MSCI

Research Insight

Eigen-Adjusted Covariance Matrices

Improving Risk Forecasts for Optimized Portfolios

Jose Menchero

Jun Wang

D.J. Orr

May 2011

Introduction

The pioneering achievement of Markowitz (1952) gives concrete meaning to the concepts of "diversification" and "risk/return tradeoff," providing a solid theoretical foundation for the edifice of modern portfolio theory. The objective of the Markowitz framework is to construct mean-variance efficient portfolios with maximum expected return for a given level of risk. The optimization process requires two inputs: (a) the asset expected returns, and (b) the asset covariance matrix.

While the Markowitz framework is sound in theory, several pitfalls may complicate successful implementation in practice. For instance, Michaud (1989) argues that optimizers are essentially "error maximizers." The basic problem is that optimizers treat the inputs as if they were exact quantities, while in reality they can only be estimated with error. Optimizers, therefore, tend to place large bets on stocks with large estimation error in expected returns, often leading to poor out-of-sample performance. Similarly, optimizers may take large offsetting positions in assets with large in-sample correlations and small return differentials. If these correlations do not persist out-of-sample, the result may be higher-than-expected portfolio risk.

Muller (1993) demonstrated empirically that risk models do, in fact, tend to underestimate the risk of optimized portfolios. He suggested, as a rule of thumb, that volatility forecasts be scaled up by 20 percent to remove the biases. It should be stressed, however, that the amount of scaling depends on the risk model under consideration. For instance, a poorly conditioned covariance matrix may require a much larger correction to remove the optimization bias.

More recently, Shepard (2009) derived an analytic result for the magnitude of the volatility bias for optimized portfolios. Assuming normality and stationarity, he showed that the true expected volatility σ_{True} of an optimized portfolio is given by the predicted volatility σ_{Pred} divided by a correction factor

$$\sigma_{True} = \frac{\sigma_{Pred}}{1-\left(N/T\right)} , \qquad (1)$$

where N is the number of assets and T is the number of observations. For instance, if 100 observations are used to compute a covariance matrix for 50 assets, the realized risk of an optimized portfolio will be about *double* the predicted risk.

In this paper, we investigate the underlying sources for the biases of optimized portfolios. We identify special portfolios, termed *eigenfactors*, that exhibit large systematic biases in the risk forecasts. We show that the magnitude of these biases can be estimated by numerical simulation, and that the covariance matrix can be adjusted to remove these biases. We further demonstrate that removing the eigenfactor biases essentially removes the optimized portfolio biases as well. Finally, we study the performance of optimized portfolios, and show that the eigenfactor methodology is effective at reducing the out-of-sample volatilities of these portfolios.

Sample Covariance Matrices

We construct the sample covariance matrix using the *relative return* of stocks,

$$f_{nt} = r_{nt} - R_t^M \ , \tag{2}$$

where r_{nt} is the return of stock n on day t, and R_t^M is the “market” return, defined as the cap-weighted estimation universe return. Note that the cap-weighted relative returns, by construction, sum to zero each period.

For simplicity, we exclude from our estimation universe all stocks with incomplete histories. More specifically, we select the 50 largest US stocks as of July 30, 2010 that have full daily return histories dating back to July 1, 1993 (a total of 4,303 trading days).

We estimate the sample covariance matrix $\mathbf{V}_0$ using rolling windows of $T = 200$ days. The individual matrix elements are given by the standard expression,

$$\mathbf{V}_0(mn) = \frac{1}{T-1}\sum_{t=1}^{T}\left(f_{mt} - \bar{f}_m\right)\left(f_{nt} - \bar{f}_n\right) \ . \tag{3}$$

Here, the index n ranges over the number of stocks. In addition, we let the $n = 0$ element correspond to the market return, i.e., $f_{0t} = R_t^M$. In this formulation, every stock has an exposure of 1 to the market. Using a variable to represent the overall market is akin to using a world factor in a global equity risk model, as described by Menchero, Morozov, and Shepard (2010).

Bias Statistics

Bias statistics are used to test the accuracy of risk forecasts. Let R_t be the return of a test portfolio on day t, and let σ_t be the predicted volatility at the start of day. The *standardized return* is given by

$$b_t = \frac{R_t}{\sigma_t} \ , \tag{4}$$

and essentially represents a z-score. The standard deviation of the standardized returns, known as the bias statistic, is given by

$$B = \sqrt{\frac{1}{\tau - 1}\sum_{t=1}^{\tau}\left(b_t - \bar{b}\right)^2} \ , \tag{5}$$

where τ is the number of days in the testing window. Since 200 trading days are required to estimate the first covariance matrix, the testing window is 4,103 trading days (from April 15, 1994 to July 30, 2010).

Roughly speaking, the bias statistic represents the ratio of realized risk to predicted risk. Therefore, we expect $B \approx 1$ for accurate risk forecasts. Of course, the bias statistic will never be *exactly* 1, even for perfect risk forecasts. Instead, it is customary to identify a *confidence interval*. Assuming normality and perfect forecasts, the 95-percent confidence interval is approximately $1 \pm \sqrt{2/\tau}$.

For real financial data, however, this confidence interval is overly strict (especially for large τ). First, it is not possible to have a perfect forecast; at best, we can have an unbiased estimate of volatility. If the denominator of Equation 4 is noisy but unbiased, then by Jensen's inequality (see Wasserman 2004, for instance) we expect the average bias statistic to be slightly greater than 1. Second, real financial data have fat tails; this further reduces the number of observations falling within the confidence interval. In this paper, however, we will not concern ourselves with minor deviations from $B = 1$. Instead, our focus is to use the bias statistic as a tool to identify *large* systematic biases in risk forecasts.

Non-Optimized Portfolios

Individual stocks represent the first set of non-optimized portfolios that we consider. In Figure 1(a), we report stock-level bias statistics for the returns of Equation 2. The stocks are sorted in ascending order by their realized volatility over the sample period. We see that most of the bias statistics are quite close to 1, indicating that the sample covariance matrix $\mathbf{V}_0$ provides accurate risk forecasts at the individual stock level.

Next, we test the accuracy of risk forecasts on a set of 100 random portfolios. The returns in this case are given by

$$R_t^l = \sum_n \varepsilon_n^l f_{nt} \quad , \tag{6}$$

where f_{nt} are the asset returns of Equation 2, ε_n^l are drawn from a standard normal distribution, and l denotes the portfolio number. The weight of the market is set to zero (i.e., $\varepsilon_0^l = 0$), so that the random portfolios are strictly dollar neutral. In Figure 1(b) we show bias statistics for these random portfolios. Again, they are quite close to 1, indicating that the sample covariance matrix provides accurate risk forecasts for random portfolios as well.

