

Performance of Portfolios Optimized with Estimation Error

Andrew F. Siegel

Business School, University of Washington, Seattle, Washington 98195-3200,

asiegel@u.washington.edu

Artemiza Woodgate

Business School, University of Washington, Seattle, Washington 98195-3200,

artemiza@u.washington.edu

We explain the poor out-of-sample performance of mean-variance optimized portfolios, developing theoretical bias adjustments for estimation risk by asymptotically expanding future returns of portfolios formed with estimated weights. We provide closed-form non-Bayesian adjustments of classical estimates of portfolio mean and standard deviation. The adjustments significantly reduce bias in international equity portfolios, increase economic gains, and are robust to sample size and to non-normality. Dominant terms grow linearly with the number of assets and decline inversely with the number of past time periods. Under suitable conditions, Sharpe-ratio maximizing tangency portfolios become more diversified. Using these methods it is possible to assess, before investing, the effect of statistical estimation error on portfolio performance.

Key words: Investments; Portfolio performance; Estimation error; Statistical noise correction; Capital market line adjustment

Andrew F. Siegel holds the Grant I. Butterbaugh Professorship at the University of Washington.

1. Introduction

Since the development of portfolio theory by Markowitz (1952), mean-variance efficient portfolios have received considerable attention, playing a key role in a variety of financial fields from investment analysis and asset pricing to topics in corporate finance. It is well known that the often-disappointing performance of mean-variance optimized portfolios, in large part, stems from the use of past performance to estimate the unknown parameters of the assets' probability distribution. Even when a statistical estimator is unbiased, bias can emerge when the estimator is used as an input to a nonlinear optimization process. Finding a solution to this combined estimation / optimization problem has been pursued in various ways, both theoretically, primarily using Bayesian methods, and empirically. Yet the current literature does not provide theoretical guidance as to the exact size of the impact of estimation error, or as to when and how mean-variance techniques should be used. This paper gives that analysis, by proving asymptotic non-Bayesian closed-form formulas for portfolio performance while accounting for estimation risk. This is the first theoretical paper in the existing literature to quantify the impact of estimation error without relying upon prior assumptions on the unknown parameters, and to give an adjustment for statistical noise that will asymptotically reflect the actual performance, within the Markovitz mean-variance framework.

We perform the following thought experiment. Suppose an investor forms classical sample estimates of asset means, variances and covariances and uses them as if they were the true assets' distribution parameters to form mean-variance efficient portfolios and to judge their anticipated performance. We refer to these classical sample performance estimates as "naïve" estimates, since they do not take into account the estimation error stemming from using a sample rather than the population in determining these parameters. If the estimation error wrongly suggests that an asset will have a high expected return, then an optimized portfolio heavily invested in this asset will be disappointing. In particular, the naïve portfolio mean estimates tend to be biased upwards and the naïve portfolio variance estimates tend to be biased downwards, resulting in nominally efficient portfolios that are "over-optimistic," in the sense that such an investor will believe she can achieve a higher mean and lower

variance than is actually available from the performance of her portfolio. We quantify this “over-optimism” bias in closed-form asymptotic formulas.

Our contributions are derived within a new theoretical framework, giving the investor an adjustment to the performance of naïvely-estimated efficient portfolios that will more accurately reflect actual portfolio performance by accounting for estimation error distortions. Achieving this closed-form bias adjustment for the mean and the risk is not an easy task because the exact functional forms of the mean and standard deviation of next-period performance of a naïvely-formed portfolio are very complex once estimation error has been used by the non-linear multivariate optimization process. We use the method of statistical differentials to find Taylor-series approximations to expectations of random variables, obtaining results that are asymptotically correct when the number of time periods is large and that remain statistically consistent when estimated values are substituted for unknown parameters. In effect, we use perturbation analysis to discover how estimation errors are misused by the mean-variance optimization technology in its attempt to improve performance while wrongly believing that the estimated parameters are correct.

A number of approaches have been proposed to study and resolve the problem of bias resulting from estimation error. Our theoretical adjustment is consistent with the empirical findings in the literature regarding the bias induced by estimation error on mean-variance efficient portfolios, including empirical studies of the estimation risk problem by Frost and Savarino (1986b) who observe through simulation that the magnitude of the problem depends upon the ratio of the number of assets to the number of observed time periods. Frankfurter, Phillips and Seagle (1971) perform an experiment in which the impact of estimation error is so strong that the usefulness of mean-variance approaches is questioned. Muller (1993) shows that optimized portfolios tend to be more risky ex-post than predicted ex-ante. Chopra and Ziemba (1993) and Merton (1980) show that the influence of the estimation error in the mean is more critical than the error in the variance. Dhingra (1980) shows that the uncertainty in optimal portfolio selection increases with the target return. Clarkson, Guedes and Thompson (1996) show that estimation risk has a meaningful and measurable non-diversifiable component. Inference methods for the estimated weights of

mean-variance portfolios may be found in Britten-Jones (1999). Solutions to the estimation risk problem include Jorion (1986, 1991) who shows how a James-Stein shrinkage estimator outperforms the sample mean, and Fomby and Samanta (1991) who study a non-Bayesian Stein-Rule approach. For an agent with quadratic utility, ter Horst, de Roon and Werker (2000) show how to adjust the level of risk aversion to compensate for estimation risk of the asset expected returns. The effects of estimation risk on market prices and returns are studied by Lewellen and Shanken (2002). Michaud (1998) reviews the problems associated with mean-variance efficient portfolios and presents a number of estimation techniques.

An estimated portfolio might be compared to the ideal portfolio that could be formed in the absence of statistical noise. Jobson and Korkie (1980) study statistical properties of the estimated Sharpe-ratio maximizing portfolio, as compared to the ideal portfolio that maximizes the Sharpe ratio for the true asset means and covariance matrix, which are unobservable to investors with noisy data. In contrast, we examine the step-ahead performance of estimated portfolios and, although this step-ahead performance depends on unknown parameters, we show how to obtain consistent estimators (with statistically significant bias reduction) using only the sample data.

A Bayesian model with a non-informative prior fails to account for bias in the mean. For example, a theoretical Bayesian approach to estimation risk (Bawa, Brown and Klein, 1979) incorporates estimation risk directly into the decision problem using predictive distributions. Using Jeffreys' non-informative invariant priors, they find that no adjustment in the mean of the predictive distribution is needed due to estimation risk. They do, however, find an estimation-risk adjustment to the sample standard deviation. Yet as Chopra and Ziemba (1993) and Merton (1980) show, the estimation error in the asset means is more critical than the estimation error in the variance, hence our second-order adjustments for estimation risk appear to be more useful than those from a Bayesian approach using non-informative priors. In related research MacLean, Foster and Ziemba (2004) develop an empirical Bayes estimate for the expected rate of return on securities by estimating the parameters in a geometric Brownian motion model. Barberis (2000) shows that it is important to incorporate estimation risk in the portfolio allocation decisions made by investors, by looking at the portfolio allocation using a Bayesian approach for

computing the predictive distribution of asset returns. He concludes that portfolio calculations can be seriously misleading if estimation error is ignored. Klein and Bawa (1976) show that, using a Bayesian procedure incorporating estimation risk directly into the decision process, the optimal portfolio choice differs considerably from the traditional analysis. Frost and Savarino (1986a) find that estimation risk is reduced using an informative prior having the property that all securities have the same expected returns, variances and pairwise correlations.

Pástor and Stambaugh (2000) compare asset pricing models and find that both margin requirements and parameter uncertainty reduce the economic loss due to holding a sub-optimal portfolio given the investor's original beliefs about the true model underlying asset returns. Tu and Zhou (2004) consider data generating process uncertainty in addition to parameter and model uncertainty, and find that the economic loss caused by data generating process uncertainty is small. Our results are consistent with this, in that while our derivations assume normality, bootstrap robustness tests show that our bias adjustments are effective with non-normal empirical data.

We use international country equity indexes to illustrate the adjustments, construct efficient frontiers, and test for significant reductions in the out-of sample bias in the mean. Other empirical studies also consider the estimation problem using international data, including Jorion (1985) who studies the international equity allocation problem under estimation risk and finds that there is a substantial increase in the out-of-sample performance of the optimal portfolio using a Stein shrinkage factor. Eun and Resnick (1994) show that there are greater gains from international diversification for US investors who control for estimation risk ex-ante.

Section 2 derives the second-order bias adjustment for the efficient frontier, while Section 3 shows how to adjust the Sharpe-ratio-maximizing tangency portfolio on the capital market line and explores the tendency of bias adjustment to enhance portfolio diversification. Empirical results in Section 4 examine the effect of the bias adjustment on internationally-diversified equity portfolios in three ways: first, showing the extent of the adjustment over the entire sample, second, testing the effectiveness of the adjustment using step-ahead performance of portfolios formed from a window of past data, and finally,

testing the finite-sample correctness of the asymptotic theory using bootstrap simulations matched to the international data. A summary and discussion follow in Section 5, with proofs in the Appendix.

2. The Efficient Frontier Adjusted for Bias

Consider $n \geq 2$ assets with rates of return observed over $T + 1$ time intervals, where R_{it} denotes the observed rate of return on asset i at time t . Asset return vectors $R_t = (R_{1t}, \dots, R_{nt})'$ are assumed to be independent and identically normally distributed with unknown true mean vector $\mu = E(R_t)$ estimated unbiasedly at time T as $\hat{\mu} = \sum_{t=1}^T R_t / T$, and with unknown true covariance matrix $V = Cov(R_t)$ estimated unbiasedly as $\hat{V} = \frac{1}{T-1} \sum_{t=1}^T (R_t - \hat{\mu})(R_t - \hat{\mu})'$. We assume that the elements of μ are not all equal and that V is nonsingular. Denote the estimation errors as $\delta = \hat{\mu} - \mu$ and $\varepsilon = \hat{V} - V$. Using these estimates based on observations at times $t = 1, \dots, T$, the classical mean-variance optimal portfolio is estimated, for a given target portfolio mean μ_0 , using weights

$$\begin{aligned} \hat{w} &= \hat{V}^{-1}(\mathbf{1} \quad \hat{\mu}) \left[(\mathbf{1} \quad \hat{\mu})' \hat{V}^{-1} (\mathbf{1} \quad \hat{\mu}) \right]^{-1} \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} \\ &= \hat{V}^{-1}(\mathbf{1} \quad \hat{\mu}) \hat{B} \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} \end{aligned} \tag{1}$$

where we define the $n \times 1$ vector $\mathbf{1} = (1, 1, \dots, 1)'$ and the 2×2 matrix $\hat{B} = \left[(\mathbf{1} \quad \hat{\mu})' \hat{V}^{-1} (\mathbf{1} \quad \hat{\mu}) \right]^{-1}$. Note

that the weights \hat{w} would be mean-variance optimal if $\hat{\mu}$ and \hat{V} were the true parameters, but are instead suboptimal as compared to the best-possible but unobservable weights

$$\begin{aligned}
 w &= V^{-1}(\mathbf{1} \quad \mu) \left[(\mathbf{1} \quad \mu)' V^{-1} (\mathbf{1} \quad \mu) \right]^{-1} \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} \\
 &= V^{-1}(\mathbf{1} \quad \mu) B \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix}.
 \end{aligned} \tag{2}$$

The next-period performance (at time $T+1$) of a portfolio with estimated weights \hat{w} is measured by its observed rate of return $\hat{w}'R_{T+1}$. To characterize this next-period performance, we seek adjustments to two naïve performance estimators. The “naïve portfolio mean” is defined as the target mean $\mu_0 = \hat{w}'\hat{\mu}$, while the “naïve portfolio risk” is the standard deviation $\hat{\sigma}_0 = \sqrt{\hat{w}'\hat{V}\hat{w}}$. Note that μ_0 would be the portfolio mean and $\hat{\sigma}_0$ would be the portfolio standard deviation if R_{T+1} were chosen from a distribution with mean $\hat{\mu}$ and covariance matrix \hat{V} instead of the true parameters μ and V . We let $\sigma_0 = \sqrt{w'Vw}$ denote the standard deviation of the portfolio with weights w based on the unobservable true parameters.

