Improved Forecasts of Higher-Order Co-moments and Implications for Portfolio Selection

Lionel Martellini
PhD, Professor of Finance and Scientific Director of the EDHEC Risk and Asset Management Research Centre

Volker Ziemann
PhD, Research Engineer at the EDHEC Risk and Asset Management Research Centre
Abstract

In the presence of non-normally distributed asset returns, optimal portfolio selection techniques require not only estimates of variance-covariance parameters, but also estimates of higher-order moments and comoments of the return distribution. This paper expands on existing literature, which has focused mostly on the covariance matrix, by introducing improved estimators for the coskewness and cokurtosis parameters. In an empirical analysis, we find that the use of these enhanced estimates leads to significantly better out-of-sample performance.
Introduction

Past research has shown that mean-variance portfolio selection techniques can involve a severe welfare loss in the presence of non-quadratic preferences and non-normally distributed asset returns (see Jondeau and Rockinger 2006 or Harvey et al. 2006 for recent references). Under relatively weak assumptions regarding the shape of the utility function, Horvath and Scott (1980) actually show that typical investors exhibit non-trivial preferences with respect to portfolio higher moments, in addition to mean and variance. In an attempt to better approximate investor preferences, and because most asset returns exhibit strong deviations from normality, optimal portfolio selection techniques should therefore involve higher-order comoments of asset return distribution as inputs, in addition to the covariance matrix. The need to estimate coskewness and cokurtosis parameters however, greatly increases the dimensionality problem, already a serious concern in the context of covariance matrix estimation. For example, it can be shown that optimising a 10 stock (25 stock) portfolio would involve the estimate of 55 (325) variance-covariance parameters, while it would require 220 (2,925) skewness-coskewness parameters and 715 (20,475) kurtosis-cokurtosis parameters! In this context, given the formidable increase in dimensionality involved in higher-order moment parameter estimation, one might wonder whether this critically useful improvement over mean-variance portfolio selection techniques can efficiently be implemented at all in realistic situations.

In the mean-variance context, it has long been recognised that the sample covariance matrix of historical returns is likely to generate high sampling error in the presence of many assets, and several methods have been introduced to improve asset return covariance matrix estimation. One solution is to impose some structure on the covariance matrix to reduce the number of parameters to be estimated. Such "structured estimators" of the variance-covariance matrix include the constant correlation forecast (Elton and Gruber 1973), the single-factor forecast (Sharpe 1963) and the multi-factor forecast (Chan et al. 1999). In these approaches, sampling error is reduced at the cost of specification error. Several authors have subsequently studied the optimal trade-off between sampling risk and model risk in the context of optimal shrinkage theory. This includes optimal shrinkage towards the grand mean (Jorion 1985, 1986), optimal shrinkage towards the single-factor model (Ledoit and Wolf 2003) or optimal shrinkage towards the constant correlation estimate (Ledoit and Wolf 2004). Also related is a paper by Jagannathan and Ma (2003), who argue that imposing no short sales constraints also entails a shrinkage effect and yields comparable results in terms of out-of-sample standard deviation.

While several methods have been introduced to reduce the number of covariance parameters in a portfolio selection context, very little, if anything, is known about improved forecasts for higher-order comoments of asset return distributions. Previous literature has in fact focused mostly on asset pricing implications of non-trivial preferences for higher moments, including seminal contributions by Kraus and Litzenberger (1976) and Harvey and Siddique (2000), who have reported empirical evidence of the presence of risk premia associated with higher-order moments of portfolio return distributions (see also Dittmar 2002 as well as Ang et al. 2006). Very few papers have actually focused on the problem of improving higher moments estimates. Kim and White (2004) have introduced robust estimates of skewness and kurtosis parameters based on inter-quartile differences but they focus merely on a univariate analysis, with no effort on improving higher-order comoments, obviously the main concern for portfolio selection techniques. Also related is a paper by Harvey et al. (2006), who introduce a semi-parametric approach to model skewness in multivariate portfolio allocation process and address the problem of parameter uncertainty in a skew normal framework. Finally, in a series of recent papers, Jondeau and Rockinger (2003, 2005) have considered asset allocation models based on conditional models for higher moments of asset return distributions. Their analysis, however, focuses on a different key aspect in risk parameter estimation, namely the issue of stationarity risk, as opposed to sample risk.

This paper expands on existing literature on improved estimates of the variance-covariance matrix by introducing enhanced estimates of the coskewness and cokurtosis parameters. More specifically, our contribution from the theoretical standpoint is to introduce suitable extensions to higher-order comoments of several models that have been found useful in improving covariance matrix forecasts, namely the constant correlation estimator, the
factor-based estimator, as well as statistical shrinkage estimators. Empirical performance of these estimators is subsequently assessed in the context of a horse race analysis similar to the one done by Chan et al. (1999) and Jagannathan and Ma (2003). Our results show that these improved estimators perform better than the sample estimators by reducing out-of-sample even moments of portfolio returns (variance, kurtosis) and increasing out-of-sample odd moments of portfolio returns (mean, skewness). Additionally, these enhanced estimators are less volatile, leading to more stable portfolio allocations.

The rest of the paper is organised as follows. In section 1, we motivate the introduction of higher moments in optimal portfolio selection. In section 2, we extend to the third and fourth orders the constant correlation and factor-based approaches originally introduced to improve estimates of the variance-covariance matrix, and also derive explicit expressions for optimal shrinkage intensities for higher-order moments in the spirit of the results of Ledoit and Wolf (2003, 2004). Section 3 assesses the empirical performance of the derived estimators. In section 4, we present our conclusions, while technical details are relegated to an appendix.

I. Portfolio Choice with Higher Moments

To better assess the impact of higher-order moments of asset returns on portfolio selection techniques, we consider a standard expected utility maximisation framework. For infinitely differentiable utility functions $U$, one can approximate the utility of terminal wealth through the following Taylor expansion:

$$ U(W) = \sum_{k=0}^{\infty} \frac{U^{(k)}(E(W))}{k!} (W - E(W))^k $$

(1)

where $W$ is a random variable representing investor wealth. Because preference theory does not reveal intuitive interpretations for additional polynomial terms (Kimball 1993 or Dittmar 2002), one typically assumes that utility is well approximated by a fourth-order Taylor expansion:

$$ U(W) = U(E(W)) + \sum_{k=2}^{4} \frac{U^{(k)}(E(W))}{k!} (W - E(W))^k $$

(2)

Applying the expectation operator to both sides of equation (2) enables us to approximate expected utility as:

$$ E[U(W)] = U(E(W)) + \frac{U^{(2)}(E(W))}{2} \mu^{(2)} + \frac{U^{(3)}(E(W))}{6} \mu^{(3)} + \frac{U^{(4)}(E(W))}{24} \mu^{(4)} $$

(3)

with $\mu^{(n)}$ being the $n$th-order centred moment:

$$ \mu^{(n)} = E\left[(W - E(W))^n\right]. $$

(4)

From (5) we learn that expected utility deriving from an investment in risky assets can be approximated by the derivatives of the utility function and the first four moments of asset return distribution. Consistent with findings reported in Horvath and Scott (1980), investors are assumed to have preferences for higher odd and lower even moments. Hence, the portfolio choice is no longer a trade-off between expected return and volatility and optimal portfolios can be regarded as tangency points in a 4-dimensional space, incorporating expected return, second, third and fourth centred moments of asset returns.
Consistent with this fourth-order approximation of the investor’s expected utility, the maximisation program can be written as:

$$\max_w \left[ U(E(R)) + \frac{U^{(2)}(E(R))}{2} \mu^{(2)} + \frac{U^{(3)}(E(R))}{6} \mu^{(3)} + \frac{U^{(4)}(E(R))}{24} \mu^{(4)} \right]$$  \( (5) \)

where \( w \) denotes the \( nxn \) weighting vector corresponding to the \( n \) available risky assets in the portfolio. The next step is to understand how the portfolio moments are related to the joint distributions of individual assets in the portfolio. As in Harvey et al. (2006) and Jondeau and Rockinger (2006), we first define third and fourth order comoments as:

$$s_{jk} = E \left[ (R_i - \mu_i)(R_j - \mu_j)(R_k - \mu_k) \right] \quad \forall \ i, j, k = \ln$$  \( (6) \)

$$k_{ijkl} = E \left[ (R_i - \mu_i)(R_j - \mu_j)(R_k - \mu_k)(R_l - \mu_l) \right] \quad \forall \ i, j, k, l = \ln$$  \( (6) \)

and we use the Kronecker product to introduce a compact notation for the coskewness and cokurtosis terms. Hence, in addition to the variance-covariance matrix (denoted by \( M_{1} \)), we define the tensor matrices \( M_{2}, M_{3}, M_{4} \) incorporating all combinations defined by (6):

$$M_{2} = E \left[ (R - E(R))(R - E(R))' \right]$$

$$M_{3} = E \left[ (R - E(R))(R - E(R))' \otimes (R - E(R))' \right]$$

$$M_{4} = E \left[ (R - E(R))(R - E(R))' \otimes (R - E(R))' \otimes (R - E(R))' \right].$$  \( (7) \)

