How Much Error is in the Tracking Error?

The Impact of Estimation Risk on Fund Tracking Error[1]

Artemiza Woodgate

Russell Investments, Seattle, Washington 98101

awoodgate@russell.com

Andrew F. Siegel

Foster School of Business, University of Washington, Seattle, Washington 98195

asiegel@uw.edu

We explain the poor out-of-sample performance of optimized portfolios (to minimize tracking-error relative to a given benchmark while achieving a specified expected excess return) in the presence of estimation error in the underlying asset means and covariances. Our theoretical bias adjustments for this estimation risk are developed by taking mathematical expectations of asymptotically expanded future returns of portfolios formed with estimated weights. We provide closed-form adjustments for estimates of the expectation and standard deviation of the portfolio's excess returns. The adjustments significantly reduce bias in global equity portfolios, reduce the costs of rebalancing portfolios, and are robust to sample size and to non-normality. Using these approximation methods it may be possible to assess, before investing, the effect of statistical estimation error on tracking-error-optimized portfolio performance.

[1] The authors gratefully acknowledge research support from the Q-Group.

1. Introduction

The performance of a fund relative to a chosen benchmark has traditionally been a key determinant of the fund's attractiveness as an investment. Hence a commonly implemented strategy throughout the investment community is to minimize the tracking error, subject to a given expected level of positive excess return, relative to a given benchmark.

In spite of their wide use, expected excess returns and tracking errors have been difficult to estimate especially in mean-variance optimized portfolios, in part due to the statistical noise contained in the means and variances of asset returns estimated from historical data (see Focardi, Kolm and Fabozzi [2004] for a review). Chan, Karceski, and Lakonishok [1999] look at forecasting the second moments of the returns in order to reduce the tracking error, while Dessislava [2006] suggests using robust optimization. Our research looks at the impact of estimation error on the expected excess return and tracking error, for portfolios optimized using mean-variance optimization. We derive asymptotic formulae that eliminate dominant terms of the bias arising from noisy estimates, hence improving the implementation power of the tracking measures. Apart from deriving these closed form bias adjustments, we test their performance using global data, and also look at the impact of the bias adjustment on the transaction costs arising from following a tracking-error minimization strategy.

Mutual funds often trade heavily to minimize tracking error, thereby incurring high transaction costs. But high transaction fees incurred from such trading can outweigh the benefits of a low tracking error, especially in stocks that are illiquid. Keim [1999] showed that a strategy that minimizes trading costs for a small-cap stock index fund provides an annual premium of 2.2% over the index. Our research relates to this literature by studying the hypothetical transaction costs arising from the implementation of the "naïve" traditional tracking error minimization strategy, versus the new adjusted tracking error that we propose. We show that the transaction costs are lower once unnecessary trades driven by estimation noise are eliminated.

We extend, to the case of tracking-error-optimized portfolios, the methods developed by Siegel and Woodgate [2007] for mean-variance-optimized portfolios to adjust for estimation error in mean-variance-efficient portfolio formation. We find asymptotic non-Bayesian closed-form formulae for tracking-error-optimized portfolio performance while accounting for estimation risk without relying upon prior assumptions on the unknown parameters, and provide an adjustment for statistical noise that asymptotically reflects the actual performance of the tracking-error-optimized portfolio.

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 a tracking-error-optimized portfolio that achieves a specified expected excess return with respect to a given benchmark. 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, estimates of the naïve portfolio's expected excess return tend to be biased upwards while estimates of the naïve portfolio's tracking-error tend to be biased downwards, resulting in nominally tracking-error-optimized portfolios that are "over-optimistic," in the sense that such an investor will believe she can achieve a higher expected excess return and lower tracking error than is actually available from the performance of her portfolio. We quantify this "over-optimism" bias in closed-form asymptotic formulae, which help in two ways when portfolio weights are influenced by statistical noise. First, there may be a systematic component that moves the portfolio expected excess return away from its target; this effect is captured by our expected excess return adjustment. Second, to the extent that the portfolio weights reflect the variability of the noise and the estimated variability, the variance of the portfolio excess returns will increase; this effect is picked up by our adjustment for the tracking error as measured by the standard deviation of the excess returns.

We show how the investor can adjust the performance of naïvely-estimated tracking-error-optimizing portfolios so that actual portfolio performance is more accurately anticipated by accounting for distortions due to estimation error. We use the method of statistical differentials to find Taylor-series approximations to expectations of random variables, obtaining theoretical 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.

We use global country equity indexes to illustrate the adjustments, construct frontiers of expectation versus standard deviation of portfolio excess return, and test for significant reductions in the out-of sample bias in the mean as well as for the impact of the adjustments on transaction costs.

2. Theoretical Bias Adjustments

Consider $n \geq 3$ 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*, and let w_B denote the column vector of weights of the *n* stocks as held in the given benchmark. Asset return vectors $R_t = \left(R_{1t},...,R_{nt}\right)'$ are assumed to be independent and identically normally distributed with unknown true mean vector $\mu = E\left(R_t\right)$ 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 $\varepsilon = \hat{V} - V$.

Let w_F denote the vector of weights chosen by the fund to track the benchmark, and let $w = w_F - w_B$ denote the perturbation of the benchmark weights w_B that produce the fund weights w_F so that $w_F = w_B + w$. Then the fund's realized excess return X at time $T+1$ may be written as

$$X = (w_F - w_B)' R_{T+1} = w' R_{T+1} \tag{1}$$

while the expected value and variance of excess returns (also known as the tracking error) denoted α and σ^2 respectively, are

$$\alpha \equiv E(X) = (w_F - w_B)' \mu = w' \mu \tag{2}$$

$$\sigma^2 = Var(X) = (w_F - w_B)' V (w_F - w_B) = w' V w \tag{3}$$

Note that the nature of the excess return, its expectation, and its variance, all have the same functional form regardless of the particular choice of benchmark weights w_B.

Let α_0 denote the given target for the expected excess return. The perturbation of the benchmark weights that achieves the smallest tracking-error σ^2 among all portfolios that have expected excess return α_0 may be written as

$$w = w_F - w_B = V^{-1} (\mathbf{1} \quad \mu) \left[(\mathbf{1} \quad \mu)' V^{-1} (\mathbf{1} \quad \mu) \right]^{-1} \begin{pmatrix} 0 \\ \alpha_0 \end{pmatrix} = V^{-1} (\mathbf{1} \quad \mu) B \begin{pmatrix} 0 \\ \alpha_0 \end{pmatrix} \tag{4}$$

where we define the $n \times 1$ vector $\mathbf{1} = (1,...,1)'$ and the 2×2 matrix $B = \left[(\mathbf{1} \quad \mu)' V^{-1} (\mathbf{1} \quad \mu) \right]^{-1}$. The resulting tracking-error is then

$$\sigma^2 = w'Vw = (0 \quad \alpha_0) B \begin{pmatrix} 0 \\ \alpha_0 \end{pmatrix} = \alpha_0^2 B_{22} \tag{5}$$

Substituting the estimates $\hat{\mu}$ and $\hat{V}$ based on observations at times $t = 1,...,T$ in place of their true unobservable values μ and V, the perturbation weights are estimated, for a given target expected excess return α_0, as

$$\hat{w} = \hat{w}_F - w_B = \hat{V}^{-1} (\mathbf{1} \quad \hat{\mu}) \left[(\mathbf{1} \quad \hat{\mu})' \hat{V}^{-1} (\mathbf{1} \quad \hat{\mu}) \right]^{-1} \begin{pmatrix} 0 \\ \alpha_0 \end{pmatrix} = \hat{V}^{-1} (\mathbf{1} \quad \hat{\mu}) \hat{B} \begin{pmatrix} 0 \\ \alpha_0 \end{pmatrix} \tag{6}$$

where we use the natural definition $\hat{B} = \left[(\mathbf{1} \quad \hat{\mu})' \hat{V}^{-1} (\mathbf{1} \quad \hat{\mu}) \right]^{-1}$. Note that the weights $\hat{w}$ would indeed be optimal if $\hat{\mu}$ and $\hat{V}$ were the true parameters, but are instead suboptimal as compared to the best-possible but unobservable perturbation weights (4).

The next-period excess return (at time $T+1$) of a portfolio with estimated perturbation weights $\hat{w}$ is $\hat{X} = \hat{w}' R_{T+1}$, a dot product of two independent random vectors. To characterize this next-period return, we seek adjustments to two naïve excess return performance measures. The "naïve expected excess return" is defined as the target excess return $\alpha_0 = \hat{w}'\hat{\mu}$, while the "naïve standard deviation" of the excess return is

$$\hat{\sigma}_0 = \sqrt{\hat{w}' \hat{V} \hat{w}} = \alpha_0 \sqrt{\hat{B}_{22}} \ . \tag{7}$$

Note that α_0 would be the expected excess return and $\hat{\sigma}_0$ would be the standard deviation of the excess return 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 also let $\sigma_0 = \sqrt{w'Vw} = \alpha_0\sqrt{B_{22}}$ denote the standard deviation of the portfolio with weights *w* based on the unobservable true parameters.

