JOIM
www.joim.com

JOURNAL OF INVESTMENT MANAGEMENT, Vol. 18, No. 3, (2020), pp. 92–108

CORRELATION SHRINKAGE: IMPLICATIONS FOR RISK FORECASTING

Jose Menchero[a] *and Peng Li*[a]

In this article, we study the impact of shrinking sample correlations toward zero. We find that while such shrinkage may be beneficial from a portfolio-construction perspective, there is virtually no benefit in terms of the accuracy of risk forecasts. In fact, we show that correlation shrinkage typically increases the errors in risk forecasts, sometimes by a large margin. Hence, we conclude that for purposes of estimating portfolio risk, the estimated correlations should not deviate significantly from the sample correlation.

1 Introduction

Asset covariance matrices are used for two primary purposes in quantitative investment management. The first is for predicting portfolio volatility, which plays a central role within the risk-management function. The second application occurs in mean–variance optimization (MVO), which represents a technique pioneered by Markowitz (1952) for constructing portfolios with maximum expected return per unit of risk.

The simplest way to estimate the asset covariance matrix is to compute the *sample covariance matrix*. The elements of this matrix are obtained by taking the time series of asset returns and applying the textbook definition of sample covariance.

Unfortunately, the sample covariance matrix suffers from an "Achilles heel." In particular, it is not robust for MVO applications. To understand the origin of this problem, we must borrow an important fact from linear algebra: if the number of assets is greater than the number of time periods T, then the sample covariance matrix is *rank deficient*, which implies the existence of zero-volatility portfolios (composed entirely of risky assets). Invariably, portfolio optimizers will latch onto these spurious "riskless" portfolios, creating what appears to be portfolios with infinite risk-adjusted performance (*ex ante*).

Note that $N > T$ is in fact the typical regime for many financial applications. For instance, global equity indices commonly contain several thousand stocks. If weekly return observations are used to estimate the sample covariance matrix, then several *decades* of return history would be

[a]Bloomberg.

required to avoid spurious "riskless" portfolios. Such long histories, of course, are typically not available.

Even when $T > N$, which precludes the possibility of "riskless" portfolios, the sample covariance matrix may still not be reliable for MVO purposes. In a recent paper, Menchero and Ji (2019) showed that using the sample covariance matrix for constructing MVO portfolios leads to poor risk-adjusted performance (out-of-sample), large turnover, and high leverage. Another well-known problem is that the sample covariance matrix tends to systematically under-forecast the risk of these optimized portfolios—often by a wide margin. All of these pitfalls are exacerbated as the covariance matrix becomes increasingly "ill conditioned," which occurs when the half-life (HL) parameter used to estimate correlations becomes small relative to the number of assets.

A statistical technique known as "shrinkage" has been applied to mitigate such problems. Shrinkage involves blending a sample estimate, which may be noisy but unbiased, with a "target" value, which may be biased but has low noise. The intuition behind shrinkage is that by optimally trading off between noise and bias, we may attain an estimate with lower mean-squared error than either the sample or the target. An early example of this technique was described by Vasicek (1973), who showed that shrinking market betas toward 1 led to a reduction in the mean-squared error of estimated betas. Another example—this one within a portfolio-construction context—was from Ledoit and Wolf (2003). They formed minimum-volatility fully invested equity portfolios using asset covariance matrices constructed using a variety of techniques. They found that blending the sample covariance matrix with an appropriate target was an effective tool for producing portfolios with lower out-of-sample volatility. The two shrinkage targets they considered were the identity matrix (which assumes uncorrelated stocks) and the covariance matrix from a one-factor market model, as described by Sharpe (1963).

In this paper, we study the implications of shrinkage for risk-forecasting purposes. In particular, we consider shrinking correlations toward the identity matrix, which reflects the view that all assets are uncorrelated. We focus on this shrinkage target for two reasons. First, it represents an *explicit* shrinkage target used in many applications. For instance, as described by Menchero (2010), in equity multi-factor risk models industry and style factors are represented by dollar-neutral factor-mimicking portfolios. The average correlation among these factor portfolios is close to zero, making the identity matrix a natural shrinkage target. The second reason we consider the identity matrix is that it represents an *implicit* shrinkage target in any model that tends to systematically under-estimate correlations. Multi-asset-class factor models, such as those described by Shepard (2007), typically fall into this second category. Hence, the results of this paper have important implications for risk forecasting of real-world portfolios.

Our main conclusions are as follows. First, we show that from a risk-forecasting perspective, the sample correlation is *nearly optimal*. Second, we find that for any realistic choice of correlation HL parameter, optimal shrinkage leads to *immaterial* improvements in the accuracy of risk forecasts. Finally, we demonstrate that excessive shrinkage of correlations may lead to large errors in risk forecasts.

Benefits of shrinkage for portfolio construction

Before discussing the *pitfalls* of correlation shrinkage from a risk-forecasting point of view, we first consider the *benefits* of shrinkage for

portfolio-construction purposes. In particular, we present a simple example that illustrates how shrinkage can improve the risk-adjusted performance of MVO portfolios.

It is convenient to write the sample asset covariance matrix $\hat{\mathbf{\Omega}}$ as a product of volatilities and correlations,

$$\hat{\mathbf{\Omega}} = \hat{\mathbf{V}}\hat{\mathbf{C}}\hat{\mathbf{V}} \tag{1}$$

where $\hat{\mathbf{V}}$ is a diagonal matrix whose elements are the predicted volatilities of each asset, and $\hat{\mathbf{C}}$ is the sample asset-correlation matrix. Volatilities and correlations are typically estimated using exponentially weighted moving averages (EWMA), which is characterized by an HL parameter. In the current example, both $\hat{\mathbf{V}}$ and $\hat{\mathbf{C}}$ are estimated using the same HL parameter, although in other cases it may prove useful to apply different HL parameters to the volatilities and correlations.

Next, we select the identity matrix $\mathbf{I}$ as our shrinkage target, which has ones along the diagonal and zeros on the off-diagonal elements. This represents the view that all assets are uncorrelated. Hence, the shrunk correlation matrix is

$$\tilde{\mathbf{C}}_\lambda = (1-\lambda)\hat{\mathbf{C}} + \lambda\mathbf{I}, \tag{2}$$

where λ is the shrinkage intensity, which varies from 0 to 1. Note that $\lambda = 0$ recovers the sample correlation matrix, whereas $\lambda = 1$ fully imposes the structure of uncorrelated assets. Using the shrunk correlation matrix computed as in Equation (2), the covariance matrix becomes

$$\tilde{\mathbf{\Omega}}_\lambda = \hat{\mathbf{V}}\tilde{\mathbf{C}}_\lambda\hat{\mathbf{V}}. \tag{3}$$

Note that the HL parameter in Equation (3) is suppressed for notational simplicity.

Suppose that we wish to construct the optimal portfolio under the view that all stocks have the same expected return. In this case, all fully invested portfolios have the same expected return. Hence, the maximum Sharpe ratio portfolio (*ex ante*) is the fully invested portfolio with minimum volatility. To evaluate the efficacy of shrinkage from an MVO perspective, we study the out-of-sample volatility of these optimized portfolios.

With the asset covariance matrix constructed as in Equation (3), we proceed to construct the minimum-volatility portfolio using MVO. As shown by Grinold and Kahn (2000), the analytic formula for the minimum-volatility fully invested portfolio ($\mathbf{w}_\lambda$) is given by

$$\mathbf{w}_\lambda = \frac{\tilde{\mathbf{\Omega}}_\lambda^{-1}\mathbf{1}}{\mathbf{1}'\tilde{\mathbf{\Omega}}_\lambda^{-1}\mathbf{1}}. \tag{4}$$

where $\mathbf{1}$ is an $N \times 1$ vector with 1 for every entry, and $\tilde{\mathbf{\Omega}}_\lambda^{-1}$ is the inverse of the asset covariance matrix.