Our main results will show how to correct for the bias of these naïve performance estimators by using the following adjusted measures of the portfolio expectation and risk:

$$\hat{\mu}_{adjusted} = \mu_0 - \frac{n-3}{T} \hat{B}_{22} (\mu_0 - \hat{\mu}_*) \quad \text{and} \tag{3}$$

$$\hat{\sigma}_{adjusted} = \left(1 + \frac{n-1.5}{T} \right) \hat{\sigma}_0, \tag{4}$$

where $\hat{\mu}_* = -\hat{B}_{12} / \hat{B}_{22}$. These adjusted measures more accurately represent the “actual portfolio mean”

$\tilde{\mu} = E(\hat{w}'R_{T+1})$ and “actual portfolio standard deviation” $\tilde{\sigma} = \sqrt{Var(\hat{w}'R_{T+1})}$.

Finding these adjustments is a difficult task because the expectations involved do not seem to have closed-form mathematical solutions. For example, the actual portfolio mean

$$\tilde{\mu} = E(\hat{w}'R_{T+1}) = \mu'E(\hat{w}) = \mu'E\left\{\hat{V}^{-1}(\mathbf{1} \ \hat{\mu})\left[(\mathbf{1} \ \hat{\mu})'\hat{V}^{-1}(\mathbf{1} \ \hat{\mu})\right]^{-1}\right\}\begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} \quad (5)$$

involves the expected value of the product of two correlated random matrices, $\hat{V}^{-1}(\mathbf{1} \ \hat{\mu})$ and $\left[(\mathbf{1} \ \hat{\mu})'\hat{V}^{-1}(\mathbf{1} \ \hat{\mu})\right]^{-1}$.

To asymptotically approximate these difficult expectations of functions of estimated values $\hat{\mu}$ and \hat{V} , we chose the *method of statistical differentials* (also called the *delta method*), which substitutes the expected value of a second-order Taylor series expansion (expanded about the unknown true values μ and V) in place of the function itself (Kotz, Johnson and Read, 1988). If we write the Taylor series approximation to a function f as $f(\hat{\mu}, \hat{V}) \cong f(\mu, V) + S_1[f(\hat{\mu}, \hat{V})] + S_2[f(\hat{\mu}, \hat{V})]$, where $S_1(f)$ denotes the first-order terms (which have expectation zero because the statistical differentials $\delta = \hat{\mu} - \mu$ and $\varepsilon = \hat{V} - V$ have mean zero due unbiasedness of $\hat{\mu}$ and \hat{V}) and $S_2(f)$ denotes the second order terms, then the delta-method asymptotic expectation, denoted $E_{\Delta}[f(\hat{\mu}, \hat{V})]$, will be

$$E[f(\hat{\mu}, \hat{V})] \cong E_{\Delta}[f(\hat{\mu}, \hat{V})] \cong f(\mu, V) + E\{S_2[f(\hat{\mu}, \hat{V})]\}. \quad (6)$$

It is reasonable to consider Taylor series expansions of the nonlinear matrix expressions that result from mean-variance optimization because the estimation error terms $\delta = \hat{\mu} - \mu$ and $\varepsilon = \hat{V} - V$ are asymptotically small, tending in distribution to zero as $T \rightarrow \infty$ for fixed n (in fact, $\sqrt{T}\delta$ and $\sqrt{T}\varepsilon$ converge in distribution to multivariate normal distributions). The main results, presented below in Theorems 1 and 2, show how an investor can adjust for the bias inherent in the use of estimated values $(\hat{\mu}, \hat{V})$ in place of the true asset parameters (μ, V) , where these adjustments reflect the future

uncertainty of the not-yet-observed time $T + 1$ asset returns R_{T+1} together with the estimation uncertainty in the weights \hat{w} .

THEOREM 1. *The target mean μ_0 is systematically biased as a measure of expected future performance because the delta-method expectation of future performance is*

$$E(\hat{w}'R_{T+1}) = E(\hat{w}'\mu) \cong E_{\Delta}(\hat{w}'\mu) = \mu_0 - \frac{n-3}{T}B_{22}(\mu_0 - \mu_*) \quad (7)$$

where $\mu_* = -B_{12}/B_{22}$ is the mean of the globally minimum variance portfolio. If we define the adjusted mean to be the estimated right-hand side of (7) so that

$$\hat{\mu}_{adjusted} = \mu_0 - \frac{n-3}{T}\hat{B}_{22}(\mu_0 - \hat{\mu}_*), \quad (8)$$

where $\hat{\mu}_* = -\hat{B}_{12}/\hat{B}_{22}$ then $\hat{\mu}_{adjusted}$ is a second-order unbiased estimator of expected future portfolio performance, eliminating the $O(1/T)$ term from the bias in the sense that

$$E(\hat{\mu}_{adjusted}) \cong E_{\Delta}(\hat{\mu}_{adjusted}) = E_{\Delta}(\hat{w}'R_{T+1}) + O\left(\frac{1}{T^2}\right) \quad (9)$$

THEOREM 2. *The naïve variance $\hat{\sigma}_0^2$ is systematically biased because its delta-method expectation is*

$$E(\hat{\sigma}_0^2) \cong E_{\Delta}(\hat{\sigma}_0^2) = \sigma_0^2 + \frac{B_{22}}{T}\sigma_0^2 - \frac{n-4}{T}B_{22}^2(\mu_0 - \mu_*)^2 - \frac{n-2}{T}\sigma_0^2 + O\left(\frac{1}{T^2}\right). \quad (10)$$

while the actual portfolio variance, evaluated using the delta-method, is

$$\begin{aligned}
\text{Var}(\hat{w}'R_{T+1}) &= E\left[(\hat{w}'R_{T+1})^2\right] - \left[E(\hat{w}'R_{T+1})\right]^2 = E[\hat{w}'V\hat{w} + \hat{w}'\mu\mu'\hat{w}] - \left[E(\hat{w}'\mu)\right]^2 \\
&\cong E_{\Delta}[\hat{w}'V\hat{w} + \hat{w}'\mu\mu'\hat{w}] - \left[E_{\Delta}(\hat{w}'\mu)\right]^2 \\
&= \sigma_0^2 + \frac{B_{22}}{T}\sigma_0^2 - \frac{n-4}{T}B_{22}^2(\mu_0 - \mu_*)^2 + \frac{n-1}{T}\sigma_0^2 + O\left(\frac{1}{T^2}\right) \\
&= E_{\Delta}(\hat{\sigma}_0^2) + \frac{2n-3}{T}\sigma_0^2 + O\left(\frac{1}{T^2}\right).
\end{aligned} \tag{11}$$

If we define the adjusted standard deviation to be

$$\hat{\sigma}_{adjusted} = \left(1 + \frac{n-1.5}{T}\right)\hat{\sigma}_0, \tag{12}$$

then $\hat{\sigma}_{adjusted}^2$ is a second-order unbiased estimator of actual portfolio variance, eliminating the $O(1/T)$ term from the bias in the sense that

$$E(\hat{\sigma}_{adjusted}^2) \cong E_{\Delta}(\hat{\sigma}_{adjusted}^2) = \text{Var}_{\Delta}(\hat{w}'R_{T+1}) + O\left(\frac{1}{T^2}\right). \tag{13}$$

Proofs are in the Appendix. Note, as expected, that the adjusted expected return $\hat{\mu}_{adjusted}$ from (8) is less than the target mean μ_0 when the target mean exceeds the estimated mean return $\hat{\mu}_*$ of the globally minimum-risk portfolio and $n > 3$, and that the adjusted standard deviation $\hat{\sigma}_{adjusted}$ from (12) always exceeds the naïve standard deviation $\hat{\sigma}_0$ that would prevail if the true parameters (μ, V) were equal to the estimated parameters $(\hat{\mu}, \hat{V})$. Note that only n and T are used to find the adjusted risk $\hat{\sigma}_{adjusted}$ from $\hat{\sigma}_0$, but that in finding the adjusted mean $\hat{\mu}_{adjusted}$ from μ_0 we also use the estimated values $\hat{\mu}$ and \hat{V} to obtain \hat{B}_{22} and $\hat{\mu}_* = -\hat{B}_{12}/\hat{B}_{22}$.

For both the mean and the standard deviation, the bias adjustment is greater when there are more assets n and when there are fewer time periods T . Adjustments are inversely proportional to the number

T of observations, which is reasonable because, with more data, we expect the estimates to be more reliable. Adjustments are directly proportional to a linear function of the number n of assets, which is reasonable because, with more assets, there is more flexibility available to mislead while optimizing over the wrong distribution. Thus our theoretical results confirm, in a more general setting, the empirical observations of Frost and Savarino (1986b).

The adjusted efficient frontier, obtained by adjusting both the mean and the standard deviation, is derived below in Theorem 3 and shown in Figure 1.

-- Insert Figure 1 here --

THEOREM 3. *The adjusted efficient frontier, expressing $\hat{\sigma}_{adjusted}$ as a function of $\hat{\mu}_{adjusted}$, is*

$$\hat{\sigma}_{adjusted} = \left(1 + \frac{n-1.5}{T}\right) \sqrt{\hat{\sigma}_*^2 + \frac{\hat{B}_{22}}{[1 - (n-3)\hat{B}_{22}/T]^2} (\hat{\mu}_{adjusted} - \hat{\mu}_*)^2} \quad (14)$$

where $\hat{\sigma}_* = \sqrt{\hat{B}_{11} - \hat{B}_{12}^2 / \hat{B}_{22}}$ is the estimated standard deviation of the global minimum-risk portfolio.

3. Capital Market Lines and Optimal Portfolio Choice

For a given risk-free rate the Capital Market Line (CML) is tangent to the efficient frontier and shows the location of all mean-variance quadratic-utility-maximizing portfolios for a given set of assets. How is the Capital Market Line affected by estimation risk? The anticipated performance of the naïve frontier adjusted for estimation error is represented by the adjusted frontier, as derived in the previous section. Therefore, the tangency portfolio on the naïve CML (tangent to the naïve frontier) will not perform as well as the naïve CML indicates, and will perform worse than the tangency portfolio on the adjusted CML (tangent to the adjusted frontier). In this section we show how to adjust the maximized Sharpe ratio of the tangency portfolio downwards to reflect estimation uncertainty. Then we derive the equations of the adjusted CML and the weights of the optimal quadratic-utility maximizing portfolios.

Moreover, we show that under suitable conditions, the tangency portfolio on the bias-adjusted CML is more diversified than the tangency portfolio on the naïve CML.

Given a riskless rate r_f , the naïve maximized Sharpe ratio $S_{naïve}$ is the maximum reward-to-risk value $(\mu_0 - r_f) / \hat{\sigma}_0$ that would be available, by varying μ_0 , if the estimated asset parameters were the true parameters.

THEOREM 4. *Assuming that $r_f < \hat{\mu}_*$, the adjusted Sharpe ratio $S_{adjusted}$, defined for the adjusted frontier and riskless rate r_f , may be expressed in terms of the naïve Sharpe ratio S as*

$$\begin{aligned} S_{adjusted} &= \frac{1}{1 + (n - 1.5)/T} \sqrt{\frac{[1 - (n - 3)\hat{B}_{22}/T]^2}{\hat{B}_{22}} + \frac{(\hat{\mu}_* - r_f)^2}{\hat{\sigma}_*^2}} \\ &= \frac{1}{1 + (n - 1.5)/T} \sqrt{S^2 - \frac{2(n - 3)}{T} + \frac{(n - 3)^2}{T^2} \hat{B}_{22}} \end{aligned} \quad (15)$$

The adjusted-Sharpe ratio is the slope of the adjusted CML. We give equations for the adjusted CML below, while the corollary that follows gives their equity weights.

THEOREM 5. *Assuming that $r_f < \hat{\mu}_*$ and given an investment weight of x in the risk-free rate, the equations of the adjusted CML are:*

$$\hat{\mu}_{CML(adjusted)}(r_f) = xr_f + (1 - x) \left(\hat{\mu}_* + \frac{\hat{\sigma}_*^2 \left(1 - \frac{n - 3}{T} \hat{B}_{22} \right)^2}{\hat{B}_{22} (\hat{\mu}_* - r_f)} \right) \quad (16)$$

$$\hat{\sigma}_{CML(adjusted)}(r_f) = (1 - x) \left(1 + \frac{n - 1.5}{T} \right) \hat{\sigma}_* \sqrt{1 + \hat{\sigma}_*^2 \left(1 - \frac{n - 3}{T} \hat{B}_{22} \right)^2 \hat{B}_{22}^{-1} (\hat{\mu}_* - r_f)^{-2}}. \quad (17)$$

Figure 2 illustrates the naïve and the adjusted Capital Market Lines, as well as optimal quadratic-utility-maximizing portfolios. The two CMLs can give very different optimal portfolio choices for an

investor. The adjusted frontier is a better approximation of the actual portfolio performance: asymptotically the portfolio on the adjusted CML that maximizes the investor's utility function is preferred to any other. In fact, the naïve-Sharpe-ratio maximizing portfolio will always have an adjusted Sharpe ratio that is less than that of the adjusted-Sharpe-ratio maximizing portfolio. To see this, take the unadjusted mean of the naïve-Sharpe-ratio maximizing portfolio of risky assets, and find its adjusted performance using Equation (3). The resulting adjusted performance must fall on the adjusted frontier, and hence cannot dominate the adjusted-Sharpe-ratio tangency portfolio. It is left to the reader to verify that these two portfolios are distinct when $n > 3$. Therefore every risky portfolio on the naïve CML will be quadratic-utility sub-optimal to some portfolio on the adjusted CML. The optimal quadratic-utility-maximizing portfolio will have some portion of the total weight invested in the adjusted-Sharpe-ratio maximizing portfolio, and the rest in the risk-free asset, depending on the risk aversion of the agent. The next corollary derives the weights for the adjusted-Sharpe-ratio maximizing portfolio.

