JOIM
www.joim.com

Journal Of Investment Management, Vol. 19, No. 3, (2021), pp. 60–80

ADVANCES IN ESTIMATING COVARIANCE MATRICES

Jose Menchero[a] *and Lei Ji*[b]

Correlation matrices are widely used in finance both for risk forecasting and for portfolio optimization. It is well known that the sample correlation matrix is unreliable for portfolio optimization. However, we show that for purposes of predicting portfolio risk, the sample correlation matrix is close to optimal. In this paper, we present a technique for estimating correlations that is well suited both for risk forecasting and for portfolio optimization. We apply our technique to estimate factor correlation matrices spanning different asset classes. We find that our technique produces improved correlation estimates compared to an alternative widely used approach.

Covariance matrices play a central role in the modern investment process. In particular, they are employed for two primary purposes. The first use-case is for risk management, which requires accurate forecasting of portfolio volatility. The second major application is for portfolio construction, with mean–variance optimization being the leading example.

Mean–variance optimization is a portfolio-construction technique pioneered by Markowitz (1952), which ushered in Modern Portfolio Theory. The objective is to construct an *efficient* portfolio, defined as having maximum expected return per unit of risk. The basic inputs required for mean–variance optimization are: (a) the expected asset returns, (b) the asset covariance matrix, and (c) a set of investment constraints.

The simplest way of estimating the asset covariance matrix is to compute it directly for each pair of assets using the textbook definition of covariance. The resulting matrix is known as the *sample covariance matrix*.

Unfortunately, the sample covariance matrix is unreliable for use in portfolio optimization. The primary reason for this shortcoming is attributable to sampling error in the correlation estimates. In statistics, sampling error (also known as "noise") arises whenever we estimate a quantity using a finite sample of data.

Menchero and Ji (2019) showed that sampling error in the covariance matrix leads to several

[a]Head of Portfolio Analytics Research, Bloomberg. Phone 415-823-7191, E-mail: jmenchero@bloomberg.net
[b]Portfolio Analytics Research, Phone 646-691-8635, E-mail: lji19@bloomberg.net

detrimental effects in mean–variance optimization. These include: (a) reduction in the out-of-sample risk-adjusted performance, (b) underestimation of risk of optimized portfolios, and (c) increased portfolio leverage and turnover.

Practitioners have long recognized the pitfalls of using the sample covariance matrix for portfolio optimization. Michaud (1989) characterized optimizers as "error maximizers," arguing that optimization tends to produce unstable and non-intuitive portfolios. He claimed that in many cases the simple equal-weighted portfolio (known as the $1/N$ portfolio) produces superior risk-adjusted performance than mean–variance optimization.

In this paper, we revisit this claim. In particular, we show that if the covariance matrix is estimated with a lot of sampling error, Michaud's claim is true. However, we also demonstrate that if the covariance matrix is properly constructed (i.e., with low noise), optimized portfolios easily outperform the $1/N$ portfolio.

One approach that practitioners have applied for overcoming these challenges is a statistical technique known as *shrinkage*. In this approach, the sample covariance matrix is first decomposed into a product of volatilities and correlations. Next, the sample correlation matrix is blended with a target correlation matrix, known in statistical parlance as the *shrinkage target*, or the *prior*. Finally, the covariance matrix is reconstructed by scaling the rows and columns of the "shrunk" correlation matrix with the original volatilities. Ledoit and Wolf (2004) showed that shrinkage helps mitigate the adverse effects of noise, thereby leading to more reliable optimized portfolios.

Although there are many shrinkage targets to choose from, a common choice is to shrink toward the identity matrix, which contains ones along the diagonal, and zeros on the off-diagonals. We refer to this approach as *naïve shrinkage*, since the identity matrix implies (rather unrealistically) that all assets are uncorrelated. Although we do not advocate using the identity matrix as a shrinkage target, we do demonstrate in this paper that such *shrinkage is beneficial from a portfolio-construction point of view*.

The problem of naïve shrinkage is that it causes systematic underprediction of correlations, which in turn leads to biased risk forecasts. In fact, Menchero and Li (2020) recently showed that shrinking correlations in this manner may produce large errors in risk forecasts. Moreover, they demonstrated that for purposes of estimating portfolio volatility, *the sample correlation matrix is very close to optimal*. We revisit some of their findings in this paper.

Hence, we face an apparent conundrum. On the one hand, the sample correlation matrix is close to optimal for risk forecasting, but unreliable for portfolio optimization. On the other hand, naïve shrinkage is beneficial for portfolio optimization, but leads to potentially large errors in risk forecasts.

In this paper, we point a way out of this conundrum. In particular, we show that by shrinking correlations toward a more realistic target, we obtain a covariance matrix that can be reliably used for *both* risk forecasting and for portfolio optimization.

The shrinkage target that we select is obtained using *principle component analysis* (PCA). Hence, we refer to our method as *PCA shrinkage*. The PCA technique, which is described in Appendix A, represents a statistical method for estimating correlations using a parsimonious set of factors (the principle components). We show that PCA shrinkage leads to estimated

correlations that never deviate far from the sample correlation (hence, accurate for risk forecasting), while producing a reliable covariance matrix for portfolio optimization.

PCA shrinkage can be readily applied to any correlation matrix. To illustrate the strength and versatility of our technique, we apply it to the construction of multi-asset-class (MAC) risk models. As described below, MAC risk models are constructed by aggregating multiple local factor models across different asset classes. As a result, MAC risk models typically contain a very large number of local factors (well over a thousand), the size of which presents a challenge for covariance matrix estimation.

Up to now, the leading technique in the industry for constructing MAC models was based on the so-called time-series method, originally developed by Shepard (2007/2008). This technique, described in Appendix A, tries to identify "global" factors capable of explaining the correlations of local factors.

In this paper, we empirically compare estimated factor correlations using the time-series method and PCA shrinkage. We show that PCA shrinkage produces estimated correlations that never deviate far from the sample correlation, which is essential for making reliable risk forecasts. By contrast, the time-series method tends to systematically underforecast factor correlations, effectively shrinking them toward zero. Moreover, the shrinkage intensity may be quite aggressive—especially across different asset classes. Such shrinkage may lead to potentially large errors in risk forecasts using the time-series method. Given the widespread use of MAC risk models across the industry, the improved technique for estimating factor correlations afforded by PCA shrinkage will likely be of great interest to practitioners.

1 Pitfalls of Using the Sample Correlation for Portfolio Optimization

In this section, we demonstrate the pitfalls of using the sample covariance matrix for mean–variance optimization. Moreover, we show how shrinkage can help mitigate these adverse effects, leading to improved out-of-sample risk-adjusted performance.

To better grasp the crux of the problem, we must borrow an important fact from linear algebra: *The sample covariance matrix is rank deficient whenever the number of assets N exceeds the number of time periods T used for estimation*. A feature of rank-deficient matrices is that they contain one or more zero eigenvalues. The eigenvalues are crucial from a risk perspective, since they represent the predicted variances of linear combinations of the original assets. In other words, rank-deficient matrices predict the existence of zero-volatility portfolios (composed entirely of risky assets). Of course, such "phantom" portfolios do not exist in reality. Given the opportunity, however, portfolio optimizers will "latch onto" these seemingly risk-free combinations of assets, leading to completely spurious portfolios with positive expected return and zero predicted risk.

Even when $T > N$, so that rank deficiency is formally averted, the sample covariance matrix may still suffer from "ill conditioning," which occurs when the smallest eigenvalues are *too small*. In this case, although there no longer exist "riskless" portfolios, the predicted risk of these small-eigenvalue portfolios may be far lower than their actual risk. While less egregious than full-blown rank deficiency, ill-conditioned matrices still cannot be reliably used for portfolio construction.

To illustrate these effects, and show the benefits of shrinkage, we briefly review some of the findings

of Menchero and Li (2020). In their study, they directly estimated the sample correlation matrix $\hat{\mathbf{C}}$ for a collection of assets. Next, they proceeded to shrink the correlation matrix toward the identity matrix $\mathbf{I}$,

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

where $\tilde{\mathbf{C}}_\lambda$ is the "shrunk" correlation matrix and λ is the *shrinkage intensity*, which varies from 0 to 1. Note that $\lambda = 0$ corresponds to the sample correlation matrix, whereas $\lambda = 1$ implies uncorrelated assets.

Next, they reconstructed the asset covariance matrix by scaling the rows and columns of $\tilde{\mathbf{C}}_\lambda$ by the estimated asset volatilities. The resulting covariance matrix was then used in mean–variance optimization to construct the minimum-volatility fully invested portfolio. The only constraint was that the asset weights sum to 100%.