As a concrete example, we select a universe of the 100 largest US equities as of 31-Mar-2016, with complete daily return history going back to the start of January 1999. We use the first two years of our sample as the "burn-in" period (for computing the initial asset covariance matrix); this produces an out-of-sample testing period spanning 27-Dec-2000 to 31-Mar-2016.

We directly estimate a family of sample covariance matrices using EWMA with an HL parameter that varies from 10 days to 150 days. We apply the same HL to both volatilities and correlations. At the start of each day, we form the optimal portfolio using Equation (4) and observe the out-of-sample portfolio return over the next day. At the end of each day, we update our asset covariance matrix and rebalance the portfolio.

In Figure 1, we plot the out-of-sample volatility (annualized) of the optimized portfolios as a function of the HL parameter. We consider three values of shrinkage intensity. For reference, we also report the realized volatility of the $1/N$ portfolio, which assigns equal weight to every stock. Note

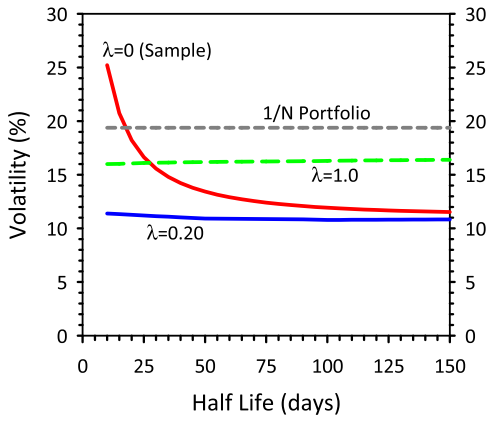


Figure 1 Out-of-sample volatility versus half-life for the minimum-volatility fully invested portfolio (containing the largest 100 US stocks). The portfolio is constructed in one of four ways. The $\lambda = 0$ case uses the sample covariance and mean–variance optimization (MVO). The $\lambda = 1$ case uses MVO under the assumption that all stocks are uncorrelated. The $\lambda = 0.2$ case uses MVO while shrinking the sample correlation by 20%. For reference, the $1/N$ portfolio (equal weighted) is also reported. The out-of-sample period spanned 27-Dec-2000 to 31-Mar-2016.

that the $1/N$ portfolio is mean–variance efficient only under the extreme assumption that all stocks are uncorrelated and have the same volatility. The $1/N$ portfolio had a realized volatility of nearly 20% annualized.

Next, we consider the $\lambda = 1$ portfolio, which is optimal under the assumption that all stocks are uncorrelated but their volatilities are given by their EWMA estimates. Rather than equally weighting all stocks, however, now the weights are inversely proportional to the predicted variance of each stock. Note that this portfolio has significantly lower volatility than the $1/N$ portfolio, demonstrating the benefit of under-weighting high-volatility stocks.

Next, we consider the $\lambda = 0$ portfolio, which is constructed using the sample covariance matrix. If the HL parameter is very short (so that the matrix is ill conditioned), the sample covariance matrix produces the worst performance (i.e., highest volatility). This type of problematic behavior has led some to view optimizers as "error maximizers," as argued by Michaud (1989). Note, however, that if the HL parameter is reasonably long (so that the matrix is reasonably well conditioned), the sample outperforms both the $1/N$ portfolio and the $\lambda = 1$ portfolio by wide margins. Hence, portfolio optimization using the sample correlation (with sufficiently long HL parameter) is able to successfully exploit asset correlations to reduce portfolio volatility.

For intermediate values of shrinkage, we obtain lower volatility than the sample covariance matrix for any choice of HL parameter. For instance using $\lambda = 0.2$, we produce portfolios that have lower out-of-sample volatility than those with no shrinkage. Note that the $\lambda = 0.2$ shrinkage intensity is not optimal across the HL spectrum. In general, we find more aggressive shrinkage is warranted for small HL parameters, while less aggressive shrinkage is suitable for longer HL parameters.

Forecasting error of asset-pair portfolios

In this section, we study the risk-forecast error for portfolios containing two correlated assets. We first consider the error from a theoretical perspective. We then discuss how these errors can be measured empirically. Finally, we present empirical results using observations from the equity and fixed income markets.

Theoretical results. Consider two assets, X and Y, with mean-zero returns, true variance of 1, and true correlation ρ. Let $x \sim N(0, 1)$ denote the return of Asset X. Similarly, let $\varepsilon \sim N(0, 1)$ be a randomly drawn return, also from a standard normal distribution. We assume that x and ε are independent. The return of the Asset Y can be

described by the return-generating process

$$y = \rho x + \sqrt{1-\rho^2}\varepsilon. \tag{5}$$

It can be easily verified that $y \sim N(0, 1)$ follows a standard normal distribution with correlation ρ to variable x.

Now consider a two-asset long/short portfolio with return given by $R = x - y$. The true portfolio variance is

$$\sigma_\rho^2 = 2(1-\rho). \tag{6}$$

In practice, the true variances and correlations are unobservable and must be estimated using a finite window of realized returns. While EWMA is typically used in practice to estimate variance, for simplicity we consider a finite look-back window containing τ periods, assigning equal weight to each period. To convert between HL parameter and τ, we use the approximation that the effective number of observations is roughly three times the HL parameter. In other words, an HL of 21 days has roughly the same sampling error as a look-back window of 63 days.

The estimated variance for Asset X is

$$\hat{\sigma}_X^2 = \frac{1}{\tau}\sum_t x_t^2, \tag{7}$$

with a similar expression $\hat{\sigma}_Y^2$ holding for Asset Y.

The sample correlation is given by the usual expression,

$$\hat{\rho} = \left(\frac{1}{\hat{\sigma}_X\hat{\sigma}_Y}\right)\frac{1}{\tau}\sum_t x_t y_t. \tag{8}$$

Next, we shrink the sample correlation toward zero,

$$\hat{\rho}_\lambda = (1-\lambda)\hat{\rho}, \tag{9}$$

where λ is the shrinkage intensity (which varies from 0 to 1). The estimated variance of the asset-pair portfolio using the shrunk correlations is given by

$$\hat{\sigma}_{\tau\rho\lambda}^2 = \hat{\sigma}_X^2 + \hat{\sigma}_Y^2 - 2\hat{\rho}_\lambda\hat{\sigma}_X\hat{\sigma}_Y. \tag{10}$$

Hence, the estimated variance depends on three parameters: (1) window length τ, (2) true correlation ρ, and (3) shrinkage intensity λ.

We define the forecast error $\delta_{\tau\rho\lambda}$ as the difference between the estimated variance and the true variance, normalized by the true variance, i.e.,

$$\delta_{\tau\rho\lambda} = \frac{\hat{\sigma}_{\tau\rho\lambda}^2 - \sigma_\rho^2}{\sigma_\rho^2}. \tag{11}$$

The root-mean-squared (RMS) error is given by

$$\varepsilon_{\tau\rho\lambda} = \sqrt{E[\delta_{\tau\rho\lambda}^2]}, \tag{12}$$

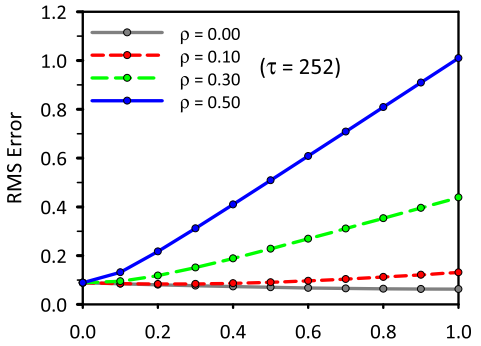
which represents the "typical" magnitude of the forecast error, expressed as a fraction of the true variance.