-- Insert Figure 2 here --

COROLLARY. *Assuming that $r_f < \hat{\mu}_*$, the portfolio weights of the adjusted-Sharpe-ratio maximizing tangency portfolio are:*

$$\hat{w}_{S-\max-adj} = \hat{V}^{-1} \begin{pmatrix} \mathbf{1} & \hat{\mu} \end{pmatrix} \hat{B} \begin{pmatrix} 1 \\ \hat{\mu}_{0,S-\max-adj} \end{pmatrix} \quad (18)$$

where

$$\hat{\mu}_{0,S-\max-adj} = \hat{\mu}_* + \frac{\hat{\sigma}_*^2}{\hat{B}_{22}(\hat{\mu}_* - r_f)} \left(1 - \frac{n-3}{T} \hat{B}_{22} \right) \quad (19)$$

is the naïve target mean whose corresponding adjusted performance is that of the adjusted Sharpe-ratio maximizing portfolio.

To assess diversification, we use a quadratic form on the portfolio weights, an approach similar to that used by Green and Hollifield (1992). We define

$$D\hat{iv} = \left(\hat{w} - \frac{1}{n} \mathbf{1} \right)' \hat{V} \left(\hat{w} - \frac{1}{n} \mathbf{1} \right) \quad (20)$$

and note that $D\hat{iv}$ will be close to 0 for a well-diversified portfolio (the most diversified portfolio being the equally-weighted portfolio) and that $D\hat{iv}$ is everywhere nonnegative and becomes larger as the portfolio becomes less diversified.

THEOREM 6. *The adjusted-Sharpe-ratio maximizing portfolio is more diversified than the naïve-Sharpe-ratio maximizing portfolio, provided $n > 3$, $\hat{\mu}_* > r_f$ and $\hat{\mu} < \hat{\mu}_* + \frac{\hat{\sigma}_*^2}{\hat{B}_{22}(\hat{\mu}_* - r_f)} \left(1 - \frac{n-3}{2T} \hat{B}_{22} \right)$, where*

$\hat{\mu} = \sum_{i=1}^n \hat{\mu}_i / n$ *is the equally weighted average of the estimated asset means .*

4. Empirical Results with International Portfolios

Empirical results are presented in this section for the theoretical bias adjustments of the efficient frontier, the CML, and the optimal utility-maximizing portfolio. Because the adjustments are asymptotic, questions arise: How well do the bias adjustments work in finite samples? How significant, statistically and economically, are the adjustments? Using international portfolio equity returns, the investment opportunity set consists of index portfolios for a set of n countries. Portfolios are formed by estimating the efficient portfolio weights based on past performance. Monthly returns then are computed from the MSCI total return country monthly index (with dividends reinvested) for the 408-month period 1/30/1970 to 12/31/2003. The risk-free rate is the CRSP nominal one-month risk-free rate based on average prices matching the index return period.

The bias adjustment is studied empirically in three ways. First, we look at the size of the adjustment while using the full data set to see how the size of the adjustment varies with the number of countries. Second, we use moving data windows of 36, 48, 60 and 120 months (3, 4, 5 and 10 years) to compare actual step-ahead performance to those predicted by the naïve and adjusted frontiers and to

compute the certainty-equivalent loss for a quadratic utility maximizing agent who ignores estimation risk. We also examine the change in diversification of Sharpe-ratio maximizing tangency portfolios implied by Theorem 6. Finally, we use bootstrap simulations to study the adjustment in a context where the actual mean-variance frontier (matched to the full sample) is known while relaxing the normality assumption used in the derivations.

4.1. Adjusted International Efficient Frontiers

Efficient frontiers for the full data set may be computed with and without adjustment to see how the size of the bias adjustment varies with the number of countries. Figure 3 shows the naïve (continuous line) and adjusted frontiers (dashed line, adjusted for estimation risk following Theorems 1 and 2) for portfolios using international equity returns from three to six countries chosen in the following order: US, UK, Japan, Hong Kong, Germany and Australia. With three assets the two frontiers are almost identical, illustrating that the variance adjustment is small and that the unadjusted mean is second-order unbiased (not needing adjustment) when $n = 3$. As the number of countries increases, there is an increase in the size of the adjustment, pushing the adjusted frontier towards the estimated minimum-variance portfolio.

-- Insert Figure 3 here --

4.2. Statistical and Economic Significance; Effect on Diversification

We present three tests of the theoretical bias adjustments. First, we assess whether the adjusted mean is closer to the actual mean than is the naïve target mean. Second, we test whether the theoretical increase in diversification implied by Theorem 6 is statistically significant. Finally, we assess economic impact by estimating the certainty-equivalent loss incurred by an agent exposed to estimation risk who neglects the opportunity to use the theoretical bias adjustments.

We test the statistical significance of the estimation-risk bias adjustment by forming portfolios, each using estimates from a fixed-length window of data, then calculating the time series of one-step-ahead performance $\hat{w}_t' R_{t+1}$, and comparing this time series of actual performance to the performance

anticipated by the naïve and the adjusted frontiers. At each step in time, using a window of the most recent 36, 48, 60 or 120 months of data, we calculate the mean-variance frontier (with and without adjustment) and estimate the portfolio weights \hat{w}_t corresponding to a fixed target mean of 2%. We then observe the rate of return $\hat{w}_t' R_{t+1}$ of this estimated portfolio for the following month. In this way we obtain three time series data sets (actual performance, naïve performance, and bias-adjusted performance) from which we compute median differences and p values from the Wilcoxon signed rank test. Cross-sectional testing was used after verifying that the series showed insignificant autocorrelation; nonparametric testing methodology was chosen to limit the influence of outliers.

Table 1 reports estimates and tests of the bias for varying numbers of countries and window sizes, measuring the bias as the median difference (in percentage points) between monthly ex-ante anticipated target mean (naïve $\mu_0 = 2\%$ or adjusted), and actual performance. In many cases the adjustment reduces the bias and decreases or eliminates its significance. For example, forming portfolios with six countries and 48 months of past data, the naïve portfolio mean shows a highly significant bias of 1.22% per month, while the adjusted portfolio mean has a non-significant bias of 0.43%.

-- Insert Table 1 here --

The naïve mean is significantly biased in 11 out of 16 cases. In theory, with three countries, the naïve mean is asymptotically unbiased for larger windows and indeed we find no evidence of bias for the larger windows of 48 months or greater ($p = 0.167, 0.051$ and 0.069), whereas with the shortest 36-month window we do find significant bias ($p = 0.00013$). With four, five, and six countries, where the naïve target mean is asymptotically biased according to theory, we find significant bias in all but two of the 12 cases.

The adjusted portfolio mean is significantly biased in only two out of 16 cases. Comparing the bias of naïve and adjusted means, we see that the significance of the bias is eliminated by the adjustment in nine of 11 cases (i.e., there were 11 cases in which the naïve mean is significantly biased and, in nine of these, the adjusted mean is not significantly biased). The two cases in which the adjustment does not

eliminate the significance of the bias (36-month data window with three and with four countries) involve the shortest window with a small number of assets, suggesting that a sample of 36 months may be too small for the asymptotic theory to provide a complete adjustment when n is small. A striking result is that the bias adjustment, although asymptotic, works well for small samples, especially when the estimation risk is large, such as for the 36-month window with five and six countries, or the 48-month window with four or more countries.

Diversification was generally enhanced when adjusted-Sharpe-ratio maximizing tangency portfolios were used instead of naïve-Sharpe-ratio maximizing tangency portfolios, confirming an implication of Theorem 6. We tested this increase in diversification by calculating a time-series of Sharpe-ratio maximizing portfolios using a 60-month rolling window and the CRSP nominal one-month risk free rate based on average prices. We computed both the naïve- and the adjusted-Sharpe-ratio maximizing portfolios, and used the diversification measure defined in Equation (20) to evaluate the degree of diversification of these portfolios. Using four countries we found that the adjusted-Sharpe-ratio maximizing portfolio was more diversified than the naïve-Sharpe-ratio maximizing portfolio 98.56% of the time (344 out of 349 times). With five countries the adjusted-Sharpe-ratio maximizing portfolio was more diversified 98.85% (345 out of 349 times), while with six countries the adjusted-Sharpe-ratio maximizing portfolio was more diversified 100% of the time. (Note that with three countries there is no adjustment in the mean and that the naïve-Sharpe-ratio maximizing portfolio and the adjusted-Sharpe-ratio maximizing portfolio have identical weights despite the adjustment for portfolio variance).

Next we consider the economic impact of the statistically significant results we have demonstrated for the bias adjustment. In particular, we ask: Is the optimal utility-maximizing portfolio on the naïve CML sub-optimal to the utility-maximizing portfolio on the adjusted CML? To test the economic value of the bias adjustments we use methodology similar to that used by Pástor and Stambaugh (2000) and by Tu and Zhou (2004). For a given risk-aversion coefficient for an agent with a derived quadratic utility and a given data window, we compute the weights of two utility-maximizing portfolios: one on the naïve CML, the other on the adjusted CML. We use fixed-size rolling windows to

obtain a bivariate time-series of actual returns from these two portfolios. Next we compute the certainty-equivalent loss incurred by a quadratic-utility agent who ignores estimation risk by holding the time-series of naïvely constructed portfolio returns instead of using the bias adjustment. The quadratic utility with risk aversion A achieved by a random return R is taken to be

$$U = E(R) - \frac{1}{2}A[\text{Var}(R)]. \quad (21)$$

Let U_{naive} be the expected utility achieved by investing in the utility maximizing portfolio on the naïve CML, and define $U_{adjusted}$ similarly but using the adjusted CML. Utility is computed for a series of actual step-ahead returns by first computing its mean and variance. Each actual performance series is determined by three inputs: the risk aversion of the investor, the assets available in computing the portfolio weights, and the adjustment (or not) for estimation error. The first two inputs are identical for both the naïve and the adjusted portfolio series. Hence any observed utility gain or loss from choosing one series of portfolio returns over another is attributable to the estimation risk bias adjustment. We compute the certainty-equivalent loss as the difference in expected utilities:

$$U_{adjusted} - U_{naive} \quad (22)$$

Table 2 reports, for varying degrees of risk-aversion, number of countries n and window sizes T , the percentage certainty-equivalent loss due to ignoring estimation error. Following Pástor and Stambaugh (2000), we report results for risk-aversion coefficient $A = 2.83$, as well as for agents with twice, and agents with half this risk aversion coefficient. For three assets, the certainty-equivalent loss is entirely due to the variance adjustment, since the mean adjustment is zero in this case. The estimated economic loss is increasing with n (the more assets used in the portfolio the larger the loss) and is decreasing in T (the larger the data window size the smaller the loss). The highest risk aversion ($A = 5.66$)

leads to the smallest economic loss, while lower risk aversion is associated with a greater economic benefit from the estimation risk adjustment.

-- insert Table 2 here --

4.3. Assessing Finite-Sample Properties through Bootstrap Simulation

We present a bootstrap simulation study to assess the finite-sample effectiveness of the asymptotic bias adjustment in a situation where we know the true generating distribution for the asset returns. By using bootstrap methodology, our results are robust to non-normality and avoid the choice of a particular parametric distribution. We began with the full data set of 408 months for four countries (US, UK, Japan, and Hong Kong) which define the population of months from which the bootstrap method samples (with replacement) vectors of country returns, drawing all four country returns simultaneously from the sampled month to preserve the empirical covariance structure.

