2001).

In this paper, we aim at improving our understanding of various tests of mean-variance spanning. HK provide a likelihood ratio test of spanning and derive its finite sample distribution under normality assumption. 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 the case of test of 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 the 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 set of $N + K$ 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. Under the nonnormality assumption, we conduct our analysis using simulation

and reach a similar conclusion.

Recently, Ferson (1992), DeSantis (1993), 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 of tests of spanning. Section II provides a comprehensive power analysis of various regression based spanning tests. Section III discusses how to generalize these spanning tests when the assets returns are not multivariate normally distributed. Section IV introduces the stochastic discount factor approach of tests of spanning and compares their performance with the regression based tests. 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

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

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

$$V = \text{Var}[R_t] \equiv \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \quad (2)$$

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

$$R_{2t} = \alpha + \beta R_{1t} + \epsilon_t, \quad (3)$$

with $E[\epsilon_t] = 0_N$ and $E[\epsilon_t R'_{1t}] = 0_{N \times K}$, where 0_N is an N -vector of zeros and $0_{N \times K}$ is an N by K matrix of zeros. It is easy to show that α and β are given by $\alpha = \mu_2 - \beta \mu_1$ and

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

$\beta = V_{21}V_{11}^{-1}$. 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

$$H_0 : \alpha = 0_N, \quad \delta = 0_N. \quad (4)$$

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 $\text{Var}[\epsilon_t]$ 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

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

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

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

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

and

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

²In 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 .

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

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

$$R_{2t} = \alpha + \beta R_{1t} + \epsilon_t, \quad t = 1, 2, \dots, T, \quad (9)$$

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:

$$R = XB + E, \quad (10)$$

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.

³Instead 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 $Qw_2 = 0_N$.

⁴Note 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.

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

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

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

Under the normality assumption, we have

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

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

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

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

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

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

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

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

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

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

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

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

$$\tilde{\Sigma} - \hat{\Sigma} = \hat{\Theta}' \hat{G}^{-1} \hat{\Theta} \quad (20)$$

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

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

where

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

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

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

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

$$W = T(\lambda_1 + \lambda_2) \stackrel{A}{\sim} \chi_{2N}^2. \quad (24)$$

and the Lagrange multiplier test is given by

$$LM = T \sum_{i=1}^2 \frac{\lambda_i}{1 + \lambda_i} \stackrel{A}{\sim} \chi_{2N}^2. \quad (25)$$

Note that although LR , W , and LM all have an asymptotic χ_{2N}^2 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.

⁵The 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 $T - K + 1$.

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 by⁸

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

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 $N \geq 2$.⁹ When $N = 1$, the correct F -test should be

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

For the special case of $N = 1$, 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

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

⁷The 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.

⁸HK'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 later studies such as Bekaert and Urias (1996) and Errunza, Hogan, and Hung (1999), but the F -test given by Jobson and Korkie (1989) is correct.

⁹Although HK assume $K \geq 2$ in their derivation, the test is nevertheless valid statistically for the case of $K = 1$. When $K = 1$, it is a joint test of whether all the test assets have the same expected return as the benchmark asset and whether the benchmark asset is the global minimum-variance portfolio of all the assets.

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)

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

where $B(\cdot, \cdot)$ 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.¹⁰ Therefore, we provide our own expression here. Let $\xi_1 = \lambda_1/(1 + \lambda_1)$ and $\xi_2 = \lambda_2/(1 + \lambda_2)$, we have,

$$P[\xi_1 + \xi_2 \leq v] = I_{\frac{v}{2}}(N-1, T-K-N+1) - \frac{\int_{\max[0, v-1]}^{\frac{v^2}{4}} u^{\frac{N-3}{2}} (1-v+u)^{\frac{T-K-N}{2}} du}{2B(N-1, T-K-N+1)}. \quad (29)$$

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 distribution of χ_{2N}^2 . 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 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

¹⁰Existing 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.

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 $\hat{a} = \hat{\mu}'\hat{V}^{-1}\hat{\mu}$, $\hat{b} = \hat{\mu}'\hat{V}^{-1}\mathbf{1}_{N+K}$, $\hat{c} = \mathbf{1}'_{N+K}\hat{V}^{-1}\mathbf{1}_{N+K}$, 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 $\hat{a}_1 = \hat{\mu}'_1\hat{V}_{11}^{-1}\hat{\mu}_1$, $\hat{b}_1 = \hat{\mu}'_1\hat{V}_{11}^{-1}\mathbf{1}_K$, $\hat{c}_1 = \mathbf{1}'_K\hat{V}_{11}^{-1}\mathbf{1}_K$. Using these constants, we can write

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

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*

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

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

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

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

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

In Figure 1, we plot the *ex post* minimum-variance frontiers of the K benchmark assets as well as the one of 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 $\sqrt{1 + \frac{\hat{d}}{\hat{c}}}/\sqrt{1 + \frac{\hat{d}_1}{\hat{c}_1}}$, 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 $\sqrt{1 + \frac{\hat{d}}{\hat{c}}}/\sqrt{1 + \frac{\hat{d}_1}{\hat{c}_1}}$ 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 as¹¹

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

Figure 1 about here

¹¹For $N = 1$, the F -test of (27) can be geometrically represented as

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

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

For the Wald and the Lagrange multiplier tests, mean-variance spanning is tested by examining different parts of the two minimum-variance frontiers. To obtain a geometrical interpretation of these two test statistics, we define $\hat{\theta}_1(r)$ and $\hat{\theta}(r)$ as the slope of the tangent lines to the sample frontier of the K benchmark assets and of all the $N + K$ 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 as¹²

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

and

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

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

¹²Note 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.

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.

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's test of spanning. Since the likelihood ratio test is not in general the uniformly most powerful invariant test, it is important for us also to understand the power of other tests and compare their power.

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 tests. 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 distribution of χ_{2N}^2 . 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 the asymptotic distribution of χ_{2N}^2 . It also deserves emphasis that the distributions of the three test statistics 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

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

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 as¹³

$$\omega = \left[\frac{c - c_1}{\hat{c}_1} + \frac{\theta^2(\hat{\mu}_{g_1}) - \theta_1^2(\hat{\mu}_{g_1})}{1 + \hat{\theta}_1^2(\hat{\mu}_{g_1})} \right], \quad (39)$$

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 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^* . Figure 2 allows us to determine what level of ω^* 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

¹³The derivation of this expression is similar to that of (36) and therefore not provided.

variances. The second term measures how close the *ex ante* tangency portfolios (when the tangent lines have a y -intercept of $\hat{\mu}_{g_1}$) of the two frontiers are in terms of the square of the slope of their tangent lines.

In determining the power of the test, the distance between the two global minimum-variance portfolios is in practice a lot more important than the distance between the two tangency portfolios. We provide an example to illustrate this. Consider the case of two benchmark assets (i.e., $K = 2$), chosen as the equally weighted and value-weighted market portfolio of the NYSE.¹⁴ Using monthly returns from 1926/1–1998/12, 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). From Figure 2, we can see that this allows us to reject the null hypothesis with a probability of 0.86 for $T - K = 60$, and the probability of rejection goes up to almost one for $T - K = 120$ and 240. In the second case, we consider a test asset that does not reduce the variance of the global minimum-variance portfolio but doubles the slope of the asymptote of the frontier from 0.0875 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, we can see that the probability of rejecting the null hypothesis is only 0.16 for $T - K = 60$, 0.29 for $T - K = 120$, and 0.54 for $T - K = 240$.

It is easy to explain why the F -test has strong power to reject the spanning hypothesis

¹⁴This example was also used by Kandel and Stambaugh (1989).