This notation allows us to remain in the matrix space when analysing high dimensional tensors. For example, we obtain for 3 assets:

$$M_{3} = \begin{bmatrix}
S_{111} & S_{112} & S_{113} & S_{121} & S_{122} & S_{123} & S_{131} & S_{132} & S_{133} \\
S_{211} & S_{212} & S_{213} & S_{221} & S_{222} & S_{223} & S_{231} & S_{232} & S_{233} \\
S_{311} & S_{312} & S_{313} & S_{321} & S_{322} & S_{323} & S_{331} & S_{332} & S_{333}
\end{bmatrix}$$  \( (8) \)

Note that \( S_{ijk} \) and \( K_{ijkl} \) denote \( nxn \) matrices. Consequently, again using the Kronecker product, the higher-order moments of portfolio return in (5) can be represented in terms of \( M_{2}, M_{3}, M_{4} \):

$$\mu^{(2)} = wM_{2}w$$

$$\mu^{(3)} = wM_{3}(w \otimes w)$$

$$\mu^{(4)} = wM_{4}(w \otimes w \otimes w)$$  \( (9) \)

Hence, supposing that the wealth \( W \) in (5) is entirely determined by the portfolio outcome \( R_{p} \), we can then rewrite the investor optimisation problem as a function of asset return comoments and portfolio weights:

$$\max_w \left[ U(E(\mu w)) + \frac{U^{(2)}(\mu w)}{2} wM_{2}w + \frac{U^{(3)}(\mu w)}{6} wM_{3}(w \otimes w) + \frac{U^{(4)}(\mu w)}{24} wM_{4}(w \otimes w \otimes w) \right]$$  \( (10) \)

We now turn to the problem of improving empirical estimates of the higher-order moment tensor matrices entering the portfolio selection problem.
II. Improved Estimators for Higher-Order Moment Matrices

As recalled in the introduction, the number of parameters involved when higher-order moments are taken into account increases greatly in the number of risky assets in the portfolio (see Table I).

Table I. Required number of parameters - Sample estimator
The numbers correspond to the amount of required parameters for each tensor. \( N \) denotes the number of assets in the portfolio.

<table>
<thead>
<tr>
<th>( N )</th>
<th>( M_2 )</th>
<th>( M_3 )</th>
<th>( M_4 )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>6</td>
<td>10</td>
<td>15</td>
<td>31</td>
</tr>
<tr>
<td>10</td>
<td>55</td>
<td>220</td>
<td>715</td>
<td>990</td>
</tr>
<tr>
<td>25</td>
<td>325</td>
<td>2,925</td>
<td>20,475</td>
<td>23,725</td>
</tr>
<tr>
<td>100</td>
<td>5,050</td>
<td>171,700</td>
<td>4,421,275</td>
<td>4,598,025</td>
</tr>
</tbody>
</table>

Given the total numbers of parameters to estimate, the higher-order moment tensor matrices will be rank-deficient even for relatively small portfolios. For 25 assets, for instance, one would need 80 years of monthly data to ensure that the number of observations exceeds the maximal possible rank of the corresponding moment tensors. Following the literature on improved estimates of the variance-covariance matrix, we now seek to impose some structure on the higher-order moment tensor matrices so as to reduce the number of parameters involved.

A. Constant Correlation Estimator

This estimator has been proposed by Elton and Gruber (1973) as a response to the statistical challenge related to estimating covariance matrices for portfolios involving a large number of assets. In a nutshell, Elton and Gruber argue that imposing the assumption of a constant correlation across assets, while obviously involving a strong model risk, leads to improved out-of-sample portfolio performance. It can easily be seen that an unbiased estimator of the common correlation coefficient is given by:

\[
\hat{r} = \frac{2}{N(N-1)} \sum_{i=2}^{N} \sum_{j=i+1}^{N} \hat{r}_{ij} \quad \text{with} \quad \hat{r}_{ij} = \frac{s_{ij}}{\sqrt{s_i s_j}} \quad (11)
\]

where \( s \) denotes sample variances and covariances. Following this approach, the covariances \( (\sigma_{ij}) \) can be estimated as a function of the constant correlation parameter and the sample standard deviations of the assets \( \sigma_y = \sqrt{s_i s_j} \), thus allowing a dramatic reduction in the number of parameters. Following the pioneering work by Elton and Gruber, empirical research has assessed out-of-sample properties of portfolio constructed upon the constant correlation estimator. In particular, Ledoit and Wolf (2003), Jagannathan and Ma (2003), and Chan et al. (1999), among others, show that the realized volatility for minimum variance portfolios constructed with the sample estimator of the covariance matrix significantly exceeds the realized volatility of portfolios constructed with the constant correlation estimator.

Following these results, we now derive the counterpart of correlation coefficients for higher-order comoments and extend the concept of a constant correlation to the context of higher-order moment tensors. We first write down explicitly all possible combinations of higher-order comoments according to the definition of the matrices \( M_3 \) and \( M_4 \) as in (6). Denoting \( X \) the centred version of the variable of \( X \), we have:

\[
\begin{align*}
  s_{ij} &= E(\overline{X}_i \overline{X}_j) \\
  s_{jk} &= E(\overline{X}_j \overline{X}_k) \\
  k_{ij} &= E(\overline{X}_i \overline{X}_j) \\
  k_{ij} &= E(\overline{X}_i \overline{X}_j) \\
  k_{ijkl} &= E(\overline{X}_i \overline{X}_j \overline{X}_k \overline{X}_l) \\
  \forall \ i \neq j \neq k \neq l
\end{align*}
\]
According to several permutations in (6), we introduce the following extended correlation coefficients:

\[
\begin{align*}
    r^{(1)}_{ij} & = \frac{E(R_i R_j)}{\sqrt{\mu_i^{(2)} \mu_j^{(2)}}} & r^{(2)}_{ij} & = \frac{E(R_i^2 R_j)}{\sqrt{\mu_i^{(4)} \mu_j^{(2)}}} \\
    r^{(3)}_{ij} & = \frac{E(R_i^3 R_j)}{\sqrt{\mu_i^{(6)} \mu_j^{(2)}}} & r^{(4)}_{ik} & = \frac{E(R_i R_j R_k)}{\sqrt{\mu_i^{(2)} E(R_j^2 R_k^2)}} \\
    r^{(5)}_{ij} & = \frac{E(R_i^5 R_j)}{\sqrt{\mu_i^{(4)} \mu_j^{(4)}}} & r^{(6)}_{jk} & = \frac{E(R_i^2 R_j^2 R_k)}{\sqrt{\mu_i^{(4)} E(R_j^2 R_k^2)}} \\
    r_{ijkl}^{(7)} & = \frac{E(R_i R_j R_k R_l)}{\sqrt{E(R_i R_j R_k R_l) E(R_i R_j R_k R_l)}}
\end{align*}
\]

Cauchy-Schwarz inequality \(|E(XY)| \leq \sqrt{E(X^2) E(Y^2)}\) ensures that the seven extended correlation coefficients are bounded:

\[
r^{(n)} \in [-1, 1] \quad n = 1 \ldots 7.
\]

Note that \(r^{(1)}\) is the standard correlation coefficient. Consistent with Elton and Gruber, we will estimate \(r^{(2)} - r^{(7)}\) and set those coefficients constant across the assets. In other words, we use the sample data to calibrate the 7 coefficients, with unbiased estimators given by:

\[
\begin{align*}
    \hat{r}^{(1)} & = \frac{2}{N(N-1)} \sum_{i < j} \sum_{t=1}^{T} \frac{(\bar{R}_{it} \bar{R}_{jt})}{\sqrt{m_i^{(2)} m_j^{(2)}}} \\
    \hat{r}^{(2)} & = \frac{2}{N(N-1)} \sum_{i < j} \sum_{t=1}^{T} \frac{(\bar{R}_{it}^2 \bar{R}_{jt})}{\sqrt{m_i^{(4)} m_j^{(2)}}} \\
    \hat{r}^{(3)} & = \frac{2}{N(N-1)} \sum_{i < j} \sum_{t=1}^{T} \frac{(\bar{R}_{it} \bar{R}_{jt}^2)}{\sqrt{m_i^{(4)} m_j^{(4)}}} \\
    \hat{r}^{(4)} & = \frac{6}{N(N-1)(N-2)} \sum_{i < j} \sum_{t=1}^{T} \frac{(\bar{R}_{it} \bar{R}_{jt})}{\sqrt{m_i^{(2)} m_j^{(5)}}} \\
    \hat{r}^{(5)} & = \frac{2}{N(N-1)} \sum_{i < j} \sum_{t=1}^{T} \frac{(\bar{R}_{it}^2 \bar{R}_{jt})}{\sqrt{m_i^{(4)} m_j^{(4)}}} \\
    \hat{r}^{(6)} & = \frac{6}{N(N-1)(N-2)} \sum_{i < j} \sum_{t=1}^{T} \frac{(\bar{R}_{it} \bar{R}_{jt}^2)}{\sqrt{m_i^{(4)} m_j^{(5)}}} \\
    \hat{r}^{(7)} & = \frac{24}{N(N-1)(N-2)(N-3)} \sum_{i < j < k < l} \sum_{t=1}^{T} \frac{(\bar{R}_{it} \bar{R}_{jt} \bar{R}_{kt} \bar{R}_{lt})}{\sqrt{m_i^{(5)} m_j^{(5)} m_k^{(5)} m_l^{(5)}}}
\end{align*}
\]

