On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model

Louis K. C. Chan
University of Illinois at Urbana-Champaign

Jason Karceski
University of Florida

Josef Lakonishok
University of Illinois at Urbana-Champaign and NBER

We evaluate the performance of models for the covariance structure of stock returns, focusing on their use for optimal portfolio selection. We compare the models' forecasts of future covariances and the optimized portfolios' out-of-sample performance. A few factors capture the general covariance structure. Portfolio optimization helps for risk control, and a three-factor model is adequate for selecting the minimum-variance portfolio. Under a tracking error volatility criterion, which is widely used in practice, larger differences emerge across the models. In general more factors are necessary when the objective is to minimize tracking error volatility.

The roots of modern investment theory can be traced back to Markowitz's (1952,1959) seminal idea that investors should hold mean-variance efficient portfolios. In the past, however, mean-variance portfolio optimization was not widely used. Instead, most investment managers focused on uncovering securities with high expected returns. At the same time, theoretical research on investments has concentrated on modeling expected returns. Similarly, empirical research focused on testing such equilibrium models, or documenting patterns in stock returns that appear to be inconsistent with these models.

Several trends suggest that professional investors are rediscovering the importance of portfolio risk management. There is mounting evidence that superior returns to investment performance are elusive. Numerous studies indicate that on average professional investment managers do not outperform passive benchmarks. In turn, the popularity of indexation [see, e.g., Chan and Lakonishok (1993)] has drawn attention to methods of optimally

We thank Ravi Jagannathan (the editor), Narasimhan Jegadeesh, Neil Pearson, Jay Ritter, and two anonymous referees for their comments. Address correspondence to Louis K. C. Chan, Department of Finance, College of Commerce and Business Administration, University of Illinois at Urbana-Champaign, Champaign, IL 61820. Partial computing support for this research was provided by an equipment grant from Intel Corporation.

The Review of Financial Studies Winter 1999 Vol. 12, No. 5, pp. 937–974

tracking a benchmark, especially when full replication of the benchmark is not desired or not practical. The recent interest in asset allocation methods, including international diversification, has also spurred interest in portfolio optimization. Another factor is the increased use of sophisticated quantitative methods in the investment industry, together with increased computing power. In short, there is an increased emphasis on risk control in the investment management industry.

Empirically, portfolio optimization can yield substantial benefits in terms of risk reduction, as the following simple experiment suggests. In order to abstract from the problems of predicting expected returns, suppose the task is to find the global minimum variance portfolio. Similarly, to sidestep for now the thorny issues of predicting return variances and covariances, the investor is assumed to have perfect foresight about the future values of these statistics. There are nonnegativity constraints on the weights since short selling is expensive for individual investors and not generally permissible for most institutional investors. Further, to ensure that the portfolio is diversified, the weight of a stock cannot exceed 2%. Every year this hypothetical investor selects the global minimum variance portfolio from a set of 250 stocks that are randomly selected from domestic NYSE and AMEX issues. The investor follows a buy-and-hold strategy for this portfolio over the next year. The experiment is repeated each year from 1973 to 1997 to give a time series of realized returns on the portfolio. The strategy yields a portfolio with a standard deviation of 6.85% per year. In comparison, a portfolio made up of all the 250 stocks (with equal amounts invested in each stock) yields a standard deviation that is more than twice as large (16.62% per year). Everything else equal, the optimized portfolio's lower volatility implies that it should have a Sharpe ratio (the ratio of excess return to standard deviation) that is more than double that of its equally weighted counterpart. To give an idea of how this might translate into returns, suppose that the return premium on stocks over the riskfree rate is expected to be 5% per year. Then levering up the optimized portfolio can result in a portfolio with the same volatility as the equally weighted portfolio, but with an expected return that is higher by 7%. Accordingly, there are large potential payoffs to portfolio risk optimization.

An additional boost to interest in optimization techniques stems from how performance is evaluated in the investment industry. While the theory of optimal portfolio choice suggests that investors should be concerned with the variance of the portfolio's return, in practice investment decisions are delegated to professional money managers. Since managers are evaluated relative to some benchmark, it has become standard practice for them to optimize with respect to tracking error volatility (the standard deviation of the difference between the portfolio's return and the benchmark return). Roll (1992) provides the analytics for this approach.

Given the increasing emphasis on risk management and its potential payoffs, there is a proliferation of portfolio optimization techniques. Yet there has been a shortage of scientific evidence evaluating the performance of different risk optimization methods. In particular, the benefits promised by portfolio optimization (either with respect to volatility or tracking error volatility) depend on how accurately the moments of the distribution of returns can be predicted. This article tackles one aspect of this issue, namely, how to forecast the second moments of returns. We start with the case where the hypothetical investor is concerned with the volatility of the portfolio. We compare the performance of different methods of forecasting variances and covariances, with an eye to judging which models improve our ability to optimize portfolio risk. The different risk models are evaluated on a statistical basis (mean absolute forecast errors and the correlations between forecast and realized values), and also on a more practical, economic basis (the realized volatility of optimized portfolios based on a particular forecast of the covariance matrix).

We focus on forecasting the second moments, rather than expected returns, for two reasons. First, several studies examine the importance of the forecasts of mean returns for mean-variance optimization [Michaud (1989), Best and Grauer (1991), Chopra and Ziemba (1993), Winston (1993)]. There is a general consensus that expected returns are notoriously difficult to predict, and that the optimization process is very sensitive to differences in expected returns. At the same time, there is a common impression that return variances and covariances are much easier to estimate from historical data [Merton (1980), Nelson (1992)]. Possibly, then, the second moments pose fewer problems in the context of portfolio optimization. Our results suggest that while future covariances are more easily predictable than future mean returns, the difficulties should not be understated. To illustrate, we replicate the previous exercise on portfolio optimization without the assumption of perfect foresight. Instead, we use the past historical covariances and variances (known as the portfolio formation date) as estimates of the future moments. In this more realistic setting the optimized portfolio's standard deviation is 12.94% per year. While optimization leads to a reduction in volatility (relative to an equally weighted portfolio, where the standard deviation was 16.62% per year), the problem of forecasting covariances poses a challenge.

Second, we bring to bear on the issue of forecasting covariances the considerable literature on the sources of return covariation. One interpretation of these sources is that they represent common risk factors, and expected returns provide compensation for bearing such risks. In this light, our results help to validate different models of the sources of systematic risk. In other words, if a factor does not help to predict return covariation, then it is less plausible that such a factor represents a source of risk that is priced. In this regard, our work is in the same spirit as Daniel and Titman (1997), who tackle the issue from the standpoint of expected returns.

Evidence on the efficacy of different models of return covariances is presented by Cohen and Pogue (1967), Elton and Gruber (1973), Elton, Gruber, and Urich (1978), and Alexander (1978). Given the state of optimization techniques and computational technology at that time, these articles examined only a very small set of stocks over short time periods. More importantly, they generally predate the large body of work on multifactor pricing models, so they examine only a limited set of risk models. A related literature predicts stock market volatility [see, e.g., Pagan and Schwert (1990) and Schwert and Seguin (1990)], but does not generally investigate its implications for the important issue of portfolio risk optimization. In recent work, Ledoit (1997) considers in detail a shrinkage estimator of covariances and applies it to portfolio optimization.

Our tests are predictive in nature: we estimate sample covariances over one period and then generate out-of-sample forecasts. The results can be summarized as follows. The future return covariance between two stocks is predictable from current attributes such as the firms' market capitalizations, market betas, and book-to-market ratios. Such models generate a slight improvement in our ability to predict future covariances compared to forecasts based on historical covariances. Introducing additional factors, however, does little to improve forecasting performance. A particularly sobering result is that in all our models the correlation between predicted and future covariances is not large. For instance, the correlation between past and future sample covariances is 34% at the 36 month horizon and much less (18%) at the 12-month horizon.

Since the models' covariance forecasts move in the same direction as the realized covariances, they help for portfolio risk optimization (as long as we impose constraints to limit the impact of estimation errors). The optimization exercises confirm that some form of portfolio optimization lowers portfolio volatility relative to passive diversification. However, they also confirm that more complicated models do not outperform (in terms of lowering portfolio volatility) simpler models.

Extending our analysis to the case where the investor is concerned with tracking error volatility, sharper distinctions arise between the different covariance forecasting models. Intuitively, the focus shifts to identifying those combinations of stocks that align most closely with the benchmark's risk exposures or attributes. As long as these exposures are attainable given the sample of stocks, the problem is substantially simplified relative to the task of identifying the global minimum variance portfolio. In this respect, the tracking error criterion is less susceptible to errors from forecasting the volatilities and covariances of the factors. Accordingly, the tracking error volatility problem (which is the most commonly encountered in practice) highlights much more dramatically the issue of how best to optimize. In general, we find that adding information from more factors helps to reduce tracking error volatility. Moreover, an approach which works well is one that circumvents the measurement errors in estimating factor loadings, but directly matches the benchmark portfolio along a number of attributes.

The remainder of this article proceeds as follows. Section 1 provides some background to our study by emphasizing the severe problems that we shall face in predicting return covariances. Section 2 outlines some of the different models we examine, and Section 3 provides evidence on their forecast performance. The results of our optimization exercises with respect to total variance are reported in Section 4. Extensions to the problem of minimizing tracking error volatility relative to several benchmarks are provided in Section 5. Section 6 concludes the article.

1. The Behavior of Stock Return Variances and Covariances

To set the stage for our analysis, we first provide some evidence on the structure of return variances and covariances. Table 1 reports the distribution of sample variances, covariances, and correlations of monthly returns on three sets of stocks. Each set is drawn in April of every year from 1968 to 1998, and sample moments of returns are calculated from the preceding 60 months. The distribution is based on the estimated statistics pooled across all years.

The first sample (panel A) comprises 500 stocks selected from the population of domestic common stock issues on the NYSE and AMEX. Closed-end funds, Real Estate Investment Trusts, trusts, American Depository Receipts, and foreign stocks are excluded. To mitigate the problems associated with low-priced stocks, only stocks with prices greater than $5 are included. The average monthly stock return variance is 0.0098, corresponding to an annualized standard deviation of about 34.29%. The average pairwise correlation is 0.28, indicating that there are potentially large payoffs to portfolio diversification.

Panels B and C check whether the second moments of returns are related to firm size. In panel B, the sample comprises all NYSE and AMEX firms with equity market capitalization in excess of the 80th percentile of the size distribution of NYSE firms. Panel C examines all NYSE and AMEX firms ranked below the 20th percentile of the NYSE size distribution. On average, small stocks display return variances that are almost three times those of large stocks. The average variance for small firms is 0.0181 (equivalent to an annualized standard deviation of 46.60%) compared to 0.0067 for large firms (or an annualized standard deviation of 28.35%). Small stocks also tend to exhibit higher average pairwise covariances compared to large stocks (0.0042 and 0.0021, respectively). However, the average correlation between small stocks is only 24%, whereas the average correlation between large stocks is 33%.

Table 1
Distributions of variances, covariances, and correlations of returns on sample stocks

	Variances	Covariances	Correlations
Panel A: 500 random stocks			
Mean	0.0098	0.0026	0.2801
Standard deviation	0.0063	0.0019	0.1477
Minimum	0.0013	−0.0062	−0.3767
25th percentile	0.0054	0.0013	0.1818
Median	0.0083	0.0023	0.2845
75th percentile	0.0126	0.0036	0.3824
Maximum	0.0474	0.0214	0.9196
Panel B: Large stocks			
Mean	0.0067	0.0021	0.3300
Standard deviation	0.0041	0.0013	0.1529
Minimum	0.0014	−0.0031	−0.3137
25th percentile	0.0042	0.0012	0.2298
Median	0.0058	0.0019	0.3353
75th percentile	0.0080	0.0028	0.4350
Maximum	0.0336	0.0144	0.8864
Panel C: Small stocks			
Mean	0.0181	0.0042	0.2438
Standard deviation	0.0140	0.0032	0.1381
Minimum	0.0011	−0.0128	−0.3767
25th percentile	0.0096	0.0021	0.1508
Median	0.0149	0.0037	0.2467
75th percentile	0.0225	0.0057	0.3397
Maximum	0.1241	0.0463	0.9175

At the end of April of each year from 1968 to 1998 three samples of stocks are selected from eligible domestic common stock issues on the NYSE and AMEX. For each set of stocks, sample variances, pairwise covariances, and correlations are calculated based on monthly returns over the prior 60 months. Summary statistics are based on the estimated values pooled over all years. In panel A, 500 stocks are randomly selected each year. In panel B, the sample of stocks includes NYSE and AMEX domestic primary firms with equity market capitalization above the 80th percentile of the size distribution of NYSE firms. In panel C, the sample of stocks includes NYSE and AMEX domestic primary firms with equity market capitalization below the 20th percentile of the size distribution of NYSE firms.