Eigenfactors

The sample covariance matrix $\mathbf{V}_0$ provides the predicted covariance between any pair of assets, as defined by Equation 2. We refer to this as the "standard basis." The assets in the standard basis are intuitive and have clear financial interpretation. In this case, for instance, the assets correspond to portfolios that go long the stock and short the market. The off-diagonal terms, being non-zero, indicate that the assets are correlated in the standard basis.

The covariance matrix, however, can always be "rotated" to an alternative basis representing different linear combinations of the original assets. One basis that is of particular interest is the *diagonal* basis, in which the off-diagonal elements are all zero. The "assets" in the diagonal basis are termed *eigenfactors*, whose predicted variances are given by the diagonal matrix elements. The off-diagonal elements, now being zero, indicate that the eigenfactors are mutually uncorrelated.

The return of eigenfactor k on day t is given by

$$R_t^k = \sum_n u_{nt}^k f_{nt} \ , \tag{7}$$

where u_{nt}^k is the weight of asset n in the eigenfactor portfolio (see Appendix A for further details). Note that the number of eigenfactors is equal to the dimensionality of the sample covariance matrix[1]. We follow the convention of sorting the eigenfactors in ascending order by predicted volatility.

Eigenfactors are not economically intuitive. However, they do play an important role in portfolio optimization. For instance, the first eigenfactor solves for the minimum variance portfolio subject to the constraint that the sum of squared weights adds up to 1. Similarly, the *last* eigenfactor solves the corresponding *maximum* variance problem.

This suggests that the problem of underestimation of risk of optimized portfolios may be related to eigenfactors. To investigate this further, we compute bias statistics for the eigenfactor portfolios defined by Equation 7. The results are plotted in Figure 1(c): *we find a very strong relationship between the bias statistic of the eigenfactor and the eigenfactor number*. More specifically, the smallest eigenfactors (i.e., those with the lowest variances) have bias statistics far greater than 1, indicating that $\mathbf{V}_0$ strongly underestimates the risk of these portfolios. For the largest eigenfactors, by contrast, the sample covariance matrix slightly *overpredicts* risk. This should not be too surprising, perhaps, since the largest eigenfactor solves the maximum variance problem and, naturally, is also subject to estimation error.

Optimized Portfolios

While the eigenfactor biases are certainly intriguing, the biases of optimized portfolios are of more direct interest for portfolio construction purposes. To study these biases, we generate random alpha signals (for $j = 1, 100$)

$$\alpha_{nt}^j = \alpha_n^j - \bar{\alpha}_t^j \ , \tag{8}$$

[1] Mathematically, we have 51 eigenfactors, which we label $k = 0, ..., 50$. It is interesting to consider the scenario in which the stock weights are constant over the 200-day estimation window. In this case, the smallest eigenfactor $(k = 0)$ would have exactly zero variance. This represents the cap-weighted asset returns of Equation 2 summing to zero; i.e., $u_{0t}^0 = 0$, with u_{nt}^0 exactly proportional to the cap weights of the stocks. The $k = 0$ eigenfactor therefore reflects an exact colinearity in the return structure; it is not an investable portfolio since the market exposure $(n = 0)$ does equal the sum of stock exposures $(n = 1, 50)$. In reality, stock weights are not constant in time, and these results are only approximate.

where α_n^j is drawn from a standard normal distribution and $\overline{\alpha}_t^j$ is the cap-weighted mean of α_n^j on day t. This form ensures that the stock alphas α_{nt}^j are cap-weighted mean zero at the start of each day. We then use the sample covariance matrix $\mathbf{V}_0$ to construct the minimum volatility portfolio under the constraint[2] that the portfolio have an alpha of 1 (i.e., $\alpha_P = 1$). The portfolio return is expressed as

$$R_t^j = \sum_n h_{nt}^j f_{nt} \ , \tag{9}$$

where the optimized weights h_{nt}^j are defined in Appendix B.

In Figure 1(d), we see that the bias statistics for the optimized portfolios range between 1.4 and 1.5, with the average being about 1.45. According to Equation 1, the expected bias statistic for $N = 50$ and $T = 200$ should be 1.33. This suggests that deviations from normality and stationarity in real financial data magnify the effects of estimation error and that stronger corrections are required to completely remove the biases of optimized portfolios.

Simulated Eigenfactor Biases

The previous section demonstrates that the sample covariance matrix $\mathbf{V}_0$ produces systematic biases in the risk forecasts of eigenfactor portfolios. In this section, we investigate whether these biases can be estimated by numerical simulation. Technical details are provided in Appendix A.

The true covariance matrix, of course, is unobservable. Instead, we are restricted to a single sample that we can only observe with estimation error. However, we will *pretend* for a moment that the sample covariance matrix $\mathbf{V}_0$ represents the "true" covariance matrix. We then use the "true" covariance matrix to simulate M sets of stock returns[3] that we use, in turn, to estimate M simulated covariance matrices $\mathbf{V}_m$ $(m = 1, M)$. We diagonalize the simulated covariance matrices to construct the simulated eigenfactors. We also use the simulated covariance matrices $\mathbf{V}_m$ to compute the predicted volatilities of the simulated eigenfactors. However, since we know the actual distribution that generated the simulated returns, we can therefore use the sample covariance matrix $\mathbf{V}_0$ to compute the "true" volatilities of the simulated eigenfactors. Comparing the true volatilities to the predicted volatilities allows us to estimate the simulated biases.

We repeat this exercise for every day t over the entire sample period and study the nature of the simulated volatility biases across time. The simulated volatility bias of eigenfactor k at time t is given by

---

[2] We also impose the investability constraint that the market exposure $(n = 0)$ equal the sum of the stock exposures $(n = 1, 50)$.

[3] Our simulations assume normality and stationarity.

$$\lambda_t(k) = \frac{1}{M} \sum_m \frac{\tilde{\sigma}_{mt}(k)}{\sigma_{mt}(k)} \ , \qquad (10)$$

where $\tilde{\sigma}_{mt}(k)$ is the "true" volatility (given by $\mathbf{V}_0$) of the simulated eigenfactor, and $\sigma_{mt}(k)$ is the predicted volatility (given by $\mathbf{V}_m$). To understand the average simulated bias across time, we compute

$$\bar{\lambda}(k) = \frac{1}{T} \sum_t \lambda_t(k) \ , \qquad (11)$$

and plot the results in Figure 2. The smallest eigenfactor has a mean simulated bias of about 1.5, whereas the largest eigenfactor has a mean bias of about 0.96. Qualitatively, Figure 2 is in excellent agreement with Figure 1(c), indicating that numerical simulation can indeed capture the main features of the empirical biases. Quantitatively, however, we see that the empirical biases in Figure 1(c) are slightly larger than the simulated biases in Figure 2. This can again be attributed to deviations from normality and stationarity, which are assumed in the simulations but are violated in practice.