In Appendix A, we derive an analytic expression for the RMS error $\varepsilon_{\tau\rho\lambda}$. In what follows, we apply this result to study how RMS error depends on the three parameters τ, ρ, and λ.

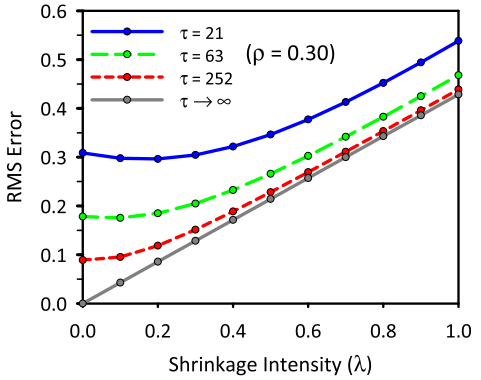
In Figure 2, we plot RMS error $\varepsilon_{\tau\rho\lambda}$ versus shrinkage intensity λ. Panel (A) considers four values of the true correlation ρ (0.0, 0.1, 0.3, and 0.5) and uses $\tau = 252$ for the length of the estimation window (corresponding to one year of daily returns). This is roughly in line with the number of effective observations for predicting correlations in practice.

Several salient features of Figure 2(A) are worth highlighting. First, as the shrinkage intensity goes to zero, we note that RMS error converges to the same level for all values of correlation. The limiting value is given by $\sqrt{2/\tau}$, which is the standard error for a variance forecast estimated from a time series of τ return observations.

Second, when the true correlation is zero (solid gray line), estimation error is minimized by fully



(a)



(b)

Figure 2 Analytic results for root-mean-squared (RMS) error versus shrinkage intensity for asset-pair portfolios. Panel (A) is for a 252-day window length, using four different values of true correlation. Panel (B) is for a true correlation of 0.30, using four different window lengths.

shrinking the estimated correlations to zero. This is not surprising, since full shrinkage in this case *eliminates* estimation error in the correlation. A crucial point, however, is that the actual reduction in error is quite small. For instance, the RMS error for no shrinkage ($\lambda = 0$) is 8.9%, versus 6.3% for full shrinkage ($\lambda = 1$). Since $\rho = 0$ represents the "best-case scenario" for shrinkage, this shows that in practice any potential improvement in the accuracy of risk forecasts due to shrinkage will be extremely small. Moreover, since in reality we never know the true underlying correlation, even this small benefit cannot be fully captured in practice.

Third, we see that shrinkage induces large RMS errors when the true correlations are non-zero. Moreover, the magnitude of these errors rises dramatically as the correlation increases. For instance, using shrinkage intensity $\lambda = 0.5$, we find in Figure 2(A) an RMS error of 9.1% for $\rho = 0.10$, which rises to 22.8% for $\rho = 0.30$, and 50.9% for $\rho = 0.50$. Furthermore, note that RMS error is exacerbated as the shrinkage intensity increases.

In Panel (B) of Figure 2, we plot RMS error versus shrinkage intensity for true correlation $\rho = 0.30$ using four different look-back windows (21, 63, 252, and infinity). The finite look-back windows correspond to 1m, 3m, and 12m of daily observations. Two features are worth highlighting. First, we note that in the limit that the sampling error goes to zero ($\tau \to \infty$), the RMS error is strictly linear in the shrinkage intensity. Moreover, for finite τ, the RMS error is *approximately* linear in shrinkage intensity for large values of λ.

This effect is explained by shrinkage-induced biases. The bias is computed by taking the expected value of the forecast error, i.e., $E[\delta_{\tau\rho\lambda}]$. This is easily obtained by substituting the estimated variance given by Equation (10) and the true variance given by Equation (6) into the definition of forecast error in Equation (11). Taking expectations, we find the resulting bias is given by

$$E[\delta_{\tau\rho\lambda}] = \frac{\lambda\rho}{1-\rho}. \tag{13}$$

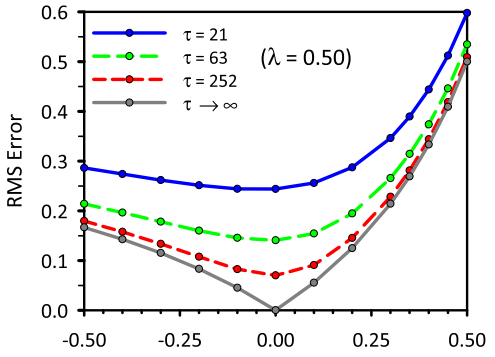
In the limit that ($\tau \to \infty$), sampling error vanishes and the bias becomes identical to the RMS error. This explains why the ($\tau \to \infty$) result in Figure 2(B) is strictly linear in λ. For any finite τ, sampling error adds noise on top of the bias, further increasing the RMS error. Nevertheless, for look-back windows of reasonable length (e.g., $\tau = 252$), we find that Equation (13) is an excellent approximation to RMS error for wide

ranges of correlation and shrinkage intensity. For instance, using $\lambda = 0.5$, $\rho = 0.5$, and $\tau = 252$, the bias is 0.50, whereas the RMS error is 0.51. In these cases, RMS error is dominated by bias.

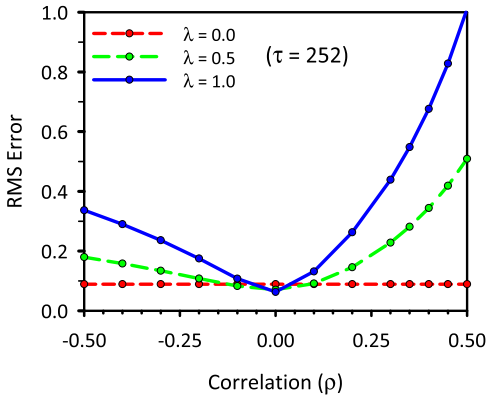
The second observation from Figure 2(B) is that for short look-back windows (e.g., $\tau = 21$), there is a clear (though shallow) minimum. For instance, using a 21-day window, the RMS error is minimized at a shrinkage intensity of $\lambda = 0.16$. Note, however, that the actual reduction in error (compared to no shrinkage) is miniscule. For instance, even with $\tau = 21$, the RMS error for no shrinkage ($\lambda = 0$) is 30.9%, versus and 29.6% at the optimal shrinkage intensity. Moreover, a look-back window of 21 days is unreasonably short for purposes of estimating correlations, so in practice any benefit would be even smaller.

Interestingly, we find that for any finite τ, RMS error is *always* minimized for a non-zero shrinkage intensity. However, two key effects occur as the window length increases: (1) the optimal shrinkage intensity converges to zero, and (2) the reduction in RMS error also converges to zero. For instance, using $\tau = 252$ and $\rho = 0.30$, we find that the optimal shrinkage intensity is $\lambda = 0.02$. The RMS error with no shrinkage ($\lambda = 0$) is 8.91%, whereas the RMS error with optimal shrinkage is 8.86%.

