A study of improved covariance matrix estimators for low and diversified volatility portfolio strategies

Benedikt Himbert[a], Julia Kapraun[b], Markus Rudolf[a]

November 2017

Abstract

The sample estimator of the covariance matrix is often found to be unstable due to estimation error. But more sophisticated covariance matrix estimators that rely on factor models, shrinkage towards some Bayesian prior or principal component analysis do not substantially improve on the sample estimator when portfolios are short-sale constrained. Moreover, all models assume the true covariance matrix to be relatively constant over time, while in reality financial asset returns are known to exhibit heteroskedasticity with volatility clustering and time-varying in-between correlations. In response, the use of multivariate Generalised Autoregressive Conditional Heteroskedasticity ("GARCH") and Exponentially Weighted Moving Average ("EWMA") models in low-dimensional covariance matrix problems for portfolio optimisation (across diversified assets) has recently been suggested. This study investigates, whether there is any value to be derived by practitioners in employing such more complex regularisation methods over the sample covariance matrix estimator in high-dimensional covariance matrix problems. We, thus, provide a comprehensive review and comparison of nine covariance matrix regularisation models and additionally test an heuristic, portfolio resampling method. Our findings suggest that only employing GARCH and EWMA methods leads to a significant improvement in the covariance matrix forecast and, thus, better out-of-sample portfolio performance when dealing with a high-dimensional covariance matrix of stock returns (i.e. when we include more than approx. 100 assets). In fact, for any smaller set of stocks none of the estimators improves much on the sample covariance matrix. We then extend our findings to popular low and diversified risk asset allocation rules and show that using GARCH and EWMA models for estimating the covariance matrix generally leads to a superior attainment of strategy specific targets and lower ex-post portfolio volatility.

Keywords: Portfolio Optimisation, Covariance Matrix Estimators, EWMA, DCC-GARCH

[a]WHU - Otto Beisheim School of Management, Burgplatz 2, 56179 Vallendar, Germany

[b]Goethe-Universität Frankfurt, Theodor-W.-Adorno-Platz 3, 60323 Frankfurt, Germany

1 Introduction

Following the wake of the financial crisis, low volatility investing has attracted a considerable amount of interest. Although low volatility portfolios tend to perform poorly relative to the cap-weighted benchmark in bullish markets, empirical evidence suggests that they provide superior long-term returns with lower risk (Chow et al, 2014). Even from a scientific point of view, it has been recognised that mean-variance investment rules can underperform naive and low volatility investment strategies (e.g. DeMiguel et al. (2009); Himbert and Kapraun (2017b)). This notion has left asset managers in the search for more robust investment rules and sparked the rush for low and diversified risk diversification methods, which led to a variety of alternatively weighted indices and ETFs. Among the most popular portfolio optimisation methods are the Minimum Variance (Markowitz, 1952; Clarke et al., 2006), Equal Risk Contribution or Risk Parity (Qian, 2006; Maillard et al., 2010b; Asness et al., 2012) and Most Diversified portfolios (Choueifaty and Coignard, 2008). All these strategies only rely on an estimate of the covariance matrix as an input to the optimisation problem.

There is a common consensus that it is mainly the vector of expected returns, which is considered notoriously hard to estimate (Merton, 1980), that causes large discrepancies between the expected and observed ex-post performance of most mean-variance investment allocations. Since the estimation error in the covariance matrix is known to be relatively smaller than in the vector of expected returns, one may be tempted to believe that low volatility investing is not sensitive to estimation error. On the contrary, Kan and Zhou (2007) show that, when the ratio of portfolio size (N) to the sample size (T) for each asset is large, the impact of misspecification of the covariance matrix estimates when using the sample estimator can be as important. We believe that more research should, thus, be devoted to reducing estimation risk in covariance matrix estimates for application in low and diversified risk investment strategies.

The standard approach to estimating the covariance matrix is the sample estimator, an equally weighted average of observed asset covariances. While it is still the most commonly employed method amongst practitioners, it has been found to be unstable (e.g. Kondor et al. (2007)). This finding motivated research into more robust methods of estimating the elements of the covariance matrix. An intuitive approach to forgo estimation error, is to reduce the number of parameters to be estimated by imposing a structure on the covariance matrix (e.g. in the form of the constant correlation model of Elton and Gruber (1973)) or by employing explicit factor models (e.g. Chan et al. (1999)). The resulting estimators are, however, biased if the true return generating model does not correspond to the assumed model. To find a compromise between bias and estimation error, Ledoit and Wolf (2003) propose shrinkage methods as introduced by James and Stein (1961). They suggest combining the properties of the sample estimator and those of a more structured estimator. In addition, implicit factor models based on principal component analysis have recently gained considerable attention (e.g. Plerou et al. (2002)). Kan and Zhou (2007) and Gosier et al. (2005), among others, respectively analyse the out-of-sample performance of such Bayesian estimators and factor models. They find that when short-sales are allowed, regularisation methods work well in improving covariance estimates and yield significantly better forecasts than the sample estimator. Pantaleo et al. (2011) and Coqueret and Milhau (2014) extend this analysis with similar results. Both studies confirm that when the mean-variance portfolios are in themselves short-sale constrained, the sample estimator, shrinkage and factor models mostly perform equally well (in line with earlier findings by Jagannathan and Ma (2003)).

Yet these approaches are only valid if one assumes the true (unobserved) covariance matrix to be constant over time or to at least vary slowly. The latter is implicit, since these methods are only targeted at reducing the estimation risk that stems from sampling error. When employing a very large T (implying a sufficiently large amount of data is available) sampling error should not be a concern and the sample covariance estimator should be reliable (Merton, 1980). A large strand of literature has, however, proven that financial asset returns exhibit heteroskedasticity with volatility clustering and the assumption of constant correlations between financial asset returns is often found to be violated. More advanced time-varying covariance matrix models as proposed by Engle (2002) have been shown to better capture the time-varying properties of the covariance matrix. Research on using forward looking properties of multivariate time-varying volatility models in portfolio optimisation, however, is sparse. Harris and Nguyen (2013), among others, only recently suggest using multivariate Generalised

Autoregressive Conditional Heteroskedasticity ("GARCH") (Engle and Bollerslev, 1986) and Exponentially Weighted Moving Average ("EWMA") (Longerstaey, 1996; Zumbach, 2004) models in portfolio optimisation to reduce the impact of estimation error in the covariance matrix. Zakamulin (2015) finds that such methods cut the forecast error compared to the sample estimator and shrinkage methods in half. This could translate into a remarkable improvement in robustness of optimal portfolio weights and benefit the whole class of low and diversified risk investment policies. The practical merits of Zakamulin's work are, however, limited to a set of Fama-French factor and index portfolios (which are taken as assets and, thus, in themselves diversified) and only proven valid for low dimensional covariance matrix problems.[1]

It lies at hand that the range of proposed estimators for the covariance matrix of asset returns is not only wide, but also highly commercialised. As a result, portfolio managers and index providers often rely on pre-packaged, black-boxed methodologies when optimising the asset portfolio towards some target.[2] Our research is set out to investigate whether there is any value to be derived by practitioners in employing such more complex regularisation methods over the simple sample covariance matrix estimator. While we are clearly not the first to conduct a comparative study of covariance matrix estimators, we complement existing literature on several levels. Firstly, we include the novel Fama-French five factor model (Fama and French, 2015) as well as fairly recent principal component and hyperbolically decaying EWMA methods in our study of covariance regularisation methods. Secondly, we extend research on the use of time-varying volatility models in portfolio optimisation by Zakamulin (2015) and Harris and Nguyen (2013) to high dimensional covariance matrix problems. We examine these in direct comparison to the heuristic "resampling" method (Michaud, 1989) that explicitly targets a reduction in sampling error and has been shown to improve on the use of more complex shrinkage estimators or factor models (Markowitz and Usmen, 1996; Becker et al., 2015a). Moreover, we are not only concerned with the question of which method works best in forecasting the covariance matrix, but with the implications for portfolio management. Diversified risk strategies such as Most Diversified, for example, are significantly more sensitive to estimation error than Minimum Variance portfolio allocations that are commonly tested for in this context (Himbert and Kapraun, 2017a). Finally, without imposing maximum weight constraints, the Minimum Variance portfolio generally is too

[1] The largest number of assets included in any dataset in Zakamulin (2015) is $N = 10$.

[2] For example, the S&P 500 Minimum Volatility Index is optimised using the covariance specifications of the "Northfields optimiser", which in itself allows for a range of unclear decisions in regards to how parameters are estimated. MSCI frequently employs their "Barra Global Equity Model" that is also commercially available.

concentrated for most investors or to be used in alternatively weighted index products. As funds are often bound by short-sale constraints and regulatory rules commonly prescribe maximum holdings of no more than 10% in any asset (while alternative indexing methods have even more restrictive weight constraints), we put additional emphasis on the impact of such natural limitations on estimation risk.

The remainder of this study is structured as follows. In section 2 we review a representative set of available covariance matrix estimation methods. In Section 3 we employ these methods to real datasets consisting of U.S. and European large-cap stocks and assess their ability to predict the volatility of the short-sale constrained minimum-variance portfolio, depending on N. Section 4 compromises an application of the selected methods to popular low and diversified risk investment strategies.

2 A review of covariance matrix estimation methods

2.1 The Sample Estimator

The sample estimator of the covariance matrix $\mathbf{\Sigma}$ serves as the benchmark model to the range of enhanced estimation methods we consider. Denote by $\mathbf{r}_t$ and r_{ft} respectively the vectors of rates of return on N risky assets and the risk-free asset at time t ($t = 1, 2, ..., T$). Then the excess returns (from here on "returns") at time t are defined as $\boldsymbol{R}_t = \mathbf{r}_t - r_{ft}\mathbf{1}_N$, where $\mathbf{1}_N$ is a $N \times 1$ vector of ones. When $\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{R}_t$ is the vector of the assets' expected returns estimated over the period of length T, the (unbiased) sample covariance estimator is given by:

$$\hat{\mathbf{\Sigma}}^{se} = \frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{R}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{R}_t - \hat{\boldsymbol{\mu}})' \tag{1}$$

A well-known problem in portfolio optimisation is that the sample covariance matrix is singular (non-invertible) when $N \geq T$. For once, the known linear solution to the short-sale unconstrained minimum-variance portfolio problem relies on the invertibility of the covariance matrix. But even for short-sale constrained portfolio problems, this causes problems in most optimisers (Kondor et al., 2017). Merton (1980) has shown that the sample covariance matrix is an unbiased and convergent estimator of the true covariance matrix when $T \to \infty$, since statistically it provides the maximum likelihood estimate under the assumption of i.i.d. returns and covariance stationarity. The availability of historical data may, however, be limited and additionally problems of non-stationarity of the variance-covariance structure have been documented (Engle, 1993). To mitigate the former and to ensure an invertible covariance matrix, we turn to higher frequency (daily) data which generally benefits the sample estimator of the covariance matrix (Gosier et al., 2005).