where \(\bar{R}_{it}\) denotes the centred return of asset \(i\) at time \(t\). \(m_i^{(n)}\) is the \(n\)-th centred sample moment of asset \(i\) given by:

\[
m_i^{(n)} = \frac{1}{T} \sum_{t=1}^{T} (\bar{R}_{it})^n.
\]
According to the concept of constant correlation, keeping the 7 correlation coefficients constant across all assets allows one to estimate all elements in $M_2$, $M_3$, and $M_4$ by using sample estimates for the first, second, third, fourth and sixth moments for all asset results without using further information on sample comoments:

$$
\begin{align*}
\hat{S}_{ij} &= \rho^{(2)} \sqrt{m_i^{(4)} m_j^{(2)}} \\
\hat{S}_{ik} &= \rho^{(4)} \sqrt{m_i^{(2)} r \sqrt{m_i^{(4)} m_j^{(2)}}} \\
\hat{k_{ij}} &= \rho^{(5)} m_i^{(4)} m_j^{(2)} \\
\hat{k_{ik}} &= \rho^{(5)} m_i^{(4)} m_j^{(2)} m_k^{(2)} \\
\hat{k_{ijk}} &= \rho^{(7)} \sqrt{m_i^{(4)} m_j^{(2)} m_k^{(2)}}
\end{align*}
$$

This approach allows us to substantially cut down the number of parameters compared to the sample estimators (see, Table II).

<table>
<thead>
<tr>
<th>N</th>
<th>$M_2$</th>
<th>$M_3$</th>
<th>$M_4$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>12</td>
<td>13</td>
<td>19</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>33</td>
<td>34</td>
<td>47</td>
</tr>
<tr>
<td>25</td>
<td>26</td>
<td>78</td>
<td>79</td>
<td>107</td>
</tr>
<tr>
<td>100</td>
<td>101</td>
<td>303</td>
<td>304</td>
<td>407</td>
</tr>
</tbody>
</table>

The numbers correspond to the amount of required parameters for each tensor. $N$ denotes the number of assets in the portfolio.

In the case of a 25-stock portfolio, the total number of parameters drops from 23,725 to 107. Assuming 5 years of monthly data, the ratio of 1,500 observations to 107 unknowns is statistically reasonable, certainly much more so than the severely under-identified system obtained when applying the sample estimator.

B. Single-Factor Estimator

The second improved estimator of the variance-covariance matrix goes back to Sharpe (1963), who introduces a single factor linear model for $n$ individual asset returns:

$$
R_t = C + \beta F_t + \epsilon_t 
$$

where $\epsilon_i$ is the idiosyncratic error term of asset $i$, and the factor $F$ is taken to be a broad-based index. The regression residuals are assumed to be homoscedastic and cross-sectionally uncorrelated:

$$
\epsilon : (0, \Psi) 
$$

According to the assumptions, all off-diagonal elements in $\Psi$ are zero. The idiosyncratic risk of the assets (residual variances) are reflected by the elements on the diagonal of $\Psi$. Then, the variance-covariance matrix of the asset universe can be written as:

$$
M_2 = \beta \beta' \mu_0^{(2)} + \Psi
$$

where $\beta$ indicates the $N \times 1$ vector of the regression coefficients in (18) and $\mu_0^{(2)}$ the second centred moment, that is the variance, of the single factor. Next, we derive the corresponding single-factor estimators for the moment tensors $M_3$ and $M_4$. We substitute (18) in (7) to obtain:

---

2 - Note that one could also introduce in a similar manner a multi-factor version of these improved estimators. This would involve, however, an increase in dimensionality that does not necessarily lead to better out-of-sample results (see Chan et al 1999), as well as (Connor and Korajczyk 1993), for evidence that after the first factor the marginal explanatory power of additional factors is relatively low.)
\[ M_1 = E \left[ (\beta F + \epsilon)(\beta F + \epsilon) \otimes (\beta F + \epsilon) \right] \]  
\[ M_2 = E \left[ (\beta F + \epsilon)(\beta F + \epsilon) \otimes (\beta F + \epsilon) \otimes (\beta F + \epsilon) \right] \]

where \( F \) is the \( T \times 1 \) vector of centred market returns \( F = F - \mu \). Out-multiplying these quantities leads, as for the covariance estimator, to two kinds of components in the estimators: one that is directly linked to the single factor and one that reflects the assumptions of the imposed model structure:

\[ M_3 = (\beta \beta \otimes \beta) \mu_0^{(2)} + \Psi \]
\[ M_4 = (\beta \beta \otimes \beta \otimes \beta) \mu_0^{(3)} + \Phi \]
\[ M_5 = (\beta \beta \otimes \beta \otimes \beta \otimes \beta) \mu_0^{(4)} + \Upsilon. \]

In order to explicitly assess the values in \( \Psi \), \( \Phi \) and \( \Upsilon \), and consistent with the spirit of the factor model approach, we assume that all cross-sectional residuals \( \epsilon_i \) and \( \epsilon_j \) \((i \neq j)\) are independent.\(^3\) As a result, we obtain:

\[ E(\epsilon_i \epsilon_j) = E(\epsilon_i)E(\epsilon_j) = 0 \]
\[ E(\epsilon_i \epsilon_j \epsilon_k) = E(\epsilon_i)E(\epsilon_j)E(\epsilon_k) = 0 \]
\[ E(\epsilon_i \epsilon_j \epsilon_k \epsilon_l) = E(\epsilon_i)E(\epsilon_j)E(\epsilon_k)E(\epsilon_l) = 0 \]

\[ E(\epsilon_i^2 \epsilon_j) = E(\epsilon_i^2)E(\epsilon_j) = 0 \]
\[ E(\epsilon_i^2 \epsilon_j \epsilon_k) = E(\epsilon_i^2)E(\epsilon_j)E(\epsilon_k) = 0 \]
\[ E(\epsilon_i^2 \epsilon_j \epsilon_k \epsilon_l) = E(\epsilon_i^2)E(\epsilon_j)E(\epsilon_k)E(\epsilon_l) = 0 \]

The relations in (26), together with the assumption of cross-sectionally independent residuals (24), allow us to specify the elements in \( \Psi \), \( \Phi \) and \( \Upsilon \). We first note that all off-diagonal elements of the \( N \times N \) matrix \( \Psi \) are zero, as for the original model of the covariance matrix estimator. The diagonal elements are defined as the estimated variances of the residuals:

\[ \Psi_{ii} = E(\epsilon_i^2) \]

with a sample estimate defined as

\[ \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t^2 \]

\[ \Psi_{ij} = 0 \quad \forall i \neq j. \] (27)

\(^3\) Note that for non-Gaussian variables the assumption of independence is stronger than absence of correlation.
The structure of the $N \times N^2$ matrix $\Phi$ is similar. All comoment elements cancel out and we obtain:

$$\phi_{ij} = \mathbb{E}(e_i^3), \text{ with a sample estimate defined as } \frac{1}{T} \sum_{t=1}^{T} e_t^3$$

$$\phi_{ij} = 0$$

$$\phi_{ijk} = 0 \quad \forall i \neq j \neq k$$

Obtaining the entries of the $N \times N^3$ matrix $\Upsilon$ is less straightforward, as indicated by the last equation in (26) and (24). Taking relevant symmetry effects into account, we obtain the inputs of $\Upsilon$ as:

$$\upsilon_{ij} = \mathbb{E}(e_i^3), \text{ with a sample estimate defined as } \frac{1}{T} \sum_{t=1}^{T} e_t^3$$

$$\upsilon_{ij} = 3 \beta_i \beta_j \mu_i \psi_{ii}$$

$$\upsilon_{ij} = \beta_i \beta_j \mu_{ij}^2 \psi_{jj} + \beta_j \beta_j \mu_{ii} \psi_{jj} + \psi_{ij} \psi_{jj}$$

$$\upsilon_{ijk} = \beta_{ij} \beta_{jk} \psi_{ii} \psi_{kk}$$

$$\upsilon_{ijk} = 0 \quad \forall i \neq j \neq k \neq l$$

We report the number of parameters that enter the estimation process of the higher moment tensors according to the single factor approach in Table III. The reduction in dimensionality, again, comes at the cost of specification risk, related to the imposed structure that all cross-sectional residuals are independent, which may not be valid if all commonality in asset returns has not been extracted by the factor model.