In summary, we find that the reduction in RMS error due to optimal shrinkage is *virtually zero* unless the estimation window is extremely short and the true correlation is very small. Hence, for the typical look-back windows used in practice, the reduction in error is *immaterial*. Moreover, in practice, the benefit of shrinkage will always be less than what is shown here, since in reality, we do not know the true correlation and so we measure the optimal shrinkage intensity with error. Another key takeaway is that excessive shrinkage may induce large biases in volatility forecasts.



(a)



(b)

Figure 3 Analytic results for root-mean-squared (RMS) error versus correlation for asset-pair portfolios. Panel (A) is for a shrinkage intensity of 0.50, and multiple values of window length. Bottom panel (B) is for an estimation window of 252 periods, with multiple shrinkage intensities.

In Figure 3, we study the functional dependency of RMS error on correlation. We use the same long/short pair portfolio as before (i.e., $R = x - y$), but now consider both positive and negative correlations. Note that the risk of the long/short pair portfolio with correlation $(-\rho)$ is equivalent to the risk of a long-only pair portfolio ($R = x+y$) with correlation $(+\rho)$.

Panel (A) of Figure 3 reports RMS error for a shrinkage intensity of $\lambda = 0.50$, using four different values for window length τ (21, 63, 252, and infinity). We first note that all curves have

a minimum at $\rho = 0$. This is intuitive, because only if the true correlation is zero will shrinkage lead to unbiased forecasts. The most striking feature of Figure 3(A) is the strong asymmetry in correlation. This occurs because RMS error is normalized using the true variance in the denominator, which is smaller for positively correlated long/short pairs than it is for negatively correlated pairs.

Also, note that for any value of correlation, RMS error increases as τ decreases. This is also intuitive, since smaller window lengths translate into higher sampling error.

In Panel (B) of Figure 3, we report RMS error versus correlation for a window length $\tau = 252$, using three different values of shrinkage intensity (0.0, 0.5, and 1.0). First, note that with no shrinkage ($\lambda = 0$), RMS error is independent of correlation. Again, the RMS error in this case is given by $\sqrt{2/\tau}$, consistent with Figure 2(A).

Second the RMS error for full shrinkage ($\lambda = 1$) is typically much greater than for no shrinkage ($\lambda = 0$), except for tiny correlations. For instance, over the very narrow range of correlations $[-0.06, 0.05]$, full shrinkage outperforms no shrinkage. For intermediate values of shrinkage ($\lambda = 0.5$), the range over which shrinkage outperforms is slightly wider $[-0.15, 0.10]$. Note, however, that the reduction in RMS error is exceedingly small, even for zero true correlation. The most significant feature of Figure 3(B) is that as the correlation increases, shrinkage induces large errors in the variance forecasts. For instance, using $\rho = 0.5$ and $\lambda = 1$, we find an RMS error exceeding 100%.

Q-statistics. Analysis of RMS error provides important theoretical insight into the key drivers of forecasting error, particularly the interplay among window length, correlation, and shrinkage intensity. Unfortunately, RMS error relies on knowing the true correlations and volatilities. Clearly, this measure cannot be applied in practice, where these quantities are unknown. What is needed is a measure that can be applied empirically.

The observable quantity that we use as a proxy for RMS error is the so-called Q-statistic, which was analyzed by Patton (2011), and further studied by Menchero and Morozov (2015). The Q-statistic is defined in terms of standardized portfolio returns (i.e., z-scores). In particular, let r_{nt} be the return of portfolio n over period t, and let $\hat{\sigma}_{nt}$ denote the start-of-period volatility forecast. The Q-statistic for this observation is given by

$$Q_{nt} = z_{nt}^2 - \ln(z_{nt}^2), \tag{14}$$

where $z_{nt} = r_{nt}/\hat{\sigma}_{nt}$. In practice, the Q-statistic is averaged across N portfolios and T periods, leading to a composite value: $\overline{Q} = (NT)^{-1} \sum_{nt} Q_{nt}$.

A critical property of the Q-statistic (derived in Technical Appendix B) is that it is minimized in expectation when the *true* volatility is used to make every forecast. This opens the possibility of using the Q-statistic as an observable proxy for the unobservable RMS error. To apply the Q-statistic in a reliable fashion, it is important to have a large number of observations to mitigate the effects of sampling error.

To develop some intuition behind the Q-statistic, note that we can interpret z_{nt}^2 as an under-forecasting penalty, whose expected value becomes large if our volatility forecasts are too low. Similarly, we can interpret $-\ln(z_{nt}^2)$ as an over-forecasting penalty, which becomes large if our risk forecasts are too high.

The expected value of the Q-statistic (for perfect risk forecasts) depends on the return distribution of the portfolio under consideration. It is easy to show via simulations that if returns are normally distributed, the expected value of the Q-statistic is approximately 2.27. For fat-tailed distributions,

again assuming perfect forecasts, the expected value of the Q-statistic is greater than 2.27, and grows with increasing kurtosis.

In practice, we are less concerned with the numerical value of the Q-statistic than we are in *differences* between Q-statistics for competing volatility estimates. Let $Q(\hat{\sigma})$ denote the Q-statistic using the estimated volatility $\hat{\sigma}$, and let $Q(\sigma)$ denote the Q-statistic using the true (unobservable) volatility. In Technical Appendix B, we show that the expected increase in Q-statistic is given by

$$E[\Delta Q] = \frac{\sigma^2}{\hat{\sigma}^2} - \ln\left(\frac{\sigma^2}{\hat{\sigma}^2}\right) - 1, \qquad (15)$$

where $\Delta Q \equiv Q(\hat{\sigma}) - Q(\sigma)$. Remarkably, Equation (15) holds *independent* of distribution. For small errors, Equation (15) is essentially symmetric. For instance, if the true volatility is $\sigma = 1.00$ and our estimated volatility is either $\hat{\sigma} = 1.01$ or $\hat{\sigma} = 0.99$, the expected increase in the Q-statistic is approximately 2.0×10^{-4}. For large errors, however, the Q-statistic penalizes underforecasting more heavily than over-forecasting. For instance, if the true volatility is $\sigma = 1.00$ and our estimated volatility is $\hat{\sigma} = 0.50$, the expected increase in the Q-statistic is 1.61. However, if the estimated volatility is $\hat{\sigma} = 2.00$, the expected increase is only 0.64. If we adopt the view that 10% forecasting error is "significant," this translates into $E[\Delta Q] \approx 0.02$. Hence, a useful rule of thumb is that a difference of 0.02 in the value of the Q-statistic is considered a "meaningful" improvement to the risk forecast.

To help understand the relationship between the Q-statistic and RMS error, we conduct a numerical simulation. As before, we consider a pair of assets, each with true volatility $\sigma = 1$ and true correlation ρ. The portfolio goes 100% long Asset X and 100% short Asset Y. To make a reliable comparison between the Q-statistics for different volatility estimates, it is necessary to have a large number of observations. We simulate one million returns for each asset using the return-generating process in Equation (5). In our simulations, we use Equation (10) to estimate portfolio volatilities using different window lengths τ and shrinkage intensities λ.

