NBER WORKING PAPER SERIES

ASSESSING THE RISK IN SAMPLE MINIMUM RISK PORTFOLIOS

Gopal K. Basak
Ravi Jagannathan
Tongshu Ma

Working Paper 10447
http://www.nber.org/papers/w10447

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2004

We thank the helpful comments and suggestions of seminar participants at CIRANO, Harvard Business School, Norwegian School of Management BI, NHH: Norwegian School of Economics and Business Administration, and the University of North Carolina. We are responsible for any errors and omissions. The views expressed herein are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

Assessing the Risk in Sample Minimum Risk Portfolios
Gopal K. Basak, Ravi Jagannathan, and Tongshu Ma
NBER Working Paper No. 10447
April 2004
JEL No. G11, G12

ABSTRACT

We show that the in-sample estimate of the variance of a global minimum risk portfolio constructed using an estimated covariance matrix of returns will on average be strictly smaller than its true variance. Scaling the in-sample estimate upward by a standard degrees-of-freedom related factor or using the Bayes covariance matrix estimator can be inadequate; the correction is likely to be twice as large as the standard correction when returns are i.i.d. multivariate Normal. We develop a Jackknife-type estimator of the optimal portfolio's variance that is valid when returns are i.i.d.; and a variation that may be better when returns exhibit volatility persistence.

We empirically demonstrate the need to correct for in-sample optimism by considering an optimal portfolio of 200 stocks that has the lowest tracking error when the S&P500 is the benchmark and three years of daily return data are used for estimating covariances. When the optimal portfolio is constructed using the sample covariance matrix, the standard deviation of the tracking error is 1.46 percent whereas its in-sample estimate is 0.94 percent. Standard degrees of freedom correction gives an estimate of 1.10 percent; our correction, 1.24 percent; and the weighted Jackknife, 1.36 percent.

Gopal K. Basak
Department of Mathematics
University of Bristol
Bristol, BS8 1TW
UK
gopal.k.basak@bristol.ac.uk

Ravi Jagannathan
J.L. Kellogg Graduate School of Management
2001 Sheridan Road
Northwestern University
Evanston, IL 60208-2001
and NBER
rjaganna@northwestern.edu

Tongshu Ma
Department of Finance
David Eccles School of Business
University of Utah
Salt Lake City, UT 84112
fintsm@business.utah.edu

I. Introduction

While mean variance portfolio theory has been around for nearly 50 years, its use has become widespread mostly during the past decade. This is primarily due to the decline in the cost of acquiring and processing financial market data. The decline in the cost of computing has also made large-scale optimization feasible for most investors, making it possible to work directly with returns on individual securities when constructing efficient portfolios instead of first grouping them into asset classes.

The use of a large collection of primitive securities has the advantage of helping construct more efficient portfolios. The disadvantage is that it becomes more difficult to assess the true variance of those portfolios when covariances of asset returns are not known and their estimates are used as inputs to the optimizer. It has been observed in the literature that the in-sample value of the variance of an optimal portfolio constructed using historical return data is an optimistic estimate of the portfolio's true variance,[1] hereafter referred to as the out-of-sample variance. This bias, commonly referred to as in-sample optimism, increases with the number of assets used to construct mean-variance efficient portfolios.

In this paper we identify the reason for this in-sample optimism and suggest methods for correcting for it. Since our focus is in assessing the risk of an efficient portfolio and not its mean, we focus on the global minimum variance and minimum benchmark tracking-error variance portfolios constructed using an estimated covariance matrix. We refer such portfolios as *sample minimum risk portfolios (SMRPs)* in this paper. We demonstrate that the in-sample optimism can be substantial in certain situations.

It might be argued that by suitably scaling up the in-sample variance by a factor that is related to the degree-of-freedom of the distribution of the estimated covariance matrix, we may be able to correct for in-sample optimism. When returns over time are drawn from an i.i.d. multivariate Normal distribution, we show that scaling up the in-sample variance by a degrees-of-freedom related factor provides an unbiased estimate of the variance of the true global minimum variance portfolio. Since the out-of-sample variance of the sample minimum risk portfolio is strictly larger on average, this procedure does not adequately correct for the in-sample optimism.

The example in Table I, Panel D, illustrates the inadequacy of the standard degrees-of-freedom correction. There we report the properties of the minimum tracking-error variance portfolio constructed using the 200 largest stocks on the NYSE/AMEX/NASDAQ, with the S&P 500 as the benchmark. The covariance matrices are estimated using three years daily return. (So the sample size T is approximately 750.) When the sample covariance matrix is used to construct the SMRP, the average out-of-sample standard deviation is 1.46 percent whereas the corresponding in-sample number is 0.94 percent, i.e., the ratio of the two variances is 2.41 ($= (\frac{1.46}{0.94})^2$). In contrast, as we show later, the classical degrees-of-freedom based scaling factor is $(\frac{T-1}{T-N})$, which is only 1.36 ($= \frac{750-1}{750-200}$). Even though the use of a three-factor model reduces the in-sample optimism, it still remains large. The in-sample estimate is 1.64 percent whereas the corresponding out-of-sample standard deviation of 1.93 percent. The ratio of the out-of-sample to in-sample variances in that case is 1.38.

In fact, when returns are i.i.d. multivariate normal and the minimum risk portfolio is constructed without portfolio weight constraints, we show that it necessary to add twice the amount of the degrees-of-freedom correction to the in-sample risk to arrive at a good estimate of of the out-of-sample risk. For example, in the first row of Table I, Panel A, the degrees-of-freedom correction is $5.42 - 4.64 = 0.78$ percent. Adding twice this amount to the in-sample risk gives us $4.64 + 2 \times 0.78 = 6.20$ percent, which is almost the realized out-of-sample risk of 6.29 percent.

In addition to being inadequate, degrees-of-freedom based corrections are difficult to derive in general. For example, there are no available degrees-of-freedom corrections for the shrinkage estimator of Ledoit or the constant correlation model of Elton and Gruber (1973). Degrees-of-freedom corrections are also unavailable when efficient portfolios are constructed subject to portfolio weight constraints, even though portfolio managers often face such constraints, and such constraints can improve the portfolios' performance (Frost and Savarino (1988), Jagannathan and Ma (2003)).

It would appear that we can avoid the in-sample optimism through the use of a Bayesian approach since it would take into account the randomness of the means and covariances used in the optimization procedure. However, with standard diffuse priors commonly used in empirical works, we find that the variance of the efficient portfolio computed using the predictive distribution can also be substantially below its true variance. This should come as no surprise since with diffuse priors, the predictive covariance matrix is the sample covariance matrix multiplied by the factor $\frac{T+1}{T-N-2}$ (see Section III). This has two implications. First, the portfolio weights would be the same as those obtained using the sample covariance matrix. Second, the variance of the global minimum variance portfolio under the predictive distribution is simply $\frac{T+1}{T-N-2}$ times the in-sample variance computed using the sample covariance matrix under the Classical method. Notice that the factor, $\frac{T+1}{T-N-2}$, is only slightly larger than the degrees-of-freedom correction, $\frac{T-1}{T-N}$, mentioned earlier, for the sample covariance matrix under the Classical method. Hence the variance of the global minimum variance portfolio computed using the predictive distribution with standard diffuse priors will be almost the same as the degree-of-freedom adjusted in-sample variance under the Classical method. The inadequacy of standard diffuse priors and the need for modifying them is discussed by Jacquier, Kane and Marcus (2002) in a related context.

There is a large literature on testing the mean-variance efficiency of a given portfolio.[2] These tests can be interpreted as examining whether the distance between a given benchmark portfolio and a particular efficient portfolio constructed using sample moments is zero after allowing for sampling errors. Although these tests also involve constructing efficient portfolios based on estimated covariance matrices, the sampling theory associated with these tests differ in important ways from that associated with the estimate of the variance of the sample minimum risk portfolio developed in this paper.

When there is no risk-free asset, these tests for the efficiency of a given portfolio rely on asymptotic theory, whereas in-sample optimism exists only in finite samples. When there is a risk-free asset and asset returns are i.i.d. multivariate Normal, these tests[3] are based on exact finite sample distribution theory. However, these results do not readily extend to address the in-sample optimism issue examined in this paper. In what follows we explain why.

Consider forming the global minimum variance portfolio based on S, an estimate of the unknown covariance matrix, Σ. The vector of portfolio weights is given by:

$$w_s = \frac{S^{-1}\mathbf{1}}{\mathbf{1}'S^{-1}\mathbf{1}}. \tag{1}$$

We use $\mathbf{1}$ to denote the column vector of ones throughout the paper. The variance of this portfolio's return is,

$$w_s'\Sigma w_s = \frac{\mathbf{1}'S^{-1}\Sigma S^{-1}\mathbf{1}}{(\mathbf{1}'S^{-1}\mathbf{1})^2}. \tag{2}$$

Since Σ is unknown, the variance of w_s is also unknown and has to be estimated.

In MacKinlay (1987) and Gibbons, Ross, and Shanken (1989), the test statistic—which provides a measure of the distance between the given benchmark portfolio and a particular sample efficient portfolio—also involves the unknown covariance matrix of returns, Σ. However, when we replace the unknown Σ with the sample covariance matrix S, the test statistic still has a known finite sample distribution when returns are i.i.d multivariate Normal. In contrast, if we replace the unknown Σ with its estimate S in (2), we just get the in-sample variance,

$$w_s'Sw_s = \frac{1}{\mathbf{1}'S^{-1}\mathbf{1}}.$$

This does not help since the in-sample variance provides a downward biased estimate of the population variance.

We therefore propose a jackknife-type method for estimating the risk in sample minimum risk portfolios. We show that this estimator is valid when returns are i.i.d. or satisfy the exchangeability condition. Simulation results indicate that this method gives more precise estimates of the sample minimum risk portfolio's risk than the in-sample estimate, the degree-of-freedom adjusted in-sample estimate, and the Bayesian estimate under diffuse priors.

There is general agreement that variances of returns and covariances among returns vary in a systematic stochastic fashion over time. To account for this persistence we consider two approaches. The first is a variation of the jackknife-type estimator for the out-of-sample variance of SMRPs that weights recent observations more heavily than those in the distant past.[4] We find that this approach provides a more precise estimate of the out-of-sample risk of the SMRPs. The second approach is based on the Dynamic Conditional Correlation Model of Engle (2002). This approach provides an accurate estimate of risks in the SMRPs constructed using one- and three-factor models. However, in our sample, SMRPs constructed using one- and three-factor models have a significantly higher risk when compared to SMRPs constructed using the sample covariance matrix. This suggests that three factors are insufficient to account for the correlation among returns, and futher work is needed to model the block diagonal structure of the correlation matrix of residual returns in a three-factor model.

II. Relation Between In-sample and Out-of-sample Variances of Minimum Risk Portfolios Constructed Using Estimated Covariance Matrices

We use the following notation: R_t is the $N \times 1$ vector of date t returns (or returns in excess of some benchmark return when the objective is to minimize the tracking-error variance) on N primitive assets; $\mu = E(R_t)$, $\Sigma = \text{Cov}(R_t)$, and S and $\hat{\mu}$ are unbiased estimates of Σ and μ based on T observations on returns, $\{R_1, R_2, \ldots, R_T\}$. With these notations, our analysis applied to both portfolio variance minimization and tracking-error variance minimization. For brevity, we will use the term "return" to mean both raw return and excess return in excess of the bentchmark, and the term "portfolio variance minimization" to mean both portfolio variance minimization and tracking-error variance minimization in our analysis.

