

NBER WORKING PAPER SERIES

A JACKKNIFE ESTIMATOR FOR TRACKING ERROR VARIANCE OF OPTIMAL  
PORTFOLIOS CONSTRUCTED USING ESTIMATED INPUTS<sup>1</sup>

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 Torben Andersen, Wayne Ferson, J. K. Ghosh, Eric Jaquier, Mark Kamstra, Raymond Kan, Nour Meddahi, Richard Roll, Pedro Santa-Clara, Rossen Valcanov, and seminar participants at CIRANO, Harvard Business School, Norwegian School of Management BI, NHH: Norwegian School of Economics and Business Administration, Northwestern University, University of California at Los Angeles, and the University of North Carolina for helpful comments and suggestions. We alone are responsible for any errors and omissions.

© 2004 by Gopal K. Basak, Ravi Jagannathan, and Tongshu Ma. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

A Jackknife Estimator for Tracking Error Variance of Optimal Portfolios Constructed Using  
Estimated Inputs I

Gopal K. Basak, Ravi Jagannathan, and Tongshu Ma

NBER Working Paper No. 10447

April 2004, Revised August 2007

JEL No. G11,G12

**ABSTRACT**

We develop a jackknife estimator for the conditional variance of a minimum-tracking- error-variance portfolio constructed using estimated covariances. We empirically evaluate the performance of our estimator using an optimal portfolio of 200 stocks that has the lowest tracking error with respect to the S&P500 benchmark when three years of daily return data are used for estimating covariances. We find that our jackknife estimator provides more precise estimates and suffers less from in-sample optimism when compared to conventional estimators.

Gopal K. Basak  
Indian Statistical Institute  
Kolkata 700108  
India  
gkb@isical.ac.in

Tongshu Ma  
SUNY-Binghamton  
tma@binghamton.edu

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

# 1 Introduction

It has been observed in the literature that the in-sample estimate of the variance of an efficient portfolio constructed using estimated inputs significantly understates the portfolio's true (*out of sample*) variance.<sup>1</sup> This downward bias, commonly referred to as *in-sample optimism*, increases with the number of assets used to construct mean-variance efficient portfolios. We demonstrate that in-sample optimism can be significantly large in certain cases, identify why that happens, and develop a jackknife-type estimator that is not subject to that bias. Since our focus is in assessing the risk associated with an efficient portfolio constructed using estimated inputs, we examine the global minimum variance and minimum benchmark tracking-error variance portfolios. We refer to such portfolios as *sample minimum risk portfolios (SMRPs)* in this paper.

It might be argued that by scaling up the in-sample estimator of the variance by the standard degree-of-freedom related correction factor, we may be able to remove in-sample optimism. When returns over time are drawn from an i.i.d. multivariate Normal distribution, and there are no portfolio weight constraints, we show that scaling up the in-sample variance by the standard degrees-of-freedom related factor provides an unbiased estimate of the variance of the population global minimum risk portfolio. Since the out-of-sample variance of the sample minimum risk portfolio is strictly larger, this procedure does not adequately correct for in-sample optimism. The intuition for this is that the optimization procedure costs additional degrees of freedom.

The example in Table 1, 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 returns, 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, the classical degrees-of-freedom based scaling factor is  $(\frac{T-1}{T-N})$ , which is only 1.36 ( $= \frac{750-1}{750-200}$ ). ( $N$  is the number of assets.) Even though using a three-factor model to estimate the covariance matrix

---

<sup>1</sup>See Jobson and Korkie (1981), Frost and Savarino (1988), and Michaud (1989).

and then construct the optimal portfolio reduces in-sample optimism, it still remains large. The in-sample estimate is 1.64 percent whereas the corresponding out-of-sample standard deviation is 1.93 percent. The ratio of the out-of-sample to in-sample variances is 1.38.

In fact, when returns are i.i.d. multivariate Normal and the global minimum risk portfolio is constructed without portfolio weight constraints, we show that it is necessary to add approximately twice the amount of the standard degree-of-freedom correction to the in-sample risk to arrive at a good estimate of what the out-of-sample risk on average would be. For example, in the first row of Table 1, Panel A, the degrees-of-freedom correction is  $5.42^2 - 4.64^2 = 7.85$ . Adding twice this amount to the in-sample variance gives us  $4.64^2 + 2 \times 7.85 = 37.22$ , i.e., a standard deviation of 6.10 percent, not much different from the realized out-of-sample standard deviation of 6.29 percent.

*In Sample Optimism with the Bayesian Covariance Matrix Estimator*

It might appear that we can avoid in-sample optimism through the use of a Bayesian approach since it would take into account the uncertainty associated with estimates of means and covariances of returns used in the optimization procedure. We find that, with standard diffuse priors commonly used in empirical works, the variance of a minimum risk portfolio computed using the predictive distribution is on average substantially below its true variance.

This should come as no surprise since with diffuse priors, the covariance matrix of returns under the predictive distribution is the sample covariance matrix multiplied by the factor  $\frac{T+1}{T-N-2}$  (see Section 3). 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 tracking error variance portfolio under the predictive distribution is  $\frac{T+1}{T-N-2}$  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 about the same as the standard degree-of-freedom correction,  $\frac{T-1}{T-N}$ , mentioned earlier, for the sample covariance matrix under the Classical method. Hence the variance of a minimum tracking error variance portfolio computed using the predictive distribution with standard diffuse priors will be almost the same as the standard degrees-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.<sup>2</sup> These tests examine 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.

To see why, consider forming the global minimum variance portfolio based on  $S$ , an estimate of the unknown covariance matrix,  $\Sigma$ . The vector of portfolio weights is given by:

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

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

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

Since  $\Sigma$  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 a given benchmark portfolio and a particular sample efficient portfolio — also involves the unknown covariance matrix of returns,  $\Sigma$ . However, when the unknown  $\Sigma$  is replaced 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  $\Sigma$  with its estimate  $S$  we just get the in-sample variance,

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

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

---

<sup>2</sup>The list of papers include Jobson and Korkie (1982); Gibbons (1982); Stambaugh (1982); Kandel (1984); Shanken (1985); Roll (1985); MacKinlay (1987); and Gibbons, Ross, and Shanken (1989).

An investor would be interested in the variance of the sample minimum risk portfolio she constructed, given by its conditional variance,  $w_s' \Sigma w_s$ . We use the term *conditional* to emphasize the fact that the elements of the portfolio weight vector,  $w_s$ , depend on the estimated covariance matrix used in forming the portfolio – therefore the variance of the portfolio will depend on the history of returns used to form that covariance matrix even when returns are i.i.d. over time. When the investor follows the same rule for forming minimum risk portfolios over time, the expected value of the conditional variance,  $E(w_s' \Sigma w_s)$ , is well defined and gives an estimate of what the conditional variance would be when averaged across all possible realizations for the history of returns used to construct the covariance matrix. Although the investor would care only about the conditional out-of-sample variance, we first analyze what that variance on average would be under the i.i.d. Normal case in order to get an insight into the issues involved. We then go on to develop a jackknife-type estimator that provides a consistent estimate of the conditional variance when returns are i.i.d. but not necessarily multivariate Normal.

When returns are not i.i.d. but instead exhibit persistence in their second moments, as in the data, we suggest two approaches.<sup>3</sup> The first approach is based on the Dynamic Conditional Correlation Model of Engle (2002). We find that empirically 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 further work is needed to model what a dynamic three factor model misses in the covariance structure of asset returns. In the second approach, we develop a modification of the jackknife-type estimator that weights more recent observations more heavily than those in the distant past.

Kan and Smith (2005) derive the exact distribution of the sample mean-variance efficient frontier of returns when there are no portfolio weight constraints and returns are drawn from an i.i.d multivariate Normal distribution. In contrast our focus is limited to global minimum variance

---

<sup>3</sup>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).

and global minimum tracking error portfolios. However, we do not require returns to be normally distributed or independent over time, and allow for the presence of portfolio weight constraints that are commonly encountered in practice. We focus on the global minimum variance portfolio and minimum tracking error portfolio for two reasons. First, minimizing tracking error variance is widely used for fund managers. Second, it is well-known in the literature that sample mean is a very poor estimate of the corresponding population mean return for portfolio optimization purposes.

In the next section, we develop some theoretical results that help understand why in-sample optimism occurs. We demonstrate in Section 3 that in-sample optimism would continue to exist even when the Bayesian approach is used to construct efficient portfolios. We develop a jackknife-type estimator of the conditional out-of-sample variance in Section 4 that is valid when returns are i.i.d. over time. In Section 5 we propose two methods for estimating the conditional out-of-sample variance when returns do not satisfy the i.i.d. assumption and empirically evaluate these methods in Section 6, and conclude in section 7. Technical derivations are provided in the Appendix.

## 2 Relation Between In-sample and Out-of-sample Variances of Sample Minimum Risk Portfolios

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;  $\Sigma = \text{Cov}(R_t)$  (may depend on the information set at date  $T$ ); and  $S$  is an unbiased estimate of  $\Sigma$  based on  $T$  observations on returns,  $\{R_1, R_2, \dots, R_T\}$ .

We make the unbiasedness assumption for analytical convenience. Even though some covariance matrix estimators do not satisfy this assumption, the results based on that assumption help understand why in-sample optimism occurs. The jackknife estimator we derive in the next section does not require the unbiasedness assumption.

With these notations, our analysis applies 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 benchmark, and the term “portfolio variance minimization” to mean

both portfolio variance minimization and tracking-error variance minimization in our analysis.

Let  $w_s$  denote the global minimum risk portfolio (i.e., the vector of portfolio weights) constructed using an estimated covariance matrix, and  $w_p$  the corresponding minimum risk portfolio constructed from the population moments of the returns. That is,  $w_s$  denotes a sample minimum risk portfolio and  $w_p$  denotes the corresponding population minimum risk portfolio. Then  $w_s' S w_s$  is the in-sample variance of  $w_s$ ,  $w_s' \Sigma w_s$  is its out-of-sample variance, and  $w_p' \Sigma w_p$  is the variance of the population minimum risk portfolio,  $w_p$ .