Table III. Required number of parameters - Single-factor estimator
The numbers correspond to the amount of required parameters for each tensor. $N$ denotes the number of assets in the portfolio.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$M_2$</th>
<th>$M_3$</th>
<th>$M_4$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>7</td>
<td>7</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>21</td>
<td>21</td>
<td>32</td>
<td>43</td>
</tr>
<tr>
<td>25</td>
<td>51</td>
<td>51</td>
<td>77</td>
<td>103</td>
</tr>
<tr>
<td>100</td>
<td>201</td>
<td>201</td>
<td>302</td>
<td>403</td>
</tr>
</tbody>
</table>

C. Bayesian Shrinkage of Higher-Order Moment Estimators
In the previous sub-sections, we introduce improved estimators for the higher-order moment tensors in an attempt to reduce the required number of parameter estimates compared to the sample estimator. While these estimators obviously involve lower estimation risk due to the imposed structure, they also involve the introduction of model risk (constant correlation across assets, and sufficient explanatory power stemming from a single factor, respectively). In an attempt to find the optimal trade-off between sample and specification risks, Ledoit and Wolf (2003) introduce in the context of the covariance matrix the asymptotically optimal linear combination of the sample estimator and a structured estimator (constant correlation or single factor), with the weight assigned to the latter being known as optimal shrinkage intensity. In this section, we extend this approach to higher-order moment tensor matrices. In the context of the covariance matrix, Ledoit and Wolf (2003) define the posterior misspecification function $L$ of such a Bayesian estimator as:

$$L(\alpha) = \| \alpha \Lambda + (1 - \alpha) S - \Omega \|_F$$

with $\Omega$ the true (unobservable) covariance matrix, $S$ the sample estimator, $\Lambda$ the shrinkage target, that is, a structured estimator for $\Omega$ (constant correlation or single-factor estimator). Additionally, $\| \|_F$ denotes the Frobenius norm of a matrix, defined as the square root of the sum of its squared entries. Minimising the expected value of this function yields the optimal linear combination of the two estimators as parameterised by the shrinkage intensity $\alpha$. Ledoit and Wolf (2003) derive an asymptotic estimator for this quantity as:
\[
\alpha^* = \frac{1}{T} \frac{\pi - \rho}{\gamma}
\]  
(31)

with
\[
\pi = \sum_{i=1}^{N} \sum_{j=1}^{K} \pi_{ij}, \quad \rho = \sum_{i=1}^{N} \sum_{j=1}^{K} \rho_{ij}, \quad \gamma = \sum_{i=1}^{N} \sum_{j=1}^{K} \gamma_{ij}
\]

where \( N \) denotes the number of rows and \( K \) the number of columns of \( \Omega \) respectively. Ledoit and Wolf (2003) further show that the \( \pi_{ij} \) represent the asymptotic variances of the sample estimator which can consistently be estimated by:
\[
\hat{\pi}_{ij} = \frac{1}{T} \sum_{t=1}^{T} \left[ s_{it} - s_{jt} \right]^2.
\]  
(32)

where \( s_{ij} = \frac{1}{T} \sum_{t=1}^{T} s_{it} \). Similarly, the \( \rho_{ij} \) represent the asymptotic covariances between the sample and the structured estimator and Ledoit and Wolf (2003, 2004) derive explicit formulas for consistent estimators of \( \rho_{ij} (\hat{\rho}_{ij}) \) when \( \Omega \) is the covariance matrix and \( \Lambda \) the single factor or the constant correlation estimator respectively (see below). Finally, the \( \gamma_{ij} \) denote the squared errors of structural estimator, with a consistent estimator given in Ledoit and Wolf (2003, 2004).

Since the Frobenius norm is not restricted to applications on squared matrices, we can directly derive consistent estimators for the asymptotic variances of the sample estimator (\( \pi \)) and the misspecification error of the structured estimator (\( \gamma \)), which are given by:

\[
\hat{\pi}_{ij} = \frac{1}{T} \sum_{t=1}^{T} \left[ (R_{it} - m_i)(R_{jt} - m_j) - M_2^{ij} \right]^2.
\]

\[
\hat{\pi}_{ik} = \frac{1}{T} \sum_{t=1}^{T} \left[ (R_{it} - m_i)(R_{kt} - m_k) - M_3^{ik} \right]^2.
\]

\[
\hat{\pi}_{ikl} = \frac{1}{T} \sum_{t=1}^{T} \left[ (R_{it} - m_i)(R_{kt} - m_k)(R_{lt} - m_l) - M_4^{ikl} \right]^2
\]  
(33)

and

\[
\hat{\gamma}_{ij} = \left( \lambda_{ij} - M_2^{ij} \right)^2.
\]

\[
\hat{\gamma}_{ik} = \left( \lambda_{ik} - M_3^{ik} \right)^2.
\]

\[
\hat{\gamma}_{ikl} = \left( \lambda_{ikl} - M_4^{ikl} \right)^2
\]  
(34)

where \( m \) denotes the sample mean of asset \( i \) and \( M_2^{ij} \) are the sample estimates of the corresponding tensor entries. \( \lambda_{ij} \) are structured estimates of the latter either using the constant correlation or the single factor approach.

The critical aspect is then to find consistent estimators for the asymptotic correlations between the structured and sample estimators (\( \rho \)). For this purpose, we use the so-called Delta-method, one of the corollaries of the central limit theorem (Green 2003):

**Delta-Method.** Let \( \hat{\theta} \) be an unbiased estimator for the unknown parameter \( N \times 1 \) vector \( \theta \) with finite asymptotic variances. Then, the central limit theorem states that
\[
\sqrt{T} (\hat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma)
\]  
(35)
where $\overset{d}{\rightarrow}$ indicates convergence in distribution and $\Sigma$ denotes the matrix of asymptotic variances and covariances of $\hat{\theta}$.

Let further $g$ be a real function with $g : \mathbb{R}^N \rightarrow \mathbb{R}^L$ and $\nabla_g$ its Jacobian matrix with respect to $\theta$. Then,

$$\sqrt{T} \{g(\hat{\theta}) - g(\theta)\} \overset{d}{\rightarrow} \mathcal{N}(0, \nabla^T \Sigma \nabla) \quad (36)$$

In the appendix, we show how to use this result and define suitable functional transformations $g$ and parameter vectors $\theta$ in order to obtain explicit expressions for the asymptotic covariances ($\rho$) of the estimators. Relation (31) enables us then to define asymptotically optimal linear combinations of the constant correlation (respectively, single factor) estimate and the sample estimate for $M_2$, as in Ledoit and Wolf (2003), but also for $M_3$ and $M_4$. Consequently, we obtain 6 shrinkage estimators, that is, for each higher-order moment tensor matrix ($M_2$, $M_3$, and $M_4$) one estimator shrunk towards the constant correlation estimate and one shrunk towards the single factor estimate. We now turn to an empirical analysis of the out-of-sample performance of the competing estimators that have been introduced in the current and previous sections.

### III. Empirical Results

Our empirical horse-race methodology is mostly inspired by Chan et al. (1999) and Jagannathan and Ma (2003), whose work we extend in several significant dimensions.

#### A. Methodology

Consistent with Chan et al. (1999) and Jagannathan and Ma (2003), we consider in our empirical analysis monthly return on common stocks listed on the NYSE and AMEX markets, with a sample period ranging from May 1973 through April 2006, a period over which we obtain from the CRSP data base valid return series for 403 stocks. While the introduction of higher moments into portfolio allocation models is often justified by the choice of a particular kind of underlying assets, such as emerging market assets (Jondeau and Rockinger 2006) or hedge funds (Agarwal and Naik 2004), it should be emphasised, however, that deviations from normality do not occur only in exotic markets, and we wish to address in this empirical analysis more common portfolio allocation practices. Table VI shows strong evidence of deviations from normality in the US equity universe.

#### Table VI. Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>5%</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.055</td>
<td>0.038</td>
<td>0.091</td>
<td>0.120</td>
<td>0.145</td>
<td>0.189</td>
<td>0.267</td>
</tr>
<tr>
<td>Std dev.</td>
<td>0.120</td>
<td>0.185</td>
<td>0.253</td>
<td>0.304</td>
<td>0.381</td>
<td>0.503</td>
<td>0.673</td>
</tr>
<tr>
<td>Skewness</td>
<td>-1.688</td>
<td>-0.224</td>
<td>0.109</td>
<td>0.337</td>
<td>0.622</td>
<td>1.334</td>
<td>4.240</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.124</td>
<td>3.656</td>
<td>4.343</td>
<td>5.115</td>
<td>6.711</td>
<td>11.939</td>
<td>60.777</td>
</tr>
</tbody>
</table>

Summary statistics are calculated for 403 assets listed on NYSE or AMEX with valid data from May 1973 through April 2006, for a total of 396 monthly returns.

According to the normality test proposed by Jarque and Bera (1980), we reject normality at the 5% confidence level for 93% of the stocks in our sample. We have also used other tests of non-normality, including the test proposed by Lilliefors (1967) and by Shapiro and Francia (1972), tests that lead to a rejection of the normality assumption for 86% and 77% of the stocks in our sample, (again at the 5% level).