We consider four time horizons: $T = 60, 120, 408$ and $1,000$ months, choosing in each case $10,000$ $T \times 4$ bootstrap samples (each sample is a time series of returns on four country indexes over T time intervals). For each bootstrap sample and each target mean μ_0 , we compute the naïve portfolio weights and find three means and three standard deviations (naïve, adjusted, and actual). The actual means and standard deviations, for a given portfolio, are computed for a very large sample from the bootstrap distribution by using the empirical mean vector and covariance matrix of the full data set. We define the three bootstrap efficient frontiers (naïve, adjusted, and actual) by averaging the $10,000$ means and averaging the $10,000$ standard deviations for each μ_0 .

Our proposed adjustment appears to effectively eliminate nearly all of the bias as shown in Figure 4. For all sample sizes the naïve frontier is biased (showing a higher mean on the upper branch than is actually available) and the adjusted frontier is considerably closer to the actual frontier. Larger sample size leads to better statistical information, and this has two consequences. First, the naïve frontiers become narrower (less biased, having smaller asymptotic slope) because, as sample size grows, the

optimization step is less distorted by the noise in the data. Second, the adjusted and actual frontiers become wider because, as sample size grows, the estimated weights move closer to the true optimal weights.

-- Insert Figure 4 here --

5. Summary and Discussion

We have derived adjustments to reduce and asymptotically eliminate the bias in performance of mean-variance efficient portfolios formed in the presence of estimation error. To do this, we used the method of statistical differentials, which is based on expected second-order Taylor series expansions of the nonlinear optimization process. The adjusted efficient frontier more realistically represents the actual performance of a portfolio that is mean-variance optimal for the estimated asset distribution. In agreement with results of previous empirical investigations, the size of our theoretical adjustment (for both the mean and the standard deviation) increases linearly with the number of assets (because, with more assets, there is more flexibility for statistical error to distort the optimization process) and decreases inversely with the size of the data set as measured by the number of time periods (because statistical error tends to decrease in larger samples).

Surprisingly, in the special case of three assets, no asymptotic adjustment is necessary for the portfolio mean. With four or more assets, the adjustment of the mean is in the direction of the mean of the globally minimum-variance portfolio, showing that the naïve mean is biased upwards along the upper efficient frontier. The adjusted standard deviation, for any number of assets, is always larger than the naïve standard deviation, showing a downward bias for the naïve risk measure. With four or more assets, the naïve-Sharpe-ratio maximizing tangency portfolio is not the portfolio on the naïvely-estimated efficient frontier with the highest actual Sharpe-ratio. Accordingly, we adjust the Sharpe-ratio maximizing tangency portfolio, which tends to move closer to the globally minimum variance portfolio and to become more diversified.

When estimating efficient portfolios of international equities, the size of the bias is statistically significant in many cases. With four or more countries, the adjustment reduces the size of the bias and, in most of these cases, eliminates the statistical significance of the bias. Although the theory is asymptotic, it seems to work well even for small and medium sample sizes. Diversification was generally increased when the bias-adjusted Sharpe ratio was maximized in place of the naïve Sharpe ratio. Tests of the economic impact of the adjustment find that an agent who ignores estimation risk will consistently experience a certainty equivalence loss of utility across a variety of risk aversion levels and sample periods. Bootstrap simulations establish robustness with respect to the normality assumption and, combined with the results of tests of the empirical step-ahead performance of international portfolios, confirm the effectiveness of the proposed bias adjustments.

References

- Anderson, T.W. 1984. *An Introduction to Multivariate Statistical Analysis*, (2d ed.). Wiley, New York.
- Barberis, N. 2000. Investing for the long run when returns are predictable. *J. Finance* **55** 225-264.
- Bawa, V. S., S. J. Brown, R. W. Klein. 1979. *Estimation Risk and Optimal Portfolio Choice*. North-Holland, Amsterdam.
- Britten-Jones, M. 1999. The sampling error in estimates of mean-variance efficient portfolio weights. *J. Finance* **54** 655-671.
- Chopra, V. K., W. T. Ziemba. 1993. The effects of errors in means, variances and covariances on optimal portfolio choice. *J. Portfolio Management*, **19** (Winter) 6-11.
- Clarkson, P., J. Guedes, R. Thompson. 1996. On the diversification, observability and measurement of estimation risk. *J. Financial Quant. Anal.* **31** 69-84.
- Dhingra, H. L. 1980. Effects of estimation risk on efficient portfolios: a Monte Carlo simulation study. *J. Bus. Finance Accounting* **7** 277-295.

- Eun, C. S., B. G. Resnick. 1994. International diversification of investment portfolios: US and Japanese perspectives. *Management Sci.* **40** 140-161.
- Fomby, T. B., S. K. Samanta. 1991. Application of Stein rules to combination forecasting. *J. Business Econom. Statist.* **9** 391-407.
- Frankfurter, G. M., H. E. Phillips, J. P. Seagle. 1971. Portfolio selection: the effects of uncertain means, variances and covariances. *J. Financial Quant. Anal.* **6** 1251-1262.
- Frost, P. A., J. E. Savarino. 1986a. An empirical Bayes approach to efficient portfolio selection. *J. Financial Quant. Anal.* **21** 293-305.
- Frost, P. A., J. E. Savarino. 1986b. Portfolio size and estimation risk; the problem that results from estimation error in portfolio selection is even more severe than has been previously suggested. *J. Portfolio Management* **12**(Summer) 60-64.
- Green, R. C., B. Hollifield. 1992. When will mean-variance portfolios be well diversified? *J. Finance* **47** 1785-1809.
- Jobson, J. D., B. Korkie. 1980. Estimation for Markowitz efficient portfolios. *J. Amer. Statist. Assoc.* **75** 544-554.
- Jorion, P. 1985. International portfolio diversification with estimation risk. *J. Bus.* **58** 259-278.
- Jorion, P. 1986. Bayes-Stein estimation for portfolio analysis. *J. Financial Quant. Anal.* **21** 279-292.
- Jorion, P. 1991. Bayesian and CAPM estimators of the means: implications for portfolio selection. *J. Banking Finance* **15** 717-727.
- Klein, R. W., V. S. Bawa. 1976. The effect of estimation risk on optimal portfolio choice. *J. Financial Econom.* **3** 215-231.
- Kotz, S., N. L. Johnson, C. B. Read. 1988. *Encyclopedia of Statistical Sciences Vol. 8*. Wiley, New York, 646-647.
- Lewellen, J., J. Shanken. 2002. Learning, asset pricing tests and market efficiency. *J. Finance* **57** 1113-1145.

- MacLean, L., M. E. Foster, W. T. Ziemba. 2004. Empirical Bayes estimation with dynamic portfolio models. Unpublished working paper. University of British Columbia.
- Markowitz, H. 1952. Portfolio selection. *J. Finance* **7** 77-91.
- Merton, R. C. 1980. On estimating the expected return on the market. *J. Financial Econom.* **8** 323-361.
- Michaud, R. O. 1998. *Efficient Asset Management, a Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Harvard Business School Press, Boston, Mass..
- Muller, P. 1993. Empirical tests of biases in equity portfolio optimization. In: Zenios, S. A. (Ed.), *Financial Optimization*. Cambridge University Press, Cambridge, UK.
- Pástor, L., R. F. Stambaugh. 2000. Comparing asset pricing models: an investment perspective. *J. Financial Econom.* **56** 335-381.
- ter Horst, J. R., F. A. de Roon, B. J. M. Werker. 2000. Incorporating estimation risk in portfolio choice, Unpublished working paper. Tilburg University.
- Tu, J., G. Zhou. 2004. Data-generating process uncertainty: What difference does it make in portfolio decisions? *J. Financial Econom.* **72** 385-421.

Table 1 Monthly bias, showing bias reduction due to adjustment with more than three countries (with p values in parentheses from the Wilcoxon signed rank test).

| | 3 countries | 4 countries | 5 countries | 6 countries |
|-------------------------|----------------------|----------------------|----------------------|----------------------|
| 36-month window | | | | |
| Naïve | 1.77% (0.00013)** | 1.34% (0.00023)** | 1.07% (0.00014)** | 1.19% (0.00054)** |
| Adjusted | 1.77% (0.00013)** | 0.96% (0.021)* | 0.46% (0.066) | 0.51% (0.195) |
| 48-month window | | | | |
| Naïve | 0.59% (0.167) | 1.23% (0.0039)** | 1.01% (0.0053)** | 1.22% (0.011)** |
| Adjusted | 0.59% (0.167) | 0.72% (0.123) | 0.30% (0.475) | 0.43% (0.837) |
| 60-month window | | | | |
| Naïve | 0.56% (0.051) | 1.13% (0.014)** | 0.71% (0.027)* | 0.87% (0.081) |
| Adjusted | 0.56% (0.051) | 0.73% (0.197) | 0.07% (0.715) | 0.07% (0.587) |
| 120-month window | | | | |
| Naïve | 0.99% (0.069) | 0.37% (0.129) | 0.77% (0.020)* | 0.67% (0.026)* |
| Adjusted | 0.99% (0.069) | 0.11% (0.437) | 0.31% (0.259) | 0.08% (0.452) |

*($p < 0.05$), **($p < 0.01$).

Table 2 Certainty equivalence loss when choosing a naïve instead of an adjusted CML portfolio.
Positive table entries indicate higher utility when using the adjusted CML.

| Window size | 3 countries | 4 countries | 5 countries | 6 countries |
|---------------------------|-------------|-------------|-------------|-------------|
| Risk Aversion $A = 1.415$ | | | | |
| 36 months | 0.0838% | 0.8425% | 1.4438% | 2.2007% |
| 48 months | 0.0428% | 0.3905% | 0.7658% | 1.1286% |
| 60 months | 0.0280% | 0.3052% | 0.5914% | 0.8594% |
| 120 months | 0.0094% | 0.1611% | 0.4112% | 0.6636% |
| Risk Aversion $A = 2.83$ | | | | |
| 36 months | 0.0419% | 0.4209% | 0.7217% | 1.0999% |
| 48 months | 0.0214% | 0.1951% | 0.3830% | 0.5640% |
| 60 months | 0.0140% | 0.1524% | 0.2958% | 0.4296% |
| 120 months | 0.0047% | 0.0805% | 0.2056% | 0.3315% |
| Risk Aversion $A = 5.66$ | | | | |
| 36 months | 0.0210% | 0.2101% | 0.3607% | 0.5494% |
| 48 months | 0.0107% | 0.0974% | 0.1916% | 0.2817% |
| 60 months | 0.0070% | 0.0760% | 0.1480% | 0.2147% |
| 120 months | 0.0024% | 0.0402% | 0.1028% | 0.1654% |

Figure 1 The bias-adjusted frontier is obtained from the naïve frontier by adjusting both the mean and the standard deviation for estimation error.

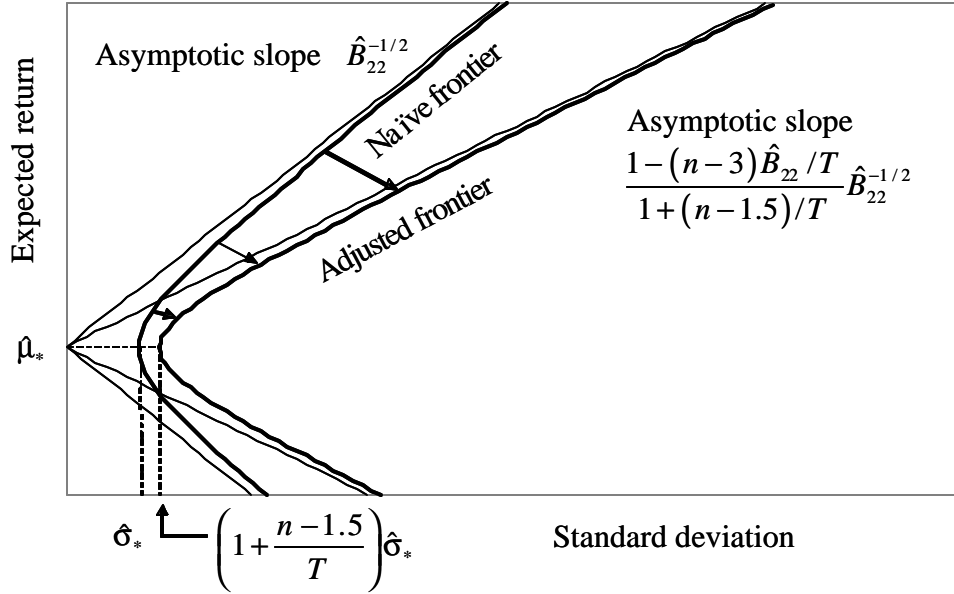


Figure 2 Naïve and adjusted capital market lines showing utility-maximizing portfolios tangent to utility indifference curves (dashed lines). Note that the maximized utility along the adjusted CML is higher than the actual utility of investing on the naïve CML.