We use w_s to denote a minimum variance portfolio constructed using sample moments of the returns. That is, w_s is a sample minimum risk portfolio. We use w_p to denote a efficient portfolio constructed from the population moments of the returns. That is, w_p is a population minimum risk portfolio. Then the in-sample variance of w_s is $w_s' S w_s$, and its out-of-sample variance is $w_s' \Sigma w_s$. $w_p' \Sigma w_p$ is the variance of the population minimum risk portfolio.

Let w_p be the population minimum risk portfolio with an expected return of ν, i.e., $w = w_p$

solves the following problem:

$$
\begin{aligned}
&\min_{w}\{w'\Sigma w\} && (3)\\
&\text{subject to} \qquad w'\mu = \nu\\
&w'\mathbf{1} = 1, \qquad \underline{w_i} \le w_i \le \bar{w}_i,\\
&\text{for some constants} \qquad \underline{w_i}, \bar{w}_i, i = 1, \ldots, N
\end{aligned}
$$

The Lagrangian function of the above minimization problem is:

$$
f_p(w) = \frac{1}{2}w'\Sigma w - \delta_1(w'\mu - \nu) - \delta_2(w'\mathbf{1} - 1) - (w - \underline{w})'\theta_1 + (w - \bar{w})'\theta_2 \tag{4}
$$

where δ_1 and δ_2 are nonnegative constants and θ_1 and θ_2 are nonnegative vectors. Since Σ is a positive definite matrix, the Lagrangian function is strictly convex with a unique minimum. Let $w = w_p$ along with $(\delta_1 = \delta_1^p,\ \delta_2 = \delta_2^p,\ \theta_1 = \theta_1^p,\ \theta_2 = \theta_2^p)$ solve the minimization problem given in (3). Then $\underline{w} \le w_p \le \bar{w}$; $(w_{p,i} - \underline{w_i})\theta_{1,i}^p = 0, i = 1 \ldots N$; $(\bar{w}_i - w_{p,i})\theta_{2,i}^p = 0, i = 1 \ldots N$; $({w_p}'\mu - \nu) = 0$; and $(w_p'\mathbf{1} - 1) = 0$, by the complimentary slackness condition for constrained minimization.

Now consider the sample analogue of the minimization problem given in equation (3) obtained by replacing Σ and μ with their sample counterparts, S and $\hat{\mu}$. The Lagrangian function of this minimization problem is:

$$
f_s(w) = \frac{1}{2}w'S\omega - \delta_1(w'\hat{\mu} - \nu) - \delta_2(w'\mathbf{1} - 1) - (w - \underline{w})'\theta_1 + (w - \bar{w})'\theta_2. \tag{5}
$$

Let w_s, together with the auxiliary parameters $(\delta_1^s,\ \delta_2^s,\ \theta_1^s,\ \theta_2^s)$, minimize the sample Lagrangian function, $f_s(w)$.

We have the following relation about the in-sample variance of w_s and the population variances of w_s and w_p :

LEMMA 1: *The following inequality relations hold:*

$$
E(w_s'Sw_s) < w_p'\Sigma w_p < E(w_s'\Sigma w_s) - 2E(\delta_1^s(w_s'\mu - \nu)). \tag{6}
$$

PROOF: See Appendix A.

This lemma says that the in-sample variance would on average under-estimate the true variance of a sample minimum risk portfolio whenever $E[\delta_1^s(w_s'\mu - \nu)]$ is positive. Our simulations indicate that for parameter values typically encountered in practice $E[\delta_1^s(w_s'\mu - \nu)]$ is likely to be small. Hence there is likely to be in-sample optimism for most sample minimum risk portfolios.

For the special case of the global minimum variance portfolio, the constraint that the expected return on the portfolio should equal some target value in the minimization problems does not apply and the last term in (6) drops out. This gives us the following proposition.

PROPOSITION 1: *Let SMRP denote the global minimum variance portfolio constructed using an unbiased estimate of the covariance matrix. Then, the in-sample estimate of the variance of SMRP will on average be strictly smaller than the variance of the true global minimum variance portfolio, which on average will be strictly smaller than the out-of-sample variance of the SMRP. That is,*

$$E(w_s'Sw_s) < w_p'\Sigma w_p < E(w_s'\Sigma w_s). \tag{7}$$

We now characterize the difference between the first two terms and the difference between the last two terms of the above inequality when there are no portfolio weight constraints. For that purpose, let q_1 denote the difference between the middle term and the left-most term in the above inequality; and q_2 denote the difference between the right-most term and the middle term; i.e., $q_1 = w_p'\Sigma w_p - E(w_s'Sw_s)$ and $q_2 = E(w_s'\Sigma w_s) - w_p'\Sigma w_p$.

PROPOSITION 2: *Let SMRP denote the global minimum variance portfolio constructed using an unbiased estimate of the covariance matrix with no portfolio weight constraints. Then on average the out-of-sample variances of SMRP equals its in-sample variance plus two strictly positive terms as given below.*

$$E(w_s'\Sigma w_s) = E(w_s'Sw_s) + q_1 + q_2, \tag{8}$$

where

$$q_1 \equiv w_p'\Sigma w_p - E(w_s'Sw_s) = E[(w_p - w_s)'S(w_p - w_s)], \tag{9}$$

$$q_2 \equiv E(w_s'\Sigma w_s) - w_p'\Sigma w_p = E[(w_p - w_s)'\Sigma(w_p - w_s)], \tag{10}$$

are two positive numbers.

Proof: Equation (8) and the first part of eqs.(9) and (10) are obvious from Proposition 1. So we only need to prove the second part of (9) and (10).

With no portfolio weight constraints, the sample minimum risk portfolio, w_s, is given by (1). The out-of-sample variance of this portfolio, $w_s'\Sigma w_s$, can be written as:

$$w_s'\Sigma w_s = w_p'\Sigma w_p + 2(w_p'\Sigma)(w_s - w_p) + (w_s - w_p)'\Sigma(w_s - w_p).$$

Notice that,

$$w_p'\Sigma w_s = \frac{\mathbf{1}'\Sigma^{-1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\Sigma\frac{S^{-1}\mathbf{1}}{\mathbf{1}'S^{-1}\mathbf{1}} = \frac{\mathbf{1}'}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\frac{S^{-1}\mathbf{1}}{\mathbf{1}'S^{-1}\mathbf{1}} = \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} = w_p'\Sigma w_p.$$

Hence we get:

$$w_s'\Sigma w_s = w_p'\Sigma w_p + (w_s - w_p)'\Sigma(w_s - w_p).$$

Taking expectation on both sides, we get:

$$E(w_s'\Sigma w_s) = w_p'\Sigma w_p + E[(w_s - w_p)'\Sigma(w_s - w_p)]. \tag{11}$$

Using the same logic as before, we can write

$$w_p'Sw_p = w_s'Sw_s + 2(w_s'S)(w_p - w_s) + (w_p - w_s)'S(w_p - w_s).$$

Notice that

$$w_s'Sw_p = \frac{\mathbf{1}'S^{-1}}{\mathbf{1}'S^{-1}\mathbf{1}}S\frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} = \frac{\mathbf{1}'}{\mathbf{1}'S^{-1}\mathbf{1}}\frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} = \frac{1}{\mathbf{1}'S^{-1}\mathbf{1}} = w_s'Sw_s.$$

This means

$$w_p'Sw_p = w_s'Sw_s + (w_p - w_s)'S(w_p - w_s)$$

Taking expectations, we get

$$w_p'\Sigma w_p = E(w_p'Sw_p) = E(w_s'Sw_s) + E[(w_p - w_s)'S(w_p - w_s)]. \tag{12}$$

The Proposition follows from combining (11) and (12). □

Both q_1 and q_2 involve the unknown covariance matrix Σ, and therefore are unknown. We now derive an unbiased estimate of q_1 under the additional assumption that the returns are drawn from an i.i.d. multivariance Normal distribution. After that, we will say something about q_2 also.

Proposition 3: *Assume that returns are drawn from an i.i.d. multivariate Normal distribution. Let SMRP denote the global minimum variance portfolio constructed using an unbiased estimate of the covariance matrix with no portfolio weight constraints. Then the quantity q_1 in Proposition 2 is*

$$q_1 = \frac{N-1}{T-N} E(w_s' S w_s). \tag{13}$$

Proof: With no portfolio weight constraints, the in-sample variance of w_s is $\frac{1}{\mathbf{1}' S^{-1} \mathbf{1}}$. The average in-sample variance of this portfolio is

$$E\left(w_s' S w_s\right) = E\left(\frac{1}{\mathbf{1}' S^{-1} \mathbf{1}}\right).$$

We know that $(T-1)S$ is a N dimensional Wishart distribution with degree of freedom $(T-N)$ and parameter matrix Σ. That is,

$$(T-1)S \sim W_N(T-N,\ \Sigma).$$

[Here we follow the notation of Muirhead (1982)]. From Theorem 3.2.11 of Muirhead (1982), we know that

$$\{\mathbf{1}'[(T-1)S]^{-1}\mathbf{1}\}^{-1} \sim W_1\left(T-N,\ (\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1}\right).$$

Now if a scalar random variable x has a Wishart$(T-N,\ \sigma^2)$ distribution then $\frac{x}{\sigma^2}$ is χ^2_{T-N} distributed (Muirhead (1982), p. 87). Since the expectation of the χ^2_{T-N} is $(T-N)$, we get,

$$E\left(\frac{(\mathbf{1}'[(T-1)S]^{-1}\mathbf{1})^{-1}}{(\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1}}\right) = T-N,$$

i.e.,

$$E\left(\frac{T-1}{T-N}\frac{1}{\mathbf{1}'S^{-1}\mathbf{1}}\right) = \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} = w_p'\Sigma w_p. \tag{14}$$

This implies that

$$q_1 = w_p'\Sigma w_p - E(w_s' S w_s) = \frac{N-1}{T-N} E(w_s' S w_s).$$

□

The above equation suggests a degrees-of-freedom adjusted in-sample variance as

$$(1+q_1) \times (\text{in-sample variance}) = \frac{T-1}{T-N} \times (\text{in-sample variance}). \tag{15}$$

This degrees-of-freedom adjusted in-sample variance gives an unbiased estimate of the variance of the population minimum risk portfolio, which is still smaller than the variance of the sample minimum variance portfolio.

Remark: Notice that the quantities q_1 and q_2 in Proposition 2 are likely to be close when the number of observations is reasonably large. Hence the in-sample estimate of the variance plus $2q_1$ is likely to provide a more precise estimate of the out-of-sample variance.

A natural question that arises at this stage is whether the use of Bayesian covariance matrix estimator would provide an unbiased estimate of the out-of-sample variance of minimum variance portfolios. In the next section we show that this is unlikely to be the case.

III. In-sample Optimism with Bayes Estimator of the Covariance Matrix

The setup of the problem and the notation are the same as described at the beginning of Section II. In addition, we assume that returns are i.i.d. multivariate Normal.