Further, the distributions of the estimated statistics show substantially larger dispersion within the group of small firms. Sample variances, for instance, range from 0.0011 to 0.1241 for small firms, while the range for large firms is from 0.0014 to 0.0336. Similarly, the covariances for small firms run from −0.0128 to 0.0463 and extend from −0.0031 to 0.0144 for large firms. The interquartile spreads of these statistics confirm the differences between small and large firms.

Conventional wisdom suggests that two stocks in the same industry are more highly correlated than two stocks in different industries, since they are likely to be affected by common events. Table 2 checks up on this intuition. Each year we classify stocks into one of 48 industries using the industry definitions from Fama and French (1997). Correlations are averaged across

Table 2
Average correlations of individual stock returns for selected industries

Panel A: All stocks

Industry	Average correlation within industry	Average correlation with firms in all other industries	Difference
Chemicals (48)	0.3579	0.3019	0.0560
Construction Materials (69)	0.3427	0.2970	0.0457
Machinery (86)	0.3338	0.2895	0.0443
Petroleum & Natural Gas (71)	0.3646	0.2283	0.1363
Utilities (133)	0.3870	0.2186	0.1684
Business services (41)	0.2913	0.2799	0.0114
Electronic equipment (54)	0.3695	0.2931	0.0764
Wholesale (58)	0.2954	0.2799	0.0155
Retail (100)	0.3260	0.2773	0.0487
Banking (61)	0.4108	0.2925	0.1183
Insurance (23)	0.3921	0.2975	0.0946

Panel B: Large stocks

Industry	Average correlation with other large firms within industry	Average correlation with large firms in all other industries	Difference	Average correlation with small firms within industry
Chemicals (21)	0.4194	0.3392	0.0802	0.3395
Construction Materials (25)	0.3920	0.3323	0.0597	0.3337
Machinery (33)	0.3917	0.3242	0.0675	0.3227
Petroleum & Natural Gas (28)	0.4228	0.2428	0.1800	0.3362
Utilities (68)	0.4570	0.2417	0.2153	0.3697
Business services (15)	0.3380	0.3142	0.0238	0.2801
Electronic equipment (18)	0.4222	0.3225	0.0997	0.3509
Wholesale (19)	0.3235	0.3086	0.0149	0.2900
Retail (40)	0.3925	0.3116	0.0809	0.3045
Banking (29)	0.4915	0.3269	0.1646	0.3705
Insurance (12)	0.4388	0.3290	0.1098	0.3616

Panel C: Small stocks

Industry	Average correlation with other small firms within industry	Average correlation with small firms in all other industries	Difference	Average correlation with large firms within industry
Chemicals (26)	0.3206	0.2764	0.0442	0.3395
Construction Materials (44)	0.3178	0.2753	0.0425	0.3337
Machinery (53)	0.3052	0.2706	0.0346	0.3227
Petroleum & Natural Gas (43)	0.3246	0.2209	0.1037	0.3362
Utilities (65)	0.3394	0.2029	0.1365	0.3697
Business services (26)	0.2588	0.2560	0.0028	0.2801
Electronic equipment (36)	0.3397	0.2746	0.0651	0.3509
Wholesale (39)	0.2741	0.2586	0.0155	0.2900
Retail (60)	0.2741	0.2512	0.0229	0.3045
Banking (33)	0.3366	0.2630	0.0736	0.3705
Insurance (11)	0.3337	0.2771	0.0566	0.3616

At the end of April of each year from 1968 to 1998 eligible domestic common stock issues on NYSE and AMEX are classified into 1 of 48 industry groups, based on the definitions of Fama and French (1997). Pairwise correlations of monthly returns between stocks in each industry are estimated from the most recent 60 months of data and then averaged across all stocks in the same industry. The average pairwise correlation between stocks in the same industry and all other stocks is also calculated. The means over all years of these average correlations are reported in this table for the 11 largest industries (in terms of equity market capitalization and the number of constituent stocks). In panels B and C, firms within these selected industries are further classified into two groups: small companies in an industry are firms with equity market capitalization below the median size of NYSE firms in the same industry, and large

Table 2
(continued)

companies in an industry are firms with equity market capitalization above the median for NYSE firms in that industry. Pairwise correlations are averaged within small firms in the same industry, within large firms in the same industry, between large and small firms in the same industry and between small and large firms in an industry and same-size firms in all other industries. Means across all years are reported in panel B for large firms and in panel C for small firms. The average number of firms in each industry is reported in parentheses.

all pairs of stocks from the same industry and across all pairs of stocks from different industries. To reduce clutter, we present results only for the industries with the highest market capitalization and with the most firms.

Panel A of Table 2 indicates that the correlation between two stocks is on average larger when they are from the same industry than when they belong to different industries. The difference is particularly striking for stocks in the utility, petroleum, banking, and insurance industries. This may be taken as evidence that these particular industries tend to be more homogeneous groupings.

In the remaining panels of Table 2 we partition firms within an industry into two sets. Large stocks (panel B) have market capitalization exceeding the median size of NYSE firms in the industry. We calculate the average pairwise correlations between a large stock and large stocks in the same industry, large stocks in all other industries, and small stocks in the same industry.

The differences between the within-industry and across-industry correlations stand out even more strongly for large firms. In the case of large utility stocks, for example, the average within-industry correlation is 0.4570, while the average correlation with large stocks in all other industries is 0.2417, yielding a difference of 0.2153. As further evidence of the strong tendency for large stocks to move together, note that in many cases the correlation between large stocks across different industries is as high as the within-industry correlation between large and small stocks.

Panel C repeats the exercise for small stocks (with market capitalization below the NYSE median in an industry). Here the results reinforce the conclusions from Table 1 about the variation across small stocks. Compared to large stocks, the correlations between small firms are lower, even when they share the same industry affiliation. The difference between the within-industry and across-industry average correlation is also less sharp. Perhaps most tellingly, on average a small stock shows less correlation with another small stock in the same industry than with a large stock in the same industry. For example, the average correlation between two small utility stocks is 0.3394, while the average correlation between a small utility stock and a large utility stock is 0.3697.

The upshot of this section is that there is good news and bad news. The good news is that stock return variances and covariances display some structure. There is encouraging evidence that certain characteristics, such as firm size and industry affiliation, tell us something about the variances and covariances of returns. The bad news is that this information may be useful only in a minority of cases. As standard textbook discussions suggest, the number of covariance terms far outweighs the number of variance terms. There are also far more pairwise combinations of stocks drawn from different industries than from the same industry. As a result, the patterns documented in this section may not help in terms of predicting most elements of the variance-covariance matrix.

2. Predicting Stock Return Covariances

Given our focus on portfolio risk optimization, we concentrate on directly predicting return covariances (forecasts of variances are discussed in a subsequent section). We use a number of different forecasting models. Each model is applied to the returns on a sample of stocks drawn from domestic common equity issues listed on the NYSE and AMEX. Given our earlier evidence on the noise in estimating covariances for small stocks, we exclude stocks that fall in the bottom 20% of market capitalization based on NYSE breakpoints. Stocks trading at prices less than $5 are also excluded. From the remaining stocks we randomly select 250 in April of each year from 1968 to 1997. Forecasts are generated based on the past 60 months of prior data (the estimation period).

2.1 Full Covariance Matrix Forecasts

The starting point for forecasting return covariances is given by the sample covariances based on the estimation period. For example, the sample covariance between stocks i and j is given by

$$\mathrm{cov}_{ij} = \frac{1}{59}\sum_{k=1}^{60}(r_{it-k} - \bar{r}_i)(r_{jt-k} - \bar{r}_j) \tag{1}$$

where r_{it} is the return in excess of the monthly Treasury bill return for stock i in month t and $\bar{r}_i$ is the sample mean excess return. The sample covariances, however, are very sensitive to outlier observations [see Huber (1977)].

2.2 Covariance Forecasts from Factor Models

Forecasts from the full covariance matrix model may reflect firm-specific events that happen to affect several stocks at the same time, but which are not expected to persist in the future. An alternative approach is to strip out the idiosyncratic components of the covariance by introducing pervasive factors that drive returns in common. One such formulation is given by the strict factor model of security returns:

$$r_{it} = \beta_{i0} + \sum_{j=1}^{K} \beta_{ij} f_{jt} + \epsilon_{it}. \tag{2}$$

Here f_{jt} is the jth common factor at time t, and ϵ_{it} is a residual term. The coefficient β_{ij} gives the loadings, or sensitivities, of stock i on each of the K factors. Assuming that these factors are uncorrelated with the residual return and that the residual returns are mutually uncorrelated, the covariance matrix V of the returns on a set of N stocks is given by

$$V = B\Omega B' + D, \tag{3}$$

where B is the matrix of factor loadings of the stocks, Ω is the covariance matrix of the factors, and D is a diagonal matrix containing the residual return variances.

We use a variety of factor models. In each case we use the 60 months of data over the estimation period to obtain the factor loadings of a stock. The factors are measured as the returns on mimicking portfolios, as in Chan, Karceski, and Lakonishok (1998). The mimicking portfolio returns over the prior 60 months also provide estimates of the covariance matrix of the factors, loadings, and residual variances that are the inputs to Equation (3).

Results are reported for the following factor models [see Chan, Karceski, and Lakonishok (1998) for full details]. A *one-factor model* uses the excess return on the value-weighted market index as the single factor. This model corresponds to the standard CAPM or single-index model. The *three-factor model* augments the value-weighted market index with size and book-to-market factors. This model has been proposed by Fama and French (1993). The remaining models add factors that have been found to work well in capturing stock return covariation [see Chan, Karceski and Lakonishok (1998)]. Specifically, a *four-factor model* includes, along with the three Fama and French factors, a momentum factor. This latter factor captures the tendency for stocks with similar values of past 6-month return (measured over the period from 7 months to 1 month ago) to behave similarly over the future with respect to their returns. An *eight-factor model* consists of the market factor as well as factors associated with firm size, book-to-market, momentum, dividend yield, cash flow yield, the term premium, and the default premium. The *10-factor model* comprises these factors along with the beta factor (the return spread between a portfolio of high-beta stocks and a portfolio of low-beta stocks) and a long-term technical factor (based on stocks' cumulative returns measured over the period from 60 months to 12 months ago).

2.3 Forecasts from a Constant Covariance Model

This forecasting model assumes that all pairwise covariances between stocks are identical. We estimate the constant covariance, $\overline{\text{cov}}$, as the simple mean across all pairwise stock return covariances from the estimation period. This model can be thought of as a version of a James–Stein estimator which shrinks each pairwise covariance to the global mean covariance (while giving no weight otherwise to the specific pair of stocks under consideration). As one motivation for this approach, the noise in stock returns and the resulting estimation error suggest that it may be unwise to make distinctions between stocks on the basis of their sample covariances. Instead, it may be more fruitful to assume that all stocks are identical in terms of their covariation.[1]

3. Empirical Results

3.1 Covariance Forecasts

At the end of April in every year of the sample period, each model's forecasts are compared to the sample covariances realized over a subsequent period. We report the results from two experiments. Realized covariances are measured over the subsequent 12 months (the test period) in the first experiment, and over the subsequent 36 months in the second experiment. The lengths of the test periods are chosen to correspond to realistic investment horizons. Since the forecasts are generated using a period disjoint from the test periods, our tests are predictive in nature.