The population global minimum risk portfolio,  $w_p$ , solves the following problem:

$$\begin{aligned}
 w_p = & \operatorname{argmin}_w \{w' \Sigma w\} & (3) \\
 \text{subject to} & & \\
 & w' \mathbf{1} = 1, & \\
 & \underline{w}_i \leq w_i \leq \bar{w}_i, \text{ for some constants, } \underline{w}_i, \bar{w}_i, i = 1, \dots, N. &
 \end{aligned}$$

Replacing  $\Sigma$  with its estimate based on historical return data,  $S$ , in Equation (3), we obtain the sample global variance minimization problem. We have the following Proposition that characterizes the relation among the expected in-sample variance of the sample global minimum variance portfolio,  $w_s$ , its expected population variance, and the variance of population counterpart,  $w_p$  :

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

$$E(w_s' S w_s) < w_p' \Sigma w_p < E(w_s' \Sigma w_s). \quad (4)$$

We now characterize the difference between the adjacent terms in Inequality (4). Let  $q_1$  denote the difference between the first two terms, and  $q_2$  denote the difference between the last two terms, i.e.,  $q_1 = w_p' \Sigma w_p - E(w_s' S w_s)$  and  $q_2 = E(w_s' \Sigma w_s) - w_p' \Sigma w_p$ .

PROPOSITION 2: *Let  $w_s$  denote the global minimum variance portfolio constructed using an un-*



biased estimate of the covariance matrix without additional portfolio weight constraints. Then on average the out-of-sample variances of  $w_s$  equals its in-sample variance plus two strictly positive terms as given below.

$$E(w'_s \Sigma w_s) = E(w'_s S w_s) + q_1 + q_2, \quad (5)$$

where

$$q_1 \equiv w'_p \Sigma w_p - E(w'_s S w_s) = E[(w_p - w_s)' S (w_p - w_s)], \quad (6)$$

$$q_2 \equiv E(w'_s \Sigma w_s) - w'_p \Sigma w_p = E[(w_p - w_s)' \Sigma (w_p - w_s)], \quad (7)$$

are two positive numbers.

Recall that the optimization problem we are discussing now only has one constraint:  $w' \mathbf{1} = 1$ . We might think that the optimal portfolio has  $N - 1$  degrees of freedom because of this constraint. However, the proof of Proposition 2 shows that the optimal portfolio  $w_s$  also satisfies

$$w'_p \Sigma w_s = w'_p \Sigma w_p.$$

That means there are only  $N - 2$  degrees of freedom in the  $N$  vector  $w_s$  – minimizing the objective function costs one more degree of freedom. That is why the standard degrees of freedom adjustment is not sufficient to get an unbiased estimate of the out-of-sample variance.

When returns are drawn from an i.i.d. multivariate Normal distribution and the covariance matrix used is the sample covariance matrix, we have the following characterization of  $q_1$  and  $q_2$ .

PROPOSITION 3: *Assume that returns are drawn from an i.i.d. multivariate Normal distribution. Let  $w_s$  denote the global minimum variance portfolio constructed using the sample covariance matrix without additional portfolio weight constraints. Then the quantities  $q_1$  and  $q_2$  in Proposition 2 are:*

$$q_1 = \frac{N - 1}{T - N} E(w'_s S w_s), \quad (8)$$

$$q_2 = \frac{N - 1}{T - N - 1} w'_p \Sigma w_p. \quad (9)$$

Since, by definition,  $w'_s S w_s$  is the in-sample variance, we can rewrite the expression for  $q_1$  in Proposition 3 to get:

$$q_1 + E(\text{in-sample variance}) = \frac{T-1}{T-N} \times E(\text{in-sample variance}). \quad (10)$$

By definition, the left side of the above equation equals  $w'_p \Sigma w_p$ , the variance of the population global minimum risk portfolio. The right side of the above equation is the expected value of the standard degrees of freedom adjusted in-sample variance. Clearly, multiplying the in-sample variance by the standard degrees of freedom related factor,  $\frac{T-1}{T-N}$ , is inadequate to correct for in-sample optimism.

From the above three equations we have

$$\begin{aligned} q_2 &= \frac{N-1}{T-N-1} w'_p \Sigma w_p \\ &= \frac{N-1}{T-N-1} \frac{T-1}{T-N} E(w'_s S w_s) \\ &= \frac{T-1}{T-N-1} \frac{N-1}{T-N} E(w'_s S w_s) \\ &= \frac{T-1}{T-N-1} q_1. \end{aligned}$$

Therefore, the total correction,  $q_1 + q_2$ , is  $\frac{2T-N-2}{T-N-1}$  times the standard degrees of freedom correction,  $q_1$ . This factor is slightly larger than 2 when the number of observations is reasonably large. Hence the in-sample estimate of the variance plus twice the standard degrees of freedom related correction 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 when standard diffuse priors are used.

### 3 In-sample Optimism with Bayes Estimator of the Covariance Matrix

The setup of the problem and the notation are the same as in Section 2. 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 investor has the standard diffuse prior about  $\mu$  and  $\Sigma$ , the vector of expected returns and return covariance matrix, given by:

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

where  $\propto$  means “positively proportional to.”

The investor’s objective is to choose the vector of portfolio weights,  $w$ , that minimizes the variance of the portfolio’s return during the next period,  $\text{var}(w'R_{T+1})$ , where  $R_{T+1}$  is the next period’s return, and  $\text{var}(\cdot)$  denotes 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  $\text{var}(w'R_{T+1})$ . Then:*

1.  $w_{sB}$  is the same as  $w_s$ , the global minimum variance portfolio constructed using the sample covariance matrix;
2.  $\text{var}(w'_{sB}R_{T+1}) = \frac{(T-1)(T+1)}{T(T-N-2)}w'_sSw_s$ .

Comparing the right side of the above equation with the degrees of freedom adjusted in-sample variance in (10), we see that

$$\begin{aligned} w'_s\text{var}(R_{T+1})w_s &= \left(\frac{T+1}{T}\right)\left(\frac{T-N}{T-N-2}\right) \\ &\times (\text{DF adjusted in-sample variance}). \end{aligned} \tag{11}$$

The factor  $\frac{T+1}{T} \frac{T-N}{T-N-2}$  will close to 1 when  $N$  is large and  $T$  is larger, but not large enough for

in-sample optimism to be negligible. For example, when  $N = 200$  and  $T = 750$  as in our empirical study,  $\frac{T+1}{T} \frac{T-N}{T-N-2} = 1.0050$ . Hence, the variance of the global minimum variance portfolio under the predictive distribution will be close in magnitude to the degree-of-freedom adjusted in-sample variance, and in-sample optimism will continue to be an issue even when we use the predictive distribution under standard diffuse prior. We reach a similar conclusion when returns have a factor structure. We omit the derivation here for brevity. It is available from the authors.

## 4 Jackknife-type Estimator of the Out-of-sample Variance

In the earlier sections we examined the relation between the *unconditional* out-of-sample and in-sample variances of a minimum risk portfolio, in order to develop an insight into the issues involved. The estimator for the out-of-sample variance we obtained there provides a natural benchmark for the other estimators we consider. In this section we develop a jackknife-type estimator (hereafter called jackknife estimator for short) for the *conditional* out-of-sample variance of a given sample minimum risk portfolio when returns are i.i.d but not necessarily Normal. We will consider the non-i.i.d. case in Section 5.

### 4.1 Portfolio Holding Period Equals the Return Observation Interval

For expositional convenience, we first consider the special case where the investor estimates covariance matrices and recomputes optimal portfolio weights at the end of each period. As before we assume that the investor 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 that estimate and  $w_T$  denote the portfolio weights formed using  $S_T$ . Here, unlike in Sections 2 and 3,  $S_T$  can be any covariance matrix estimator, including the shrinkage estimator of Ledoit and Wolf (2003a, 2003b). Further, any optimization procedure may be used to construct  $w_T$  based on  $S_T$ , even though in this paper we only consider  $w_T$  that minimizes the portfolio variance. Let  $w_T' R_{T+1}$  denote the return on the portfolio,  $w_T$ , during the period  $T+1$ . For notational convenience, let  $Var(w_T' R_{T+1})$  denote the conditional variance of the portfolio return,  $w_T' R_{T+1}$ . We have dropped the subscript  $T$  from

$Var_T(w'_T R_{T+1})$  for notational convenience, although the variance of the portfolio will in general depend on the history of realized returns even when returns are i.i.d. over time.

We first drop the  $i$ 'th return observation, for an arbitrary  $i$  between 1 and  $T$ , and re-estimate the return covariance matrix. Let  $S_{(T-1,i)}$  denote that covariance matrix, i.e., the one obtained using data  $\{R_1, \dots, R_{i-1}, R_{i+1}, \dots, R_T\}$ . We then construct the global minimum variance portfolio using  $S_{(T-1,i)}$ , using the same method that gave us  $w_T$  from  $S_T$ . Denote the resulting portfolio as  $w_{(T-1,i)}$ . For large enough  $T$ , the return,  $w'_{(T-1,i)} R_i$ , will have approximately the same distribution as the out-of-sample return,  $w'_T R_{T+1}$ , that the investor is interested in, if the time series of returns on the primitive assets are independently and identically distributed over time. Hence a natural estimator of the out-of-sample variance,  $Var(w'_T R_{T+1})$ , would be the sample variance of the  $T$  portfolio returns,  $[w'_{(T-1,1)} R_1, w'_{(T-1,2)} R_2, \dots, w'_{(T-1,T)} R_T]$ . 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

$$\hat{q}^{JK} = \frac{1}{T} \sum_{i=1}^T (w'_{(T-1,i)} R_i)^2. \quad (12)$$

In Appendix A we show that this jackknife estimator and the one in the next sub-section provide consistent estimates of the conditional variance.