The most rigorous way to evaluate the efficacy of covariance matrices for mean–variance optimization is to measure the *out-of-sample* volatility of optimized portfolios. In particular, under a common set of constraints, the covariance matrix that produces the lowest *out-of-sample* volatility is deemed most efficient for portfolio construction.

Note that the out-of-sample volatility of optimized portfolios is inversely proportional to the risk-adjusted performance. For instance, suppose an investor holds the view that all stocks have the same expected return. In this case, the maximum Sharpe ratio portfolio is the fully invested portfolio with lowest out-of-sample volatility. Higher volatility therefore translates into lower risk-adjusted performance.

To illustrate, we consider the 100 largest US equities (as of 31-Mar-2016) with complete return history back to January 1999. We use the first two years of data to compute the initial asset covariance matrix, which therefore leads to an out-of-sample testing period that spans 27-Dec-2000 to 31-Mar-2016.

To estimate the sample correlation matrix, we use exponentially weighted moving averages (EWMA), which assigns more weight to recent observations than to distant ones. The responsiveness of the model is determined by the half-life (HL) parameter. For instance, a 10-day HL means that we assign half the weight to the previous 10 days, one quarter of the weight to days 11–20, etc. We vary the HL parameter from 10 days to 500 days, and apply the same HL to both volatilities and correlations.

As we move deeper back in time, the weights exponentially decline, but never vanish. Hence, we always assign non-zero weight to observations more than 100 days ago, which implies that the correlation matrix never becomes rank deficient (i.e., $T > N$). However, as the HL parameter becomes smaller, we have fewer *effective* observations to estimate correlations, which leads to higher sampling error and more ill-conditioned (i.e., noisier) correlation matrices.

At the end of each trading day, we update the asset covariance matrix and use mean–variance optimization to construct the minimum-volatility fully invested portfolio. We then observe the out-of-sample portfolio return over the subsequent trading day. We repeat this process every day over the full-sample period, thus generating a long history of out-of-sample portfolio returns, from which we compute the realized volatility.

In Figure 1, we plot the annualized out-of-sample volatility of the optimized portfolios as a function of the HL parameter, for three different values of shrinkage intensity λ. We first consider the $\lambda = 0$ portfolio, which corresponds to the sample correlation. In this case, we find that the volatility depends strongly on the HL parameter. In particular, using short HL parameters (which cause

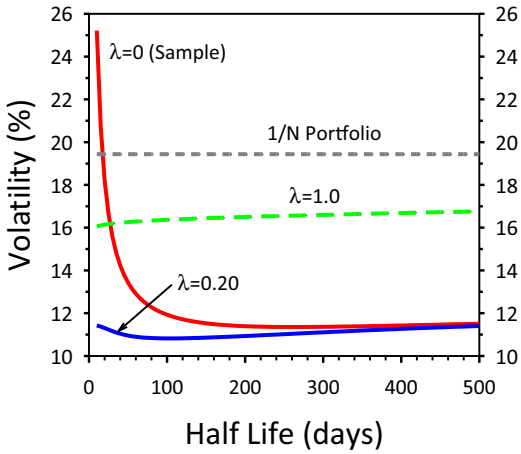

Figure 1 Out-of-sample volatility (annualized) versus HL for the minimum-volatility fully invested portfolio consisting of the largest 100 US equities. The optimal HL for the sample correlation ($\lambda = 0$) was 260 days, which produced a volatility of 11.35%. The optimal HL for $\lambda = 0.20$ was 105 days, which produced a volatility of 10.82%.

the correlation matrix to become ill-conditioned), the volatility rises sharply, reaching more than 25% for a 10-day HL. The optimal HL parameter for $\lambda = 0$ was 260 days, which produced an out-of-sample volatility of 11.35%.

Next, we examine the $1/N$ portfolio, which assigns equal weight (1%) to each of the 100 stocks. This portfolio would be considered optimal only under the unrealistic assumption that all stocks had the same volatility and zero correlation. The realized portfolio volatility was nearly 20%. For the vast majority of HL parameters, the $1/N$ portfolio had the *highest* volatility, which translates into the *lowest* risk-adjusted performance. The only exception occurred for very short HL parameters, in which case the sample correlation produced even higher volatilities. This corresponds to the regime in which optimizers behave as "error maximizers," as discussed by Michaud (1989). Note, however, that as the HL becomes sufficiently long and the correlation matrix gets reasonably well-conditioned, mean–variance optimization using the sample correlation easily outperforms the $1/N$ portfolio.

We now consider the $\lambda = 1$ portfolio, whose weights are inversely proportional to the variances. This portfolio is optimal under the assumption that all stocks are uncorrelated. For all ranges of HL parameters, the $\lambda = 1$ portfolio has significantly lower volatility than the $1/N$ portfolio, thus illustrating the benefit of underweighting high-volatility stocks. However, for HL parameters greater than 25 days, the sample correlation ($\lambda = 0$) produced even lower volatility. This shows that using well-conditioned covariance matrices, mean–variance optimization can effectively exploit the sample correlations to reduce portfolio risk. Finally, note that the $\lambda = 1$ portfolio has a slight upward slope. For instance, at an HL of 10 days, the annualized volatility is approximately 16%, versus nearly 17% using HL = 500 days. The reason for this gradual increase is that as the HL becomes longer, the volatility estimate uses increasingly stale data, which translates into less-efficient weights.

Finally, we consider the $\lambda = 0.20$ portfolio, which corresponds to 20% shrinkage. This portfolio had lower volatility than the sample correlation for all values of HL parameters. If the correlation matrix is well conditioned (e.g., 500-day HL), the benefit of shrinkage is fairly marginal. However, the real power of shrinkage is revealed when applied to ill-conditioned correlation matrices. For example, if the HL parameter is below say 50 days, the $\lambda = 0.20$ portfolio outperformed all other methods by a wide margin. The optimal HL parameter for $\lambda = 0.20$ was 105 days, which produced an out-of-sample volatility of 10.82%.

The astute reader may wonder why we make such a fuss over shrinkage. For instance, why not just estimate the sample correlation matrix with a 260-day HL and be done with it? Although this is a viable option for a 100-stock portfolio, it fails for

larger numbers of assets. The reason is that the conditioning of the matrix depends on the ratio of N/T, where N is the number of assets and T is the effective number of time periods. Hence, if a 260-day HL produces a well-conditioned correlation matrix for 100 stocks, then a 7,800-day HL (more than 30 years) would be required to produce a similarly well-conditioned matrix for a universe containing 3,000 stocks. In practice, such deep histories are not readily available. Moreover, even if the data were available, it would be of dubious value due to the staleness of information. Hence, shrinkage is an essential technique for solving real-world portfolio optimization problems with large numbers of assets.

The main conclusions of this section are two fold. First, for mean–variance optimization to produce reliable portfolios, it is necessary to feed in a well-conditioned correlation matrix. Second, shrinkage is an effective tool for producing well-conditioned correlation matrices when the number of assets is large relative to the number of effective time periods.

2 Pitfalls of Using Naïve Shrinkage for Risk Forecasting

The previous section showed that using the sample correlation matrix produced poor results for portfolio optimization, while naïve shrinkage led to more reliable optimal portfolios. In this section, we show that the opposite is true when it comes to risk forecasting. That is, *the sample correlation is close to optimal for risk forecasting*, whereas naïve shrinkage leads to large errors.

Menchero and Li (2020) studied the impact of naïve shrinkage on the accuracy of risk forecasts. In this section, we apply their findings to illustrate the errors introduced by naïve shrinkage on the risk forecasts of asset-pair portfolios. We construct the portfolios by taking a long position in one asset and a short position in the other asset. Each asset is assumed to have unit volatility ($\sigma = 1$), true correlation ρ, and follow a stationary return distribution.

The true variance of the asset-pair portfolio is easily computed as $\sigma_\rho^2 = 2(1-\rho)$. Of course, true variances and correlations are unobservable. In practice, they must be estimated using a finite number of return observations, which introduces sampling error.

We simulate τ periods of stock returns from the true distribution, and use the simulated returns to *estimate* volatilities and correlations. Let $\hat{\sigma}_A^2$ denote the estimated variance of asset A, $\hat{\sigma}_B^2$ denote the corresponding variance for asset B, and let $\hat{\rho}$ be the sample correlation.