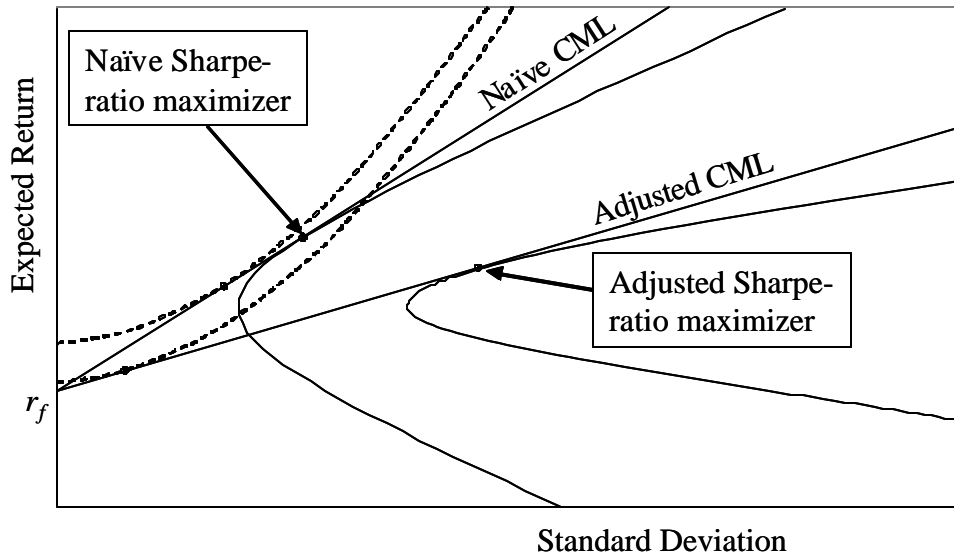


Figure 3 Efficient frontiers, with and without bias adjustment, for portfolios formed from three through six international index funds chosen in the following order: US, UK, Japan, Hong Kong, Germany and Australia, using 408 months of returns data from 1/30/1970 to 12/31/2003. With three countries the adjustment is negligible. As the number of assets increases, so does the size of the adjustment.

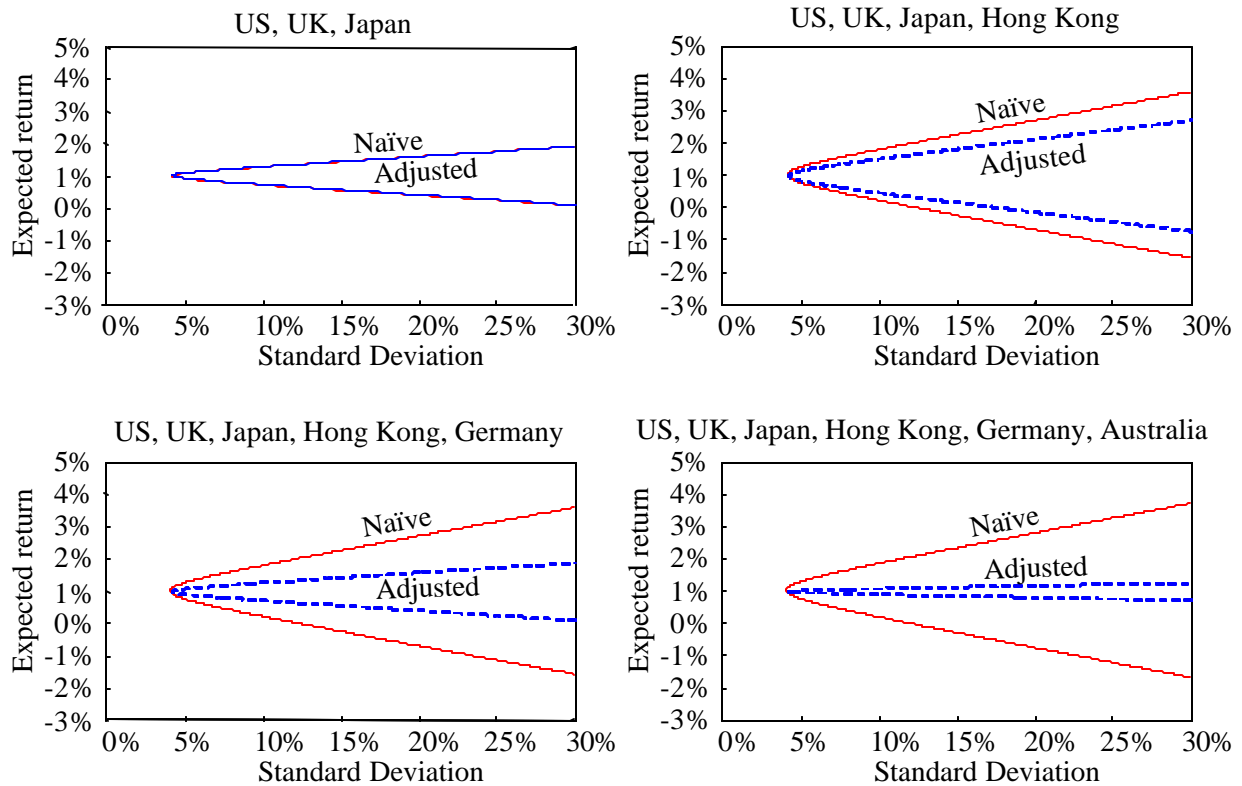
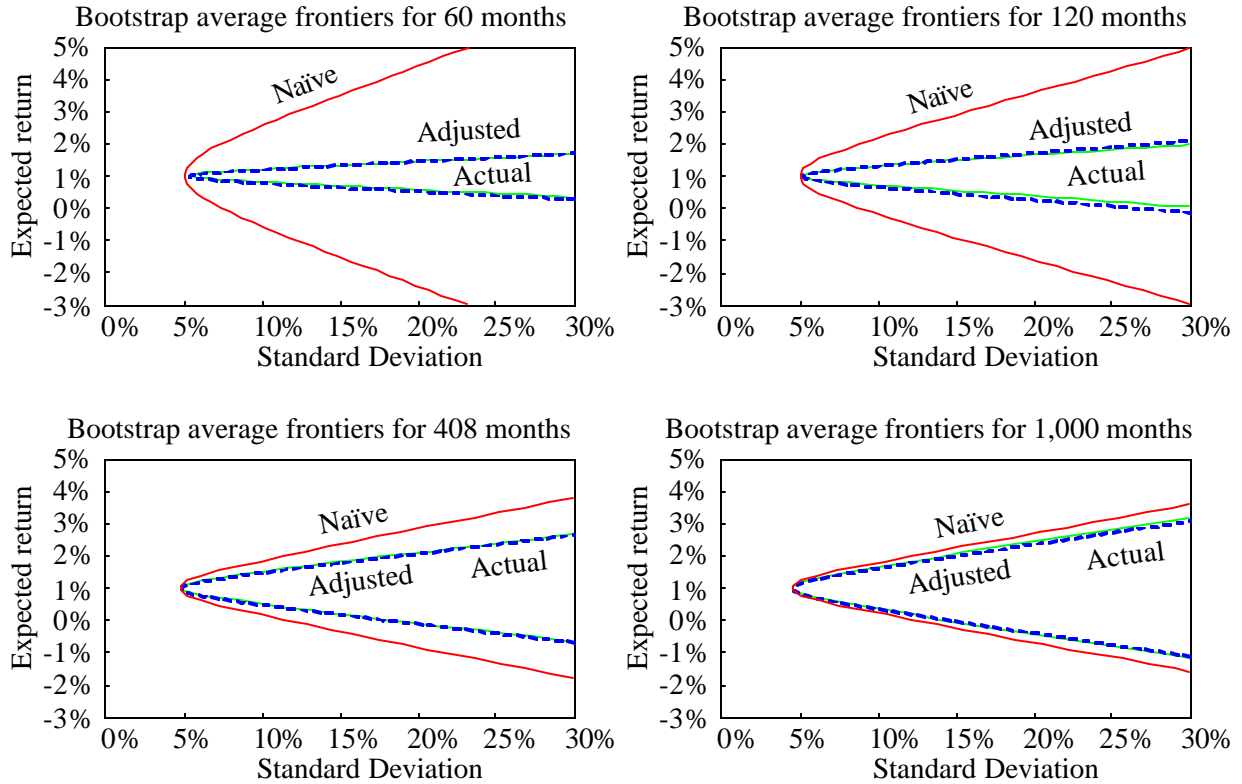


Figure 4 Averages of bootstrap simulations of the naïve, adjusted and actual frontiers to illustrate the effectiveness of the bias adjustment using four countries (US, UK, Japan, and Hong Kong) with monthly returns over 60, 120, 408 and 1,000 months, using 408 months of returns data from 1/30/1970 to 12/31/2003.



Appendix: Proofs

This Appendix briefly reviews the notation and definitions and establishes basic lemmas before proving the results of the paper. Basic second-order expectations are given in Lemma 1, Lemma 2 gives properties of the statistical differentials δ and ε , Lemma 3 gives the expansion of a matrix inverse, Lemma 4 shows first- and second-order expansions of $(\mathbf{1} \ \hat{\mu})$ and \hat{V}^{-1} , Lemma 5 shows expansions of \hat{B} , Lemma 6 shows expansions of \hat{w} , Lemma 7 derives matrix traces, and Lemma 8 derives the expectations that are needed to derive the adjusted performance measures and their properties.

We define the first-order statistical differential terms δ and ε so that $\hat{\mu} = \mu + \delta$ and $\hat{V} = V + \varepsilon$, while B is defined as $B = \left[(\mathbf{1} \ \mu)' V^{-1} (\mathbf{1} \ \mu) \right]^{-1}$ so that $w = V^{-1} (\mathbf{1} \ \mu) B \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix}$, while B is estimated as $\hat{B} = \left[(\mathbf{1} \ \hat{\mu})' \hat{V}^{-1} (\mathbf{1} \ \hat{\mu}) \right]^{-1}$. We also define the scalar $\xi = (0 \ 1) B (\mathbf{1} \ \mu_0)' = B_{22} (\mu_0 - \mu_*)$, where $\mu_* = -B_{12} / B_{22}$. We will let $S_i(f)$ denote the term of order i in the Taylor series expansion of f , while $S_i(f, \delta)$ denotes the term of order i including only δ (i.e., setting $\varepsilon = 0$), and similarly $S_i(f, \varepsilon)$ is obtained by including only ε terms with $\delta = 0$. We will use $E_{\Delta} \left[f(\hat{\mu}, \hat{V}) \right]$ to denote the expectation of the second-order expansion of f at (μ, V) with respect to δ and ε so that

$$\begin{aligned} E_{\Delta} \left[f(\hat{\mu}, \hat{V}) \right] &= f(\mu, V) + E \left\{ S_2 \left[f(\hat{\mu}, \hat{V}) \right] \right\} \\ &= f(\mu, V) + E \left\{ S_2 \left[f(\hat{\mu}, \hat{V}), \delta \right] \right\} + E \left\{ S_2 \left[f(\hat{\mu}, \hat{V}), \varepsilon \right] \right\} \end{aligned} \tag{23}$$

because any second-order term that includes both δ and ε will have expectation zero.

LEMMA 1. For any symmetric $n \times n$ matrix Q , $E(\varepsilon Q \varepsilon) = [VQV + V \operatorname{tr}(QV)] / (T-1)$ and $E(\delta' Q \delta) = \operatorname{tr}(QV) / T$.