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 to provide a consistent estimate of the *conditional* variance of the return,  $w'_T R_{T+1}$ . That is because we only need  $w'_{T-1,i} R_i$  and  $w'_T R_{T+1}$  to have approximately the same distribution. In Appendix A we show that  $w_{T-1,i}$  converges to  $w_T$  at the rate  $O(1/T)$ . Even though the estimate of the variance of  $R_{T+1}$  itself converges only at the standard asymptotic rate, i.e.,  $O(1/\sqrt{T})$ , as we will see later, the jackknife estimator provides a good estimate of the conditional variance even for moderately large values of  $T$ .

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 of 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. Revising portfolio holdings once every month instead of once every day would be more reasonable when transactions costs are nontrivial. 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.

## 4.2 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 investor 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 form the optimal portfolio weights,  $w_T$ ; and  $m = T/l$  denote the number of months of observation. Let  $q_{T,l}$  denote the conditional expectation of the sample variance of the  $l$  post-formation daily returns on the minimum risk portfolio, i.e., the conditional expectation of the sample variance of  $(w'_T R_{T+1}, \dots, w'_T R_{T+l})$ . Our objective is to estimate  $q_{T,l}$ .

As before, 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 minimum risk portfolio,  $(w'_{T-l,i} R_{(i-1)l+1}, \dots, w'_{T-l,i} R_{il})$ . 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  $w_{(T-l,i)}$  will also have approximately the same distribution as  $w_T$ . Hence, each sequence,  $(w'_{T-l,i} R_{(i-1)l+1}, \dots, w'_{T-l,i} R_{il})$ , will have approximately the same distribution as the sequence  $(w'_T R_{T+1}, \dots, w'_T R_{T+l})$ , and the average of the  $m$  such variances,  $q_{T,l}(i), i = 1, 2, \dots, m$ , provides a consistent estimate of  $q_{T,l}$ , the object of our interest. This gives us the following jackknife estimator:

$$\widehat{q}_{T,l}^{JK} = \frac{1}{m} \sum_{i=1}^m q_{T,l}(i). \quad (13)$$

As we mentioned earlier, even though the estimator has only asymptotic validity, we expect the convergence to be faster than the standard  $\sqrt{T}$  asymptotic theory.

## 5 Time Varying Covariances

The jackknife estimator we derived in Section 4 assumes that returns are independently and identically distributed over time. 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 i.i.d. assumption. We therefore consider the following two additional estimators of the out-of-sample variance of SMRPs.

### 5.1 Multivariate GARCH

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

$$R_t = \alpha_0 + \alpha_1 D_t + \beta_0 F_t + \beta_1 F_{t-1} + \beta_2 F_{t-2} + \epsilon_t, \quad (14)$$

where  $R_t$  is defined before,  $F_t$  is the column-vector of returns of  $k$  factors in period  $t$ , the  $\beta$ 's are  $N \times k$  matrices of factor loadings, and  $\epsilon_t$  is the residual vector and is i.i.d. over time. Because we are using daily returns, we allow 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 factor is the daily excess return on the stock index portfolio; in the latter, the factors correspond to the three Fama and French (1993) factors.

We model the evolution of the factor returns using a VAR model with error terms following the

DCC multivariate GARCH process given below:

$$F_t = \phi_0 + \phi_1 D_t + \sum_{k=1}^p \Phi_k F_{t-k} + \delta_t, \quad (15)$$

$$\delta_t \sim \text{DCC multivariate GARCH}, \quad (16)$$

where  $\phi_0$  and  $\phi_1$  are  $k$  vectors, the  $\Phi_k$ 's are  $k \times k$  matrices,  $\delta_t$  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 and its estimation are described in Appendix B.

We use daily data, with optimal portfolios re-formed every month. As in the previous sections, assume that the optimal portfolio weight,  $w_T$ , is already constructed. So our task is to estimate,

$$\text{Sample Variance of } (w_T' R_{T+1}, w_T' R_{T+2}, \dots, w_T' R_{T+l}). \quad (17)$$

the sample variance of the optimal portfolio's returns in the  $l$  days of the next month.

To do that we can, using the estimated model parameters, simulate forward the sample paths of  $F_{T+1}, \dots, F_{T+l}$ . Then we use (14) to generate the sample path of future returns,  $R_{T+1}, R_{T+2}, \dots, R_{T+l}$ , and calculate (17). When simulating 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 (14) are generated using the bootstrap method. Specifically, given the parameter estimates of (14), we calculate the realized residuals,  $\hat{\epsilon}_1, \dots, \hat{\epsilon}_T$ . When we use (14) to forecast the returns from the factors, the error terms  $\epsilon_{T+1}, \dots, \epsilon_{T+l}$  are randomly drawn (with replacement) from these  $T$  realized residuals. We use this bootstrap method to catch any non-Normality 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 (17) as our final estimate of the optimal portfolio's out-of-sample variance.

When choosing the number of lags for the VAR model in (15), 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.<sup>4</sup>

## 5.2 Weighted Jackknife

In this subsection, we modify the jackknife estimator in Section 4 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 (12). If returns do not satisfy the i.i.d. assumption, specifically, if returns have persistent covariance structure, it would make intuitive sense to use a weighted average instead of straight average in (12), giving more weight to the recent observations. One such estimator is given by:

$$\hat{q}^{JK} = \frac{\sum_{i=1}^T e^{\alpha i} (w'_{(T-1,i)} R_i)^2}{\sum_{i=1}^T e^{\alpha i}}, \quad (18)$$

for some positive  $\alpha$ . Since the covariance matrix,  $S_{T-1,i}$ , is computed by deleting the  $i$ 'th observation as before, the corresponding vector of portfolio weights,  $w_{T-1,i}$  will be close to  $w_T$  for moderately large values of  $T$ ; and the return,  $w'_{T-1,i} R_i$  will be almost the same as the return,  $w'_T R_i$ . Therefore, based on the results in Foster and Nelson (1996), we believe that the realized sample variance of the sequence  $[w'_{T-1,i} R_i, i = 1, 2, \dots, T]$  computed by weighting those observations with  $i$  closer to  $T + 1$  more than those observations that are farther away using an exponentially declining weighting scheme as given above will provide a better estimate of the conditional variance of

---

<sup>4</sup>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  $\ln |\tilde{\Sigma}| + \ln(T)/T \cdot K$ , respectively, 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.

$w_T' R_{T+1}$ .

When the optimal weights are recomputed once every month using historical daily returns data, the jackknife estimator under the i.i.d. assumption is given by (13). When variances and covariances vary over time in a systematic stochastic fashion, we modify the average in (13) into a weighted average, yielding the weighted jackknife estimate of the out-of-sample variance:

$$\widehat{q_{T,l}}^{JK} = \frac{\sum_{i=1}^m e^{\alpha i} q_{T,l}(i)}{\sum_{i=1}^m e^{\alpha i}}, \quad (19)$$

for some positive  $\alpha$ .

## 6 Empirical Evaluation of the Estimators

### 6.1 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. Therefore, there are two types of estimators in our empirical work: One type consists of the covariance matrix estimators. Different covariance matrix estimators will give rise to different SMRPs. The second type consists of estimators of the out-of-sample risk of the SMRPs. Our focus is on evaluating the relative performance of different estimators of the second type.

Starting from 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 preceding 36 months. We estimate the covariance matrix of the 200 stocks using daily return of the preceding 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.<sup>5</sup> We then estimate 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.<sup>6</sup> 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 tracks the return of the S&P 500 index. Again, we construct two tracking-error minimizing portfolios for each covariance matrix estimator, one with and one without portfolio weight constraints.

## 6.2 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  $\bar{\mu}$  is the vector of average returns.

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

---

<sup>5</sup>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 those of the second one and hence are omitted.

<sup>6</sup>For convenience of interpretation, we work with standard deviations instead of variances.

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

$$S_1 = s_m^2 BB' + D. \tag{20}$$

Here  $B$  is the  $N \times 1$  vector of  $\beta$ '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 et al. (2003a). 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.

### 6.3 Examining the Performance of the Estimators Using Simulated Data

We first examine how the different estimators of out-of-sample risks perform when returns are i.i.d. multivariate Normal, by working with synthetic return data generated by simulation using Monte Carlo methods.

For that purpose, we first read in the real returns (or excess returns for the case of tracking error minimization) on the 200 stocks over a given three-year calibration period, and then read in the realized values of the corresponding three Fama and French (1993) factors. We estimate the betas for all the 200 stocks with respect to the factors, calculate the stocks' residual variances and the factors' covariance matrix. By assuming that returns have an exact three-factor structure and using

the stocks' betas and residual variances and the factor-covariance matrix, we generate the factor returns and stock returns for each day in the three years plus one month period (the total time length of in-sample and out-of-sample) from an i.i.d. multivariate Normal distribution. We then use the simulated data as though they were the historical data and apply the different estimators to form the covariance matrices, optimal portfolio weights, and estimate out-of-sample variances of these SMRPs. We then move forward by one month, and recalibrate using the most recent three years of daily data, and repeat the procedure. This gives one out-of-sample variance and one set of estimates of it, based on simulated data, for each hypothetical month over the 1967/5 to 2000/4 period. We then examine how the different estimators perform by comparing the realized out-of-sample variance with the corresponding estimates.

Next, we repeat the exercise using realized historical returns instead of simulated returns over the 1967/5 to 2000/4 period, and compare the results with the i.i.d. multivariate Normal case.

## 6.4 Empirical Results

Panels A and B of Tables 1–2 and Table 3 display the empirical results for the simulated data. Panels C and D of Tables 1–2 and Tables 4 and 5 give the results obtained using historical data. 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.