Diffuse priors are often used in empirical studies in finance. Therefore, we assume the portfolio manager has the standard diffuse prior about μ and Σ, the vector of expected returns and return covariance matrix, given by:

$$p(\mu, \Sigma) \propto |\Sigma|^{-(N+1)/2}.$$

We assume that the Bayesian portfolio manager's objective is to choose a portfolio w to minimize the variance of the portfolio's return in the next period,

$$\mathrm{var}(w' R_{T+1}),$$

where R_{T+1} is the next period's return, and $\mathrm{var}(\cdot)$ is the variance under the predictive distribution. We have the following result.

PROPOSITION 4: *Let returns be drawn from an i.i.d. multivariate Normal distribution and S be the sample covariance matrix. Let w_{sB} be the portfolio that minimizes* $\mathrm{var}(w' R_{T+1})$*. Then:*

1. *w_{sB} is the same as w_s, the global minimum variance portfolio constructed using the sample covariance matrix;*

2. $\mathrm{var}(w'_{sB}R_{T+1}) = \frac{(T-1)(T+1)}{T(T-N-2)} w'_s S w_s$.

The proof is in Appendix B.

It is worth noting that the right side of the above equation is very close to the degree-of-freedom (DF) adjusted in-sample variance in (15). In fact, comparing with (15), we see that

$$\begin{aligned} w'_s var(R_{T+1}) w_s &= \left(\frac{T+1}{T}\right)\left(\frac{T-N}{T-N-2}\right) \\ &\times \text{ (DF adjusted in-sample variance)} \end{aligned} \tag{16}$$

The factor $\frac{T+1}{T}\ \frac{T-N}{T-N-2}$ is very close to 1 since in practice the N and T are usually quite large. Hence, the variance of the global minimum variance portfolio under the predictive distribution will be very close in magnitude to the degree-of-freedom adjusted in-sample variance.

The fact that the Bayesian estimate is very close to the degree-of-freedom adjusted in-sample estimate implies there is still in-sample optimism using the Bayes estimates with the standard diffuse priors.

When the covariance matrix has a factor structure, we can also examine the estimation of the covariance matrix and the out-of-sample variance of the optimal portfolio under the predictive distribution. The details are in Appendix C.

IV. Jackknife-type Estimator of the Out-of-sample Variance

A. Portfolio Holding Period and Return Observation Interval Are the Same

For expositional convenience, we first consider the special case where the portfolio manager estimates covariance matrices and computes optimal portfolilo weights at the end of each period, and the manager's investment horizon is also one period. As before we assume that the portfolio manager has time series data for T periods in the immediate past, and estimates the return covariance matrix using these T observations. Let S_T denote this estimate and w_T denote the portfolio weights formed using S_T. The return on this portfolio during period $T+1$ is $w_T' R_{T+1}$. Consider the following estimator of $Var(w_T' R_{T+1})$.

We first drop the $i'th$ return observation, for an arbitrary i in the interval $[1, T]$, and estimate the return covariance matrix. Let $S_{(T-1,\, i)}$ denote the covariance matrix estimate so obtained using the data $\{R_1, \ldots, R_{i-1}, R_{i+1}, \ldots, R_T\}$. We then construct the global minimum variance portfolio using $S_{(T-1,i)}$. Denote the resulting portfolio as $w_{(T-1,i)}$. Note that the portfolio returns, $w'_{(T-1,i)} R_i$ and $w'_T\, R_{T+1}$, have approximately the same distribution for large enough T, if the time series of returns on the primitive assets are i.i.d. Hence a natural estimator of the out-of-sample variance, $Var\left(w_T' R_{T+1}\right)$, would be the sample variance of the T portfolio returns, $[w'_{(T-1,\,1)} R_1,\ w'_{(T-1,\,2)} R_2,\ \ldots,\ w'_{(T-1,\,T)} R_T]$. With daily data, we can ignore the effect of sample mean return on the calculation of the sample variance, and hence the jackknife estimator is

$$\hat{q}^{JK} = \frac{1}{T} \sum_{i=1}^{T} (w'_{(T-1,i)} R_i)^2. \tag{17}$$

While the validity of the jackknife also requires T to be large, it need not be as large as what may be required for the standard central limit theorem. This is because we only need $w'_{T-1,i} R_i$ and $w_T' R_{T+1}$ to have approximately the same distribution.

We can relax the i.i.d. assumption with the assumption that returns have the *exchangeability* property, i.e., for any fixed positive integer k, the joint density of $\{R_1, \ldots, R_k\}$, say $f(r_1, \ldots, r_k)$, is the same as $f(r_{\sigma(1)}, \ldots, r_{\sigma(k)})$ for all permutations $\sigma : \{1, \ldots, k\} \to \{1, \ldots, k\}$.

Jagannathan and Ma (2003) show that using daily return data instead of monthly return data improves the performance of minimum risk portfolios. The empirical findings reported in Liu (2003) suggest that there may not be much advantage to using higher-frequency return data (higher than daily) if the holding period is one month or longer, when a year or more of historical daily return data are available. In view of this, in our empirical analysis we form minimum risk portfolios based on covariance matrices estimated using the past three years of daily return data and hold the portfolio for one month. At the end of each month, we recompute the portfolio weights based on the most recent three years of historical daily returns. Monthly instead of daily rebalancing of portfolio weights is justified in the presence of transactions costs. For risk management and portfolio performance evaluation purposes, examine the portfolios' performance in the month ahead is also more relevant. In what follows we therefore modify the estimator of the out-of-sample variance we discussed to allow for the holding period to be longer than the interval over which returns are measured.

B. Portfolio Holding Period Is Longer Than Return Observation Interval

We now consider the case where returns are measured more frequently, say once a day, but the portfolio manager revises the holdings once every few days, say once a month, with there being l days in a month. Let T denote the number of days of daily return data used to estimate the covariance matrix, S_T, and the efficient portfolio weights, w_T; and $m = T/l$ denote the number of months of observation. Let $q_{T,l}$ denote the sample variance of the l post-formation daily returns on the efficient portfolio, i.e., the sample variance of $(w'_T R_{T+1}, \ldots, w'_T R_{T+l})$. Our objective is to estimate $q_{T,l}$.

We assume that monthly returns (i.e., blocks of l returns) satisfy the exchangeability property. Then, for each month $1 \leq i \leq m$, we estimate the covariance matrix $S_{(T-l,i)}$ by deleting the return data for the l days in month i and construct the global minimum variance portfolio using $S_{(T-l,i)}$. Let the resulting portfolio be $w_{(T-l,i)}$. Compute the sample variance of the sequence of l returns on the efficient portfolio, $\left(w'_{T-l,i} R_{(i-1)l+1}, \ldots, w'_{T-l,i} R_{il}\right)$. Let $q_{T,l}(i)$ denote this sample variance.

Note that $S_{(T-l,\,i)}$ will have approximately the same distribution as S_T, for moderately large T, and so each $w_{(T-1,\,i)}$ will have approximately the same distribution as w_T. Hence, each sequence, $\left(w'_{T-l,i} R_{(i-1)l+1}, \ldots, w'_{T-l,i} R_{il}\right)$, will have approximately the same distribution as the sequence $(w'_T R_{T+1}, \ldots, w'_T R_{T+l})$. Since $q_{T,l}(i)$ will have approximately the same distribution as $q_{T,l}$, the average of the m such variances, $q_{T,l}(i), i = 1, 2, \ldots, m$, provides a consistent estimate of $q_{T,l}$, the object of our interest. This gives us the following jackknife-type estimator:

$$\widehat{q_{T,l}}^{JK} = \frac{1}{m} \sum_{i=1}^{m} q_{T,l}(i). \tag{18}$$

V. Assessing the Optimal Portfolio's Risk When the Covariance Structure Is Time-varying

In Section IV, our jackknife method of assessing the sample optimal portfolio's risk is based on the assumption that the time series of returns satisfy the exchangeability property. However, there is ample evidence in the literature that the first and second moments of returns vary over time in a systematic stochastic fashion, violating the exchangeability condition. We therefore consider the following two additional estimators for assessing the out-of-sample variance of SMRPs.

A. Method Based on the Multivariate GARCH Model

For computational feasibility and statistical reliability we assume that the stock returns have a k factor structure:

$$R_t = \alpha_0 + a_1 D_t + \beta_0 F_t + \beta_1 F_{t-1} + \beta_2 F_{t-2} + \epsilon_t, \tag{19}$$

where R_t is defined before, F_t is the column-vector of returns of k factors in period t, the β's are $N \times k$ matrices of factor loadings, and ϵ_t is the residual vector and is assumed to be i.i.d. over time. Because we are using daily returns, we allow the lagged factor returns to affect the current asset returns. In addition to the factors, we also allow a weekend-and-holiday dummy, D_t, which is one if the previous day has no trading and zero otherwise. We use this dummy variable to catch the weekend and holiday effect in daily stock returns. In our empirical work, we will use a one-factor model and a three-factor model. (That is, k is either 1 or 3.) In the former, the daily market excess return is the factor; in the latter, the daily returns of the Fama-French three-factors are the factors.

We model the evolution of the factor returns using a VAR model with error terms follow the DCC multivariate GARCH:

$$\begin{aligned} F_t &= \phi_0 + \phi_1 D_t + \sum_{k=1}^{p} \Phi_k F_{t-k} + \delta_t, && (20) \\ \delta_t &\sim \text{DCC multivariate GARCH}, && (21) \end{aligned}$$

where ϕ_0 and ϕ_1 are k vectors, the Φ_k's are $k \times k$ matrices, δ_t is a k vector of error term and DCC multivariate GARCH denotes the Dynamic Conditional Correlation multivariate GARCH model developed in Engle (2002) and Tse and Tsui (2002). Details of this model are described in Appendix D.

We use daily data, with optimal portfolios balanced every month. Assume that the optimal portfolio weight, w_T, is already constructed. So our task is to estimate the sample variance of the optimal portfolio's returns in the l days of the next month:

$$Sample\ Variance\ of\ (w_T' R_{T+1}, w_T' R_{T+2}, ..., w_T' R_{T+l}). \tag{22}$$

To do that we can simulate forward the sample paths of $F_{T+1}, ..., F_{T+l}$. Then we use (19) to generate the sample path of future returns, $R_{T+1}, R_{T+2}, ..., R_{T+l}$, and calculate (22). When generating the future factor returns, the standardized residuals in the multivariate GARCH model are drawn from i.i.d. standard normal distributions. When generating the future asset returns from factor returns, the error terms in (19) are generated using the bootstrap method. Specifically, given the parameter estimates of (19), we calculate the realized residuals, $\hat{\epsilon}_1, ..., \hat{\epsilon}_T$. When we use (19) to forecast the returns from the factors, the error terms $\epsilon_{T+1}, ..., \epsilon_{T+l}$ are randomly drawn (with replacement) from these T realized residuals. We use this bootstrap method to catch any nonnormality and any covariance structure in the asset returns that is left out of the factor model.

We repeat this procedure M times and use the average of the M estimates of (22) as our final estimate of the optimal portfolio's out-of-sample variance.

When choosing the number of lags for the VAR model in (20), we first did some experiments using the AIC and BIC criteria. With the Fama-French three factors, we found that over the sample period, the AIC tended to choose the maximum lag length of 20, while the BIC tended to choose very low lag lengths such as one, two, and at most four. With only the market factor, the results are qualitatively similar. Therefore, we decided to choose the order of the VAR (or AR when using only one factor) to be five as a compromise. We allow for five lags also because there are usually as many trading days in a week, and we hope five lags will be sufficient to catch the dynamics of the daily factor returns while avoiding the use of too many parameters.[5]

B. The Weighted Jackknife Method

In this subsection, we modify the jackknife estimator in Section IV to take into account the persistence in the covariance structure.