2.2 Explicit Factor Model Estimators

Chan et al. (1999) suggest that estimates of the sample covariance matrix are biased by specific events that affect several stocks at a time but may not persist in the future. They, thus, propose to "strip out the idiosyncratic components of the covariance structure" by introducing common factors that drive asset returns. In the spirit of Ross (1976) it is assumed that the return of each stock i ($i = 1, 2, ..., N$) can be decomposed into a combination of K factors and an idiosyncratic term:

$$R_{i,t} = \beta_{i0} + \sum_{k=1}^{K} \beta_{ik} f_{k,t} + \epsilon_{i,t}, \quad for\, k = 1, 2, ...K. \tag{2}$$

Here $R_{i,t}$ denotes the excess return of an asset i and $f_{k,t}$ the return of the k-th common factor at time t, while $\epsilon_{i,t}$ is a residual term. The coefficients β_{ik} can be understood as the loadings of stock i on each of the K factors. When $\hat{\mathbf{B}}$ is the $(N \times K)$ matrix of factor loadings of the assets that contains the least-squares estimators β_{ik} and $\hat{\mathbf{\Omega}}$ denotes the sample covariance matrix of factor returns, the factor estimator of the covariance matrix is given by:

$$\hat{\mathbf{\Sigma}}^{fac} = \hat{\mathbf{B}}\hat{\Omega}\hat{\mathbf{B}}' + \hat{\mathbf{D}}, \tag{3}$$

with $\hat{\mathbf{D}}$ being a diagonal matrix that contains the residual return variances (from (2)). Employing factor models significantly decreases the number of parameters to be estimated. While the sample covariance matrix estimator requires the estimation of $N \times (N+1)/2$ off-diagonal elements, the number of parameters required in the factor model is limited to $N \times K$ betas, $K \times (K+1)/2$ factor covariances and N idiosyncratic variances. Reducing the number of covariance matrix elements decreases the chance of estimation outliers that can dominate an optimised portfolio, something that is often referred to as error maximisation (Michaud, 1989). Factor model, thus, appear particularly attractive for high-dimensional covariance matrix problems where the risk of extremely erroneous parameter estimates is higher.

The use of a single factor model for estimating the covariance matrix goes back to Sharpe (1964), who considers the stock market index as the unique factor. But Fama and French (1993), among others, show that a single factor does not sufficiently capture stock price movements and suggest using a multi-factor model. When choosing these factors, it is important that they are rewarded in the long-run. Only factors that help to explain asset returns can be expected to reduce the unexplained variance (Chan et al., 1999). We, thus, employ the most widely recognised three-factor asset pricing model of Fama and French (1993) and their most recent extension of a five-factor model (Fama and French, 2015) that, to our knowledge, has so far not been considered in this context. The respective factors are the cap-weighted market index, the small-minus-big capitalisation factor and the high-minus-low book-to-market factor that are complemented by the high-minus-low profitability and the conservative-minus-aggresive investment factor in Fama-French's five factor model. Products available to portfolio managers such as the "Barra Risk Factor Analysis" incorporate over 40 factors. While incorporating a larger number of factors can help to reduce the unexplained variance, it also increases the number of parameters to be estimated. In this context, Chan et al. (1999) test a variety of factor models that include up to 10 factors and find that when lower and upper weight constraints are imposed in portfolio optimisation, a one-factor model can perform as well as a multi-factor model.

2.3 Shrinkage Estimators

The sample covariance matrix is known to be (asymptotically) unbiased, but exhibits a significant degree of estimation error. At the other extreme one finds structured estimators such factor or constant correlation matrix models (e.g. Chan et al. (1999); Elton and Gruber (1973)) that have little estimation error but tend to be biased. James and Stein (1961) suggest that there must exist an optimal trade-off between bias and estimation error and propose compromise between the two by computing a convex linear combination between the sample matrix and a Bayesian prior. What is referred to as shrinkage can, thus, be understood as taking a weighted average of the sample covariance matrix with some target matrix $\mathbf{F}$ that has less estimation error:

$$\hat{\mathbf{\Sigma}}^{shrink} = (1-b) \cdot \hat{\mathbf{\Sigma}}^{se} + b \cdot \mathbf{F}. \tag{4}$$

Here $b \epsilon (0,1)$ denotes the shrinkage intensity that controls how much structure is imposed to the shrunk covariance matrix estimator. Michaud (1989) reviews a variety of such shrinkage estimators and finds that most of them break down when $N \geq T$, since the derivation of the optimal shrinkage intensity b in its original form relies on the inverse of the sample covariance matrix that is not attainable when the number of assets exceeds the number of data points in the estimation set. In response to this finding, Ledoit and Wolf (2003, 2004) propose a set of shrinkage estimators where the optimal shrinkage intensity is derived under the Frobenius norm.[3]

In this study, we explore two alternative models for the shrinkage target - shrinkage towards a single factor estimator and shrinkage towards the constant correlation matrix. As laid out in 2.2, there is yet

[3]The shrinkage coefficient, thus also depends on the shrinkage target $\mathbf{F}$. For a detailed derivation we refer the interested reader to Ledoit and Wolf (2003) and Ledoit and Wolf (2004).

no consensus on the identity and the number of factors K to be used in a factor model since these can be very different across datasets. In fact, Chan et al. (1999) suggest that a single index model with the market as the only factor (which corresponds to the model of Sharpe (1964)) may be a sufficiently structured estimator in most applications. Ledoit and Wolf (2003), thus, propose an estimator of the form (3) as a shrinkage target with the single factor being the return of the market portfolio (where we take the Fama French market factor as a proxy of the return on the market portfolio). The second, more structured, shrinkage target we consider is the covariance matrix derived under the constant correlation matrix model as introduced by Elton and Gruber (1973) and used in Ledoit and Wolf (2004). Here all pairwise correlations are assumed to be identical and thus the average of all the sample correlations is the estimator of the common constant correlation. The shrinkage target is then obtained by multiplying the constant correlation matrix with the sample standard deviations. One major drawback of shrinkage remains the fact that the choice of shrinkage target is somewhat arbitrary.

2.4 Random Matrix Theory Principal Component Estimator

Methods of estimating the covariance matrix through principal component analysis avoid pre-specified, explicit factor models in favour of extracting relevant factors from the data. Hence, one does not assume returns to be described by a factor model that includes risk factors such as profitability or size. Principal component analysis instead relies on the eigendecomposition of the sample covariance (or correlation) matrix. The extracted eigenvectors ("factors") are orthogonal and, by design, their variances are the eigenvalues of the covariance matrix. It can be shown that the ratio of each eigenvalue to the sum of eigenvalues is the proportion of the total variance of observations which is explained by the corresponding factor (Coqueret and Milhau, 2014). Factors that exhibit the highest eigenvalue, thus, explain the largest share of total variance. In order to decide, which factors are relevant and should be kept, eigenvectors are ranked by order of the corresponding decreasing eigenvalues. There is yet no consensus in literature on how many factors should be kept (Bai and Ng, 2002). Most research concludes that six to seven factors can sufficiently capture the co-movements between the assets. (e.g. Alessi et al. (2010)). In our study we use a more recent principal component estimator that was first proposed by Laloux et al. (2000) and is also used by Pantaleo et al. (2011). The estimator requires the eigendecomposition of the sample correlation matrix $\hat{\mathbf{C}}$ such that:

$$\hat{\mathbf{C}} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}', \tag{5}$$

where $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues λ_i i $(i = 1, 2, ..., N)$, sorted by decreasing order, and $\mathbf{E}$ is an orthogonal matrix that contains the corresponding eigenvectors. As in Laloux et al. (2000) we make use of Random Matrix Theory ("RMT") to determine the number of factors that are kept. Subject to RMT, if the N variables of the system are i.i.d. with unit variance, in the limit $N, T \to \infty$ the ratio N/T converges to a constant ν different from 1 and all eigenvalues of the sample correlation matrix are bounded from above by:[4]

$$\lambda_{max} = (1 + \sqrt{\nu})^2. \tag{6}$$

When applied to finite samples, the constant ν is estimated by the ratio N/T. We then follow Plerou et al. (2002) in applying RMT to the covariance estimation problem. In doing so, we factorise the sample correlation matrix as in (5) and define the diagonal matrix $\mathbf{\Lambda}^*$ of which the diagonal elements are given by:

$$\lambda_i^* = \lambda_i \; if \; \lambda_i > \lambda_{max} \quad and \quad \lambda_i^* = 0 \; otherwise. \tag{7}$$

Thus, one essentially replaces all eigenvalues smaller than λ_{max} by 0 in order to eliminate their corresponding eigenvector. We then plug the modified diagonal matrix back into (5) and estimate the new matrix product:

$$\hat{\mathbf{C}}^* = \mathbf{E}\mathbf{\Lambda}^*\mathbf{E}'. \tag{8}$$

In order to transform the factor correlation matrix into a covariance matrix estimate, we set the diagonal elements of $\hat{\mathbf{C}}^*$ equal to one. We then obtain the covariance matrix estimator by multiplying $\hat{\mathbf{C}}^*$ with a diagonal matrix $\hat{\mathbf{\Sigma}}_D$, whose diagonal elements are the sample volatilities of the individual assets:[5]

$$\hat{\mathbf{\Sigma}}^{pca} = \hat{\mathbf{\Sigma}}_D \mathbf{\Lambda}^* \hat{\mathbf{\Sigma}}_D. \tag{9}$$

In unreported results we also examine the eigendecomposition of the covariance matrix and retain the eigenvectors with the six largest eigenvalues. We find that this estimator consistently underperforms the eigendecomposition of the correlation matrix as proposed by Plerou et al. (2002) and, thus, omit its results in the interest of brevity.