### Variance Minimization

Table 1 panels A and C give the result for the global minimum variance portfolios. The \*'s in the last two columns indicate that the corresponding numbers are significantly smaller than one.<sup>7</sup>

---

<sup>7</sup> In this section, we often report that the average of  $x_1, x_2, \dots, x_N$  is significantly smaller than the average of  $y_1, y_2, \dots, y_N$ , 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  $\log(x_1/y_1), \dots, \log(x_N/y_N)$  is significantly smaller than zero using a one-tailed test. We use Newey-West standard

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. 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 estimator of the out-of-sample risk is reasonably accurate.

In Section 2 we showed that 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 about twice as large as the standard degrees of freedom based correction. In Panel A we see that the standard degrees-of-freedom correction for the variance is  $5.42^2 - 4.64^2 = 7.85$ . If we add twice of this correction to the in-sample variance, we get  $4.64^2 + 2 \times 7.85 = 37.22$ , i.e., a standard deviation of 6.10, not much different from the average 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 1 gives the results when historical daily return data on stocks traded on NYSE, AMEX and NASDAQ are used. In that case, 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.<sup>8</sup> These results suggest the inadequacy of the three factor model to

---

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

<sup>8</sup>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 7 for the method of test used.

capture the correlation structure in historical returns and the non i.i.d. nature of the data.

As we expected, in-sample optimism comes down substantially with portfolio weight constraints. 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 7). 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, supporting our earlier observation regarding the non-i.i.d. nature of the return data. The estimate that we obtain by adding twice the standard degrees of freedom based correction to the in-sample estimate is also dominated by the jackknife estimates.

### **Tracking Error Variance Minimization**

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. In-sample optimism is most pronounced for the sample covariance matrix. Adding twice the degree of freedom correction to the in-sample variance gives us  $1.45^2 + 2 \times (1.69^2 - 1.45^2) = 3.61$ , i.e., an estimate of 1.90% for the out-of-sample standard deviation, not much different from the average realized standard deviation of 1.93%. Jackknife estimates are also equally accurate. The three-factor-model based covariance matrix estimator performs the best in terms of both lowest out-of-sample risk and absence of 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 (i.e., Panel B). 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 standard deviation for the three-factor model is 1.32 ( $= 1.93/1.46$ ) times that for the sample covariance matrix. Statistical tests (see footnote 7) 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.

The results in Panels C and D show that, with historical 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  $(5.56^2 + 2 \times (6.49^2 - 5.56^2))^{0.5} = 7.30\%$  for an estimate of the out-of-sample standard deviation, which is still smaller than its average realized value of 8.44%. This is because the historical return data are not i.i.d. multivariate Normal.

The results in Panels C and D suggest that the 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. Clearly there is nontrivial correlation among the residuals from the factor models.

Further, 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.

### **Tracking the Realized Out of Sample Variances**

While Table 1 compares the averages of various estimates of the out-of-sample standard deviation with the average realized out-of-sample standard deviation, Table 2 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 7 for test method).



As in Table 1, there are four panels in Table 2: the first two panels are for the simulated return data, and the last two 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 historical 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 historical data (Panels C and D), the performance becomes significantly worse when compared to estimates made using simulated data (Panels A and B), i.e., the mean absolute differences between the various estimates and realized out of sample variances increased significantly. This suggests that relaxing the i.i.d. assumption may be important.

### **Estimating the Conditional Variance**

Table 3 shows a simulation study where the true covariance matrix is fixed, and not matched to the most recent three years of data every month. We first estimate the covariance matrix of the 200 largest stocks using their daily return from 1963/05 to 1968/04, together with the daily Fama-French three-factor returns, and assuming the covariance matrix has a strict three-factor structure. We use the estimated covariance matrix as the true covariance matrix for the simulation. We simulate i.i.d. Normal daily return data over a three-year plus one month period (number of days set to that over the 1963/05-1968/05 period). Then we use the first three years' simulated data to estimate the covariance matrix and form the optimal portfolios, and examine these portfolios' performance over the last month of the simulated data. This procedure is repeated 396 times and Table 3 shows some summary statistics.

Panels A and B report the results of the global minimum variance portfolios. Several patterns are worth noting. First, the in-sample standard deviation with the standard degree of freedom correction is 3.20, which is exactly the standard deviation of the population global minimum variance

portfolio. This verifies the theoretical results presented earlier. Second, on average, the in-sample standard deviation with twice of the degree-of-freedom correction is 3.65, which is almost the corresponding true standard deviation of the sample optimal portfolio of 3.73, again verifying our earlier analysis.

Panel B shows that jackknife tracks the true standard deviation of the sample optimal portfolio much better than the in-sample estimate, with the mean absolute difference of the former being only a fraction of the latter. The last column of Panel B shows that the standard deviation (calculated over the 396 repetitions) of the risk of sample optimal portfolio is quite small. This means the dependency of the sample optimal portfolio's risk on the particular realized returns used in estimating the covariance matrix is rather small. Hence the twice-of-the-degree-of-freedom-correction rule which was derived for the unconditional out-of-sample risk should work well for the conditional one as well (i.e., the one conditional on a particular realized sample of returns used to estimate the covariance matrix), when returns satisfy the i.i.d multivariate Normality assumption.

Panels C and D report the simulation results for the minimum tracking error variance portfolios. We reach qualitatively similar conclusions here as in Panels A and B. The twice-the-degrees-of-freedom correction rule performs rather well for the tracking error minimization problem as well.

### **Dynamic Conditional Correlation Model and Weighted Jackknife**

In Sections 5.5.1 and 5.5.2, we proposed a multivariate GARCH-based method and a weighted-jackknife method to estimate the out-of-sample standard deviations of the sample minimum risk portfolios to allow for time variations in variances and covariances of returns. Tables 4 and 5 report the performance of these methods. In using the weighted-jackknife, we choose the decay rate  $\alpha = 0.21$  in Equation (19), which corresponds to about 0.01 per day since there are on average about 21 trading days per month.<sup>9</sup>

The layout of Tables 4 and 5 is similar to that of Tables 1 and 2, except we only examine

---

<sup>9</sup>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), and 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 (19).

the results with historical data. That is because our purpose here is to see how these estimators perform when the return data exhibits persistence in variances and covariances across time.

Table 4 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 process 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 5 reports mean absolute differences between the various estimators and the realized out-of-sample standard deviation. As in Table 2, a \* in the last two columns indicates that the associated number is statistically larger than the corresponding number in the first column (see footnote 7 for test method). Under this criterion, the weighted-jackknife often gives significantly smaller mean absolute tracking error. Considering both the unconditional bias in Table 4 and the mean absolute difference here, the weighted-jackknife method performs better than the other methods we considered.

## 7 Conclusion

It has been observed in the literature that the in-sample variance of a mean-variance efficient portfolio constructed using estimated inputs typically understates its true variance, i.e., there is *in-sample optimism*. In this paper we identify why in-sample optimism occurs and suggest a jackknife type estimator for the out-of-sample conditional variance of minimum risk portfolios constructed using estimated inputs.

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 on average about 64 percent of the out-of-sample standard deviation. The corresponding number is 92 percent for the jackknife-type estimator.

The use of one- and three-factor models lead to increased-tracking error variance suggesting the presence of 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).

An 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, the value of this advantage of factor models comes down.

The performance of the jackknife-type estimator worsens when evaluated using realized historical data instead of synthetic data obtained through simulation from a suitably calibrated i.i.d. multivariate Normal distribution. This is probably due to volatility persistence in the historical return data. We show that the use of the Dynamic Conditional Correlation Model of Engle (2002) or a weighted-jackknife method to account for such persistence leads to substantial improvement.

## Appendix. Mathematical Derivations

### A. Proofs

PROOF OF PROPOSITION 1.

Since  $w_p$  is the global minimum variance portfolio with respect to the population covariance matrix  $\Sigma$ , obviously

$$w_p' \Sigma w_p \leq w_s' \Sigma w_s$$

for any portfolio  $w_s$  that satisfies the same constraint as  $w_p$ , and the inequality becomes strict when  $w_s$  is different from  $w_p$ . Since  $w_s$  is the optimal portfolio constructed from the sample covariance matrix,  $w_s \neq w_p$  with probability one. Taking expectations on both sides of the above, we get the second half of (4).

Similarly, we have

$$w_s' S w_s < w_p' S w_p$$

with probability one. Taking expectations on both sides, and noticing that  $E(S) = \Sigma$ , we get the first half of (4).

PROOF OF PROPOSITION 2. Equation (5) and the first part of Equations (6) and (7) are obvious from Proposition 1. So we only need to prove the second part of (6) and (7).

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:

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

Using the fact that,

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

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:

$$E(w_s' \Sigma w_s) = w_p' \Sigma w_p + E[(w_s - w_p)' \Sigma (w_s - w_p)]. \quad (\text{A1})$$

This proves the second part of (7). The proof for the second part of (6) is similar.