Our main results will suggest an approximate correction for the bias of these naïve performance measures α_0 and $\hat{\sigma}_0$ by using the following adjusted measures of the excess return expectation and risk:

$$\hat{\alpha}_{adjusted} = \alpha_0\left(1 - \frac{n-3}{T}\hat{B}_{22}\right) \quad \text{and} \tag{8}$$

$$\hat{\sigma}_{adjusted} = \left(1 + \frac{n-1.5}{T}\right)\hat{\sigma}_0 . \tag{9}$$

We will show that the adjusted measures (8) and (9) more accurately represent the actual expected excess return and the actual standard deviation as they would be experienced by the investor.

Finding these adjustments is a difficult task because the expectations involved do not seem to have closed-form mathematical solutions. For example, the bias in the expected excess return (the expected value of the target expected excess return minus the actual excess return),

$$\begin{aligned} E\left(\alpha_0 - \hat{w}'R_{T+1}\right) &= \alpha_0 - \mu' E\left(\hat{w}\right) \\ &= \alpha_0 - \mu' E\left\{\hat{V}^{-1}\left(\mathbf{1} \quad \hat{\mu}\right)\left[\left(\mathbf{1} \quad \hat{\mu}\right)'\hat{V}^{-1}\left(\mathbf{1} \quad \hat{\mu}\right)\right]^{-1}\right\}\begin{pmatrix} 0 \\ \alpha_0 \end{pmatrix} \end{aligned} \tag{10}$$

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

Following Siegel and Woodgate (2007) we asymptotically approximate these difficult expectations of functions of estimated values $\hat{\mu}$ and $\hat{V}$, using 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 $f\left(\hat{\mu},\hat{V}\right) \cong f\left(\mu,V\right) + S_1\left[f\left(\hat{\mu},\hat{V}\right)\right] + S_2\left[f\left(\hat{\mu},\hat{V}\right)\right]$, where $S_1\left(f\right)$ 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 $S_2\left(f\right)$ denotes the second order terms, then the delta-method asymptotic expectation, denoted $E_\Delta\left[f\left(\hat{\mu},\hat{V}\right)\right]$, will be

$$E_\Delta\left[f\left(\hat{\mu},\hat{V}\right)\right] \equiv f\left(\mu,V\right) + E\left\{S_2\left[f\left(\hat{\mu},\hat{V}\right)\right]\right\} \cong E\left[f\left(\hat{\mu},\hat{V}\right)\right]. \tag{11}$$

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 $T \to \infty$ 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 $\left(\hat{\mu},\hat{V}\right)$ in place of the true asset parameters $\left(\mu,V\right)$, where these adjustments approximately reflect the future uncertainty of the not-yet-observed time $T+1$ asset returns R_{T+1} together with the estimation uncertainty in the perturbation weights $\hat{w}$.

Theorem 1. *The target expected excess return α_0 is systematically biased as a measure of expected future portfolio performance because the delta-method expectation of the expected improvement upon the benchmark's expected return is $E_\Delta\left(\hat{w}'\mu\right)$* where

$$E\left(\hat{w}'R_{T+1}\right) = E\left(\hat{w}'\mu\right) \cong E_\Delta\left(\hat{w}'\mu\right) = \alpha_0\left(1 - \frac{n-3}{T}B_{22}\right). \tag{12}$$

If we define the adjusted expected excess return to be the estimated right-hand side of (12) *so that*

$$\hat{\alpha}_{adjusted} \equiv \alpha_0\left(1 - \frac{n-3}{T}\hat{B}_{22}\right), \tag{13}$$

then $\hat{\alpha}_{adjusted}$ is a second-order unbiased estimator of expected future excess return, eliminating the $O(1/T)$ term from the bias in the sense that

$$E\left(\hat{\alpha}_{adjusted}\right) \cong E_\Delta\left(\hat{\alpha}_{adjusted}\right) = E_\Delta\left(\hat{w}'R_{T+1}\right) + O\left(\frac{1}{T^2}\right) \tag{14}$$

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

$$E_\Delta\left(\hat{\sigma}_0^2\right) = E_\Delta\left(\hat{w}'\hat{V}\hat{w}\right) = \sigma_0^2\left(1 - \frac{n-5}{T}B_{22} - \frac{n-2}{T}\right) + O\left(\frac{1}{T^2}\right) \tag{15}$$

while the actual tracking-error, evaluated using the delta-method, is

$$Var\left(\hat{w}'R_{T+1}\right) \cong \sigma_0^2\left(1-\frac{n-5}{T}B_{22}+\frac{n-1}{T}\right)+O\left(\frac{1}{T^2}\right)=E_\Delta\left(\hat{\sigma}_0^2\right)+\frac{2n-3}{T}\sigma_0^2+O\left(\frac{1}{T^2}\right). \qquad (16)$$

If we define the adjusted standard deviation to be

$$\hat{\sigma}_{adjusted} \equiv \left(1+\frac{n-1.5}{T}\right)\hat{\sigma}_0, \qquad (17)$$

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

$$E\left(\hat{\sigma}^2_{adjusted}\right) \cong E_\Delta\left(\hat{\sigma}^2_{adjusted}\right)=Var_\Delta\left(\hat{w}'R_{T+1}\right)+O\left(\frac{1}{T^2}\right). \qquad (18)$$

Proofs are in the Appendix. Note, as expected, that the adjusted expected excess return $\hat{\alpha}_{adjusted}$ from (13) is less than the target expected excess return α_0 when we seek a higher target mean than the benchmark and $n>3$ because $\hat{B}$ is positive definite, and that the adjusted standard deviation $\hat{\sigma}_{adjusted}$ from (17) 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 standard deviation $\hat{\sigma}_{adjusted}$ from $\hat{\sigma}_0$, but that in finding the adjusted expected excess return $\hat{\alpha}_{adjusted}$ from α_0 we also use the estimated values $\hat{\mu}$ and $\hat{V}$ to obtain $\hat{B}_{22}$.

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.

3. Empirical Results with Global Portfolios

Because the adjustments are asymptotic, questions arise: How well do the bias adjustments work in finite (small) samples? How significant, statistically and economically, are the adjustments? Our investment opportunity set consists of index portfolios for each of 18 developed countries (Australia, Austria, Belgium, Canada, Denmark, France, Germany, Hong Kong, Italy, Spain, Netherlands, Norway, Singapore, Spain, Sweden, Switzerland, UK, and USA). Monthly returns are computed from the total return country index with dividends reinvested, obtained from Morgan Stanley Capital International (MSCI), for the 464-month period 1/30/1970 to 9/30/2008.

The benchmark used in the studies below is an equal-weighted portfolio of the country returns. The key reason for constructing an equal-weighted benchmark rather than using a given capitalization-weighted benchmark is that we want to see how sensitive the adjustment formulae are to small samples, and we don't want to have the benchmark being dominated by a few large countries.

The bias adjustment is studied empirically in four 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 bootstrap simulations to study the adjustment while relaxing the normality assumption used in the derivations. Third, we test the effectiveness of the adjustment using step-ahead performance of portfolios formed from a window of past data, and fourth, we analyze the costs associated with rebalancing portfolios with and without taking into consideration estimation error.

A. Estimation Error Adjustment in Global Equity Portfolios

Using the full data set of 18 country indexes, we form portfolios of excess returns using 6, 12, and 18 country indexes, relative to the equal-weighted benchmark in each case. Figure 1 presents the efficient frontiers of excess returns versus the standard deviation of these excess returns. For the “Naïve” frontier the α_0 naïve excess return is graphed versus the naïve standard deviation given by equation (7), which ignores the existence of estimation error. The “Adjusted” frontier uses the adjusted expected excess return, $\hat{\alpha}_{adjusted}$ from (13), graphed versus the adjusted standard deviation, $\hat{\sigma}_{adjusted}$ from (17).