Our empirical test protocol resembles that used by Chan et al. (1999) or Jagannathan and Ma (2003) in the sense that the optimised portfolio is held through the subsequent 12 months until a new allocation decision takes place. Consequently, we obtain 28 years of monthly out-of-sample data. At the end of April of each year, the higher-order moment tensor matrices $M_3$, $M_4$, and $M_4$ are estimated based on the prior 60 months for each estimator involved in the horse race. On the other hand, we deviate from Chan et al. or Jagannathan and Ma in the following two dimensions. First, instead of considering a single out-of-sample portfolio based on a randomly chosen basket of stocks, we propose to construct 100 randomly chosen baskets of $N = 25$ assets each, and report...
the distribution of the various statistical indicators across the 100 portfolios. Secondly, for each portfolio, we hold constant the universe of stocks as opposed to re-sampling every year. We argue that this methodology has some important advantages. For one, allowing multiple asset universes allows us to alleviate concerns over results, being driven by a specific outcome of the asset selection process at a given point in time. For another, holding constant the sub-universe of stocks for each portfolio across time allows us to measure the stability of each of the constructed portfolios, a key indicator of the forecast ability of the underlying estimators, which is not possible when the menu of assets changes through time. It can actually been argued that portfolio turnover might even be a more insightful indicator than reported out-of-sample performance. In fact, in our analysis as in Chan et al. (1999) or Jagannathan and Ma (2003), the portfolios are held in a buy-and-hold strategy for one year until the next allocation decision. The momentum effect induced by the buy-and-hold allocation scheme leads to converging allocations in portfolios within each yearly period, even when initial optimal allocations differ significantly across competing estimators. The turnover rate, on the other hand, is a direct indicator of the quality of past allocation decisions in view of ex-post return realisations. Note, finally, that obtaining a low turnover portfolio is a desirable property from an implementation standpoint, in addition to signalling enhanced robustness in parameter estimates.

We find the optimal allocation by assuming CARA preferences in the objective function (5), which breaks down to the expression 39, where we have further assumed that a budget and short-selling constraints hold:

\[
\min \left[ \frac{\lambda}{2} w^T M \omega - \frac{\lambda}{6} w^T (M \otimes \omega) + \frac{\lambda}{24} w^T (M \otimes \omega \otimes \omega) \right] \\
\text{s.t.:} \quad w^T I_n = 1 \\
\omega_i \geq 0
\] 

(37)

Note that we have neutralised in this expression the impact of the expected return parameter, and set \( \mu = 0 \) for all assets. Indeed, it has long been recognised that the sample mean is a poor estimator for the true population mean (Jorion 1986), with the consequence that in a mean-variance portfolio choice, for instance, global minimum variance portfolios often achieve higher out-of-sample Sharpe ratios than other tangency portfolios (Brandt 2005). Hence, like Chan et al. or Jagannathan and Ma, who focus on minimum variance portfolio, we generate global minimum risk portfolios, here represented by a 3-dimensional risk space. The absolute risk aversion coefficient \( \lambda \), which defines the risk-trade-off between variance, skewness and kurtosis, is taken equal to 10 in our base case, and we also try \( \lambda = 5 \) and \( \lambda = 15 \) in subsequent comparative static analysis.

In order to analyse the out-of-sample performance of the constructed portfolios, we define the certainty equivalent return \( (CER) \) as the deterministic return that makes the investor indifferent with respect to an investment in the risky portfolio (Cass and Stiglitz 1972, Frost and Savarino 1986 and Brandt 2005). This condition can be written as:

\[
-\exp(-\lambda (CER_i)) = \frac{1}{T} \sum_{t=1}^{T} -\exp(-\lambda R^t_i) 
\]

(38)

so that the certainty equivalent return for a given portfolio \( i \) is given as:

\[
CER_i = -\frac{1}{\lambda} \log \left( \frac{1}{T} \sum_{t=1}^{T} \exp(-\lambda R^t_i) \right)
\]

(39)

with \( t \) corresponding to out-of-sample dates as of May 1978 through April 2006 for a total of 336 monthly return observations. The quantity of interest for our purpose is the difference in certainty equivalent returns for competing estimators (see Frost and Savarino 1986), and also the difference with respect to the sample estimator portfolios in particular. Accordingly, we define this difference as monetary unit gains (MUG):

---

5 - We also report results with \( N = 0 \).
6 - The situation would have been different if a fixed-mix analysis had been implemented.
\( \text{MUG}_i = \text{CER}_i - \text{CER}_0 \), where the lower index 0 indicates the sample estimator and lower index i one of the improved estimators. Obviously, this is equivalent to defining \( \text{MUG} \) by the relationship:

\[
\frac{1}{T} \sum_{t=1}^{T} - \exp(-\lambda R'_{it}) = \frac{1}{T} \sum_{t=1}^{T} - \exp(-\lambda (R'^{0}_{it} + \text{MUG}_i))
\] (40)

So as to assess whether the reported out-performances in terms of differences in mean, standard deviation, third and fourth central moment and certainty equivalent returns are significant across estimators, we use a two-sample permutation test (see Good 2005). This non-parametric test based on resampling analysis is very suitable for our purpose as it allows us to avoid distributional assumptions about realised portfolio return moments and certainty equivalent returns.

We now turn to the analysis of the empirical results.

B. Base Case Results

The estimators involved in the horse race are the sample estimator, the constant correlation estimator, the single factor estimator and the two corresponding shrinkage estimators, namely the asymptotically optimal linear combination of the constant correlation estimator and the sample estimator (labeled shrinkage towards CC), and the asymptotically optimal linear combination of the single factor estimator and the sample estimator (labeled shrinkage towards SF), using expressions for the shrinkage intensity derived in section II.C. Consistent with Jagannathan and Ma, the single factor is represented by the CRSP value-weighted index of NYSE and AMEX listed assets.

Table V. Optimal shrinkage intensities

<table>
<thead>
<tr>
<th></th>
<th>Constant Correlation</th>
<th>Single Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_2 )</td>
<td>77.38 (15.27)</td>
<td>68.14 (12.95)</td>
</tr>
<tr>
<td>( M_1 )</td>
<td>98.83 (6.88)</td>
<td>97.61 (7.00)</td>
</tr>
<tr>
<td>( M_0 )</td>
<td>18.24 (22.35)</td>
<td>2.40 (9.78)</td>
</tr>
</tbody>
</table>

Optimal shrinkage intensities for each moment tensor have been calculated. 28 time windows and 100 sub-universes for a total of 2800 observations have been considered. Average values and, in parentheses, standard deviations are reported.

Table V reports average values of the shrinkage intensities and their standard deviations across time and across the 100 baskets of stocks. For the second-moment case (the covariance matrix), we obtain very stable optimal shrinkage intensities that are consistent with the findings in Ledoit and Wolf (2003), who find shrinkage intensities around 80 percent on average for the single factor estimator. The optimal linear combinations for the coskewness tensor matrix exhibit an even more pronounced tilt towards the structured estimators, as evidenced by the very high shrinkage intensities, while shrinkage intensities corresponding to the 4th moment tensors are much lower on average and less stable. Overall, shrinkage intensities appear to be lower for the single-factor approach than for the constant correlation approach. Table VI summarises the distribution of the monetary utility gains across the 100 portfolios.

Table VI. Optimal shrinkage intensities (%)

<table>
<thead>
<tr>
<th></th>
<th>Avg</th>
<th>Prob</th>
<th>Min</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const. Correlation</td>
<td>0.37</td>
<td>61</td>
<td>-2.55</td>
<td>-1.67</td>
<td>-0.53</td>
<td>0.26</td>
<td>1.29</td>
<td>2.89</td>
<td>3.37</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.49</td>
<td>71</td>
<td>-1.83</td>
<td>-0.99</td>
<td>-0.07</td>
<td>0.38</td>
<td>1.01</td>
<td>2.19</td>
<td>2.65</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.55</td>
<td>70</td>
<td>-1.49</td>
<td>-0.95</td>
<td>-0.26</td>
<td>0.52</td>
<td>1.27</td>
<td>2.53</td>
<td>3.14</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.58</td>
<td>86</td>
<td>-1.10</td>
<td>-0.45</td>
<td>0.14</td>
<td>0.53</td>
<td>0.99</td>
<td>1.80</td>
<td>2.51</td>
</tr>
</tbody>
</table>

The monetary utility gain is defined as the supplementary amount of return that is required by a CARA investor to achieve the same level of utility with portfolios based upon the sample estimator as compared to portfolios based upon the corresponding competing estimator. Average values and percentiles of its distribution are reported. Additionally, the probability that the corresponding estimator generates a higher out-of-sample utility than the sample estimator, that is, a positive monetary utility gain, is reported (Prob). All numbers are annualised and in percentages.
The constant correlation (single factor) estimator out-performs the sample estimator for 61 (71) of the 100 portfolios. Using the corresponding shrinkage estimators allows a significant increase in this out-performance probability, which reaches 70% for the constant correlation estimator and 86%, for the single-factor estimator. Similar results can be observed in terms of median and average monetary utility gains. The single factor approach and the corresponding linear combination (shrinkage towards single factor) perform better than their constant correlation counterparts in terms of extreme observations (minimum and maximum values) as well as out-performance probability with respect to the sample estimator and median monetary utility gains. Additionally, the distributions display an asymmetric pattern, as evidenced by average values exceeding median monetary utility gains. In other words, not only do improved estimators out-perform the sample estimators on average and in probability but also in magnitude: in cases where the sample estimator outperforms another estimator, this out-performance is less pronounced.