PROOF OF PROPOSITION 3: With no portfolio weight constraints, the in-sample variance of  $w_s$  is  $\frac{1}{\mathbf{1}' S^{-1} \mathbf{1}}$ , here  $S$  is the sample covariance matrix. The average in-sample variance of this portfolio is

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

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

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

When a scalar random variable  $x$  has a Wishart( $T-N, \sigma^2$ ) distribution,  $\frac{x}{\sigma^2}$  has a  $\chi_{T-N}^2$  distribution (Muirhead (1982), p. 87). Since the expectation of a  $\chi_{T-N}^2$  distribution 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.,

$$E\left(\frac{T-1}{T-N} \frac{1}{\mathbf{1}' S^{-1} \mathbf{1}}\right) = \frac{1}{\mathbf{1}' \Sigma^{-1} \mathbf{1}} = w_p' \Sigma w_p. \quad (\text{A2})$$

Therefore,

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

To get the expression for  $q_2$ , we first simplify  $E(w_s' \Sigma w_s)$ , by following Kan and Zhou (2005). Define  $A$  as  $A = ([\nu, Q]' W^{-1} [\nu, Q])^{-1}$  where  $W = \Sigma^{-1/2} S \Sigma^{-1/2}$  (with  $TW$  following a Wishart distribution with  $(T-1)$  d.f.) and  $[\nu, Q]$  is a  $N \times N$  orthogonal matrix with  $\nu = (\Sigma^{-1/2} \mathbf{1}) / (\mathbf{1}' \Sigma^{-1} \mathbf{1})^{1/2}$ . From equation (A48) of Kan and Zhou (2005), it follows that,

$$w_s' \Sigma w_s = (1 + z' A_{22}^{-1} z) / (\mathbf{1}' \Sigma^{-1} \mathbf{1})$$

where  $z = -A_{22}^{-1/2} A_{21}$  and  $A_{ij}$  is the  $ij$ th sub-matrix of  $A$ . From Theorem 3.2.10 of Muirhead (1982)  $z$  and  $A_{22}$  are independent. From Theorem 3.2.12 of Muirhead (1982)  $z' A_{22}^{-1} z = v_1/v_2$  where  $v_1 \sim \chi_{N-1}^2$  and  $v_2 \sim \chi_{T-N+1}^2$  and they are independent. Therefore,

$$E(z' A_{22}^{-1} z) = E(v_1/v_2) = (N-1)/(T-N-1)$$

since for  $v \sim \chi_p^2$ ,  $E(1/v) = 1/(p-2)$ . Hence

$$E(w_s' \Sigma w_s) = \frac{1 + (N-1)/(T-N-1)}{\mathbf{1}' \Sigma^{-1} \mathbf{1}} = (1 + (N-1)/(T-N-1)) w_p' \Sigma w_p$$

It follows that,

$$q_2 \equiv E(w_s' \Sigma w_s) - w_p' \Sigma w_p = \frac{N-1}{T-N-1} w_p' \Sigma w_p.$$

PROOF OF PROPOSITION 4. It is well-known that the predictive distribution for  $R_{T+1}$  is multivariate Student  $t$  (Zellner (1971), p.235-236) given by:

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

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

$$w_s' \text{var}(R_{T+1}) w_s = \frac{(T+1)(T-1)}{T(T-N-2)} w_s' S w_s = \frac{(T-1)(T+1)}{T(T-N-2)} \frac{1}{\mathbf{1}' S^{-1} \mathbf{1}}. \quad (\text{A3})$$

This proves Proposition 4.

#### CONSISTENCY OF THE JACKKNIFE ESTIMATORS.

We first prove a lemma on the closeness of  $w_{i,T}$  and  $w_T$ . To simplify the notation, we use  $w_{i,T}$  to denote  $w_{(T-l,i)}$  and  $S_{i,T}$  to denote  $S_{(T-l,i)}$ .

LEMMA 1: *Assume that monthly returns (i.e., blocks of  $l$  returns) follow a stationary process, so that all sample variances computed from them are finite random variables. Let the portfolios ( $w_T$  and  $w_{i,T}$ ) be constructed from the sample covariance matrices ( $S_T$  and  $S_{i,T}$  resp.). Assume further that  $S_T$  and its population counterpart are positive definite. Then, for any  $i = 1, \dots, m$ ,  $w_T - w_{i,T}$  is  $O(1/m)$ , in the sense that, for every  $\epsilon > 0$ , there exists  $N$ , such that*

$$P(m \|w_T - w_{i,T}\| > N) \leq \epsilon.$$

(I.e.,  $m(w_T - w_{i,T})$  is a finite random variable for any  $m$ .)

PROOF. The Lagrangian function of the variance minimization problem with respect to  $S$  is:

$$f_s(w) = \frac{1}{2} w' S w - \delta(w' \mathbf{1} - 1) - (w - \underline{w})' \theta_1 + (w - \bar{w})' \theta_2. \quad (\text{A4})$$



Taking the derivative of  $f_s(w)$  with respect to  $w$  in (A4) and equating with zero we obtain

$$\mathbf{0} = \text{grad}f_s(w) = Sw - \delta\mathbf{1} - \theta_1 + \theta_2. \quad (\text{A5})$$

Observe that  $w_T$  is a solution to the above equation. By the same token,  $w_{i,T}$  would also solve the above equation when  $S$  (or,  $S_T$ ) is replaced by  $S_{i,T}$ , where the ‘ $i$ ’ in the subscripts indicates that the  $i$ th block (month) is omitted. For the full sample problem, let us denote the solution of the parameters corresponding to  $w_T$  as  $\delta_T$ ,  $\theta_{1,T}$ , and  $\theta_{2,T}$ , respectively, and for the reduced sample problem (i.e., with the  $i$ th block deleted) the solution of the parameters corresponding to  $w_{i,T}$  as  $\delta_{i,T}$ ,  $\theta_{1,i,T}$  and  $\theta_{2,i,T}$ , respectively. Thus,

$$\begin{aligned} \mathbf{0} &= S_T w_T - \delta_T \mathbf{1} - \theta_{1,T} + \theta_{2,T} - [S_{i,T} w_{i,T} - \delta_{i,T} \mathbf{1} - \theta_{1,i,T} + \theta_{2,i,T}] \\ &= (S_T - S_{i,T}) w_{i,T} + S_T (w_T - w_{i,T}) \\ &\quad - (\delta_T - \delta_{i,T}) \mathbf{1} - (\theta_{1,T} - \theta_{1,i,T}) + (\theta_{2,T} - \theta_{2,i,T}). \end{aligned} \quad (\text{A6})$$

Therefore,

$$\begin{aligned} S_T (w_{i,T} - w_T) &= (S_T - S_{i,T}) w_{i,T} \\ &\quad - (\delta_T - \delta_{i,T}) \mathbf{1} - (\theta_{1,T} - \theta_{1,i,T}) + (\theta_{2,T} - \theta_{2,i,T}). \end{aligned} \quad (\text{A7})$$

We first observe that

$$\begin{aligned} S_T - S_{i,T} &= \frac{(l-1)S_i}{ml-1} + \frac{(m-1)l \bar{X}_i \bar{X}_i'}{(ml-1)m} - \frac{(m-1)l (\bar{R}_{i,T} \bar{X}_i' + \bar{X}_i \bar{R}_{i,T}')}{(ml-1)m} \\ &\quad + \frac{(m-1)l \bar{R}_{i,T} \bar{R}_{i,T}'}{(ml-1)m} - \frac{l S_{i,T}}{ml-1}, \end{aligned} \quad (\text{A8})$$

where  $\bar{R}_{i,T} = (1/[(m-1)l])(\sum_{j=1}^{ml} R_j - \sum_{j=(i-1)l+1}^{il} R_j)$ ,  $\bar{X}_i = (1/l) \sum_{j=(i-1)l+1}^{il} R_j$  and  $S_i = (1/[l-1]) \sum_{j=(i-1)l+1}^{il} (R_j - \bar{X}_i)(R_j - \bar{X}_i)'$ . These notations will be used in Proposition 5 below. Defining  $\|B\|_{tr} = (\text{Trace}(BB'))^{1/2}$  and  $\|x\| = (\sum_j x_j^2)^{1/2}$  for any matrix  $B$  and vector  $x$  and observing that  $\|\bar{X}_i\|$ ,  $\|S_i\|_{tr}$ ,  $\|\bar{R}_{i,T}\|$  and  $\|S_{i,T}\|_{tr}$  all are finite random variable, from (A8) one gets

$S_T - S_{i,T} = O(1/m)$ .

Now, observe that,

$$(w_{i,T} - w_T)'(\theta_{1,T} - \theta_{1,i,T}) = (w_{i,T} - \underline{w})'\theta_{1,T} - (w_{i,T} - \underline{w})'\theta_{1,i,T} - (w_T - \underline{w})'\theta_{1,T} + (w_T - \underline{w})'\theta_{1,i,T}$$

since  $(w_{i,T} - w_T) = (w_{i,T} - \underline{w}) - (w_T - \underline{w})$ . Also, since

$$(w_{i,T} - \underline{w})'\theta_{1,i,T} = 0 = (w_T - \underline{w})'\theta_{1,T} \quad (\text{A9})$$

one obtains

$$(w_{i,T} - w_T)'(\theta_{1,T} - \theta_{1,i,T}) = (w_{i,T} - \underline{w})'\theta_{1,T} + (w_T - \underline{w})'\theta_{1,i,T} \geq 0 \quad (\text{A10})$$

since  $(w_{i,T} - \underline{w})$ ,  $(w_T - \underline{w})$ ,  $\theta_{1,i,T}$  and  $\theta_{1,T}$  are all non-negative vectors.

Similarly,

$$\begin{aligned} (w_{i,T} - w_T)'(\theta_{2,T} - \theta_{2,i,T}) &= -(\bar{w} - w_{i,T})'\theta_{2,T} + (\bar{w} - w_{i,T})'\theta_{2,i,T} + (\bar{w} - w_T)'\theta_{2,T} - (\bar{w} - w_T)'\theta_{2,i,T} \\ &= -(\bar{w} - w_{i,T})'\theta_{2,T} - (\bar{w} - w_T)'\theta_{2,i,T} \\ &\leq 0. \end{aligned} \quad (\text{A11})$$

The second equality follows, since

$$(\bar{w} - w_{i,T})'\theta_{2,i,T} = 0 = (\bar{w} - w_T)'\theta_{2,T} \quad (\text{A12})$$

and the last inequality follows, since  $(\bar{w} - w_{i,T})$ ,  $(\bar{w} - w_T)$ ,  $\theta_{2,i,T}$  and  $\theta_{2,T}$  are all non-negative vectors.