Panel A of Table 3 provides summary statistics on the forecasted covariances from each model. Forecasts from the historical full covariance model display the largest standard deviation (0.18%) of all the models, suggesting that straightforward extrapolation from the past may be overly accommodative of the data. In comparison, the factor models tend to smooth out the covariances, yielding less extreme forecasts. The standard deviation of forecasts from a one-factor model is 0.14%, while the multifactor models share similar standard deviations of about 0.16%. Since the historical full covariance model is a limiting case where there are as many factors as stocks, the general impression is that the high-dimensional factor models are prone to making overly bold predictions. As the dimensionality of the factor model grows, there is an increasing chance that the model snoops the data, resulting in overfitting.

The remaining panels of Table 3 compare each model's forecasts with sample covariances estimated over the subsequent 12 months (panel B) or over the subsequent 36 months (panel C). Forecast performance is first evaluated in terms of the absolute difference between the realized and forecast values. The single-factor model turns in the lowest mean and median absolute error. Since it is the most conservative of the factor models in terms of the dispersion of its forecasts, the one-factor model is apparently not penal-

[1] Frost and Savarino (1986) and Jobson and Korkie (1981) provide evidence that using the common sample mean return as the expected return for each stock improves the out-of-sample performance of optimized portfolios relative to assuming that historical average returns will persist.

Table 3
Performance of covariance forecasting models

Panel A: Properties of forecasted covariances

Model	Mean	Standard deviation	Minimum	5-th percentile	95-th percentile	Maximum
Full covariance	0.0027	0.0018	−0.0051	0.0003	0.0061	0.0211
1-factor	0.0025	0.0014	−0.0001	0.0007	0.0051	0.0130
3-factor	0.0026	0.0015	−0.0017	0.0006	0.0055	0.0150
4-factor	0.0026	0.0016	−0.0018	0.0006	0.0055	0.0151
8-factor	0.0026	0.0016	−0.0028	0.0006	0.0056	0.0159
10-factor	0.0026	0.0016	−0.0032	0.0005	0.0057	0.0162
Average covariance	0.0027	0.0	0.0027	0.0027	0.0027	0.0027

Panel B: Forecast performance based on subsequent 12 months

Model	Absolute forecast error Mean	Median	95-th percentile	Maximum	Correlation	Slope
Full covariance	0.0040	0.0028	0.0115	0.1631	0.1792	0.3589
1-factor	0.0037	0.0026	0.0107	0.1409	0.1643	0.5435
3-factor	0.0038	0.0027	0.0110	0.1461	0.1994	0.4868
4-factor	0.0038	0.0027	0.0110	0.1462	0.1987	0.4808
8-factor	0.0038	0.0027	0.0111	0.1518	0.1963	0.4573
10-factor	0.0038	0.0027	0.0112	0.1535	0.1962	0.4488
Average covariance	0.0039	0.0030	0.0105	0.1403	0.0000	0.0000

Panel C: Forecast performance based on subsequent 36 months

Model	Absolute forecast error Mean	Median	95-th percentile	Maximum	Correlation	Slope
Full covariance	0.0019	0.0016	0.0051	0.0224	0.3394	0.3658
1-factor	0.0018	0.0014	0.0046	0.0215	0.3416	0.4863
3-factor	0.0018	0.0015	0.0047	0.0214	0.3590	0.4599
4-factor	0.0018	0.0015	0.0047	0.0214	0.3583	0.4566
8-factor	0.0018	0.0015	0.0048	0.0211	0.3593	0.4405
10-factor	0.0019	0.0015	0.0048	0.0211	0.3599	0.4340
Average covariance	0.0019	0.0017	0.0042	0.0217	0.0000	0.0000

At the end of April of each year from 1973 onward a random sample of 250 firms is drawn from eligible domestic common stock issues on the NYSE and AMEX. Forecasts of monthly return covariances are generated from seven models, based on the prior 60 months of data for each stock. Summary statistics for the distribution of forecasted values are reported in panel A. Forecasts are then compared against the realized sample covariances estimated over the subsequent 12 months (panel B) and over the subsequent 36 months (panel C). The last estimation period ends in 1997 in panel B, and ends in 1995 in panel C. Summary statistics are provided for the distribution of the absolute difference between realized and forecasted values of pairwise covariances. Also reported is the Pearson correlation between forecasts and realizations, and the slope coefficient in the regression of realizations on forecasts.

The full covariance model uses the return covariance estimated over the most recent past 60 months prior to portfolio formation as the forecast. Covariance forecasts from the factor models are based on Equation (3) in the text. The one-factor model uses the excess return on the value-weighted CRSP index over the monthly Treasury–bill rate as the factor. The three-factor model includes the excess return on the value-weighted index as well as size and book-to-market factors. The four-factor model includes these three as well as a momentum factor (based on the rate of return beginning 7 months and ending 1 month before portfolio formation). The eight-factor model uses as factors the market, size, book-to-market, momentum, dividend yield, cash flow yield, the term premium and the default premium. The 10-factor model includes these as well as the mimicking portfolio for the beta factor and a long-term technical factor (based on the rate of return beginning 60 months and ending 12 months before portfolio formation). In the average covariance model the forecast is the average across all pairwise covariances of stocks in the sample.

ized as heavily as the other models. By the same token, while the average covariance model has a large mean absolute error, it nonetheless does not do disastrously. Its performance is on a par with the full covariance model, which has the highest mean absolute forecast error (the median absolute error gives roughly the same ordering). The differences across the models' performances are slight, but the striking message is that more factors do not necessarily give rise to smaller forecast errors.

The last two columns of panel B provide additional measures of forecast performance. Specifically, we regress the realized values on the predicted values and recover the slope coefficient and the correlation coefficient from the regression. While the mean absolute error criterion penalizes a model for making overly bold predictions, the correlation coefficient focuses more on whether the predictions tend to be in the same direction as the realizations.[2] As the high-dimensional models are more likely to overfit the data, the slope coefficients on their forecasts are more attenuated relative to the simpler models. For example, the slope coefficient from the 1-factor model is 0.5435, compared to 0.3589 for the full covariance model and 0.4488 for the 10-factor model. The correlations suggest there is little distinction between a 3-factor and a 10-factor model. The correlations for the full covariance and one-factor models stand apart from the others, but for different reasons. In the case of the one-factor model, the correlation is only 0.1643, suggesting that exposures to market risk do not fully capture variation in realized covariances. For the full covariance model, the large dispersion of its forecasts pull down the correlation to 0.1792.

When the forecasts are compared to covariances realized over the subsequent 36-month test period (panel C), the average absolute forecast errors are reduced. For instance, the mean absolute error for the full covariance model is 0.0040 for the 12-month test period, compared to 0.0019 for the 36-month test period. The drop in forecast errors suggests that there is a lot of noise in covariances measured over a period as short as 12 months. As additional confirmation, the correlations between forecasts and realizations are higher over the 36-month test period than over the 12-month test period. Nonetheless, as in panel B, extending the number of factors beyond a relatively small set does not lead to superior forecasting performance.[3]

3.2 Additional Results

In additional, unreported work (available from the authors upon request) we extend our results in two directions: forecasting correlations, and modifying

[2] Using Spearman rank correlations between forecasts and realizations does not alter the main findings.

[3] To help in assessing the magnitude of the forecast errors in panel C, we also generate forecasts from a randomized model. That is, the forecast of the future 36-month covariance between two stocks is given by the historical covariance between a randomly selected pair of stocks. This randomized model yields a mean absolute error of 0.0030.

the factor models. When we repeat the forecasting exercises for correlations, we find that the relative performance of the different models is quite similar to their performance in forecasting covariances. The various factor models differ only slightly in terms of mean absolute errors. In particular, the constant correlation model actually generates the lowest mean absolute errors in forecasting correlations. In general the results suggest that it is harder to predict correlations than covariances. For the 36-month test period, for example, the correlation between past and future correlations averages 24% across the models, compared to an average correlation of about 34% between past and future covariances.[4]

The overall verdict from Table 3 is that the various models for forecasting covariances generally perform quite similarly. The factor models yield somewhat smaller absolute errors than the full covariance model, but the forecast errors provide little discrimination between a 3-factor model and a 10-factor model. Indeed, assuming that all pairwise covariances are constant would not lead to much worse performance. Yet, as the correlations in Table 3 suggest, the factors are able to capture the direction in which realized covariances vary with beta, size, and other risk exposures. Why then isn't there a more appreciable improvement from models with multiple factors? Factor (or index) models are so widely used in financial research and investment practice that we feel compelled to dig deeper and examine what implementation aspects may be hurting the models' performance. In additional unreported work we explore two such aspects: the timeliness of stocks' estimated exposures, and the functional form implied by the factor model in Equation (3). In general, we find that modifications to address these potential problems do not alter our conclusions.

3.3 Variance Forecasts

We also apply several models (described in the appendix) to forecast return variances. The results for forecasting variances (Table A.1), when placed alongside the results for covariances, suggest that variances are relatively more stable, and hence easier to predict, than covariances. Past and future return covariances (measured over the subsequent 36 months) have a correlation of 33.94% (panel C of Table 3), while the same correlation for variances is 52.25%. As in our earlier results, however, higher dimensional models do not necessarily raise forecasting power.

In terms of forecasting return covariances and variances, our bottom line is as follows. There is some stable underlying structure in return covari-

[4] Using additional factors based on size or market beta may help less for forecasting correlations than for forecasting covariances due to offsetting effects in predicting covariances and variances. In Table 1, for example, return covariances are generally decreasing with firm size, while return variances also decrease with firm size. As a result, what might otherwise be a strong relation between firm size (or other factors) and covariances is dampened when the covariances are scaled by standard deviations.

ances. Factor models help to improve forecast power, but there is little to distinguish between the performance of a 3-factor model and a 10-factor model. Modifications of the factor model structure, such as relaxing the model's linear structure or having more timely information on loadings and attributes, do not salvage the models. The situation improves somewhat when it comes to forecasting variances. Future variances are relatively more predictable from past variances, and so the models' forecast power is relatively stronger.

4. Applying the Forecasts: Portfolio Optimization

4.1 The Global Minimum Variance Portfolio

An important reason for forecasting the variances and covariances of returns is to provide inputs into the portfolio mean-variance optimization problem. The perils of forecasting expected returns are well known. In terms of forecasting the second moments, our results from the previous sections suggest that there are relatively minor differences between the various models' performance. In this sense the choice of a particular model for the second moments may matter less from the standpoint of optimization. The weights for an optimized portfolio (with constraints imposed) are complicated functions of the forecasts, however. It is thus not straightforward to assess the economic impact of errors in forecasting the second moments. In this section we report the results of several portfolio optimization exercises. These let us judge how the models' forecasting performance translates into the variance of the optimized portfolio's returns. Since we place constraints on the portfolio weights, the optimization exercises also let us tame the occasional bold forecasts from some of the models. In this respect a more meaningful comparison across the models can emerge. From both the technical and practical standpoint, therefore, the optimization experiments provide perhaps the most important metric for evaluating the models.

The setup of our optimization experiments is as follows. To highlight the role of the second moments, our goal is to form the global minimum variance portfolio (any other point on the efficient frontier would put some emphasis on forecasts of expected returns). In April of each year from 1973 to 1997, we randomly select 250 stocks listed on the NYSE and AMEX.[5] We use our different models to forecast future variances and covariances for these stocks' returns. These forecasts are the inputs to a quadratic programming routine. In order to make the experiments correspond as closely as possible to actual practice, several other constraints are imposed. Portfolio weights are required to be nonnegative, since short selling is not generally

[5] As in the earlier sections, we consider only domestic common equity issues above the second decile of market capitalization based on NYSE breakpoints, and with prices greater than $5.