In Figure 4, we plot the average Q-statistic as a function of shrinkage intensity. Panel (A) considers a 252-day window length, with four different values of correlation. Panel (B) considers a true correlation of 0.30, with four different window lengths. Note that the results for infinite look-back window were not determined by simulation. In this case, we substitute $\hat{\sigma}^2 = 2 - 2\rho(1 - \lambda)$

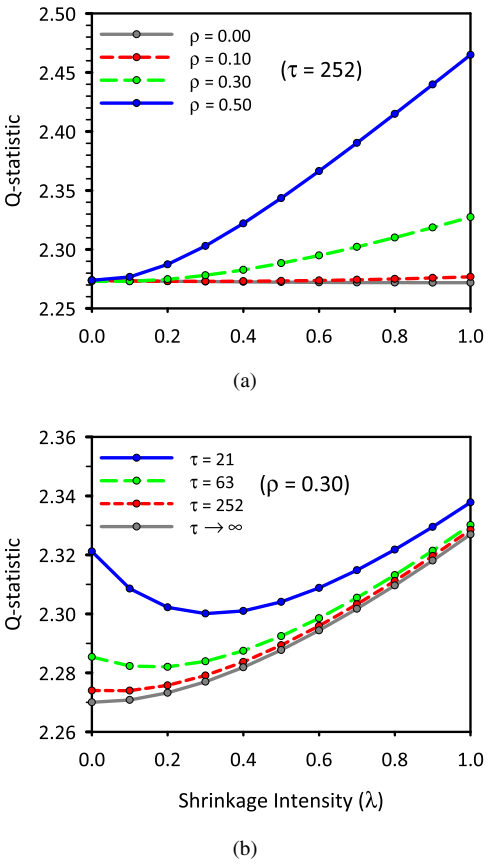


Figure 4 Simulated results for Q-statistic versus shrinkage intensity for asset-pair portfolios. Panel (A) is for a 252-day window, using four values of the "true" correlation. Panel (B) is for a true correlation of 0.30 with four different look-back windows.

and $\sigma^2 = 2(1 - \rho)$ into Equation (15) to obtain an analytic result.

We wish to compare the behavior of RMS error with the Q-statistic. Note that the parameter settings in Figures 2 and 4 are identical. Comparing Figure 2(A) with Figure 4(A), we see that qualitatively, RMS error and the Q-statistic exhibit virtually identical behavior. For instance, if the true correlation is zero (solid gray line), we see a slight reduction in both RMS error and the Q-statistic as we increase the shrinkage intensity. As another example, if the true correlation is 0.50, then both RMS error and the Q-statistic increase quite dramatically (almost linearly) with increasing shrinkage intensity.

Figure 2(B) is again qualitatively similar to Figure 4(B). For instance, using relatively long windows (e.g., $\tau = 252$), we find that both RMS error and the Q-statistics increase with increasing shrinkage intensity. By contrast, for short window length (e.g., $\tau = 21$), there is an evident benefit to shrinkage. For instance, with a 21-day window length, we see from Figure 2(B) that the optimal shrinkage intensity for RMS error is roughly 20%. In Figure 4(B), we see that the Q-statistic is minimized near a shrinkage intensity of 30%.

In summary, we find remarkable similarity between Figures 2 and 4. This shows that the Q-statistic may serve as a viable proxy for the unobservable RMS error.

Empirical results (Q-statistics). Up to now, all of our results were based on idealized examples where the return-generating process was normally distributed, stationary, and fully specified. In this section, we apply our concepts to empirical observations where all of these conditions are violated.

Our first exercise illustrates how the Q-statistic can be applied as a tool for calibrating model parameters. Accurate risk forecasting involves finding an optimal tradeoff between responsiveness and sampling error. Responsive forecasts employ a *short* HL parameter to assign more weight to recent observations, which is desirable since the recent past is the best indicator of the immediate future. On the other hand, to mitigate the effects of sampling error, a long HL parameter is preferable. The Q-statistic allows us to identify the optimal tradeoff between these two effects.

As our data set, we take the 20 largest US equities as of 31-Mar-2016, with complete daily return history going back to 05-Jan-1993. The first two years form the burn-in period (for obtaining our initial volatility estimates), with the out-of-sample testing period running from 30-Dec-1994 to 31-Mar-2016. We use EWMA with a variable HL parameter to estimate stock volatility. We then average the Q-statistic across all 20 stocks over more than 20 years of daily returns (more than 100,000 total observations).

In Figure 5, we plot the mean Q-statistic of the individual stocks (solid blue line) as a function of volatility HL. We see that the curve has a pronounced minimum near a 20-day HL, indicating that this represents the optimal HL for risk-forecasting purposes. We round the optimal HL parameter to one month (21 days). If the HL parameter is well below 21 days, the forecast becomes more responsive, but also more noisy, causing an increase in the Q-statistic. Alternatively, if the HL parameter is well above 21 days, then sampling error (noise) is reduced, but the use of "stale" data causes the Q-statistic to rise.

Next, we form stock-pair portfolios where the weights are inversely proportional to the predicted stock volatility. This means that each "leg" of the portfolio has the same stand-alone volatility (as in our previous simulations). To estimate the stock volatility, we use the optimal 21-day HL. To estimate correlations, however, we use a variable HL. We then compute Q-statistics for

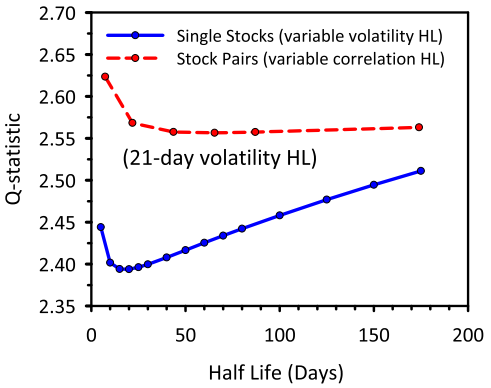


Figure 5 Empirical results for Q-statistic versus HL for single stocks and stock pairs. For single stocks (solid blue line), we varied the volatility HL. For the stock pairs, we kept the volatility HL constant at 21 days (near optimal), and varied the correlation HL. The selected stocks were the 20 largest US equities as of 31-Mar-2016 with complete daily return history back to 05-Jan-1993. The out-of-sample period spans 30-Dec-1994 to 31-Mar-2016. The average pairwise correlation between stock pairs was 0.34.

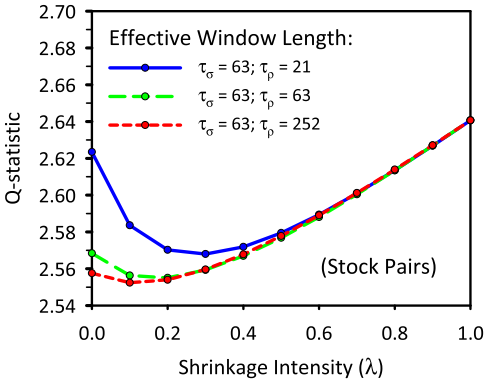


Figure 6 Empirical results for Q-statistic versus shrinkage intensity for stock-pair portfolios using three different correlation HL (effective look-back windows). For the stock pairs, we kept the volatility HL constant at 21 days (near optimal), and varied the correlation HL. The selected stocks were the 20 largest US equities as of 31-Mar-2016 with complete daily return history back to 05-Jan-1993. The out-of-sample period spans 30-Dec-1994 to 31-Mar-2016.

all 190 pair portfolios over 20+ years of daily returns (roughly one million total observations). The result is shown by the dashed red line in Figure 5. The optimal HL in this case is near 60 days (about three months). However, note that the minimum is extremely shallow. In other words, using a longer HL parameter for correlations leads to only a very slight (immaterial) increase in the Q-statistic. Moreover, the longer correlation HL leads to a better-conditioned covariance matrix, which proves useful for portfolio construction. This analysis provides empirical evidence to support the view that short HL parameters are useful to estimate volatilities, whereas longer HL parameters are preferable for correlations.