To shed light on this utility out-performance, table IIIV reports average values for mean, standard deviation, third and fourth central moments of the out-of-sample portfolios as well as out-performance probabilities (with high values for mean and third central moment and low values for standard deviation and fourth central moment considered desirable).

Table VII. Average out-of-sample portfolio moments

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. dev.</th>
<th>3rd moment</th>
<th>4th moment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>13.58</td>
<td>12.51</td>
<td>-1.67</td>
<td>1.02</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>13.63</td>
<td>(53)</td>
<td>12.41</td>
<td>(58)</td>
</tr>
<tr>
<td>Single Factor</td>
<td>13.77</td>
<td>(60)</td>
<td>12.38</td>
<td>(68)</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>13.68</td>
<td>(56)</td>
<td>12.25</td>
<td>(82)</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>13.82</td>
<td>(64)</td>
<td>12.26</td>
<td>(88)</td>
</tr>
</tbody>
</table>

Realised returns from May 1978 through April 2006 are observed for each portfolio. Mean returns, standard deviations, 3rd and 4th central moments are calculated for each portfolio. Average annualised values are reported. Means and standard deviations are in percentages, higher central moments in 1e-04. The numbers in parentheses correspond to out-performance probabilities (in percentages) with respect to the sample estimator where high values for mean and 3rd central moment and low values for standard deviation and 4th central moment are considered desirable.

Sample estimators are out-performed by all competing estimators across the 100 portfolios in terms of average mean return, average volatility, and average third and fourth central moments of the returns. The shrinkage estimators perform better than their structured counterparts (those with shrinkage intensity equal to 100%, as opposed to the optimal value) in terms of average mean return, average volatility, and average fourth central moment. It is only in terms of average third central moments that the structured estimators perform better than the corresponding shrinkage estimators.

Table VIII. Significance tests

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. dev.</th>
<th>3rd moment</th>
<th>4th moment</th>
<th>CER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const. Correlation</td>
<td>0.424</td>
<td>0.311</td>
<td>0.000</td>
<td>0.077</td>
<td>0.063</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.207</td>
<td>0.246</td>
<td>0.002</td>
<td>0.026</td>
<td>0.022</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.339</td>
<td>0.078</td>
<td>0.001</td>
<td>0.031</td>
<td>0.011</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.145</td>
<td>0.092</td>
<td>0.127</td>
<td>0.020</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Permutation tests are run in order to assess the significance of differences in means, standard deviations, 3rd and 4th central moment and certainty equivalent returns (CER) as compared to the sample estimator. P-values, based on the resampling distribution are given (see above for test methodology).

Table VIII analyses whether the reported out-performance in terms of mean, standard deviation, third and fourth central moments, and certainty equivalent returns are significant. The p-values suggest that the null-hypothesis that average mean returns are identical cannot be rejected for reasonable confidence levels. Nor, more surprisingly, are standard deviations. On the other hand, higher-order portfolio return moments improve significantly when applying a structural or shrinkage estimator. This suggests that improved estimates of higher-moments might generate greater benefits than improved estimates of the variance-covariance matrix. Besides, the single-factor approach seems to perform slightly better than the constant correlation approach, which is consistent with the results in table VI. Table VIII suggests that this is due mostly to better performance in terms of even moments of

---

7 Note that the shrinkage estimator for the third central moment is very much shrunk towards the sample estimator (especially in the case of the single-factor approach), because of high values for the shrinkage intensity factor (see Table VI).
portfolio returns. Overall, there is convincing evidence that certainty equivalent returns are significantly positive, as indicated by low p-values (0.01 to 0.06).

We also compare these results to the naive diversification obtained by investing in an equally-weighted portfolio. In only 5% of the cases, does the equally-weighted portfolio generate higher monetary utility gains than the portfolio based on structured estimates (constant correlation and factor-based), a figure that drops to 3% for both shrinkage estimators, with an average monetary utility loss as high as 3.49. Analysing the out-of-sample moments of the equally weighted portfolios shows that the naive diversification leads in fact to higher mean returns (16.17% on average) but to consistently worse standard deviations (15.68% on average), lower central moments (-8.52e-04 on average) and central moments (3.15e-04 on average), than does to portfolio optimisation based on improved estimators (see table VII). Our results regarding the first two moments are consistent with those of DeMiguel, Garlappi, and Uppal (2007), who find that none of the methods based on mean-variance portfolio optimisation with improved forecasts of the variance-covariance matrix perform consistently better than the equally-weighted portfolios in terms of Sharpe ratio or mean-variance certainty-equivalent return. On the other hand, they suggest that equally-weighted portfolios are dominated by more sophisticated portfolio construction techniques when higher moments are incorporated.

Ledoit and Wolf (2003) as well as Jagannathan and Ma relate the out-performance in terms of out-of-sample standard deviation to a shrinkage effect on the covariance matrix caused by the structure in the estimators (constant correlation and single factor respectively). Accordingly, we report the shrinkage intensity related to higher moment tensors in table IX.

Table IX. Distribution of co-moment estimates

<table>
<thead>
<tr>
<th></th>
<th>Sample</th>
<th>(2.40)</th>
<th>2.76</th>
<th>(1.70)</th>
<th>3.94</th>
<th>(3.49)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant Correlation</td>
<td>5.35</td>
<td>(2.34)</td>
<td>1.16</td>
<td>(1.25)</td>
<td>3.76</td>
<td>(3.18)</td>
</tr>
<tr>
<td>Single Factor</td>
<td>5.19</td>
<td>(2.22)</td>
<td>0.93</td>
<td>(0.98)</td>
<td>2.25</td>
<td>(1.75)</td>
</tr>
</tbody>
</table>

Frobenius norms are calculated for each moment tensor by taking the square root of the sum all squared elements of the corresponding matrix. The N entries that correspond to central moments are omitted. 28 time windows and 100 sub-universes, for a total of 2,800 matrices for each method are considered. Average values and, in parentheses, standard deviations are reported. All numbers are in percentages.

Not surprisingly, the Frobenius norm of the covariance matrix estimated with the sample estimator exceeds the norm estimated with structured estimators. This result is consistent with the findings in Ledoit and Wolf (2003) or Jagannathan and Ma. Interestingly, our analysis suggests that a similar shrinkage effect is present for higher-order co-moments matrices, with an effect significantly more pronounced for the third moment tensor. In all cases, the single-factor approach reveals the lowest variability. Overall, the high variability of sample covariance estimates, and the resulting unstable portfolio decisions, are driving force for imposing structure on the covariance matrix. Consequently, portfolios based on structured and shrinkage estimators ought to be more stable over time. Table X confirms this intuition.

Table X. Optimal portfolio weights - Turnover rate

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>5%</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>1.39</td>
<td>1.65</td>
<td>1.89</td>
<td>2.06</td>
<td>2.27</td>
<td>2.52</td>
<td>2.66</td>
</tr>
<tr>
<td>Constant Correlation (CC)</td>
<td>0.99</td>
<td>1.03</td>
<td>1.17</td>
<td>1.29</td>
<td>1.38</td>
<td>1.59</td>
<td>1.84</td>
</tr>
<tr>
<td>Single Factor (SF)</td>
<td>0.96</td>
<td>1.02</td>
<td>1.19</td>
<td>1.33</td>
<td>1.49</td>
<td>1.66</td>
<td>1.89</td>
</tr>
<tr>
<td>Shrinkage towards CC (SCC)</td>
<td>1.10</td>
<td>1.20</td>
<td>1.36</td>
<td>1.48</td>
<td>1.60</td>
<td>1.80</td>
<td>1.95</td>
</tr>
<tr>
<td>Shrinkage towards SF (SSF)</td>
<td>1.11</td>
<td>1.25</td>
<td>1.42</td>
<td>1.57</td>
<td>1.72</td>
<td>1.85</td>
<td>1.90</td>
</tr>
</tbody>
</table>

Mean absolute deviations (MAD) across time of the target allocations are computed for each asset in a portfolio. The table reports the distribution of the average MAD of all assets in the portfolio across the 100 portfolios.

The table reports the average turnover rate, which for portfolio \( i \) (\( TR^i \)) is defined as the average value of the mean absolute deviations of the assets in the portfolio:

\[
TR^i = \frac{1}{N(T-1)} \sum_{n=1}^{N} \sum_{t=2}^{T} |w^i_{n,t} - w^i_{n,t-1}| \tag{41}
\]
The results are striking. Portfolios constructed upon the constant correlation and single-factor estimators yield median turnover rate reductions of 40% and 35% respectively. For shrinkage estimators the improvement in turnover is less pronounced but still significant (30% and 20%, respectively). Overall, these results suggest that enhanced out-of-sample portfolio performance is achieved by the use of improved parameter estimates.

C. Additional Results

C.1. Changes in Parameter Values

In order to assess whether the aforementioned results are robust with respect to the chosen parameters, we conduct several additional optimisations and report the results in this sub-section. We consider in table XI the distribution of monetary utility gains for portfolios of smaller size $N=10^9$.