Recall that when the interval over which returns are measured and the holding period match, the jackknife estimator for i.i.d. data is given by equation (17). If returns do not satisfy the exchangeability condition, specifically, if returns have persistent covariance structure, it would make intuitive sense to use a weighted average instead of straight average in (17), giving more weight to the recent observations. One such estimator is given by:

$$\hat{q}^{JK} = \frac{\sum_{i=1}^{T} e^{\alpha i}(w'_{(T-1,i)}R_i)^2}{\sum_{i=1}^{T} e^{\alpha i}}, \tag{23}$$

for some positive α.

When the efficient portfolio weights are recomputed once every month using historical daily returns data, the jackknife estimator under the i.i.d. assumption is given by (18). When variances and covariances vary over time in a systematic stochastic fashion, we modify the average in (18) into a weighted average, yielding the weighted jackknife estimate of the out-of-sample variance:

$$\widehat{q_{T,l}}^{JK} = \frac{\sum_{i=1}^{m} e^{\alpha i} q_{T,l}(i)}{\sum_{i=1}^{m} e^{\alpha i}}, \tag{24}$$

for some positive α. The motivation for using exponentially declining weights comes from the results in Foster and Nelson (1996).

VI. Empirical Evaluation of the Estimators

A. Data and Methodology

We evaluate the performance of the jackknife and other estimators by examining the mean and standard deviation of the forecast errors as well as the mean absolute forecast error over a large number of out-of-sample holding periods, as in Chan, Karceski, and Lakonishok (1999), Ledoit and Wolf (2003a, 2003b) and Jagannathan and Ma (2003). These papers compare the out-of-sample risks of SMRPs constructed from different covariance matrix estimators. In contrast, we compare how well the jackknife, the in-sample variance, and other estimators perform in estimating the out-of-sample risk of the SMRPs.

At the end of April 1967, we choose 200 stocks with the largest market capitalization from all common domestic stocks traded on the NYSE, the AMEX and NASDAQ, and with monthly return data for all the immediately preceeding 36 months. We estimate the covariance matrix of the 200 stocks using daily return of the preceeding 36 months, and form the global minimum variance and minimum tracking-error variance portfolios. When a daily return is missing, the equally weighted market return of that day is used instead.

When portfolio variance minimization is the objective, we form two global minimum variance portfolios using each covariance matrix estimator. The first portfolio is constructed without imposing any restrictions on portfolio weights, the second is subject to the constraint that portfolio weights should be nonnegative.[6] We then compute estimates of the out-of-sample variance of these portfolios using the jackknife and other estimators. The optimal portfolios are held for one month. Their daily returns are recorded, and the daily return sample standard deviations within that month are calculated.[7] At the end of the next month, the same process is repeated. This gives a time series of post-formation daily return standard deviations for each of the 396 nonoverlapping one month intervals during the period May 1967 to April 2000 for each portfolio, as well as estimates of these out-of-sample standard deviations according to the different estimators.

For tracking-error minimization, following Chan, Karceski, and Lakonishok (1999), we assume the investor is trying to track the return of the S&P 500 index. We construct two tracking-error minimizing portfolios for each covariance matrix estimator in the same manner as the case of portfolio variance minimization. Notice that constructing the minimum tracking-error variance portfolio is the same as constructing the minimum variance portfolio using returns in excess of the benchmark.

B. Covariance Matrix Estimators

The first estimator is the sample covariance matrix:

$$S_N = \frac{1}{T-1}\sum_{t=1}^{T}(R_t - \bar{\mu})(R_t - \bar{\mu})',$$

where T is the sample size, R_t is a N vector of stock returns in period t, and $\bar{\mu}$ is the vector of average returns.

The second estimator assumes that returns are generated according to Sharpe's (1963) one-factor model:

$$r_{it} = \alpha_i + \beta_i r_{mt} + \epsilon_{it},$$

where r_{mt} is the period t return on the value-weighted portfolio of stocks traded on the NYSE, AMEX, and Nasdaq. Then the covariance estimator is

$$S_1 = s_m^2 BB' + D. \tag{25}$$

Here B is the $N \times 1$ vector of β's, s_m^2 is the sample variance of r_{mt}, and D has the sample variances of the residuals in its diagonal, and zeros elsewhere.

The third estimator is the Fama and French (1993) three-factor model, which is similar to the one-factor model.

We also examined the shrinkage estimator proposed by Ledoit (1996). The Ledoit covariance matrix is a weighted average of the sample covariance matrix and the one-factor model based covariance matrix. With three years of daily return data we find that the optimal weight assigned to the sample covariance matrix is in excess of 0.95. Hence the results for the Ledoit covariance matrix are almost the same as those for the sample covariance matrix and therefore not reported.

C. Examining the Performance of the Estimators Using Simulated Data

In order to examine how the different estimators of out-of-sample risks would perform when returns are i.i.d. multivariate Normal, we generate return data using Monte Carlo methods. For that purpose, every time we read in the returns (or excess returns for the case of tracking error minimization) on the 200 stocks over the three-year estimation period, we also read in the corresponding three Fama and French (1993) factors. We then estimate the betas for all the 200 stocks with respect to the factors, calculate the residual variances and the covariance matrix of the factors. Using the betas, the residual variances, and the covariance matrix of the factors, we generate the stock returns and factor returns for each day in the three years and one month period (the total time length of in-sample and out-of-sample) from an i.i.d. multivariate Normal distribution and assuming an exact three-factor structure. Then we apply the same procedure to the simulated data as to the real data. This gives one out-of-sample variance and one set of estimates of it for each month over the 1967/5 to 2000/4 period, just as for the real return data.

D. Empirical Results

Tables I – IV display the empirical results. Different covariance matrix estimators result in different sample minimum risk portfolios (SMRPs). These are displayed on different rows. We also have different estimators that estimate the SMRPs' out-of-sample standard deviations. These are displayed on different columns, along with the realized out-of-sample standard deviations.

Table I panels A and C give the result for the global minimum variance portfolios. The $*$'s in the last two columns indicate whether the related numbers are significantly smaller than one.[8] When returns have an exact three-factor structure and are i.i.d. multivariate Normal (Panel A), the in-sample risk accurately reflects the out-of-sample risk provided a three-factor model is used to estimate covariance matrices, as is to be expected. However, when the sample covariance matrix or the one-factor model based covariance matrix estimator is used to construct the minimum risk portfolios, the in-sample risk is substantially smaller than its out-of-sample counterpart. Neither the use of the standard degrees of freedom based correction nor the correction based on the Bayesian estimate leads to a sufficient reduction in the in-sample optimism. In all cases the jackknife type estimator of the out-of-sample risk is reasonably accurate.

In Section II we showed when returns are generated from an i.i.d. multivariate Normal distribution and the optimal portfolio is constructed from the sample covariance matrix with no portfolio weight constraints, the right correction is likely to be twice as large as the standard degrees of freedom based correction. In Panel A we see that the degrees-of-freedom correction in this case is $5.42 - 4.64 = 0.78$. If we add twice of this correction to the in-sample risk, we get $4.64 + 2 \times 0.78 = 6.20$, which is almost the out-of-sample standard deviation of 6.29.

When portfolio weights are constrained to be nonnegative, the out-of-sample risks of the minimum variance portfolios associated with the three covariance matrix estimators are about the same, as observed by Jagannathan and Ma (2003). However, the in-sample optimism comes down substantially with portfolio weight constraints. While the use of constraints imposes a penalty—the out of sample variance goes up—the investor is able to more accurately assess the out-of-sample risk.

Panel C of Table I gives the results when daily return data on stocks traded on NYSE, AMEX and NASDAQ are used. The sample covariance matrix outperforms the one- and three-factor models—the out-of-sample risk as well as the in-sample optimism are the least for the sample covariance matrix.[9] As is to be expected, with portfolio weight constraints the in-sample optimism comes down substantially. As is the case for simulated data, neither the standard degrees-of-freedom correction nor the use of the Bayesian estimator of the covariance matrix help much in reducing the in-sample optimism, while the jackknife estimator performs better. In fact, t-tests show that the average jackknife estimates are not significantly lower than the average realized out-of-sample standard deviations (see footnote 8). However, the fact that the numbers in the last column are all smaller than one indicates that, unlike with simulated data, the jackknife too exhibits some in-sample optimism, suggesting the non-i.i.d. nature. The jackknife also provides a better estimate than that obtained by adding twice the standard degrees of freedom based correction to the in-sample estiamte. in daily returns.

Panel B gives the results for tracking-error-variance minimization with simulated data. The patterns are similar to the variance minimization case in Panel A. The in-sample optimism is the highest for the sample covariance matrix. Adding twice the degree of freedom correction to the in-sample risk gives us $1.45 + 2 \times (1.69 - 1.45) = 1.93$, which is exactly the out-of-sample risk!Jackknife estimates are accurate. The three-factor-model based covariance matrix estimator performs the best in terms of both lowest out-of-sample risk and absence of no in-sample optimism. This is probably because on the one hand, the data is simulated using a three-factor model, hence there is no specification error when we use a three-factor model to estimate the covariance matrix of returns. On the other hand, the sampling error is also very low when estimating a factor model, because factor models have far less parameters than the sample covariance matrix. Unlike the results in Panel A, imposing portfolio weight constraints has little effect.

Panel D gives the results for tracking-error minimization with daily return data. The patterns are quite different from what we observed when simulated data were used. The sample covariance matrix provides the lowest out-of-sample risk as measured by the standard deviation of the tracking-error. However, it has the most in-sample optimism. The out-of-sample risk for the three-factor model is 1.32 (= 1.93/1.46) times that for the sample covariance matrix. Statistical tests (see footnote 8) show that the average out-of-sample risk for the factor models are significantly higher than that of the sample covariance matrix. Hence there is substantial benefit to using the sample covariance matrix if tracking-error minimization is the objective. This is true with or without portfolio weight constraints. Results in the last two columns indicate that the jackknife method has the lowest in-sample optimism.

The results in Panels C and D show that, with real return data, adding twice the degrees of freedom adjustment to the in-sample risk still underestimates the out-of-sample risk. For example, in Panel C, doing that will only give us $5.56 + 2 \times (6.49 - 5.56) = 7.42$, which is still smaller than the out-of-sample risk of 8.44. This is because the real return data are not i.i.d. multivariate Normal.

The results in Panels C and D suggest that the Sharpe and Fama-French one- and three-factor models do not capture the full covariance structure of the returns. Indeed, standard likelihood ratio tests using nonoverlapping three-year periods consistently reject the hypothesis that the residuals from the factor models have diagonal covariance matrix. Furthermore, over the sample period, for the residuals from the one-factor model, the sum of all the variances and covariances is 2.52 times as large as the sum of all the variances. This ratio is 1.95 for the residuals from the three-factor model, suggesting the inadequacy of the one- and three-factor model to capture the covariance structure of returns.

These results also suggest that although the jackknife type estimator (based on the assumption that returns are i.i.d.) performs rather surprisingly well compared to other estimators of out-of-sample risks, there is scope for improvement through relaxing the i.i.d. assumption.