In Figure 6, we plot the mean Q-statistic as a function of shrinkage intensity for the 190 stock-pair portfolios. As before, we estimate stock volatility using a 21-day HL, which corresponds to roughly $\tau_\sigma = 63$ effective observations. We estimated the stock correlations using three different HL parameters (7d, 21d, and 84d), which correspond to an effective number of observations τ_ρ equal to those in Figure 4(B), (i.e., 21d, 63d, and 252d).

Note that the average pairwise correlation of the 190 pairs was 0.34, close to the correlation of 0.30 used to generate Figure 4(B). Qualitatively, we see that Figures 6 and 4(B) are very similar. For instance, using an effective window length of $\tau_\rho = 21$ days leads to an optimal shrinkage intensity of 0.30 in both the simulated results of Figure 4(B) and the empirical results of Figure 6.

While the qualitative behavior is quite similar, there are some notable differences. First, the mean Q-statistics in Figure 6 (empirical) are significantly larger than the corresponding values in Figure 4(B). Part of the reason for this increase is that the empirical distribution is fat tailed, whereas the simulated results are normally distributed. Another contributing factor is that the empirical distribution is non-stationary, while the simulations were stationary.

Another notable difference is that in Figure 6, all lines appear to converge to the same value as the shrinkage intensity approaches 1. By contrast, in Figure 4(B), the short HL case clearly has a larger Q-statistic for $\lambda = 1$. This is because in Figure 4, the volatility HL is the same as correlation HL, whereas in Figure 6 the volatility HL is 21-days for all curves.

Bias statistics. Another useful measure of risk-forecasting accuracy is the so-called bias statistic, which is also defined in terms of z-scores. The bias statistic B is defined by

$$B^2 = \frac{1}{NT} \sum_{nt} z_{nt}^2. \tag{16}$$

The bias statistic essentially represents the ratio of realized risk to predicted risk. In particular, if the true volatility is used to make every forecast, then $E[B^2] = 1$. Over-forecasting of risk leads to bias statistics less than 1, whereas under-forecasting leads to bias statistics greater than 1.

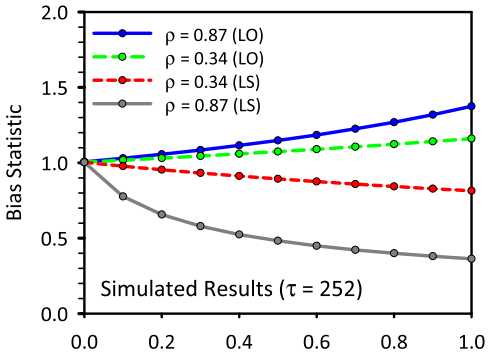
An advantage of the bias statistic over the Q-statistic is that the former has a more intuitive interpretation. A disadvantage of the bias statistic is that it is less reliable than the Q-statistic for purposes of evaluating the accuracy of risk forecasts. In particular, the bias statistic is subject to *error cancelation*. For instance, suppose we over-forecast risk for six months and under-forecast risk for the subsequent six months. Measured over the entire year, the bias statistic may be very close to 1, but this does not imply that the risk forecasts were accurate. By contrast, the Q-statistic is not susceptible to error cancelation; each forecasting error increases the expected value of the Q-statistic.

Empirical results (bias statistics). Next, we explicitly consider the biases in asset-pair portfolios induced by correlation shrinkage. We first simulate returns for two assets with equal volatility and true correlation ρ, using the same joint-normal assumptions as before. We considered two values for the true correlation: an "intermediate" correlation ($\rho = 0.34$) which is typical of individual stocks, and a "high" correlation $\rho = 0.87$ that is typical of a set of bond indices (described below). Asset weights were inversely proportional to the estimated volatility using a 21-day HL. Hence, within estimation error, the two legs of our portfolios have equal stand-alone volatility.

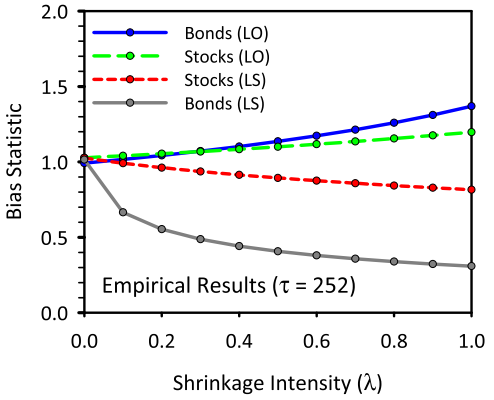
In Panel (A) of Figure 7, we plot simulated results for the bias statistics of the pair portfolios as a function of shrinkage intensity. The HL used to estimate correlations was 84 days (corresponding to an effective window size of 252 days). For the long-only (LO) portfolios, we see that shrinkage causes *under-prediction* of risk. Furthermore, the greater the correlation, the greater the bias. Note that the theoretical maximum bias is bounded by $\sqrt{2}$, which occurs for full shrinkage of two perfectly correlated assets.

By contrast, for long/short (LS) pairs, shrinkage causes significant *over-forecasting* of portfolio risk. Again, the biases are more severe for the highly correlated asset pairs. In this scenario, the bias statistic can reach as low as zero. This occurs when the assets are perfectly correlated, which leads to a true volatility of zero.

In Panel (B) of Figure 7, we plot the bias statistics for long/short and long-only pair portfolios of stocks and bonds. Again, the weights were inversely proportional to the estimated volatility, which was computed using a 21-day HL. To form the stock-pair portfolios, we used the same top 20 US stocks considered in Figure 5 (whose average correlation was 0.34). The out-of-sample period spanned 30-Dec-1994 to 31-Mar-2016. We considered all 190 combinations of stock-pair portfolios. Note that the empirical biases for the stock-pair portfolios are extremely similar



(a)



(b)

Figure 7 Bias statistic versus shrinkage intensity for long-only (LO) and long/short (LS) asset pairs. Panel (A) is for joint-normal simulated data in which both assets are assumed to have the same volatility and true correlation of 0.34 or 0.87. In Panel B, we present empirical results for stock pairs (mean correlation 0.34) and bond pairs (mean correlation 0.87). The out-of-sample period for stocks was from 30-Dec-1994 to 31-Mar-2016, whereas for bonds the period spanned 23-Nov-2004 to 05-Jan-2018. In both Panel (A) and Panel (B), an 84d HL (252d effective window) was used to estimate correlations and asset weights were inversely proportional to the estimated volatility using a 21d HL (63d effective window).

to the simulated biases in Figure 7(A) using a correlation of 0.34.

To form our bond-pair portfolios, we took 20 fixed income indices derived from the Bloomberg-Barclays US Aggregate Bond Index. In particular, we considered various carve-outs of the index, such as US Treasuries (of various maturity buckets), the US Aggregate with various quality/maturity buckets, US Agency bonds, and MBS. The average correlation among these 190 possible bond-index pairs was 0.87. Again, we see that the empirical biases in Figure 7(B) match up very closely to the simulated biases in Figure 7(A).

Summary

Shrinking correlations toward zero—either explicitly or implicitly—is a common practice in many financial risk model applications. In this paper, we studied the impact of such shrinkage on the accuracy of risk forecasts for asset-pair portfolios. We found that while shrinkage may be helpful in portfolio optimization, there is typically little benefit to shrinkage for risk-forecasting purposes. In fact, for any reasonable choice of HL parameter, the optimal shrinkage intensity is close to zero and we find that the benefit due to optimal shrinkage is *immaterial*. However, if the shrinkage intensity is too high—as is often the case in practice—shrinkage may result in large errors (biases) in volatility forecasts. Hence, we conclude that for risk-forecasting purposes, estimated correlations should not deviate appreciably from the sample correlation.