Table 4
Performance and characteristics of minimum variance portfolios based on forecasting models

Panel A: Performance of portfolios

Model	Mean	Standard deviation	Sharpe ratio	Tracking error	Correlation with market	Average number of stocks with weights above 0.5%
(1) Full covariance	0.1554	0.1294	0.6405	0.0739	0.8764	53
(2) 1 factor	0.1610	0.1280	0.6907	0.0926	0.7972	54
(3) 3 factor	0.1569	0.1266	0.6659	0.0877	0.8197	54
(4) 9 factor	0.1529	0.1292	0.6224	0.0772	0.8638	54
(5) Product of standard deviations	0.1571	0.1263	0.6693	0.0879	0.8188	55
(6) Industry, size	0.1612	0.1281	0.6919	0.0852	0.8309	61
(7) Combination	0.1560	0.1259	0.6624	0.0841	0.8358	54
(8) 250 stocks, value-weighted	0.1431	0.1554	0.4539	0.0304	0.9807	45
(9) 250 stocks, equally-weighted	0.1727	0.1662	0.6027	0.0616	0.9287	0

Panel B: Characteristics of portfolios

Model	Beta (Rank)	LogSize (Rank)	BM (Rank)	DP (Rank)	Percent invested in: SIC 35, 36	SIC 49
(1) Full covariance	0.5533	20.46	0.7982	0.0626	3.82	46.66
	(1.76)	(7.48)	(5.84)	(8.04)		
(2) 1 factor	0.5071	20.36	0.8491	0.0643	1.97	57.85
	(1.28)	(7.40)	(6.20)	(8.52)		
(3) 3 factor	0.5141	20.36	0.8234	0.0642	2.41	55.74
	(1.24)	(7.40)	(5.88)	(8.40)		
(4) 9 factor	0.5357	20.38	0.8106	0.0625	3.83	47.76
	(1.56)	(7.48)	(5.96)	(8.08)		
(5) Product of standard deviations	0.5948	20.81	0.8118	0.0652	3.20	57.96
	(2.04)	(8.20)	(5.88)	(8.56)		
(6) Industry, size	0.6490	20.37	0.8258	0.0674	0.46	61.09
	(2.28)	(7.32)	(6.12)	(8.52)		
(7) Combination	0.5330	20.40	0.8178	0.0654	2.55	53.82
	(1.40)	(7.40)	(6.04)	(8.36)		
(8) 250 stocks, value-weighted	0.9897	22.34	0.5844	0.0408	13.20	8.66
	(4.60)	(10.0)	(3.88)	(6.32)		
(9) 250 stocks, equally-weighted	1.0686	20.28	0.7593	0.0413	11.02	15.31
	(5.20)	(7.12)	(5.24)	(6.32)		

At the end of April of each year from 1973 through 1997 a random sample of 250 firms is drawn from eligible NYSE and AMEX domestic common stock issues. Forecasts of covariances and variances of monthly excess returns (over the monthly Treasury bill rate) are generated from different models, using the prior 60 months of data for each stock. Based on each model's forecasts of variances and covariances, a quadratic programming procedure is used to find the global minimum variance portfolio. The portfolio weights are constrained to be nonnegative and not larger than 2%. These weights are then applied to form buy-and-hold portfolio returns until the next April, when the forecasting and optimization steps are repeated and the portfolios are reformed. For each procedure, summary statistics are presented in panel A for the annualized mean return and annualized standard deviation for the returns realized on the portfolio; the annualized average Sharpe ratio (the mean return in excess of the Treasury bill rate, divided by the standard deviation); the correlation between the monthly portfolio return and the return on the S&P 500 index; the annualized tracking error (standard deviation of the portfolio return in excess of the S&P 500 return); and the average number of stocks each year with portfolio weights above 0.5%. Panel B reports average characteristics of stocks in each portfolio: the beta relative to the value-weighted CRSP index (based on the 60 months of data prior to portfolio formation); market value of equity (in natural logarithms); book-to-market equity ratio (denoted BM); and dividend yield (denoted DP). Each characteristic is also measured as a decile ranking (from 1, the lowest, to 10, the highest). Also reported is the average proportion of the portfolio invested in firms with two-digit SIC codes of 35 and 36 (Industrial, Commercial Machinery, Computer Equipment and Electrical Equipment excluding Computers) and in firms with a two-digit SIC code of 49 (Electric, Gas and Sanitary Services). Each characteristic is measured as of the portfolio formation date and averaged across all portfolio formation years.

Table 4
(continued)

The full covariance model (model 1) uses the return covariance estimated over the most recent past 60 months prior to portfolio formation as the forecast. Covariance forecasts from the factor models (models 2–4) are based on Equation (3) in the text. The one-factor model uses the return on the value-weighted CRSP index as the factor. The three-factor model includes the return on the value-weighted index as well as size and book-to-market factors. The nine-factor model uses as factors the market, size, book-to-market, momentum, dividend yield, cash flow yield, the term premium, the default premium, and the second principal component. Model 5 is based on a regression model that uses the most recent past 8 years of data. Return covariances between two stocks are measured over the most recent 3 years and are regressed on the product of the standard deviations of the stocks' returns measured over the earliest 5 years. The estimated model is used to generate the covariance forecasts. In model 6, stocks are assigned to an industry-size group. There are 48 industries, as in Fama and French (1997). Each industry is divided into a set of large firms (with market value of equity exceeding the median capitalization of NYSE firms in that industry) and a set of small firms (with market value of equity below the NYSE median in that industry). The covariance forecast for any two stocks is given by the average of all pairwise covariances between stocks in their respective industry-size groups, based on the most recent past 60 months. Model 7 uses an equally weighted average of the forecasts from four models: the full covariance model, the three-factor model, the nine-factor model, and the industry-size model. Models 8 and 9 are the value-weighted and equally weighted portfolios, respectively, of all 250 stocks available at the portfolio formation date.

a common practice for most investors. We also impose an upper bound of 2% on the portfolio weights, so as to mitigate the effects of forecast errors.[6] Given the optimized weights, we calculate buy-and-hold returns on the portfolio for the subsequent 12 months, at the end of which the forecasting and optimization procedures are repeated. The resulting time series of monthly returns lets us characterize the performance and other properties of optimized portfolios based on each of our models.

Table 4 summarizes the optimization results. For the sake of brevity, the results in the table are all based on one model for forecasting variances, namely, a model using regression-adjusted historical variances. Panel A of the table evaluates each portfolio's performance in terms of its average monthly return, standard deviation, Sharpe ratio, and its tracking error volatility (the volatility of the monthly difference between the portfolio's return and the return on the S&P 500 index). These are all expressed on an annualized basis. Also reported is the correlation between the portfolio's return and the return on the S&P 500 index, and the average number of stocks in each portfolio with weights above 0.5%.

To provide some background, we present results for two simple diversification strategies that involve no optimization, namely the value-weighted and equally weighted portfolios made up of the same stocks available to the optimizer. An investor who diversifies by investing equal amounts in each of the 250 available stocks would experience an annualized standard deviation of 16.62%. A value-weighted portfolio takes larger positions in some stocks than in others, but the tendency for larger stocks to have low volatilities pulls down the overall standard deviation to 15.54%. In comparison, it is clear that some form of optimization helps. The annualized

[6] Since our forecasts are based on the past 60 months of returns, and there are 250 stocks, the covariance matrix is singular. Imposing the constraints guarantees a solution to the variance minimization problem.

standard deviation of the optimized portfolio based on the full covariance model is 12.94%, yielding a Sharpe ratio of 0.6405 (compared to a Sharpe ratio of 0.6027 for the equally weighted portfolio).

As in our earlier forecasting exercises, the various models generally provide similar results. To single out two cases, for example, the full covariance model yields an annualized standard deviation of 12.94%, while the corresponding statistic for the three-factor model is 12.66%. As a further illustration of the general point that more complicated models do not necessarily do better, the prize for the single model generating the lowest prospective standard deviation goes to model 5, which assumes that covariances are proportional to the product of the stocks' return volatilities. This model essentially sets all pairwise correlations between stocks to a constant and takes advantage of the relative stability of return variances. When covariance forecasts are generated by a composite model (model 7 in Table 4), the standard deviation of the optimized portfolio drops to 12.59%.[7]

Panel B of Table 4 provides further clues as to why the models perform so similarly. We report four characteristics of each portfolio: its beta relative to the value-weighted CRSP market index, the average size (in logarithms), average book-to-market ratio, and the average dividend yield of the stocks in the portfolio. To ease comparison, each characteristic is also expressed as a decile ranking (with 1 being the lowest and 10 being the highest). In addition, we report the percentage of the portfolio invested in two industries, namely firms whose first two digits of the SIC code are 35 or 36 (Industrial, Commercial Machinery, Computer Equipment, and Electrical Equipment excluding Computers) and firms with a two-digit SIC code of 49 (Electric, Gas and Sanitary Services). These two industries display considerable differences in terms of features such as return volatilities and market betas. Each of the reported characteristics is measured as of the portfolio formation date and are averaged over all portfolio formation years.

A striking feature of the optimized portfolios is that they all select stocks with low betas. While the equally weighted portfolio of all the candidate stocks has an average beta of 1.07 (placing it roughly in the fifth decile), the betas of the optimized portfolios fall between 0.5 and 0.7 (the average decile ranking is below 2). This emphasis on stocks with low betas may help to explain why the portfolios tilt somewhat toward larger stocks and toward value stocks. For the same reason, the portfolios all include a preponderance of utility stocks, which tend to have low betas and low return volatilities. At least 40 percent of each optimized portfolio is invested in the utility industry (SIC code 49). This is compared with a weight of 8.66% for utilities in the value-weighted portfolio and 15.31% in the equally weighted portfolio. In

[7] The composite model is an equally weighted average of the forecasts from four other models. The component models are the full covariance model, the three-factor model, the nine-factor model, and an industry-size model.

other words, the optimizer selects basically every utility it is presented. On the other hand the weight given to stocks in SIC codes 35 and 36 is relatively puny (never more than 4% for the optimized portfolios). As Fama and French (1997) document, industries 35 and 36 tend to have higher than average market betas.

4.2 Interpretation

The results of the variance-minimization exercises suggest the following interpretation. One key conclusion is that there is a major factor which is more important than the other influences on returns.[8] This dominant influence, the market, tends to overpower the remaining factors, so that their incremental informativeness becomes very difficult to detect. This accounts for why the different factor models generate similar forecast results.

Put another way, portfolio p's variance ν_p from the factor model [Equation (2)] is

$$\nu_p = \sum_{j=1}^{K} \beta_{pj}^2 \omega_j + \delta_p. \tag{4}$$

Here ω_j is the variance of the jth factor, δ_p is the variance of the idiosyncratic return, and (solely for expository convenience) it is assumed that the factors are mutually uncorrelated. If the largest part of the variance is due to the variance of the market factor, then what the optimizer tends to do in each case generally is not very surprising. Namely, the optimizer selects stocks with low market betas, such as utility stocks. Further, as an empirical matter, stocks with low market betas generally tend to have low residual variances and low total return volatilities as well. Any covariance model that exploits these patterns will tend to yield similar results (and tend to do better than the passively diversified equally weighted or value-weighted portfolios). As a result, the optimized portfolios do not differ dramatically in terms of their performance. Since the role of the remaining factors is obscured by the major factor, there appears to be little benefit from obtaining finer breakdowns of the systematic component of the volatility.

To calibrate the relative importance of the different factors, we use the asymptotic principal components method of Connor and Korajzyck (1988). Five factors are extracted from the monthly returns of all NYSE-AMEX issues over the 5-year period ending on April 1998. To capture the average situation, we partition the sample variance of the equally weighted CRSP index into the proportions attributable to each principal component. The first factor accounts for 75% of the variability of the index. The proportions

[8] A similar interpretation is offered by Green and Hollifield (1992), who provide conditions under which well-diversified portfolios are mean-variance efficient. The empirical evidence in Connor and Korajczyk (1993) also suggests that after the first factor the marginal explanatory power of additional factors is relatively low.

captured by the remaining four principal components pale in comparison (they amount to 6.9%, 3.6%, 8.8% and 5.6%, respectively). In short, after accounting for the dominant influence of the first factor, further refinements do not offer a great deal of improvement.

This line of thinking also suggests a way to structure our experiment to yield sharper differences between the different covariance models. In particular, if we can remove the impact of the dominant market factor the importance of the remaining factors may show up more clearly. As it turns out, this problem is a specific case of tracking a benchmark portfolio, which occurs much more commonly in practice than the variance-minimization problem.