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.¹⁵ The first term of ω involves $c - c_1 = 1'_{N+K}V^{-1}1_{N+K} - 1'_KV_{11}^{-1}1_K$ which is determined by V but not μ . Since estimates of V are in general a lot more accurate than estimates of μ , 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 of the power of the spanning tests is an extremely difficult task for $N > 1$. 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 of the three tests in small samples, and finally examples illustrating the power of the three tests 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

¹⁵Due to the imprecision of using $\hat{\mu}$ as an estimate of expected return, Jorion (1985) advocates the use of a shrinkage estimator that relies heavily on the sample mean of the *ex post* global minimum-variance portfolio.

written as

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

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

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

To understand why λ_1 and λ_2 are essential in testing the null hypothesis, note that the null hypothesis $H_0 : \Theta = \text{O}_{2 \times N}$ 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 $\Theta = \text{O}_{2 \times N}$ 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 sample 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 2 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 powers of the three tests. It can be shown that the power of the three tests are monotonically increasing function in $T\omega_1$ and $T\omega_2$.¹⁶ This means that for fixed ω_1 and ω_2 , we have the powers of the three tests as increasing functions of T , which is expected. The more interesting question is how the powers of these tests are 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

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

where $a = \mu'V^{-1}\mu$, $b = \mu'V^{-1}\mathbf{1}_{N+K}$, $c = \mathbf{1}'_{N+K}V^{-1}\mathbf{1}_{N+K}$, $a_1 = \mu'_1V_{11}^{-1}\mu_1$, $b_1 = \mu'_1V_{11}^{-1}\mathbf{1}_K$, and $c_1 = \mathbf{1}'_KV_{11}^{-1}\mathbf{1}_K$ 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 powers of the three tests also depend 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

¹⁶It 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.

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 we introduce additional test assets that do indeed 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 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 $\omega_2 \approx 0$, we have the Wald test as the most powerful among the three tests. However, when $\omega_1 \approx \omega_2$, the Lagrange multiplier test is more powerful than the other two tests. There are only a few cases when the likelihood ratio test is the most powerful test. 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 to help us to identify alternative hypotheses that give $\omega_2 \approx 0$ and $\omega_1 \approx \omega_2$.

Lemma 2 *Define*

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

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

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

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 we discuss 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, the distance between the standard deviations of the two global minimum-variance portfolios is the primary determinant of the powers of the three spanning tests. 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. Obviously, a small difference in the global minimum-variance portfolios, while statistically significant, is not economically as important as a big difference in the tangency portfolios, even though

the latter may be more difficult to show up 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.¹⁷ A low p -value does not always imply there is an economically significant difference between the two frontiers, and a high p -value does not always imply 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.¹⁸ This procedure is seldom used in the finance literature but we believe this alternative procedure is potentially more flexible and provides more information than the three tests that we discuss earlier. The step-down procedure is a sequential test. We first test $\alpha = 0_N$ using an F -test. This test is similar to the GRS F -test. Denote

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

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. The second test is to test $\delta = 0_N$ but conditional on the constraint $\alpha = 0_N$. This test is also an F -test, and it is given by

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

¹⁷For example, Errunza, Hogan, and Hung (1999) use the p -value from the HK test to measure the degree to which one can reject mean-variance spanning.

¹⁸See Section 8.4.5 of Anderson (1984) for a discussion of the step-down procedure. It should be noted that the step-down procedure there applies to each of the test assets but not to each component of the hypothesis as in our case.

where $\tilde{\Sigma}$ is the constrained estimate of Σ by imposing both the constraints of $\alpha = 0_N$ and $\delta = 0_N$. In the appendix, we show that under the null hypothesis, F_2 has a central F -distribution with N and $T - K - N + 1$ degrees of freedom, and it is independent of F_1 . Suppose the level of significance of the first test is α_1 and that of the second test is α_2 . Under the step-down procedure, we will accept the spanning hypothesis if we accept both tests. Therefore, the significance level of this step-down test is $1 - (1 - \alpha_1)(1 - \alpha_2) = \alpha_1 + \alpha_2 - \alpha_1\alpha_2$.¹⁹ 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 more significant 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 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

¹⁹Alternatively, 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.

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%.²⁰ 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 distributed as χ_{2N}^2 under the null hypothesis. However, if ϵ_t is not normally

²⁰Under 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.

distributed, their finite sample distributions will not be the same as the ones presented in Section I. Nevertheless, the small sample distribution under the normality assumption can still provide a very good approximation for the small sample distribution under the nonnormality case. To illustrate this, we simulate the returns on the test assets under the null hypothesis but with ϵ_t independently drawn from a multivariate Student- t distribution with five degrees of freedom.²¹ 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.

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.²² 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).

²¹Due 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.

²²It 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.

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

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

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

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

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

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

where

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

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 the constrained estimate of B is involved in calculating 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 or 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 special case that the returns have a multivariate elliptical distribution. Except for the special case of multivariate normal distribution, the conditional variance of ϵ_t is no longer a constant, but a function of R_{1t} . The use of the multivariate elliptical distribution to model returns can be motivated both empirically and theoretically. Empirically, Blatteberg and Gonedes (1974) find that stock returns have excess kurtosis compared with the normal distribution, and such a feature can often be captured by some members in the elliptical distribution like the multivariate Student- t distribution. Theoretically, multivariate elliptical distribution 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.²³

Proposition 2: *Suppose R_t is independently and identically distributed as a multivariate elliptical distribution with finite fourth moments. Define its kurtosis parameter as*

$$\kappa = \frac{E[(R_{it} - E[R_{it}])^4]}{3\text{Var}[R_{it}]^2} - 1, \quad (51)$$

where R_{it} is an element of R_t . The GMM Wald test of spanning is given by

$$W_a = T\text{tr}(\hat{H}\hat{G}_a^{-1}) \stackrel{A}{\sim} \chi_{2N}^2, \quad (52)$$

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

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

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

²³We thank Chris Geczy for suggesting an improvement to this proposition. See Geczy (2000) for a similar conditional heteroskedasticity adjustment for the tests of mean-variance efficiency.

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 .²⁴ Since $\hat{G}_a - \hat{G}$ does not go to zero asymptotically when $\kappa > 0$, using W will lead to over-rejection problem when returns follow a multivariate elliptical distribution with excess kurtosis. In the following, we study a popular member of the multivariate elliptical distribution – the multivariate Student- t distribution.²⁵ To assess the impact of the multivariate Student- t distribution on tests of spanning, we perform a simulation experiment using the same two benchmark assets given in Figure 3. For different choices of N , we simulate returns of the benchmark assets and the test assets jointly from a multivariate Student- t distribution with mean and variance satisfying the null hypothesis. In Table V, we present the actual size of W and W_a 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. In addition, we also present the actual size of the small sample version of both Wald tests using the distribution in (28), which is exact under the normality assumption.

Table V about here

As we can see from Table V, the regular Wald tests reject far too often, regardless of using the asymptotic distribution χ_{2N}^2 or the small sample distribution under normality assumption. 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 based on the χ_{2N}^2 distribution, it works reasonably well except when N is large, and its probability of rejection gets closer to the size of the test as T increases. The GMM Wald test based on the small sample distribution under the normality assumption works well when $\nu = 10$. When $\nu = 5$, the small sample GMM Wald tests under-reject the null hypothesis, especially when N is large. This under-rejection problem,

²⁴It 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 it can be much bigger than \hat{G} when the return distribution has fat tails.

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

however, disappears as T increases.

Table V also reports the average ratio of W to W_a . It shows that for $\nu = 5$, the regular Wald test is on average about twice as big as the GMM Wald test, but for $\nu = 10$, the ratio goes down to roughly 1.167. To understand how this average ratio is determined, we show in the Appendix that the expected bias of the regular Wald test under the multivariate Student- t distribution is

$$E \left[\frac{W}{W_a} \right] - 1 \approx \frac{\kappa}{2} = \frac{1}{\nu - 4} \quad (54)$$

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 very small compared with one (which is usually the case for monthly data). Therefore, when $\nu = 5$, the expected bias is about 100%, and when $\nu = 10$, 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. In MacKinlay and Richardson (1991), they find that for $\nu = 5$, the bias of the regular Wald test of mean-variance efficiency of a given portfolio is less than 35% even when the squared Sharpe ratio of the benchmark portfolio is very large, and the bias 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 $(1 + a_1)\Sigma$ and $\left[1 + \left(\frac{\nu-2}{\nu-4}\right) a_1\right]\Sigma$, respectively.²⁶ When the squared Sharpe ratio of the benchmark portfolio, a_1 , is small compared with one, $1 + a_1$ is very close to $1 + \left(\frac{\nu-2}{\nu-4}\right) a_1$, and hence the impact of the conditional heteroskedasticity adjustment on test of $\alpha = 0_N$ is minimal. For the case of test of spanning, it is a joint test of $\alpha = 0_N$ and $\delta = 0_N$. The asymptotic variance of $\hat{\delta}$ with and without the conditional heteroskedasticity adjustment are $c_1\Sigma$ and $\left(\frac{\nu-2}{\nu-4}\right) c_1\Sigma$, respectively, and the ratio of the two is always equal to $(\nu - 2)/(\nu - 4)$. 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

²⁶The asymptotic variance of $\hat{\alpha}$ is given in (A28) in the Appendix. For the special case of $K = 1$, this expression is given in MacKinlay and Richardson (1991).