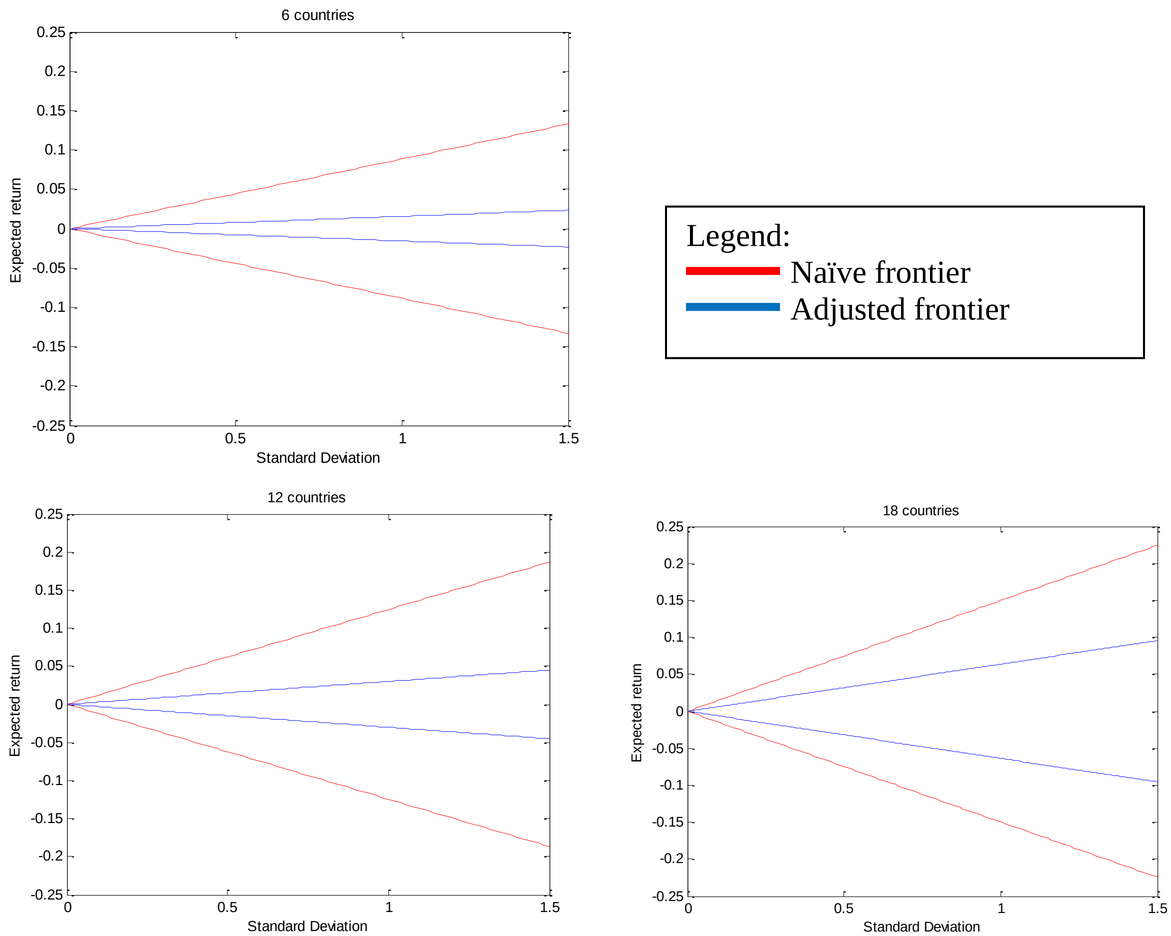

Figure 1
Efficient Frontiers in Excess Return Space, With and Without Bias Adjustment, for Portfolios Formed from Six to Eighteen Country Index Funds Chosen in the Following Order: Australia, Austria, Belgium, Canada, Denmark, France, Germany, Hong Kong, Italy, Spain, Netherlands, Norway, Singapore, Spain, Sweden, Switzerland, UK, USA, Using 464 Months of Returns Data from 1/30/1970 to 9/30/2008. In each case the frontier becomes smaller when adjusted for bias due to estimation error.

With six assets the naïve frontier is markedly wider than the adjusted frontier, while the adjusted frontier is quite narrow. These indicate that estimation error plays a substantial role even with a small number of assets. As the number of countries increases, both frontiers widen. If we hold α_0 constant across the three charts, we note that both the naïve and adjusted standard deviations decrease as the number of assets increases, indicating that the additional assets have been helpful in controlling risk. In all cases the estimation error adjustment seems to have an economically meaningful impact. The subsequent sections analyze the robustness of this adjustment as well as the statistical significance in more detail.

B. Bootstrap Simulations

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 begin with the full data set of 464 months for eighteen countries which define the population of months. From this population the bootstrap method samples (with replacement) vectors of country returns. We consider 6, 12, and 18 country portfolios, in each instance drawing all country returns simultaneously from the sampled month to preserve the empirical covariance structure.

We simulate portfolios using four time intervals: T = 60, 120, 464, and 1,000 months, choosing in each case 10,000 bootstrap samples (each sample is a time series using T months, of country index excess returns over the equal-weighted benchmark for that number of countries). We average each frontier (Actual, Naïve, and Adjusted) across the 10,000 simulations, to analyze the robustness of the adjustments for estimation error. The Actual frontier is computed using the empirical means and covariances (the exact bootstrap distribution) with the estimated weights. The Naïve frontier is computed using the estimated means and covariances with the estimated weights, while the Adjusted frontier is computed using formulae (13) and (17).

Figure 2 presents the bootstrap averages using a relatively short window of data, 60 months.

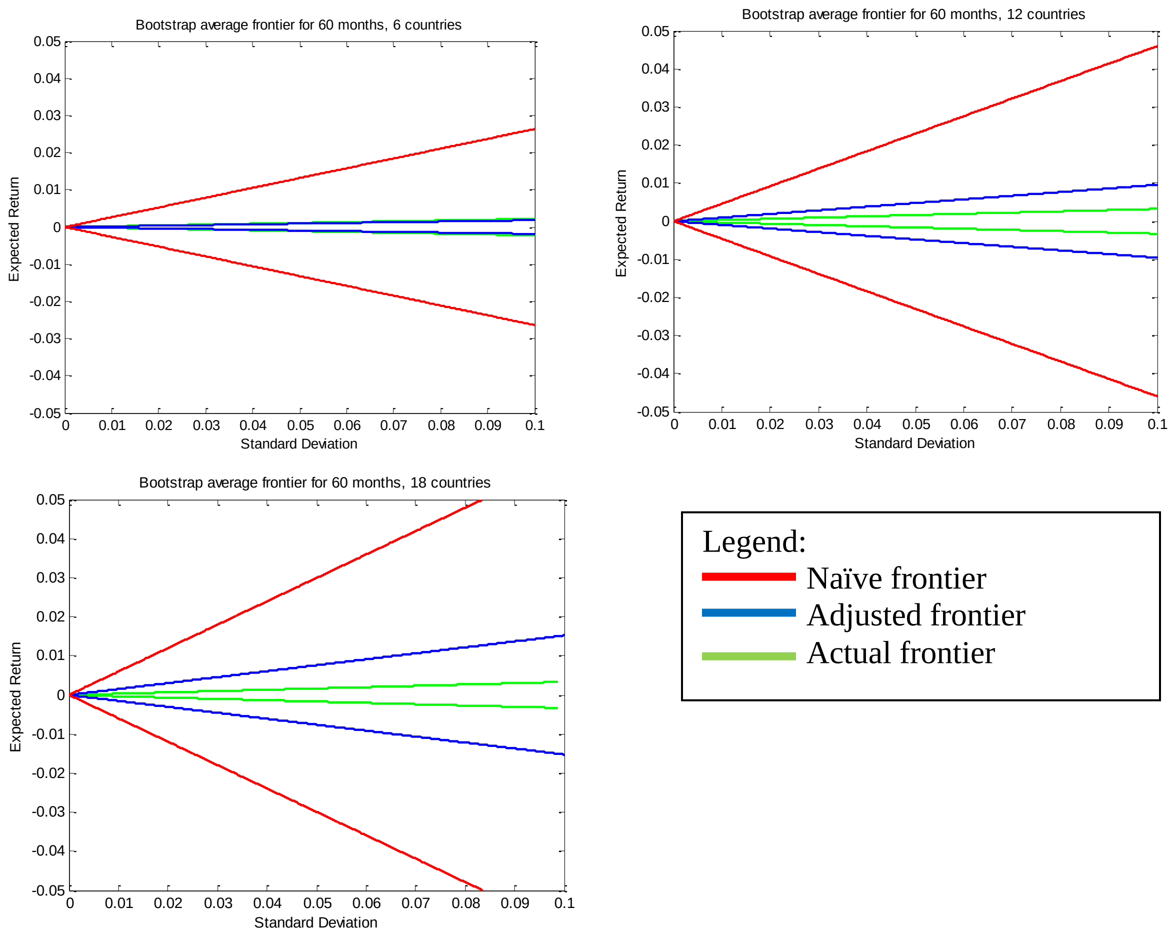

Figure 2
Bootstrap averages of Actual, Naïve, and Adjusted Efficient Frontiers in Excess Return Space, for Portfolios Formed from Six to Eighteen Country Index Funds using windows of 60 months of data. Excess Returns are computed relative to an equal-weighted benchmark. In each case the naïve frontier is widest and the actual frontier is smallest.

With this window of 60 months, there is considerable estimation error with small numbers n of assets (six countries) and the adjustment seems to nearly perfectly correct the bias by aligning the adjusted and actual frontiers. As the number of countries grows to 12 and to 18, the estimation error remains large and the bias adjustment moves closer to the actual frontier; however, the bias adjustment starts deviating from the actual frontier, as the asymptotic term in the bias adjustment for the standard deviation is quadratic in n for a given T.