Appendix A: Derivation of RMS error formula

In this Appendix, we derive the formula for the root-mean-square (RMS) error of the variance estimate for a long/short portfolio of two assets (X, Y) with correlation ρ. The portfolio return is $R = x - y$, where x and y each has a standard normal distribution.

The error $\delta_{\tau\rho\lambda}$ is defined by Equation (11). Substituting Equation (10) into Equation (11), the error

may be written as

$$\delta_{\tau\rho\lambda} = \frac{1}{\sigma_\rho^2}[A + B + C - \sigma_\rho^2], \tag{A1}$$

where $\sigma_\rho^2 = 2(1-\rho)$ is the true variance of the portfolio, A is the sample variance of X,

$$A \equiv \frac{1}{\tau}\sum_t x_t^2, \tag{A2}$$

τ is the number of periods in the sample, B is the sample variance of Y,

$$B \equiv \frac{1}{\tau}\sum_t y_t^2, \tag{A3}$$

and C is the off-diagonal covariance term,

$$C \equiv \frac{-2(1-\lambda)}{\tau}\sum_{t=1}^{\tau} x_t y_t, \tag{A4}$$

where λ is the shrinkage intensity for the correlation. Hence, the mean-squared error is

$$\begin{aligned} E[\delta_{\tau\rho\lambda}^2] = \frac{1}{\sigma_\rho^4}E[&A^2 + B^2 + C^2 + \sigma_\rho^4 \\ &+ 2AB + 2AC - 2A\sigma_\rho^2 \\ &+ 2BC - 2B\sigma_\rho^2 - 2C\sigma_\rho^2]. \end{aligned} \tag{A5}$$

We now directly compute the expected values of each term. For the first term, we have

$$E[A^2] = \frac{1}{\tau^2}\sum_{t=1}^{\tau}\sum_{t'=1}^{\tau} E[x_t^2 x_{t'}^2]. \tag{A6}$$

Note that if $t \neq t'$, then x_t^2 and $x_{t'}^2$ are independent, implying $E[x_t^2 x_{t'}^2] = E[x_t^2]E[x_{t'}^2]$, which gives $E[x_t^2 x_{t'}^2] = 1$. If $t = t'$, then we have $E[x_t^4] = 3$, which is just the kurtosis of a standard normal distribution. Hence, the general expression is $E[x_t^2 x_{t'}^2] = 1 + 2\delta_{tt'}$, where $\delta_{tt'}$ is the delta function, which is equal to 1 if $t = t'$, and is equal to zero otherwise. Substituting this result into Equation (A6) gives

$$E[A^2] = \frac{1}{\tau^2}\sum_{t=1}^{\tau}\sum_{t'=1}^{\tau}(1 + 2\delta_{tt'}). \tag{A7}$$

Carrying out the sum over t', we obtain

$$E[A^2] = \frac{1}{\tau^2}\sum_{t=1}^{\tau}(\tau + 2). \tag{A8}$$

Finally, carrying out the sum over t gives

$$E[A^2] = \frac{\tau + 2}{\tau}. \tag{A9}$$

The second term is derived in identical fashion, which leads to

$$E[B^2] = \frac{\tau + 2}{\tau}. \tag{A10}$$

The next term is

$$E[C^2] = \frac{4(1-\lambda)^2}{\tau^2}\sum_{t=1}^{\tau}\sum_{t'=1}^{\tau} E[x_t y_t x_{t'} y_{t'}]. \tag{A11}$$

To compute this term, we write the standard expression for two correlated standard normal random variables, i.e., $y_t = \rho x_t + \sqrt{1-\rho^2}\varepsilon_t$, where ε_t follows a standard normal distribution and is independent of x_t. Substituting these expressions into Equation (A11), we obtain

$$\begin{aligned} E[C^2] = \frac{4(1-\lambda)^2}{\tau^2}\sum_{t=1}^{\tau}\sum_{t'=1}^{\tau} E\big[x_t\big(&\rho x_t \\ &+ \sqrt{1-\rho^2}\varepsilon_t\big)x_{t'}\big(\rho x_{t'} + \sqrt{1-\rho^2}\varepsilon_{t'}\big)\big]. \end{aligned} \tag{A12}$$

We have already shown $E[x_t^2 x_{t'}^2] = 1 + 2\delta_{tt'}$. Next, we must compute $E[x_t^2 x_{t'}\varepsilon_{t'}]$. Since $\varepsilon_{t'}$ is independent, we have $E[x_t^2 x_{t'}\varepsilon_{t'}] = E[x_t^2 x_{t'}]E[\varepsilon_{t'}]$, which is equal to zero. Finally, we must compute $E[x_t\varepsilon_t x_{t'}\varepsilon_{t'}]$. If $t \neq t'$, then all terms are independent, which gives zero. If $t = t'$, we

have $E[x_t^2\varepsilon_t^2] = E[x_t^2]E[\varepsilon_t^2]$, which is equal to 1. Hence, the general expression is given by $E[x_t\varepsilon_t x_{t'}\varepsilon_{t'}] = \delta_{tt'}$. Plugging these expressions into Equation (A12), we obtain

$$E[C^2] = \frac{4(1-\lambda)^2}{\tau^2}\sum_{t=1}^{\tau}\sum_{t'=1}^{\tau}[\rho^2(1+2\delta_{tt'}) + (1-\rho^2)\delta_{tt'}]. \quad \text{(A13)}$$

Carrying out the sums, we find

$$E[C^2] = \frac{4(1-\lambda)^2(\tau\rho^2+\rho^2+1)}{\tau}. \quad \text{(A14)}$$

Next, we must evaluate the cross-terms,

$$E[AB] = \frac{1}{\tau^2}\sum_{t=1}^{\tau}\sum_{t'=1}^{\tau}E[x_t^2 y_{t'}^2]. \quad \text{(A15)}$$

We use the return-generating process to express $y_{t'}^2$ in terms of $x_{t'}$ and $\varepsilon_{t'}$, i.e.,

$$y_{t'}^2 = \left(\rho x_{t'} + \sqrt{1-\rho^2}\varepsilon_{t'}\right)^2. \quad \text{(A16)}$$

Substituting Equation (A16) into Equation (A15), and using $E[x_t^2 x_{t'}^2] = 1 + 2\delta_{tt'}$, $E[x_t^2\varepsilon_{t'}^2] = 1$, and $E[x_t^2 x_{t'}\varepsilon_{t'}] = 0$, we find

$$E[AB] = \frac{1}{\tau^2}\sum_{t=1}^{\tau}\sum_{t'=1}^{\tau}[\rho^2(1+2\delta_{tt'}) + (1-\rho^2)]. \quad \text{(A17)}$$

Carrying out the sums, we obtain

$$E[AB] = 1 + \frac{2\rho^2}{\tau}. \quad \text{(A18)}$$

The next cross-term is

$$E[AC] = \frac{-2(1-\lambda)}{\tau^2}\sum_{t=1}^{\tau}\sum_{t'=1}^{\tau}E[x_t^2 x_{t'} y_{t'}]. \quad \text{(A19)}$$

Substituting $y_{t'} = \rho x_{t'} + \sqrt{1-\rho^2}\varepsilon_{t'}$ into Equation (A19) and carrying out the algebra, we find

$$E[AC] = -2(1-\lambda)\left(\rho + \frac{2\rho}{\tau}\right). \quad \text{(A20)}$$