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.

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 case of multivariate normality. 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 since the asymptotic variance of $\hat{\delta}$ under multivariate Student- t returns will be much higher than in the case of multivariate normal returns, especially when the degrees of freedom is small.

Figure 5 about here

In Figure 5, we plot the power function of W_a 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 of W_a under multivariate Student- t returns, the power functions are obtained by simulation. 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. In fact, the power of W_a improves slightly when returns follow a multivariate Student- t distribution with $\nu = 5$. This is because in finite samples, $\hat{\alpha}$ is not normally distributed when we have

multivariate Student- t returns, so it is possible that the power of W_a can be slightly improved for the case of multivariate Student- t returns, especially when the asymptotic variance of \hat{a} does not change by much 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 when returns follow a multivariate Student- t distribution, as compared with the case of multivariate normal.

IV. SDF Based Tests of Spanning

A. Equivalent Hypotheses of Spanning

Ferson (1992), DeSantis (1993), 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. Starting with the BU approach, they project a stochastic discount factor m_t with mean c on the returns of $N + K$ assets as

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

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

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

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

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

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

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

where $Q = [O_{N \times K}, I_N]$. 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.²⁷ To understand what $\beta(c_1)$ and $\beta(c_2)$ represent, we note

²⁷See also Ferson (1992) for yet another alternative proof and discussion.

that

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

where w_1 and w_2 are the weights of two frontier portfolios defined in (5) and (6). 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 $\delta = 0_N$ as we show earlier.

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

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

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

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

where

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

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

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

To see that H_2 is equivalent to H_0 , we write U^{-1} as

$$\begin{aligned} U^{-1} &= V^{-1} - \frac{V^{-1}(\mathbf{1}_{N+K} + \mu)(\mathbf{1}_{N+K} + \mu)'V^{-1}}{1 + (\mathbf{1}_{N+K} + \mu)'V^{-1}(\mathbf{1}_{N+K} + \mu)} \\ &= V^{-1} - (d_1w_1 + d_2w_2)(\mathbf{1}_{N+K} + \mu)'V^{-1}, \end{aligned} \quad (64)$$

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

$$d_1 = \frac{1'_{N+K} V^{-1} \mu}{1 + (1_{N+K} + \mu)' V^{-1} (1_{N+K} + \mu)}, \quad (65)$$

$$d_2 = \frac{1'_{N+K} V^{-1} 1_{N+K}}{1 + (1_{N+K} + \mu)' V^{-1} (1_{N+K} + \mu)}. \quad (66)$$

Therefore, we have

$$\begin{aligned} \gamma(c) &= U^{-1}[(1-c)1_{N+K} - c\mu] \\ &= V^{-1}[(1-c)1_{N+K} - c\mu] - (d_1 w_1 + d_2 w_2)(1_{N+K} + \mu)' V^{-1}[(1-c)1_{N+K} - c\mu] \\ &= (1-c)(1'_{N+K} V^{-1} 1_{N+K}) w_2 - c(1'_{N+K} V^{-1} \mu) w_1 \\ &\quad - (1_{N+K} + \mu)' V^{-1} [(1-c)1_{N+K} - c\mu] (d_1 w_1 + d_2 w_2). \end{aligned} \quad (67)$$

Like $\beta(c_1)$ and $\beta(c_2)$, $\gamma(c_1)$ and $\gamma(c_2)$ are just 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 differences between these two approaches. As a statistical model, the regression approach assumes α and β are constant parameters. This assumption can be satisfied without μ and V being constant over time. Therefore, under the regression approach, we do not need to assume the weights of the frontier portfolios are constant over time. However, the statistical models used by the SDF approach assume that 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 constant over time and hence the weights of the frontier portfolios must also be constant over time. While the regression approach still does not allow for arbitrary changes in μ and V over time, it is more flexible than the SDF approach.

With these remarks in mind, we now turn to the GMM estimation and tests by using

the SDF approach. The sample moment conditions used by BU are

$$\bar{h}_T(\beta(c_1), \beta(c_2)) = \begin{bmatrix} \frac{1}{T} \sum_{t=1}^T R_t(c_1 + (R_t - \hat{\mu})' \beta(c_1)) - (1 - c_1)1_{N+K} \\ \frac{1}{T} \sum_{t=1}^T R_t(c_2 + (R_t - \hat{\mu})' \beta(c_2)) - (1 - c_2)1_{N+K} \end{bmatrix}. \quad (68)$$

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

$$\bar{h}_T^*(\beta(c_1), \beta(c_2), \mu) = \begin{bmatrix} \frac{1}{T} \sum_{t=1}^T R_t(c_1 + (R_t - \mu)' \beta(c_1)) - (1 - c_1)1_{N+K} \\ \frac{1}{T} \sum_{t=1}^T R_t(c_2 + (R_t - \mu)' \beta(c_2)) - (1 - c_2)1_{N+K} \\ \frac{1}{T} \sum_{t=1}^T (R_t - \mu) \end{bmatrix}. \quad (69)$$

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

$$\hat{\beta}(c_1) = \hat{V}^{-1}[(1 - c_1)1_{N+K} - c_1 \hat{\mu}], \quad (70)$$

$$\hat{\beta}(c_2) = \hat{V}^{-1}[(1 - c_2)1_{N+K} - c_2 \hat{\mu}]. \quad (71)$$

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 of $\hat{b} = (\hat{\beta}(c_1)', \hat{\beta}(c_2)')$ 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

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad (72)$$

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. The following proposition summarizes the results.

Proposition 3: *The GMM estimator of $(\beta(c_1)', \beta(c_2)')$ in (69) has asymptotic variance*

$$\text{Avar}(\hat{b}) = C^{-1}(S_{11} - F'S_{21} - S_{12}F + F'S_{22}F)C^{-1}, \quad (73)$$

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

$$J_1 = T(\hat{b}'Q_2'[Q_2\hat{C}^{-1}(\hat{S}_{11} - \hat{F}'\hat{S}_{21} - \hat{S}_{12}\hat{F} + \hat{F}'\hat{S}_{22}\hat{F})\hat{C}^{-1}Q_2']^{-1}Q_2\hat{b}) \stackrel{A}{\sim} \chi_{2N}^2, \quad (74)$$

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

$$J_2 = T(\hat{b}'Q_2'(Q_2\hat{C}^{-1}\hat{S}_{11}\hat{C}^{-1}Q_2')^{-1}Q_2\hat{b}) \stackrel{A}{\sim} \chi_{2N}^2, \quad (75)$$

the GMM Wald test obtained by BU from (68) 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

$$\bar{m}_T(\gamma(c_1), \gamma(c_2)) = \begin{bmatrix} \frac{1}{T} \sum_{t=1}^T (1 + R_t)(c_1 + (1 + R_t)'\gamma(c_1)) - 1_{N+K} \\ \frac{1}{T} \sum_{t=1}^T (1 + R_t)(c_2 + (1 + R_t)'\gamma(c_2)) - 1_{N+K} \end{bmatrix}. \quad (76)$$

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

$$\hat{\gamma}(c_1) = \hat{U}^{-1}[1_{N+K} - c_1(1_{N+K} + \hat{\mu})], \quad (77)$$

$$\hat{\gamma}(c_2) = \hat{U}^{-1}[1_{N+K} - c_2(1_{N+K} + \hat{\mu})], \quad (78)$$

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 not 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

$$J_3 = T(\hat{g}'Q_2'[(I_2 \otimes Q\hat{U}^{-1})\hat{S}_m(I_2 \otimes \hat{U}^{-1}Q)']^{-1}Q_2\hat{g}) \stackrel{A}{\sim} \chi_{2N}^2, \quad (79)$$

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 $c_1 = -(\hat{\mu}'\hat{V}^{-1}\mathbf{1}_{N+K} + \mathbf{1}'_{N+K}\hat{V}^{-1}\mathbf{1}_{N+K})$ and $c_2 = 1 + \hat{\mu}'\hat{V}^{-1}\hat{\mu} + \hat{\mu}'\hat{V}^{-1}\mathbf{1}_{N+K}$. With this choice of c_1 and c_2 , we can verify that $\hat{\gamma}(c_1) = \hat{V}^{-1}\mathbf{1}_{N+K}$ 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 .²⁸ 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

²⁸When 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.