Figure 3 compares the size of the adjustment as the window of data is increased in bootstrap simulations. Bootstrap averages for Actual, Adjusted, and Naïve frontiers using 12 countries are computed, varying the window of data from 120 to 464 and to 1000 months

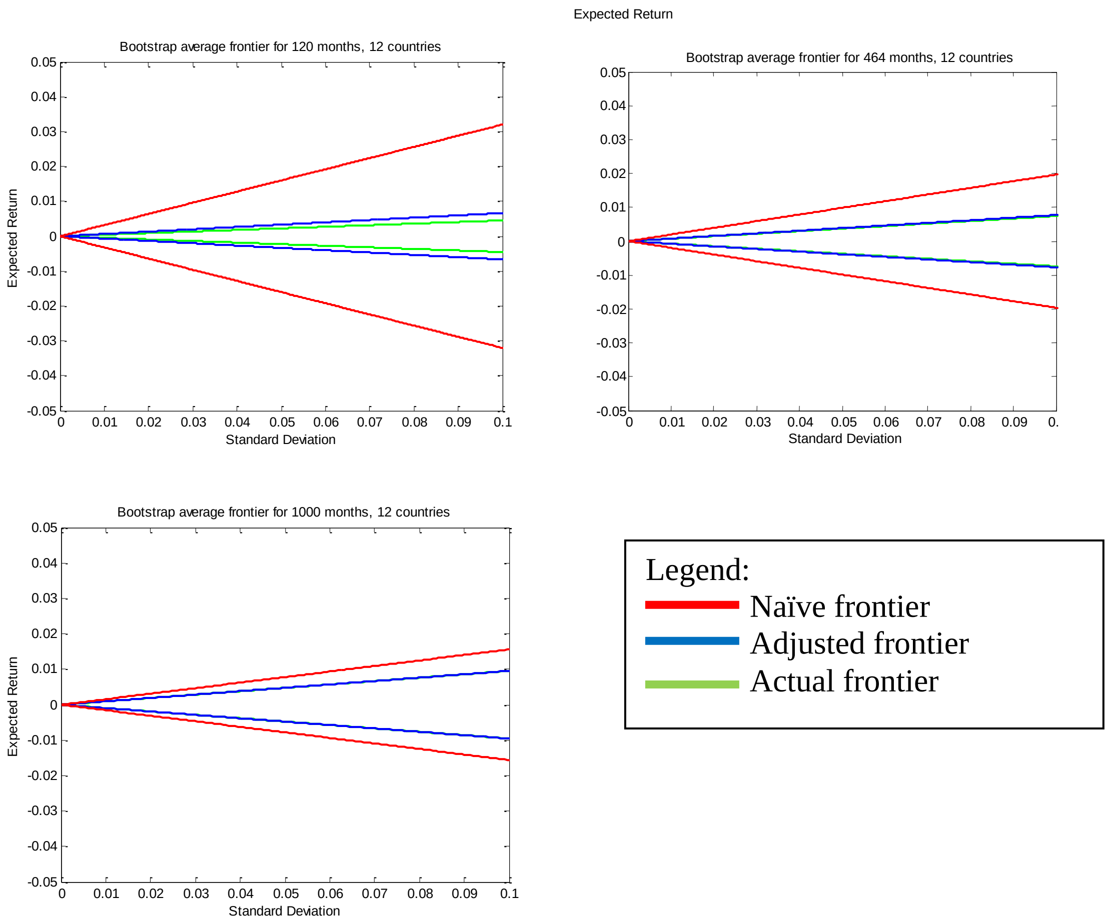

Figure 3
Bootstrap averages of Actual, Naïve, and Adjusted Efficient Frontiers in Excess Return Space, for Portfolios Formed from 12 Country Index Funds using windows of 120, 464, and 1000 months of data. Excess Returns are computed relative to an equal-weighted benchmark. As the number of months grows, the asymptotic bias adjustment improves, with the adjusted frontier becoming indistinguishable from the actual frontier.

While the bias adjustments are derived assuming that returns are normally distributed, overall the bootstrap results of Figure 3 indicate that the proposed adjustment is robust to this return normality assumption, effectively eliminating nearly all of the bias shown in Figure 2, as the time interval increases. Figure 3 shows that larger interval 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 the interval size grows, the optimization step is less distorted by the noise in the data. Second, the adjusted and actual frontiers become wider because, as the interval size grows, the estimated weights move closer to the true optimal weights.

C. Rolling Window Tests

The bootstrap results of the previous section indicate that the Adjusted frontier corrects a large part of the estimation bias, coming close to the Actual frontier. In this section we analyze statistical significance of the bias adjustment.

We test the statistical significance of the estimation risk bias adjustment by forming portfolios, each using estimates from a fixed-length time 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 expected performance anticipated by the naïve and the adjusted frontiers. At each step in time, using a window of the most recent 36, 60, or 120 months of data, we calculate the mean-variance frontier (with and without adjustment) and estimate the naïve portfolio weights $\hat{w}_t$ corresponding to a fixed target excess return of 2% per annum converted to a monthly target of 2% / 12 for computing purposes. Generally, all reported excess returns are then annualized by multiplying monthly values by 12. 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 $\hat{w}_t R_{t+1}$, the constant naïve target performance, and the time-varying bias-adjusted target). We then compute median differences (naïve minus actual, and bias-adjusted minus actual) and p-values from the Wilcoxon signed rank test (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) of the annualized monthly ex ante anticipated excess return (naïve = monthly equivalent of 2% annual, or adjusted) minus actual monthly performance.

Table 1. Median monthly bias, in annualized percentage points, showing bias reduction due to adjustment with six to eighteen countries (targeting 2% excess return per annum)

	Six countries	Nine countries	Twelve countries	Fifteen countries	Eighteen countries
36-month window					
Naïve	1.44 (8E-08)**	1.99 (3E-10)**	1.74 (1E-12)**	1.43 (1E-11)**	1.53 (4E-12)**
Adjusted	0.40 (0.72)	0.79 (0.29)	0.68 (0.05)*	0.43 (0.11)	0.63 (0.01)**
48-month window					
Naïve	1.84 (4E-07)**	1.96 (3E-10)**	1.69 (1E-11)**	1.95 (8E-14)**	1.81 (4E-15)**
Adjusted	0.57 (0.54)	0.65 (0.22)	0.47 (0.19)	0.77 (0.05)*	0.69 (0.02)**
60-month window					
Naïve	1.94 (5E-08)**	2.02 (1E-09)**	1.91 (1E-10)**	1.77 (7E-13)**	2.10 (5E-14)**
Adjusted	0.37 (0.11)	0.49 (0.33)	0.54 (0.26)	0.45 (0.14)	0.74 (0.08)
120-month window					
Naïve	2.50 (8E-04)**	2.89 (2E-09)**	2.13 (2E-06)**	2.08 (6E-09)**	1.95 (3E-10)**
Adjusted	0.06 (0.72)	0.35 (0.43)	0.26 (0.78)	0.18 (0.57)	0.02 (0.82)

The median difference between actual portfolio excess return and naïve respectively adjusted portfolios is computed, with the associated *p*-values from the Wilcoxon signed rank test in parentheses. Target excess return is 2% per annum. The median differences and *p*-values are shown by number of countries (6 to 18) and by the rolling window employed (36 to 120 months).
*(p<0.05), **(p<0.01)

In every case, the adjustment reduces the size of the bias, and decreases its significance as measured by the *p*-values. The naïve anticipated portfolio performance is statistically significantly different from that of the actual performance at all twenty combinations of horizon and sample size, indicating that the effect of estimation error is significant. The adjusted portfolio has nonsignificant bias in sixteen of the twenty cases, tending to show significant bias with more countries and a smaller window size. However, reflecting the asymptotic nature of the bias adjustment, as the estimation window is increased to 60 months or above, the bias in the adjusted portfolio becomes non-significant across all country portfolios.

D. Rebalancing Costs

One of the issues with forming portfolios ignoring estimation error is that they would theoretically require more frequent rebalancing due to noise in the estimated weights, especially if they are required to be within a given tracking error band. We analyze the rebalancing costs and size of rebalance for naïve and adjusted portfolios formed targeting 2% annual excess returns (where the adjusted portfolio targets an adjusted return of 2%). The analysis is done while varying the number of countries and using estimates from a fixed-length rolling window of data.

For both the naïve and adjusted portfolios the following rebalancing rule was applied. First, the portfolios are rebalanced if their respective standard deviation jumped from one period to the next. The “jump” that would trigger a rebalance was set to the monthly equivalent of 3.5% annual standard deviation. Second, the portfolios are rebalanced if the realized return is less than the monthly equivalent of ½ the target of 2% (i.e. 1%) annually. Since the monthly volatility can be quite significant leading to outlier portfolios (monthly standard deviations are in the range of 4-7%), another rebalancing rule was added. If the adjusted portfolio has a negative alpha ex-ante (as the adjustment would be so large, given the estimation noise, that it changes the sign), then the investor using the adjusted rule will choose to hold the equally-weighted portfolio (i.e. have zero perturbation weights). As the investor using the naïve portfolio has no ex-ante ability to judge whether their 2% annual target is achievable, they cannot implement a similar approach. If the portfolio was not rebalanced, the weights were not changed. If it was rebalanced, then the optimal weights for that point in time were computed and used. The investor using the naïve formulae would target 2% annualized, and the adjusted investor would compute the weights that would correspond to an adjusted portfolio return of 2%, using equation (13).