It is also informative to study the *stability* of the simulated volatility bias across time. To this end, we sort the values $\lambda_t(k)$ in ascending order for each eigenfactor k, and let $\lambda^p(k)$ denote the *p*-percentile value. In Figure 2, we plot $\lambda^p(k)$ for $p = 1$ and $p = 99$ percentiles. We see that the eigenfactor biases are remarkably stable across time. For instance, over roughly a 16-year period, the simulated volatility bias for the smallest eigenfactor varies within the narrow range of 1.4 to 1.6 about 98 percent of the time. Similarly, for the largest eigenfactor, the range varies from 0.9 to 1.0 over this same time period.

Eigen-Adjusted Covariance Matrices

In the previous section, we showed that numerical simulation could be used to estimate the volatility biases of eigenfactors. The simulations, however, assume normality and stationarity, which are violated in practice. Consequently, additional scaling is required to fully remove the biases of the eigenfactors, as described in Appendix A (see Equation A8).

Once we have estimated the size of the eigenfactor biases, we assume that the sample covariance matrix $\mathbf{V}_0$, which uses the same estimator as the simulated covariance matrices $\mathbf{V}_m$, also suffers from the same biases. We then rotate the sample covariance matrix to the diagonal basis and de-bias the eigenvariances. The final step is to "rotate back" to the standard basis representing the individual stocks; this leads to the eigen-adjusted covariance matrix that we denote $\tilde{\mathbf{V}}_0$. The procedure is described in more detail in Appendix A.

Note that the diagonal elements of $\mathbf{V}_0$ and $\tilde{\mathbf{V}}_0$, in general, differ. Therefore, it is possible that the eigen-adjustment procedure will induce biases in the volatility forecasts of individual stocks. To investigate this, we compute stock-level bias statistics for the returns of Equation 2. The results are reported in Figure 3(a), with the stocks again rank-ordered from low volatility to high volatility. We find a definite relationship between the bias statistic and the volatility of the stock. More specifically, the lowest volatility stocks have bias statistics of about 0.90, whereas the highest volatility stocks have bias statistics of about 1.10.

We can understand these biases as follows. First, note that Equation 7 expresses every eigenfactor as a linear combination of stocks. This relationship can also be inverted. That is, every stock can be expressed as linear combination of eigenfactors,

$$f_{nt} = \sum_k u_{kt}^n R_t^k \quad , \tag{12}$$

where u_{kt}^n is the "weight" of eigenfactor k in stock n[4]. Mathematically, the coefficient u_{kt}^n measures how strongly a given stock "projects" onto each eigenfactor. We find that low-volatility stocks tend to project more strongly on the small eigenfactors, whereas the high-volatility stocks preferentially project on the large eigenfactors.

To illustrate this effect, we consider the projections on analysis date July 30, 2010 for two particular stocks: Chevron and Citigroup. The former had the lowest predicted volatility (13.7 percent) on the analysis date, whereas the latter had the highest predicted volatility (31.7 percent). We compute the squared projection coefficients $(u_{kt}^n)^2$ and aggregate them within eigenfactor bins[5]. The results are presented in Figure 4. We clearly see that Citigroup projects primarily on the largest eigenfactors, whereas Chevron projects mostly on the smallest ones. Since the smallest (or largest) eigenfactors have their volatilities increased (or decreased) by the eigen-adjustment methodology, this explains the observed biases in Figure 3(a).

While the eigenvariance adjustment does induce small biases at the stock level, note that *the average bias is about zero*. In other words, some stocks are biased slightly upward, while others are biased slightly downward. This is in contrast to Figure 1(d), which shows that for optimized portfolios *all* risk forecasts are systematically biased downward by a large amount.

Of course, the main purpose of a risk model is to compute *portfolio* risk, not individual stock risk. It is likely that the biases observed at the stock level largely cancel at the portfolio level. To test this hypothesis, we compute bias statistics using the same random portfolios given in Equation 6. The results are shown in Figure 3(b). We see that the average bias statistics are indeed very close to 1, indicating that the eigen-adjusted covariance matrix produces accurate risk forecasts for non-optimized portfolios.

We now study the bias statistics of the eigenfactor portfolios using $\tilde{\mathbf{V}}_0$. Note that the eigen-adjustment method does not affect the composition of the eigenfactors themselves, whose returns are still given by Equation 7. The bias statistics are plotted in Figure 3(c). We see that the eigen-adjusted covariance matrix has essentially removed the eigenfactor biases across the entire spectrum.

Next, we consider optimized portfolios with returns given by

$$\tilde{R}_t^j = \sum_n \tilde{h}_{nt}^j f_{nt} \quad . \tag{13}$$

---

[4] Note that u_{kt}^n can also be interpreted as the weight of stock k in eigenfactor n. Mathematically, this represents the u_{nt}^k element of the "transpose" matrix.

[5] An important property of the projection coefficients is that $\sum_k (u_{kt}^n)^2 = 1$.

To obtain these holdings, we use the same alphas given by Equation 8, but now construct the holdings using the eigen-adjusted covariance matrix $\tilde{\mathbf{V}}_0$. In Figure 3(d), we report the bias statistics for the optimized portfolios; we see that the biases have been essentially removed. Also note that the results are virtually indistinguishable from Figure 3(b), indicating that $\tilde{\mathbf{V}}_0$ provides equally accurate risk forecasts for both optimized and non-optimized portfolios.

Out-of-Sample Volatilities

Removing the biases of optimized portfolios resolves a long-outstanding problem in risk modeling and quantitative investing. Nevertheless, of even greater importance to the quantitative manager is the out-of-sample *performance* of optimized portfolios.

The first optimized portfolio that we consider is the minimum-risk fully invested portfolio. This is crucial for portfolio construction purposes, since it constitutes the left-most point on the Markowitz efficient frontier. Reducing the out-of-sample volatility of this portfolio effectively raises the efficient frontier.