Next, we shrink the estimated correlation toward zero, i.e., $\hat{\rho}_\lambda = (1-\lambda)\hat{\rho}$, where λ is the shrinkage intensity. Using the variance estimates and the shrunk correlation, the predicted variance of the asset-pair portfolio becomes

$$\hat{\sigma}_{\tau\rho\lambda}^2 = \hat{\sigma}_A^2 + \hat{\sigma}_B^2 - 2\hat{\rho}_\lambda \hat{\sigma}_A \hat{\sigma}_B. \tag{2}$$

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

We define the relative forecast error as the difference between the estimated variance and the true variance, normalized by the true variance. Hence, the root-mean-squared (RMS) error is given by

$$\varepsilon_{\tau\rho\lambda} = \sqrt{E\left[\left(\frac{\hat{\sigma}_{\tau\rho\lambda}^2 - \sigma_\rho^2}{\sigma_\rho^2}\right)^2\right]}, \tag{3}$$

which measures the expected magnitude of the relative forecast error.

Menchero and Li (2020) derived an analytic expression for both the RMS error ($\varepsilon_{\tau\rho\lambda}$) and the optimal shrinkage intensity ($\tilde{\lambda}$). In Figure 2, we apply this analytic formula to plot the RMS error

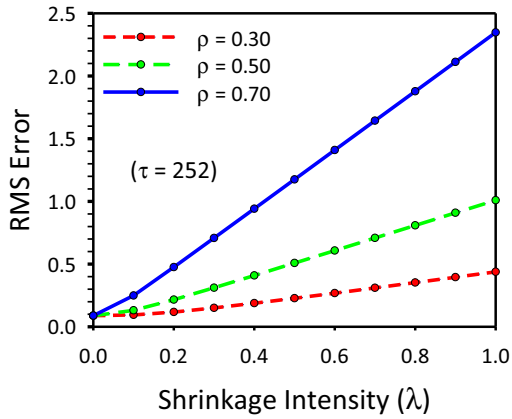

Figure 2 RMS error versus shrinkage intensity for asset-pair portfolios. The estimation window was 252 days, and three different values of correlation are considered.

$\varepsilon_{\tau\rho\lambda}$ versus shrinkage intensity λ. In this exercise, we considered three values of the true correlation ρ (0.3, 0.5, and 0.7), and used $\tau = 252$ 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 used for predicting correlations in practice.

One interesting feature of Figure 2 is that, independent of correlation, the RMS error converges to the same value as the shrinkage intensity goes to zero. This limiting value is given by $\sqrt{2/\tau}$, which represents the standard error for a variance forecast estimated using τ return observations. For $\tau = 252$, this works out to an RMS error of $\varepsilon = 0.08909$ (i.e., 8.9%).

Casual inspection of Figure 2 immediately points to an optimal shrinkage intensity of zero. That is, the sample correlation ($\lambda = 0$) appears to minimize RMS error. Although this is *nearly* true, it is not *exactly* correct. In fact, Menchero and Li (2020) showed that for any finite value of τ, there is always an optimal (non-zero) shrinkage intensity. Moreover, they provided an analytic formula for the optimal shrinkage intensity. Applying their result using $\rho = 0.30$ and $\tau = 252$, we find an optimal shrinkage intensity of $\tilde{\lambda} = 0.021$, which in turn produces a minimum error of $\tilde{\varepsilon} = 0.08863$. Hence, the reduction in RMS error resulting from optimal shrinkage (versus no shrinkage) is only $\Delta\varepsilon = 0.00046$. This represents an error reduction of less than 0.05% (i.e., completely immaterial).

For larger correlations, the reduction in error from optimal shrinkage is even smaller. For instance, using $\rho = 0.70$ and $\tau = 252$, the optimal shrinkage intensity is $\tilde{\lambda} = 0.007$, which leads to a minimum error of $\tilde{\varepsilon} = 0.08907$. This represents a microscopic reduction in error of only $\Delta\varepsilon = 0.00002$ relative to the sample correlation. These results show that for all practical purposes, *the sample correlation is the optimal estimate for purposes of predicting risk*.

Two other features of Figure 2 are worth highlighting. First, we see that for large values of shrinkage intensity, the RMS error is approximately linear in λ. Second, we observe that for a given shrinkage intensity, the RMS error rises dramatically with increasing correlation. For example, using $\tau = 252$ and $\lambda = 0.6$, we find an RMS error of 0.27 for $\rho = 0.3$, which rises to 0.61 for $\rho = 0.5$, which further increases to 1.41 (i.e., 141%) for $\rho = 0.7$.

The common explanation for both effects is that for large shrinkage intensity, the RMS error is dominated by bias. Taking the limit $\tau \to \infty$ (which eliminates sampling error), we find,

$$\lim_{\tau\to\infty}(\varepsilon_{\tau\rho\lambda}) = \frac{\lambda\rho}{1-\rho}. \tag{4}$$

This result shows that when the noise is small relative to the bias, the RMS error is proportional to λ, with the constant of proportionality increasing with increasing correlation. For small values of τ and λ, the bias and noise may be of comparable magnitude, in which case the RMS error is determined by the interplay between noise and bias.

To summarize, these results show that when it comes to predicting volatility, the sample correlation is for all practical purposes the optimal estimator. Moreover, the effect of naïve shrinkage is to introduce large biases in the estimated correlations, which in turn translates into large errors in risk forecasts.

3 PCA Shrinkage

In this section, we describe our technique for estimating correlation matrices. Our method can be applied to any correlation matrix, regardless of the number of assets N or the number of time periods T. In particular, our method is applicable even in the case of highly rank-deficient correlation matrices ($N \gg T$).

We develop our technique with two explicit objectives in mind. First, the correlation matrix must be well-conditioned so that it can be reliably used for portfolio optimization. Second, in order to produce accurate risk forecasts, the estimated correlations must deviate minimally from the sample correlation.

Our method represents an extension of the approach taken by Ledoit and Wolf (2004), who applied shrinkage techniques to obtain robust estimates of the covariance matrix for portfolio optimization. The basic technique is to estimate the sample correlation matrix and blend it with a target correlation matrix, called the shrinkage target.

Ledoit and Wolf (2004) considered two shrinkage targets. The first target was the constant correlation matrix, which assumes that all assets have the same correlation (given by the average pair-wise correlation across all assets). The second shrinkage target was the one-factor model of Sharpe (1963), which assumes that asset correlations are fully explained by common exposure to the market factor.

Although these shrinkage targets may work well for portfolio optimization, they lead to potentially large biases in correlation estimates. As we saw in the previous section, such biases may be very detrimental to the accuracy of risk forecasts.

To avoid such biases, we select a different shrinkage target for our correlation matrix. More specifically, our shrinkage target is based upon *principal component analysis* (PCA), which is a statistical technique for estimating correlation matrices.

As described in Appendix A, the "principle components" represent statistical factors that are extracted directly from the return time series. The return component not explained by the principal components is assumed to be idiosyncratic and uncorrelated with all other factors and idiosyncratic returns.

The correlation matrix using PCA shrinkage may be expressed as

$$\tilde{\mathbf{C}}(\lambda, J) = (1 - \lambda)\mathbf{C}_0 + \lambda\mathbf{C}_P(J), \qquad (5)$$

where now $\mathbf{C}_0$ is the sample correlation matrix, $\mathbf{C}_P(J)$ is the PCA correlation matrix, J is the number of principle components, and λ is the shrinkage intensity.

To illustrate the benefits of PCA shrinkage from a portfolio-construction perspective, we revisit our previous example. Namely, we consider the largest 100 US stocks as of 31-Mar-2016, with complete daily return history back to 13-Jan-1999. We directly compute the 100×100 sample covariance matrix using EWMA with daily returns for two different half-life parameters (20 days and 100 days). The first covariance matrix is produced on 27-Dec-2000, and the out-of-sample period runs from 27-Dec-2000 to 31-Mar-2016. We shrink the sample correlation matrix toward the PCA correlation matrix, which was constructed using a single principal component ($J = 1$). We update the covariance

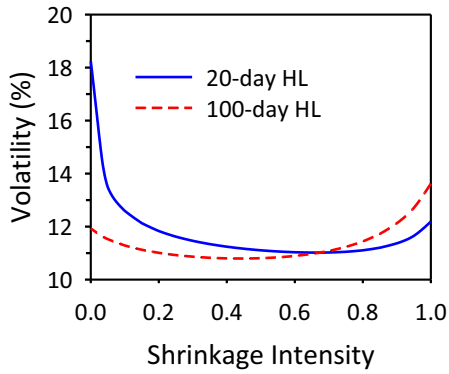

Figure 3 Out-of-sample volatility (annualized) versus shrinkage intensity for the minimum-volatility fully invested portfolio containing the 100 largest US stocks. The optimal shrinkage intensity for 20d HL was $\lambda = 0.65$, which produced a volatility of 11.02%. The optimal shrinkage intensity for 100d HL was $\lambda = 0.45$, which produced a volatility of 10.79%.

matrix on a daily basis, and use it to form the minimum-volatility fully invested portfolio each day.