By inspection, $E[BC] = E[AC]$, and $E[A] = E[B] = 1$. Finally,

$$E[C] = -2\rho(1-\lambda). \quad \text{(A21)}$$

Substituting the 10 expectation values into Equation (A5) and using $\sigma_\rho^2 = 2(1-\rho)$, we find after a bit of algebra,

$$E[\delta_{\tau\rho\lambda}^2] = a_0 + a_1(1-\lambda) + a_2(1-\lambda)^2, \quad \text{(A22)}$$

where the constant term is given by

$$a_0 = \frac{1}{(1-\rho)^2}\left(\rho^2 + \frac{1+\rho^2}{\tau}\right), \quad \text{(A23)}$$

the linear coefficient is given by

$$a_1 = \frac{-2}{(1-\rho)^2}\left(\rho^2 + \frac{2\rho}{\tau}\right), \quad \text{(A24)}$$

and the quadratic coefficient is given by

$$a_2 = a_0. \quad \text{(A25)}$$

Finally, taking the square root of Equation (A22), we obtain the RMS error $\varepsilon_{\tau\rho\lambda}$, which is Equation (12) of the main text. To find the optimal shrinkage intensity λ^*, we take the derivative of Equation (A22) and set it equal to zero. The result is

$$\lambda^* = \frac{(1-\rho)^2}{1+\rho^2+\tau\rho^2}. \quad \text{(A26)}$$

It is easily verified that the second derivative is everywhere positive, so Equation (A26) represents the global optimal shrinkage intensity that minimizes mean-squared error.

Appendix B: Q-statistics

Let r be the portfolio return and let $\hat{\sigma}$ denote the predicted start-of-period volatility. Express the out-of-sample return as a z-score, i.e., $z = r/\hat{\sigma}$.

The Q-statistic for the return observation is defined in terms of the z-score,

$$Q = z^2 - \ln(z^2). \tag{B1}$$

We first show that the Q-statistic is minimized in expectation when the true volatility is used for the forecast. We assume that returns are mean zero. This is an excellent approximation for daily stock returns and still quite valid even for monthly returns. The expected value of the Q-statistic is given by

$$E[Q] = E\left[\frac{r^2}{\hat{\sigma}^2}\right] - E\left[\ln\left(\frac{r^2}{\hat{\sigma}^2}\right)\right]. \tag{B2}$$

Let $\sigma^2 \equiv E[r^2]$ denote the *true* portfolio variance (returns are assumed mean zero). Hence, Equation (B2) can be rewritten as

$$E[Q] = \frac{\sigma^2}{\hat{\sigma}^2} - 2E[\ln(r)] + 2\ln(\hat{\sigma}). \tag{B3}$$

The derivative of $E[Q]$ with respect to $\hat{\sigma}$ is given by

$$\frac{dE[Q]}{d\hat{\sigma}} = -\frac{2\sigma^2}{\hat{\sigma}^3} + \frac{2}{\hat{\sigma}}. \tag{B4}$$

Let $\tilde{\sigma}$ denote the value $\hat{\sigma}$ of that produces an extremum. This is found by setting the derivative equal to zero, which gives

$$\tilde{\sigma}^2 = \sigma^2. \tag{B5}$$

Note that the *second* derivative evaluated at the extremum is equal to $4/\tilde{\sigma}^2$. Since the second derivative is positive, this proves that $E[Q]$ is minimized when we use the true volatility for the forecast.

Next, we solve for the expected increase in the Q-statistic due to forecasting error. Let σ^2 denote the true variance of the portfolio, and let $\hat{\sigma}^2$ be the estimated variance. The expected value of the Q-statistic using the true variance is given by

$$E[Q(\sigma)] = 1 - E\left[\ln\left(\frac{r^2}{\sigma^2}\right)\right], \tag{B6}$$

where we have used the fact that $E[r^2/\sigma^2] = 1$. Similarly, the expected value of the Q-statistic using the estimated variance is given by

$$E[Q(\hat{\sigma})] = E\left[\left(\frac{r^2}{\sigma^2}\frac{\sigma^2}{\hat{\sigma}^2}\right)\right] - E\left[\ln\left(\frac{r^2}{\sigma^2}\frac{\sigma^2}{\hat{\sigma}^2}\right)\right], \tag{B7}$$

which reduces to

$$E[Q(\hat{\sigma})] = \frac{\sigma^2}{\hat{\sigma}^2} - E\left[\ln\left(\frac{r^2}{\sigma^2}\right) + \ln\left(\frac{\sigma^2}{\hat{\sigma}^2}\right)\right]. \tag{B8}$$

Letting $\Delta Q \equiv Q(\hat{\sigma}) - Q(\sigma)$, we find

$$E[\Delta Q] = \frac{\sigma^2}{\hat{\sigma}^2} - \ln\left(\frac{\sigma^2}{\hat{\sigma}^2}\right) - 1. \tag{B9}$$

Note that this result is independent of distribution. Define the error in variance

$$\delta = \frac{\hat{\sigma}^2 - \sigma^2}{\sigma^2}. \tag{B10}$$

Equation (B9) may be written as

$$E[\Delta Q] = \frac{1}{1+\delta} + \ln(1+\delta) - 1. \tag{B11}$$

Next, we do a Taylor Series expansion, keeping terms second order in δ. The result is

$$E[\Delta Q] = \frac{\delta^2}{2}. \tag{B12}$$

Hence, the expected increase in Q-statistic is proportional to the mean-squared error (MSE). For small values of δ, $E[\Delta Q]$ is symmetric in δ. However, for larger errors the higher-order terms become important. Note that the penalty is asymmetric. If we under-forecast volatility by 50%, we find $E[\Delta Q] = 1.61$, whereas if we over-forecast by 50%, we obtain $E[\Delta Q] = 0.26$. If we over-forecast by 100%, we get $E[\Delta Q] = 0.64$.

References

Grinold, R. and Kahn, R. (2000). *Active Portfolio Management* (Second Edition). McGraw-Hill.

Ledoit, O. and Wolf, M. (2003). "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection," *Journal of Empirical Finance* **10**(5), 603–621.

Markowitz, H. (1952). "Portfolio Selection," *Journal of Finance* **7**(1), 77–91.

Menchero, J. (2010). "The Characteristics of Factor Portfolios," *Journal of Performance Measurement* **15**(1) (Fall), 52–62.

Menchero, J. and Ji, L. (2019). "Porfolio Optimization with Noisy Covariance Matrices," *Journal of Investment Management* **17**(1) (Q1), 77–91.

Menchero, J. and Morozov, A. (2015). "Improving Risk Forecasts Through Cross-sectional Observations," *Journal of Portfolio Management* **41**(3) (Spring), 84–96.

Michaud, R. (1989). "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?" *Financial Analysts Journal* **45**(1) (January/February), 31–42.

Patton, A. (2011). "Volatility Forecast Comparison Using Imperfect Volatility Proxies," *Journal of Econometrics* **160**(1), 246–256.

Sharpe, W. (1963). "A Simplified Model for Portfolio Analysis," *Management Science* **9**(2) (January), 277–293.

Shepard, P. (2007/2008). "Integrating Multi-market Risk Models," *Journal of Risk* **10**(2) (Winter), 25–45.

Vasicek, O. (1973). "A Note on Using Cross-sectional Information in a Bayesian Estimation of Security Betas," *Journal of Finance* **28**(5) (December), 1233–1239.

Keywords: Shrinkage; correlation matrices; estimation error; volatility forecasting