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 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 we cannot make the general statement that the regression based GMM Wald test is better than the SDF based GMM Wald tests for all cases, our simulation evidence appears to prefer the spanning test under the regression approach, especially when returns have a multivariate Student- t distribution.

Figure 7 about here

V. 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.

Appendix

Proof of (29): From Anderson (1984, p.529) and using the duplication formula

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

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

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

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

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

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

$$\begin{aligned} & P[a_1 \leq v] \\ &= \frac{n+1}{2\text{B}(2m+2, 2n+3)} \int_0^{\frac{v^2}{4}} \int_{2\sqrt{a_2}}^{\min[v, 1+a_2]} a_2^m (1 - a_1 + a_2)^n da_1 da_2 \\ &= \frac{n+1}{2\text{B}(2m+2, 2n+3)} \int_0^{\frac{v^2}{4}} a_2^m \left[\frac{(1 - a_1 + a_2)^{n+1}}{n+1} \right] \Big|_{\min[v, 1+a_2]}^{2\sqrt{a_2}} da_2 \\ &= \frac{1}{2\text{B}(2m+2, 2n+3)} \left[\int_0^{\frac{v^2}{4}} a_2^m (1 - \sqrt{a_2})^{2n+2} da_2 - \int_{\max[0, v-1]}^{\frac{v^2}{4}} a_2^m (1 - v + a_2)^{n+1} da_2 \right] \\ &= I_{\frac{v}{2}}(2m+2, 2n+3) - \frac{1}{2\text{B}(2m+2, 2n+3)} \int_{\max[0, v-1]}^{\frac{v^2}{4}} a_2^m (1 - v + a_2)^{n+1} da_2. \quad (\text{A4}) \end{aligned}$$

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

$$\hat{V}^{-1} = \begin{bmatrix} \hat{V}_{11}^{-1} + \hat{\beta}' \hat{\Sigma}^{-1} \hat{\beta} & -\hat{\beta}' \hat{\Sigma}^{-1} \\ -\hat{\Sigma}^{-1} \hat{\beta} & \hat{\Sigma}^{-1} \end{bmatrix} = \begin{bmatrix} \hat{V}_{11}^{-1} & \mathbf{O}_{K \times N} \\ \mathbf{O}_{N \times K} & \mathbf{O}_{N \times N} \end{bmatrix} + \begin{bmatrix} -\hat{\beta}' \\ \mathbf{I}_N \end{bmatrix} \hat{\Sigma}^{-1} \begin{bmatrix} -\hat{\beta} & \mathbf{I}_N \end{bmatrix}. \quad (\text{A5})$$

Therefore,

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

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

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

It is easy to see that

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

where the last equality follows from the fact that

$$\hat{\theta}^2(r) - \hat{\theta}_1^2(r) = (\hat{a} - 2\hat{b}r + \hat{c}r^2) - (\hat{a}_1 - 2\hat{b}_1r + \hat{c}_1r^2) = \Delta \hat{a} - 2\Delta \hat{b}r + \Delta \hat{c}r^2. \tag{A11}$$

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

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

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 $Y_1^*(T\hat{\Sigma})^{-1}Y_1^{*'} = \hat{G}^{-\frac{1}{2}}\hat{\Theta}\hat{\Sigma}^{-1}\hat{\Theta}'\hat{G}^{-\frac{1}{2}}$ 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

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

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

$$\Omega = T\hat{G}^{-\frac{1}{2}}\hat{\Theta}\hat{\Sigma}^{-1}\hat{\Theta}'\hat{G}^{-\frac{1}{2}}. \quad (\text{A14})$$

It is well known that the hypergeometric function only depends on the eigenvalues of Ω , which is the same as the eigenvalues of $T\hat{H}\hat{G}^{-1}$. Therefore, the joint density function of λ_1 and λ_2 depends only on the eigenvalues of $T\hat{H}\hat{G}^{-1}$ 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 $L_{11} \sim \sqrt{\chi_{T-K-N+1}^2}$, $L_{22} \sim \sqrt{\chi_{T-K-N}^2}$, and $L_{12} \sim N(0, 1)$. 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 $\text{vec}(Z) \sim N(\text{vec}(D^{\frac{1}{2}}), I_4)$, then we have $Z'Z \sim W_2(2, I_2, D)$ and $A = S + Z'Z \sim W_2(N, I_2, D)$.

This completes the proof. *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)

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

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

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

where the last equality follows from the fact that

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

(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 $\theta^2(\mu_z) - \theta_1^2(\mu_z) = 0$. (ii) Using the inequality $(a + b)^2 \geq 4ab$ for a and b nonnegative and the definition of μ_z , we have

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

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

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

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),

$$\frac{|\bar{\Sigma}|}{|\tilde{\Sigma}|} \sim U_{N,1,T-K} \quad (\text{A20})$$

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

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

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 $\text{vec}(\hat{B}')$ is given by $(D'_0 S_0^{-1} D_0)^{-1}$, where

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

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

$$\begin{aligned} (A \otimes I_N)(D'_0 S_0^{-1} D_0)^{-1}(A' \otimes I_N) &= (A \otimes I_N)D_0^{-1}S_0D_0^{-1}(A' \otimes I_N) \\ &= (A(E[x_t x_t']^{-1} \otimes I_N)S_0((E[x_t x_t'])^{-1}A' \otimes I_N)) \end{aligned} \quad (\text{A23})$$

Using the partitioned matrix inverse formula, we have

$$\begin{aligned} A(E[x_t x_t']^{-1})^{-1} &= \begin{bmatrix} 1 & 0'_K \\ 0 & -1'_K \end{bmatrix} \begin{bmatrix} 1 & \mu'_1 \\ \mu_1 & V_{11} + \mu_1 \mu'_1 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 0'_K \\ 0 & -1'_K \end{bmatrix} \begin{bmatrix} 1 + \mu'_1 V_{11}^{-1} \mu_1 & -\mu'_1 V_{11}^{-1} \\ -V_{11}^{-1} \mu_1 & V_{11}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} 1 + \mu'_1 V_{11}^{-1} \mu_1 & -\mu_1 V_{11}^{-1} \\ 1'_K V_{11}^{-1} \mu_1 & -1'_K V_{11}^{-1} \end{bmatrix}. \end{aligned} \quad (\text{A24})$$

Replacing S_0 and $A(E[x_t x_t']^{-1})^{-1}$ 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

$$E[R_{1t} R'_{1t} \otimes \epsilon_t \epsilon'_t] = \mu_1 \mu'_1 \otimes \Sigma + (1 + \kappa) V_{11} \otimes \Sigma = (V_{11} + \mu_1 \mu'_1) \otimes \Sigma + \kappa V_{11} \otimes \Sigma, \quad (\text{A25})$$

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

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

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

$$(D'_0)^{-1}S_0D_0^{-1} = \begin{bmatrix} 1 + a_1 & -\mu'_1V_{11}^{-1} \\ -V_{11}^{-1}\mu_1 & V_{11}^{-1} \end{bmatrix} \otimes \Sigma + \kappa \begin{bmatrix} a_1 & -\mu'_1V_{11}^{-1} \\ -V_{11}^{-1}\mu_1 & V_{11}^{-1} \end{bmatrix} \otimes \Sigma. \quad (\text{A27})$$