We form the minimum-risk fully invested portfolio in two ways: one uses the sample covariance matrix $\mathbf{V}_0$; the other uses the eigen-adjusted covariance matrix $\tilde{\mathbf{V}}_0$. The portfolio holdings, described in Appendix B, are rebalanced daily. The cap-weighted estimation universe has a realized volatility of 19.84 percent (annualized) over the out-of-sample period April 15, 1994 to July 30, 2010. Over this same period, the minimum-risk fully invested portfolio constructed using $\mathbf{V}_0$ has a realized volatility of only 14.64 percent, showing that the sample covariance matrix is an effective tool for hedging risk. Nonetheless, the minimum-risk fully invested portfolio constructed using $\tilde{\mathbf{V}}_0$ has an even lower realized volatility at 13.98 percent. In other words, the eigenfactor approach reduces the out-of-sample volatility by about 4.5 percent relative to the sample covariance matrix.

Another important class of optimized portfolios are the minimum-risk portfolios subject to $\alpha_P = 1$. We construct two such sets of optimized portfolios: one is obtained using $\mathbf{V}_0$, which results in the portfolios of Equation 9; the other is constructed using $\tilde{\mathbf{V}}_0$, which produces the portfolios of Equation 13. Let $\tilde{\sigma}_l^R$ denote the realized out-of-sample volatility for optimized portfolio l using the eigen-adjusted covariance matrix $\tilde{\mathbf{V}}_0$, and let σ_l^R denote the corresponding volatility using the sample covariance matrix $\mathbf{V}_0$. We define the realized volatility ratio as

$$v_l^R = \frac{\tilde{\sigma}_l^R}{\sigma_l^R} , \tag{14}$$

where $l = 1,100$ denotes the optimized portfolio number. In Figure 5, we plot the resulting histogram of realized volatility ratios. We see that for every portfolio, the eigenfactor method leads to lower out-of-sample volatility. The mean realized volatility ratio is 0.936, which represents a 6.4 percent reduction in risk on average. Since all portfolios by construction have the same alpha $(\alpha_P = 1)$, this translates into a nearly 7 percent increase in Information Ratio.

N/T Effects

Thus far, we have considered $N = 50$ stocks with covariance matrices estimated over $T = 200$ periods. In this section, we vary the number of periods used to compute the covariance matrices from $T = 60$ to $T = 500$.

We again consider minimum-risk optimized portfolios with $\alpha_P = 1$. We form one set of optimized portfolios using $\mathbf{V}_0$, with holdings given by Equation 9. The other set of optimized portfolios is constructed using $\tilde{\mathbf{V}}_0$, leading to the holdings given by Equation 13.

In Table 1, we present bias statistics and mean-realized volatilities for the two sets of optimized portfolios. We make several interesting observations. First, the portfolios constructed using the eigen-adjusted covariance matrix have lower realized volatilities for every value of T. Second, the performance gap in realized volatility between the two methods increases in inverse proportion to the number of periods used to construct the covariance matrix. For instance, using $T = 500$ observations results in a difference of only 6 bps, whereas for $T = 60$ the realized volatilities using $\mathbf{V}_0$ are more than double those obtained using the eigen-adjusted covariance matrix $\tilde{\mathbf{V}}_0$. Third, in all cases the bias statistics are much closer to 1 using the eigen-adjusted covariance matrices. Fourth, the bias statistics using the sample covariance matrix increase dramatically as we decrease the number of periods in the estimation window. This is in accordance with Equation 1. Note, however, that the Equation 1 consistently underpredicts the actual bias. For example, for $T = 100$ days, Equation 1 predicts a bias statistic of 2.0, whereas Table 1 shows the actual bias statistic is 2.25.

Conclusion

We demonstrate that the sample covariance matrix produces biased risk forecasts for eigenfactor portfolios. This in turn leads to the classic problem of underestimation of risk for optimized portfolios. By de-biasing the eigenvariances, we effectively remove the biases in risk forecasts for optimized portfolios. We also demonstrate that the eigenfactor methodology is effective at reducing the out-of-sample volatilities of these portfolios.

Appendix A: Eigenvariance De-biasing

Let $\mathbf{V}_0$ denote the $N\text{x}N$ sample covariance matrix, computed as

$$\mathbf{V}_0 = \frac{\mathbf{f} \cdot \mathbf{f}'}{T-1} \ , \tag{A1}$$

where $\mathbf{f}$ is the $N\text{x}T$ matrix of realized asset returns and T is the number of periods. The "assets" could represent factor portfolios, asset classes, or individual securities. We assume that the sample covariance matrix is full rank, which requires that the number of periods T exceeds the number of assets N.

The sample covariance matrix can be expressed in diagonal form as

$$\mathbf{D}_0 = \mathbf{U}_0'\mathbf{V}_0\mathbf{U}_0 \ , \tag{A2}$$

where $\mathbf{U}_0$ is the $N\text{x}N$ rotation matrix whose columns are given by the eigenvectors of $\mathbf{V}_0$. The n^{th} element of the k^{th} column of $\mathbf{U}_0$ gives the weight of stock n in eigenfactor portfolio k, as described by Equation 7 of the main text. These eigenfactors represent mutually uncorrelated portfolios of assets. The diagonal elements of $\mathbf{D}_0$ represent the predicted variances of the eigenfactors.

Although the true covariance matrix is unobservable, we suppose for simulation purposes that the sample covariance matrix $\mathbf{V}_0$ governs the "true" return-generating process. We generate a set of asset returns for simulation m as

$$\mathbf{f}_m = \mathbf{U}_0\mathbf{b}_m \ , \tag{A3}$$

where $\mathbf{b}_m$ is an $N\text{x}T$ matrix of simulated eigenfactor returns. The elements of row k of $\mathbf{b}_m$ are drawn from a random normal distribution with mean zero and variance given by the diagonal element $D_0(k)$ of matrix $\mathbf{D}_0$. It can be easily verified that the simulated returns in Equation A3 have a true covariance matrix given by $\mathbf{V}_0$. Due to sampling error, however, the *estimated* covariance matrix

$$\mathbf{V}_m = \frac{\mathbf{f}_m \cdot \mathbf{f}_m'}{T-1} \ , \tag{A4}$$

will differ from the true covariance matrix $\mathbf{V}_0$. Nevertheless, $\mathbf{V}_m$ is unbiased in the sense that $E\left[\mathbf{V}_m\right] = \mathbf{V}_0$. We diagonalize the simulated covariance matrix

$$\mathbf{D}_m = \mathbf{U}_m'\mathbf{V}_m\mathbf{U}_m \ , \tag{A5}$$

where $\mathbf{U}_m$ denotes the simulated eigenfactors with estimated variances given by the diagonal elements of $\mathbf{D}_m$, i.e., $D_m(k)$.