Table 2 shows the mean, median, standard deviation, information ratio, and *t*-statistic of the means of the actual portfolio annualized excess return (for the naïve and adjusted portfolios respectively). This is to see if the portfolio's excess return relative to the alpha target was higher under the naïve or the adjusted frontier. Then we computed the percent of changes required (i.e. percent rebalances) for the naïve and adjusted portfolio respectively. This is to see if using the naïve rule leads to more adjustments (hence higher costs using a linear transaction cost estimate) than using the adjustment. We also computed the average size of the rebalancing change required (the sum of the absolute changes in the weights, averaged over all months that involved rebalancing), again to see which of the naïve or the adjusted portfolios required larger rebalances (and therefore would be more costly).

Table 2. Annualized portfolio performance, standard deviation, and rebalance statistics, for the naïve and adjusted portfolios. All numbers except the *t*-statistic and the Information Ratio are in percentage points.

	Naïve rule			**Adjusted rule**		
	Six countries	Twelve countries	Eighteen countries	Six countries	Twelve countries	Eighteen countries
36 month window						
Mean excess return	0.31	0.46	0.29	0.63	1.09	0.61
Med excess return	-0.22	0.10	0.28	0.00	0.00	0.00
Standard deviation	2.38	1.44	1.38	2.62	2.86	3.13
IR	0.13	0.32	0.21	0.24	0.38	0.19
t-stat of mean	0.23	0.55	0.37	0.41	0.66	0.34
% rebalances	57.91	55.12	54.65	45.81	44.19	49.07
Mean rebalance size	1.32	1.23	1.72	1.85	3.19	5.39
48 month window						
Mean excess return	-0.03	0.16	0.14	0.16	0.28	0.58
Med excess return	-0.17	0.25	0.33	0.00	0.00	0.00
Standard deviation	2.50	1.48	1.42	2.62	2.65	3.28
IR	-0.01	0.11	0.10	0.06	0.11	0.18
t-stat of mean	-0.02	0.18	0.17	0.10	0.18	0.30
% rebalances	55.02	55.98	56.94	40.19	45.93	47.85
Mean rebalance size	1.30	1.06	1.37	1.60	2.54	4.64
60 month window						
Mean excess return	0.50	0.09	-0.05	-0.05	-0.21	-0.26
Med excess return	-0.06	0.32	-0.05	0.00	0.00	0.00
Standard deviation	2.71	1.71	1.58	2.39	2.94	3.29
IR	0.18	0.05	-0.03	-0.02	-0.07	-0.08
t-stat of mean	0.31	0.09	-0.05	-0.04	-0.12	-0.13
% rebalances	56.40	55.42	56.90	43.60	47.54	44.33
Mean rebalance size	1.13	1.08	1.33	1.45	2.66	3.99
120 month window						
Mean excess return	-0.16	-0.27	-0.22	-0.41	-0.33	-0.76
Med excess return	-0.75	0.05	-0.15	0.00	0.00	0.00
Standard deviation	3.22	2.17	1.90	2.45	2.68	2.67
IR	-0.05	-0.12	-0.12	-0.17	-0.12	-0.29
t-stat of mean	-0.08	-0.19	-0.18	-0.26	-0.19	-0.44
% rebalances	57.80	55.78	58.38	43.35	37.28	34.10
Mean rebalance size	0.97	0.99	1.17	1.23	2.43	3.43

Note that the zero median excess returns, for the adjusted rule across all windows, are due to the fact that more than half of the time periods show such a strong adjustment that the adjusted alpha becomes negative. In these cases the adjusted rule chooses to hold the equally-weighted portfolio, leading to excess returns of zero for these periods.

For the shorter windows of 36 and 48 months, the adjusted portfolios have a higher mean excess return than the naïve portfolios, with generally higher information ratio in spite of their generally higher standard deviations. The adjusted portfolio rebalances less frequently, but then adjusts by more on average when it does adjust. While there is one case (6 assets with the 48 month window) for which the adjusted portfolio seems less costly to maintain than the naïve portfolio (multiplying frequency and average size of rebalance), the opposite is more generally true.

As the windows get large, 60 and 120 months, the naïve portfolio does better in achieving a higher mean, though that comes at the expense of a negative (worse) median than the adjusted portfolio for 6 and for 18 countries. Standard deviations are higher with naïve than with adjusted portfolios for 6 countries, and are lower for 12 and for 18. The naïve portfolio gets rebalanced more frequently, with a smaller mean adjustment. Neither the naïve nor the adjusted portfolio achieves close to its target, which may be due to a lack of available alpha in the data, which occurs when country means are similar to one another (so that all portfolios have approximately the same mean as the equally-weighted benchmark).

We expanded the experiment above by increasing the available alpha, via scaling the mean returns of the assets to be between -2% and 2% annualized, equally-spaced by the number of countries (for example with 6 countries country number one would subtract the monthly equivalent of 2% annually from its each of its excess returns, and country number two would subtract 1.33% respectively, and so forth). Table 3 presents the results using these new scaled assets.

Table 3. Annualized portfolio performance, standard deviation, and rebalance statistics, for the naïve and adjusted portfolios, with added excess returns spreads. All numbers except the *t*-statistic and the Information Ratio are in percentage points.

	Naïve rule			**Adjusted rule**		
	Six countries	Twelve countries	Eighteen countries	Six countries	Twelve countries	Eighteen countries
36 month window						
Mean excess return	1.62	1.72	1.76	1.89	2.07	2.00
Med excess return	1.56	1.75	1.83	1.74	2.05	2.00
Standard deviation	0.83	0.69	0.61	1.09	0.90	0.70
IR	1.94	2.48	2.91	1.74	2.29	2.84
t-stat of mean	3.35	4.28	5.02	2.99	3.96	4.90
% rebalances	37.91	35.12	28.84	37.21	31.16	26.51
Mean rebalance size	0.25	0.42	0.68	0.33	0.60	0.84
48 month window						
Mean excess return	1.72	1.71	1.73	2.00	2.12	1.95
Med excess return	1.58	1.69	1.85	1.78	1.97	2.03
Standard deviation	0.85	0.67	0.56	1.05	0.92	0.65
IR	2.01	2.54	3.09	1.90	2.29	2.98
t-stat of mean	3.42	4.31	5.25	3.23	3.89	5.07
% rebalances	39.23	33.73	31.10	36.60	31.10	29.43
Mean rebalance size	0.18	0.32	0.43	0.23	0.42	0.52
60 month window						
Mean excess return	1.79	1.76	1.79	1.98	2.05	2.01
Med excess return	1.64	1.70	1.91	1.82	1.97	2.09
Standard deviation	0.81	0.66	0.54	0.93	0.80	0.62
IR	2.21	2.67	3.33	2.14	2.57	3.25
t-stat of mean	3.71	4.47	5.58	3.59	4.30	5.45
% rebalances	37.93	32.51	29.06	36.45	29.80	27.34
Mean rebalance size	0.15	0.26	0.35	0.18	0.32	0.42
120 month window						
Mean excess return	1.92	1.96	1.90	2.03	2.12	2.04
Med excess return	1.80	2.01	2.02	1.89	2.15	2.12
Standard deviation	0.78	0.60	0.48	0.83	0.66	0.52
IR	2.47	3.28	3.98	2.45	3.22	3.90
t-stat of mean	3.82	5.08	6.15	3.78	4.97	6.03
% rebalances	35.55	26.88	26.30	34.68	25.72	25.14
Mean rebalance size	0.07	0.14	0.17	0.08	0.16	0.20

The results in Table 3 show that, with increased alpha added to the data, the adjusted portfolio outperforms the naïve portfolio by having higher mean and higher median across all assets and all windows, although the standard deviations are also higher (which may be necessary in order to come closer to the target alpha of 2%). The rebalances are less frequent and larger on average for the adjusted portfolio relative to the naïve portfolio, and the overall impact of transaction costs (taken as percent rebalances times the mean rebalance size as a rough measure) is generally slightly higher for the adjusted portfolio than for the naïve portfolio.

4. Summary and Discussion

We have derived adjustments to reduce and asymptotically eliminate the bias in performance of tracking-error optimized 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. When this approximation is effective, the adjusted efficient frontier more realistically represents the actual performance of a portfolio that seeks a target mean performance, while minimizing the tracking error in a mean-variance framework using 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).

Our derivations are based on Taylor-series expansions and are therefore not finite-sample exact. The approximations will tend to degrade when T is smaller (because the statistical estimates will tend to be farther from the true parameter values about which they are expanded) and when n is larger (because the optimization procedure may tend to choose a combination of assets with high estimation error).

Our empirical results illustrated the economic importance of the size of the bias adjustment for international portfolios, the efficacy of the bias adjustment using bootstrap simulations while relaxing the normal approximation, the statistical significance of the bias adjustment, and the slightly higher rebalancing costs incurred while using the adjustment to improve portfolio performance.