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

$$(A \otimes I_N)(D_0^{-1})'S_0D_0^{-1}(A' \otimes I_N) = \begin{bmatrix} 1 + (1 + \kappa)a_1 & (1 + \kappa)b_1 \\ (1 + \kappa)b_1 & (1 + \kappa)c_1 \end{bmatrix} \otimes \Sigma. \quad (\text{A28})$$

By replacing $a_1, b_1, c_1, \kappa, \Sigma$ 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 $\text{vec}(\hat{\Theta}')$ is $\hat{G}_a \otimes \hat{\Sigma}$. Therefore, the GMM Wald test is

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

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

Proof of (54): From (A10), we have

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

Using a similar proof, we have

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

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

$$\text{tr}(\hat{H}\hat{G}^{-1}) \approx X_1 + (1 + \kappa)X_2. \quad (\text{A32})$$

Therefore,

$$\frac{W}{W_a} - 1 \approx \frac{X_1 + (1 + \kappa)X_2}{X_1 + X_2} - 1 = \kappa \left(\frac{X_2}{X_1 + X_2} \right). \quad (\text{A33})$$

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

$$E \left[\frac{W}{W_a} \right] - 1 \approx \frac{\kappa}{2}. \quad (\text{A34})$$

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

Proof of Proposition 3: Define

$$D_0 = E \left[\frac{\partial \bar{h}_T^*}{\partial \beta(c_1)'}, \frac{\partial \bar{h}_T^*}{\partial \beta(c_2)'}, \frac{\partial \bar{h}_T^*}{\partial \mu'} \right] = \begin{bmatrix} C & -F' \\ \text{O}_{(N+K) \times 2(N+K)} & -I_{N+K} \end{bmatrix}, \quad (\text{A35})$$

we have the asymptotic variance of $(\hat{b}, \hat{\mu})$ as $(D_0' S^{-1} D_0)^{-1} = D_0^{-1} S (D_0')^{-1}$. Using

$$D_0^{-1} = \begin{bmatrix} C^{-1} & -C^{-1} F' \\ \text{O}_{(N+K) \times 2(N+K)} & -I_{N+K} \end{bmatrix}, \quad (\text{A36})$$

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

$$\begin{bmatrix} C^{-1}(S_{11} - F' S_{21} - S_{12} F + F' S_{22} F) C^{-1} & -C^{-1}(S_{12} - F' S_{22}) \\ -(S_{21} - S_{22} F) C^{-1} & S_{22} \end{bmatrix}, \quad (\text{A37})$$

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

$$Q_2 C^{-1} (S_{11} - F' S_{21} - S_{12} F + F' S_{22} F) C^{-1} Q_2', \quad (\text{A38})$$

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

$$Q_2 C^{-1} F' = \begin{bmatrix} QV^{-1}\mu\beta(c_1)' \\ QV^{-1}\mu\beta(c_2)' \end{bmatrix} = \text{O}_{2N \times (N+K)}. \quad (\text{A39})$$

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.*

References

- Anderson, Theodore W., 1984, *An introduction of multivariate statistical analysis*, 2nd edition (Wiley, New York).
- Black, Fischer, 1972, Capital market equilibrium with restricted borrowing, *Journal of Business* 45, 444–454.
- 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* forthcoming.
- DeSantis, Giorgio, 1993, Volatility bounds for stochastic discount factors: tests and implications from international financial markets, Ph.D. dissertation, Graduate School of Business, University of Chicago.
- 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., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
- Ferson, Wayne, 1992, 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, 2000, Some generalized tests of mean-variance efficiency, 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 S. 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.
- Jorion, Philippe, 1985, International portfolio diversification with estimation risk, *Journal of Business* 58, 259–278.
- Kandel, Shmuel, and Robert F. Stambaugh, A mean-variance framework for tests of asset pricing models, *Review of Financial Studies* 2, 125–156.
- Lintner, John, 1965, The valuation of risky assets and the selection of risky investments in the portfolios and capital budgets, *Review of Economics and Statistics* 47, 13–37.
- 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., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.
- Merton, Robert C., 1972, An analytic derivation of the efficient portfolio frontier, *Journal of Financial and Quantitative Analysis* 7, 1851–1872.
- 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.
- 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.
- 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).
- Shanken, Jay, 1992, On the estimation of beta-pricing models, *Review of Financial Studies* 5, 1–33.
- Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.
- Zellner, Arnold, 1971, *An introduction to Bayesian Inference in Econometrics*, (Wiley, New York).
- Zhou, Guofu, 1993, Asset pricing test under alternative distributions, *Journal of Finance* 48, 1925–1942.
- Zhou, Guofu, 1994, Analytical GMM tests: Asset pricing with time-varying risk premiums, *Review of Financial Studies* 7, 687–709.

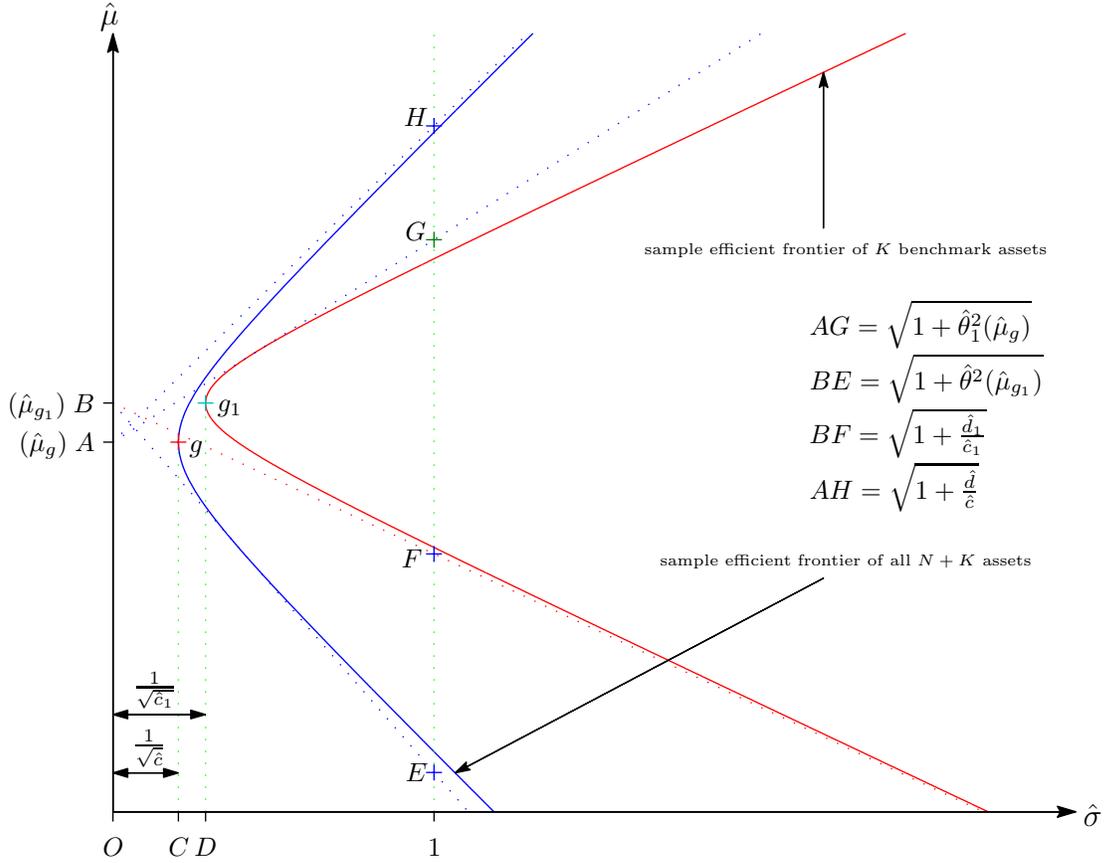


Figure 1
The Geometry of Mean-Variance Spanning Tests

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 $\hat{a}_1 = \hat{\mu}'_1 \hat{V}_{11} \hat{\mu}_1$, $\hat{b}_1 = \hat{\mu}'_1 \hat{V}_{11} \mathbf{1}_K$, $\hat{c}_1 = \mathbf{1}'_K \hat{V}_{11} \mathbf{1}_K$, 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 $\hat{a} = \hat{\mu}' \hat{V} \hat{\mu}$, $\hat{b} = \hat{\mu}' \hat{V} \mathbf{1}_{N+K}$, $\hat{c} = \mathbf{1}'_{N+K} \hat{V} \mathbf{1}_{N+K}$, 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 $\hat{\mu}_{g_1}$.