In Figure 3, we plot the out-of-sample volatilities as a function of shrinkage intensity. For zero shrinkage (i.e., using the sample correlation), we find that the reasonably well-conditioned 100-day HL produced much lower volatility (11.93%) than the poorly conditioned 20-day HL (18.23%). These results are consistent with Figure 1, and can be understood by considering the interplay between noise and bias. Although the 20-day HL has lower bias (being based on recent data) than the 100-day HL (which uses stale data), the dominant effect for the sample correlation is *noise*. Hence, the 100-day HL, which is less plagued by noise, easily outperforms the 20-day HL.

For full shrinkage ($\lambda = 1$), which corresponds to pure PCA, the results are flipped. That is, the 20-day HL produced a lower volatility (12.19%) than the 100-day HL (13.62%). These results can be understood by again considering the effects of noise and bias. First, the noise associated with the first principle component is quite small for both the 20-day HL and the 100-day HL. That is because even a 20-day HL is very long relative to the number of factors, which is equal to one in this case. Second, the bias due to non-stationarity is considerably larger for the 100-day HL (which uses stale data) than for the 20-day HL. For pure PCA using a single principle component, the dominant effect is *bias*, which explains why the 20-day HL produced lower volatility.

The full benefit of PCA shrinkage is found by considering intermediate values of λ. First we study the behavior of the 20-day HL. We see that for small shrinkage intensity, the volatility of the 20-day HL drops very quickly. For instance, the volatility drops from 18.23% for the sample correlation to only 13.48% using a shrinkage intensity $\lambda = 0.05$. The reason for the sudden drop is that the 20-day HL is dominated by noise, and hence sees the biggest benefit in noise reduction through even a small PCA blending. Note that the optimal shrinkage intensity for the 20-day HL is $\lambda = 0.65$, which produced a volatility of 11.02%.

For the 100-day HL, we find that PCA shrinkage produces a more modest drop in volatility. For example, the volatility drops from 11.93% for no shrinkage to 11.53% for $\lambda = 0.05$. The reason for the smaller drop in volatility is that the 100-day HL has relatively low noise to begin with, so does not benefit as much as the 20-day HL. The optimal shrinkage intensity for the 100-day HL was $\lambda = 0.45$—significantly smaller than the $\lambda = 0.65$ optimal shrinkage found using the 20-day HL. This makes intuitive sense, since the 100-day HL was better conditioned to start with; hence, it requires less shrinkage. Finally, note that the 100-day HL with optimal PCA shrinkage produced a volatility of 10.79%, slightly below the 10.82% volatility obtained using naïve shrinkage in Figure 1.

4 Application of PCA Shrinkage to MAC Models

In this section, we discuss the construction of multi-asset-class (MAC) risk models. We begin with a brief overview of individual local factor models. We then discuss how these local factor models are linked together to create an MAC model. We consider two distinct ways for linking the local models. The first is based on PCA shrinkage, whereas the second approach uses the time-series method of Shepard (2007/2008). Finally, we describe the process of *model integration*, which ensures that the local blocks of the MAC model are consistent with the local models.

Local factor models. Before delving into MAC models, it is useful to briefly review traditional factor models. Traditional factor models, also known as *local* factor models, were pioneered by Rosenberg (1974) for the purpose of constructing robust estimates of the asset covariance matrix (i.e., suitable for portfolio optimization). Direct estimates of the sample covariance matrix are not viable, since the number of assets N may far exceed the number of time periods T, thus leading to rank-deficient matrices.

Factor models posit that security returns are driven by a parsimonious set of factors, which capture the co-movement of asset returns. The number of factors K in a local model typically numbers around several dozen. The component of return unexplained by the factors is called the *idiosyncratic* or *specific* return, which is assumed to be uncorrelated with all factors and the specific returns of all other assets. Using factor models, the full asset covariance matrix can be estimated using the far smaller *factor* covariance matrix, together with the factor exposure matrix and specific risk forecasts for each asset. As long as the factors have been properly identified and the number of factors is "small" compared with the number of time periods ($K \ll T$), the resulting asset covariance matrix is well-conditioned and suitable for use in portfolio optimization.

Structure of MAC models. MAC models are constructed by linking the covariance matrices of multiple *local* factor models. In particular, this requires estimating the off-diagonal correlations between local factors in different local models. The local models are designed to explain the risk of individual markets within a particular asset class. For example, within the equity asset class, we have local models corresponding to markets in the US, Europe, Japan, and so on. Each local model is described by a set of several dozen local factors, which includes market factors, styles factors, industry factors, and in the case of multi-country markets (e.g., Europe), country factors.

The factor correlation matrix associated with an individual local model is termed a *local block*. Collectively, the full suite of local models covers all global equities. Hence, we refer to the correlation matrix of all local factors across all equity models as the *global equity block*, which typically contains several hundred local equity factors.

In addition to the global equity block, MAC models contain global blocks for fixed income, commodities, and currencies. For instance, the fixed-income block contains hundreds of local factors to capture the risk of multiple bond markets, such as US Municipals, Emerging Markets, and European Credit.

The collection of all local factors across all asset classes is called the *universal block*, which typically contains a large number of factors (over a thousand). The main task in building an MAC model is to compute the correlation matrix of the universal block. Since the total number of factors is large relative to the effective number of time periods, the sample correlation will be rank deficient and unreliable for use in portfolio

optimization. Our goal is to find an estimate of the universal correlation matrix that deviates minimally from the sample correlation matrix (to ensure accurate risk forecasts), while still producing a well-conditioned matrix suitable for portfolio optimization.

We use PCA shrinkage to estimate the factor correlation matrices of the individual local blocks. Let m denote an individual local block (containing K_m factors) and let $\tilde{\mathbf{C}}^L_{mm}$ be the $K_m \times K_m$ local factor correlation matrix, estimated using Equation (5). The matrix $\tilde{\mathbf{C}}^L_{mm}$ represents our best estimate of the factor correlation matrix for the individual local model m. In particular, this matrix deviates minimally from the sample correlation, yet is well-conditioned for use in portfolio optimization. We repeat this process for all local blocks.

Next, we must estimate the correlations between factors belonging to different local blocks. Let $K = \sum_{m=1}^{M} K_m$ be the total number of factors in the global equity block, where M is the total number of local equity models. Let $\tilde{\mathbf{C}}_G$ denote the $K \times K$ global equity correlation matrix. Let $\tilde{\mathbf{C}}^G_{mn}$ denote the $K_m \times K_n$ correlation matrix for the off-diagonal block linking models m and n.

In this paper, we compare and contrast two different ways of estimating the global equity correlation matrix $\tilde{\mathbf{C}}_G$. The first method is based on PCA shrinkage. In this approach, we simply compute the $K \times K$ sample correlation matrix of the global equity block, and blend it with the corresponding $K \times K$ PCA correlation matrix. We establish the model parameters λ and J by minimizing the out-of-sample volatilities of portfolios optimized to have lowest risk, subject to the constraint of unit exposure to a given factor. We find this approach to model calibration to be very robust, producing similar optimal parameters across different sub-periods and different equity markets.

Time-series method. The second approach that we analyze is based on the time-series method, which was introduced by Shepard (2007/2008) as a way of estimating the global correlation matrix $\tilde{\mathbf{C}}_G$. In this approach, global factor returns (also called "core" factors) are specified as explanatory variables for purposes of estimating the correlation of local factors. In particular, a time-series regression of local factor returns against the global factor returns is used to estimate the sensitivities (loadings) of the local factors onto the global factors. Given the covariance matrix of the global factor returns, and the sensitivities of the local factors to the global factors, the factor-pair correlations can be inferred. Technical details are provided in Appendix A.

For equities, the core factors used in the time-series method are given by the factor returns from a global equity factor model containing a market factor, countries, industries, and styles. Global factor returns are estimated by weekly cross-sectional regression. To better understand the intuition behind the time-series method, consider the correlation between the US and European Momentum factors. Clearly, both of these local factors will be correlated with the Global Momentum factor, which will partially capture the correlation between the two local factors. As another example, consider the correlation between the US and Japanese Automobile factors. Again, part of this correlation will surely be explained by common exposure to the Global Automobile factor.