Since we know the true distribution that governs the simulated asset returns, we can compute the true covariance matrix of the simulated eigenfactors,

$$\tilde{\mathbf{D}}_m = \mathbf{U}'_m \mathbf{V}_0 \mathbf{U}_m \ . \tag{A6}$$

Note that since $\mathbf{U}_m$ does not represent the "true" eigenfactors, the matrix $\tilde{\mathbf{D}}_m$ is not diagonal. In principle, risk forecasts could be improved by adjusting the directions of the eigenfactors. In practice, we do not find such a benefit. Our focus here is on adjusting the *variances* of the eigenfactors.

As described in the main text, the predicted variances of the simulated eigenfactors are biased. That is, $E\left[D_m(k)\right] \neq \tilde{D}_m(k)$. We compute the simulated volatility biases

$$\lambda(k) = \frac{1}{M}\sum_m \sqrt{\frac{\tilde{D}_m(k)}{D_m(k)}} \ , \tag{A7}$$

where M is the total number of simulations.

Our simulations assume both normality and stationarity. Real financial data, of course, violate both of these assumptions. In practice, therefore, additional scaling is required to remove the biases of the eigenfactors. To obtain the "empirical volatility bias," we simply scale the simulated volatility biases based on their deviation from 1,

$$\gamma(k) = a\left[\lambda(k) - 1\right] + 1 \ , \tag{A8}$$

where a is a constant. Empirically, we find that $a = 1.4$ is effective at removing the biases across the entire eigenfactor spectrum. In Table 1, we also use $a = 1.4$ for all values of T.

We now assume that the sample covariance matrix $\mathbf{V}_0$, which uses the same estimator as the simulated covariance matrices $\mathbf{V}_m$, also suffers from the same biases. Let $\tilde{\mathbf{D}}_0$ denote the diagonal covariance matrix whose eigenvariances have been de-biased

$$\tilde{\mathbf{D}}_0 = \boldsymbol{\gamma}^2 \mathbf{D}_0 \ , \tag{A9}$$

where $\boldsymbol{\gamma}^2$ is a diagonal matrix whose elements are given by $\gamma^2(k)$. The de-biased covariance matrix is now rotated back from the diagonal basis to the standard basis using the sample eigenfactors,

$$\tilde{\mathbf{V}}_0 = \mathbf{U}_0 \tilde{\mathbf{D}}_0 \mathbf{U}'_0 \ . \tag{A10}$$

This represents the eigen-adjusted covariance matrix.

Appendix B: Holdings of Optimal Portfolios

Let $\mathbf{V}$ denote an asset covariance matrix. As shown by Grinold and Kahn (2000), the holdings of the minimum-risk fully invested portfolio are given by

$$\mathbf{h}_C = \frac{\mathbf{V}^{-1}\mathbf{e}}{\mathbf{e}'\mathbf{V}^{-1}\mathbf{e}} \ , \tag{B1}$$

where $\mathbf{e}$ is an N-dimensional column vector with 1 in every entry. Similarly, the minimum-risk portfolio with $\alpha_P = 1$ is given by

$$\mathbf{h}_A = \frac{\mathbf{V}^{-1}\boldsymbol{\alpha}}{\boldsymbol{\alpha}'\mathbf{V}^{-1}\boldsymbol{\alpha}} \ , \tag{B2}$$

where $\boldsymbol{\alpha}$ is an N-dimensional column vector of stock alphas.

REFERENCES

Grinold, R., and R. Kahn. 2000. *Active Portfolio Management.* New York: McGraw-Hill.

Markowitz, Harry. 1952. “Portfolio Selection.” *Journal of Finance*, vol. 7: 77-91.

Menchero, Jose, Andrei Morozov, and Peter Shepard. 2010. “Global Equity Risk Modeling.” In J. Guerard, ed., *The Handbook of Portfolio Construction: Contemporary Applications of Markowitz Techniques.* New York, NY: Springer.

Michaud, Richard. 1989. “The Markowitz Optimization Enigma: Is ‘Optimized’ Optimal?” *Financial Analysts Journal,* vol. 45, no. 1 (January/February): 31-42.

Muller, Peter. 1993. “Financial Optimization.” *Cambridge University Press* (Zenios): 80-98.

Shepard, Peter. 2009. “Second Order Risk.” *Working paper, http://arxiv.org/abs/0908.2455v1*.

Wasserman, Larry. 2004. *All of Statistics.* New York: Springer.

Figure 1

Bias statistics for (a) stocks, (b) random portfolios, (c) eigenfactors, and (d) optimized portfolios. Return sources are defined by Equation 2, Equation 6, Equation 7, and Equation 9, respectively. Bias statistics were computed using the sample covariance matrix $\mathbf{V}_0$; the out-of-sample testing period comprised 4103 trading days from April 15, 1994 to July 30, 2010.