<table>
<thead>
<tr>
<th>Table XI. Monetary utility gains - $N=10$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Const. Correlation</strong></td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Const. Correlation</td>
</tr>
<tr>
<td>Single Factor</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
</tr>
</tbody>
</table>

The monetary utility gain (MUG) is the additional return required by a CARA investor to achieve the same level of utility when using portfolios based on the sample estimator instead of those based on the corresponding competing estimator. Average MUGs and percentiles of their distributions are reported. Additionally, the probability that the corresponding estimator generates a higher out-of-sample utility than the sample estimator, that is, a positive monetary utility gain, is reported (Prob). All numbers are annualised and in percentages.

The results are consistent with the results corresponding to portfolios of size $N=25$, with a slightly higher out-performance probability across all estimators in absolute terms.

<table>
<thead>
<tr>
<th>Table XII. Significance tests</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Const. Correlation</td>
</tr>
<tr>
<td>Single Factor</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
</tr>
</tbody>
</table>

Permutation tests are run in order to assess the significance of differences in means, standard deviations, 3rd and 4th central moment and certainty equivalent returns (CER) as compared to the sample estimator. P-values, based on the resampling distribution are given (see above for test methodology).

Interestingly, the reported out-performance, while higher in absolute terms than in the base case ($N=25$), appears less significant. Obviously, this out-performance is mostly due to better performances in terms of average mean returns, which is the only moment displaying lower p-values than to the base case. This suggests that the benefits of improved estimators increase with the number of assets, a phenomenon arguably explained by the trade-off between model and estimation risk that is tilted towards the latter as the number of parameters increases.

<table>
<thead>
<tr>
<th>Table XIII. Optimal shrinkage intensities - $N=10$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>M₁</td>
</tr>
<tr>
<td>M₂</td>
</tr>
<tr>
<td>M₃</td>
</tr>
</tbody>
</table>

Optimal shrinkage intensities for each moment tensor are calculated. 28 time windows and 100 sub-universes for a total of 2800 observations are considered. Average values and, in parentheses, standard deviations are reported.

---

We have also looked at portfolios of size $N=50$, and obtained qualitatively similar results, which are available upon request. Because of the increased computing time, we have, in this case, conducted the analysis for a lower number of sub-universes (10 instead of 100).
Table XIII shows the average shrinkage intensities. The numbers are consistent with those in the base case methodology \((N=25)\) but shrinkage intensities are less extreme for higher-order moment tensor matrices. The results suggest that, as the portfolios size grows, structured estimators are asymptotically more suitable to estimate the third moment tensor matrix than the sample estimator. The converse seems to hold for the fourth moment tensor as shrinkage intensities seem to decrease in portfolio size.

Table XIV. Optimal portfolio weights - Turnover rate - \(N=10\)

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>5%</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>2.46</td>
<td>2.70</td>
<td>3.18</td>
<td>3.54</td>
<td>3.91</td>
<td>4.62</td>
<td>4.92</td>
</tr>
<tr>
<td>Constant Correlation (CC)</td>
<td>1.71</td>
<td>2.03</td>
<td>2.29</td>
<td>2.56</td>
<td>2.85</td>
<td>3.32</td>
<td>4.04</td>
</tr>
<tr>
<td>Single Factor (SF)</td>
<td>1.34</td>
<td>1.87</td>
<td>2.31</td>
<td>2.60</td>
<td>2.99</td>
<td>3.69</td>
<td>3.79</td>
</tr>
<tr>
<td>Shrinkage towards CC (SCC)</td>
<td>2.10</td>
<td>2.21</td>
<td>2.52</td>
<td>2.79</td>
<td>3.11</td>
<td>3.51</td>
<td>4.23</td>
</tr>
<tr>
<td>Shrinkage towards SF (SSF)</td>
<td>2.02</td>
<td>2.23</td>
<td>2.65</td>
<td>2.94</td>
<td>3.21</td>
<td>3.87</td>
<td>4.20</td>
</tr>
</tbody>
</table>

Mean absolute deviations (MAD) across time of the target allocations are computed for each asset in a portfolio. The table reports the distribution of the average MAD of all assets in the portfolio across the 100 paths.

Finally, Table XIV shows the percentiles of the turnover rates across the 100 portfolios. Obviously, in absolute terms, numbers have increased compared to the base case \((N=25)\), which can be explained by the fact that the average allocation to an asset has increased. In relative terms, however, the turnover rate reduction is identical to the base case. We also test different risk aversion parameters, so as to allow for increased or decreased importance to higher moments in the objective function. The higher the risk aversion parameter \(\lambda\), the higher the impact of higher-order moments on investor choices.

Table XV. Monetary utility gains - \(\lambda = 5\)

<table>
<thead>
<tr>
<th></th>
<th>Avg</th>
<th>Prob</th>
<th>Min</th>
<th>5%</th>
<th>25%</th>
<th>Med</th>
<th>75%</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant Correlation</td>
<td>0.11</td>
<td>57.00</td>
<td>-2.55</td>
<td>-1.63</td>
<td>-0.68</td>
<td>0.24</td>
<td>0.91</td>
<td>1.76</td>
<td>2.72</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.24</td>
<td>61.00</td>
<td>-1.57</td>
<td>-0.91</td>
<td>-0.30</td>
<td>0.22</td>
<td>0.70</td>
<td>1.46</td>
<td>1.88</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.20</td>
<td>62.00</td>
<td>-1.79</td>
<td>-1.23</td>
<td>-0.47</td>
<td>0.26</td>
<td>0.81</td>
<td>1.51</td>
<td>2.47</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.30</td>
<td>70.00</td>
<td>-1.04</td>
<td>-0.51</td>
<td>-0.08</td>
<td>0.26</td>
<td>0.65</td>
<td>1.19</td>
<td>1.88</td>
</tr>
</tbody>
</table>

The monetary utility gain (MUG) is the additional return required by a CARA investor to achieve the same level of utility when using portfolios based on the sample estimator instead of those based on the corresponding competing estimator. Average MUGs and percentiles of their distributions are reported. Additionally, the probability that the corresponding estimator generates a higher out-of-sample utility than the sample estimator, that is, a positive monetary utility gain, is reported \((\text{Prob})\). All numbers are annualised and in percentages.

Table XVI. Significance tests

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. dev.</th>
<th>3rd moment</th>
<th>4th moment</th>
<th>CER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant Correlation</td>
<td>0.481</td>
<td>0.267</td>
<td>0.000</td>
<td>0.102</td>
<td>0.280</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.245</td>
<td>0.240</td>
<td>0.002</td>
<td>0.034</td>
<td>0.091</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.440</td>
<td>0.084</td>
<td>0.000</td>
<td>0.054</td>
<td>0.137</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.234</td>
<td>0.079</td>
<td>0.082</td>
<td>0.024</td>
<td>0.047</td>
</tr>
</tbody>
</table>

Permutation tests are run in order to assess the significance of differences in means, standard deviations, 3rd and 4th central moment and certainty equivalent returns (CER) as compared to the sample estimator. P-values, based on the resampling distribution, are given (see above for test methodology).

We observe a drop in the level of significance for improvements in certainty equivalent returns when compared to the base case \((\lambda =10)\). Intuitively, this is explained by the relative increase in the weight put on lower portfolio moments (mean and standard deviation) relative to that on. In other words, the use of improved estimators for higher moments is more critical for more-risk averse investors.

Table XVII. Optimal portfolio weights - Turnover rate - \(\lambda = 5\)

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>5%</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>1.41</td>
<td>1.67</td>
<td>1.88</td>
<td>2.06</td>
<td>2.25</td>
<td>2.50</td>
<td>2.66</td>
</tr>
<tr>
<td>Constant Correlation (CC)</td>
<td>0.99</td>
<td>1.01</td>
<td>1.15</td>
<td>1.26</td>
<td>1.35</td>
<td>1.54</td>
<td>1.76</td>
</tr>
<tr>
<td>Single Factor (SF)</td>
<td>0.95</td>
<td>1.01</td>
<td>1.19</td>
<td>1.32</td>
<td>1.47</td>
<td>1.63</td>
<td>1.85</td>
</tr>
<tr>
<td>Shrinkage towards CC (SCC)</td>
<td>1.09</td>
<td>1.16</td>
<td>1.31</td>
<td>1.44</td>
<td>1.57</td>
<td>1.75</td>
<td>1.89</td>
</tr>
<tr>
<td>Shrinkage towards SF (SSF)</td>
<td>1.09</td>
<td>1.23</td>
<td>1.39</td>
<td>1.54</td>
<td>1.68</td>
<td>1.80</td>
<td>1.87</td>
</tr>
</tbody>
</table>

Mean absolute deviations (MAD) across time of the target allocations are computed for each asset in a portfolio. The table reports the distribution of the average MAD of all assets in the portfolio across the 100 paths.
The monetary utility gain (MUG) is the additional return required by a CARA investor to achieve the same level of utility when using portfolios based on the sample estimator instead of those based on the corresponding competing estimator. Average MUGs and percentiles of their distributions are reported. Additionally, the probability that the corresponding estimator generates a higher out-of-sample utility than the sample estimator, that is, a positive monetary utility gain, is reported (Prob). All numbers are annualised and in percentages.