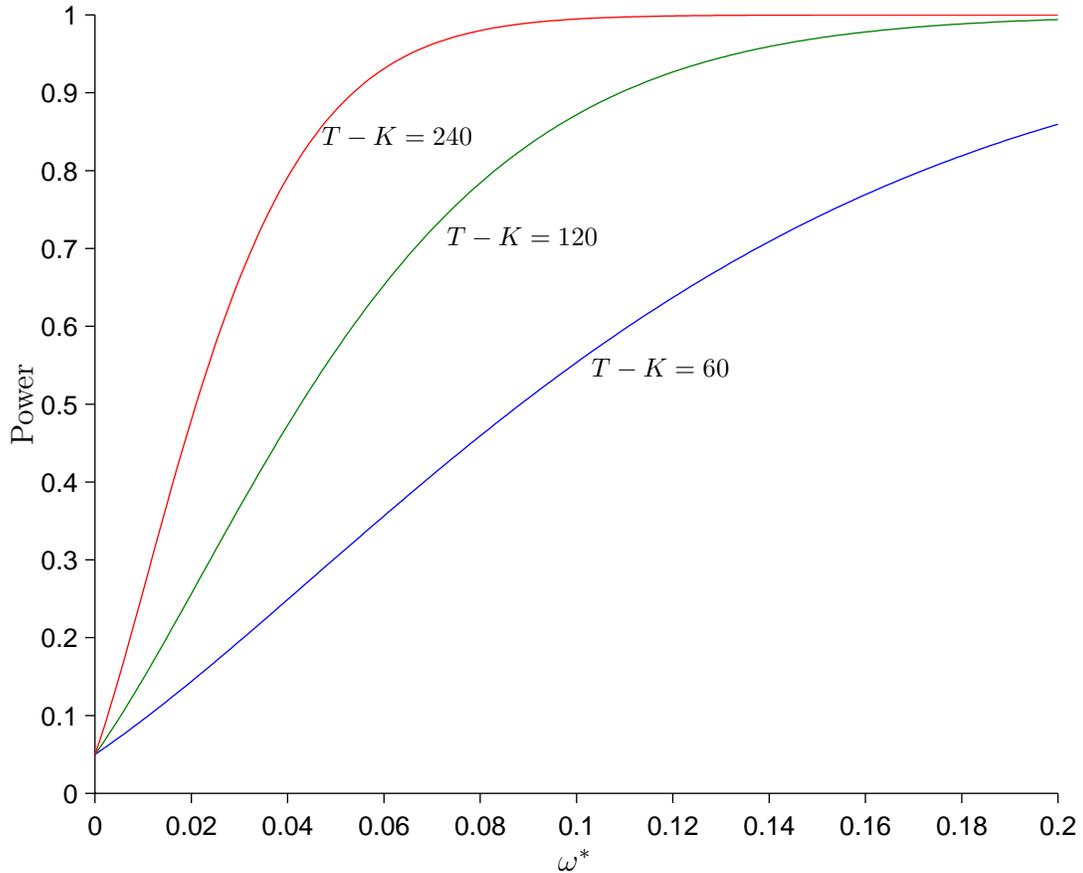


Figure 2

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.

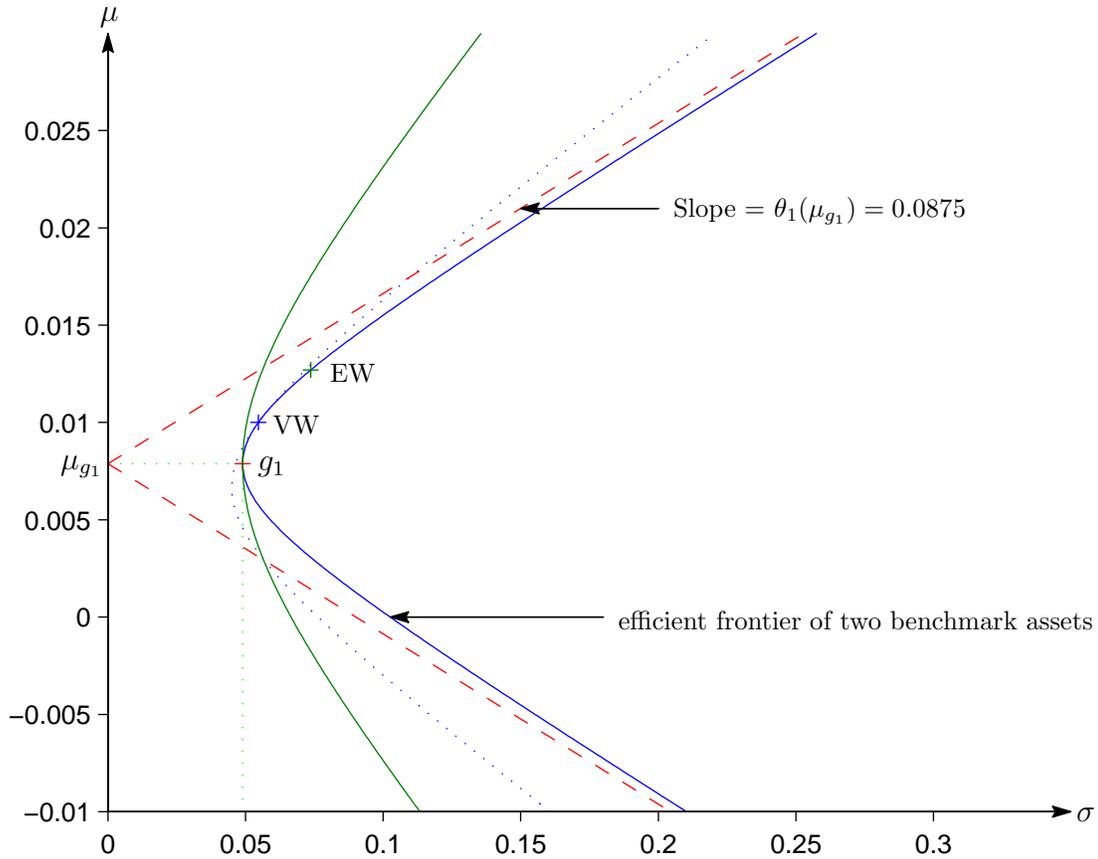


Figure 3
Minimum-Variance Frontier of Two Benchmark Assets

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.