Appendix: Proofs

PROOF OF THEOREM 1. The expected excess return of the estimated portfolio, using iterated conditional expectations and the fact that R_{T+1} is independent of the previous observations, may be written as $E\left(\hat{w}'R_{T+1}\right)=E\left(\hat{w}'\mu\right)=E\left[\hat{w}'\hat{\mu}-\hat{w}'\left(\hat{\mu}-\mu\right)\right]=\alpha_0-E\left[\hat{w}'\left(\hat{\mu}-\mu\right)\right]=\alpha_0-E\left(\hat{w}'\delta\right)$. Using the method of statistical differentials, the delta-method approximation to this expectation is $E_\Delta\left(\hat{w}'R_{T+1}\right)=\alpha_0-E_\Delta\left(\hat{w}'\delta\right)=\alpha_0-E\left[S_1\left(\hat{w}',\delta\right)\delta\right]$. Substituting for $\hat{w}$, we find

$$
\begin{aligned}
E_\Delta\left(\hat{w}'R_{T+1}\right)&=\alpha_0-E\left\{\delta'S_1\left[\hat{V}^{-1}\left(\mathbf{1}\quad\hat{\mu}\right)\hat{B}\begin{pmatrix}0\\ \alpha_0\end{pmatrix},\delta\right]\right\}\\
&=\alpha_0-E\left\{\delta'\left[S_1\left(\hat{V}^{-1},\delta\right)\left(\mathbf{1}\quad\mu\right)B\right]\right\}\begin{pmatrix}0\\ \alpha_0\end{pmatrix}\\
&\quad-E\left\{\delta'\left(V^{-1}S_1\left[\left(\mathbf{1}\quad\hat{\mu}\right),\delta\right]B\right)\right\}\begin{pmatrix}0\\ \alpha_0\end{pmatrix}\\
&\quad-E\left\{\delta'\left[V^{-1}\left(\mathbf{1}\quad\mu\right)S_1\left(\hat{B},\delta\right)\right]\right\}\begin{pmatrix}0\\ \alpha_0\end{pmatrix}
\end{aligned}
\tag{19}
$$

Next, we note that $S_1\left(\hat{V}^{-1},\delta\right)=0$ because δ does not appear in the covariance estimate, $S_1\left[\left(\mathbf{1}\quad\hat{\mu}\right),\delta\right]=\left(\mathbf{0}\quad\delta\right)$ by definition of δ, and

$$
S_1\left(\hat{B},\delta\right)=-B\left[\begin{pmatrix}0\\1\end{pmatrix}\delta'V^{-1}\left(\mathbf{1}\quad\mu\right)+\left(\mathbf{1}\quad\mu\right)'V^{-1}\delta\left(0\quad 1\right)\right]B \tag{20}
$$

from Lemma 4 of Siegel and Woodgate (Online Appendix, 2007). This leads to

$$
\begin{aligned}
E_{\Delta}\left(\hat{w}'R_{T+1}\right) &= \alpha_0 - E\left\{\delta'\left[V^{-1}\left(\mathbf{0} \quad \delta\right)B\right]\right\}\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\
&\quad + E\left\{\delta'\left[V^{-1}\left(\mathbf{1} \quad \mu\right)B\left[\begin{pmatrix}0\\1\end{pmatrix}\delta'V^{-1}\left(\mathbf{1} \quad \mu\right) + \left(\mathbf{1} \quad \mu\right)'V^{-1}\delta\left(0 \quad 1\right)\right]B\right]\right\}\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\
&= \alpha_0 - E\left\{\left[\left(0 \quad \delta'V^{-1}\delta\right)B\right]\right\}\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\
&\quad + E\left\{\delta'V^{-1}\left(\mathbf{1} \quad \mu\right)B\begin{pmatrix}0\\1\end{pmatrix}\delta'V^{-1}\left(\mathbf{1} \quad \mu\right)B\right\}\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\
&\quad + E\left\{\delta'V^{-1}\left(\mathbf{1} \quad \mu\right)B\left(\mathbf{1} \quad \mu\right)'V^{-1}\delta\left(0 \quad 1\right)B\right\}\begin{pmatrix}0\\ \alpha_0\end{pmatrix}
\end{aligned}
\tag{21}
$$

Next, we transpose an embedded bracketed scalar to find

$$
\begin{aligned}
E_{\Delta}\left(\hat{w}'R_{T+1}\right) &= \alpha_0 - E\left\{\left[\left(0 \quad \delta'V^{-1}\delta\right)B\right]\right\}\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\
&\quad + E\left\{\delta'V^{-1}\left(\mathbf{1} \quad \mu\right)B\begin{pmatrix}0\\1\end{pmatrix}\left[\delta'V^{-1}\left(\mathbf{1} \quad \mu\right)B\begin{pmatrix}0\\ \alpha_0\end{pmatrix}\right]\right\} \\
&\quad + E\left\{\delta'V^{-1}\left(\mathbf{1} \quad \mu\right)B\left(\mathbf{1} \quad \mu\right)'V^{-1}\delta\left(0 \quad 1\right)B\right\}\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\
&= \alpha_0 - E\left\{\left[\left(0 \quad \delta'V^{-1}\delta\right)B\right]\right\}\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\
&\quad + E\left\{\delta'V^{-1}\left(\mathbf{1} \quad \mu\right)B\begin{pmatrix}0\\1\end{pmatrix}\left(0 \quad \alpha_0\right)B\left(\mathbf{1} \quad \mu\right)'V^{-1}\delta\right\} \\
&\quad + E\left\{\delta'V^{-1}\left(\mathbf{1} \quad \mu\right)B\left(\mathbf{1} \quad \mu\right)'V^{-1}\delta\left(0 \quad 1\right)B\right\}\begin{pmatrix}0\\ \alpha_0\end{pmatrix}
\end{aligned}
\tag{22}
$$

Now use the fact that $E\left(\delta'Q\delta\right) = tr\left(QV\right)/T$ for any symmetric $n \times n$ matrix Q (see, e.g., page 14 of Seber (1984)) to establish

$$
\begin{aligned}
E_{\Delta}\left(\hat{w}'R_{T+1}\right) = \alpha_0 - &\left\{\left[\begin{pmatrix}0 & n/T\end{pmatrix}B\right]\right\}\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\
&+\frac{1}{T}tr\left\{V^{-1}\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}B\begin{pmatrix}0\\ 1\end{pmatrix}\begin{pmatrix}0 & \alpha_0\end{pmatrix}B\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}'\right\} \\
&+\frac{1}{T}tr\left\{V^{-1}\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}B\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}'\right\}\begin{pmatrix}0 & 1\end{pmatrix}B\begin{pmatrix}0\\ \alpha_0\end{pmatrix}
\end{aligned}
\tag{23}
$$

Evaluating, while using commutativity of matrices within the trace operator, we find

$$
\begin{aligned}
E_{\Delta}\left(\hat{w}'R_{T+1}\right) &= \alpha_0 - \frac{n\alpha_0 B_{22}}{T} + \frac{1}{T}tr\left\{\begin{pmatrix}0 & \alpha_0\end{pmatrix}B\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}'V^{-1}\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}B\begin{pmatrix}0\\ 1\end{pmatrix}\right\} \\
&\quad+\frac{1}{T}tr\left\{\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}'V^{-1}\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}B\right\}\begin{pmatrix}0 & 1\end{pmatrix}B\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\
&= \alpha_0 - \frac{n\alpha_0 B_{22}}{T} + \frac{1}{T}\left\{\begin{pmatrix}0 & \alpha_0\end{pmatrix}BB^{-1}B\begin{pmatrix}0\\ 1\end{pmatrix}\right\} + \frac{1}{T}tr\left\{B^{-1}B\right\}\begin{pmatrix}0 & 1\end{pmatrix}B\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\
&= \alpha_0 - \frac{n\alpha_0 B_{22}}{T} + \frac{1}{T}\left\{\begin{pmatrix}0 & \alpha_0\end{pmatrix}B\begin{pmatrix}0\\ 1\end{pmatrix}\right\} + \frac{1}{T}tr\left\{I_2\right\}\begin{pmatrix}0 & 1\end{pmatrix}B\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\
&= \alpha_0 - \frac{n\alpha_0 B_{22}}{T} + \frac{\alpha_0 B_{22}}{T} + \frac{2\alpha_0 B_{22}}{T} \\
&= \alpha_0\left(1 - \frac{n-3}{T}B_{22}\right)
\end{aligned}
\tag{24}
$$

which completes the proof of Equation (12). To establish (14) we expand $E_{\Delta}\left(\hat{\alpha}_{adjusted}\right)$ to second order as follows:

$$
\begin{aligned}
E_{\Delta}\left(\hat{\alpha}_{adjusted}\right) &= E_{\Delta}\left[\alpha_0\left(1-\frac{n-3}{T}\hat{B}_{22}\right)\right] = \alpha_0\left(1-\frac{n-3}{T}B_{22}\right) + E\left[\alpha_0\frac{n-3}{T}S_2\left(\hat{B}_{22}\right)\right] \\
&= E_{\Delta}\left(\hat{w}'R_{T+1}\right) + O\left(\frac{1}{T^2}\right)
\end{aligned}
\tag{25}
$$

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 of Siegel and Woodgate (Online Appendix, 2007). □