5. Minimizing Tracking Error Volatility

5.1 The Minimum Tracking Error Volatility Portfolio

In practice, mean-variance optimization is much more commonly applied in a different context. Since professional investment managers are paid to outperform a benchmark, they are in general not concerned with the absolute variance of their portfolios, but are concerned with how their portfolio deviates from the benchmark. Underperforming the benchmark for several years typically results in the termination of a manager's appointment. The objective in this case is to minimize the portfolio's tracking error volatility, or the volatility of the difference between a portfolio's return and the return on the benchmark. In the context of the factor model, the generating process for the return difference is

$$r_{pt} - r_{Bt} = \beta_{p0} - \beta_{B0} + \sum_{j=1}^{K} (\beta_{pj} - \beta_{Bj}) f_{jt} + (\epsilon_{pt} - \epsilon_{Bt}), \qquad (5)$$

where r_{pt} and r_{Bt} are the returns on the portfolio and on the benchmark, respectively, at time t, and their factor loadings are given by β_{pj} and β_{Bj} for $j = 0, 1, \ldots, K$. Accordingly, a portfolio whose loadings come close to matching the benchmark's would do a good job in minimizing the volatility of excess returns, τ_p:

$$\tau_p = \sum_{j=1}^{K} (\beta_{pj} - \beta_{Bj})^2 \omega_j + \psi_p, \qquad (6)$$

where for ease of exposition it is assumed that the factors are mutually uncorrelated.

The benchmark's loadings can be estimated and used as anchors for the desired portfolio's exposures. Then the problem of minimizing tracking error volatility is, at least potentially, much simpler than the problem of minimizing the global portfolio variance. Whether this turns out to be the case depends on the benchmark having exposures that can be matched by some feasible portfolio of the sample of candidate stocks. The variance-minimization criterion discussed in the previous section sets the benchmark to be the riskless asset with zero exposures on all the factors. In this instance, as long as we insist on nonnegative weights, and as long as most stocks have exposures to the important factors that are of the same sign, all the different models will have an equally hard time. Similarly, there will not be notable differences across the models if the problem were to use, for example, utility stocks to track a single issue such as Netscape.

Given that the market factor is the most important, suppose the benchmark is a portfolio whose market exposure is not too unrepresentative of those of the underlying set of stocks. It is then a less challenging problem to minimize the role of this dominant source of return variability, since the difference between the market betas of the portfolio and the benchmark can be set close to zero. As a result, the incremental importance of any remaining factors in the decomposition [Equation (6)] becomes easier to detect, so there may be more opportunities to discriminate between the different covariance models.

We implement the tracking error optimization problem as follows. We choose the S&P 500 index as the benchmark. The minimum tracking error volatility portfolio can be found by expressing every stock's return in excess of the return on the benchmark, and solving for the portfolio with the lowest variance of excess returns. The constraints are as in the previous set of optimization exercises (nonnegative weights and an upper bound of 2%).

Panel A of Table 5 summarizes the performance of the minimum tracking error volatility portfolios. The results are based on several models for forecasting the covariance of excess returns. In every case, the forecasted variance is set to be the regression-adjusted variance of the most recent past 60 monthly excess returns (over the S&P 500). As we had hoped, simplifying the problem generates more discrimination between the covariance models. The annualized tracking error volatility of the equally weighted portfolio is 6.16%, while a three-factor model reduces the tracking error volatility to 4.53%. Adding more factors gives rise to lower volatility relative to the benchmark. For example, a nine-factor model results in a tracking error volatility of 4.01%.

The link between this finding and our earlier forecasting results is actually quite natural. In the earlier case of minimizing portfolio variance, return covariances between stocks are generally positive and the portfolio weights are nonnegative by design. Hence the dominant component of the portfolio's variance comes from the covariance terms. As our forecasting exercises suggest, the various models all have roughly the same degree of success in predicting future covariances. Accordingly, the models do not differ much when it comes to minimizing portfolio variance. In the case of tracking error volatility, however, the covariances of returns in excess of the market return can be positive as well as negative. There is therefore more scope, even

Table 5
Performance and characteristics of portfolios with minimum tracking error volatility based on forecasting models

Panel A: Performance of portfolios

Model	Mean	Standard deviation	Sharpe ratio	Tracking error	Correlation with market	Average number of stocks with weights above 0.5%
(1) Full covariance	0.1565	0.1498	0.5607	0.0403	0.9648	63
(2) 1 factor	0.1649	0.1496	0.6175	0.0512	0.9429	67
(3) 3 factor	0.1557	0.1454	0.5719	0.0453	0.9552	77
(4) 9 factor	0.1558	0.1474	0.5648	0.0401	0.9651	76
(5) Product of standard deviations	0.1597	0.1510	0.5773	0.0497	0.9467	87
(6) Industry, size	0.1590	0.1472	0.5878	0.0416	0.9624	68
(7) 250 stocks, value-weighted	0.1431	0.1554	0.4539	0.0304	0.9807	45
(8) 250 stocks, equally-weighted	0.1727	0.1662	0.6027	0.0616	0.9287	0
(9) 1000 stocks, value-weighted	0.1484	0.1520	0.4994	0.0138	0.9959	36

Panel B: Characteristics of portfolios

Model	Beta (Rank)	MAD	LogSize (Rank)	MAD	BM (Rank)	MAD	DP (Rank)	MAD	Percent invested in: SIC 35, 36	SIC 49
(1) Full covariance	0.9802	0.0251	21.28	1.0018	0.6604	0.0603	0.0432	0.0060	10.67	15.81
	(4.52)		(9.00)		(4.48)		(6.60)			
(2) 1 factor	0.9731	0.0277	20.64	1.6468	0.7513	0.1440	0.0466	0.0088	9.67	23.20
	(4.44)		(7.92)		(5.28)		(7.20)			
(3) 3 factor	0.9701	0.0287	20.97	1.3156	0.7093	0.1021	0.0471	0.0080	8.64	26.20
	(4.44)		(8.52)		(4.92)		(7.16)			
(4) 9 factor	0.9734	0.0267	21.10	1.1848	0.6982	0.0909	0.0452	0.0061	9.75	21.65
	(4.44)		(8.84)		(4.84)		(7.04)			
(5) Product of standard deviations	0.9814	0.0328	20.63	1.6500	0.7359	0.1287	0.0466	0.0076	8.74	23.80
	(4.44)		(7.92)		(5.12)		(7.20)			
(6) Industry, size	0.9531	0.0456	21.42	0.8591	0.6546	0.0492	0.0450	0.0070	8.95	15.09
	(4.24)		(9.24)		(4.40)		(6.96)			
(7) 250 stocks, value-weighted	0.9897	0.0394	22.34	0.2279	0.5844	0.0332	0.0408	0.0056	13.20	8.66
	(4.60)		(10.00)		(3.88)		(6.32)			
(8) 250 stocks, equally-weighted	1.0686	0.0970	20.28	2.0020	0.7593	0.1521	0.0413	0.0064	11.02	15.31
	(5.20)		(7.12)		(5.24)		(6.32)			
(9) 1000 stocks, value-weighted	0.9873		22.28		0.6072		0.0419		13.26	8.57
	(4.60)		(10.00)		(4.00)		(6.48)			

Table 5
(continued)

At the end of April of each year from 1973 through 1997 a random sample of 250 firms is drawn from eligible NYSE and AMEX domestic common stock issues. Forecasts of covariances and variances of monthly returns in excess of the benchmark's return are generated from different models, using the prior 60 months of data for each stock. Based on each model's forecasts of excess return variances and covariances, a quadratic programming procedure is used to find the portfolio with minimum tracking error volatility (standard deviation of portfolio return in excess of the return on the benchmark) where the benchmark portfolio is the Standard & Poor's 500 index. The portfolio weights are constrained to be nonnegative and not larger than 2%. These weights are then applied to form buy-and-hold portfolio returns until the next April, when the forecasting and optimization steps are repeated and the portfolios are reformed. For each procedure summary statistics are presented in panel A for the annualized mean return and annualized standard deviation for the returns realized on the portfolio; the annualized average Sharpe ratio (the mean return in excess of the Treasury bill rate, divided by the standard deviation); the correlation between the monthly portfolio return and the return on the S&P 500 index; the annualized standard deviation of the portfolio return in excess of the S&P 500 return; and the average number of stocks each year with portfolio weights above 0.5%. Panel B reports average characteristics of stocks in each portfolio: the beta relative to the value-weighted CRSP index (based on the 60 months of data prior to portfolio formation); market value of equity (in natural logarithms); book-to-market equity ratio (denoted BM); and dividend yield (denoted DP). Each characteristic is also measured as a decile ranking (from 1, the lowest, to 10, the highest). The column denoted MAD reports the mean across years of the absolute difference between the optimized portfolio's characteristic and the benchmark's characteristic. Also reported is the average proportion of the portfolio invested in firms with two-digit SIC codes of 35 and 36 (Industrial, Commercial Machinery, Computer Equipment and Electrical Equipment excluding Computers) and in firms with a two-digit SIC code of 49 (Electric, Gas and Sanitary Services). Each characteristic is measured as of the portfolio formation date and averaged across all portfolio formation years.

The full covariance model (model 1) uses the return covariance estimated over the most recent past 60 months prior to portfolio formation as the forecast. Covariance forecasts from the factor models (models 2–4) are based on Equation (3) in the text. The one-factor model uses the return on the value-weighted CRSP index as the factor. The three-factor model includes the return on the value-weighted index as well as size and book-to-market factors. The nine-factor model uses as factors the market, size, book-to-market, momentum, dividend yield, cash flow yield, the term premium, the default premium, and the second principal component. Model 5 uses the most recent past 8 years of data. Return covariances between two stocks are measured over the most recent 3 years and are regressed on the product of standard deviations of the stocks' returns measured over the earliest 5 years. The regression model is used to generate the covariance forecasts. In model 6, stocks are assigned to an industry-size group. There are 48 industries, as in Fama and French (1997). Each industry is divided into a set of large firms (with market value of equity exceeding the median capitalization of NYSE firms in that industry) and a set of small firms (with market value of equity below the NYSE median in that industry). The covariance forecast for any two stocks is given by the average of all pairwise covariances between stocks in their respective industry-size groups, based on the most recent past 60 months. Models 7 and 8 are the value-weighted and equally weighted portfolios, respectively, of all the 250 candidate stocks available at each portfolio formation date. Model 9 is the value-weighted portfolio of the largest 1000 stocks at each portfolio formation date.

with nonnegative weights, for the optimizer to cancel out the covariance component of the portfolio's tracking error volatility.[9] Furthermore, the models display more dispersion in terms of the association between the forecasted and realized excess return covariances. At the 36-month horizon, for instance, the correlation between forecasts and realized excess return covariances is 0.13 for the full covariance model. For the one-factor model the correlation is 0.14, and it is 0.20 for the three-factor model. As a result, there is more room to differentiate between the models when the objective is to minimize tracking error volatility.

To ensure that our findings are not sample specific, we also replicate the experiment in Table 5 200 times for each of our models. In each replication we draw a different set of stocks to carry out the minimization with respect to tracking error volatility. The results confirm the general pattern of differences across the models in Table 5. For example, the tracking error volatility averaged over all 200 replications is 3.99% for the full covariance model, compared to 5.11% for the one-factor model and 4.54% for the three-factor model. The average differences across the factor models in tracking error volatility are large and statistically significant. Comparing the one- and three-factor models, for example, the mean difference in tracking error volatility is 0.57% (with a standard error of the mean of 0.01%).

Panel B examines how closely each model comes to the benchmark along several attributes, and whether the degree of alignment is associated with the realized tracking error volatility. In this way we can pinpoint potentially relevant determinants of a portfolio's risk profile. We take the characteristics of the value-weighted index of the largest thousand stocks to be our proxy for the benchmark portfolio's characteristics.[10] For each model we report in the column denoted MAD the average (across portfolio formation years) absolute difference between the characteristics of the optimized portfolio and the benchmark.