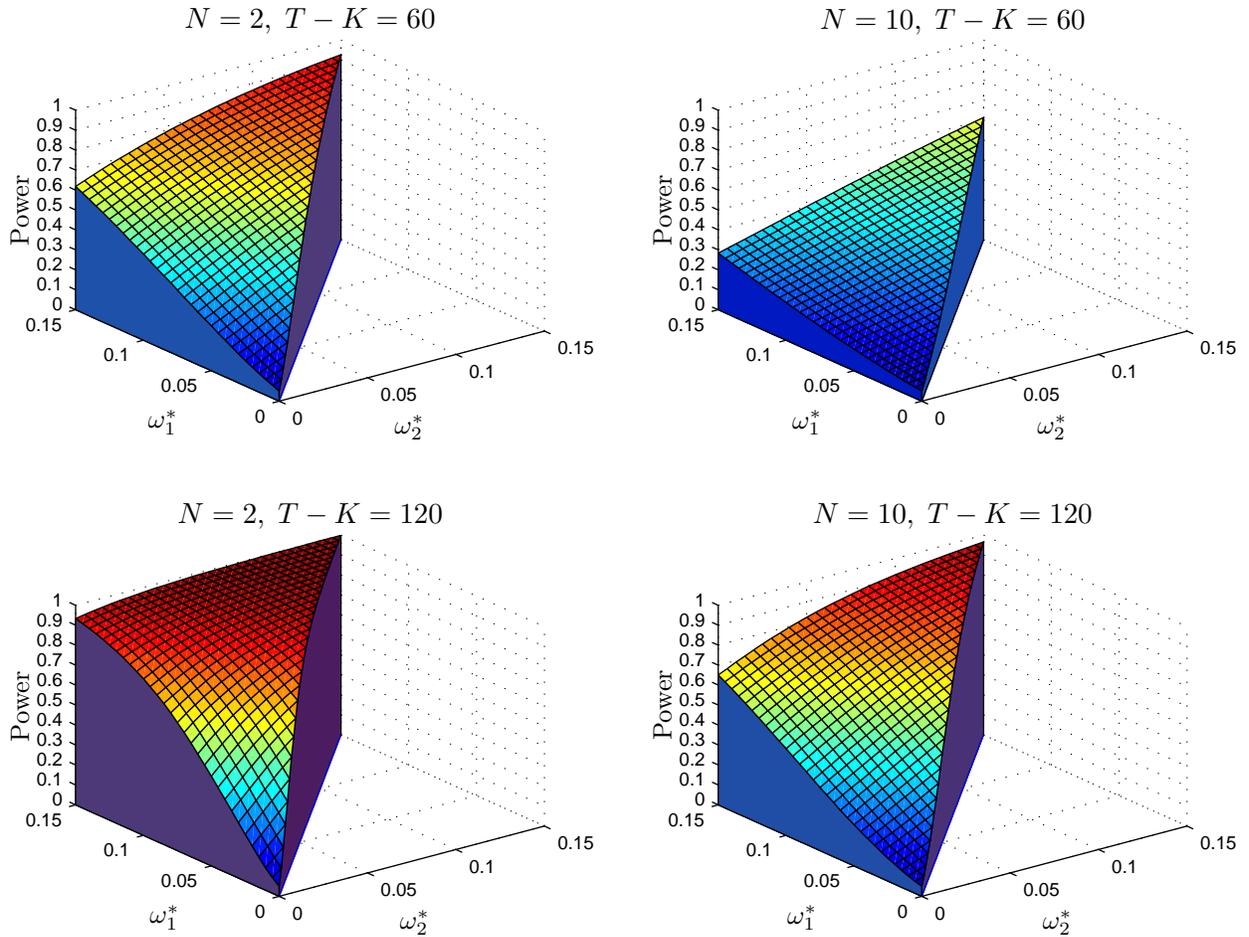


Figure 4
Power Function of Likelihood Ratio Test

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.

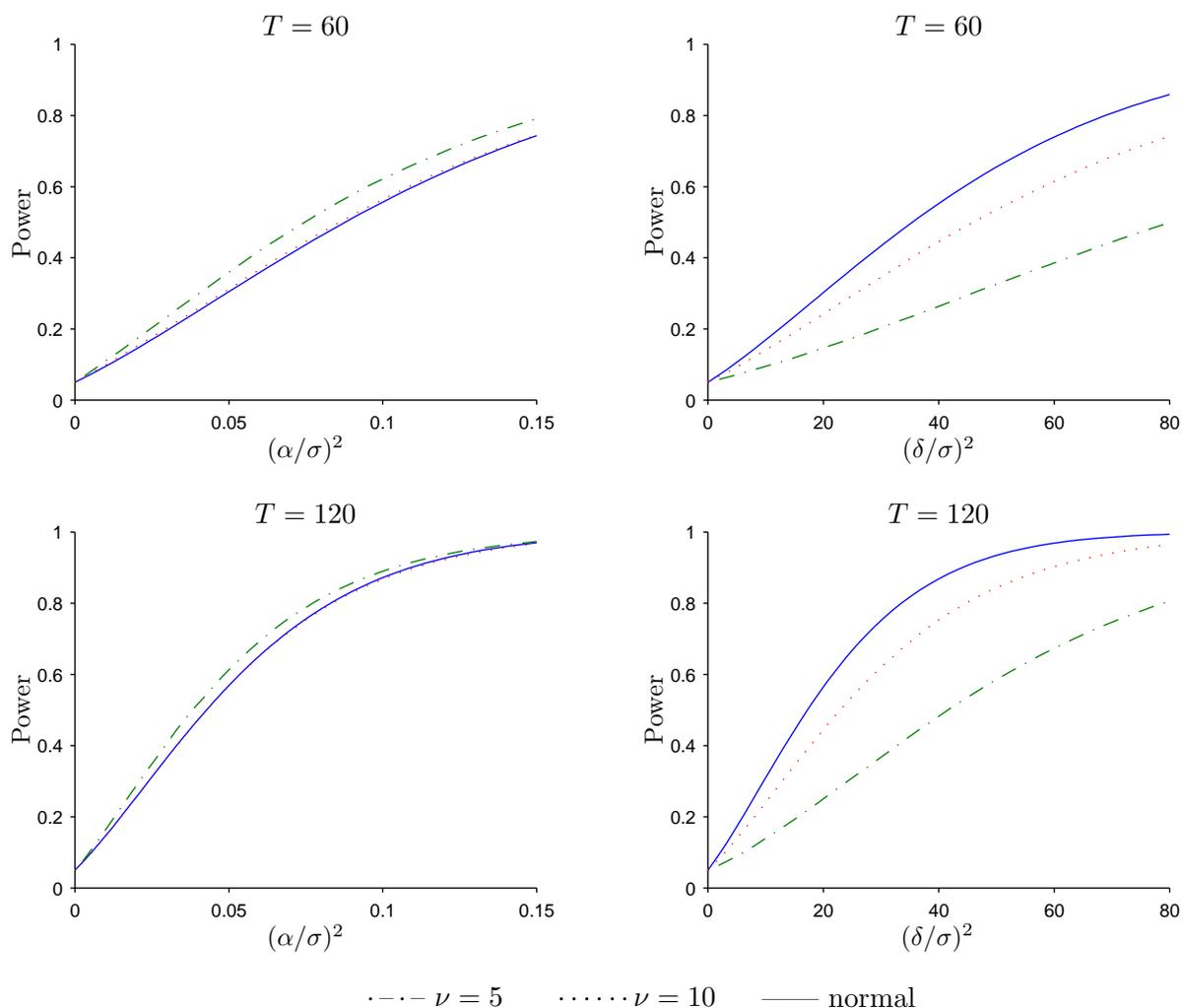


Figure 5

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.