Proof of Theorem 2. Using the method of statistical differentials, the delta-method approximation to the expectation of the naïve tracking-error is $E_\Delta\left(\hat{w}'\hat{V}\hat{w}\right) = w'Vw + E\left[S_2\left(\hat{w}'\hat{V}\hat{w}\right)\right]$. We substitute, using the definitions of $\hat{w}$ and of σ_0^2, to find:

$$\begin{aligned}
E_\Delta\left(\hat{w}'\hat{V}\hat{w}\right) &= \sigma_0^2 + E\left\{S_2\left[\begin{pmatrix}0 & \alpha_0\end{pmatrix}\hat{B}\begin{pmatrix}\mathbf{1} & \hat{\mu}\end{pmatrix}'\hat{V}^{-1}\hat{V}\hat{V}^{-1}\begin{pmatrix}\mathbf{1} & \hat{\mu}\end{pmatrix}\hat{B}\begin{pmatrix}0\\ \alpha_0\end{pmatrix}\right]\right\} \\
&= \sigma_0^2 + E\left\{S_2\left[\begin{pmatrix}0 & \alpha_0\end{pmatrix}\hat{B}\begin{pmatrix}\mathbf{1} & \hat{\mu}\end{pmatrix}'\hat{V}^{-1}\begin{pmatrix}\mathbf{1} & \hat{\mu}\end{pmatrix}\hat{B}\begin{pmatrix}0\\ \alpha_0\end{pmatrix}\right]\right\} \\
&= \sigma_0^2 + E\left\{S_2\left[\begin{pmatrix}0 & \alpha_0\end{pmatrix}\hat{B}\hat{B}^{-1}\hat{B}\begin{pmatrix}0\\ \alpha_0\end{pmatrix}\right]\right\} = \sigma_0^2 + \begin{pmatrix}0 & \alpha_0\end{pmatrix}E\left[S_2\left(\hat{B}\right)\right]\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\
&= \sigma_0^2 + \begin{pmatrix}0 & \alpha_0\end{pmatrix}E\left[S_2\left(\hat{B},\delta\right) + S_2\left(\hat{B},\varepsilon\right)\right]\begin{pmatrix}0\\ \alpha_0\end{pmatrix}
\end{aligned} \tag{26}$$

We use Equations (EC25) and (EC26) from Lemma 7 of Siegel and Woodgate (Online Appendix, 2007) together with the fact that $\sigma_0^2 = \alpha_0^2 B_{22}$ to obtain

$$\begin{aligned}
E_\Delta\left(\hat{w}'\hat{V}\hat{w}\right) &= \sigma_0^2 + \begin{pmatrix}0 & \alpha_0\end{pmatrix}\left[\frac{1}{T}B_{22}B - \frac{n-4}{T}B\begin{pmatrix}0\\ 1\end{pmatrix}\begin{pmatrix}0 & 1\end{pmatrix}B - \frac{n-2}{T-1}B\right]\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\
&= \sigma_0^2 + \left[\frac{1}{T}B_{22}\left(\alpha_0^2 B_{22}\right) - \frac{n-4}{T}\left(\alpha_0 B_{22}\right)^2 - \frac{n-2}{T-1}\alpha_0^2 B_{22}\right] \\
&= \sigma_0^2 + \sigma_0^2\left(\frac{1}{T}B_{22} - \frac{n-4}{T}B_{22} - \frac{n-2}{T-1}\right) = \sigma_0^2\left(1 - \frac{n-5}{T}B_{22} - \frac{n-2}{T-1}\right) \\
&= \sigma_0^2\left(1 - \frac{n-5}{T}B_{22} - \frac{n-2}{T}\right) + O\left(\frac{1}{T^2}\right)
\end{aligned} \tag{27}$$

which establishes Equation (15).

The actual tracking error may be approximated as follows

$$\begin{aligned} Var\left(\hat{w}'R_{T+1}\right) &\cong E_{\Delta}\left[\left(\hat{w}'R_{T+1}\right)^{2}\right]-\left[E_{\Delta}\left(\hat{w}'R_{T+1}\right)\right]^{2}=E_{\Delta}\left(\hat{w}'V\hat{w}+\hat{w}'\mu\mu'\hat{w}\right)-\left[E_{\Delta}\left(\hat{w}'\mu\right)\right]^{2} \\ &=E_{\Delta}\left(\hat{w}'\hat{V}\hat{w}\right)-E_{\Delta}\left(\hat{w}'\varepsilon\hat{w}\right)+E_{\Delta}\left(\hat{w}'\mu\mu'\hat{w}\right)-\left[E_{\Delta}\left(\hat{w}'\mu\right)\right]^{2} \end{aligned} \tag{28}$$

In (28) we recognize the first of the four final terms, $E_{\Delta}\left(\hat{w}'\hat{V}\hat{w}\right)$, which has already been evaluated in (15), and use the fact that $E\left(\delta\varepsilon\right)=0$ while expanding the second term of (28) as follows:

$$E_{\Delta}\left(\hat{w}'\varepsilon\hat{w}\right)=E\left[S_{2}\left(\hat{w}'\varepsilon\hat{w}\right)\right]=2w'E\left[\varepsilon S_{1}\left(\hat{w},\varepsilon\right)\right] \tag{29}$$

Substituting for $\hat{w}$ and for w, we expand to find:

$$\begin{aligned} E_{\Delta}\left(\hat{w}'\varepsilon\hat{w}\right) &= 2\begin{pmatrix}0 & \alpha_0\end{pmatrix}B\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}'V^{-1}E\left\{\varepsilon S_1\left[\hat{V}^{-1}\begin{pmatrix}\mathbf{1} & \hat{\mu}\end{pmatrix}\hat{B},\varepsilon\right]\right\}\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\ &= 2\begin{pmatrix}0 & \alpha_0\end{pmatrix}B\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}'V^{-1}E\left[\varepsilon S_1\left(\hat{V}^{-1},\varepsilon\right)\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}B+\varepsilon V^{-1}\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}S_1\left(\hat{B},\varepsilon\right)\right]\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \end{aligned} \tag{30}$$

We use Equations (EC8) and (EC13) from Lemmas 3 and 4 of Siegel and Woodgate (Online Appendix, 2007) to obtain

$$\begin{aligned} &E_{\Delta}\left(\hat{w}'\varepsilon\hat{w}\right) \\ &= 2\begin{pmatrix}0 & \alpha_0\end{pmatrix}B\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}'V^{-1}E\left[-\varepsilon V^{-1}\varepsilon V^{-1}\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}B+\varepsilon V^{-1}\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}B\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}'V^{-1}\varepsilon V^{-1}\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}B\right]\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\ &= 2\begin{pmatrix}0 & \alpha_0\end{pmatrix}B\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}'V^{-1}E\left[-\varepsilon V^{-1}\varepsilon+\varepsilon V^{-1}\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}B\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}'V^{-1}\varepsilon\right]V^{-1}\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}B\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\ &= 2\begin{pmatrix}0 & \alpha_0\end{pmatrix}B\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}'V^{-1}E\left\{\varepsilon\left[-V^{-1}+V^{-1}\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}B\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}'V^{-1}\right]\varepsilon\right\}V^{-1}\begin{pmatrix}\mathbf{1} & \mu\end{pmatrix}B\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \end{aligned} \tag{31}$$

The expectation in (31) may be evaluated using Lemma 1 of Siegel and Woodgate (Online Appendix, 2007) which follows from Theorem 3.1 (iii) of Haff (1979), and also using (EC22) from this Online Appendix to evaluate a trace and obtain

$$\begin{aligned}
&E\left\{\varepsilon\left[-V^{-1}+V^{-1}(\mathbf{1}\quad\mu)B(\mathbf{1}\quad\mu)'V^{-1}\right]\varepsilon\right\}\\
&=\left\{\left[-V+(\mathbf{1}\quad\mu)B(\mathbf{1}\quad\mu)'\right]+V\,tr\left[-I_n+V^{-1}(\mathbf{1}\quad\mu)B(\mathbf{1}\quad\mu)'\right]\right\}/(T-1)\\
&=\left[-V+(\mathbf{1}\quad\mu)B(\mathbf{1}\quad\mu)'-(n-2)V\right]/(T-1)\\
&=\left[(\mathbf{1}\quad\mu)B(\mathbf{1}\quad\mu)'-(n-1)V\right]/(T-1)
\end{aligned}\tag{32}$$

Using this expectation (32) in the formula (31) for $E_\Delta\left(\hat{w}'\varepsilon\hat{w}\right)$, the evaluation of the second term of (28) is completed as follows:

$$\begin{aligned}
E_\Delta\left(\hat{w}'\varepsilon\hat{w}\right)&=\frac{2}{T-1}(0\quad\alpha_0)B(\mathbf{1}\quad\mu)'V^{-1}\left[(\mathbf{1}\quad\mu)B(\mathbf{1}\quad\mu)'-(n-1)V\right]V^{-1}(\mathbf{1}\quad\mu)B\begin{pmatrix}0\\ \alpha_0\end{pmatrix}\\
&=\frac{2}{T-1}(0\quad\alpha_0)B\left[B^{-1}-(n-1)B^{-1}\right]B\begin{pmatrix}0\\ \alpha_0\end{pmatrix}\\
&=-\frac{2(n-2)}{T-1}(0\quad\alpha_0)B\begin{pmatrix}0\\ \alpha_0\end{pmatrix}=-\frac{2(n-2)}{T-1}\alpha_0^2B_{22}=-\frac{2(n-2)}{T-1}\sigma_0^2
\end{aligned}\tag{33}$$