Comparing the full covariance model and the one-factor model brings out clearly the importance of nonmarket sources of return covariation. The one-factor model concentrates on coming close to the benchmark's market exposure (the average absolute difference is 0.0277). It does so at the expense of deviating considerably from the benchmark in terms of size, book-to-market, and dividend yield (the average absolute differences are the largest of all the optimized models in Table 5). The focus on beta is evidently not enough, for the resulting tracking error volatility is relatively large (5.12%). In comparison, the full covariance model delivers small de-

[9] This may also help explain the tendency to have a relatively large portion of the optimized portfolios invested in utility stocks (SIC code 49). The bulk of negative excess return covariances is clustered in the utility industry, where 47% of the pairwise covariances between utility stocks and other stocks is negative.

[10] Difficulties in identifying which firms belong to the S&P 500, especially in the early years of the sample period, force us to work with a proxy for the index for this calculation.

viations from the benchmark on all four dimensions and comes up with a lower tracking error volatility (4.03%). More generally, the results suggest that a higher-dimensional model should provide more opportunities to narrow any potential divergence from the benchmark's risk exposures. For example, the tracking error volatility under the nine-factor model (4.01%) is lower than under the one-factor model. Given, however, the possibility of data snooping, requiring the optimized portfolio to be aligned with the benchmark on a host of dimensions, may ultimately become counterproductive. This may explain why the tracking error volatilities for the full covariance model and for the nine-factor model are roughly comparable.

The results for the value-weighted portfolio of all the 250 stocks provide further testimony to the strong covariation among large stocks. This portfolio comes closest to the benchmark on all the reported attributes, except for beta, in terms of mean absolute differences. The resulting tracking error volatility is 3.04%, the lowest of all the models in panel A. Since the benchmark in this case comprises large stocks, however, the performance of the value-weighted portfolio may not hold for other choices of benchmark.[11]

Roll (1992) shows that a portfolio that is optimized with respect to tracking error volatility will in general not be mean-variance efficient. A comparison of Tables 4 and 5 lets us quantify the difference between the two criteria. Optimizing with respect to tracking error volatility (Table 5) yields portfolios that have larger standard deviations than the portfolios that are optimized with respect to return volatility (Table 4). The average difference is quite large and is on the order of 2%. Insofar as the emphasis in practice on tracking error volatility reflects a mismatch between the objectives of the portfolio manager and those of the ultimate investor, then the 2% difference is an estimate of the cost of this mismatch.

5.2 Tracking Error Volatility Minimization by Matching on Attributes

The models of the previous section attempt to minimize tracking error volatility by forming portfolios whose factor loadings come close to the benchmark portfolio's loadings. Since the loadings must be estimated, this approach may be handicapped by estimation errors. For example, there is nothing to prevent the optimizer from choosing small stocks to track the S&P 500 Index if these small stocks happen to have past estimated loadings which coincide with the benchmark's. An approach which is less prone to measurement errors may give better results. In this regard, it is intriguing to note that the value-weighted portfolio of the 250 sample stocks yields the lowest tracking error volatility in Table 5. As one way to formalize this observation,

[11] For example, when the benchmark is an equally weighted portfolio of 500 randomly selected stocks, the value-weighted portfolio has a tracking error volatility of 5.59%, compared to a tracking error volatility of 2.02% under the full covariance model.

Table 6
Performance and characteristics of portfolios based on matching the benchmark by attributes

Panel A: Performance of portfolios

Attribute	Mean	Standard deviation	Sharpe ratio	Tracking error	Correlation with market	Average number of stocks with weights above 0.5%
(1) Industry	0.1771	0.1669	0.6263	0.0579	0.9383	59
(2) Industry, size	0.1641	0.1637	0.5594	0.0460	0.9600	51
(3) Size, residual variance	0.1491	0.1488	0.5107	0.0353	0.9731	78
(4) Size, book-to-market, residual variance	0.1477	0.1521	0.4940	0.0325	0.9774	79
(5) 9 attributes, residual variance	0.1476	0.1565	0.4798	0.0301	0.9813	77
(6) 1000 stocks, value-weighted	0.1484	0.1520	0.4994	0.0138	0.9959	36

Panel B: Characteristics of portfolios

Attribute	Beta (rank)	MAD	LogSize (rank)	MAD	BM (rank)	MAD	DP (rank)	MAD	Percent invested in: SIC 35, 36	SIC 49
(1) Industry	1.0938	0.1184	20.55	1.7315	0.7361	0.1289	0.0403	0.0057	13.07	9.37
	(5.44)		(7.92)		(5.08)		(6.24)			
(2) Industry, size	1.0752	0.0961	21.10	1.1782	0.6841	0.0769	0.0406	0.0050	13.67	9.78
	(5.36)		(8.80)		(4.56)		(6.28)			
(3) Size, residual variance	0.9379	0.0567	21.76	0.5192	0.6482	0.0452	0.0443	0.0059	8.55	20.69
	(4.04)		(9.80)		(4.24)		(6.76)			
(4) Size, book-to-market,	0.9539	0.0442	21.77	0.5120	0.6007	0.0076	0.0415	0.0038	9.26	17.64
residual variance	(4.24)		(9.80)		(3.92)		(6.52)			
(5) 9 attributes, residual variance	0.9828	0.0055	21.78	0.4969	0.6021	0.0075	0.0393	0.0035	10.20	13.53
	(4.48)		(9.88)		(3.92)		(6.24)			
(6) 1000 stocks, value-weighted	0.9873		22.28		0.6072		0.0419		13.26	8.57
	(4.60)		(10.00)		(4.00)		(6.48)			

At the end of April of each year from 1973 through 1997 a random sample of 250 firms is drawn from eligible NYSE and AMEX domestic common stock issues. Using a quadratic programming procedure, portfolios are formed that minimize the sum of squared deviations between the portfolio's attributes and the benchmark's attributes. The benchmark portfolio is the Standard & Poor's 500 index. The portfolio weights are constrained to be nonnegative and not larger than 2%. These weights are then applied to form buy-and-hold portfolio returns until the next April, when the optimization is repeated and the portfolios are reformed. For each procedure summary statistics are presented in panel A for the annualized mean return and annualized standard deviation for the returns realized on the portfolio; the annualized average Sharpe ratio (the mean return in excess of the Treasury bill rate, divided by the standard deviation); the

Table 6
(continued)

correlation between the monthly portfolio return and the return on the S&P 500 index; the annualized standard deviation of the portfolio return in excess of the S&P 500 return; and the average number of stocks each year with portfolio weights above 0.5%. Panel B reports average characteristics of stocks in each portfolio: the beta relative to the value-weighted CRSP index (based on the sixty months of data prior to portfolio formation); market value of equity (in natural logarithms); book-to-market equity ratio (denoted BM); and dividend yield (denoted DP). Each characteristic is also measured as a decile ranking (from 1, the lowest, to 10, the highest). The column denoted MAD reports the mean across years of the absolute difference between the optimized portfolio's characteristic and the benchmark's characteristic. Also reported is the average proportion of the portfolio invested in firms with two-digit SIC codes of 35 and 36 (Industrial, Commercial Machinery, Computer Equipment and Electrical Equipment excluding Computers) and in firms with a two-digit SIC code of 49 (Electric, Gas and Sanitary Services). Each characteristic is measured as of the portfolio formation date, and averaged across all portfolio formation years.

In the *industry model*, there are nine industry sectors. Each sector's percentage contribution to the total market capitalization of the benchmark is calculated. The fraction is then equally allocated across the stocks belonging to that sector and that are available for portfolio formation, up to a maximum allocation of 2%. The weight for a stock is then its allocation, scaled so that the weights sum to one across all the candidate stocks. In the *industry, size model* each of the nine sectors is divided into two sets containing stocks that are above (below) the median market capitalization of NYSE firms in the sector. Within each sector-size classification the portfolio formation procedure is as for the industry model.

For the other models the general procedure is as follows. In each case we choose the portfolio weights x_j for stock $j = 1, \ldots, N$ to minimize the portfolio's residual variance and the sum of squared deviations between the portfolio's attributes and the benchmark's attributes:

$$\sum_{j=1}^{N} x_j^2 \omega_j^2 + \sum_{i=1}^{K} \left(\sum_{j=1}^{N} x_j Z_{ij} - Z_{iB} \right)^2,$$

where ω_j^2 is the residual variance for stock j and Z_{ij} is its ith attribute and Z_{iB} is the corresponding attribute for the benchmark. The portfolio weights are constrained to be nonnegative and not larger than 2%. The attributes (measured as of the portfolio formation date) are ordered and expressed as percentile ranks (between zero and one).

In the *size, residual variance model* the attribute is firm size (equity market capitalization). The *size, book-to-market, residual variance model* augments firm size with the ratio of book-to-market value of equity. The *nine attribute, residual variance model* includes firm size, book-to-market ratio, dividend yield, rate of return beginning seven months and ending 1 month before portfolio formation, rate of return beginning 60 months and ending 12 months before portfolio formation, and loadings on the default premium factor, on the term premium factor, on the equally weighted CRSP market index and on the second principal component. Model 6 is the value-weighted portfolio of the largest 1000 stocks at each portfolio formation date.

suppose that the cross section of returns is given by the following model:[12]

$$r_i - r_f = \sum_{j=1}^{K} Z_{ij} f_j + u_i \tag{7}$$

where Z_{ij} is stock i's risk descriptor (or exposure) with respect to the jth common factor f_j. The difference between portfolio p's return and the benchmark return is thus

$$r_p - r_B = \sum_{j=1}^{K} (Z_{pj} - Z_{Bj}) f_j + (u_p - u_B). \tag{8}$$

From this perspective, matching the benchmark with respect to the observable attributes Z_{Bj}, such as size or dividend yield, is another way to reduce tracking error volatility.[13]

Table 6 reports the results from this alternative approach to minimizing tracking error volatility. In general, matching the benchmark by attributes produces lower tracking error volatilities, compared to the results in Table 5. For example, the approach using loadings from a nine-factor model (model 4 in Table 5) generates a tracking error volatility of 4.01%, while a portfolio that matches the benchmark along the corresponding nine attributes (model 5 in Table 6) generates a tracking error volatility of 3.01%. To place this in perspective, it is quite common for investment managers to be evaluated in terms of their information ratios (the portfolio alpha divided by tracking error volatility). In the above example, the nine-attribute matching procedure raises the information ratio by as much as a third.

While the factor loading approach benefits from being directly based on the behavior of past returns, its advantage is more than offset by the measurement errors in the loadings. As a result, a portfolio's current attributes provide more reliable indicators of its future tracking error. In some cases the attribute matching procedure even compares favorably with the ideal case of perfect foresight. In particular, the tracking error volatility under the full covariance model assuming perfect foresight about the future covariance matrix of excess returns is 1.57% (compared to roughly 3.25% and 3.01% for models 4 and 5 in Table 6). We also repeated the exercise in Table 6 200 times, drawing different sets of stocks each time. The overall conclusions are very similar. Notably, in every repetition the nine-attribute matching procedure generates a lower tracking error volatility than the cor-

[12] See, for example, BARRA (1990). Jagannathan and Wang (1992) also provide evidence on the potential usefulness of firm characteristics such as size for estimating betas.

[13] In practice this approach is commonly used among hedge funds and for long-short investment strategies, where short positions are feasible. The underlying idea is that the long and short positions should be closely aligned in terms of attributes such as market risk, size, dividend yield, and industry composition.

responding nine-factor covariance model. The average reduction across the replications is 0.95% (the standard error of the mean is 0.01%).

One widely used approach to replicating an index is to match the benchmark's industry composition. Model 1 in Table 6 does this and yields an annualized tracking error volatility of 5.79%. Compared to the other models in Table 6, it clearly does not do well. Requiring that the portfolio match the benchmark's size composition as well as its industry composition reduces the tracking error volatility to 4.60% (model 2). Indeed, of all the different attributes, size turns out to be the critical dimension on which to match. The size-matched minimum residual variance portfolio (model 3) has a tracking error volatility of 3.53%, which is close to the performance (3.01%) of the nine attribute model (model 5).