[4]Coqueret and Milhau (2014) points out that the assumption of unit variance across all assets explains why one considers the eigendecomposition of the correlation matrix, as opposed to that of the covariance matrix.

[5]This corresponds to $\hat{\mathbf{\Sigma}}^{se}$ in which the off-diagonal elements are set to zero.

2.5 Exponentially Weighted Moving Average Estimators

All of the estimators considered so far are subject to the assumption that one is modelling a stable return process and, therefore, each point in time contains equally relevant information. This implies stable variances and correlations among assets in a diversified portfolio. If this held true, estimation risk stemming from sampling error could easily be controlled for by assuming a sufficiently large T. There is, however, considerable evidence that asset returns exhibit heteroskedasticity with volatility clustering and the assumption of constant correlation among assets is violated (Engle, 1993). Imposing a factor model of the form (2) does not resolve this issue, since even between factor returns, changing correlations and volatilities persist. This induces an additional component of estimation risk, which is independent from sampling error. To capture the time-varying nature of volatilities, Longerstaey (1996) proposes the exponentially weighted moving average model. The EWMA model for estimating variances better captures the dynamics of return shocks by putting more weight on recent observations and can easily be transformed into a multivariate context for estimating covariances. In what follows we assume that $\mathbf{R}_t$ is given by:

$$\mathbf{R}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \tag{10}$$

where $\boldsymbol{\mu}_t$ denotes the Nx1 vector of time-varying ("conditional") expectations of $\mathbf{R}_t$ (that is the vector of mean asset returns) and $\boldsymbol{\varepsilon}_t$ the Nx1 vector of random disturbances at time t such that $E_{t-1}[\boldsymbol{\varepsilon}_t] = \mathbf{0}$. We then assume that the conditional returns $\mathbf{R}_t$ follow a multivariate normal distribution with mean zero ($\boldsymbol{\mu}_t = \mathbf{0}$) and conditional covariance matrix $\mathbf{H}_t$.[6] The EWMA covariance matrix is then defined recursively over t ($t = 1, 2, ..., T$):

$$\mathbf{H}_t = (1-\kappa)\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1} + \kappa\mathbf{H}_{t-1}, \tag{11}$$

for a constant $\kappa\epsilon(0,1)$ that determines the intensity of reaction of covariances to return shocks in the first term as well as persistence in covariances in the second term. κ is usually set to 0.94 for daily frequency data as recommended by Longerstaey (1996), while the full-sample covariance matrix is chosen as the required, initial value $\mathbf{H}_0$. By definition, extreme returns, therefore, become less important in the average as the data window slides along. The input covariance estimator for portfolio

[6] Note that the elements of $\boldsymbol{\mu}_t$ and $\mathbf{H}_t$ must be measurable w.r.t. the σ-field $\mathcal{F}_{t-1}$ generated by $\mathbf{R}_{t-j}$ for $j \geq 1$. In practice, the zero-mean assumption is satisfied by de-meaning $\mathbf{R}_t$.

optimisation is then simply defined as:

$$\hat{\mathbf{\Sigma}}^{ewma1996} = \mathbf{H}_T, \tag{12}$$

the conditional covariance matrix recursively estimated over T days.

Zumbach (2006) suggest an improved version of the EWMA estimator in which the weights on past covariances decay hyperbolically rather than exponentially. It is estimated as a weighted sum of several realistic EWMAs instead of a single EWMA as in (11). Formally, it is given by:

$$\mathbf{H}_t = \sum_{l=1}^{L} w_l \mathbf{H}_{l,t} \tag{13}$$

$$\mathbf{H}_{l,t} = (1-\kappa_l)\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1} + \kappa_l \mathbf{H}_{l,t-1}$$

$$w_l = \frac{1}{g}(1 - \frac{ln(\tau_l)}{ln(\tau_0)})$$

$$\kappa_l = exp(-\frac{1}{\tau_l})$$

$$\tau_l = \tau_1 \rho^{l-1} \; for \; l = 1, 2, ..., L$$

where g is a normalisation constant that ensures $\sum_{l=1}^{L} w_l = 1$. τ_1 denotes the lower cut-off (i.e. the shortest time scale at which covariances are measured), whereas τ_L defines the upper cut-off (which increases exponentially in L) and τ_0 is the logarithmic decay factor. The parameter ρ, required to operationalise the model, is set to $\sqrt{2}$, $\tau_0 = 1560$ (days), $\tau_1 = 4$ (days) and $\tau_L = 504$ (days) corresponding to Zumbach (2006) or Harris and Nguyen (2013).[7] Depending on its parametrisation, the estimator generally exhibits a "longer memory" of past covariances (Zumbach, 2004). Again, the input covariance estimator for portfolio optimisation, now subject to problem (13), is recursively defined as defined as:

$$\hat{\mathbf{\Sigma}}^{ewma2006} = \mathbf{H}_T. \tag{14}$$

[7]Zumbach (2006) actually assumes $\tau_L = 512$ which is equivalent to L = 15. We employ an estimation period of only 504 days, and, thus set $\tau_L = 504$. The impact of this adjustment may be considered negligible.

2.6 Dynamic Conditional Correlation GARCH Estimator

In the EWMA model, the forecast of average covariances over any forecast horizon is set equal to the current estimate of conditional covariances (just as the sample covariance matrix estimator). Since conditional volatilities are assumed constant by the model, so are the estimates of conditional correlations. Alexander (2008), thus, argues that the EWMA model can be reduced to a constant correlation model without any superior forecasting ability. Multivariate Generalised Autoregressive Conditional Heteroskedasticity models, on the other hand, specify the dynamics volatilities and correlations to estimate the parameters. Such models have been shown to better capture volatility clusters by allowing volatility to increase following periods of high realised volatility, or below-normal returns, and allowing volatility to decrease following periods of low realised volatility, or above-normal returns (e.g. Engle and Bollerslev (1986)). As for the EWMA method, we assume that $\mathbf{R}_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{H}_t)$. This implies that:

$$\mathbf{R}_t = \boldsymbol{\varepsilon}_t = \mathbf{H}_t^{1/2}\mathbf{z}_t, \tag{15}$$

where $\mathbf{H}_t^{1/2}$ is any square matrix such that $\mathbf{H}_t = \mathbf{H}_t^{1/2}(\mathbf{H}_t^{1/2})'$ and $\mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$ with $\mathbf{I}_N$ being the identity matrix of order N. The quasi-log-likelihood function under the assumed multivariate Gaussian pdf of $\mathbf{z}_t$ for a sample of T observed vectors $\mathbf{R}_t$ is then given by:

$$\ell_T(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}(N\,ln(2\pi) + ln \mid \mathbf{H}_t| + \boldsymbol{\varepsilon}_t'\mathbf{H}_t^{-1}\boldsymbol{\varepsilon}_t), \tag{16}$$

where $\boldsymbol{\theta}$ denotes the vector of parameters appearing in $\boldsymbol{\mu}_t$, $\mathbf{H}_t$ and (if any) in the pdf of $\mathbf{z}_t$.

In conditional correlation models, conditional covariances are decomposed into the conditional standard deviations ($\sqrt{h_{i,t}}$) for $i = 1, 2, ..., N$ and the conditional correlations $q_{ij,t}$ for $j = 2, 3, ..., N$ and $i < j$. The conditional covariance matrix may, thus, also be expressed as:

$$\mathbf{H}_t = \mathbf{D}_t\boldsymbol{\Gamma}_t\mathbf{D}_t \tag{17}$$

where $\mathbf{D}_t$ is a diagonal matrix of time-varying standard deviations with $\sqrt{h_{i,t}}$ as its i^{th} diagonal element, and $\boldsymbol{\Gamma}_t$ the time-varying correlation matrix of order N. The specification of $\mathbf{H}_t$ is, thus, a function of model choice for each conditional variance and a choice of the conditional correlation matrix. By substituting (13) into (12) one obtains the log-likelihood function of the estimator and can define $\tilde{\boldsymbol{\varepsilon}}_t = \mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_t)$, the vector of residuals standardised by their conditional standard deviation, where:[8]

$$\ell_T(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}(N\, ln(2\pi) + ln \mid \mathbf{D}_t| + log \mid \boldsymbol{\Gamma}_t| + \tilde{\boldsymbol{\varepsilon}}_t'\boldsymbol{\Gamma}_t^{-1}\tilde{\boldsymbol{\varepsilon}}_t). \tag{18}$$

Hence, the conditional correlation is given through the conditional covariance between the standardised disturbances. The conditional variances $h_{i,t}$ are modelled as univariate GARCH processes, which in its most commonly assumed GARCH(1,1) form is given by:

$$h_{i,t} = \omega_i + \delta_i \varepsilon_{i,t-1}^2 + \gamma_i h_{i,t-1}. \tag{19}$$

Here ω_i denotes the weighted long-run variance, $\delta_i \varepsilon_{i,t-1}^2$ the moving average term with weight δ_i assigned to the lagged squared innovation and $\gamma_i h_{i,t-1}$ the autoregressive term with assigned weight γ_i for the lagged variance .[9]

Literature suggest several specifications for choosing the conditional correlation matrix $\boldsymbol{\Gamma}_t$, with one of the most popular being the Dynamic Constant Correlation model proposed by Engle (2002). He proposes the following relationship:

$$\boldsymbol{\Gamma}_t = \mathbf{Q}_t^{*-1}\mathbf{Q}_t\mathbf{Q}_t^{*-1} \tag{20}$$

$$\mathbf{Q}_t = (1-a-c)\bar{\mathbf{Q}} + a\tilde{\boldsymbol{\varepsilon}}_{t-1}\tilde{\boldsymbol{\varepsilon}}_{t-1}' + c\mathbf{Q}_{t-1},$$

where $\mathbf{Q}_t^* = (diag(\mathbf{Q}_t))^{\frac{1}{2}}$. $\bar{\mathbf{Q}}$ is the unconditional covariance of the standardised residuals ($\bar{\mathbf{Q}} = E[\tilde{\boldsymbol{\varepsilon}}_t\tilde{\boldsymbol{\varepsilon}}_t']$), while a and c are scalars and assumed constant across different pairs of assets.[10] (20) thereby ensures that $\boldsymbol{\Gamma}_t$ is positive definite and $\mid q_{ij,t} \mid \leq 1$.