While Table I compares the averages of various estimates of the out-of-sample standard deviation with the average realized out-of-sample standard deviation, Table II reports the mean absolute difference between the various estimates of the out-of-sample standard deviation and the realized out-of-sample standard deviation. The $*$ in the last three columns indicates the associated number is significantly larger than the corresponding number in the first column (see footnote 8 for test method). Similar to Table I, there are four panels in Table II: the first two panels are for the simulated return data, and the last two panels are for the real return data. Panel A indicates that the jackknife method often leads to significant improvement compared to the in-sample and other alternatives when there are no portfolio weight constraints. As is to be expected, when the covariance matrix is estimated using a three-factor model, there is little difference between the in-sample and jackknife estimates, since the simulated data indeed has an exact three-factor structure. Panel B, which reports the tracking-error variance minimization results using simulated data, has essentially the same pattern.

Panels C and D report the results for real return data. Again, the mean absolute difference is less for the jackknife estimator when compared to the in-sample estimator with and without the degrees-of-freedom correction, and the differences are often statistically significant.

Note that with real data (Panels C and D), the performance becomes significantly worse when compared to estimates made using simulated data (Panels A and B). The mean absolute difference between the various estimates and realized out-of-sample standard deviation are much larger. This suggests that relaxing the i.i.d. assumption may be important.

In Sections V.A and V.B, we proposed a multivariate GARCH-based method and a weighted-jackknife method to estimate the out-of-sample standard deviations of the optimal portfolios to allow for time variations in variances and covariances of returns. Tables III and IV report the performance of these methods. In using the weighted-jackknife, we choose the decay rate of the weights to be 0.01 per day and 0.21 per month (there are about 21 trading days in a month). So $\alpha = 0.21$ in eq. (24).[10]

The layout of Tables III and IV is similar to that of Tables I and II, except we only examine the results with real data. This is because our purpose here is to see how these estimators perform when the return data has covariance structure persistence. Therefore, the results when data are i.i.d. Normal are irrelevant.

Table III shows that the weighted-jackknife method out-performs the alternatives when the sample covariance matrix is used to construct optimal portfolios. In all other cases, the performance of the weighted-jackknife is about the same as the multivariate GARCH method. This suggests that the multivariate GARCH does a good job of capturing the temporal dependence in factor return variances and covariances. The limitation is that we cannot implement the multivariate GARCH procedure for the sample covariance matrix because the dimensions involved are too large making computations difficult. Since the use of the sample covariance matrix leads to lower tracking-error variance on average, the weighted-jackknife has advantages.

Table IV reports mean absolute differences between the various estimators and the realized out-of-sample standard deviation. As in Table II, the $*$ in the last two columns indicates the associated number is statistically larger than the corresponding number in the first column (see footnote 8 for test method). Under this criterion, the weighted-jackknife often gives significantly smaller mean absolute tracking error. Considering both the unconditional bias and the mean absolute difference, the weighted-jackknife method performs better than the other methods we considered.

VII. Conclusion

It has been observed in the literature that the in-sample variance of a minimum risk portfolio constructed from the sample moments of returns typically understates its true variance in the population, i.e., there is *in-sample optimism.* In this paper we show that this is what one should expect to find.

We show that the in-sample estimate of the variance of a global minimum-risk portfolio constructed using an estimated covariance matrix of returns will, on average, be strictly smaller than the variance of the corresponding population global minimum-risk portfolio. Scaling the in-sample estimate upward by a standard degrees-of-freedom related factor or using the Bayes covariance matrix estimator is inadequate. It only provides an unbiased estimate of the variance of the population global minimum-risk portfolio when returns are drawn from an i.i.d. multivariate Normal distribution, and the right amount of correction is likely to be twice as large for i.i.d. multivariate Normal case, and will need to be larger than that for real return data. Therefore, we suggest a jackknife-type estimator that is valid when returns are i.i.d, and a variation that may be better when returns exhibit volatility persistence. The jackknife type estimator has the advantage that it is widely applicable. For example, it can be readily applied to the cases when factor models or shrinkage covariance matrix estimators are used, and when portfolios are constructed subject to portfolio-weight constraints.

We empirically demonstrate the need to correct for in-sample optimism by considering an optimal portfolio of 200 stocks that has lowest tracking error when the S&P500 is the benchmark. We consider three covariance matrix estimators constructed using three years of historical daily return data: the sample covariance matrix, Sharpe's single-index model and the Fama and French (1993) three-factor model. We find that the in-sample estimate of the tracking-error standard deviation is only 64 percent of the out-of-sample standard deviation. The corresponding numbers are 75 percent when corrected for the degrees of freedom and 92 percent for the jackknife-type estimator.

The use of one- and three-factor models lead to increased-tracking error variance. The out-of-sample tracking-error for the three-factor model is 1.93 percent, which is significantly higher than the 1.46 percent associated with the sample covariance matrix. This suggests that there is nontrivial correlation among the residuals in the three-factor model that needs to be taken into account when the objective is to construct efficient portfolios. An alternative to increasing the number of factors to capture the correlation structure of returns would be to identify the block diagonal structure of the covariance matrix of the residuals in a three-factor model using cluster analysis as in Ahn, Conrad and Dittmar (2003).

The advantage of using factor models is that they have smaller in-sample optimism when compared to the sample covariance matrix. With the use of the jackknife-type estimator for assessing the out-of-sample tracking-error variance, this advantage of the three-factor model disappears.

The performance of the jackknife-type estimator worsens when evaluated using real data instead of data simulated from i.i.d. multivariate Normal distribution. We show that the use of Dynamic Conditional Correlation Model of Engle (2002) or a weighted-jackknife method to account for this persistence leads to substantial improvement.

Appendix A: The Relation Between In-sample and Out-of-sample Variance of Sample Mean-variance Efficient Portfolios

The goal of this appendix is to prove Lemma 1. The setup and notations are described from the beginning of Section II until Lemma 1. For brevity we do not repeat them.

When $w_p \neq w_s$ we obtain the following inequalities:

$$\begin{aligned}
\frac{1}{2}w_p'\Sigma w_p &= f_p(w_p) \\
&< f_p(w_s) \\
&= \frac{1}{2}w_s'\Sigma w_s - \delta_1^s(w_s'\mu - \nu) - (w_s - \underline{w})'\theta_1^s + (w_s - \bar{w})'\theta_2^s \\
&= \frac{1}{2}w_s'\Sigma w_s - \delta_1^s(w_s'\mu - \nu).
\end{aligned} \tag{A1}$$

The first (equality) follows from the definition of the Lagrangian function given in equation (4) and the observation that $(w_p - \underline{w})'\theta_1^p = 0$; $(\bar{w}_i - w_{p,i})'\theta_2^p = 0$; $(w_p{}'\mu - \nu) = 0$; and $(w_p{}'\mathbf{1} - 1) = 0$. The second (inequality) follows from the uniqueness of the minimization problem and the fact that $w_p \neq w_s$. The third (equality) follows from the definition of the function $f_p(w_s)$ and the observation that $w_s{}'\mathbf{1} = 1$. The last (equality) follows from the fact $(w_s, \delta_1^s, d_2^s, \theta_1^s, \theta_2^s)$ minimizes the Lagrangian function $f_s(w)$ given by equation (5); and therefore both $(w_s - \underline{w})'\theta_1^s$ and $(\bar{w} - w_s)'\theta_2^s$ are zero by the complementary slackness conditions.

By a similar logic, it follows that:

$$\begin{aligned}
\frac{1}{2}w_s'Sw_s &= f_s(w_s) \\
&< f_s(w_p) \\
&= \frac{1}{2}w_p'Sw_p - \delta_1^p(w_p'\hat{\mu} - \nu) - (w_p - \underline{w})'\theta_1^p + (w_p - \bar{w})'\theta_2^p \\
&= \frac{1}{2}w_p'Sw_p - \delta_1^p(w_p'\hat{\mu} - \nu).
\end{aligned} \tag{A2}$$

Since $Prob\{w_s = w_p\} < 1$, $E(S) = \Sigma$, and $E(\hat{\mu}) = \mu$, taking the expectation on both sides of

(A2) gives:

$$
\begin{aligned}
E(f_s(w_s)) < E(f_s(w_p)) &= \frac{1}{2}w_p'E(S)w_p - \delta_1^p(w_p'E(\hat{\mu}) - \nu)) \\
&= \frac{1}{2}w_p'\Sigma w_p - \delta_1^p(w_p'\mu) - \nu) \\
&= \frac{1}{2}w_p'\Sigma w_p \\
&= f_p(w_p). \qquad \text{(A3)}
\end{aligned}
$$

Similarly, by taking the expectation of both sides of (A1) we get:

$$
E(f_p(w_p)) < E(f_p(w_s)) = \frac{1}{2}E(w_s'\Sigma w_s) - E(\delta_1^s(w_s'\mu - \nu)).
$$

Note that the expected values of the in-sample and the out-of-sample variances of the efficient portfolio are $E(w_s'Sw_s)$ and $E(w_p'\Sigma w_p)$, respectively. The former is also equal to $2E(f_s(w_s))$ (see (A2)). We are now in a position to relate the two.

From (A2) we get, $\frac{1}{2}E(w_s'Sw_s) = E(f_s(w_s)) < E(f_s(w_p))$; from (A3) we get, $E(f_s(w_p)) = \frac{1}{2}w_p'\Sigma w_p f_p(w_p) = E(f_p(w_p))$; and from (A1) we get $E(f_p(w_p)) < E(f_p(w_s))$, which in turn equals $\frac{1}{2}E(w_s'\Sigma w_s) - E(\delta_1^s(w_s'\mu - \nu))$. By combining these inequalities and equalities and multiplying by 2 we get:

$$
E(w_s'Sw_s) < w_p'\Sigma w_p < E(w_s'\Sigma w_s) - 2E(\delta_1^s(w_s'\mu - \nu)). \qquad \text{(A4)}
$$

This proves Lemma 1.

Appendix B: Proof of Proposition 4

PROOF: It is well-known that the predictive distribution for R_{T+1} is multivariate Student t (Zellner (1971), p.235-236) given by:

$$
\begin{aligned}
& p(R_{T+1}|R_1, ..., R_T) \\
\propto \quad & \left[1 + \frac{T}{T+1}(R_{T+1} - \hat{\mu})[(T-1)S]^{-1}(R_{T+1} - \hat{\mu})'\right]^{-T/2} \\
\propto \quad & \left[(T-N) + (R_{T+1} - \hat{\mu})\left[\frac{(T+1)(T-1)}{T(T-N)}S\right]^{-1}(R_{T+1} - \hat{\mu})'\right]^{-(N+(T-N))/2}.
\end{aligned}
$$

This is a $t\left(\hat{\mu},\ [\frac{(T+1)(T-1)}{T(T-N)}S]^{-1},\ T-N,\ N\right)$ distribution, using Zellner's notation (Zellner (1971), p.383). The $\propto$ sign means positively proportional. So the posterior covariance matrix of R_{T+1} is:

$$
\text{var}(R_{T+1}) = \frac{T-N}{T-N-2}\ \frac{(T+1)(T-1)}{T(T-N)}\ S = \frac{(T+1)(T-1)}{T(T-N-2)}\ S.
$$

Since the posterior covariance matrix is proportional to S, the global minimum portfolio weights are the same as w_s, the portfolio weights constructed using the sample covariance matrix S. Therefore, the variance of the portfolio under the predictive distribution is:

$$
w_s' var(R_{T+1}) w_s = \frac{(T+1)(T-1)}{T(T-N-2)} w_s' S w_s = \frac{(T-1)(T+1)}{T(T-N-2)}\ \frac{1}{\mathbf{1}'S^{-1}\mathbf{1}} \tag{B1}
$$

This proves Proposition 4.