For other asset classes, such as fixed income or commodities, the time-series method typically specifies the global factor returns as weighted averages of local factor returns. For instance, a core factor describing a parallel shift in US interest rates can be obtained by taking the average change in key rates (local factors) at selected points along the US yield curve.

A potential pitfall of the time-series approach is that there may be a large subjective element in specifying the global factors. If important factors have been omitted, correlations may be underestimated.

The time-series technique increases in complexity when applied *across* asset classes. In this case, an even more parsimonious set of factors, called "core-of-core" factors, are used to estimate the correlations between core factors in different asset classes. These core-of-core factors are typically constructed by taking weighted averages of the core factors. Again, a large subjective element is involved in identifying the core-of-core factors, and omission of important factors may lead to underestimation of correlations.

Model integration. We have considered two distinct ways of estimating $\tilde{\mathbf{C}}_G$. The first method is based on PCA shrinkage, whereas the second approach uses the time-series method.

Regardless of which method we employ, we face a significant problem. More specifically, the diagonal blocks of the global correlation matrix $\tilde{\mathbf{C}}_G$ will not match our best estimates of the correlation matrices for these local blocks: i.e., $\tilde{\mathbf{C}}^G_{mm} \neq \tilde{\mathbf{C}}^L_{mm}$.

This is problematic for two reasons. First, it introduces inconsistencies between risk forecasts using the global model and the local model. Second, $\tilde{\mathbf{C}}_G$ contains inferior estimates along the diagonal blocks, which is clearly unappealing to model users. This problem is resolved by the technique of *model integration*, as we now discuss.

The aim of model integration is to recover the local correlation matrices $\tilde{\mathbf{C}}^L_{mm}$ along the diagonal blocks of the global correlation matrix $\tilde{\mathbf{C}}_G$. One easy way of accomplishing this is to simply "cut-and-paste" $\tilde{\mathbf{C}}^L_{mm}$ onto the diagonal blocks of $\tilde{\mathbf{C}}_G$. Unfortunately, this simplistic method does not ensure that the resulting matrix would be positive definite, thus violating the most fundamental property of a covariance matrix.

In Appendix B, we show the proper way to "integrate" the local correlation matrices $\tilde{\mathbf{C}}^L_{mm}$ into the global correlation matrix $\tilde{\mathbf{C}}_G$, such that we recover our best estimates along the diagonal blocks. This procedure ensures that the reconstituted matrix is positive definite, which is crucial for any risk model. After the global correlation matrix has been reconstituted in this fashion, the resulting matrix becomes our best estimate for the global block.

The same technique is applied to integrate the global blocks into the universal block. That is, we first employ either PCA shrinkage or the time-series method to estimate the universal correlation matrix. This matrix is then reconstituted using the technique of model integration to replicate our best estimates of the global blocks for each asset class. Since our best estimate of the global block already contains $\tilde{\mathbf{C}}^L_{mm}$ along the diagonal blocks, this ensures that the MAC risk model uses our best estimates of the correlation matrix at any level of the hierarchy.

5 Empirical Results

We have seen that to ensure accurate risk forecasts, estimated correlations should deviate minimally from the sample correlation. In this section, we investigate the magnitude of these deviations. In particular, we provide side-by-side comparisons between estimated correlations using PCA shrinkage versus the time-series method.

In our empirical study, the sample correlation was estimated using a 100-week HL. The time-series method also employed a 100-week HL, both to estimate the factor loadings in the time-series regression and to estimate the covariance matrix of core factors.

Cross-market equity factor correlations. We first study cross-market factor correlations within the equity asset class. As our data set, we use the nine local equity blocks found in the Bloomberg second-generation MAC risk model (MAC2). The local models include: (1) Asia, (2) Australia, (3) Canada, (4) China, (5) Emerging Europe, Middle East, and Africa, (6) Developed Europe, (7) Japan, (8) Latin America, and (9) USA. All together, these nine models contain 387 factors and span the global equity markets.

Each single-country model (e.g., Australia) contains three types of factors: (1) a market factor, to which all stocks have unit exposure, (2) industry factors, whose exposures are given by (0,1), and (3) style factors, whose exposures are standardized to be cap-weighted mean zero with unit cross-sectional variance. In addition to the three types of factors found in single-country models, multi-country models (e.g., Europe) also contain country factors, with exposures given by (0,1).

We estimate factor returns using weekly cross-sectional regression, with regression weights proportional to the square root of market capitalization. Note that the sum of the industry factor exposures yields a column of ones, which is identical to the market factor exposure. Hence, the factor structure contains an exact collinearity. For multi-country models, there is a second exact collinearity, since the sum of country factor exposures also yields a column of ones. To obtain a unique regression solution, we apply the method of Heston and Rouwenhorst (1994). Namely, we constrain the cap-weighted sum of industry factor returns to equal zero. For multi-country models, we impose the additional constraint that the cap-weighted sum of country factor returns is zero.

In Figure 4, we present scatterplots (as of 21-Mar-2018) of the estimated correlations versus the sample correlation for the off-diagonal equity blocks. The left panel contains results for the time-series method, whereas the right panel is obtained using PCA shrinkage. Each dot represents a pair of factors across different equity markets (e.g., Japan Automobiles versus US Momentum). There are roughly 66,000 such factor pairs.

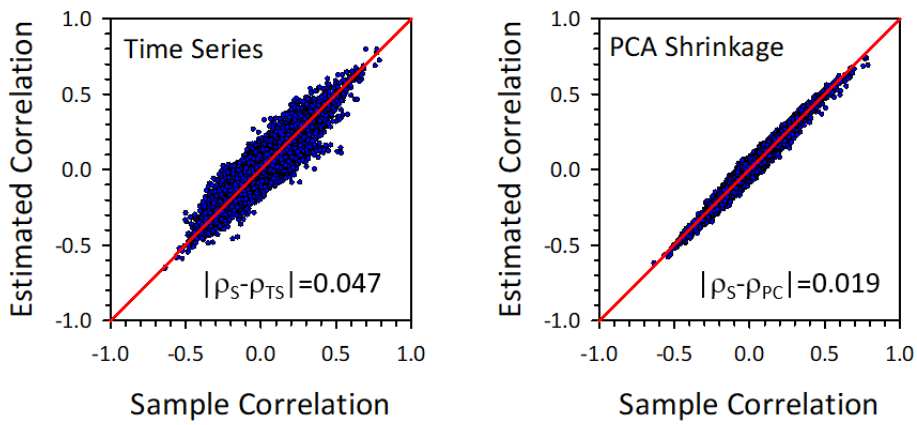

Figure 4 Scatterplots of estimated correlation versus sample correlation. Left panel is for the off-diagonal equity blocks using the time-series method. Right panel is for the off-diagonal equity blocks using PCA shrinkage. The analysis date for all plots was 21-Mar-2018. The correlation HL was 100 weeks. The mean absolute deviation between the estimated correlation and the sample correlation was 0.047 for the time-series method, versus 0.019 using PCA shrinkage.

The most obvious feature of Figure 4 is that PCA shrinkage provides a much better fit to the sample correlation than does the time-series method. That is, the time-series estimated correlations often deviate significantly from the sample correlation, which in turn translates into higher risk-forecasting errors.

We also computed the mean (across all factor pairs) absolute value of the difference between the estimated correlation and the sample correlation. Using the time-series method, this mean absolute deviation was 0.047. By contrast, using PCA shrinkage, the mean absolute deviation was only 0.019. Hence, PCA shrinkage produced estimated correlations with a much tighter fit to the sample correlation.

Although obscured by the high density of points, another important feature from Figure 4 is that the time-series method tends to systematically underforecast the magnitude of the correlations. For example, of the roughly 33,000 factor pairs with positive sample correlation, 81% of the points lie below the 45-degree line using the time-series method. By contrast, using PCA shrinkage, 52% of the points lie above the 45-degree line, and 48% lie below. In other words, the time-series method underestimated 81% of the correlations, whereas using PCA shrinkage the likelihood of overestimation and underestimation is evenly balanced.

Equity/commodity factor correlations. We now consider correlations across different asset classes. In this example, we study the correlation between the Brent Crude Oil factor from the MAC2 Commodity model, and the Energy factor from the MAC2 US equity risk model. Naturally, we expect these two factors to exhibit a strong positive correlation.