It would be premature, however, to conclude from this exercise that size is the only important attribute to match. In particular, this finding may be specific to the nature of the index chosen, namely large stocks making up the S&P 500. To check up on this, Table 7 provides results for minimizing tracking error volatility relative to two other benchmarks. The benchmark is either the value-weighted portfolio of the 250 stocks that are ranked highest by the ratio of book-to-market value of equity (in part I), or the value-weighted portfolio of the 250 stocks that are ranked lowest by the book-to-market ratio (part II). These two reference portfolios (the value stock benchmark and the growth stock benchmark, respectively) correspond in spirit to indexes that are widely used in practice to evaluate the performance of value- and growth-oriented investment managers.

Using the value or growth benchmarks generally leads to larger tracking error volatilities compared to the case of the S&P 500 benchmark. For example, under the full covariance model, the tracking error volatility is 4.72% with respect to the value benchmark (part I, panel A) and 5.00% with respect to the growth benchmark (part II, panel A), compared to 4.03% with respect to the S&P 500 index (Table 5). Matching by size only (model 4 in Table 7) would not be the most successful procedure, as it produces large differences from the benchmark with respect to other attributes. In particular, the average absolute difference between the book-to-market ratio of the size-matched portfolio and the value stock benchmark is 0.3881, while the average absolute difference with respect to dividend yield is 0.0156. The corresponding differences for the growth stock benchmark are 0.2286 and 0.0120, respectively. Consequently, the tracking error volatility under the size-matching procedure is 4.61% under the value benchmark and 4.78% under the growth benchmark. Instead, matching on nine attributes yields the lowest tracking error volatilities (3.79% and 3.97% for the value and growth benchmarks, respectively).[14]

[14] The attribute-matching procedure also works well when we use an equally weighted index of 500 randomly selected stocks as the benchmark. In this case, matching on nine attributes generates a tracking error volatility of 1.66%, compared to 2.02% using the full covariance model.

Table 7
Performance and characteristics of portfolios with minimum tracking error volatility with respect to value and growth benchmarks, based on forecasting models

Part I: Value stock benchmark
Panel A: Performance of portfolios

Model	Mean	Standard deviation	Sharpe ratio	Tracking error volatility	Correlation with benchmark	Average number of stocks with weights above 0.5%
(1) Full covariance	0.1736	0.1465	0.6900	0.0472	0.9509	62
(2) 250 stocks, value-weighted	0.1431	0.1554	0.4539	0.0624	0.9179	45
(3) 250 stocks, equally-weighted	0.1727	0.1662	0.6027	0.0564	0.9410	0
(4) Matched on size, residual variance	0.1556	0.1491	0.5570	0.0461	0.9533	83
(5) Matched on 9 attributes, residual variance	0.1593	0.1594	0.5446	0.0379	0.9724	54
(6) Value stock benchmark	0.1790	0.1522	0.7000	0.0000	1.0000	45

Panel B: Characteristics of portfolios

Model	Beta (rank)	MAD	LogSize (rank)	MAD	BM (rank)	MAD	DP (rank)	MAD	Percent invested in: SIC 35, 36	SIC 49
(1) Full covariance	0.9295 (4.00)	0.0550	20.75 (8.12)	0.7676	0.8320 (6.00)	0.2521	0.0553 (7.64)	0.0160	8.01	27.58
(2) 250 stocks, value-weighted	0.9897 (4.60)	0.0684	22.34 (10.00)	0.8447	0.5844 (3.88)	0.4996	0.0408 (6.32)	0.0234	13.20	8.66
(3) 250 stocks, equally-weighted	1.0686 (5.20)	0.1035	20.28 (7.12)	1.2309	0.7593 (5.24)	0.3248	0.0413 (6.32)	0.0212	11.02	15.31
(4) Matched on size, residual variance	0.9340 (4.04)	0.0623	21.35 (9.08)	0.1666	0.6959 (4.76)	0.3881	0.0460 (7.16)	0.0156	8.76	23.85
(5) Matched on 9 attributes, residual variance	0.9662 (4.40)	0.0116	21.12 (8.84)	0.3862	1.0859 (8.00)	0.0259	0.0642 (8.32)	0.0126	6.65	29.46
(6) Value stock benchmark	0.9745 (4.40)	0.0000	21.51 (9.44)	0.0000	1.0841 (7.96)	0.0000	0.0616 (8.28)	0.0000	8.81	22.86

Table 7
(continued)

Part II: Growth stock benchmark
Panel A: Performance of portfolios

Model	Mean	Standard deviation	Sharpe ratio	Tracking error volatility	Correlation with benchmark	Average number of stocks with weights above 0.5%
(1) Full covariance	0.1505	0.1556	0.5013	0.0500	0.9520	60
(2) 250 stocks, value-weighted	0.1431	0.1554	0.4539	0.0452	0.9611	45
(3) 250 stocks, equally-weighted	0.1727	0.1662	0.6027	0.0760	0.8938	0
(4) Matched on size, residual variance	0.1453	0.1519	0.4788	0.0478	0.9565	69
(5) Matched on 9 attributes, residual variance	0.1376	0.1643	0.3961	0.0397	0.9706	52
(6) Growth stock benchmark	0.1427	0.1633	0.4296	0.0000	1.0000	57

Panel B: Characteristics of portfolios

Model	Beta (rank)	MAD	LogSize (rank)	MAD	BM (rank)	MAD	DP (rank)	MAD	Percent invested in: SIC 35, 36	SIC 49
(1) Full covariance	1.0124	0.0506	21.26	1.1486	0.5919	0.1996	0.0399	0.0085	9.34	13.30
	(4.80)		(9.04)		(3.88)		(6.20)			
(2) 250 stocks, value-weighted	0.9897	0.0546	22.34	0.2905	0.5844	0.1921	0.0408	0.0094	13.20	8.66
	(4.60)		(10.00)		(3.88)		(6.32)			
(3) 250 stocks, equally-weighted	1.0686	0.0865	20.28	2.1305	0.7593	0.3670	0.0413	0.0100	11.02	15.31
	(5.20)		(7.12)		(5.24)		(6.32)			
(4) Matched on size,	0.9507	0.0800	21.99	0.4242	0.6209	0.2286	0.0424	0.0120	8.73	17.07
residual variance	(4.20)		(10.00)		(4.08)		(6.72)			
(5) Matched on 9 attributes,	1.0036	0.0112	21.81	0.5973	0.4111	0.0188	0.0321	0.0021	11.08	5.31
residual variance	(4.60)		(9.84)		(2.20)		(5.04)			
(6) Growth stock benchmark	1.0040	0.0000	22.41	0.0000	0.3923	0.0000	0.0321	0.0000	14.09	4.15
	(4.68)		(10.00)		(2.16)		(4.96)			

At the end of April of each year from 1973 through 1997 a random sample of 250 firms is drawn from eligible NYSE and AMEX domestic common stock issues. Forecasts of covariances and variances of monthly returns in excess of the benchmark's return are generated from different models, using the prior 60 months of data for each stock. Based on each model's forecasts of excess return variances and covariances, a quadratic programming procedure is used to find the portfolio with minimum tracking error volatility (standard deviation of portfolio return in excess of the return on the benchmark) where the benchmark portfolio is either a value-weighted index of value stocks (part I) or a value-weighted index of growth stocks (part II). The portfolio weights are constrained to be nonnegative and not larger than 2%. These weights are then applied to form buy-and-hold portfolio returns until the next April, when the forecasting and optimization steps are repeated and the portfolios are reformed. For each procedure, summary statistics are presented in panel A for the annualized mean return and annualized standard deviation for the

Table 7
(continued)

returns realized on the portfolio; the annualized average Sharpe ratio (the mean return in excess of the Treasury bill rate, divided by the standard deviation); the correlation between the monthly portfolio return and the return on the benchmark; the annualized standard deviation of the portfolio return in excess of the benchmark's return; and the average number of stocks each year with portfolio weights above 0.5%. Panel B reports average characteristics of stocks in each portfolio: the beta relative to the value-weighted CRSP index (based on the 60 months of data prior to portfolio formation); market value of equity (in natural logarithms); book-to-market equity ratio (denoted BM); and dividend yield (denoted DP). Each characteristic is also measured as a decile ranking (from 1, the lowest, to 10, the highest). The column denoted MAD reports the mean across years of the absolute difference between the optimized portfolio's characteristic and the benchmark's characteristic. Also reported is the average proportion of the portfolio invested in firms with two-digit SIC codes of 35 and 36 (Industrial, Commercial Machinery, Computer Equipment and Electrical Equipment excluding Computers) and in firms with a two-digit SIC code of 49 (Electric, Gas and Sanitary Services). Each characteristic is measured as of the portfolio formation date, and averaged across all portfolio formation years.

The full covariance model (model 1) uses the return covariance estimated over the most recent past 60 months prior to portfolio formation as the forecast. Models 2 and 3 are the value-weighted and equally weighted portfolios, respectively, of all the 250 candidate stocks available at each portfolio formation date. Details on implementing the matching procedure used for models 4 and 5 are given in the notes to Table 6. The value (growth) stock benchmark is a value-weighted portfolio comprising the 250 stocks that are ranked highest (lowest) each year by the ratio of book-to-market value of equity.

6. Summary and Conclusion

Although the concept of portfolio mean-variance optimization forms the backbone of modern portfolio theory, it has come into widespread use only fairly recently. With the recent emphasis on risk management there has been a proliferation of portfolio optimization techniques. Yet there is very little scientific evidence on the performance of alternative risk optimization procedures. This article provides evidence with respect to forecasting the return covariances and variances that are key inputs to the optimizer. We compare the forecasting performance of different models of covariances, and we assess the out-of-sample performance of optimized portfolios based on each model.

Factor models of security returns were originally proposed as parsimonious ways to predict return covariances and simplify portfolio optimization. They remain at the center stage of portfolio analysis and have also been extensively used in modeling the behavior of expected returns. Accordingly, the bulk of our analysis focuses on applying such models.

We find that a few factors such as the market, size, and book-to-market value of equity capture the general structure of return covariances. For example, a model based on these three factors generates a correlation of 0.1994 between covariance forecasts and subsequent covariances (measured over a 12-month horizon). Expanding the number of factors does not necessarily improve our ability to predict covariances. Instead, the higher-dimensional models tend to overfit the data. For example, the full covariance model (which essentially assumes that there are as many factors as stocks) generates a correlation of 0.1792 between forecasted and realized covariances. However, there is substantial imprecision in the forecasts, so that the factor models yield mean absolute forecast errors that are not notably different from a simple model which assumes that all stocks share the same average pairwise covariance. Relaxing the linearity assumption underlying the factor models and using more updated estimates of factor loadings does not improve forecast power.

The true test of the models' forecasting ability is in the context of optimized portfolios. We conduct two types of experiments. First, in the spirit of the work of Markowitz we generate the global minimum variance portfolio under each model. In practice, however, investment managers who use portfolio optimization methods are evaluated relative to benchmarks. Accordingly, in the second set of experiments we compare the models in terms of minimizing tracking error volatility (the standard deviation of the difference between the returns on the portfolio and a benchmark). We highlight the global minimum variance or global minimum tracking error volatility portfolios, as any other point on the efficient frontier dilutes the importance of the second moments and concentrates more on expected returns.

The good news is some form of portfolio optimization helps for risk control. The various global minimum variance portfolios have future annualized standard deviations between 12.59% and 12.94%. In contrast, a passively diversified portfolio which invests equal amounts in each stock has a much higher standard deviation (16.62% per year). As the results on forecasting covariances suggest, however, there is very little discrimination between the models under a minimum variance criterion. A one-factor model is as good as a nine-factor model. All the models exploit the idea that the biggest benefits arise from reducing exposure to the market. The historical betas of the portfolios average about 0.6, compared to 1.1 if stocks were equally weighted. The objective of keeping the beta low explains why utility stocks are so heavily favored in the global minimum variance portfolios. In essence, the optimizer grabs every single utility.