Appendix C: Bayes Estimator of the Covariance Matrix When Returns Follow a Factor Structure

In this appendix we derive the Bayes' estimator of the covariance matrix when it has an exact k factor structure. Let R_t, denote the $N \times 1$ vector of date t returns as before, and f_t, denote the $k \times 1$ vector of date t factors. We assume that returns and factors together are drawn from an i.i.d. multivariate Normal distribution and the manager has observations on a time series of T returns and factors.

Let the true mean return vector be μ_r, the true covariance matrix of returns be Σ_r. We use subscript f to denote the corresponding parameters of the vector of factors.

We assume that returns have the following factor structure:

$$R_t = \alpha + \beta f_t + U_t, \tag{C1}$$

where α is a $N \times 1$, vector, β is a $N \times k$, matrix, and U_t is $N \times 1$ vector of residuals that have a multivariate normal distribution with mean zero and a diagonal covariance matrix Ω.

Let R denote the $T \times N$ matrix of returns. Then equation (C1)can be written in matrix form as follows:

$$R = FB + U,$$

where

$$F = \begin{pmatrix} 1 & f_{11} & ... & f_{1k} \\ ... & ... & ... & ... \\ 1 & f_{T1} & ... & f_{Tk} \end{pmatrix}$$

$$B = \begin{pmatrix} \alpha' \\ \beta' \end{pmatrix}$$

and U is defined similarly.

As before, we assume that the Bayesian portfolio manager does not observe $\mu_f, \Sigma_f, \Omega, \alpha$, and β but believes that they are drawn from a prior distribution. From the posterior distribution for $\mu_f, \Sigma_f, \Omega, \alpha$, and β, the portfolio manager computes Σ_r, the covariance matrix of R_t under the predictive distribution. Our derivation below follows that of Wang (2000). But our prior specifications are somewhat different, as will be pointed out below.

Let the prior be the usual diffuse prior:

$$p(\mu_f, \Sigma_f, \alpha, \beta, \Omega) \propto |\Omega|^{-1}|\Sigma_f|^{-(k+1)/2}. \tag{C2}$$

Notice that since Ω is diagonal, $|\Omega|$ is the product of its diagonal elements. Therefore the prior for $|\Omega|$ is a product of the independent priors for its diagonal elements, with the ith one given by:

$$p(\Omega_{ii}) \propto \frac{1}{\Omega_{ii}},$$

which is the commonly used diffuse prior. This prior is equivalent to the following prior for the square roots of the diagonal elements:

$$p(\Omega_{ii}^{1/2}) \propto \frac{1}{\Omega_{ii}^{1/2}}.$$

In what follows we will adopt this latter parameterization for convenience.

Our prior specification (C2) differs from that of Wang's (2000) in two aspects: First, we assume Ω is diagonal while Wang does not. Second, we start with a diffuse prior for α, while Wang assumes that the investor believes the asset pricing model to a certain degree, i.e., the prior for α is Normal centered at the zero vector.

Given our prior, the posterior distributions are standard: $p(\Sigma_f|F)$ is inverted Wishart, $p(\mu_f|F, \Sigma_f)$ is Normal, $p(\Omega_{ii}^{1/2}|F, R)$ is inverted Gamma, and $p(B_i|F, R, \Omega_{ii})$ is Normal, as given below:

$$p(\Sigma_f|F) \propto |\Sigma_f|^{-(T+k)/2} \exp\left\{-\frac{T}{2}\mathrm{tr}(\hat{\Sigma}_f\Sigma_f^{-1})\right\} \tag{C3}$$

$$p(\mu_f|F, \Sigma_f) \propto |\Sigma_f|^{-1/2} \exp\left\{-\frac{T}{2}\mathrm{tr}(\mu_f - \hat{\mu}_f)(\mu_f - \hat{\mu}_f)'\Sigma_f^{-1}\right\} \tag{C4}$$

$$p(\Omega_{ii}^{1/2}|F, R) \propto \Omega_{ii}^{-(\nu+1)/2} \exp\left(-\frac{\nu S_{ii}}{2\Omega_{ii}}\right), \quad i = 1, 2, ..., N. \tag{C5}$$

$$p(B_i|F, R, \Omega_{ii}) \propto \Omega_{ii}^{-1/2} \exp\left(-\frac{(B_i - \hat{B}_i)'F'F(B_i - \hat{B}_i)}{2\Omega_{ii}}\right), \forall i. \tag{C6}$$

where $\hat{\Sigma}_f$ and $\hat{\mu}_f$ refer to the maximum likelihood estimates of the respective quantities, B_i is the ith column of B, $\hat{B}_i$ is the least squares estimate of B_i, S_{ii} is the least squares estimate of the residual variance in the factor model regression for the return on stock i, (i.e., sum of squared residuals divided by $T-k-1$) and $\nu = T-k-1$.

These distribution functions imply that the posterior distribution of (B, Ω) is independent of that of (μ_f, Σ_f), as pointed out by Wang (2000), with the following posterior moments:

$$E(\Sigma_f|F) = \frac{T}{T-k-2}\hat{\Sigma}_f \equiv \tilde{\Sigma}_f \tag{C7}$$

$$E(\mu_f|F) = \hat{\mu}_f \tag{C8}$$

$$Var(\mu_f|F) = \frac{1}{T-k-2}\hat{\Sigma}_f \tag{C9}$$

$$E(\Omega_{ii}|F,R) = \frac{\nu}{\nu-2}S_{ii} \equiv \tilde{\Omega}_{ii} \tag{C10}$$

$$E(B|F,R,\Omega) = \hat{B} \tag{C11}$$

$$Var(vec(B)) = \tilde{\Omega} \otimes (F'F)^{-1} \tag{C12}$$

Since the posterior distributions of (B, Ω) and (μ_f, Σ_f) are independent, the posterior mean of the return covariance matrix is

$$E(\Sigma_r|F,R) = E(\beta\Sigma_f\beta'|F,R) + \tilde{\Omega}.$$

From the law of iterated expectation, we have

$$E(\beta\Sigma_f\beta'|F,R) = \hat{\beta}\tilde{\Sigma}_f\hat{\beta}' + \mathrm{tr}(G\tilde{\Sigma}_f)\tilde{\Omega}$$

where G is the $k \times k$ submatrix in the lower-right corner of $(F'F)^{-1}$.

From the inverse of partitioned matrices, we know that $G = \frac{1}{T}(\hat{\Sigma}_f)^{-1}$, hence the last term in the above equation is given by $\frac{k}{T-k-2}\tilde{\Omega}$.

Hence the posterior mean of the covariance matrix of the returns is

$$E(\Sigma_r|F,R) = \hat{\beta}\tilde{\Sigma}_f\hat{\beta}' + (1 + \frac{k}{T-k-2})\tilde{\Omega} = \hat{\beta}\tilde{\Sigma}_f\hat{\beta}' + \frac{T-2}{T-k-2}\tilde{\Omega}. \tag{C13}$$

Given the posterior moments for the model parameters, we can calculate the covariance matrix of R_{T+1} under the predictive distribution. From the law of iterated expectations, we have

$$Var(R_{T+1}|R,F) = E(\Sigma_r|R,F) + Var(\mu_r|R,F). \tag{C14}$$

The first term on the RHS is given in (C13). The second term is

$$
\begin{aligned}
& Var(\alpha + b\mu_f | F, R) \\
= \; & Var(E[\alpha + \beta\mu_f | \mu_f, R, F] | R, F) + E[Var(\alpha + \beta\mu_f | \mu_f, R, F) | R, F] \\
= \; & \frac{1}{T-k-2}\hat{\beta}\hat{\Sigma}_f\hat{\beta}' + E[Var(\alpha + \beta\mu_f | \mu_f, R, F) | R, F] \qquad \text{(C15)}
\end{aligned}
$$

Since $\alpha + \beta\mu_f = (I_N \otimes (1 \ \mu_f'))vec(B)$, it follows from equation (C12) that

$$
\begin{aligned}
& Var(\alpha + \beta\mu_f | \mu_f, R, F) && \text{(C16)} \\
= \; & (I_N \otimes (1 \ \mu_f'))(\tilde{\Omega} \otimes (F'F)^{-1})(I_N \otimes (1 \ \mu_f'))' && \text{(C17)} \\
= \; & \tilde{\Omega} \otimes [(1 \ \mu_f')(F'F)^{-1}(1 \ \mu_f')'] && \text{(C18)} \\
= \; & \rho\tilde{\Omega} && \text{(C19)}
\end{aligned}
$$

where $\rho = (1 \ \mu_f')(F'F)^{-1}(1 \ \mu_f')'$. The posterior mean of ρ is,

$$
E(\rho | F, R) = \mathrm{tr}\left\{ (F'F)^{-1} \begin{pmatrix} 1 & \hat{\mu}_f' \\ \hat{\mu}_f & (T-k-2)^{-1}\hat{\Sigma}_f + \hat{\mu}_f\hat{\mu}_f' \end{pmatrix} \right\}.
$$

Notice that,

$$
F'F = T \begin{pmatrix} 1 & \hat{\mu}_f' \\ \hat{\mu}_f & \hat{\Sigma}_f \end{pmatrix},
$$

and

$$
\begin{pmatrix} 1 & \hat{\mu}_f' \\ \hat{\mu}_f & (T-k-2)^{-1}\hat{\Sigma}_f + \hat{\mu}_f\hat{\mu}_f' \end{pmatrix} = \begin{pmatrix} 1 & \hat{\mu}_f' \\ \hat{\mu}_f & \hat{\Sigma}_f \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & -\frac{(T-k-3)}{(T-k-2)}\hat{\Sigma}_f + \hat{\mu}_f\hat{\mu}_f' \end{pmatrix}.
$$

Hence $E(\rho | F, R)$ is the sum of two terms. The first term is $\frac{k+1}{T}$. To calculate the second term, we need to get the $k \times k$ lower-right submatrix of $(F'F/T)^{-1}$. Using the inverse of partitioned matrix, this submstrix is

$$
(\hat{\Sigma}_f - \hat{\mu}_f\hat{\mu}_f')^{-1} = \hat{\Sigma}_f^{-1} + \frac{1}{1 - \hat{\mu}_f'\hat{\Sigma}_f^{-1}\hat{\mu}_f}\hat{\Sigma}_f^{-1}\hat{\mu}_f\hat{\mu}_f'\hat{\Sigma}_f^{-1}.
$$

It can be verified that the second term in the *trace* is therefore,

$$
-\frac{k(T-k-3)}{T(T-k-2)} + \frac{1}{T(T-k-2)}\frac{\hat{\mu}_f'\hat{\Sigma}_f^{-1}\hat{\mu}_f}{1 - \hat{\mu}_f'\hat{\Sigma}_f^{-1}\hat{\mu}_f} \qquad \text{(C20)}
$$

Combining (C13), (C14),(C15), (C19), and (C20), and adding the term $\frac{k+1}{T}$ from above, we find that the posterior variance of R_{T+1} is given by:

$$\begin{aligned} & Var(R_{T+1}|R,F) \\ = \; & \frac{T+1}{T-k-2}\hat{\beta}\hat{\Sigma}_f\hat{\beta}' + \frac{(T+1)(T-2)}{T(T-k-2)}\tilde{\Omega} \\ & + \frac{1}{T(T-k-2)}\frac{\hat{\mu}_f'\hat{\Sigma}_f^{-1}\hat{\mu}_f}{1-\hat{\mu}_f'\hat{\Sigma}_f^{-1}\hat{\mu}_f}\tilde{\Omega}. \end{aligned} \tag{C21}$$

As observed by Wang (2000), the term $\hat{\mu}_f'\hat{\Sigma}_f^{-1}\hat{\mu}_f$ is the square of the highest Sharpe ratio of the frontier spanned by the sample mean $\hat{\mu}_f$ and variance $\hat{\Sigma}_f$. For the U.S. data, it is definitely less than 0.5. So the last term in (C21) is less than $\frac{1}{T(T-k-2)}\tilde{\Omega}$, which is less than $0.0003\tilde{\Omega}$ if $T \geq 60$ and $k \leq 3$. Hence this term can be ignored for practical purposes.