Thus, observing  $w'_{i,T}\mathbf{1} = 1 = w'_T\mathbf{1}$ , one obtains from (A7)

$$\begin{aligned}
(w_{i,T} - w_T)'S_T(w_{i,T} - w_T) &= (w_{i,T} - w_T)'(S_T - S_{i,T})w_{i,T} \\
&\quad - (w_{i,T} - \underline{w})'\theta_{1,T} \\
&\quad - (w_T - \underline{w})'\theta_{1,i,T} - (\bar{w} - w_{i,T})'\theta_{2,T} - (\bar{w} - w_T)'\theta_{2,i,T} \\
&\leq (w_{i,T} - w_T)'(S_T - S_{i,T})w_{i,T}.
\end{aligned} \tag{A13}$$

Here we used the complementary slackness conditions (A9) and (A12) in the first equality, and (A10) and (A11) in the next inequality.

Therefore,

$$\begin{aligned}
&C_s \|w_{i,T} - w_T\|^2 \\
&\leq \|w_{i,T} - w_T\| \|w_{i,T}\| \|S_T - S_{i,T}\|_{tr},
\end{aligned} \tag{A14}$$

where  $C_s$  is the minimum eigenvalue of  $S_T$  and is therefore positive. Now, using (A8) and the fact that  $\|w_{i,T}\|$  and  $\|w_T\|$  are all  $O(1)$  random variables whereas  $C_s$  will be uniformly bounded from 0 since they will be close to their population counterparts for large samples one gets

$$\begin{aligned}
\|w_{i,T} - w_T\| &\leq \frac{1}{C_s} \|w_{i,T}\| \|S_T - S_{i,T}\|_{tr} \\
&= O\left(\frac{1}{m}\right).
\end{aligned} \tag{A15}$$

Hence the proof.

REMARK 1: *The Lemma assumes that the portfolios are constructed from the sample covariance matrices of the returns. However, it is clear from the above proof that any estimator  $\hat{\Sigma}$  of  $\Sigma$  that satisfies:*

$$\hat{\Sigma}_T - \hat{\Sigma}_{i,T} = O\left(\frac{1}{m}\right)$$

*with  $\hat{\Sigma}_T$  and its population counterpart being positive definite matrices, will give rise to portfolios ( $w_T$ , and  $w_{i,T}$  resp.) that would satisfy the Lemma, i.e.,  $w_T - w_{i,T} = O(1/m)$ .*

We now consider the general case discussed in Section 4.4.2.

PROPOSITION 5: *Assume that monthly returns (i.e., blocks of  $l$  returns) follow a stationary process, so that all sample variances computed from them are finite random variables. Let the portfolios ( $w_T$  and  $w_{i,T}$ ) be constructed from the sample covariance matrices ( $S_T$  and  $S_{i,T}$  resp.), where  $S_T$  and its population counterpart are positive definite. Assume further that the returns and portfolios have finite fourth moment. (Assumption of fourth moment on portfolios are immediate whenever they are constructed subject to some lower bounds as they add up to one.) For a fixed  $w_T$ , the conditional expectation of the sample variance of  $\{w_T' R_{T+1}, \dots, w_T' R_{T+l}\}$  is consistently estimated by  $(\sum_{i=1}^m \frac{1}{a_i}) \sum_{i=1}^m a_i w_{i,T}' S_i w_{i,T}$ , provided  $(\bar{S} - \Sigma_T)$  converges to zero in mean (or mean square), where  $S_i$  is the sample covariance matrix estimated using the daily data of month  $i$  (i.e.,  $S_i = (1/[l - 1]) \sum_{j=(i-1)l+1}^{il} (R_j - \bar{X}_i)(R_j - \bar{X}_i)'$ ), and  $\Sigma_T$  is the conditional expectation of sample covariance matrix of  $\{R_{T+1}, \dots, R_{T+l}\}$  (i.e., conditional expectation of  $S_{m+1}$ ) given returns up to time  $T$ , and  $\bar{S} = (\sum_{i=1}^m \frac{1}{a_i}) \sum_{i=1}^m a_i S_i$ , where  $\{a_i\}$  are a sequence of positive numbers. The convergence in mean is of order  $O(1/\sqrt{m})$ , whenever the order of the convergence of  $(\bar{S} - \Sigma_T)$  is  $O(1/\sqrt{m})$ .*

PROOF.

Observe that,

$$\begin{aligned}
& \sum_{i=1}^m a_i w_{i,T}' S_i w_{i,T} - \sum_{i=1}^m a_i w_T' \Sigma_T w_T \\
= & \sum_{i=1}^m a_i (w_{i,T} - w_T)' (S_i - \Sigma_T) (w_{i,T} - w_T) + 2 \sum_{i=1}^m a_i w_T' (S_i - \Sigma_T) (w_{i,T} - w_T) \\
& + \sum_{i=1}^m a_i (w_{i,T} - w_T)' \Sigma_T (w_{i,T} - w_T) + 2 \sum_{i=1}^m a_i w_T' \Sigma_T (w_{i,T} - w_T) \\
& + \sum_{i=1}^m a_i w_T' (S_i - \Sigma_T) w_T
\end{aligned} \tag{A16}$$

After dividing throughout by  $\sum_{i=1}^m a_i$ , we will show each term goes to zero in the mean and establish

our consistency result. Observe that the first term,

$$\begin{aligned}
& E \left| \frac{1}{\sum_{i=1}^m a_i} \sum_{i=1}^m a_i (w_{i,T} - w_T)' (S_i - \Sigma_T) (w_{i,T} - w_T) \right| \\
& \leq \left[ \frac{\sum_{i=1}^m a_i E[\|w_{i,T} - w_T\|^2 \|S_i - \Sigma_T\|_{tr}]}{\sum_{i=1}^m a_i} \right] \\
& \leq \left[ \frac{\sum_{i=1}^m a_i (E[\|w_{i,T} - w_T\|^4])^{\frac{1}{2}} \times E[\|S_i - \Sigma_T\|_{tr}^2]^{\frac{1}{2}}}{\sum_{i=1}^m a_i} \right] \\
& \leq \left[ \frac{\sum_{i=1}^m a_i E[\|w_{i,T} - w_T\|^4]}{\sum_{i=1}^m a_i} \right]^{\frac{1}{2}} \times \left[ \frac{\sum_{i=1}^m a_i E[\|S_i - \Sigma_T\|_{tr}^2]}{\sum_{i=1}^m a_i} \right]^{\frac{1}{2}}. \tag{A17}
\end{aligned}$$

The last inequality follows from Cauchy-Schwarz inequality using  $a_i/(\sum_{i=1}^m a_i)$  as probability for outer expectation. Using Lemma 1 above and the assumption of the moments of portfolios and returns in this Proposition one gets

$$\left[ \frac{1}{\sum_{i=1}^m a_i} \sum_{i=1}^m a_i E[\|w_{i,T} - w_T\|^4] \right]^{1/2} \rightarrow 0$$

in the order of  $O(1/m^2)$ . Now the second term of (A17) is bounded by

$$\left[ \frac{1}{\sum_{i=1}^m a_i} \sum_{i=1}^m a_i E[\|S_i - \Sigma_T\|_{tr}^2] \right]^{1/2},$$

which is bounded by a constant. Thus the first term of (A16) is of order  $O(1/m^2)$ .

Now the next term in (A16)

$$\begin{aligned}
& 2E \left| \frac{\sum_{i=1}^m a_i w_T' (S_i - \Sigma_T) (w_{i,T} - w_T)}{\sum_{i=1}^m a_i} \right| \\
& \leq 2 \left[ \frac{\sum_{i=1}^m a_i E[\|w_T\| \|w_{i,T} - w_T\| \|S_i - \Sigma_T\|_{tr}]}{\sum_{i=1}^m a_i} \right] \\
& \leq 2 \left[ \frac{\sum_{i=1}^m a_i E[\|w_T\| \|w_{i,T} - w_T\|^2]}{\sum_{i=1}^m a_i} \right]^{\frac{1}{2}} \times \left[ \frac{\sum_{i=1}^m a_i E[\|S_i - \Sigma_T\|_{tr}^2]}{\sum_{i=1}^m a_i} \right]^{\frac{1}{2}}. \tag{A18}
\end{aligned}$$

Since  $E[\|w_T\| \|w_{i,T} - w_T\|^2] \leq (E[\|w_T\|^4])^{1/2} (E[\|w_{i,T} - w_T\|^4])^{1/2}$  both term in (A18) can be bounded by arguing exactly as in (A17) and the order of convergence is  $O(1/m)$ .

Observe that the third term in (A16)

$$\begin{aligned}
& E \left| \frac{\sum_{i=1}^m a_i (w_{i,T} - w_T)' \Sigma_T (w_{i,T} - w_T)}{\sum_{i=1}^m a_i} \right| \\
& \leq \left[ \frac{\sum_{i=1}^m a_i E [\|w_{i,T} - w_T\|^2 \|\Sigma_T\|_{tr}]}{\sum_{i=1}^m a_i} \right] \\
& \leq \left[ \frac{\sum_{i=1}^m a_i (E[\|w_{i,T} - w_T\|^4])^{\frac{1}{2}} (E[\|\Sigma_T\|_{tr}^2])^{\frac{1}{2}}}{\sum_{i=1}^m a_i} \right]
\end{aligned}$$

and it converges to zero as in (A17) in the order of  $O(1/m^2)$ .

Now the next term of (A16)

$$2E \left| \frac{\sum_{i=1}^m a_i w_T' \Sigma_T (w_{i,T} - w_T)}{\sum_{i=1}^m a_i} \right| \leq 2 \left[ \frac{\sum_{i=1}^m a_i E [\|w_T\| \|w_{i,T} - w_T\| \|\Sigma_T\|_{tr}]}{\sum_{i=1}^m a_i} \right].$$

Since  $E[\|w_T\| \|w_{i,T} - w_T\| \|\Sigma_T\|_{tr}] \leq (E[\|w_{i,T} - w_T\|^2])^{1/2} (E[\|w_T\| \|\Sigma_T\|_{tr}^2])^{1/2}$  arguing as in (A18) one gets order of the 4th term of (A16) is  $O(1/m)$ .