For portfolio optimisation, it is necessary to derive a p-step ahead forecasts of the conditional covariance matrix that is determined by the desired investment horizon or rebalancing interval. The

[8]For a derivation and description of the estimation process we refer the interested reader to Engle (2002).
[9]$\omega_i \geq 0$, $\gamma_i \geq 0$, $\delta_i \geq 0$ and $\gamma_i + \delta_i \leq 1$.
[10]Note that $\mathbf{Q}_0$ must be positive definite.
$0 \leq a + c \leq 1$.

univariate conditional variances (that are the diagonal elements of $\mathbf{D}_t$) can be forecast through:

$$h_{i,t+p} = \sum_{m=1}^{p-2} \omega_i(\delta_i + \gamma_i)^m + (\delta_i + \gamma_i)^{p-1} h_{i,t-1}. \tag{21}$$

The DCC evolution process, however, is non-linear and Engle and Sheppard (2001) suggest using a direct forecast of the conditional correlation matrix:

$$E_t[\mathbf{\Gamma}_{t+f}] = \sum_{m=0}^{p-2} (1 - a - c)\bar{\mathbf{\Gamma}}(a + c)^m + (a + c)^{p-1}\mathbf{\Gamma}_{t+1}, \tag{22}$$

based on the approximation $\bar{\mathbf{Q}} \approx \bar{\mathbf{\Gamma}}$ (where the latter denotes the unconditional correlation matrix) and $E_t[\mathbf{Q}_{t+1}] \approx E_t[\mathbf{\Gamma}_{t+1}]$. The p-step ahead forecast of the conditional covariance matrix is then given by:

$$\mathbf{H}_{T+p} = \mathbf{D}_{T+p}\mathbf{\Gamma}_{T+p}\mathbf{D}_{T+p}. \tag{23}$$

Assuming a rebalancing period of length P, we obtain the input covariance matrix through:

$$\hat{\mathbf{\Sigma}}^{dcc-garch} = \frac{1}{P} \sum_{p=0}^{P-1} \mathbf{H}_{T+p}. \tag{24}$$

2.7 Incorporating heuristic methods: Resampled Efficiency

It is established in literature that the solution to the mean-variance portfolio optimisation problem preferentially includes assets with high return and low risk. A number of studies (e.g. Nawrocki (1991)) show that estimation error in exactly these assets tends to be highest, increasing the impact of estimation error on portfolio weights and leading to what is known as error maximisation. The problem is magnified since the optimal weights in mean-variance portfolios are known to be very sensitive to the input estimates.[11] Michaud (1989) first suggests resampled efficiency, a method aimed at reducing the impact of estimation error by smoothing weights in the asset allocation (i.e. preventing outliers that stem from a large sampling error) as a remedy. The method does not directly provide an alternative covariance matrix estimation method, but is aimed at finding the average optimal weights subject to a large number of realistic covariance matrix estimates. It follows a straightforward algorithm, based on Monte-Carlo methods:

[11] Such problem is even more of a concern for mean-variance portfolios other than the minimum-variance portfolio, since they additionally rely on noisy estimates of return.

1. Given a multivariate return series for N assets of length T, estimate the unknown parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by their empirical moments $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$.[12]

2. Resample from a multivariate normal distribution $\mathcal{N}(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}})$ taking T independent draws and generate N return series of length T. Estimate $\hat{\boldsymbol{\Sigma}}$ by its empirical moment $\hat{\hat{\boldsymbol{\Sigma}}}$.

3. Find the optimal portfolio composition $\hat{\hat{\boldsymbol{w}}}$, subject to the specified portfolio optimisation problem, based on $\hat{\hat{\boldsymbol{\Sigma}}}$.

4. Repeat steps 2. to 3. 499 times.

5. Let $\hat{\hat{\boldsymbol{w}}}^s$ be the portfolio of draw $s = 1, 2, ...500$, calculate the average optimal portfolio weight vector $\hat{\boldsymbol{w}}$ through $\hat{\boldsymbol{w}} = \frac{1}{500}\sum_{d=1}^{500} \hat{\hat{\boldsymbol{w}}}^s$ and allocate portfolio weights accordingly.

Portfolio resampling, thus, works by generating a large number of statically equivalent samples of the actual time series of returns with the same length that average optimal portfolio weights are based upon. Averaging, in fact, ensures that the weights in the portfolio $\hat{\boldsymbol{w}}$ also sum up to one. Some authors argue that the method is heuristic and has no economic justification based on the optimising behaviour of rational agents. Scherer (2002) specifically points to the problem of "optionality" for the allocation of assets that arises from combination of long-only constraints and the averaging procedure. He finds that an increase in volatility of an asset (assuming its expected return remains constant) leads to an increase in the average weight allocation to this asset - a result which cannot be aligned with the assumption of risk-aversion of rational agents. In addition, we want to point out that the method also assumes return and covariance stationarity and, thus, only works in reducing estimation risk that stems from sampling error. Evidence of the value of resampled efficiency in mean-variance portfolio problems is, therefore, mixed. Most notably, Markowitz and Usmen (1996) compare Bayesian covariance matrix estimation methods with resampled efficiency in a simulation study and, against their expectations, find portfolio resampling to work relatively better. Similar results are reported by Fletcher and Hillier (2001) in an empirical study, while Becker et al. (2015a) show shrinkage estimators to outperform Michaud's resampling procedure. Some researchers even argue that resampled efficiency can simply not be compared with Bayesian methods of covariance matrix estimation as for "every distribution a prior will be found that outperforms resampling and vice versa" (Frahm, 2015). Our study is set out to also investigate its value in application to low and diversified volatility investment strategies.

[12]We employ the unbiased sample estimator (1) but other covariance estimators may also be used in conjunction with the resampling method.

Table 2.1: Summary of employed covariance matrix estimators and methods

Method	**Estimated Covariance Matrix ($\hat{\mathbf{\Sigma}}^{y}$)**
Sample Estimator	$\hat{\mathbf{\Sigma}}^{se}$
Fama French 3-factor Model Estimator	$\hat{\mathbf{\Sigma}}^{fac-3}$
Fama French 5-factor Model Estimator	$\hat{\mathbf{\Sigma}}^{fac-5}$
Shrinkage Estimator - Constant Correlation Matrix	$\hat{\mathbf{\Sigma}}^{shrink-cc}$
Shrinkage Estimator - Single Index/Factor Model	$\hat{\mathbf{\Sigma}}^{shrink-si}$
Random Matrix Theory Principal Component Estimator	$\hat{\mathbf{\Sigma}}^{pca}$
EWMA Estimator (RiskMetrics 1996)	$\hat{\mathbf{\Sigma}}^{ewma1996}$
EWMA Estimator (RiskMetrics 2006)	$\hat{\mathbf{\Sigma}}^{ewma2006}$
DCC-GARCH Estimator	$\hat{\mathbf{\Sigma}}^{dcc-garch}$
Resampled Efficiency - Portfolio Resampling	n/a*

**Resampled Efficiency is a heuristic method that relies on resampling of the sample covariance matrix to extract sampling error (as described in section 2.7).*

3 A test of covariance matrix estimation methods

3.1 Data and research design

Our primary U.S. and European datasets respectively consist of daily returns on all assets that have been part of the S&P 500 and Euro Stoxx large cap indices over the 17-year period between January 2000 and December 2016 with little concern for selection bias.[13] From each dataset we randomly select $N = \{25, 50, 100, 150, 200, 250\}$ stocks for analysis. We then employ a rolling window analysis, where at the beginning of each quarter (every $P = 63$ days) from January 2002 through December 2016 we estimate as an input to the portfolio optimisation the covariance matrix, subject to the chosen method, using the $T = 504$ preceding days (which also avoids the issue of facing a non-invertible covariance structure). $\hat{\boldsymbol{\Sigma}}^y$ may, thus, be labelled the ex-ante estimated covariance matrix, subject to the chosen regularisation method y. The optimal weights of a Minimum Variance portfolio are then determined by the solution to the following problem:

$$\hat{\boldsymbol{w}}^y_{MV} = arg\ \min_{\boldsymbol{w}} \boldsymbol{w}'\hat{\boldsymbol{\Sigma}}^y\boldsymbol{w} \quad s.t.\ \mathbf{1}'_{\mathbf{N}}\boldsymbol{w} = 1\ and\ \boldsymbol{w} \geq 0. \tag{25}$$

By focusing on the Minimum Variance portfolio problem, where the vector of expected asset returns is not involved, we can exclude estimation risk that stems of erroneous estimates of expected returns (Ingersoll, 1987). As in Zakamulin (2015) we impose short-sale constraints since most private investors cannot engage in short-selling activities due to budget constraints and it is commonly forbidden by fund rules. The portfolio is held over the following $O = 63$ days and then rebalanced.[14] Following this procedure yields a 15-years long out-of-sample return series from which the annualised ex-post volatility is determined. Any covariance estimation method that yields the lowest ex-post portfolio volatility may then be considered the superior method.[15] We additionally test the effect of including a maximum weight constraint of 10% for each individual asset and hope to contribute to the discussion

[13]The sample, thus, consists of 361 stocks (S&P 500) and 345 stocks (Euro Stoxx large cap) respectively.

[14]Hence, we set $P = O$.

[15]In unreported results we follow Pantaleo et al. (2011) in averaging the ex-post portfolio volatilities, each estimated over their respective O-days out-of-sample period by $\sqrt{(\hat{\boldsymbol{w}}^y_{mv})'\mathring{\boldsymbol{\Sigma}}\hat{\boldsymbol{w}}^y_{mv}}$, where $\mathring{\boldsymbol{\Sigma}}$ denotes the ex-post covariance matrix estimated over O days. Results of this analysis correspond to the results reported in this paper.

of whether (and when) the inclusion of such constraint effectively reduces ex-post portfolio volatility.