PROOF. For the expectation involving ε , using the Cholesky decomposition, let C be any matrix such that $CC' = V$. Let $Z_t \sim N_n(\mathbf{0}, I)$ be independent of each other. Then $CZ_t \sim N_n(\mathbf{0}, V)$ and (e.g., Anderson, 1984, p. 71, Theorem 3.3.2)

$$\hat{V} \stackrel{d}{=} \frac{1}{T-1} \sum_{t=1}^{T-1} CZ_t Z_t' C'. \quad (24)$$

Define $\varepsilon^* = C^{-1} \varepsilon (C')^{-1}$. Then $\varepsilon = C \varepsilon^* C'$ and

$$\varepsilon^* = C^{-1} (\hat{V} - V) (C')^{-1} \stackrel{d}{=} \frac{1}{T-1} \sum_{t=1}^{T-1} Z_t Z_t' - I_n. \quad (25)$$

Using this representation, we compute

$$\begin{aligned} E(\varepsilon Q \varepsilon) &= E(C \varepsilon^* C' Q C \varepsilon^* C') \\ &= E \left[C \left(\frac{1}{T-1} \sum_{t=1}^{T-1} Z_t Z_t' - I_n \right) C' Q C \left(\frac{1}{T-1} \sum_{t=1}^{T-1} Z_t Z_t' - I_n \right) C' \right] \\ &= \frac{1}{(T-1)^2} E \left[C \left(\sum_{t=1}^{T-1} Z_t Z_t' \right) C' Q C \left(\sum_{t=1}^{T-1} Z_t Z_t' \right) C' \right] - V Q V. \end{aligned} \quad (26)$$

Multiplying out, grouping similar terms, using identical distributions for Z_t and the fact that

$$E(Z_t Z_t') = I_n, \text{ we find:}$$

$$\begin{aligned}
 E(\varepsilon Q \varepsilon) &= \frac{1}{(T-1)^2} \left[(T-1) CE \left(Z Z_1' C' Q C Z_1 Z_1' \right) C' + (T-1)(T-2) CE \left(Z_1 Z_1' C' Q C Z_2 Z_2' \right) C' \right] - V Q V \\
 &= \frac{1}{T-1} CE \left(Z Z_1' C' Q C Z_1 Z_1' \right) C' + \frac{T-2}{T-1} C C' Q C C' - V Q V \\
 &= \frac{1}{T-1} CE \left(Z Z_1' C' Q C Z_1 Z_1' \right) C' - \frac{1}{T-1} V Q V.
 \end{aligned} \tag{27}$$

To see that

$$E \left(Z_1 Z_1' C' Q C Z_1 Z_1' \right) = 2 C' Q C + I_n \operatorname{tr} (C' Q C), \tag{28}$$

note (by moving an embedded scalar) that $E \left(Z_1 Z_1' C' Q C Z_1 Z_1' \right) = E \left[\left(Z_1' C' Q C Z_1 \right) Z_1 Z_1' \right]$

$$= E \left\{ \left[\sum_{i,j=1}^n Z_{1i} (C' Q C)_{ij} Z_{1j} \right] Z_1 Z_1' \right\}, \text{ with diagonal } (k, k) \text{ entry}$$

$$\begin{aligned}
 \left[E \left(Z_1 Z_1' C' Q C Z_1 Z_1' \right) \right]_{kk} &= E \left\{ \left[\sum_{i,j=1}^n Z_{1i} (C' Q C)_{ij} Z_{1j} \right] Z_{1k}^2 \right\} \\
 &= E \left[(C' Q C)_{kk} Z_{1k}^4 \right] + E \left\{ \left[\sum_{i \neq k} Z_{1i}^2 (C' Q C)_{ii} \right] Z_{1k}^2 \right\} \\
 &= 3 (C' Q C)_{kk} + \sum_{i \neq k} (C' Q C)_{ii} = 2 (C' Q C)_{kk} + \operatorname{tr} (C' Q C) (I_n)_{kk},
 \end{aligned} \tag{29}$$

where we have used $E(Z^4) = 3$ for a standard normal deviate Z and the fact that $E(Z_{1i} Z_{1j} Z_{1k} Z_{1m})$ must be zero unless each index occurs an even number of times. The off-diagonal entry (k, m) with $k \neq m$ has the same form, establishing (28) because

$$\begin{aligned}
 \left[E \left(Z_1 Z_1' C' Q C Z_1 Z_1' \right) \right]_{km} &= E \left\{ \left[\sum_{i,j=1}^n Z_{1i} (C' Q C)_{ij} Z_{1j} \right] Z_{1k} Z_{1m} \right\} \\
 &= E \left\{ (C' Q C)_{km} Z_{1k}^2 Z_{1m}^2 \right\} + E \left\{ (C' Q C)_{nk} Z_{1k}^2 Z_{1m}^2 \right\} \\
 &= 2 (C' Q C)_{km} = 2 (C' Q C)_{km} + \operatorname{tr} (C' Q C) (I_n)_{km}.
 \end{aligned} \tag{30}$$

Using (28) in (27), with matrix commutativity within the trace operator, completes the proof because

$$\begin{aligned}
 E(\varepsilon Q \varepsilon) &= \frac{1}{T-1} C [2C' Q C + I_n \operatorname{tr}(C' Q C)] C' - \frac{1}{T-1} V Q V \\
 &= \frac{1}{T-1} [2C C' Q C C' + C C' \operatorname{tr}(C' Q C) - V Q V] \\
 &= \frac{1}{T-1} [2V Q V + V \operatorname{tr}(Q C C') - V Q V] \\
 &= \frac{1}{T-1} [V Q V + V \operatorname{tr}(Q V)].
 \end{aligned} \tag{31}$$

To see that $E(\delta' Q \delta) = \operatorname{tr}(Q V) / T$, use matrix commutativity within the trace operator as follows:

$$E(\delta' Q \delta) = E[\operatorname{tr}(\delta' Q \delta)] = E[\operatorname{tr}(Q \delta \delta')] = \operatorname{tr}[E(Q \delta \delta')] = \operatorname{tr}(Q V) / T. \tag{32}$$

Q.E.D. (Lemma 1)

LEMMA 2. The error terms δ and ε are independent and satisfy the following conditions:

$$\sqrt{T} \delta \sim N_n(\mathbf{0}, V) \quad \text{and} \quad E(\delta \delta') = V / T \tag{33}$$

$$\sqrt{T} \varepsilon \xrightarrow{d} N(\mathbf{0}, \Omega). \tag{34}$$

PROOF. Independence of δ and ε is a standard result in multivariate statistics (e.g., Anderson, 1984, p. 71, Theorem 3.3.2), as is (33). To see that $\sqrt{T} \varepsilon$ is asymptotically normal with finite covariance structure Ω , apply the central limit theorem to representation (24), establishing (34).

Q.E.D. (Lemma 2)

LEMMA 3. The first- and second-order expansion terms of the inverse of the matrix $Q + \theta$, where Q and θ are $n \times n$ matrices with both Q and $Q + \theta$ invertible, are given by

$$S_1[(Q + \theta)^{-1}, \theta] = -Q^{-1}\theta Q^{-1} \quad (35)$$

$$S_2[(Q + \theta)^{-1}, \theta] = Q^{-1}\theta Q^{-1}\theta Q^{-1}. \quad (36)$$

PROOF. Begin by observing directly that the identity matrix may be written as

$$I_n = (Q + \theta)(Q^{-1} - Q^{-1}\theta Q^{-1} + Q^{-1}\theta Q^{-1}\theta Q^{-1}) - \theta Q^{-1}\theta Q^{-1}\theta Q^{-1}. \quad (37)$$

Next, premultiply both sides by $(Q + \theta)^{-1}$ to obtain

$$\begin{aligned} (Q + \theta)^{-1} &= Q^{-1} - Q^{-1}\theta Q^{-1} + Q^{-1}\theta Q^{-1}\theta Q^{-1} - (Q + \theta)^{-1}\theta Q^{-1}\theta Q^{-1}\theta Q^{-1} \\ &= Q^{-1} - Q^{-1}\theta Q^{-1} + Q^{-1}\theta Q^{-1}\theta Q^{-1} + o(\|\theta\|^3), \end{aligned} \quad (38)$$

from which we can identify the first- and second-order terms.

Q.E.D. (Lemma 3)

LEMMA 4. First- and second-order expansions of $(\mathbf{1} \quad \hat{\mu})$ and \hat{V}^{-1} are as follows:

$$S_1[(\mathbf{1} \quad \hat{\mu}), \delta] = \delta(0 \quad 1) \quad (39)$$

$$S_2[(\mathbf{1} \quad \hat{\mu}), \delta] = (\mathbf{0} \quad \mathbf{0}) \quad (40)$$

$$S_1(\hat{V}^{-1}, \varepsilon) = -V^{-1}\varepsilon V^{-1} \quad (41)$$

$$S_2(\hat{V}^{-1}, \varepsilon) = V^{-1} \varepsilon V^{-1} \varepsilon V^{-1} \quad (42)$$

$$S_1[(\mathbf{1} \ \hat{\mu}), \varepsilon] = S_2[(\mathbf{1} \ \hat{\mu}), \varepsilon] = (\mathbf{0} \ \mathbf{0}) \quad \text{and} \quad S_1(\hat{V}^{-1}, \delta) = S_2(\hat{V}^{-1}, \delta) = \mathbf{0}. \quad (43)$$

PROOF. Equations (39) and (40) follow directly from the definition $\hat{\mu} = \mu + \delta$, while (41) and (42) follow from Lemma 3 using $\hat{V} = V + \varepsilon$. Equation (43) follows because $\hat{\mu}$ does not depend on ε and \hat{V} does not depend on δ .

Q.E.D. (Lemma 4)

LEMMA 5. First- and second-order expansions of \hat{B} with respect to δ and ε are as follows:

$$S_1(\hat{B}, \delta) = -B \left[\begin{pmatrix} 0 \\ 1 \end{pmatrix} \delta' V^{-1} (\mathbf{1} \ \mu) + (\mathbf{1} \ \mu)' V^{-1} \delta (0 \ 1) \right] B \quad (44)$$

$$\begin{aligned} S_2(\hat{B}, \delta) &= \delta' \left[V^{-1} (\mathbf{1} \ \mu) B (\mathbf{1} \ \mu)' V^{-1} - V^{-1} \right] \delta B \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0 \ 1) B \\ &\quad + B_{22} B (\mathbf{1} \ \mu)' V^{-1} \delta \delta' V^{-1} (\mathbf{1} \ \mu) B \\ &\quad + B \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0 \ 1) B (\mathbf{1} \ \mu)' V^{-1} \delta \delta' V^{-1} (\mathbf{1} \ \mu) B \\ &\quad + B (\mathbf{1} \ \mu)' V^{-1} \delta \delta' V^{-1} (\mathbf{1} \ \mu) B \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0 \ 1) B \end{aligned} \quad (45)$$

$$S_1(\hat{B}, \varepsilon) = B (\mathbf{1} \ \mu)' V^{-1} \varepsilon V^{-1} (\mathbf{1} \ \mu) B \quad (46)$$

$$S_2(\hat{B}, \varepsilon) = B (\mathbf{1} \ \mu)' V^{-1} \varepsilon \left[V^{-1} (\mathbf{1} \ \mu) B (\mathbf{1} \ \mu)' V^{-1} - V^{-1} \right] \varepsilon V^{-1} (\mathbf{1} \ \mu) B. \quad (47)$$

PROOF. Equations (44) and (45) follow from identifying first- and second-order terms in the expansion of \hat{B} with respect to δ , using Lemma 3 while setting $\varepsilon = 0$ (because we are expanding with respect to δ only), as follows:

$$\begin{aligned}
 [(\mathbf{1} \ \hat{\mu})' V^{-1} (\mathbf{1} \ \hat{\mu})]^{-1} &= \left\{ \left[(\mathbf{1} \ \mu)' + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \delta' \right] V^{-1} [(\mathbf{1} \ \mu) + \delta(0 \ 1)] \right\}^{-1} \\
 &= \left\{ B^{-1} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \delta' V^{-1} (\mathbf{1} \ \mu) + (\mathbf{1} \ \mu)' V^{-1} \delta(0 \ 1) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \delta' V^{-1} \delta(0 \ 1) \right\}^{-1} \\
 &\cong B - B \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \delta' V^{-1} (\mathbf{1} \ \mu) + (\mathbf{1} \ \mu)' V^{-1} \delta(0 \ 1) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \delta' V^{-1} \delta(0 \ 1) \right\} B \\
 &+ B \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \delta' V^{-1} (\mathbf{1} \ \mu) + (\mathbf{1} \ \mu)' V^{-1} \delta(0 \ 1) \right\} B \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \delta' V^{-1} (\mathbf{1} \ \mu) + (\mathbf{1} \ \mu)' V^{-1} \delta(0 \ 1) \right\} B.
 \end{aligned} \tag{48}$$