In Figure 5, the dashed green line represents the realized sample correlation versus time. As

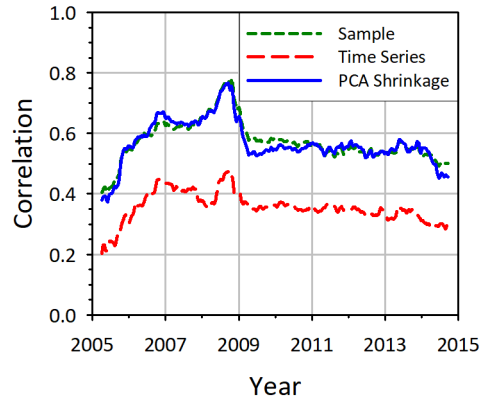

Figure 5 Correlation versus time between the Brent Crude Oil factor from the MAC2 Commodity model and the energy factor from the MAC2 US equity risk model.

expected, the realized correlation was quite high. For instance, at the start of the period it was 0.40 and reached nearly 0.80 in late 2008.

The solid blue line in Figure 5 is the estimated correlation using PCA shrinkage. As we see, the estimated correlation never deviated far from the sample correlation. This result held during all market cycles, whether in times of placid markets or turbulent markets.

The dashed red line shows the estimated correlation using the time-series method. This plot makes clear that the time-series method persistently underforecasts correlations. For instance, at the start of the period, the sample correlation was 0.40, whereas the estimated correlation using the time-series method was only 0.20. This represents an effective shrinkage intensity of 50%. As we have seen from Figure 2, such aggressive shrinkage may lead to large errors in risk forecasts.

Equity/bond factor correlations. We now consider correlations between the 387 equity factors described above and a set of 431 fixed-income factors from the Bloomberg MAC2 risk model. The number of pair-wise correlations is very large (166,797). The 431 fixed-income factors

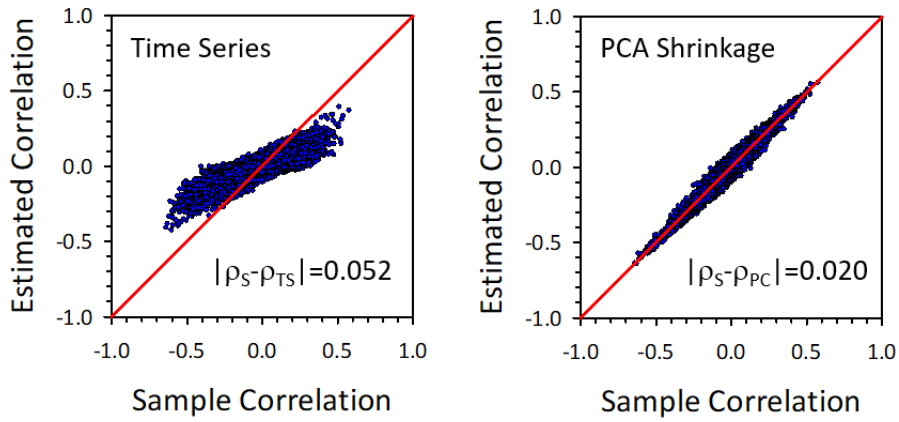

Figure 6 Scatterplots of estimated correlation versus sample correlation for 387 equity factors and 431 fixed-income factors. Left panel is for the time-series method. Right panel is using PCA shrinkage. The analysis date was 21-Mar-2018 and the correlation HL was 100 weeks. The mean absolute deviation between the estimated correlation and the sample correlation was 0.052 for the time-series method, versus 0.020 using PCA shrinkage.

span the following bond markets: (1) Australia, (2) Canada, (3) Emerging Markets, (4) Europe, (5) UK, (6) Japan, and (7) USA.

In Figure 6, we show correlation scatterplots of equity factors against fixed-income factors on analysis date 21-Mar-2018. The left panel shows empirical results for the time-series method. This figure dramatically illustrates the extent to which the time-series method systematically underforecasts correlations. In effect, the time-series method amounts to an aggressive shrinkage of correlations toward zero. Of the more than 82,000 factor pairs that had positive sample correlation, the time-series method underestimated 93% of them.

On the right panel, we show the corresponding scatterplot using PCA shrinkage. Again, the points all line up neatly along the 45-degree line, with the estimated correlations never deviating far from the sample correlation. The mean absolute deviation of the estimated correlation from the sample correlation was 0.020 for PCA shrinkage, versus 0.052 for the time-series model.

Skeptical readers may wonder whether perhaps we are "cherry picking" the examples. For instance, perhaps the time-series method "just happened" to perform poorly on analysis date 21-March-2018. To investigate this possibility, we fit a regression line (with no intercept) through the cloud of points,

$$\hat{C}_{nt} = \beta_t C_{nt} + \delta_{nt}, \tag{6}$$

where $\hat{C}_{nt}$ is the estimated correlation for factor-pair n at time t, C_{nt} is the corresponding sample correlation, β_t is the slope coefficient at time t, and δ_{nt} is the unexplained residual. Ideally, we would like to find a slope coefficient close to 1, which indicates minimal shrinkage.

In Figure 7, we plot the slope coefficient β_t versus time for PCA shrinkage and for the time-series method. We find that for PCA shrinkage, the slope coefficient is very close to 1 during all periods of time. By contrast, the slope coefficient using the time-series method ranges from 0.2 to 0.3. In other words, the time-series method represents a persistent shrinkage intensity of between 70% and 80%. From Figure 2, we see that shrinkage of this magnitude may induce large errors in risk forecasts.

Skeptical readers may also wonder whether the time-series method "just happens" to perform

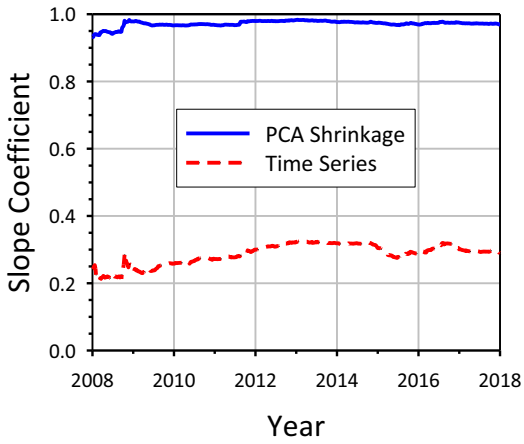

Figure 7 Slope coefficients versus time for correlations between 387 equity factors and 431 fixed-income factors. The correlation HL was 100 weeks.

poorly for correlations between equity factors and fixed-income factors. To explore this possibility, we made scatterplots on 21-Mar-2018 of factors across all major asset classes in the Bloomberg MAC2 model, including 387 equity factors, 431 fixed-income factors, and 222 commodity factors. The commodity model includes Shift, Twist, and Butterfly factors to explain the return variability of futures contracts for Agriculture, Oil, Electricity, Metals, and Natural Gas. For each pair of asset classes, we fit a regression through the cloud of points, as in Equation (6).

In Table 1, we report the slope coefficients across all asset classes using both the time-series method and PCA shrinkage. The first three columns report slope coefficients for the three asset classes using the time-series method. The time-series approach performs least badly on the equity/equity correlations, but even in this case the slope coefficient is equal to 0.75, which amounts to a shrinkage intensity of 25%. For bond/bond correlations, the time-series method produced a slope coefficient of 0.62, which represents a nearly 40% shrinkage intensity. The commodity/commodity slope coefficient was only 0.31. Finally, note that the slope coefficients *across* asset classes ranged from 0.2 to 0.3, which represents a shrinkage intensity of roughly 70% to 80%.

By contrast, using PCA shrinkage, the slope coefficients were all very close to 1. The smallest slope coefficient was 0.91, between equities and commodities. These results show the power and versatility of PCA shrinkage to accurately forecast correlations, even across different asset classes.

6 Summary

Covariance matrices are used for two basic purposes in finance: (1) to forecast portfolio risk, and (2) to construct portfolios using mean–variance optimization. We have shown that the sample correlation matrix is unreliable for portfolio optimization, but is very close to ideal for purposes

Table 1 Slope coefficients of scatterplot of sample correlation (x-axis) versus estimated correlation (y-axis), for analysis date 21-Mar-2018. The equity block contained 387 factors, the fixed-income block had 431 factors, while the commodity block had 222 factors. The correlation HL was 100 weeks in all cases.