Finally, the last term of (A16),

$$\begin{aligned}
E \left| \frac{\sum_{i=1}^m a_i w_T' (S_i - \Sigma_T) w_T}{\sum_{i=1}^m a_i} \right| &= E |w_T' (\bar{S} - \Sigma_T) w_T| \\
&\leq E (\|w_T\|^2 \|\bar{S} - \Sigma_T\|_{tr}) \\
&\leq (E[\|w_T\|^4])^{\frac{1}{2}} (E[\|\bar{S} - \Sigma_T\|_{tr}^2])^{\frac{1}{2}},
\end{aligned}$$

where the first term is bounded by a constant and the second term converges to zero in the order of  $O(1/\sqrt{m})$ .

Hence the proof as the lowest order of convergence is  $O(1/\sqrt{m})$ .

REMARK 2: *From Lemma 1 it is important to observe that, since  $w_T - w_{i,T} = O(1/m)$ , all but the last term converge at least in the order of  $O(1/m)$  (the first and third terms of (A16) are of  $(1/m^2)$ ). The last term converges in the order of  $O(1/\sqrt{m})$  because the convergence of  $\bar{S}$  to  $\Sigma$  is in the order of  $O(1/\sqrt{m})$ . Further, any estimator  $\hat{\Sigma}$  of  $\Sigma$  and  $\hat{\mu}$  of  $\mu$  satisfying the conditions stated in Remark 1 give rise to portfolios ( $w_T$ , and  $w_{i,T}$  respectively) that satisfy the Lemma, i.e.,*

$w_T - w_{i,T} = O(1/m)$ . Thus, the conclusion of Proposition 5 also holds for such cases as well.

REMARK 3: For *i.i.d.* returns, we can set  $a_i = 1$  for all  $i$ , then the assumption of convergence of  $(\bar{S} - \Sigma_T)$  holds. For returns that follow GARCH type models, we can set  $a_i = e^{\alpha i}$  for certain positive decay rate  $\alpha$ , and the assumption of convergence of  $(\bar{S} - \Sigma_T)$  holds.

## B. The DCC Multivariate GARCH Model and Its Estimation

The Dynamic Conditional Correlation (DCC) multivariate GARCH model was proposed by Engle (2002) and Tse and Tsui (2002). Here we follow Engle and Sheppard (2001). Let  $\{\delta_t\}$  be a  $k$ -variate zero mean and conditionally normal time series, and  $H_t$  be the conditional covariance matrix of  $\delta_t$ . The DCC model describes the evolution of  $H_t$  over time.

The covariance matrix  $H_t$  is decomposed into variances and correlations:

$$H_t = \Lambda_t P_t \Lambda_t,$$

where  $\Lambda_t$  is the diagonal matrix of standard deviations of  $\delta_t$ , and  $P_t$  is the correlation matrix.

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

$$Q_t = \left(1 - \sum_{i=1}^I \gamma_i - \sum_{j=1}^J \eta_j\right) \bar{Q} + \sum_{i=1}^I \gamma_i (\tilde{\delta}_{t-i} \tilde{\delta}'_{t-i}) + \sum_{j=1}^J \eta_j Q_{t-j}, \quad (\text{A19})$$

$$P_t = Q_t^{*-1} Q_t Q_t^{*-1}. \quad (\text{A20})$$

Here each element of  $\tilde{\delta}_t$  is the corresponding element of  $\delta_t$  standardized by its conditional standard deviation, so it has unit variance.  $\bar{Q}$  is the unconditional covariance of the standardized residuals, and

$$Q_t^* = \begin{bmatrix} \sqrt{q_{11}} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{q_{22}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{q_{kk}} \end{bmatrix}, \quad (\text{A21})$$

so the normalization in (A20) guarantees that  $P_t$  is a matrix of correlations.

Following Engle and Sheppard (2001), we set the lags  $I$  and  $J$  to one.

We assume that the parameter estimates in Equation (15) and those in Equation (16) are



asymptotically uncorrelated and first estimate Equation (15) using OLS (i.e., ignoring the conditional heteroskedasticity in  $\delta_t$ ), and get the residuals  $\{\hat{\delta}_t\}$ . We then use  $\{\hat{\delta}_t\}$  as if they are the true  $\{\delta_t\}$  to estimate the parameters in the DCC multivariate GARCH model.

To estimate a DCC multivariate GARCH model for  $\{\delta_t\}$ , we first estimate a univariate GARCH(1,1) process for each component, given that our choice of number of lags  $I$  and  $J$  are both one. This is achieved using the standard maximum likelihood method. The estimated GARCH(1,1) process produces estimates of the conditional variances for each component at every  $t$ . For each  $t$ , we standardize each component of  $\delta_t$  by the square root of its conditional variance, and this gives us the standardized residual vector  $\tilde{\delta}_t$ .

To estimate the parameters that govern the evolution of the conditional correlation matrix  $P_t$ , we can use the maximum likelihood method. The log likelihood is given by:

$$L(\gamma_1, \eta_1) = -\frac{1}{2} \sum_{t=1}^T (\log(|P_t|) + \tilde{\delta}_t' P_t^{-1} \tilde{\delta}_t), \quad (\text{A22})$$

Notice that  $\bar{Q}$  in Equation (A19) can be estimated by the sample covariance matrix of  $\tilde{\delta}_t$ 's. Also, asymptotically the estimate of  $(\gamma_1, \eta_1)$  does not depend on the initial value of  $Q_0$ . So in practice we use the sample covariance matrix of the first few standardized residuals as the value of  $Q_0$ . We then calculate  $L(\gamma_1, \eta_1)$  using Equations (A19) - (A22). Maximizing the log likelihood is achieved using Newton-Raphson method with numerical derivatives.

## References

- Ahn, D., Conrad, J., Dittmar, R., 2005. Basis assets. Working paper. Kenan-Flagler Business School, University of North Carolina.
- Chan, L., Karceski, J., Lakonishok, J., 1999. On portfolio optimization: forecasting covariances and choosing the risk model. *Review of Financial Studies* 12, 937-974.
- Engle, R., 2002. Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20, 339-350.
- Engle, R., Sheppard, K., 2001. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Working paper. University of California at San Diego.
- Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3-56.
- Flemming, J., Kirby, C., Ostdiek, B., 2003. The economic value of volatility timing using 'realized' volatility. *Journal of Financial Economics* 67, 473-509.
- Foster, D., Nelson, D., 1996. Continuous record asymptotics for rolling sample variance estimators. *Econometrica* 64, 139-174.
- Frost, P., Savarino, J., 1988. For better performance: constrain portfolio weights. *Journal of Portfolio Management* 15, 29-34.
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2003. Predicting volatility: getting the most out of return data sampled at different frequencies. Working paper. University of North Carolina and UCLA.
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2004a. There is a risk-return trade-off after all. Working paper. University of North Carolina and UCLA.
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2004b. The MIDAS touch: mixed data sampling regression models. Working paper. University of North Carolina and UCLA.

- Gibbons, M., 1982. Multivariate test of financial models: a new approach. *Journal of Financial Economics* 10, 3-27.
- Gibbons, M., Ross, S., Shanken, J., 1989. A test of the efficiency of a given portfolio. *Econometrica* 57, 1121-1152.
- Jacquier, E., Kane, A., Marcus, A., 2002. Optimal forecasts of long term returns and asset allocation: geometric, arithmetic, or other means? Working paper. Boston University.
- Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: why imposing the wrong constraints helps. *Journal of Finance* 58, 1651-1683.
- Jobson, J., Korkie, B., 1981. Putting Markowitz theory to work. *Journal of Portfolio Management*, 70-74.
- Jobson, J., Korkie, B., 1982, Potential performance and tests of portfolio efficiency. *Journal of Financial Economics* 10, 433-466.
- Kan, R., Smith, D., 2005. The distribution of the sample minimum-variance frontier. Working paper. The University of Toronto.
- Kan, R., Zhou, G., 2005. Optimal estimation for economic gains: portfolio choice with parameter uncertainty. Working paper. Washington University in St. Louis.
- Kandel, S., 1984. The likelihood ratio test of mean-variance efficiency without a riskless asset. *Journal of Financial Economics* 13, 575-592.
- Künsch, H., 1989. The jackknife and the bootstrap for general stationary distributions. *Annals of Statistics* 17, 1217-1241.
- Lahiri, S., 2002. On the jackknife-after-bootstrap method for dependent data and consistency properties. *Econometric Theory* 18, 79-98.
- Ledoit, O., Wolf, M., 2003a. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10, 603-621.
- Ledoit, O., Wolf, M., 2003b. Honey, I shrunk the sample covariance matrix. Working paper. Credit Suisse First Boston and Universitat Pompeu Fabra.

- Liu, Q., 2003. On portfolio optimization: how do we benefit from high-frequency data? Working paper. Kellogg Graduate School of Management, Northwestern University.
- Lütkepohl, H., 1993. Introduction to Multiple Time Series Analysis (2nd ed.). Springer-Verlag, New York.
- MacKinlay, A., 1987. On multivariate tests of the CAPM. *Journal of Financial Economics* 18, 341-372.
- Michaud, R., 1989. The Markowitz optimization enigma: is optimized optimal? *Financial Analysts Journal* 45, 31-42.
- Muirhead, R., 1982. *Aspects of Multivariate Statistical Theory*. John Wiley and Sons, New York.
- Roll, R., 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, W., 1963. A simplified model for portfolio analysis. *Management Science* 9, 277-293.
- Shanken, J., 1985. Multivariate tests of the zero-beta CAPM. *Journal of Financial Economics* 14, 327-348.
- Stambaugh, R., 1982. On the exclusion of assets from tests of the two parameter model. *Journal of Financial Economics* 10, 235-268.
- Tse, Y., Tsui, A., 2002. A multivariate generalized autoregressive conditional heteroskedasticity model with time-varying correlations. *Journal of Business and Economic Statistics* 20, 351-362.
- Zellner, A., 1971. *An Introduction to Bayesian Inference in Econometrics*. John Wiley and Sons, New York.