Pantaleo et al. (2011) finds that Minimum Variance portfolios based on regularised covariance matrix estimates are on average more diversified in terms of portfolio holdings than those based on the sample covariance matrix. In order to confirm such effect we assess the average "Effective Number of Stocks", which is the inverse of the Herfindahl Index and over 60 ($d = 1, 2, ..., 60$) out-of-sample periods of length O for i ($i = 1, 2, ..., N$) defined as:

$$\overline{ES}^{y}_{MV} = \frac{1}{60}\sum_{d=1}^{60} ES^{y,d}_{MV} = \frac{1}{60}\sum_{d=1}^{60} \frac{1}{\sum_{i=1}^{N}(\hat{w}^{y,d}_{MV,i})^2}, \tag{26}$$

where $\hat{w}^{y,d}_{MV,i}$ denotes the ex-ante optimal weight of an asset i that is estimated over the respective, preceding estimation period. In addition, we are interested in the average portfolio diversification in terms of individual asset volatilities and, thus, track the average "Diversification ratio" as proposed by Choueifaty and Coignard (2008):

$$\overline{DR}^{y}_{MV} = \frac{1}{60}\sum_{d=1}^{60} DR^{y,d}_{MV} = \frac{1}{60}\sum_{d=1}^{60} \frac{(\hat{\boldsymbol{w}}^{y,d}_{MV})'\mathring{\boldsymbol{e}}^{d}}{\bar{\sigma}^{y,d}_{MV}}. \tag{27}$$

Here $\mathring{\boldsymbol{e}}^{d}$ represents the $N \times 1$ vector of individual assets' ex-post sample volatilities and $\bar{\sigma}^{y,d}_{MV}$ is the realised portfolio volatility (both estimated over the d-th out-of-sample period).

3.2 Results from studying selected covariance matrix estimators

In this section we present the results of repeated portfolio optimisation towards the Minimum Variance target using the range of covariance matrix estimators from section 2. Table 3.1 summarises the annualised, ex-post volatilities over the 15-years out-of-sample period in dependency of the chosen covariance matrix regularisation method and the number of assets used as an input to the optimiser. The volatility obtained from using the sample estimator, reported in the first line of each panel, serves as a benchmark to all covariance matrix estimation methods. Our findings are best illustrated in Figure 3.1 that evidences differences in ex-post portfolio volatility subject to the choice of the covariance matrix estimator.

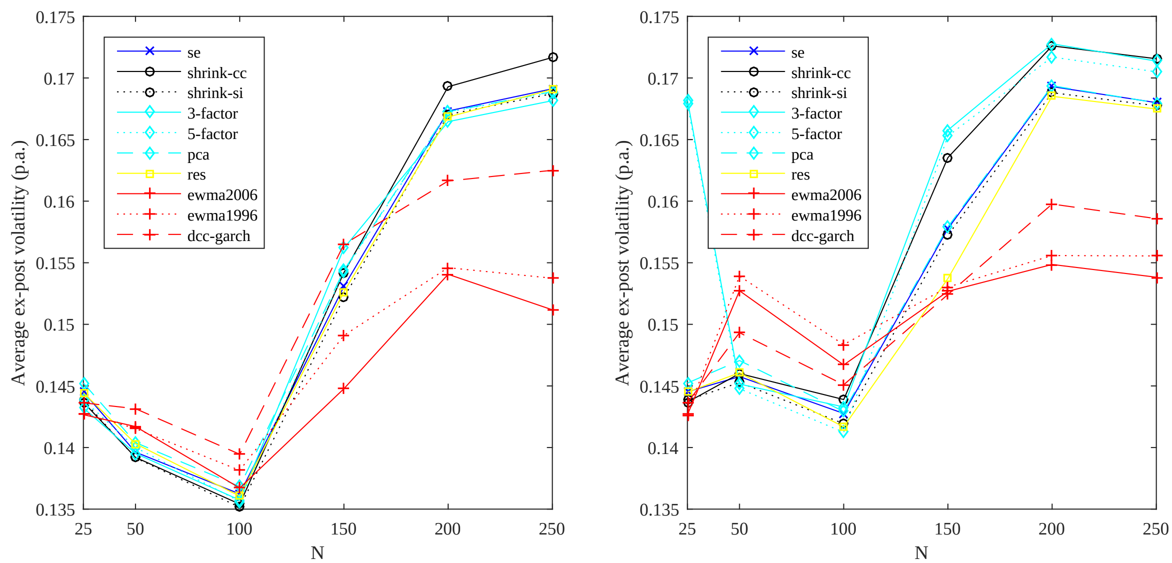
Table 3.1: Out-of-sample (annualised) volatilities for estimated Minimum Variance portfolios

$\hat{\boldsymbol{\Sigma}}^y$	N					
	25	**50**	**100**	**150**	**200**	**250**
U.S.						
se	0.1445	0.1396	0.1362	0.1531	0.1673	0.1691
shrink-cc	0.1437	0.1392	0.1354	0.1542	0.1693	0.1717
shrink-si	0.1438	0.1392	0.1352	0.1522	0.1670	0.1687
fac-3	0.1432	0.1395	0.1357	0.1562	0.1665	0.1682
fac-5	0.1439	0.1400	0.1356	0.1543	0.1672	0.1688
pca	0.1452	0.1404	0.1368	0.1544	0.1673	0.1689
res	0.1444	0.1403	0.1361	0.1526	0.1668	0.1691
ewma1996	0.1428	0.1415	0.1381	0.1491*	0.1546***	0.1537***
ewma2006	0.1427	0.1417	0.1367	0.1448***	0.1541***	0.1512***
dcc-garch	0.1436	0.1432	0.1395	0.1565	0.1616**	0.1625**
Europe						
se	0.1445	0.1458	0.1428	0.1578	0.1693	0.1680
shrink-cc	0.1437	0.1460	0.1439	0.1635	0.1726	0.1716
shrink-si	0.1438	0.1453	0.1419	0.1572	0.1688	0.1677
fac-3	0.1682	0.1452	0.1433	0.1658	0.1728	0.1714
fac-5	0.1680	0.1448	0.1412	0.1653	0.1717	0.1705
pca	0.1452	0.1471	0.1429	0.1579	0.1694	0.1680
res	0.1445	0.1461	0.1417	0.1537	0.1685	0.1675
ewma1996	0.1428	0.1539***	0.1484**	0.1530*	0.1556***	0.1556***
ewma2006	0.1426	0.1527***	0.1467*	0.1527**	0.1549***	0.1538***
dcc-garch	0.1436	0.1493	0.1451	0.1524**	0.1598***	0.1586***

*Ex-post annualised average volatility of the Minimum Variance portfolio between January 2002 and December 2016, subject to the respective covariance matrix estimator. Values are estimated from daily excess returns and portfolios are rebalanced quarterly. F-test for equality of variances between sample covariance portfolio and portfolio optimised subject to the respective estimator: ***p-value of the null hypothesis <1%, **p-value of the null hypothesis <5%, *p-value of the null hypothesis <10%. Explanations to all abbreviations are given in Table 2.1. "U.S." and "Europe" relate to a subset of N stocks drawn from the S&P 500 and Euro Stoxx large cap indices respectively.*

Figure 3.1: Out-of-sample, annualised volatilities for estimated Minimum Variance portfolios

(a) Portfolios drawn from U.S. stocks (S&P 500) (b) Portfolios drawn from European stocks (Euro Stoxx Large)



This figure illustrates the annualised average volatility of the Minimum Variance portfolio between January 2002 and December 2016, subject to the respective covariance matrix estimator. Values are estimated from daily excess returns and portfolios are rebalanced quarterly. Explanations to all abbreviations are given in Table 2.1.

These results stand in stark contrast to Pantaleo et al. (2011), Jagannathan and Ma (2003) or Coqueret and Milhau (2014), who do not include time-varying volatility models in their comparison and find that none of the other regularisation methods performs much better than the sample covariance estimator when short-sales are constrained and $T > N$. Our findings instead suggest that when dealing with a high-dimensional covariance matrix of stock returns (that is approx. $N > 100$), employing *ewma1996, ewma2006* or *dcc-garch* methods for modelling volatilities and covariances reduces estimation risk and, thus, ex-post volatility of the Minimum Variance portfolio. For example, in a set of $N = 200$ assets drawn from the S&P 500 stock universe, the average annualised out-of-sample portfolio volatility using the sample estimator is relatively 8.6% larger than when using the *ewma2006* estimator. *ewma1996 and ewma2006* methods that forecast the covariance matrix under the assumption of constant volatilities and conditional correlations in fact outperform the more sophisticated *dcc-garch* method, with the *ewma2006* estimator as suggested by Zumbach (2006) yielding the lowest ex-post portfolio volatility. While this finding may appear surprising at first, it is in line with Harris and Nguyen (2013) and Zakamulin (2015) who can neither attest any superior forecasting abilities of GARCH methods over EWMA covariance matrix models.

One must point out that the value from using such methods can only be derived when facing a large N, where the sample estimator is particularly noisy. When N is rather small, ex-post portfolio volatility from employing time-varying covariance matrix models can be higher than from using the sample estimator. This in turn means that the benefit from using EWMA and GARCH methods in portfolio optimisation may not be substantial (or in fact negative) to private investors, who tend to hold a smaller number of assets in their portfolio. Furthermore, our findings are not very sensitive to variations of T, which can be explained by the daily frequency of the data.[16] When using daily returns, T is generally substantially larger than N. Hence, increasing N affects the ratio N/T (often referred to as the determinant of estimation risk) relatively more.

Zakamulin (2015), in fact, proves a significantly superior forecasting ability of *ewma1996* and *dcc-garch* methods over the sample estimator in small portfolios where N is no larger than 10. All of his datasets are, however, composed of Fama French, industry or stock and bond index portfolios (all taken as assets) and with $N \leq 10$ the number of assets effectively held by the Minimum Variance portfolio routine is small. To reconcile his findings with ours we, in unreported results, replicated the analysis using 25 value-weighted Fama French portfolios formed on Size and Book-to-Market ratio and obtain similar results. For once, this evidence suggests that EWMA and GARCH models generally work relatively better in forecasting covariance matrix structures of diversified portfolios and indices.[17]

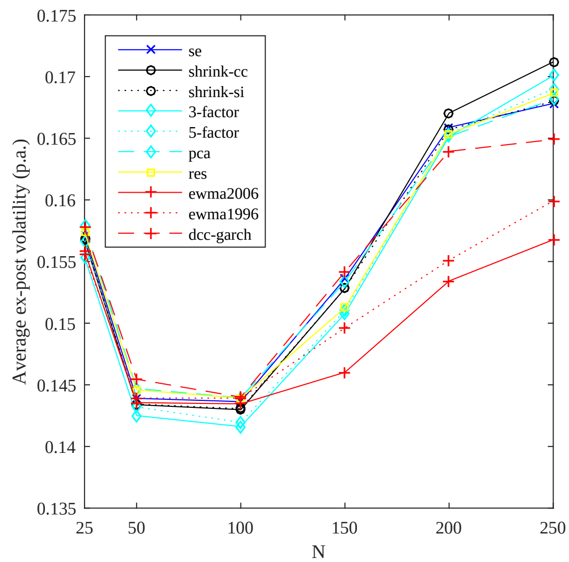
In line with existing literature, we show that estimation risk is not substantially reduced compared to the sample estimator by employing shrinkage methods, factor or principal component models. At least when short-sale constraints are in place, they perform equally well in forecasting the covariance matrix and lead to similar ex-post portfolio volatility. Resampled efficiency as suggested by Michaud (1989), does neither yield any significant improvement over the simple sample covariance estimator, supporting results by Fletcher and Hillier (2001). In all these respects, our results hold when imposing a 10% upper bound on individual portfolio weights. Pantaleo et al. (2011) further demonstrate that portfolios based on covariance regularisation methods are on average more diversified. We can

[16] In unreported results, we test for $T = 252$ and $T = 1008$.