We evaluate the third term of (28) as follows:

$$\begin{aligned}
E_\Delta\left(\hat{w}'\mu\mu'\hat{w}\right)&=E_\Delta\left\{\left[\hat{w}'\hat{\mu}-\hat{w}'\left(\hat{\mu}-\mu\right)\right]^2\right\}=E_\Delta\left[\left(\alpha_0-\hat{w}'\delta\right)^2\right]\\
&=\alpha_0^2-2\alpha_0E_\Delta\left[\hat{w}'\delta\right]+E_\Delta\left[\left(\hat{w}'\delta\right)^2\right]\\
&=\alpha_0^2-2\alpha_0E\left[\delta'S_1\left(\hat{w},\delta\right)\right]+E\left\{S_2\left[\left(\hat{w}'\delta\right)^2,\delta\right]\right\}\\
&=\alpha_0^2-2\alpha_0E\left[\delta'S_1\left(\hat{w},\delta\right)\right]+\frac{\sigma_0^2}{T}
\end{aligned}\tag{34}$$

where the final term of (34) was evaluated using Lemma 1 of Siegel and Woodgate (Online Appendix, 2007) along with commutativity of matrix multiplication within the trace function. To complete the evaluation of (34) we will also need

$$\begin{aligned} S_1(\hat{w},\delta) &= S_1\left[\hat{V}^{-1}(\mathbf{1}\quad\hat{\mu})\hat{B}\begin{pmatrix}0\\ \alpha_0\end{pmatrix},\delta\right] = V^{-1}\left\{S_1\left[(\mathbf{1}\quad\hat{\mu}),\delta\right]B\begin{pmatrix}0\\ \alpha_0\end{pmatrix} + (\mathbf{1}\quad\mu)S_1(\hat{B},\delta)\begin{pmatrix}0\\ \alpha_0\end{pmatrix}\right\} \\ &= V^{-1}\left\{(\mathbf{0}\quad\delta)B\begin{pmatrix}0\\ \alpha_0\end{pmatrix} + (\mathbf{1}\quad\mu)S_1(\hat{B},\delta)\begin{pmatrix}0\\ \alpha_0\end{pmatrix}\right\} \end{aligned} \tag{35}$$

Using EC11 of Siegel and Woodgate (Online Appendix, 2007) we find

$$S_1(\hat{w},\delta) = V^{-1}\left\{(\mathbf{0}\quad\delta)B\begin{pmatrix}0\\ \alpha_0\end{pmatrix} - (\mathbf{1}\quad\mu)B\left[\begin{pmatrix}0\\ 1\end{pmatrix}\delta'V^{-1}(\mathbf{1}\quad\mu) + (\mathbf{1}\quad\mu)'V^{-1}\delta(0\quad 1)\right]B\begin{pmatrix}0\\ \alpha_0\end{pmatrix}\right\} \tag{36}$$

and therefore

$$\begin{aligned} E\left[\delta' S_1(\hat{w},\delta)\right] &= E\left[\delta'V^{-1}(\mathbf{0}\quad\delta)\right]B\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\ &\quad - E\left\{\delta'V^{-1}(\mathbf{1}\quad\mu)B\begin{pmatrix}0\\ 1\end{pmatrix}\left[\delta'V^{-1}(\mathbf{1}\quad\mu)B\begin{pmatrix}0\\ \alpha_0\end{pmatrix}\right]\right\} \\ &\quad - E\left[\delta'V^{-1}(\mathbf{1}\quad\mu)B(\mathbf{1}\quad\mu)'V^{-1}\delta\right](0\quad 1)B\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \\ &= \alpha_0 B_{22}E\left(\delta'V^{-1}\delta\right) \\ &\quad - E\left[\delta'V^{-1}(\mathbf{1}\quad\mu)B\begin{pmatrix}0\\ 1\end{pmatrix}(0\quad\alpha_0)B(\mathbf{1}\quad\mu)'V^{-1}\delta\right] \\ &\quad - E\left[\delta'V^{-1}(\mathbf{1}\quad\mu)B(\mathbf{1}\quad\mu)'V^{-1}\delta\right](0\quad 1)B\begin{pmatrix}0\\ \alpha_0\end{pmatrix} \end{aligned} \tag{37}$$

where we have transposing an embedded scalar in the second term. Applying results from Siegel and Woodgate (Online Appendix, 2007; Lemma 1 to evaluate the expectations and EC22 for a trace), and using commutativity of matrices within the trace function we find

$$E\left[\delta' S_1(\hat{w},\delta)\right] = \frac{n\alpha_0 B_{22}}{T} - \frac{\alpha_0 B_{22}}{T} - \frac{2\alpha_0 B_{22}}{T} = (n-3)\frac{\alpha_0 B_{22}}{T} \tag{38}$$

Finally, using this in (34) and substituting $\sigma_0^2 = \alpha_0^2 B_{22}$ we find

$$E_\Delta\left(\hat{w}'\mu\mu'\hat{w}\right) = \alpha_0^2 - (2n-7)\frac{\sigma_0^2}{T} \tag{39}$$

which completes the evaluation of the third term of (28). Putting (15), (33), (39), and (12) together, we can evaluate (28) as follows:

$$\begin{aligned}
Var\left(\hat{w}'R_{T+1}\right) &\cong E_\Delta\left(\hat{w}'\hat{V}\hat{w}\right) - E_\Delta\left(\hat{w}'\varepsilon\hat{w}\right) + E_\Delta\left(\hat{w}'\mu\mu'\hat{w}\right) - \left[E_\Delta\left(\hat{w}'\mu\right)\right]^2 \\
&= \left[\sigma_0^2\left(1 - \frac{n-5}{T}B_{22} - \frac{n-2}{T}\right) + O\left(\frac{1}{T^2}\right)\right] + \sigma_0^2\frac{2(n-2)}{T-1} \\
&\quad + \left[\alpha_0^2 - (2n-7)\frac{\sigma_0^2}{T}\right] - \left[\alpha_0\left(1 - \frac{n-3}{T}B_{22}\right)\right]^2 \\
&= \sigma_0^2\left(1 - \frac{n-5}{T}B_{22} - \frac{n-2}{T}\right) + \sigma_0^2\frac{2(n-2)}{T} + \frac{\sigma_0^2}{T} + O\left(\frac{1}{T^2}\right) \\
&= \sigma_0^2\left(1 - \frac{n-5}{T}B_{22} + \frac{n-1}{T}\right) + O\left(\frac{1}{T^2}\right)
\end{aligned} \tag{40}$$

which, together with (15), establishes (16).

To establish (18) we expand $E_\Delta\left(\hat{\sigma}^2_{adjusted}\right)$ to second order as follows:

$$\begin{aligned}
E_\Delta\left(\hat{\sigma}^2_{adjusted}\right) &= E_\Delta\left\{\left[\left(1 + \frac{n-1.5}{T}\right)\hat{\sigma}_0\right]^2\right\} = E_\Delta\left[\hat{\sigma}_0^2\left(1 + \frac{2n-3}{T}\right)\right] + O\left(\frac{1}{T^2}\right) \\
&= E_\Delta\left(\hat{\sigma}_0^2\right) + \frac{2n-3}{T}\sigma_0^2 + O\left(\frac{1}{T^2}\right)
\end{aligned} \tag{41}$$

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 of Siegel and Woodgate (Online Appendix, 2007). Finally, use (16) to see that

$$E_{\Delta}\left(\hat{\sigma}^2_{adjusted}\right) = Var_{\Delta}\left(\hat{w}' R_{T+1}\right) + O\left(\frac{1}{T^2}\right). \tag{42}$$

completing the proof. □

References

Chan, Louis K.C., Jason Karceski, and Josef Lakonishok, "On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model," 1999, NBER Working Paper No. 7039.

Dessislava, P., "Handling Parameter Uncertainty in Portfolio Risk Minimization," **The Journal of Portfolio Management** Summer 2006, Vol. 32, No. 4: pp. 70-78.

Focardi, Sergio M., Petter N. Kolm , and Frank J. Fabozzi, "New Kids on the block," **The Journal of Portfolio Management** 2004, 30th Anniversary, Vol. 30, No. 5: pp. 42-54.

Siegel, Andrew F., and A. Woodgate, "Performance of Portfolios Optimized with Estimation Error," **Management Science**, June 2007, Vol. 53, No. 6: pp. 1005-1015.

Siegel, Andrew F., and A. Woodgate, Online Appendix to "Performance of Portfolios Optimized with Estimation Error," **Management Science**, June 2007, http://mansci.journal.informs.org/cgi/data/53/6/1005/DC1/1.