	Time series method			PCA shrinkage		
Asset Class	Equities	Fixed	Comm	Equities	Fixed	Comm
Equities	0.75	0.29	0.29	1.00	0.97	0.91
Fixed income	0.29	0.62	0.21	0.97	1.01	0.94
Commodities	0.29	0.21	0.31	0.91	0.94	0.95

of forecasting risk. On the other hand, we have found that naively shrinking correlations toward zero is beneficial for portfolio optimization, but detrimental for purposes of forecasting portfolio risk.

In this paper, we have presented a new technique for estimating correlation matrices. Our approach is based on shrinking the sample correlation matrix toward the PCA correlation matrix. Using this technique—known as PCA shrinkage—results in correlation matrices that are well-conditioned and can be reliably used for portfolio optimization. Moreover, the estimated correlations never deviate far from the sample correlation, which means that it can be reliably used for risk forecasting as well.

As an illustration, we applied our technique to the construction of MAC risk models, which contain large numbers of local factors. Up to now, the prevailing technique for estimating factor correlations in MAC models was the so-called time-series method. In this paper, we showed that the time-series technique effectively shrinks correlations aggressively toward zero, which in turn may induce large errors in risk forecasts. By contrast, we showed that the correlation estimates using PCA shrinkage deviate minimally from the sample correlation, thus translating into more accurate risk forecasts.

Appendix A

In this Appendix, we provide mathematical details on two techniques for estimating correlation matrices. The first approach is based on PCA shrinkage, while the second technique employs the time-series method.

A.1. PCA shrinkage

We first describe how the sample correlation matrix is estimated. Let f_{jt} be the return to factor j over period t. Over a short horizon, such as one week, it is generally a good approximation to assume that the mean factor return is insignificant relative to the standard deviation. Under this assumption, the sample covariance matrix elements are computed as

$$F_{jk} = \sum_t w_t f_{jt} f_{kt}, \tag{A.1}$$

where w_t is an exponential weight. The exponential weights, which sum to 1, are characterized by their HL parameter. The sample variance of factor j is given by

$$\sigma_j^2 = \sum_t w_t f_{jt}^2, \tag{A.2}$$

with a corresponding expression holding for factor k. The sample correlation matrix elements are computed as

$$C_{jk} = \frac{F_{jk}}{\sigma_j \sigma_k}, \tag{A.3}$$

where the diagonal elements, by construction, are equal to 1. Note that j and k vary from 1 to K, where K is the total number of factors. In matrix notation, we use $\mathbf{C}_0$ to denote the $K \times K$ sample correlation matrix whose elements are given by Equation (A.3).

Next, we discuss how to estimate the PCA correlation matrix. Let $\mathbf{U}_0$ denote the $K \times K$ rotation matrix whose column vectors are given by the eigenvectors of sample correlation matrix $\mathbf{C}_0$, arranged in decreasing order. These eigenvectors, also known as *eigenfactors*, represent portfolios of local factors. The sample correlation matrix, expressed in the diagonal basis, is given by

$$\mathbf{D}_0 = \mathbf{U}_0' \mathbf{C}_0 \mathbf{U}_0, \tag{A.4}$$

where the diagonal elements (i.e., the eigenvalues) of $\mathbf{D}_0$ represent the predicted variances of the eigenfactor portfolios. Note that if the number of factors K exceeds the number of time periods T, then matrix $\mathbf{D}_0$ will be rank deficient, with $(K-T)$ zero eigenvalues.

Assume that only J principal components are required to effectively capture the cross-sectional variation of local factor returns, where $J < K$. Let $\tilde{\mathbf{D}}_0$ denote the $J \times J$ diagonal block of $\mathbf{D}_0$, which consists of the J largest eigenvalues arranged in decreasing order. Similarly, let $\tilde{\mathbf{U}}_0$ represent the $K \times J$ matrix obtained by keeping only the first J columns of $\mathbf{U}_0$. The portion of the sample correlation matrix explained by the J principal components is given by

$$\tilde{\mathbf{C}}_0(J) = \tilde{\mathbf{U}}_0 \tilde{\mathbf{D}}_0 \tilde{\mathbf{U}}_0'. \tag{A.5}$$

Note that if $J = K$, we recover the sample correlation matrix $\mathbf{C}_0$. For $J < K$, however, the principle components do not fully explain the cross-sectional variation of local factors. In this case, the diagonal elements of $\tilde{\mathbf{C}}_0(J)$ are less than 1, which implies that $\tilde{\mathbf{C}}_0(J)$ cannot be interpreted as a correlation matrix.

This discrepancy is resolved by noting that we must also account for the portion of return variation *not* explained by the principal components. Since each of the original K factors has unit variance by construction, the unexplained variance of an individual factor k is given by

$$\Delta_{kk} = 1 - \text{diag}_k[\tilde{\mathbf{C}}_0(J)]. \tag{A.6}$$

Let $\mathbf{\Delta}$ denote the $K \times K$ diagonal matrix whose elements are given by Δ_{kk}. The PCA correlation matrix is expressed as

$$\mathbf{C}_P(J) = \tilde{\mathbf{U}}_0 \tilde{\mathbf{D}}_0 \tilde{\mathbf{U}}_0' + \mathbf{\Delta}. \tag{A.7}$$

Note that $\mathbf{C}_P(J)$ has 1's along the diagonal and is guaranteed to be positive definite so long as the number of principal components is less than the rank of the matrix.

The correlation matrix using PCA shrinkage is found by blending the sample correlation matrix $\mathbf{C}_0$ with the PCA correlation matrix $\mathbf{C}_P(J)$. More specifically, the correlation matrix using PCA shrinkage is given by the two-parameter formula,

$$\tilde{\mathbf{C}}(\lambda, J) = (1 - \lambda)\mathbf{C}_0 + \lambda \mathbf{C}_P(J), \tag{A.8}$$

where λ is the shrinkage intensity.

A.2. Time-series method

Let f_{tk} denote the return to local factor k over period t. Arrange these factor returns into a $T \times K$ matrix $\mathbf{f}$. The k^{th} column of this matrix is a $T \times 1$ vector that represents the time series of returns to local factor k

$$\mathbf{f}_k = \mathbf{f}\boldsymbol{\delta}_k, \tag{A.9}$$

where $\boldsymbol{\delta}_k$ is a $K \times 1$ column vector with entry 1 for element k and zeros elsewhere. Similarly, let $\mathbf{f}_t$ denote the t^{th} row of matrix $\mathbf{f}$. This represents the $1 \times K$ row vector of factor returns over period t,

$$\mathbf{f}_t = \boldsymbol{\delta}_t' \mathbf{f}, \tag{A.10}$$

where $\boldsymbol{\delta}_t'$ is a $1 \times T$ row vector with entry 1 in element t and zeros elsewhere.

Now consider a set of global factors designed to capture the co-movement of local factors. These global factors are sometimes called *core* factors. Let g_{tj} be the return to global factor j over period t. Arrange these global factor returns into a $T \times J$ matrix $\mathbf{g}$. As before, the time series of returns to factor j are given by the columns of matrix $\mathbf{g}$, i.e.,

$$\mathbf{g}_j = \mathbf{g}\boldsymbol{\delta}_j, \tag{A.11}$$

where $\boldsymbol{\delta}_j$ is a $J \times 1$ column vector with a 1 in entry j and zeros elsewhere. Similarly, the rows of matrix $\mathbf{g}$ give the global factor returns over period t,

$$\mathbf{g}_t = \boldsymbol{\delta}_t' \mathbf{g}, \tag{A.12}$$

where $\mathbf{g}_t$ is a $1 \times J$ row vector of factor returns for period t, and $\boldsymbol{\delta}_t'$ is a $1 \times T$ row vector with entry 1 for element t and zeros elsewhere.