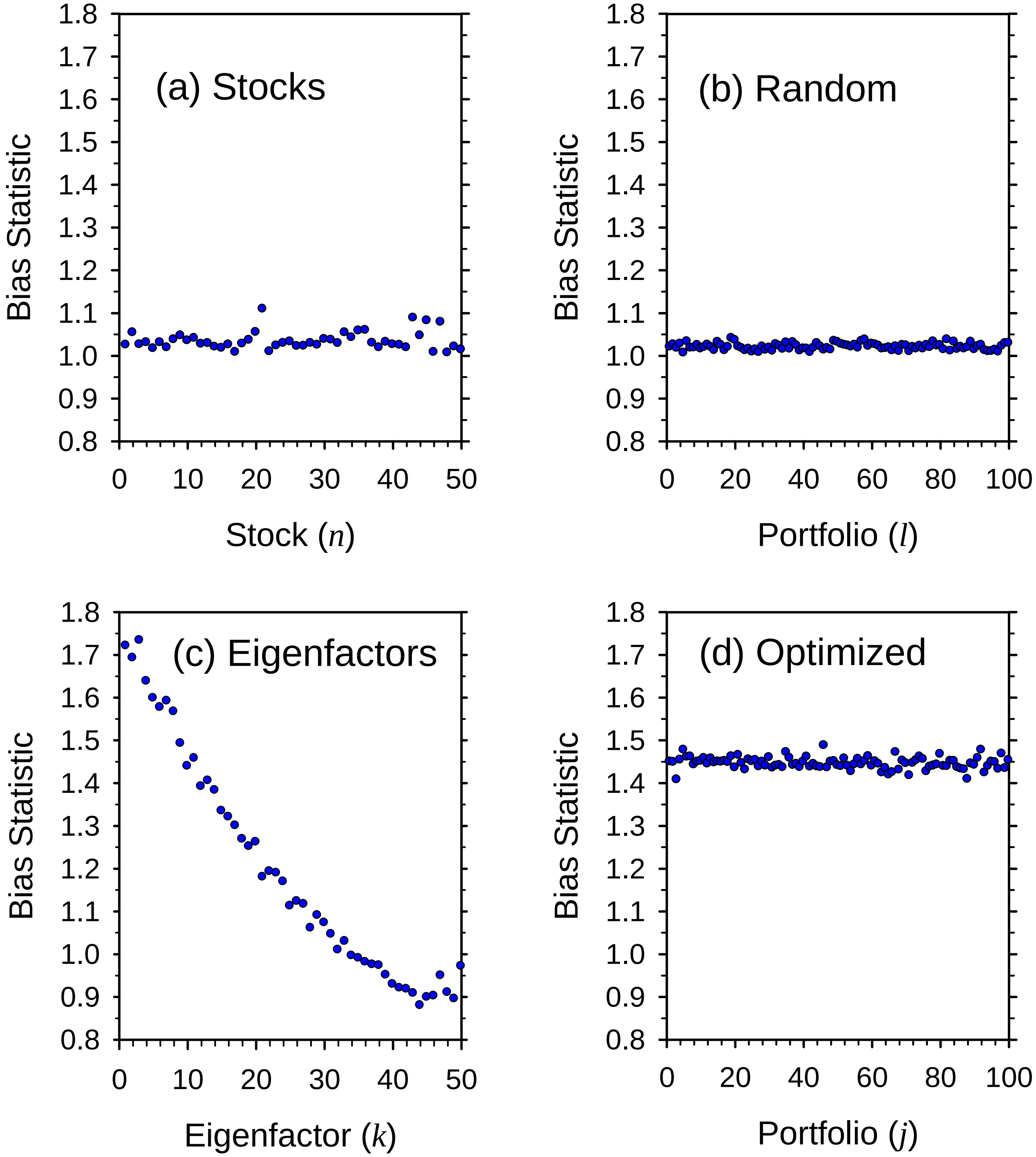

Figure 2

Simulated volatility biases for eigenfactors. The circles give the mean-simulated volatility bias over 4,103 trading days (April 15, 1994 to July 30, 2010). The up triangles give the 1-percentile bias, and the down triangles give the 99-percentile bias.

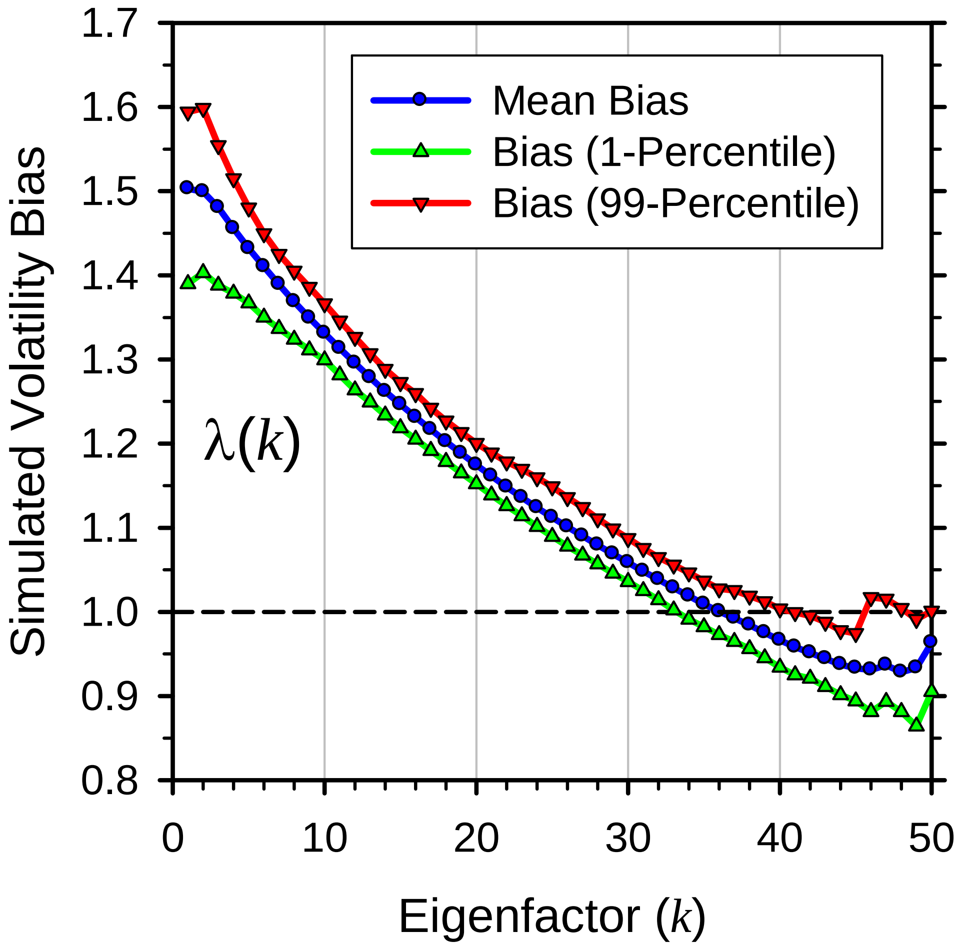

Figure 3

Bias statistics for (a) stocks, (b) random portfolios, (c) eigenfactors, and (d) optimized portfolios. Return sources are defined by Equation 2, Equation 6, Equation 7, and Equation 13, respectively. Bias statistics were computed using the eigen-adjusted covariance matrix $\tilde{\mathbf{V}}_0$; the out-of-sample testing period comprised 4103 trading days from April 15, 1994 to July 30, 2010.