[17] Which they have also initially been intended for.
We also find Zakamulin's analysis to be biased towards time-varying volatility models, since his estimation of the 1-month covariance matrix forecast relies on 120 months of past daily data. A larger T only benefits the sample covariance estimator when correlations and volatilities are stationary, which is unlikely over such long period.

only support this notion for portfolios based on *ewma1996, ewma2006* and *dcc-garch* methods, drawn from the S&P 500 stock universe, since using the sample estimator generally yields portfolios that are comparably (if not more) diversified in terms of average holdings and volatilities (Table 3.2).

Figure 3.2: Out-of-sample, annualised volatilities for 10% maximum weight constrained estimated Minimum Variance portfolios (U.S. stocks)



This figure illustrates the annualised average volatility of the Minimum Variance portfolio (with 10% constraint on the weight of each asset) between January 2002 and December 2016, subject to the respective covariance matrix estimator. Values are estimated from daily excess returns and portfolios are rebalanced quarterly. Explanations to all abbreviations are given in Table 2.1.

Looking at volatility in dependency of N in the S&P 500 stock universe, we observe that it decreases with an expanding opportunity set up until approx. $N = 100$. This corresponds to the natural expectation in the absence of estimation error, as lower volatility arises from a higher diversification benefit. For any larger N, the loss that arises from forecasting error, however, outweighs the gain from a bigger opportunity set and ex-post volatility increases. We additionally find that the effect of limiting asset weights to 10% is dissimilar for small and large N, as revealed by the comparison of Figure 3.2 with Figure 3.1. Consistent with the shrinkage interpretation of weight constraints (e.g. Jagannathan and Ma (2003)), upper limits slightly reduce ex-post volatility for all but EWMA and GARCH estimators when N is large and estimation risk is relatively higher. When is N is small, imposing additional structure to the optimisation problem leads to higher ex-post volatility (which concurs with the ex-ante expectation in the absence of estimation risk).

Table 3.2: Average ex-post Diversification ratio and Effective Number of Stocks for estimated Minimum Variance portfolios (U.S. stocks)

	N					
	25	**50**	**100**	**150**	**200**	**250**
$\overline{\mathbf{DR}}^{\mathbf{y}}_{\mathbf{MV}}$						
se	1.4748	1.6044	1.6492	1.7747	1.8559	1.8592
shrink-cc	1.4558	1.5740	1.6222	1.7534	1.8260	1.8341
shrink-si	1.4674	1.5990	1.6408	1.7684	1.8536	1.8599
fac-3	1.4521	1.5766	1.6014	1.7480	1.8439	1.8471
fac-5	1.4543	1.5804	1.6088	1.7445	1.8359	1.8402
pca	1.4606	1.5941	1.6513	1.7774	1.8616	1.8677
res	1.4848	1.6227	1.6791	1.8026	1.8673	1.8719
ewma1996	1.4769	1.6030	1.6754	1.7901	1.8603	1.8639
ewma2006	1.4824	1.6217	1.6806	1.7926	1.8567	1.8728
dcc-garch	1.4491	1.5420	1.5965	1.7650	1.8115	1.8261
$\overline{\mathbf{ES}}^{\mathbf{y}}_{\mathbf{MV}}$						
se	6.4853	9.1730	16.8818	81.8553	142.8479	183.9476
shrink-cc	6.4679	8.9895	16.9129	82.5847	152.2784	197.5327
shrink-si	6.4599	9.2257	17.1018	81.9928	142.8477	182.0813
fac-3	6.6475	9.3534	17.0108	88.0195	143.7146	181.8329
fac-5	6.6084	9.2560	16.3221	82.7142	143.3930	181.6845
pca	6.0803	8.8346	16.7381	83.4074	141.7944	180.8351
res	6.7763	10.1439	15.4117	78.9028	141.4517	180.8423
ewma1996	5.5821	8.2288	18.7030	97.2466	154.0454	192.6962
ewma2006	5.6520	9.0039	20.1556	88.7622	150.4771	188.8731
dcc-garch	5.6987	7.7341	15.8215	103.0192	164.3124	214.5769

Average Diversification ratio ($\overline{DR}^{y}_{MV}$) and Effective Number of Stocks ($\overline{ES}^{y}_{MV}$) of the Minimum Variance portfolio between January 2002 and December 2016, subject to the respective covariance matrix estimator. Values are estimated from daily excess returns and portfolios are rebalanced quarterly. Explanations to all abbreviations are given in Table 2.1.

4 Application to low and diversified risk investment strategies

4.1 Employed low and diversified risk strategies

Recent years have seen increasing demand for new forms of passive investment benchmarks other than the traditional capitalisation weighted indices. So-called Smart Beta strategies, alternatively weighted index products, systematically re-weight an existing index portfolio subject to quantitative investment rules. The majority of quantitatively optimised Smart Beta indices only relies on portfolio risk optimisation and as a result on an estimate of the covariance matrix. In this section we, therefore, explore the implications of our findings for portfolio managers and index providers of low and diversified risk strategies. A representative set of covariance matrix estimation methods from section 2 is respectively applied to the Minimum Variance, Most Diversified , Equal Risk Contribution and Inverse Volatility (index) portfolio strategies. Optimal portfolio allocations derived from Minimum Variance and Most Diversified optimisation methods are, however, often too concentrated for most institutional investors to hold or to represent an alternatively weighted index. Moreover, funds are mostly bound by short-sale constraints and regulatory requirements that commonly prescribe maximum holdings of no more than 10% in any asset. To adhere with these constraints, we limit the maximum weight of a single asset to 10%. Strategies, tested in addition to the Minimum Variance portfolio allocation can be summarised as follows.

Most Diversified portfolio - Choueifaty and Coignard (2008) propose the Most Diversified portfolio that maximises the ex-ante Diversification ratio of a portfolio and, thus, the distance between portfolio volatility and individual components' volatility. The Most Diversified portfolio is defined as:

$$\hat{\boldsymbol{w}}^{y}_{MDP-C} = arg\ \max_{\boldsymbol{w}}\ \frac{\boldsymbol{w}'\hat{\boldsymbol{e}}^{y}}{\sqrt{\boldsymbol{w}'\hat{\boldsymbol{\Sigma}}^{y}\boldsymbol{w}}} \quad s.t.\ \mathbf{1}'_{\mathbf{N}}\boldsymbol{w} = 1\ and\ 0 \leq \boldsymbol{w} \leq 0.1, \tag{28}$$

where $\hat{\boldsymbol{e}}^{y}$ constitutes the Nx1 vector of estimated, individual assets' volatilities subject to method y. Himbert and Kapraun (2017a) show that Most Diversified portfolios, for example, are significantly more sensitive to estimation error than e.g. Minimum Variance portfolio allocations.

Equal Risk Contribution/Risk Parity portfolio - Qian (2006) and Maillard et al. (2010b) define the marginal risk contribution of an asset to total portfolio volatility that is given by:

$$\frac{\partial \hat{\sigma}_p}{\partial w_i} = \frac{w_i \hat{\sigma}_i^2 + \sum_{j \neq i} w_j \hat{\sigma}_{ij}}{\hat{\sigma}_p}, \tag{29}$$

where $\hat{\sigma}_i^2$ is the estimated sample variance of asset i, $\hat{\sigma}_{ij}$ is the estimated sample covariance of assets i and j (for $i, j = (1, 2, ..., N)$) and $\hat{\sigma}_p = \sqrt{\boldsymbol{w}'\hat{\boldsymbol{\Sigma}}\boldsymbol{w}}$ denotes the sample volatility of the index portfolio. The total risk contribution ψ_i of an asset to portfolio risk is then:

$$\psi_i = w_i \frac{\partial \hat{\sigma}_p}{\partial w_i}. \tag{30}$$

In other words, the index volatility can be decomposed into the sum of total risk contributions. To create portfolios that are more diversified, in the sense of having a more balanced contribution to risk by its constituents, Maillard et al. (2010b) suggest Equal Risk Contribution portfolios - i.e. $\psi_i = \psi_j \, \forall \, i, j$. The Risk Parity portfolio is given by the solution to the following optimisation problem:

$$\hat{\boldsymbol{w}}^y_{ERC-C} = arg \min_{\boldsymbol{w}} g^y(\boldsymbol{w}) \quad s.t. \, \mathbf{1}'_{\mathbf{N}}\boldsymbol{w} = 1 \, and \, 0 \leq \boldsymbol{w} \leq 0.1 \tag{31}$$

$$g^y(\boldsymbol{w}) = \sum_{i=1}^{N} \sum_{j=1}^{N} (w_i(\hat{\boldsymbol{\Sigma}}^y \boldsymbol{w})_i - w_j(\hat{\boldsymbol{\Sigma}}^y \boldsymbol{w})_j)^2,$$

where $(\hat{\boldsymbol{\Sigma}}^y \boldsymbol{w})_i$ denotes the i-th row vector from the product of $\hat{\boldsymbol{\Sigma}}^y$ and $\boldsymbol{w}$ and the covariance matrix estimates subject to method y are used. Hence, the strategy optimal index portfolio is described by the vector of constituent weights that minimises the variance of the rescaled risk contributions. Equal Risk Contribution strategies diversify across all assets of the opportunity set while taking their single and joint risk contributions into account.