The tracking error volatility criterion provides an important test case for distinguishing between the risk models. In particular, the tracking error (return in excess of the benchmark) tends to diminish the dominant influence of the market factor, thereby improving our ability to sort out the relative importance of any remaining factors. The different models do stand apart more when it comes to minimizing the tracking error volatility (as long as the benchmark's risk exposures are not too unrepresentative of the exposures of the underlying set of stocks). As an example of a naive approach, an equally weighted portfolio has a tracking error volatility of 6.16% per year. A one-factor model for forecasting tracking error covariances generates a tracking error volatility of 5.12% per year, while a nine-factor model pushes this down to 4.01% per year. This sizable reduction in tracking error volatility highlights the importance of optimization for investment management practice.

When the objective is to minimize tracking error volatility, we find that a simple heuristic procedure does better than the standard optimization approach. The standard approach is handicapped by the noisiness in forecasting covariances as well as the complex structure of the optimization. Instead, the alternative procedure chooses a portfolio which matches the benchmark along attributes such as size or book-to-market ratio. In general, the choice of benchmark and the set of available stocks determine the number of attributes necessary for matching. When we use the S&P 500 index as the benchmark, for example, the results suggest that two attributes, size and book-to-market, suffice (the tracking error volatility drops to 3.25%). In cases where a value or growth stock index is the benchmark, a larger number of attributes is needed to produce superior results.

In a realistic setting, such operational issues (the choice of forecasting model for covariances, the use of attributes or loadings) matter more in practice when the objective is to minimize portfolio tracking error volatility, rather than minimizing portfolio variance. Nonetheless, the imprecision with which return covariances are forecast does not detract from the main message that portfolio optimization helps substantially in risk reduction. At the same time, the low correlation between past and future covariances suggests that a dose of humility may not be the least important part of any risk optimization procedure.

Appendix: Variance Forecasting Models

We use several different models to forecast return variances. In each case, attributes of a stock are measured over one period. These attributes are related to return variances measured over a disjoint subsequent period (to maintain the predictive flavor of our tests). The estimated model is then used to generate forecasts, using the most recently observed attributes of the stock. Consider, for instance, the first set of forecasts made at the end of April 1968 (time t). The forecasting model is formulated using the most recent past 8 years of data (in this example extending back to April 1960). For each stock i, firm attributes, denoted $X_{i,t-1}$, are measured over the earliest 5 years (in the example these are measured over the period 1960–1965). These are related to variances $v_{i,t}$ measured over the 3 years (e.g., from 1965 to 1968) immediately prior to the forecast date using the model

$$v_{i,t} = X_{i,t-1}\phi + \epsilon_{i,t}. \tag{9}$$

We then step forward to the forecast date and update the firm attributes, $X_{i,t}$, using the most recent past 5 years (e.g., from 1963 to 1968). Forecasts are generated as $X_{i,t}\hat{\phi}$, where $\hat{\phi}$ is the least squares estimate from Equation (9).[15]

Model 1 forecasts future variances from past variances (after adjusting for realized variances' tendency to regress to the mean). Model 2 predicts a stock's variance from its loading on three factors: the market, the size factor, and the book-to-market factor. The loadings are estimated for each stock from a multiple regression using the past 60 months of returns. Note that while the model lets variances depend on the levels

[15] In Tables 1 and 2, we estimate variances based on the most recent 60 months of returns. To get corresponding results, we calibrate the variance forecasting models by adding an adjustment factor so that the mean of the forecast distribution matches the mean of the actual variances measured over the 60 months immediately prior to the forecast date. Note that the model estimation and calibration makes no use of information about realizations in the period subsequent to the forecast date.

Table A.1
Performance of variance forecasting models

Panel A: Properties of forecasted variances

Model	Mean	Standard deviation	Minimum	5th percentile	95th percentile	Maximum
(1) Regression-adjusted historical variance	0.0115	0.0050	0.0054	0.0064	0.0210	0.0456
(2) Loadings on 3 factors (market beta, size, book-to-market)	0.0115	0.0043	−0.0003	0.0047	0.0185	0.0248
(3) Dummy variables, 6 categories (3 attributes) (market beta, size, book-to-market)	0.0115	0.0043	0.0036	0.0050	0.0193	0.0220
(4) 3 attributes, dummy variable for zero dividends (market beta, size, dividend yield)	0.0115	0.0051	−0.0068	0.0038	0.0212	0.0259
(5) Dummy variables, 6 categories (4 attributes) (market beta, size, book-to-market, dividend yield, dummy variable for zero dividend yield)	0.0115	0.0048	0.0027	0.0049	0.0211	0.0247
(6) Combination (models 1, 2, 4)	0.0115	0.0044	0.0008	0.0053	0.0194	0.0300

Table A.1
(continued)

Panel B: Forecast performance based on subsequent 36 months

Model	Absolute forecast error: Mean	Median	95th percentile	Maximum	Correlation	Slope
(1) Regression-adjusted historical variance	0.0062	0.0043	0.0173	0.0913	0.5225	1.1930
(2) Loadings on 3 factors (market beta, size, book-to-market)	0.0065	0.0048	0.0167	0.0871	0.4758	1.1669
(3) Dummy variables, 6 categories (3 attributes) (market beta, size, book-to-market)	0.0066	0.0048	0.0175	0.0870	0.4467	1.1096
(4) 3 attributes, dummy variable for zero dividends (market beta, size, dividend yield)	0.0064	0.0047	0.0168	0.0878	0.5109	1.0619
(5) Dummy variables, 6 categories (4 attributes) (market beta, size, book-to-market, dividend yield, dummy variable for zero dividend yield)	0.0063	0.0045	0.0170	0.0852	0.5066	1.0759
(6) Combination (models 1, 2, 4)	0.0062	0.0045	0.0163	0.0868	0.5447	1.3202

At the end of April of each year from 1973 until 1995 a random sample of 250 firms is drawn from eligible domestic common stock issues on the NYSE and AMEX. Forecasts of monthly return variances are generated from six models, based on the prior 60 months of data for each stock. Summary statistics for the distribution of forecasted values are reported in panel A. Forecasts are then compared against the realized sample variances estimated over the subsequent 36 months in panel B. Summary statistics are provided for the distribution of the absolute difference between realized and forecasted values of variances. Also reported is the Pearson correlation between forecasts and realizations, and the slope coefficient in the regression of realizations on forecasts. The specification of the different models is described in the appendix.

of the loadings, variances are not restricted to be proportional to the squared loadings. Model 3 uses indicator variables based on the same set of estimated loadings on the three factors. Corresponding to each loading we define two dummy variables. If the loading is high (above the 80th percentile of the distribution of loading estimates of NYSE firms), the first dummy variable takes the value of one and the second dummy variable takes the value of zero; if the loading is low (below the 20th percentile of loading estimates of NYSE firms), the first dummy variable equals zero and the second dummy variable is set to one. In total, then, six dummy variables are defined. Model 4 uses the values of three stock characteristics: market beta, size, and dividend yield to predict future variances. Model 5 uses dummy variables based on four attributes (market beta, size, book-to-market, and dividend yield). The dummy variables are defined as in Model 3, yielding eight variables. Both models 4 and 5 also include an additional dummy variable to handle the case of zero-dividend yield (the dummy variable equals one if the dividend yield is zero, and is zero otherwise). Finally, model 6 is a composite forecasting model that gives equal weights to the forecasts from models 1, 3, and 5.[16]

Table A.1 reports the results for a subset of our variance forecasting models. The model based on regression-adjusted historical variance (model 1) is associated with the largest dispersion in forecast values. The range of forecasts from this model is 0.0402, and the standard deviation of forecasts is 0.50%. Nonetheless, its forecasting performance is as good as the other, more elaborate models in the table. Of all the individual forecasting models, the regression-adjusted historical variance model produces the lowest average absolute error (0.0062) as well as the highest correlation between forecasts and realizations (52.25%). Again, the poor out-of-sample performance of the higher-dimensional models may be attributable to data snooping. Combining the information in the loadings and attributes along with the information in historical variance to yield a composite forecast (model 6) raises the correlation to 54.47%, representing only a 4% relative improvement.

References

Alexander, G. J., 1978, “A Reevaluation of Alternative Portfolio Selection Models Applied to Common Stocks,” *Journal of Financial and Quantitative Analysis*, 13, 71–78.

BARRA, 1990, *The United States Equity Model: Handbook*, BARRA, Berkeley.

Best, M. J., and R. R. Grauer, 1991, “On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results,” *Review of Financial Studies*, 4, 315–342.

Chan, L. K. C., and J. Lakonishok, 1993, “Are the Reports of Beta’s Death Premature?,” *Journal of Portfolio Management*, 19, 51–62.

Chan, L. K. C., J. Karceski, and J. Lakonishok, 1998, “The Risk and Return from Factors,” *Journal of Financial and Quantitative Analysis*, 33, 159–188.

Chopra, V. K., and W. T. Ziemba, 1993, “The Effect of Errors in Means, Variances and Covariances on Optimal Portfolio Choice,” *Journal of Portfolio Management*, 19, 6–11.

Cohen, K. J., and J. A. Pogue, 1967, “An Empirical Evaluation of Alternative Portfolio Selection Models, *Journal of Business*, 40, 166–193.

Connor, G., and R. A. Korajczyk, 1988, “Risk and Return in an Equilibrium APT: Application of a New Test Methodology,” *Journal of Financial Economics*, 21, 255–289.

[16] All models also include a constant term. There is nothing in our models that rules out negative forecast values.

Connor, G., and R. A. Korajczyk, 1993, "A Test for the Number of Factors in an Approximate Factor Model," *Journal of Finance*, 48, 1263–1291.

Daniel, K., and S. Titman, 1997, "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns," *Journal of Finance*, 52, 1–33.

Elton, E. J., and M. J. Gruber, 1973, "Estimating the Dependence Structure of Share Prices — Implications for Portfolio Selection," *Journal of Finance*, 28, 1203–1232.

Elton, E. J., M. J. Gruber, and T. Urich, 1978, "Are Betas Best?," *Journal of Finance*, 23, 1375–1384.

Fama, E. F., and K. R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Fama, E. F., and K. R. French, 1997, "Industry Costs of Equity," *Journal of Financial Economics*, 43, 153–193.

Frost, P. A., and J. E. Savarino, 1986, "An Empirical Bayes Approach to Efficient Portfolio Selection," *Journal of Financial and Quantitative Analysis*, 21, 293–305.

Green, R. J., and B. Hollifield, 1992, "When Will Mean-Variance Efficient Portfolios be Well Diversified?," *Journal of Finance*, 47, 1785–1809.

Huber, P. J., 1977, "Robust Covariances," in S. S. Gupta and Dennis S. Moore (eds.), *Statistical Decision Theory and Related Topics II*, Academic Press, New York.

Jagannathan, R., and Z. Wang, 1992, "The Cross-Section of Expected Stock Returns: Do Size and Book to Market Equity Measure Systematic Risk Better than Beta?," working paper, University of Minnesota.

Jobson, J. D., and B. Korkie, 1981, "Putting Markowitz Theory to Work," *Journal of Portfolio Management*, 7, 70–74.

Ledoit, O., 1997, "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection," working paper, UCLA.

Markowitz, H. M., 1952, "Portfolio Selection," *Journal of Finance*, 7, 77–91.

Markowitz, H. M., 1959, *Portfolio Selection: Efficient Diversification of Investments*, Yale University Press, New Haven, CT.

Merton, R. C., 1980, "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.

Michaud, R. O., 1989, "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?," *Financial Analysts Journal*, 45, 31–42.

Nelson, D. B., 1992, "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model," *Journal of Econometrics*, 52, 61–90.

Pagan, A., and G. W. Schwert, 1990, "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics*, 45, 267–290.

Roll, R., 1992, "A Mean-Variance Analysis of Tracking Error," *Journal of Portfolio Management*, 18, 13–22.

Schwert, G. W., and P. J. Seguin, 1990, "Heteroskedasticity in Stock Returns," *Journal of Finance*, 45, 1129–1155.

Winston, Kenneth, 1993, "The 'Efficient Index' and Prediction of Portfolio Variance," *Journal of Portfolio Management*, 19, 27–34.