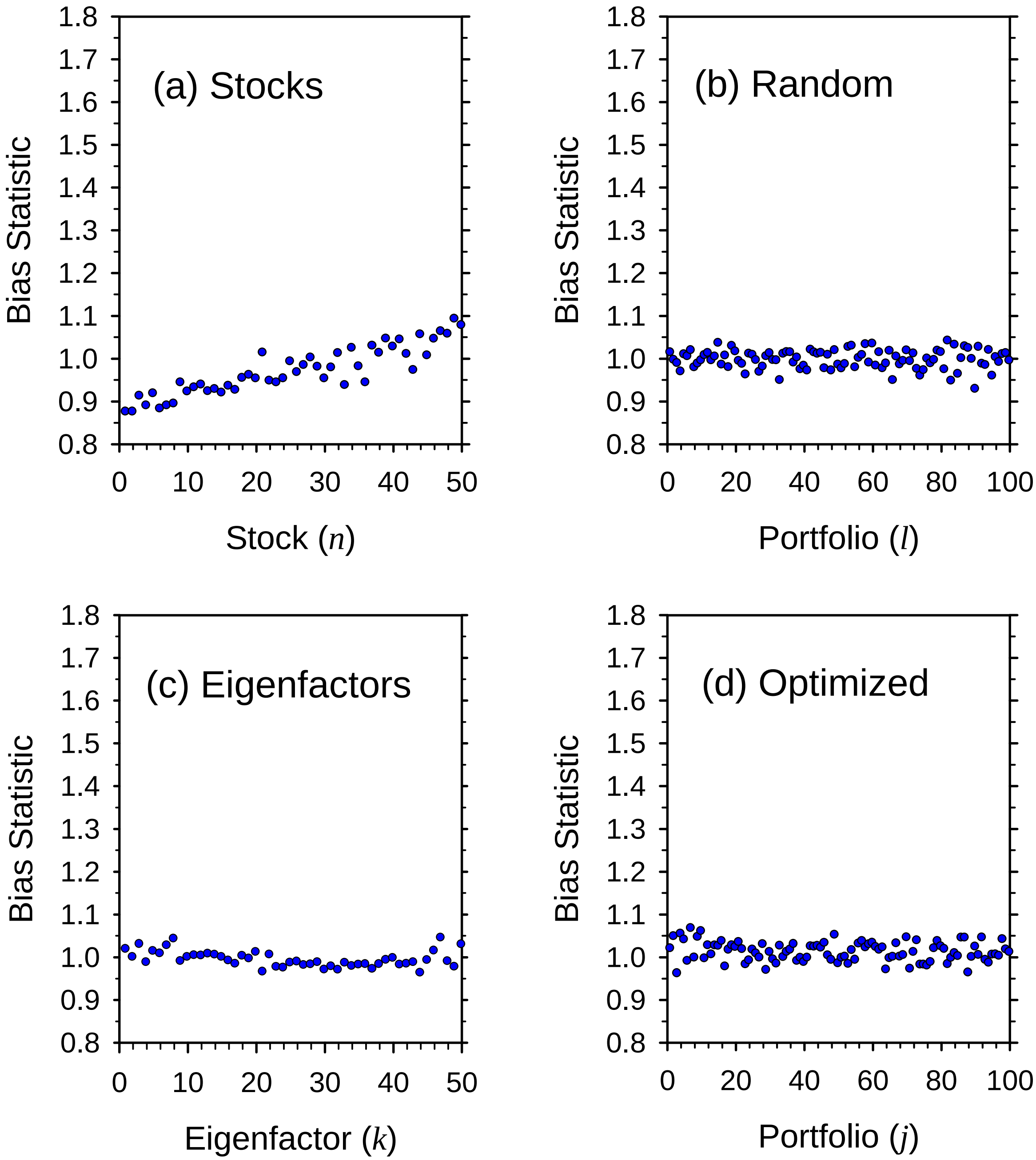

Figure 4

Histogram of squared projection coefficients for Chevron (lowest-volatility stock) and Citigroup (highest-volatility stock), as of July 30, 2010. The results are aggregated into eigenfactor bins. Low-volatility stocks tend to project on small eigenfactors, whereas high-volatility stocks project primarily on the largest eigenfactors.