Inverse Volatility portfolio - Finally, we employ a heuristic low volatility strategy that does not rely on covariance estimates, but only on direct estimates of volatility. To find the Inverse Volatility weighted portfolio, we estimate the standard deviation of each stock in the underlying portfolio and find its inverse (hence, the stock with the highest ex-ante volatility will have the lowest inverted ex-ante volatility). The portfolio weight of stock i ($i = 1, 2, ..., N$) is then calculated by dividing its inverse volatility $\hat{\sigma}_i^{-1}$ by the sum of all inverted standard deviations and is, thus, given by:

$$\hat{w}^y_{IV-C,i} = \frac{(\hat{\sigma}^y_i)^{-1}}{\sum_{i=1}^N (\hat{\sigma}^y_i)^{-1}}. \tag{32}$$

Note that since the strategy naturally diversifies across all assets of the underlying portfolio, with increasing N it more closely resembles an equally weighted portfolio. It must also hold that $0 \leq \hat{w}^y_{IV-C,i} \leq 0.1$, which is never violated in our dataset (but overall weights would be rescaled where $\hat{w}^y_{IV-C,i} > 0.1$).

4.2 Data and research design

We follow the rolling-window procedure laid out in section 3.1 for assessing the ex-post performance of various strategies, subject to the respective covariance matrix regularisation method over the complete 15-years out-of-sample period from January 2002 through December 2016. N is set equal to the number of all assets that have been part of the S&P 500 large cap index over the 17-year period between January 2000 and December 2016 (i.e. $N = 361$). Every $P = 63$ days we estimate as an input to the portfolio strategy the covariance matrix estimator $\hat{\boldsymbol{\Sigma}}^y$, subject to the chosen regularisation method y, using the $T = 504$ preceding days and the strategy optimal weights are determined. The out-of-sample return series from quarterly rebalancing then serves as the basis for performance evaluation. From an investors point of view, one of the covariance matrix estimation methods can be said to be strictly dominant over another when it allows for an improvement in a strategy's explicit target (i.e. lowest ex-post portfolio volatility or better portfolio diversification in terms of volatilities), while at the same time attaining a similar or better return-to-risk performance (i.e. constant or improved Sharpe ratio). In contrast to common belief, these objectives cannot always be aligned and are, thus, difficult to balance. For example, as pointed out in Himbert and Kapraun (2017a), strategies that (ex-ante) do not explicitly optimise the portfolio towards having the highest Sharpe ratio may not benefit in terms of ex-post Sharpe ratio when reducing estimation risk. Hence, bringing the ex-ante portfolio allocations of the Most Diversified portfolio closer to the ex-post optimal weights ensures an improved Diversification ratio but not necessarily a higher ex-post Sharpe ratio.

4.3 Results from studying low and diversified volatility investment strategies

Our results, presented in Table 4.1, support findings from section 3. Employing *ewma2006* and *dcc-garch* methods for estimating the covariance matrix largely leads to superior attainment of strategy specific targets. The out-of-sample volatility of Minimum Variance, Risk Parity and Inverse Volatility portfolios constructed from *ewma2006* or *dcc-garch* based estimates of the covariance matrix is smaller than ex-post volatility from using the sample estimator. It is striking that employing any other structured estimator in these strategies does not lead to a significant improvement over the sample estimator or even has an adverse affect on out-of-sample volatility. The additional structure that comes from using such regularisation methods does clearly not outweigh the accepted bias that arises. For the Most Diversified Portfolio, on the other hand, the improvement in Diversification ratio from employing any estimators other than the sample estimator is not substantial. Here only portfolio resampling appears to have an edge over other methods, which may be explained by the particular strength of this method in controlling estimation risk that purely arises from sampling error.[18]

Consistent with Chan et al. (1999) we find the heuristic Inverse Volatility weighting of constituents to perform not much worse than Minimum Variance portfolio allocation, both in terms of volatility and performance ratios. It, in fact, attains a lower out-of-sample portfolio volatility when using the sample estimator but is on average less diversified than portfolios from Minimum Variance asset allocation. In this regard, note that Risk Parity portfolios are most diversified in terms of average portfolio holdings but also have a lower Diversification ratio than e.g. Minimum Variance portfolios (which proponents of Equal Risk Contribution strategies often fail to mention).

With respect to overall portfolio performance we find the Most Diversified portfolio to post the highest Sharpe ratio in our set of strategies, which comes at the cost of highest ex-post volatility. In consideration of the Omega ratio, the share of realised positive returns yet outweighs the realised lower partial moment of the return series the most, compared to all other strategies. Achieving the largest ex-post diversification through improved covariance matrix estimation does, however, not align with the best ex-post portfolio performance as measured by both Sharpe and Omega ratios. When managers are benchmarked by such portfolio performance measures, they may be better off using

[18] As pointed out by Himbert and Kapraun (2017a) the MDP strategy is subject to significantly more estimation risk that stems from sampling error than all other low risk strategies. It so appears that resampling only benefits where sampling error is high and otherwise induces unnecessary, additional estimation error.

Table 4.1: Out-of-sample performance statistics for portfolios based on selected covariance matrix estimators - January 2002 to December 2016 (U.S. stocks)

	Return	**Volatility**	$\overline{\mathbf{SR}}^{\mathbf{y}}$	$\overline{\mathbf{OR}}^{\mathbf{y}}$	$\overline{\mathbf{DR}}^{\mathbf{y}}$	$\overline{\mathbf{ES}}^{\mathbf{y}}$	$\overline{\mathbf{GC}}^{\mathbf{y}}$
Minimum Variance							
se	0.0568	0.1730	0.3282	1.0640	1.8868	268.6064	0.2491
shrink-si	0.0557	0.1732	0.3217	1.0627	1.8868	268.3366	0.2484
fac-5	0.0560	0.1733	0.3234	1.0631	1.8831	269.2987	0.2516
pca	0.0559	0.1741	0.3209	1.0626	1.8867	268.2286	0.2435
res	0.0577	0.1754	0.3292	1.0644	1.8847	282.2497	0.2296
ewma2006	0.0595	0.1650	0.3604	1.0682	1.8945	270.5425	0.2508
dcc-garch	0.0621	0.1633	0.3804	1.0705	1.8605	298.0290	0.2232
Most Diversified							
se	0.1631	0.2601	0.6270	1.1228	2.0079	15.4293	0.9698
shrink-si	0.1608	0.2653	0.6061	1.1184	2.0036	15.7442	0.9692
fac-5	0.1603	0.2749	0.5832	1.1145	1.9873	16.5568	0.9678
pca	0.1637	0.2629	0.6225	1.1221	2.0044	15.3918	0.9697
res	0.1248	0.2252	0.5541	1.1081	2.1014	35.7721	0.9262
ewma2006	0.1185	0.2216	0.5346	1.1021	2.0161	14.2232	0.9728
dcc-garch	0.1413	0.2561	0.5517	1.1041	2.0005	15.0371	0.9712
Risk Parity							
se	0.0631	0.1857	0.3397	1.0668	1.8505	324.8385	0.1800
shrink-si	0.0628	0.1853	0.3388	1.0666	1.8510	323.6529	0.1803
fac-5	0.0635	0.1854	0.3426	1.0673	1.8508	323.0893	0.1824
pca	0.0631	0.1857	0.3397	1.0667	1.8503	324.8618	0.1799
res	0.0630	0.1858	0.3390	1.0666	1.8501	325.7256	0.1796
ewma2006	0.0601	0.1824	0.3296	1.0643	1.8550	324.7892	0.1875
dcc-garch	0.0615	0.1823	0.3373	1.0657	1.8442	325.5477	0.1738
Inverse Volatility							
se	0.0583	0.1729	0.3373	1.0660	1.7877	260.0170	0.2775
shrink-si	0.0583	0.1729	0.3373	1.0660	1.7877	260.0170	0.2775
fac-5	0.0583	0.1729	0.3371	1.0660	1.7874	260.0736	0.2776
pca	0.0583	0.1729	0.3373	1.0660	1.7877	260.0170	0.2775
res	0.0583	0.1729	0.3370	1.0660	1.7888	260.7114	0.2756
ewma2006	0.0566	0.1671	0.3388	1.0656	1.7893	251.7814	0.2839
dcc-garch	0.0556	0.1668	0.3335	1.0643	1.7804	246.7405	0.2985

$\overline{SR}^y$: Sharpe ratio; $\overline{OR}^y$: Omega ratio - the ratio of upper to lower partial moments of the respective ex-post return series as suggested by Keating and Shadwick (2002); $\overline{DR}^y$: Average Diversification ratio; $\overline{ES}^y$: Average Effective Number of Stocks; $\overline{GC}^y$: Average Gini coefficient of risk contributions - a measure of differences within a distribution of risk contributions as suggested by Maillard et al. (2010b), 0 (no differences) - 1 (large dispersion). Values are estimated from daily excess returns and portfolios are rebalanced quarterly.

the sample estimator.[19] For Minimum Variance portfolios, however, the covariance matrix forecasts derived from using *ewma2006* and *dcc-garch* methods lead to both, a reduction in portfolio volatility and an improvement in portfolio performance. The implications of this finding are far-reaching, since the benefits of using these models as a general input in mean-variance portfolio optimisation could be substantial.

5 Conclusion

It has been established in literature that low and diversified volatility portfolio optimisation is significantly affected by estimation errors in the covariance matrix. This lead to a variety of commercialised methodologies becoming available, which promise more robust estimation of the input parameters. In section 2 we provide a comprehensive overview of such improved estimation methods, available to investors today. We then investigate (i) whether there is any value to be derived by practitioners in employing these regularisation methods over the simple sample covariance matrix estimator and if so, (ii) under which conditions (with respect to the size of the asset universe) they are enhancing the out-of-sample performance of the corresponding optimal portfolios? From studying the Minimum Variance portfolio, the majority of literature (e.g. Pantaleo et al. (2011); Jagannathan and Ma (2003)) finds that when portfolios are short-sale constrained estimation risk is not substantially reduced compared to the sample estimator by using a variety of covariance matrix regularisation methods. But Zakamulin (2015) recently proposed adding EWMA and GARCH methods to the comparison and attests a superior forecasting ability in low-dimensional (diversified assets) covariance matrix problems over the sample estimator. Our findings suggest that employing *ewma1996, ewma2006* or *dcc-garch* methods only leads to a significant improvement in the covariance matrix forecast and, thus, better ex-post portfolio performance when dealing with a high-dimensional covariance matrix of stock returns. Across nine tested covariance regularisation methods, the *ewma2006* estimator as suggested by Zumbach (2006) yields the lowest ex-post portfolio volatility of the Minimum Variance portfolio (across both U.S. and European datasets) when the set of assets includes more than approx. 100 stocks. For any smaller set of stocks, none of the estimators improves much on the sample covariance matrix. Furthermore, consistent with the shrinkage interpretation of maximum weight constraints, we find that imposing a weight limit can reduce ex-post volatility in high-dimensional covariance matrix estimation problems. When portfolio are small, however, imposing additional structure to optimisation problem generally

[19] The same goes for the Equal Risk Contribution and Inverse Volatility strategies.

leads to worse out-of-sample performance.