Write the local factor returns as a linear combination of global factor returns plus an unexplained

residual,

$$\mathbf{f}_k = \mathbf{g}\mathbf{b}_k + \mathbf{e}_k, \tag{A.13}$$

where $\mathbf{b}_k$ is the $J \times 1$ vector of slope coefficients that define the loadings of local factor k on the global factors, and $\mathbf{e}_k$ is the $T \times 1$ vector of residual returns. Using weighted least squares (WLS) regression to solve for the slope coefficients, we obtain

$$\mathbf{b}_k = (\mathbf{g}'\mathbf{W}\mathbf{g})^{-1}\mathbf{g}'\mathbf{W}\mathbf{f}_k, \tag{A.14}$$

where $\mathbf{W}$ is $T \times T$ diagonal matrix whose elements are given by the EWMA weights. Arrange these column vectors into a $J \times K$ matrix $\mathbf{B}$ of factor loadings, so that

$$\mathbf{B} = (\mathbf{g}'\mathbf{W}\mathbf{g})^{-1}\mathbf{g}'\mathbf{W}\mathbf{f}. \tag{A.15}$$

Hence, Equation (A.13) can be written in compact form for all factors,

$$\mathbf{f} = \mathbf{g}\mathbf{B} + \mathbf{e}, \tag{A.16}$$

where $\mathbf{e}$ is a $T \times K$ matrix of residual returns. Let $\mathbf{F}_0$ denote the $K \times K$ sample covariance matrix of local factor returns,

$$\mathbf{F}_0 = \mathbf{f}'\mathbf{W}\mathbf{f}. \tag{A.17}$$

Similarly, let $\mathbf{G}$ denote the $J \times J$ covariance matrix of global factor returns,

$$\mathbf{G} = \mathbf{g}'\mathbf{W}\mathbf{g}, \tag{A.18}$$

and let $\mathbf{E}$ denote the $K \times K$ covariance matrix of residual factor returns,

$$\mathbf{E} = \mathbf{e}'\mathbf{W}\mathbf{e}. \tag{A.19}$$

Substituting Equation (A.16) into Equation (A.17), we obtain

$$\mathbf{F}_0 = (\mathbf{B}'\mathbf{g}' + \mathbf{e}')\mathbf{W}(\mathbf{g}\mathbf{B} + \mathbf{e}). \tag{A.20}$$

Since $\mathbf{e}$ is orthogonal to $\mathbf{g}$, (i.e., $\mathbf{g}'\mathbf{W}\mathbf{e} = \mathbf{0}$), this simplifies to

$$\mathbf{F}_0 = \mathbf{B}'\mathbf{G}\mathbf{B} + \mathbf{E}. \tag{A.21}$$

Note that matrix $\mathbf{E}$ is not diagonal, since the global factors do not explain all of the return covariation of the local factors. In fact, Equation (A.21) is just an alternative representation of the sample covariance matrix.

Recall that the sample covariance matrix is rank deficient whenever $K > T$. To obtain a positive definite covariance matrix, we make the usual assumption that the residual returns are mutually uncorrelated. Letting E_{ij} denote element ij of matrix $\mathbf{E}$, we define a new matrix $\tilde{\mathbf{E}}$, with elements given by

$$\tilde{E}_{ij} = E_{ij}\delta_{ij}, \tag{A.22}$$

where $\delta_{ij} = 1$ if $i = j$, and zero otherwise. The estimated local factor covariance matrix using the time-series method is given by

$$\mathbf{F}_T = \mathbf{B}'\mathbf{G}\mathbf{B} + \tilde{\mathbf{E}}, \tag{A.23}$$

which is full column rank.

To recover a proper correlation matrix (i.e., with 1's along the diagonals), we must scale by the inverse of the factor volatilities. Let $\mathbf{S}_T$ denote the diagonal matrix whose elements are equal to the factor volatilities derived from matrix $\mathbf{F}_T$. Multiplying the factor covariance matrix on each side by the inverse of the factor volatility matrix, we obtain

$$\mathbf{C}_T = \mathbf{S}_T^{-1}\mathbf{F}_T\mathbf{S}_T^{-1}, \tag{A.24}$$

which represents the correlation matrix in the time-series method.

Appendix B

In this Appendix, we describe the technique of model integration. Our aim is to replicate the local correlation matrices along the diagonal blocks. For simplicity in the derivation, we consider

only two models. The extension to M models is obvious and straightforward.

Let $\tilde{\mathbf{C}}^L_{mm}$ denote our best estimate of the local factor correlation matrix for model m, and let $\tilde{\mathbf{C}}^L_{nn}$ denote our best estimate of the local factor correlation matrix for model n. As a concrete example, m might represent the US equity model, and n might denote the Europe equity model.

Let $\tilde{\mathbf{C}}_G$ denote the "global" (i.e., spanning models m and n) factor correlation matrix, which can be computed using either PCA shrinkage, as in Equation (A.8), or using the time-series method, as in Equation (A.24). This matrix can be written in block form,

$$\tilde{\mathbf{C}}_G = \begin{bmatrix} \tilde{\mathbf{C}}^G_{mm} & \tilde{\mathbf{C}}^G_{mn} \\ \tilde{\mathbf{C}}^G_{nm} & \tilde{\mathbf{C}}^G_{nn} \end{bmatrix}. \tag{B.1}$$

Unfortunately, the diagonal blocks of matrix $\tilde{\mathbf{C}}_G$ differ from our best estimates obtained in the local model. That is, $\tilde{\mathbf{C}}^G_{mm} \neq \tilde{\mathbf{C}}^L_{mm}$. Hence, if we were to use matrix $\tilde{\mathbf{C}}_G$ as our correlation matrix, we would be utilizing sub-optimal estimates for the diagonal blocks.

The objective of model integration is to obtain a final correlation matrix $\mathbf{C}_F$ in which the diagonal blocks exactly match the diagonal blocks of the local models. One way to achieve this is to simply "drop" the local correlation matrices into the diagonal blocks of $\tilde{\mathbf{C}}_G$. Unfortunately, there is no guarantee that the resulting correlation matrix would be positive definite. In order to preserve positive definiteness, we must make minor adjustments to the off-diagonal blocks.

This task is accomplished by means of the Integration Matrix $\mathbf{M}$, defined as

$$\mathbf{M} = \begin{bmatrix} \mathbf{Q}_{mm} & 0 \\ 0 & \mathbf{Q}_{nn} \end{bmatrix}, \tag{B.2}$$

where

$$\mathbf{Q}_{mm} \equiv (\tilde{\mathbf{C}}^L_{mm})^{1/2}(\tilde{\mathbf{C}}^G_{mm})^{-1/2}, \tag{B.3}$$

and

$$\mathbf{Q}_{nn} \equiv (\tilde{\mathbf{C}}^L_{nn})^{1/2}(\tilde{\mathbf{C}}^G_{nn})^{-1/2}. \tag{B.4}$$

Note that if the diagonal blocks agree between the global and local models, i.e., $\tilde{\mathbf{C}}^L_{mm} = \tilde{\mathbf{C}}^G_{mm}$, then the integration matrix becomes simply the identity matrix. The final correlation matrix is found by applying the transformation,

$$\mathbf{C}_F = \mathbf{M}\tilde{\mathbf{C}}_G\mathbf{M}'. \tag{B.5}$$

Carrying out the matrix multiplication, we obtain

$$\mathbf{C}_F = \begin{bmatrix} \tilde{\mathbf{C}}^L_{mm} & \mathbf{Q}_{mm}\tilde{\mathbf{C}}^G_{mn}\mathbf{Q}'_{nn} \\ \mathbf{Q}_{nn}\tilde{\mathbf{C}}^G_{nm}\mathbf{Q}'_{mm} & \tilde{\mathbf{C}}^L_{nn} \end{bmatrix}. \tag{B.6}$$

Note that the diagonal blocks of the reconstituted matrix $\mathbf{C}_F$ equal the diagonal blocks of the local models. Moreover, in practice we find $\tilde{\mathbf{C}}^L_{mm} \approx \tilde{\mathbf{C}}^G_{mm}$, which implies that the off-diagonal blocks of $\mathbf{C}_F$ are nearly equal to the off-diagonal blocks of $\tilde{\mathbf{C}}_G$, i.e., $\mathbf{Q}_{nn}\tilde{\mathbf{C}}^G_{nm}\mathbf{Q}'_{mm} \approx \tilde{\mathbf{C}}^G_{nm}$.

References

Heston, S. L. and Rouwenhorst, K. G. (1994). "Does Industrial Structure Explain the Benefits of International Diversification?" *Journal of Financial Economics* **36**(1) (August), 3–27.

Ledoit, O. and Wolf, M. (2004). "Honey, I Shrunk the Covariance Matrix," *Journal of Portfolio Management* **30**(4) (Summer), 110–119.

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

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

Menchero, J. and Li, P. (2020). "Correlation Shrinkage: Implications for Risk Forecasting," *Journal of Investment Management* **18**(3) (Third Quarter): 92–108.

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

Rosenberg, B. (1974). "Extra-Market Component of Covariance in Security Returns," *Journal of Financial and Quantitative Analysis* **9**(2) (March), 263–274.

Sharpe, W. F. (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.

Keywords: Portfolio optimization; covariance matrices; sampling error; shrinkage; principle component analysis; integrated factor models; multi-asset-class risk models.