By identifying, moving, and transposing selected embedded scalars, the second-order term may be re-expressed as

$$\begin{aligned}
 S_2(\hat{B}, \delta) &= -B \begin{pmatrix} 0 \\ 1 \end{pmatrix} (\delta' V^{-1} \delta)(0 \ 1) B + B \begin{pmatrix} 0 \\ 1 \end{pmatrix} \left[\delta' V^{-1} (\mathbf{1} \ \mu) B \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right] \delta' V^{-1} (\mathbf{1} \ \mu) B \\
 &+ B \begin{pmatrix} 0 \\ 1 \end{pmatrix} \left[\delta' V^{-1} (\mathbf{1} \ \mu) B (\mathbf{1} \ \mu)' V^{-1} \delta \right] (0 \ 1) B \\
 &+ B (\mathbf{1} \ \mu)' V^{-1} \delta \left[(0 \ 1) B \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right] \delta' V^{-1} (\mathbf{1} \ \mu) B \\
 &+ B (\mathbf{1} \ \mu)' V^{-1} \delta \left[(0 \ 1) B (\mathbf{1} \ \mu)' V^{-1} \delta \right] (0 \ 1) B \\
 &= \delta' \left[V^{-1} (\mathbf{1} \ \mu) B (\mathbf{1} \ \mu)' V^{-1} - V^{-1} \right] \delta B \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0 \ 1) B \\
 &+ B_{22} B (\mathbf{1} \ \mu)' V^{-1} \delta \delta' V^{-1} (\mathbf{1} \ \mu) B \\
 &+ B \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0 \ 1) B (\mathbf{1} \ \mu)' V^{-1} \delta \delta' V^{-1} (\mathbf{1} \ \mu) B \\
 &+ B (\mathbf{1} \ \mu)' V^{-1} \delta \delta' V^{-1} (\mathbf{1} \ \mu) B \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0 \ 1) B.
 \end{aligned} \tag{49}$$

Equations (46) and (47) follow from identifying first- and second-order terms in the expansion of \hat{B} with respect to ε , using Lemma 3 twice while setting $\delta = 0$, as follows:

$$\begin{aligned}
 \left[(\mathbf{1} \ \mu)' \hat{V}^{-1} (\mathbf{1} \ \mu) \right]^{-1} &= \left[(\mathbf{1} \ \mu)' (V + \varepsilon)^{-1} (\mathbf{1} \ \mu) \right]^{-1} \\
 &\cong \left[(\mathbf{1} \ \mu)' (V^{-1} - V^{-1} \varepsilon V^{-1} + V^{-1} \varepsilon V^{-1} \varepsilon V^{-1}) (\mathbf{1} \ \mu) \right]^{-1} \\
 &= \left[B^{-1} - (\mathbf{1} \ \mu)' V^{-1} \varepsilon V^{-1} (\mathbf{1} \ \mu) + (\mathbf{1} \ \mu)' V^{-1} \varepsilon V^{-1} \varepsilon V^{-1} (\mathbf{1} \ \mu) \right]^{-1} \quad (50) \\
 &\cong B - B \left[-(\mathbf{1} \ \mu)' V^{-1} \varepsilon V^{-1} (\mathbf{1} \ \mu) + (\mathbf{1} \ \mu)' V^{-1} \varepsilon V^{-1} \varepsilon V^{-1} (\mathbf{1} \ \mu) \right] B \\
 &\quad + B \left[-(\mathbf{1} \ \mu)' V^{-1} \varepsilon V^{-1} (\mathbf{1} \ \mu) \right] B \left[-(\mathbf{1} \ \mu)' V^{-1} \varepsilon V^{-1} (\mathbf{1} \ \mu) \right] B.
 \end{aligned}$$

Q.E.D. (Lemma 5)

LEMMA 6. First-order expansions of \hat{w} with respect to δ and ε are as follows:

$$S_1(\hat{w}, \delta) = \left[\xi V^{-1} - V^{-1} (\mathbf{1} \ \mu) B \begin{pmatrix} 0 \\ 1 \end{pmatrix} w' - \xi V^{-1} (\mathbf{1} \ \mu) B (\mathbf{1} \ \mu)' V^{-1} \right] \delta \quad (51)$$

$$S_1(\hat{w}, \varepsilon) = \left[V^{-1} (\mathbf{1} \ \mu) B (\mathbf{1} \ \mu)' V^{-1} - V^{-1} \right] \varepsilon w. \quad (52)$$

PROOF. Note that $\hat{w} = \hat{V}^{-1} (\mathbf{1} \ \hat{\mu}) \hat{B} \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix}$. Expanding with respect to δ , we may set $\varepsilon = 0$ using (39)

and (44) to find

$$\begin{aligned}
 S_1(\hat{w}, \delta) &= V^{-1} S_1[(\mathbf{1} \ \hat{\mu}) \hat{B}, \delta] \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} = V^{-1} \left\{ S_1[(\mathbf{1} \ \hat{\mu}), \delta] B + (\mathbf{1} \ \mu) S_1(\hat{B}, \delta) \right\} \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} \\
 &= V^{-1} \left\{ \delta \begin{pmatrix} 0 & 1 \end{pmatrix} B - (\mathbf{1} \ \mu) B \begin{bmatrix} 0 \\ 1 \end{bmatrix} \delta V^{-1} (\mathbf{1} \ \mu) + (\mathbf{1} \ \mu)' V^{-1} \delta \begin{pmatrix} 0 & 1 \end{pmatrix} \right\} B \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} \quad (53) \\
 &= \xi V^{-1} \delta - V^{-1} (\mathbf{1} \ \mu) B \begin{bmatrix} 0 \\ 1 \end{bmatrix} \left[\delta V^{-1} (\mathbf{1} \ \mu) B \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} \right] - \xi V^{-1} (\mathbf{1} \ \mu) B (\mathbf{1} \ \mu)' V^{-1} \delta.
 \end{aligned}$$

Transposing an embedded scalar and factoring establishes (51). To prove (52), expand with respect to ε while setting $\delta = 0$, so that $\hat{\mu} = \mu$, using (41) and (46) to find

$$\begin{aligned}
 S_1(\hat{w}, \varepsilon) &= S_1[\hat{V}^{-1} (\mathbf{1} \ \mu) \hat{B}, \varepsilon] \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} = S_1(\hat{V}^{-1}, \varepsilon) (\mathbf{1} \ \mu) B \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} + V^{-1} (\mathbf{1} \ \mu) S_1(\hat{B}, \varepsilon) \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} \quad (54) \\
 &= -V^{-1} \varepsilon V^{-1} (\mathbf{1} \ \mu) B \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} + V^{-1} (\mathbf{1} \ \mu) B (\mathbf{1} \ \mu)' V^{-1} \varepsilon V^{-1} (\mathbf{1} \ \mu) B \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix},
 \end{aligned}$$

which simplifies to complete the proof.

Q.E.D. (Lemma 6)

LEMMA 7. Some useful traces are given by

$$\text{tr} \left[V^{-1} (\mathbf{1} \ \mu) B (\mathbf{1} \ \mu)' \right] = 2 \quad \text{and} \quad (55)$$

$$\text{tr} \left[V^{-1} (\mathbf{1} \ \mu) B \begin{pmatrix} 0 \\ 1 \end{pmatrix} w' V \right] = B_{22} (\mu_0 - \mu_*). \quad (56)$$

PROOF. Equation (55) follows from commutativity of matrices within the trace operator and the

definition of B as follows: $tr \left[V^{-1} \begin{pmatrix} \mathbf{1} & \boldsymbol{\mu} \end{pmatrix} B \begin{pmatrix} \mathbf{1} & \boldsymbol{\mu} \end{pmatrix}' \right] = tr \left[\begin{pmatrix} \mathbf{1} & \boldsymbol{\mu} \end{pmatrix}' V^{-1} \begin{pmatrix} \mathbf{1} & \boldsymbol{\mu} \end{pmatrix} B \right] = tr[I_2] = 2.$

Equation (56) follows for similar reasons as follows:

$$\begin{aligned} tr \left[V^{-1} \begin{pmatrix} \mathbf{1} & \boldsymbol{\mu} \end{pmatrix} B \begin{pmatrix} 0 \\ 1 \end{pmatrix} (w'V) \right] &= tr \left[(w'V) V^{-1} \begin{pmatrix} \mathbf{1} & \boldsymbol{\mu} \end{pmatrix} B \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right] \\ &= tr \left[\begin{pmatrix} 1 & \boldsymbol{\mu}_0 \end{pmatrix} B \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right] = \xi = B_{22}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_*). \end{aligned} \quad (57)$$

Q.E.D. (Lemma 7)

LEMMA 8. Some useful expectations may be computed as follows:

$$E \left[S_2(\hat{B}, \boldsymbol{\varepsilon}) \right] = -\frac{n-2}{T-1} B \quad (58)$$

$$E \left[S_2(\hat{B}, \boldsymbol{\delta}) \right] = \frac{1}{T} B_{22} B - \frac{n-4}{T} B \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0 \quad 1) B \quad (59)$$

$$E \left[S_2(\hat{\boldsymbol{\sigma}}_0^2) \right] = \frac{B_{22}}{T} \boldsymbol{\sigma}_0^2 - \frac{n-4}{T} B_{22}^2 (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_*)^2 - \frac{n-2}{T-1} \boldsymbol{\sigma}_0^2 \quad (60)$$

$$E \left[S_2(\hat{w}'\boldsymbol{\varepsilon}\hat{w}) \right] = -\frac{2(n-2)}{T-1} \boldsymbol{\sigma}_0^2 \quad (61)$$

$$E \left[\delta' S_1(\hat{w}) \right] = \frac{n-3}{T} B_{22} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_*) \quad (62)$$

$$E \left[S_2(\hat{w}'V\hat{w}) \right] = \frac{B_{22}}{T} \boldsymbol{\sigma}_0^2 - \frac{n-4}{T} B_{22}^2 (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_*)^2 + \frac{n-2}{T-1} \boldsymbol{\sigma}_0^2 \quad (63)$$

$$E[S_2(\hat{w}'\mu\mu'\hat{w})] = \frac{\sigma_0^2}{T} - 2\mu_0 \frac{n-3}{T} B_{22}(\mu_0 - \mu_*). \quad (64)$$

PROOF. To prove Equation (58), use Lemma 1 to evaluate the expectation of (47), simplify while recognizing that $(\mathbf{1} \ \mu)' V^{-1}(\mathbf{1} \ \mu) = B^{-1}$, and using (55) to evaluate a trace as follows:

$$\begin{aligned} E[S_2(\hat{B}, \varepsilon)] &= B(\mathbf{1} \ \mu)' V^{-1} E\left\{\varepsilon \left[V^{-1}(\mathbf{1} \ \mu)B(\mathbf{1} \ \mu)' V^{-1} - V^{-1} \right] \varepsilon\right\} V^{-1}(\mathbf{1} \ \mu)B \\ &= \frac{1}{T-1} B(\mathbf{1} \ \mu)' V^{-1} \left\{ (\mathbf{1} \ \mu)B(\mathbf{1} \ \mu)' - V + V \operatorname{tr} \left[V^{-1}(\mathbf{1} \ \mu)B(\mathbf{1} \ \mu)' - I_n \right] \right\} V^{-1}(\mathbf{1} \ \mu)B \quad (65) \\ &= -\frac{n-2}{T-1} B. \end{aligned}$$

For Equation (59), apply Lemma 1 to (45), and use (33) to find:

$$\begin{aligned} E[S_2(\hat{B}, \delta)] &= \operatorname{tr} \left[V^{-1}(\mathbf{1} \ \mu)B(\mathbf{1} \ \mu)' - I_n \right] B \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0 \ 1)B/T \\ &\quad + B_{22}B(\mathbf{1} \ \mu)' V^{-1}(\mathbf{1} \ \mu)B/T + B \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0 \ 1)B(\mathbf{1} \ \mu)' V^{-1}(\mathbf{1} \ \mu)B/T \quad (66) \\ &\quad + B(\mathbf{1} \ \mu)' V^{-1}(\mathbf{1} \ \mu)B \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0 \ 1)B/T. \end{aligned}$$

Using (55) and the definition of B , (66) simplifies to establish (59). For Equation (60), use (58) and (59)

with the fact that $\hat{\sigma}_0^2 = (\mathbf{1} \ \mu_0)\hat{B}(\mathbf{1} \ \mu_0)'$ to find

$$\begin{aligned} E[S_2(\hat{\sigma}_0^2)] &= (\mathbf{1} \ \mu_0) E \left[S_2(\hat{B}, \delta) + S_2(\hat{B}, \varepsilon) \right] \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} \\ &= (\mathbf{1} \ \mu_0) \left[\frac{1}{T} B_{22}B - \frac{n-4}{T} B \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0 \ 1)B - \frac{n-2}{T-1} B \right] \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix}, \quad (67) \end{aligned}$$

which, recognizing that $(\mathbf{1} \ \mu_0)B(0 \ 1)' = \xi = B_{22}(\mu_0 - \mu_*)$, simplifies to (60). For Equation (61), by symmetry and using (52) we have

$$E[S_2(\hat{w}'\varepsilon\hat{w})] = 2w'E[\varepsilon S_1(\hat{w}, \varepsilon)] = 2w'E\left\{\varepsilon\left[V^{-1}(\mathbf{1} \ \mu)B(\mathbf{1} \ \mu)'V^{-1} - V^{-1}\right]\varepsilon\right\}w. \quad (68)$$