We then extend our findings to popular low and diversified risk asset allocation rules and show that employing *ewma2006* and *dcc-garch* methods for estimating the covariance matrix generally leads to a superior attainment of strategy specific targets and lower realised portfolio volatility. In addition, we confirm that employing any other structured estimator in these strategies does not lead to a significant improvement over the sample covariance matrix. With regard to portfolio performance (beyond strategy specific targets) we find that only for Minimum Variance portfolios the covariance matrix forecasts derived from using *ewma2006* and *dcc-garch* models lead to both, a reduction in portfolio volatility and an improvement in portfolio performance. This underlines our hypothesis that in contrast to common belief, objectives of reducing estimation risk and attaining best performance in terms of measures such as the Sharpe or Omega ratios cannot always be aligned. Or in other words, there may persist positive effects of estimation risk for benchmarked portfolio managers.

References

Alessi, L., Barigozzi, M., and Capasso, M. (2008). A robust criterion for determining the number of static factors in approximate factor models. *European Central Bank Working Paper Series*, (903).

Alessi, L., Barigozzi, M., Capasso, M., Alessi, L., Barigozzi, M., and Capasso, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters*, 80(23-24):1806–1813.

Alexander, C. (2001). Principal Component Models for Generating Large GARCH Covariance Matrices.

Alexander, C. (2008). Moving Average Models for Volatility and Correlation , and Covariance Matrices. In *Handbook of Finance*.

Asness, C. S., Frazzini, A., and Pedersen, L. H. (2012). Leverage Aversion and Risk Parity. *Financial Analysts Journal*, 68(1):47–59.

Bai, B. Y. J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1):191–221.

Baker, M. and Wurgler, J. (2011). Benchmarks as Limits to Arbitrage : Understanding the Low Volatility Anomaly. *Financial Analysts Journal*, 67(1):40–54.

Bauwens, L. U. C., Ebastien, S., and Rombouts, J. V. K. (2006). Multivariate GARCH models : A survey. 109(April 2003):79–109.

Becker, F., Gürtler, M., and Hibbeln, M. (2015a). Markowitz versus Michaud : portfolio optimization strategies reconsidered. *The European Journal of Finance*, 21(4):269–291.

Becker, R., Clements, A. E., Doolan, M. B., and Hurn, A. S. (2015b). Selecting volatility forecasting models for portfolio allocation purposes. *International Journal of Forecasting*, 31:849–861.

Benartzi, S. and Thaler, R. (2001). Naive Diversification Strategies in Defined Contribution Saving Plans. *The American Economic Review*, 91(1):79–98.

Chan, L., Karceski, J., and Lakonishok, J. (1999). On Portfolio Optimization : Forecasting Covariances and Choosing the Risk Model. *The Review of Financial Studies*, 12(5):937–974.

Chopra, V. K. and Ziemba, W. T. (1993). The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice. *The Journal of Portfolio Management*, 19(2):6–11.

Choueifaty, Y. and Coignard, Y. (2008). Toward Maximum Diversification. *The Journal of Portfolio Management*, 35(1):35–41.

Choueifaty, Y., Froidure, T., and Reynier, J. (2013). Properties of the Most Diversified Portfolio. *Journal of Investment Strategies*, 2(2):49–70.

Chow, T.-m., Hsu, J., Kalesnik, V., and Little, B. (2011). A survey of alternative equity index strategies. *Financial Analysts Journal*, 67(5):37–57.

Chow, T.-M., Hsu, J., Kuo, L.-L., and Li, F. (2014). A Study of Low-Volatility Portfolio Construction Methods. *The Journal of Portfolio Management*, 40(4).

Clarke, R., Silva, H. D., and Thorley, S. (2006). Minimum-Variance Portfolios in the U.S. Equity Market. *The Journal of Portfolio Management*, 33(1):10–24.

Coqueret, G. and Milhau, V. (2014). Estimating Covariance Matrices for Portfolio Optimization.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal Versus Naive Diversification : How Inefficient is the 1 / N Portfolio Strategy ? *The Review of Financial Studies*, 22(5):1916–1953.

Elton, E. J. and Gruber, M. J. . (1973). Estimating the Dependence Structure of Share Prices - Implications for Portfolio Selection. *The Journal of Finance*, 28(5):1203–1232.

Engle, R. (2002). Dynamic Conditional Correlation : A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economics Statistics*, 20(3):339–350.

Engle, R. F. (1993). Statistical Models for Financial Volatility. *Financial Analysts Journal*, 49(1):72–78.

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometrics Review*, 5(1):1–50.

Engle, R. F. and Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH.

Fama, E. F. and French, K. R. (1993). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Fletcher, J. and Hillier, J. (2001). An Examination of Resampled Portfolio Efficiency An Examination of Resampled Portfolio Efficiency. *Financial Analysts Journal*, 57(5):66–74.

Frahm, G. (2015). A theoretical foundation of portfolio resampling. *Theory and Decision*, (June 2014):107–132.

Gomez, K. and Gallon, S. (2011). Comparison among High Dimensional Covariance Matrix Estimation Methods. *Revista Colombiana de Estadistica*, 34(3):567–588.

Gosier, K., Madhavan, A., Serbin, V., and Yang, J. (2005). Toward Better Risk Forecast. *The Journal of Portfolio Management*, 31(3):82–91.

Harris, R. and Nguyen, A. (2013). Long Memory Conditional Volatility and Asset Allocation. *International Journal of Forecasting*, 29(2):258–273.

Haugen, R. A. and Baker, N. L. (1991). The efficient market inefficiency of capitalization-weighted stock portfolios. *The Journal of Portfolio Management*, 17(3):45–50.

Herold, U. and Maurer, R. (2006). Portfolio Choice and Estimation Risk. A Comparison of Bayesian to Heuristic Approaches. *Astin Bulletin*, 36(1):135–160.

Himbert, B. and Kapraun, J. (2017a). On the trade-off between optimality and estimation risk in alternatively weighted equity index strategies.

Himbert, B. and Kapraun, J. (2017b). The sensitivity of estimation risk to portfolio size: why having more choice is not always better.

Hlouskova, J., Schmidheiny, K., and Wagner, M. (2009). Multistep predictions for multivariate GARCH models : Closed form solution and the value for portfolio management. *Journal of Empirical Finance*, 16(2):330–336.

Ingersoll, J. E. (1987). *Theory of Financial Decision Making*.

Jagannathan, R. and Ma, T. (2003). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *The Journal of Finance*, 58(4):1651–1683.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Fourth Berkeley Symposium*.

Jorion, P. (1986). Bayes-Stein Estimation for Portfolio Analysis. *The Journal of Financial and Quantitative Analysis*, 21(3):279–292.

Kan, R. and Zhou, G. (2007). Optimal Portfolio Choice with Parameter Uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656.

Keating, C. and Shadwick, W. F. (2002). A Universal Performance Measure. *Journal of Performance Measurement*, 6(3):59–84.

Kondor, I., Gabor, P., and Caccioli, F. (2017). Analytic solution to variance optimization with no short positions.

Kondor, I., Pafka, S., and Gabor, N. (2007). Noise sensitivity of portfolio selection under various risk measures. *Journal of Banking & Finance*, 31(5):1545–1573.

Laloux, L., Cizeau, P., Potters, M., and Bouchaud, J.-p. (2000). Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance*, 3(3):391–397.

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.

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

Longerstaey, J. (1996). RiskMetrics Technical Document.

Maillard, S., Roncalli, T., and Demey, P. (2010a). Risk-Based Indexation.

Maillard, S., Roncalli, T., and Teiletche, J. (2010b). The properties of equally-weighted risk contribution portfolios. *The Journal of Portfolio Management*, 36(4):60–70.

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

Markowitz, H. and Usmen, N. (1996). The likelihood of various stock market return distributions, Part 2: Empirical results. *Journal of Risk and Uncertainty*, 13(3):221–247.

Merton, R. C. (1980). On Estimating the Expected Return on the Market : An Exploratory Investigation. *Journal of Financial Economics*, 8(4):323–361.

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

Nawrocki, D. N. (1991). Optimal algorithms and lower partial moment : ex post results. *Applied Economics*, 3.

Pantaleo, E., Tumminello, M., Lillo, F., and Mantegna, R. N. (2011). When do improved covariance matrix estimators enhance portfolio optimization ? An empirical comparative study of nine estimators When do improved covariance matrix estimators enhance portfolio optimization ? An empirical comparative study of nine estimato. *Quantitative Finance*, 11(7):1067–1080.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, A. N., Guhr, T., and Stanley, H. E. (2002). A Random Matrix Approach to Cross-Correlations in Financial Data. *Physical Review E*, 65(6):66–126.

Qian, E. (2006). On the financial interpretation of risk contribution: Risk budgets do add up. *Journal of Investment Management*, 4(4):41–51.

Ross, S. A. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, 13:341–360.

Scherer, B. (2002). Portfolio Resampling : Review and Critique. *Financial Analysts Journal*, 58(6):98–109.

Scherer, B. (2007). *Portfolio Construction and Risk Budgeting*. Risk Books.

Sharpe, W. F. (1964). A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, 19(3):425–442.

Zakamulin, V. (2015). A Test of Covariance-Matrix Forecasting Methods. *The Journal of Portfolio Management*, 41(3):97–108.

Zumbach, G. (2004). Volatility processes and volatility forecast with long memory. *Quantitative Finance*, 4:70–86.

Zumbach, G. (2006). The RiskMetrics 2006 methodology.