The portfolio manager's problem is to choose portfolio weights, w, to

$$\min_w var(w'R_{T+1}).$$

given the predictive covariance matrix of R_{T+1} in (C21).

We can compare Bayes' estimator of the covariance matrix given in equation (C21) with the sample estimate of the covariance matrix under factor model structure. The sample estimate is given by:

$$S_r = \hat{\beta}S_f\hat{\beta}' + S. \tag{C22}$$

The term S_f, the sample covariance matrix of the factors, differs from the first term of $Var(R_{T+1})$ by a the scale factor, $\frac{(T+1)(T-1)}{T(T-k-2)}$. When $k \leq 3$ and $T \geq 60$, this scale factor is less than 1.091. The term, S, the least-squares estimate of the (diagonal) residual covariance matrix (i.e., the sum of squared-residuals divided by $T-k-1$), differs from the second term of (C21) by the scale factor, $\frac{(T+1)(T-2)}{T(T-k-2)}\frac{T-k-1}{T-k-3}$, which is less than 1.11 if $k \leq 3$ and $T \geq 60$. Therefore, it appears that the difference may be small whether we use Bayes' estimator or the sample covariance matrix estimator for factor models.

Also, since the first and second terms of the RHS of the above two equations differ by different factors, the portfolio weights constructed from the sample covariance matrix will differ from the one constructed using Bayes' covariance matrix estimator. However, since the factors are very similar, we expect the difference in the portfolio weights to be small.

Under Bayes' estimator, the unconstrained global minimum variance portfolio will have a variance of $(\mathbf{1}'[Var(R_{T+1}|F,R)]^{-1}\mathbf{1})^{-1}$.

Appendix D: The DCC Multivariate GARCH Model

The Dynamic Conditional Correlation (DCC) multivariate GARCH model is proposed by Engle (2002) and Tse and Tsui (2002). Let R_t be the vector of time t returns on k assets, and H_t be the conditional covariance matrix of r_t. The model describes the evolution of H_t over time.

The covariance matrix H_t is decomposed into variances and correlations:

$$H_t = D_t P_t D_t,$$

where D_t is the diagonal matrix of standard deviations of R_t, and P_t is the correlation matrix.

Each diagonal element of D_t is assumed to follow the well-known univariate GARCH process. So different elements of D_t will have different persistences and unconditional means. The dynamic correlation structure is

$$\begin{aligned} Q_t &= (1 - \sum_{i=1}^{I} \alpha_i - \sum_{j=1}^{J} \beta_j)\bar{Q} + \sum_{i=1}^{I} \alpha_i (\epsilon_{t-i}\epsilon'_{t-i}) + \sum_{j=1}^{J} \beta_j Q_{t-j} & \text{(D1)} \\ P_t &= Q_t^{*-1} Q_t Q_t^{*-1}. & \text{(D2)} \end{aligned}$$

Here ϵ_t is the vector of standardized residuals, $\bar{Q}$ is the unconditional covariance of the standardized residuals, and

$$Q_t^* = \begin{bmatrix} \sqrt{q_{11}} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{q_{22}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{q_{kk}} \end{bmatrix},$$

so the normalization in (D2) guarantees that P_t is a matrix of correlations.

Usually setting the lags I and J to one is a good choice.

REFERENCES

Ahn, Dong-Hyun, Jennifer Conrad, and Robert F. Dittmar, 2003, Basis assets, Working paper, Kenan-Flagler Business School, University of North Carolina.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2003, Modeling and forecasting realized volatility, *Econometrica* 71, 579-625.

Chan, Louis K. C., Jason Karceski, and Josef Lakonishok, 1999, On portfolio optimization: Forecasting covariances and choosing the risk model, *Review of Financial Studies* 12, 937-974.

Efron, Bradley, 1982, *The Jackknife, the Bootstrap, and Other Resampling Plans,* Society for Industrial and Applied Mathematics, Philadelphia, PA.

Elton, Edwin J., and Martin J. Gruber, 1973, Estimating the dependence structure of share prices—Implications for portfolio selection, *Journal of Finance* 28, 1203-1232.

Engle, Robert F., 2002, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics* 20, 339-350.

Fama, Eugene, and Kenneth French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Flemming, Jeff, Cris Kirby, and Barbara Ostdiek, 2003, The economic value of volatility timing using "realized" volatility, *Journal of Financial Economics* 67, 473-509.

Foster, D. P., and D. B. Nelson, 1996, Continuous record asymptotics for rolling sample variance estimators, *Econometrica* 64, 139-174.

Foster, F. D., Tom Smith, and Robert Whaley, 1997, Assessing goodness-of-fit of asset pricing models: The distribution of the maximal R^2, *Journal of Finance* 52, 591-607.

Frost, Peter A., and James E. Savarino, 1988, For better performance: Constrain portfolio weights, *Journal of Portfolio Management* 15, 29-34.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov, 2003, Predicting volatility: Getting the most out of return data sampled at different frequencies, Working paper, University of North Carolina and UCLA.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov, 2004a, There is a risk-return trade-off after all, Working paper, University of North Carolina and UCLA.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov, 2004b, The MIDAS touch: Mixed data sampling regression models, Working paper, University of North Carolina and UCLA.

Gibbons, Michael, 1982, Multivariate test of financial models: A new approach, *Journal of Financial Economics* 10, 3-27.

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

Hall, P. and Jing, B. Y., 1996, On sample reuse methods for dependent data, *Journal of Royal Statistical Society,* Series B, 58, 727-737.

Hall, P., Jing, B. Y. and Lahiri, S. N., 1998, On the sampling window method for long-range dependent data, *Statistica Sinica*, 8, 4, 1189-1204.

Jacquier, Eric, Alex Kane, and Alan Mrcus, 2002, Optimal forecasts of long term returns and asset allocation: Geometric, arithmetic, or other means? Working paper, Boston College.

Jagannathan, Ravi, and Tongshu Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraint helps, *Journal of Finance* 58, 1651-1683.

Jobson, J. D., and Bob Korkie, 1981, Putting Markowitz theory to work, *Journal of Portfolio Management,* 70-74.

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

Jorion, Philippe, 1986, Bayes-Stein estimation for portfolio analysis, *Journal of Financial and Quantitative Analysis* 21, 279-292.

Kandel, Shmuel, 1984, The likelihood ratio test of mean-variance efficiency without a riskless asset, *Journal of Financial Economics* 13, 575-592.

Künsch, Hans R. 1989, The jackknife and the bootstrap for general stationary distributions, *Annals of Statistics* 17, 1217-1241.

Lahiri, S. N. 2002, On the jackknife-after-bootstrap method for dependent data and consistency properties, *Econometric Theory* 18, 79-98.

Ledoit, Olivier, and Michael Wolf, 2003a, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of Empirical Finance* 10, 603-621.

Ledoit, Olivier, and Michael Wolf, 2003b, Honey, I shrunk the sample covariance matrix, Working paper, Credit Suisse First Boston and Universitat Pompeu Fabra.

Liu, Qianqiu, 2003, On portfolio optimization: How do we benefit from high-frequency data? Working paper, Kellogg Graduate School of Management, Northwestern University.

Lütkepohl, Helmut, 1993, *Introduction to multiple time series analysis* (2nd ed.), Springer-Verlag, New York.

MacKinlay, A. Craig, 1987, On multivariate tests of the CAPM, *Journal of Financial Economics* 18, 341-372.

Merton, Robert C., 1980, On estimating the expected return on the market: An exploratory investigation *Journal of Financial Economics* 8, 323-361.

Michaud, Richard O., 1989, The Markowitz optimization enginma: Is optimized optimal? *Financial Analysts Journal* 45, 31-42.

Michaud, Richard O., 1998, *Efficient Asset Management,* Harvard Business School Press, Boston, MA.

Muirhead, Robb J., 1982, *Aspects of Multivariate Statistical Theory,* John Wiley and Sons, New York.

Roll, Richard, 1985, A note on the geometry of Shanken's CSR T^2 test for mean/variance efficiency, *Journal of Financial Economics* 14, 349-357.

Sharpe, William, 1963, A simplified model for portfolio analysis, *Management Science* 9, 277-293.

Schwert, G. Williams, and Paul L. Seguin, 1990, Heteroskedasticity in Stock Returns, *Journal of Finance* 45, 1129-1155.

Shanken, Jay, 1985, Multivariate tests of the zero-beta CAPM, *Journal of Financial Economics* 14, 327-348.

Stambaugh, Robert, 1982, On the exclusion of assets from tests of the two parameter model, *Journal of Financial Economics* 10, 235-268.

Tse, Y. K., and Albert K. C. Tsui, 2002, A multivariate generalized autoregressive conditional heterskedasticity model with time-varying correlations, *Journal of Business and Economic Statistics* 20, 351-362.

Wang, Zhenyu, 2000, A shrinkage theory of asset allocation using asset-pricing models, Working paper, Columbia University.

Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics,* John Wiley and Sons, New York.

Footnotes

[1] See Jobson and Korkie (1981).

[2] The notable papers on the topic include Jobson and Korkie (1982); Gibbons (1982); Stambaugh (1982); Kandel (1984); Shanken (1985); Roll (1985); MacKinlay (1987); and Gibbons, Ross, and Shanken (1989).

[3] See MacKinlay (1987) and Gibbons, Ross and Shanken (1989).

[4] Some theoretical support for such an approach can be found in Foster and Nelson (1996), Flemming et al. (2003), and Liu (2003) and the mixed data sampling approach based estimator in Ghysels et al. (2003, 2004a, 2004b).

[5] Choosing the optimal number of lags for a VAR is described in Lütkepohl (1993). The AIC and BIC criteria are: $\ln|\tilde{\Sigma}| + 2/T \cdot k$ and $\ln|\tilde{\Sigma}| + \ln(T)/T \cdot k$, where T is the number of observations, $\tilde{\Sigma}$ is the estimated covariance matrix of the error term, k is the number of freely estimated parameters.

[6] We also formed a third global minimum variance portfolio, which also faces the restriction that no more than five percent (i.e., 10 times of the equal weight) of the investment can be in any one stock. But the results for this portfolio are very similar to the second one and hence are omitted.