Using Lemma 1 to evaluate the expectation and (55) to evaluate a trace, we simplify to find

$$\begin{aligned} E[S_2(\hat{w}'\varepsilon\hat{w})] &= 2w'\left\{\left[(\mathbf{1} \ \mu)B(\mathbf{1} \ \mu)' - V\right] + Vtr\left[V^{-1}(\mathbf{1} \ \mu)B(\mathbf{1} \ \mu)' - I_n\right]\right\}w/(T-1) \\ &= 2w'\left\{\left[(\mathbf{1} \ \mu)B(\mathbf{1} \ \mu)' - V\right] + (2-n)V\right\}w/(T-1) \\ &= 2\left[(\mathbf{1} \ \mu_0)B\begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} - (n-1)\sigma_0^2\right]/(T-1) = \frac{2}{T-1}\left[\sigma_0^2 - (n-1)\sigma_0^2\right] \\ &= -\frac{2(n-2)}{T-1}\sigma_0^2. \end{aligned} \quad (69)$$

For Equation (62), first note that $E[\delta'S_1(\hat{w})] = E[\delta'S_1(\hat{w}, \delta)]$ because $E[\delta'S_1(\hat{w}, \varepsilon)] = 0$ by independence of δ and ε , then substitute using (51), evaluate the expectation using Lemma 1, and find the trace using (55) and (56) to obtain

$$\begin{aligned} E[\delta'S_1(\hat{w}, \delta)] &= E\left\{\delta'\left[\xi V^{-1} - V^{-1}(\mathbf{1} \ \mu)B\begin{pmatrix} 0 \\ 1 \end{pmatrix}w' - \xi V^{-1}(\mathbf{1} \ \mu)B(\mathbf{1} \ \mu)'V^{-1}\right]\delta\right\} \\ &= \frac{1}{T}tr\left[\xi I_n - V^{-1}(\mathbf{1} \ \mu)B\begin{pmatrix} 0 \\ 1 \end{pmatrix}w'V - \xi V^{-1}(\mathbf{1} \ \mu)B(\mathbf{1} \ \mu)'\right] \\ &= \frac{1}{T}[n\xi - \xi - 2\xi], \end{aligned} \quad (70)$$

which simplifies to (62). To prove (63), observe that

$$E[S_2(\hat{w}'V\hat{w})] = E[S_2(\hat{w}'\hat{V}\hat{w} - \hat{w}'\varepsilon\hat{w})] = E[S_2(\hat{\sigma}_0^2)] - E[S_2(\hat{w}'\varepsilon\hat{w})]. \quad (71)$$

Substituting using (60) and (61), then simplifying, establishes (63). To prove (64), note using $\mu_0 = \hat{w}'\hat{\mu}$ that

$$\begin{aligned} S_2(\hat{w}'\mu\mu'\hat{w}) &= S_2\left[\hat{w}'(\hat{\mu}-\mu)(\hat{\mu}-\mu)'\hat{w} + 2\mu_0\hat{w}'\mu - \mu_0^2\right] = S_2\left[\hat{w}'\delta\delta'\hat{w} + 2\mu_0\hat{w}'\mu - \mu_0^2\right] \\ &= w'\delta\delta'w + 2\mu_0 S_2(\hat{w}'\mu). \end{aligned} \quad (72)$$

The expectation may be found using Lemma 2 and Theorem 1, as follows:

$$\begin{aligned} E\left[S_2(\hat{w}'\mu\mu'\hat{w})\right] &= w'E(\delta\delta')w + 2\mu_0 E\left[S_2(\hat{w}'\mu)\right] \\ &= \frac{w'Vw}{T} - 2\mu_0 \frac{n-3}{T} B_{22}(\mu_0 - \mu_*), \end{aligned} \quad (73)$$

completing the proof.

Q.E.D. (Lemma 8)

PROOF OF THEOREM 1. Since R_{T+1} is independent of \hat{w} (and observing that second-order terms containing both δ and ε will have expectation zero) we may write

$$\begin{aligned} E_\Delta(\hat{w}'\mu) &= E_\Delta(\hat{w}'\hat{\mu}) - E_\Delta[\hat{w}'(\hat{\mu}-\mu)] = \mu_0 - E_\Delta(\delta'\hat{w}) \\ &= \mu_0 - E\left[S_2(\delta'\hat{w})\right] = \mu_0 - E\left[\delta'S_1(\hat{w})\right], \end{aligned} \quad (74)$$

which simplifies using (62) from Lemma 8 to establish (7). To prove Equation (9), we expand

$E_\Delta(\hat{\mu}_{adjusted})$ to second order as follows:

$$\begin{aligned} E_\Delta\left\{\mu_0 - \frac{n-3}{T}\hat{B}_{22}(\mu_0 - \hat{\mu}_*)\right\} &= \mu_0 - \frac{n-3}{T}B_{22}(\mu_0 - \mu_*) + E\left\{S_2\left[\mu_0 - \frac{n-3}{T}\hat{B}_{22}(\mu_0 - \hat{\mu}_*)\right]\right\} \\ &= E_\Delta(\hat{w}'R_{T+1}) + O\left(\frac{1}{T^2}\right) \end{aligned} \quad (75)$$

where the last equality was obtained by observing that expectations of second-order expansion terms with respect to either δ or ε will be $O(1/T)$ by Lemma 1.

Q.E.D. (Theorem 1)

PROOF OF THEOREM 2. Equation (10) is obtained by applying (60) from Lemma 8 to the delta-method expansion $E_{\Delta}(\hat{\sigma}_0^2) = \sigma_0^2 + E[S_2(\hat{\sigma}_0^2)]$. Having used iterated expectation to establish the second equality of (11), we next expand as follows :

$$E_{\Delta}[\hat{w}'V\hat{w} + \hat{w}'\mu\mu'\hat{w}] - [E_{\Delta}(\hat{w}'\mu)]^2 = \sigma_0^2 + E[S_2(\hat{w}'V\hat{w})] + \mu_0^2 + E[S_2(\hat{w}'\mu\mu'\hat{w})] - [E_{\Delta}(\hat{w}'\mu)]^2 \quad (76)$$

and then use (63), (64), and Theorem 1 to evaluate the resulting delta-method expectations, finally simplifying to complete the proof. To prove (13), note that $E_{\Delta}(\hat{\sigma}_{adjusted}^2) = [1 + (2n - 3)/T]E_{\Delta}(\hat{\sigma}_0^2) + O(1/T^2)$, then substitute from (10) and simplify.

Q.E.D. (Theorem 2)

PROOF OF THEOREM 3. Begin with the naïve frontier that expresses $\hat{\sigma}_0$ as a function of μ_0 :

$$\hat{\sigma}_0 = \sqrt{\hat{\sigma}_*^2 + \hat{B}_{22}(\mu_0 - \hat{\mu}_*)^2} \quad (77)$$

then substitute for $\mu_0 - \hat{\mu}_*$ by solving Equation (8), and substitute for $\hat{\sigma}_0$ by solving (12).

Q.E.D. (Theorem 3)

PROOF OF THEOREM 4. When $r_f < \hat{\mu}_*$, the naïve Sharpe ratio S_{naive} , which maximizes $(\mu_0 - r_f)/\hat{\sigma}_0$ for the naïve frontier (77) may be written as

$$S_{naive} = \sqrt{\frac{1}{\hat{B}_{22}} + \frac{(\hat{\mu}_* - r_f)^2}{\hat{\sigma}_*^2}} \quad (78)$$

Since the adjusted frontier remains geometrically a hyperbola with equation given by (14) instead of (77), it follows that we may make the corresponding substitutions in the naïve Sharpe ratio formula (78). These substitutions, $\hat{\sigma}_* \rightarrow [1 + (n-1.5)/T] \hat{\sigma}_*$ and $\hat{B}_{22} \rightarrow \hat{B}_{22} [1 + (n-1.5)/T]^2 [1 - (n-3)\hat{B}_{22}/T]^{-2}$, establish the first equality in (15). The second equality follows by solving (20) for $(\hat{\mu}_* - r_f)^2 / \hat{\sigma}_*^2$, substituting, and simplifying.

Q.E.D. (Theorem 4)

PROOF OF THEOREM 5. From the definition of the Sharpe-ratio maximizing portfolio the naïve CML is defined as:

$$\hat{\mu}_{CML(naive)}(r_f) = xr_f + (1-x) \left(\hat{\mu}_* + \frac{\hat{\sigma}_*^2}{\hat{B}_{22}(\hat{\mu}_* - r_f)} \right) \quad (79)$$

$$\hat{S}_{CML(naive)}(r_f) = (1-x) \hat{S}_* \sqrt{1 + \frac{\hat{S}_*^2}{\hat{B}_{22}(\hat{\mu}_* - r_f)^2}} \quad (80)$$

Since the adjusted frontier remains geometrically a hyperbola with equation given by (14) instead of (77), it follows that (as in the proof of Theorem 4) we may make corresponding substitutions in the naïve CML formulae to obtain (16) and (17).

Q.E.D. (Theorem 5)

PROOF OF COROLLARY TO THEOREM 5. This follows by reversing equation (3) to define the naïve target mean as function of the adjusted target mean and then substituting the risky component of equation (16) for the adjusted-Sharpe-ratio maximizing portfolio performance.

Q.E.D. (Corollary to Theorem 5)

PROOF OF THEOREM 6. For \hat{w} on the estimated efficient frontier (1) we have $\hat{w}'\hat{V}\mathbf{1} = n \left(1 \quad \hat{\mu} \right) \hat{B} \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix}$

and $\hat{\sigma}_0^2 = \hat{w}'\hat{V}\hat{w} = \left(1 \quad \mu_0 \right) \hat{B} \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix}$, where $\hat{\mu} = \sum_{i=1}^n \hat{\mu}_i / n$ is the equally weighted average of the

estimated asset means. The diversification measure $D\hat{i}v = \left(\hat{w} - \frac{1}{n} \mathbf{1} \right)' \hat{V} \left(\hat{w} - \frac{1}{n} \mathbf{1} \right)$ can be expanded as

follows:

$$\begin{aligned} D\hat{i}v &= (1 \quad \mu_0) \hat{B} \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} - 2(1 \quad \mu_0) \hat{B} \begin{pmatrix} 1 \\ \hat{\mu} \end{pmatrix} + \frac{1}{n^2} \mathbf{1}' \hat{V} \mathbf{1} \\ &= \hat{B}_{22} (\mu_0 - \hat{\mu})^2 + \frac{1}{n^2} \mathbf{1}' \hat{V} \mathbf{1} - (1 \quad \hat{\mu}) \hat{B} \begin{pmatrix} 1 \\ \hat{\mu} \end{pmatrix}. \end{aligned} \quad (81)$$

Hence $D\hat{i}v$ is closer to 0 the closer μ_0 is to $\hat{\mu}$, ceteris paribus. The target mean that maximizes the naïve Sharpe ratio can be written as follows:

$$\hat{\mu}_{0,S-\max-naive} = \hat{\mu}_* + \frac{\hat{\sigma}_*^2}{\hat{B}_{22} (\hat{\mu}_* - r_f)}. \quad (82)$$

Using equation (19) for the naïve mean corresponding to the adjusted-Sharpe-ratio maximizing portfolio, one can see that the target mean for the naïve-Sharpe ratio maximizing portfolio will always be greater than the naïve mean corresponding to the adjusted-Sharpe-ratio maximizing portfolio, provided $n > 3$ and $\hat{\mu}_* > r_f$. Hence the adjusted-Sharpe-ratio maximizing portfolio is more diversified whenever

$$\hat{\mu} < \frac{\hat{\mu}_{0,S-\max-naive} + \hat{\mu}_{0,S-\max-adj}}{2}. \quad (83)$$

Equation (83) leads to the third restriction in the theorem statement and completes the proof.

Q.E.D. (Theorem 6)