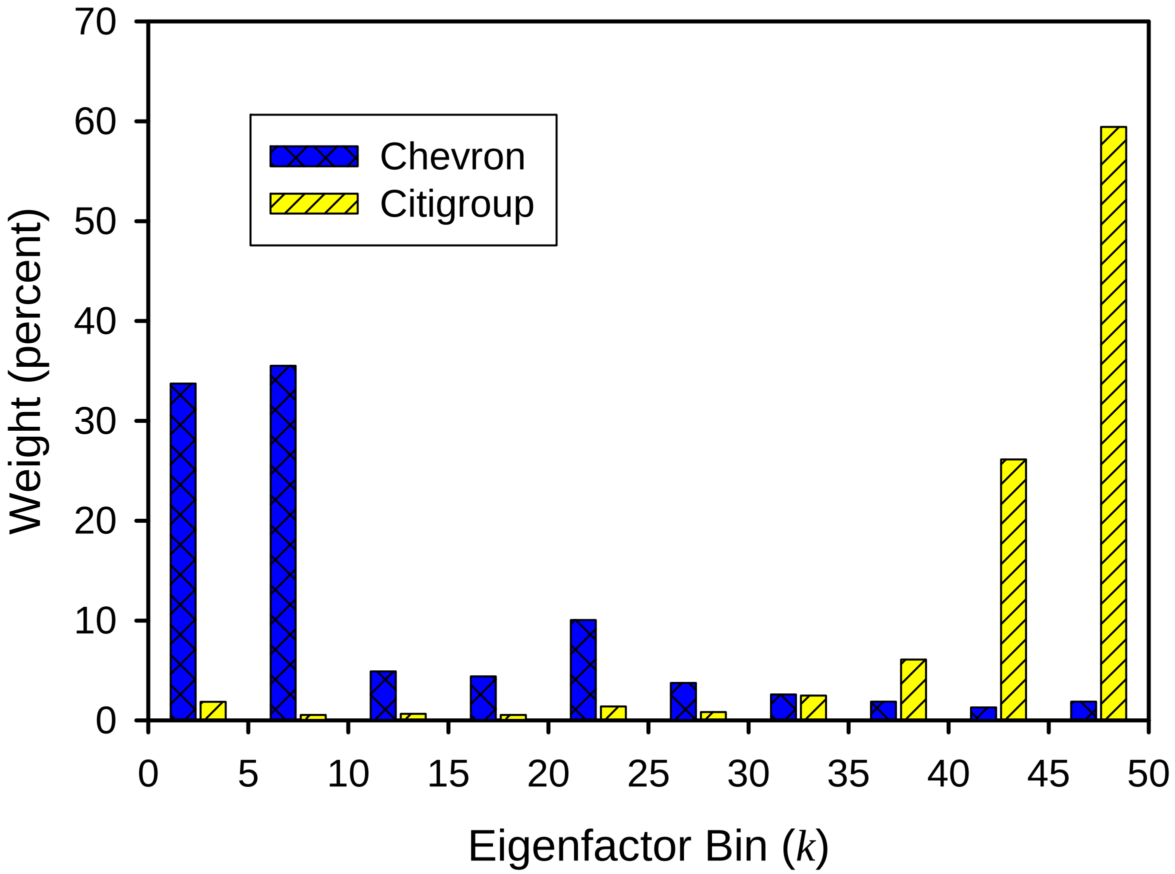

Figure 5

Histogram of realized volatility ratio, $\tilde{\sigma}_l^R / \sigma_l^R$. Here, $\tilde{\sigma}_l^R$ is the realized volatility of optimized portfolio l using the eigen-adjusted covariance matrix $\tilde{\mathbf{V}}_0$, and σ_l^R is the realized volatility using the sample covariance matrix $\mathbf{V}_0$. The out-of-sample testing period is 4,103 trading days (from April 15, 1994 to July 30, 2010). In all cases, the eigen-adjusted approach leads to lower out-of-sample volatilities.

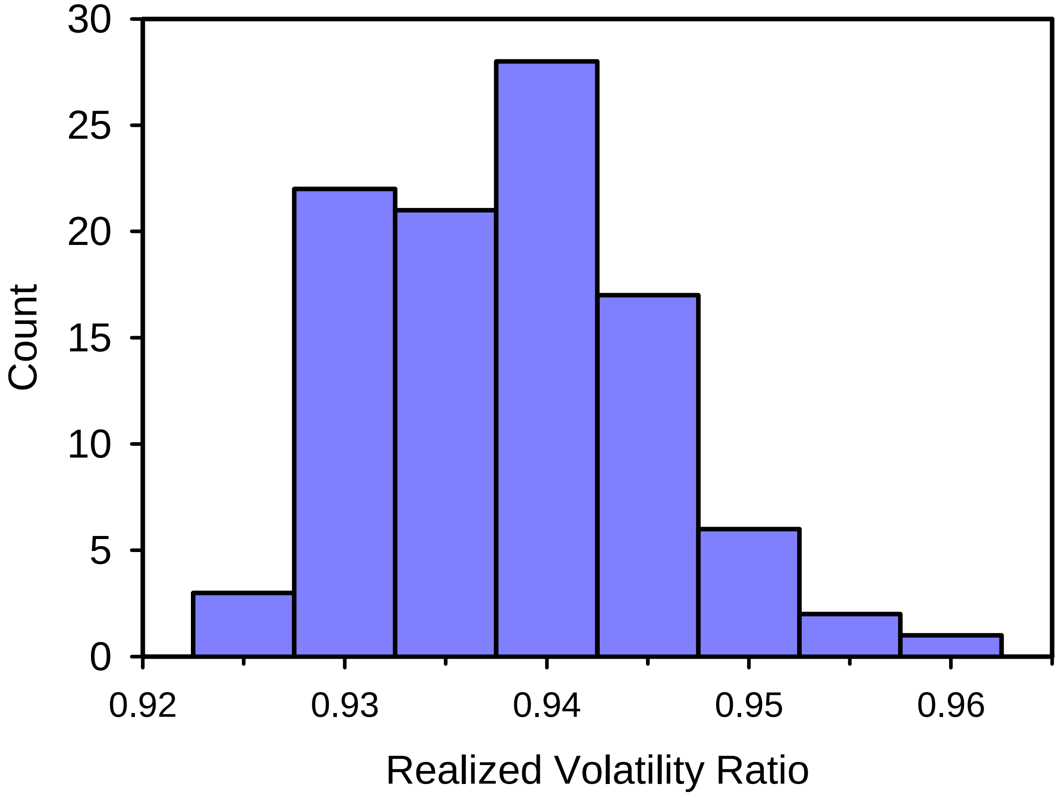

Table 1

N/T effects for optimized portfolios with $\alpha_P = 1$. The number of stocks in the estimation universe is fixed at $N = 50$, while the number of days used to estimate the covariance matrix varies from $T = 60$ to $T = 500$. The out-of-sample testing period is 3797 trading days (from July 3, 1995 to July 30, 2010).

	(Sample Covariance Matrix)		(Eigen-adjusted Matrix)	
Days (T)	Bias Stats	Realized Vol	Bias Stats	Realized Vol
60	7.78	8.04	0.89	3.86
100	2.25	4.57	0.97	3.66
200	1.45	3.77	1.02	3.53
300	1.30	3.60	1.03	3.48
400	1.23	3.53	1.04	3.46
500	1.20	3.51	1.05	3.45

Client Service Information is Available 24 Hours a Day

clientservice@msci.com

Americas

Americas	1.888.588.4567 (toll free)
Atlanta	+ 1.404.551.3212
Boston	+ 1.617.532.0920
Chicago	+ 1.312.675.0545
Montreal	+ 1.514.847.7506
Monterrey	+ 52.81.1253.4020
New York	+ 1.212.804.3901
San Francisco	+ 1.415.836.8800
Sao Paulo	+ 55.11.3706.1360
Stamford	+1.203.325.5630
Toronto	+ 1.416.628.1007

Europe, Middle East & Africa

Amsterdam	+ 31.20.462.1382
Cape Town	+ 27.21.673.0100
Frankfurt	+ 49.69.133.859.00
Geneva	+ 41.22.817.9777
London	+ 44.20.7618.2222
Madrid	+ 34.91.700.7275
Milan	+ 39.02.5849.0415
Paris	0800.91.59.17 (toll free)
Zurich	+ 41.44.220.9300

Asia Pacific

China North	10800.852.1032 (toll free)
China South	10800.152.1032 (toll free)
Hong Kong	+ 852.2844.9333
Seoul	+827.0768.88984
Singapore	800.852.3749 (toll free)
Sydney	+ 61.2.9033.9333
Tokyo	+ 81.3.5226.8222

Notice and Disclaimer

About MSCI

MSCI Inc. is a leading provider of investment decision support tools to investors globally, including asset managers, banks, hedge funds and pension funds. MSCI products and services include indices, portfolio risk and performance analytics, and governance tools.

The company's flagship product offerings are: the MSCI indices which include over 148,000 daily indices covering more than 70 countries; Barra portfolio risk and performance analytics covering global equity and fixed income markets; RiskMetrics market and credit risk analytics; ISS governance research and outsourced proxy voting and reporting services; FEA valuation models and risk management software for the energy and commodities markets; and CFRA forensic accounting risk research, legal/regulatory risk assessment, and due-diligence. MSCI is headquartered in New York, with research and commercial offices around the world.