[7] For convenience of interpretation, we work with standard deviations instead of variances in the sequel.

[8] In this section, we often report that the average of $x_1, x_2, ..., x_N$ is significantly smaller than the average of $y_1, y_2, ..., y_N$, where both sequences are usually measures of standard deviations. For example, the x's may be in-sample standard errors, and the y's may be out-of-sample standard deviations. N will be 396 since the various estimators are evaluated over 396 months. To test this hypothesis, we test whether the mean of the sequence $\log(x_1/y_1), ..., \log(x_N/y_N)$ is significantly smaller than zero using a one-tailed test. We use Newey-West standard deviations

with 6 and 12 lags for this test. Unless indicated, the use of 6 or 12 lags does not change the conclusion.

[9] Standard t test shows that the average out-of-sample standard deviation of the unconstrained optimal portfolio from the one-factor model is significantly higher than that from the sample covariance matrix. Or loosely speaking, the difference between the two numbers in the table, 9.16 and 8.44, is significant. See footnote 8 for the method of test used.

[10] When using the optimal rolling sample variance estimators, people generally uses decay rate between 0.01 per day to 0.05 per day (see Flemming et al. (2003), Liu (2003), Ghysels et al. (2003, 2004a, 2004b)). We feel numbers in the higher end of the range may be too high for our purpose. For example, a daily decay rate of 0.04 would imply a monthly decay rate of 0.84, meaning the second-to-the-last month is only given a weight that is 16% of the last month in (24).

Table I

Average standard deviations

Each month 200 stocks with the largest market capitalization are selected. Thirty-six months of daily data are used to estimate the covariance matrices and form the global minimum-variance portfolios or minimum-tracking-error portfolios. The portfolios are rebalanced every month. Results are based on 396 months of out-of-sample returns from 1967/5 to 2000/4. Standard deviations are annualized and expressed in percentage points. The degrees of freedom adjustment factors are computed using the formula in Section II. The Bayesian covariance matrix estimators are described in Section III and Appendix C. Panels A and B are for simulated i.i.d. normal data (see Section VI.C). Panels C and D are for the real return data. The letter C after each covariance matrix estimator denotes the optimal portfolio is constructed subject to nonnegativity constraint. The $*$ in the last two columns indicates the associated number is significantly smaller than one.

	In-sample			Jackknife	Out-of-sample	Normalized Std.Dev.	
		DF adjusted	Bayesian			In/Out	Jackknife/Out
Panel A. Global minimum variance portfolio: Simulated data							
Sample covariance matrix	4.64	5.42	5.42	6.26	6.29	0.74*	1.00
Sample covariance matrix C	7.78			8.03	8.00	0.97	1.00
One-factor model	4.70		4.70	6.66	6.65	0.71*	1.00
One-factor model C	7.32			8.05	8.01	0.91*	1.00
Three-factor model	5.23		5.25	5.52	5.55	0.94	0.99
Three-factor model C	7.82			7.99	7.95	0.98	1.01
Equally-weighted portfolio	13.93				13.68	1.02	

Table I—continued

	In-sample			Jackknife	Out-of-sample	Normalized Std.Dev.	
		DF adjusted	Bayesian			In/Out	Jackknife/Out
Panel B. Minimum tracking error variance portfolio: Simulated data							
Sample covariance matrix	1.45	1.69	1.69	1.96	1.93	0.75*	1.02
Sample covariance matrix C	1.46			1.90	1.88	0.78*	1.01
One-factor model	1.65		1.66	1.81	1.78	0.93*	1.02
One-factor model C	1.65			1.81	1.78	0.93*	1.02
Three-factor model	1.68		1.68	1.68	1.65	1.02	1.02
Three-factor model C	1.68			1.68	1.65	1.02	1.02
Equally-weighted portfolio	2.19				2.11	1.04	

Table I—continued

	In-sample			Jackknife	Out-of-sample	Normalized Std.Dev.	
		DF adjusted	Bayesian			In/Out	Jackknife/Out
Panel C. Global minimum variance portfolio: Real data							
Sample covariance matrix	5.56	6.49	6.50	7.76	8.44	0.66*	0.92
Sample covariance matrix C	8.07			7.99	8.67	0.93	0.92
One-factor model	4.48		4.49	8.42	9.16	0.49*	0.92
One-factor model C	7.16			8.16	8.91	0.80*	0.92
Three-factor model	4.97		4.99	7.77	8.59	0.58*	0.90
Three-factor model C	7.52			8.02	8.77	0.86*	0.91
Equally-weighted portfolio	13.80				13.07	1.06	

Table I—continued

	In-sample			Jackknife	Out-of-sample	Normalized Std.Dev.	
		DF adjusted	Bayesian			In/Out	Jackknife/Out
Panel D. Minimum tracking error variance portfolio: Real data							
Sample covariance matrix	0.94	1.10	1.10	1.30	1.46	0.64*	0.89*
Sample covariance matrix C	0.95			1.26	1.42	0.67*	0.89*
One-factor model	1.61		1.61	2.15	2.24	0.72*	0.96
One-factor model C	1.61			2.15	2.24	0.72*	0.96
Three-factor model	1.64		1.64	1.86	1.93	0.85*	0.96
Three-factor model C	1.64			1.86	1.93	0.85*	0.96
Equally-weighted portfolio	2.80				2.54	1.10	

Table II

Mean absolute difference from out-of-sample standard deviations

Each month 200 stocks with the largest market capitalization are selected. Thirty-six months of daily data are used to estimate the covariance matrices and form the global minimum variance portfolios or minimum tracking error portfolios. The portfolios are rebalanced every month. Results are based on 396 months of out-of-sample returns from 1967/5 to 2000/4. Standard deviations are annualized and expressed in percentage points. The degrees of freedom adjustment factors are computed using the formula in Section 2.1. The Bayesian covariance matrix estimators are described in Section III and Appendix C. Panels A and B are for simulated i.i.d. normal data (see Section VI.C). Panels C and D are for the real return data. The letter C after each covariance matrix estimator denotes the optimal portfolio is constructed subject to nonnegativity constraint. The $*$ in the last three columns indicates the related number is significantly larger than the corresponding number in the first column.

	Jackknife	In-sample	DF adjusted	Bayesian
Panel A. Global minimum variance portfolio: Simulated data				
Sample covariance matrix	0.851	1.669*	1.076*	1.071*
Sample covariance matrix C	1.034	1.001		
One-factor model	0.879	2.016*		2.010*
One-factor model C	1.088	1.262*		
Three-factor model	0.703	0.730		0.726
Three-factor model C	1.038	1.004		
Equally-weighted portfolio		1.687		
Panel B. Minimum tracking error variance portfolio: Simulated data				
Sample covariance matrix	0.263	0.490*	0.314*	0.313*
Sample covariance matrix C	0.258	0.432*		
One-factor model	0.255	0.270*		0.269*
One-factor model C	0.255	0.270*		
Three-factor model	0.235	0.224		0.225
Three-factor model C	0.236	0.224		
Equally-weighted portfolio		0.291		

Table II—continued

	Jackknife	In-sample	DF adjusted	Bayesian
Panel C. Global minimum variance portfolio: Real data				
Sample covariance matrix	2.29	3.06*	2.53	2.53
Sample covariance matrix C	2.61	2.73		
One-factor model	2.81	4.76*		4.76*
One-factor model C	2.87	3.22		
Three-factor model	2.64	3.79*		3.78*
Three-factor model C	2.78	2.95		
Equally-weighted portfolio		4.50		
Panel D. Minimum tracking error variance portfolio: Real data				
Sample covariance matrix	0.352	0.533*	0.421	0.420
Sample covariance matrix C	0.346	0.487*		
One-factor model	0.635	0.757		0.756
One-factor model C	0.635	0.757		
Three-factor model	0.536	0.550		0.550
Three-factor model C	0.536	0.550		
Equally-weighted portfolio		0.740		

Table III

Average standard deviations when real return data are used: Weighted-jackknife and dynamic factor models

Each month 200 stocks with the largest market capitalization are selected. Thirty-six months of daily data are used to estimate the covariance matrices and form the global minimum variance portfolios. The portfolios are rebalanced every month. Results are based on 396 months of out-of-sample returns from 1967/5 to 2000/4. The letter C after each covariance matrix estimator denotes the optimal portfolio is constructed subject to nonnegativity constraint. The $*$ in the last two columns indicates the associated number is significantly smaller than one.

	Estimated			Out-of-sample	Normalized estimated		
	Weighted Jackknife	One-factor model	Three factor model		Weighted Jackknife	One-factor model	Three factor model
Panel A. Global minimum variance portfolio							
Sample covariance matrix	8.09	5.36	5.44	8.44	0.96	0.64*	0.64*
Sample covariance matrix C	8.33	7.22	7.63	8.67	0.96	0.83*	0.88*
One-factor model	8.77	8.61	8.66	9.16	0.96	0.94	0.95
One-factor model C	8.55	7.86	8.25	8.91	0.96	0.88	0.93
Three-factor model	8.18	7.73	7.78	8.59	0.95	0.90	0.91
Three-factor model C	8.42	7.58	7.96	8.77	0.96	0.86	0.91

Table III—continued

	Estimated			Out-of-sample	Normalized estimated		
	Weighted Jackknife	One-factor model	Three factor model		Weighted Jackknife	One-factor model	Three factor model
Panel B. Minimum tracking error variance portfolio							
Sample covariance matrix	1.35	0.93	0.93	1.46	0.92*	0.64*	0.64*
Sample covariance matrix C	1.31	0.94	0.94	1.42	0.92*	0.66*	0.66*
One-factor model	2.15	2.19	2.20	2.24	0.96	0.98	0.98
One-factor model C	2.15	2.19	2.20	2.24	0.96	0.98	0.98
Three-factor model	1.84	1.88	1.89	1.93	0.95	0.97	0.98
Three-factor model C	1.84	1.88	1.89	1.93	0.95	0.97	0.98

Table IV

Mean absolute difference from out-of-sample standard deviations when real return data are used

Each month 200 stocks with the largest market capitalization are selected. Thirty-six months of daily data are used to estimate the covariance matrices and form the global minimum variance portfolios. The portfolios are rebalanced every month. Results are based on 396 months of out-of-sample returns from 1967/5 to 2000/4. Standard deviations are annualized and expressed in percentage points. The letter C after each covariance matrix estimator denotes the optimal portfolio is constructed subject to nonnegativity constraint. The $*$ in the last two columns indicates the related number is significantly larger than the corresponding number in the first column.

	Weighted Jackknife	One-factor model	Three-factor model
Panel A. Global minimum variance portfolio			
Sample covariance matrix	2.07	3.20*	3.12*
Sample covariance matrix C	2.30	2.49	2.53
One-factor model	2.51	2.83	2.74
One-factor model C	2.50	2.72	2.80
Three-factor model	2.29	2.64*	2.62
Three-factor model C	2.42	2.64*	2.66
Panel B. Minimum tracking error variance portfolio			
Sample covariance matrix	0.306	0.543*	0.548*
Sample covariance matrix C	0.302	0.497*	0.500*
One-factor model	0.488	0.639*	0.614*
One-factor model C	0.488	0.639*	0.614*
Three-factor model	0.427	0.550*	0.542*
Three-factor model C	0.426	0.550*	0.542*