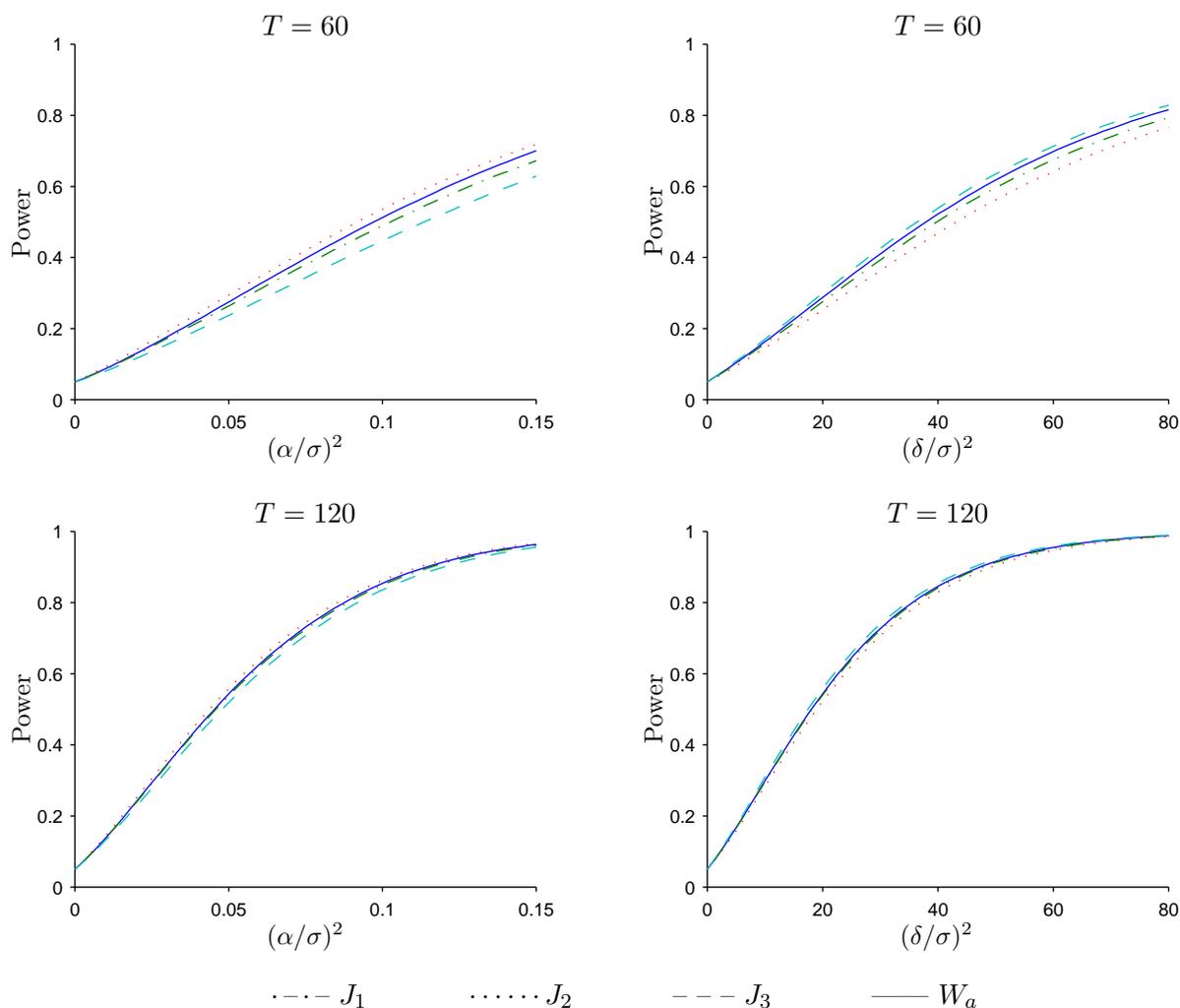


Figure 6
Power Function of GMM Wald Tests Under Multivariate Normal Returns

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.

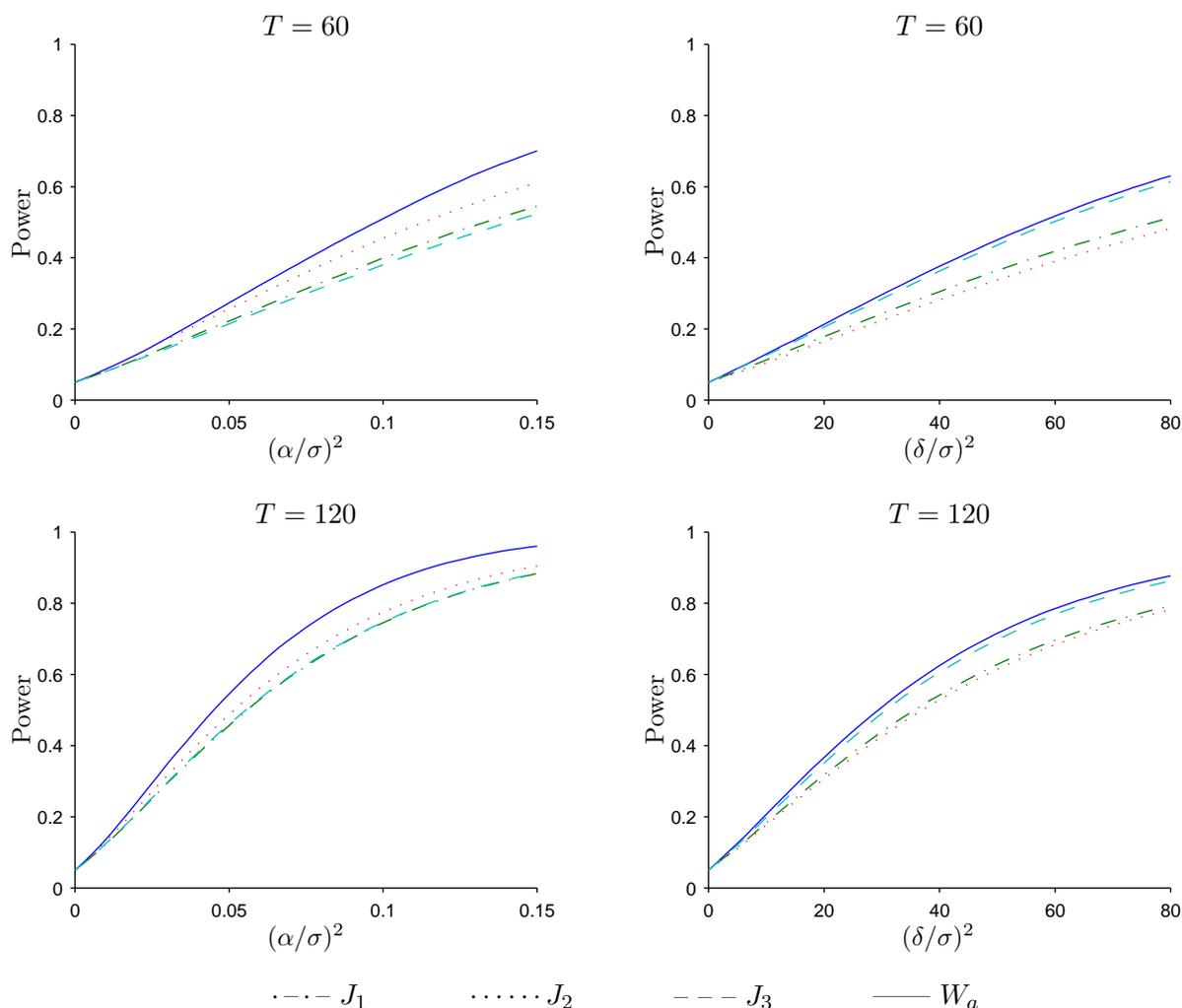


Figure 7

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.

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.

K	N	T	Actual Probabilities of Rejection				
			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	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	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		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		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	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	
		$\Delta a = 0.0229$ $\Delta b, \Delta c = 0$	$\Delta a, \Delta b = 0$ $\Delta c = 76.69$
α_1	α_2		
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.

K	N	T	Actual Probabilities of Rejection			
			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 (W) and GMM Wald (W_a) tests 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 GMM Wald tests are based on 95th percentile of χ^2_{2N} . The table also presents the average ratio of the two Wald tests and probabilities of rejection using the small sample version of the Wald tests with the rejection decision based on 95th percentile of their exact distributions under normality. Results for different values of number of test assets (N) and time series observations (T) are based on 100,000 simulations.

N	T	Average W/W_a	Actual Probabilities of Rejection			
			W	W_a	Small Sample	
			W	W_a	W	W_a
Degrees of Freedom = 5						
2	60	1.862	0.197	0.040	0.150	0.024
	120	1.877	0.197	0.036	0.173	0.028
	240	1.893	0.208	0.037	0.195	0.033
5	60	1.814	0.329	0.048	0.199	0.018
	120	1.834	0.308	0.036	0.242	0.022
	240	1.855	0.317	0.035	0.283	0.028
10	60	1.787	0.560	0.086	0.246	0.013
	120	1.808	0.472	0.044	0.320	0.017
	240	1.835	0.459	0.035	0.383	0.022
25	60	1.762	0.979	0.553	0.270	0.008
	120	1.775	0.848	0.118	0.417	0.009
	240	1.803	0.755	0.044	0.538	0.013
Degrees of Freedom = 10						
2	60	1.168	0.116	0.073	0.079	0.046
	120	1.167	0.099	0.060	0.082	0.047
	240	1.167	0.094	0.055	0.086	0.048
5	60	1.167	0.192	0.109	0.091	0.043
	120	1.166	0.147	0.076	0.099	0.047
	240	1.167	0.128	0.062	0.106	0.049
10	60	1.166	0.372	0.221	0.103	0.043
	120	1.166	0.241	0.116	0.120	0.046
	240	1.167	0.185	0.079	0.128	0.048
25	60	1.165	0.943	0.851	0.113	0.039
	120	1.165	0.635	0.377	0.150	0.041
	240	1.166	0.407	0.166	0.175	0.044

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.

K	N	T	Actual Probabilities of Rejection				
			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	