Table 1  
Average standard deviations

	In-sample		Jackknife		Out-of-sample	Normalized Std.Dev.	
	DF adjusted	Bayesian	KS	KS		In/Out	Jackknife/Out
Panel A. Global minimum variance portfolio: Simulated i.i.d. normal data							
Sample covariance matrix	4.64	5.42	6.26	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	
Panel B. Minimum tracking error variance portfolio: Simulated i.i.d. normal data							
Sample covariance matrix	1.45	1.69	1.96	1.97	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 1—continued

	In-sample		Jackknife		Normalized Std.Dev.	
	DF adjusted	Bayesian	KS	Out-of-sample	In/Out	Jackknife/Out
Panel C. Global minimum variance portfolio: Real data						
Sample covariance matrix	5.56	6.49	6.50	7.55	8.44	0.66*
Sample covariance matrix C	8.07				8.67	0.93
One-factor model	4.48		4.49		9.16	0.49*
One-factor model C	7.16				8.91	0.80*
Three-factor model	4.97		4.99		8.59	0.58*
Three-factor model C	7.52				8.77	0.86*
Equally-weighted portfolio	13.80				13.07	1.06
Panel D. Minimum tracking error variance portfolio: Real data						
Sample covariance matrix	0.94	1.10	1.10	1.28	1.46	0.64*
Sample covariance matrix C	0.95				1.42	0.67*
One-factor model	1.61		1.61		2.24	0.72*
One-factor model C	1.61				2.24	0.72*
Three-factor model	1.64		1.64		1.93	0.85*
Three-factor model C	1.64				1.93	0.85*
Equally-weighted portfolio	2.80				2.54	1.10

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, which are held for one month. This procedure is repeated every month. We compare the out-of-sample standard deviations of the optimal portfolios with the estimates of the out-of-sample standard deviations. 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 degree-of-freedom adjustment factors are computed using the formula in Section 2. The Bayesian covariance matrix estimators are described in Section 3. KS refer to the Kan and Smith (2005) estimate of out-of-sample standard deviation. Panels A and B are for simulated i.i.d. normal data (see Section 6.6.3). 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. The column “In/Out” (“Jackknife/Out”) means the ratio of in-sample standard deviation (Jackknife estimate of standard deviation) to out-of-sample standard deviation.

**Table 2**  
**Mean absolute difference**  
**from out-of-sample standard deviations**

	Jackknife	In-sample	DF adjusted	Bayesian	KS
Panel A. Global minimum variance portfolio: Simulated i.i.d. normal data					
Sample covariance matrix	0.851	1.669*	1.076*	1.071*	0.789
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 i.i.d. normal data					
Sample covariance matrix	0.263	0.490*	0.314*	0.313*	0.239
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 2—continued

	Jackknife	In-sample	DF adjusted	Bayesian	KS
Panel C. Global minimum variance portfolio: Real data					
Sample covariance matrix	2.29	3.06*	2.53	2.53	2.24
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	0.355
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			

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, which are held for one month. This procedure is repeated every month. We report the summaries of the difference between the out-of-sample standard deviations of the optimal portfolios and the estimates of the out-of-sample standard deviations. 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 degree-of-freedom adjustment factors are computed using the formula in Section 2. The Bayesian covariance matrix estimators are described in Section 3. KS refers to the Kan and Smith (2005) estimate of the out-of-sample standard deviation. Panels A and B are for simulated i.i.d. normal data (see Section 6.6.3). 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.



Table 3  
**Comparison between conditional out-of-sample variances  
and unconditional out-of-sample variances**

Panel A. Global minimum variance portfolios									
	Population		True	in-sample			Difference from True		
	3.20	3.73		0.986	jackknife	In-sample w/DF	In-sample w/2*DF	KS	
Sample covariance matrix			3.73	0.986	0.029	0.530	0.080	0.005	
Sample covariance matrix C			4.68	0.342	0.088				
One-factor model			3.37	0.310	0.042				
One-factor model C			4.62	0.257	0.085				
Three-factor model			3.27	0.143	0.038				
Three-factor model C			4.62	0.231	0.087				

  

Panel B. Global minimum variance portfolios: Other statistics									
	Population –		Jackknife	Absolute difference from True			Stdev of True		
	In-sample w/DF	In-sample w/DF		In-sample	In-sample w/ 2*DF	KS			
Sample covariance matrix	0.069	0.102	0.986	0.114	0.095	0.050			
Sample covariance matrix C		0.127	0.342			0.030			
One-factor model		0.071	0.310			0.027			
One-factor model C		0.118	0.257			0.021			
Three-factor model		0.068	0.143			0.010			
Three-factor model C		0.120	0.231			0.021			

Table 3—continued

Panel C. Minimum tracking error variance portfolios

	Population		True		Difference from True	
	In-sample	Jackknife	In-sample w/DF	In-sample w/2*DF	KS	KS
Sample covariance matrix	1.555	0.475	0.012	1.557	1.777	1.813
Sample covariance matrix C	1.759	0.410	0.012			
One-factor model	1.656	0.205	0.019			
One-factor model C	1.656	0.205	0.019			
Three-factor model	1.573	0.037	0.016			
Three-factor model C	1.573	0.037	0.016			

Panel D. Minimum tracking error variance portfolios: Other statistics

	Population –		Absolute difference from True		Stddev of True	
	In-sample w/DF	Jackknife	In-sample	In-sample w/2*DF	KS	KS
Sample covariance matrix	0.029	0.040	0.475	0.048	0.041	0.025
Sample covariance matrix C		0.038	0.410			0.018
One-factor model		0.034	0.205			0.005
One-factor model C		0.034	0.205			0.005
Three-factor model		0.032	0.037			0.003
Three-factor model C		0.032	0.037			0.003

In the end of April 1963, we select the largest 200 stocks from the CRSP NYSE/AMEX/NASDAQ universe. We use these stocks daily returns over the 196305-196804 period and the daily Fama-French factor returns to estimate the covariance matrix of these stocks, assuming a strict three-factor structure. Then we use these covariance structure together with the factor returns to simulate i.i.d. normal daily returns of 200 stocks. Each time, we simulate 36 months of daily returns and use these returns to estimate the covariance matrix and form the optimal portfolios. Then we examine the conditional and unconditional out-of-sample risks of these portfolios, as well as the in-sample, jackknife, and degree-of-freedom adjusted in-sample risks. This procedure is repeated 396 times and the summary statistics are reported. The letter C after each covariance matrix estimator denotes the optimal portfolio is constructed subject to nonnegativity constraint. “Population” means the population standard deviation of the population optimal portfolio; “True” means the population standard deviation of the sample optimal portfolio; “w/DF” means with degree-of-freedom correction; “w/2\*DF” means with twice of the DF correction; KS means the Kan and Smith (2005) estimate of out-of-sample standard deviation.

**Table 4**  
**Average standard deviations when real return data are used:**  
**Weighted-jackknife and dynamic factor models**

	Estimated			Out-of-sample			Normalized estimated		
	Weighted Jackknife	One-factor model	Three factor model	KS	Out-of-sample		Weighted Jackknife	One-factor model	Three factor model
					8.44	8.67			
<b>Panel A. Global minimum variance portfolio</b>									
Sample covariance matrix	8.09	5.36	5.44	7.55	8.44	0.96	0.64*	0.64*	0.64*
Sample covariance matrix C	8.33	7.22	7.63		8.67	0.96	0.83*	0.83*	0.88*
One-factor model	8.77	8.61	8.66		9.16	0.96	0.94	0.94	0.95
One-factor model C	8.55	7.86	8.25		8.91	0.96	0.88	0.88	0.93
Three-factor model	8.18	7.73	7.78		8.59	0.95	0.90	0.90	0.91
Three-factor model C	8.42	7.58	7.96		8.77	0.96	0.86	0.86	0.91
<b>Panel B. Minimum tracking error variance portfolio</b>									
Sample covariance matrix	1.36	0.93	0.93	1.28	1.46	0.92*	0.64*	0.64*	0.64*
Sample covariance matrix C	1.31	0.94	0.94		1.42	0.92*	0.66*	0.66*	0.66*
One-factor model	2.15	2.19	2.20		2.24	0.96	0.98	0.98	0.98
One-factor model C	2.15	2.19	2.20		2.24	0.96	0.98	0.98	0.98
Three-factor model	1.84	1.88	1.89		1.93	0.95	0.97	0.97	0.98
Three-factor model C	1.84	1.88	1.89		1.93	0.95	0.97	0.97	0.98

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, which are held for one month. This procedure is repeated every month. We compare the out-of-sample standard deviations of the optimal portfolios with the estimates of the out-of-sample standard deviations. 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. KS refers to the Kan and Smith (2005) estimate of the out-of-sample standard deviation. Standard deviations are annualized and expressed in percentage points. The last three columns are the ratios of the first three columns to the fourth column. The \* in the last two columns indicates the associated number is significantly smaller than one.

**Table 5**  
**Mean absolute difference from out-of-sample standard deviations**  
**when real return data are used**

	Weighted Jackknife	One-factor model	Three-factor model	KS
Panel A. Global minimum variance portfolio				
Sample covariance matrix	2.07	3.20*	3.12*	2.24
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*	0.355
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*	

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, which are held for one month. This procedure is repeated every month. We compare the out-of-sample standard deviations of the optimal portfolios with the estimates of the out-of-sample standard deviations. Results are based on 396 months of out-of-sample returns from 1967/5 to 2000/4. KS refers to the Kan and Smith (2005) estimate of out-of-sample standard deviations. 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.