The results in Tables XVIII and XIX confirm the aforementioned observations. Indeed, the level and significance of the out-performance of improved estimators increases with increasing risk aversion.\textsuperscript{10}

### Table XIX. Significance tests

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. dev.</th>
<th>3rd moment</th>
<th>4th moment</th>
<th>CER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const. Correlation</td>
<td>0.408</td>
<td>0.334</td>
<td>0.000</td>
<td>0.053</td>
<td>0.016</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.200</td>
<td>0.248</td>
<td>0.001</td>
<td>0.020</td>
<td>0.010</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.282</td>
<td>0.076</td>
<td>0.003</td>
<td>0.016</td>
<td>0.005</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.112</td>
<td>0.096</td>
<td>0.143</td>
<td>0.019</td>
<td>0.011</td>
</tr>
</tbody>
</table>

Permutation tests are run in order to assess the significance of differences in means, standard deviations, 3rd and 4th central moments, and certainty equivalent returns (CER) as compared to the sample estimator. P-values, based on the resampling distribution are given (see above for test methodology).

Finally, we also consider a different objective, minimisation of the portfolio Value-at-Risk (VaR), a risk measure often used in practice in spite of its alleged inconsistency (Artzner et al 1999). When returns are normally distributed, the \((1-\alpha)\%\) VaR estimate is given by:

\[
VaR(1-\alpha) = -(z_\alpha \sigma + \mu).
\]

where \(z_\alpha\) denotes the \(\alpha\)-percentile of the standard normal density function, \(\sigma\) the return volatility and \(\mu\) the expected return. When the return distribution is unknown, one may use a Cornish-Fisher transformation as a suitable approximation of the true density (Cornish and Fisher 1938). Accordingly, the modified \(\alpha\)-percentile \((\bar{z}_\alpha)\) is defined as:

\[
\bar{z}_\alpha = z_\alpha + \frac{1}{6}(z_\alpha^3 - 1)S + \frac{1}{24}(z_\alpha^4 - 3z_\alpha^2)K - \frac{1}{36}(2z_\alpha^5 - 5z_\alpha^3)S^2
\]

where \(S\) denotes the standardised sample-skewness, \(K\) the standardised sample-excess-kurtosis. The modified Cornish-Fisher VaR is then estimated by:

\[
VaR_{\text{mod}}(1-\alpha) = -(\bar{z}_\alpha \sigma + \mu).
\]

Hence, minimising this modified version of the value-at-risk allows us to incorporate higher-order moments of asset returns in the optimal allocation decision.

\textsuperscript{10} Note that the comparison in terms of monetary utility gains is misleading since certainty equivalent returns are functions of risk aversion parameters.
Table XXI reports the out-of-sample value-at-risk distribution across the 100 portfolios for each of the estimators. The table confirms the results obtained in the base case. Portfolios constructed using the sample estimators display higher out-of-sample Value-at-Risk estimates than those built upon improved estimators. The out-performance probabilities are even more significant in this context than in the maximum utility framework.

Table XXI. Out-of-sample Value-at-Risk

<table>
<thead>
<tr>
<th></th>
<th>Avg</th>
<th>Prob</th>
<th>Min</th>
<th>5%</th>
<th>25%</th>
<th>Med</th>
<th>75%</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>5.50</td>
<td>3.68</td>
<td>4.16</td>
<td>4.76</td>
<td>5.59</td>
<td>6.11</td>
<td>7.05</td>
<td>7.68</td>
<td></td>
</tr>
<tr>
<td>Single Factor</td>
<td>5.10</td>
<td>82</td>
<td>3.21</td>
<td>3.75</td>
<td>4.32</td>
<td>5.16</td>
<td>5.76</td>
<td>6.83</td>
<td>7.87</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>4.81</td>
<td>96</td>
<td>3.12</td>
<td>3.47</td>
<td>4.17</td>
<td>4.91</td>
<td>5.37</td>
<td>5.98</td>
<td>6.40</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>5.06</td>
<td>87</td>
<td>3.30</td>
<td>3.66</td>
<td>4.25</td>
<td>5.11</td>
<td>5.72</td>
<td>6.69</td>
<td>7.78</td>
</tr>
</tbody>
</table>

Realised returns from May 1978 through April 2006 are observed for each portfolio. The first column shows the average value of the empirical monthly Cornish-Fisher Value-at-Risk (VaR) that is calculated for each portfolio. The second column gives the out-performance probability of the estimator with respect to the sample estimator where low VaR is considered desirable. The remaining columns report percentiles of the VaR-distribution across the 100 portfolios. All numbers are in percent.

Table XXII. Significance tests

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. dev.</th>
<th>3rd moment</th>
<th>4th moment</th>
<th>VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const. Correlation</td>
<td>0.923</td>
<td>0.000</td>
<td>0.170</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.925</td>
<td>0.000</td>
<td>0.830</td>
<td>0.003</td>
<td>0.006</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.871</td>
<td>0.000</td>
<td>0.276</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.977</td>
<td>0.000</td>
<td>0.714</td>
<td>0.006</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Permutation tests are run in order to assess the significance of differences in means, standard deviations, 3rd and 4th central moments and certainty equivalent returns (CER) as compared to the sample estimator. P-values, based on the resampling distribution, are given (see above for test methodology).

Table XXII, however, suggests that this relative performance is due mostly to the higher focus on the portfolio volatility that can be seen in equation (44).

Table XXIII. Optimal portfolio weights - Turnover rate - minVaR portfolios

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>5%</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>1.91</td>
<td>2.21</td>
<td>2.52</td>
<td>2.87</td>
<td>3.16</td>
<td>3.62</td>
<td>3.87</td>
</tr>
<tr>
<td>Constant Correlation (CC)</td>
<td>1.17</td>
<td>1.24</td>
<td>1.37</td>
<td>1.54</td>
<td>1.78</td>
<td>2.22</td>
<td>2.55</td>
</tr>
<tr>
<td>Single Factor (SF)</td>
<td>1.02</td>
<td>1.11</td>
<td>1.28</td>
<td>1.51</td>
<td>1.69</td>
<td>2.06</td>
<td>2.14</td>
</tr>
<tr>
<td>Shrinkage towards CC (SCC)</td>
<td>1.21</td>
<td>1.34</td>
<td>1.56</td>
<td>1.68</td>
<td>1.94</td>
<td>2.34</td>
<td>2.48</td>
</tr>
<tr>
<td>Shrinkage towards SF (SSF)</td>
<td>1.18</td>
<td>1.30</td>
<td>1.53</td>
<td>1.71</td>
<td>2.00</td>
<td>2.42</td>
<td>3.07</td>
</tr>
</tbody>
</table>

Mean absolute deviations (MAD) across time of the target allocations are computed for each asset in a portfolio. The table reports the distribution of the average MAD of all assets in the portfolio across the 100 paths.
IV. Conclusion
This paper introduces improved estimators for higher-order moments/co-moments of asset returns and discusses the implications in terms of optimal asset allocation decisions. In particular, we extend to coskewness and cokurtosis tensor matrices the constant correlation approach (Elton and Gruber 1973) and the single-factor approach (Sharpe 1963) originally introduced for the covariance matrix. We further extend the concept of optimal shrinkage intensities to the presence of higher-order moments; that is, we define asymptotically optimal linear combinations of structured and sample estimators.

Consistent with the methodology used in the related literature (Jagannathan and Ma or Ledoit and Wolf, this study performs an empirical horse-race analysis of the competing estimators, using rolling window optimisations in a buy-and-hold framework. Portfolios built upon improved estimators are found to outperform portfolios built upon sample estimators in terms of out-of-sample expected utility. The probability of a positive monetary utility gain, that is, higher expected utility, lies between 60% and 90% for all estimators and all parameter specifications (portfolio size, risk aversion parameter, objective function). Shrinkage estimators, defined as linear combinations of structured estimators (constant correlation and single-factor estimators) and sample estimators, perform best. Another striking result is a significant reduction of turnover rates when using improved estimators rather than to sample estimators. Indeed, on average, the mean absolute deviation of asset weights over time is reduced by 35%-40%, suggesting superior forecasting ability associated with improved estimators, most likely due to the shrinkage effects for higher-order moment matrices that are consistent with the findings of Ledoit and Wolf (2003) or (Jagannathan and Ma 2003 with respect to covariance matrices. In general, with respect to the sample estimator, the single-factor approach and the estimator shrunk towards the single-factor estimator outperform their constant correlation counterparts in terms of monetary utility gains and expected utility out-performance, whereas portfolios based upon the constant correlation estimator are slightly more stable. Overall, our results suggest that when an investor’s objective incorporates higher moments, improved estimators generate added-value in an out-of-sample context, a result that expands on previous similar findings about the covariance matrix.
Appendix

In this appendix, we derive explicit expressions for the optimal shrinkage intensity in the context of higher-order moment tensor matrices. We denote $\rho^{cc}_y$ the asymptotic covariance between the constant correlation estimate and the sample estimate of the covariance between $i$ and $j$ ($\sigma_{ij}$). In the same line, $\rho^{sf}_y$ denotes the asymptotic covariance between the single factor estimate and the sample estimate for $\sigma_{ij}$. The sample estimate of $\sigma_{ij}$ is denoted by ($s_{ij}$) and the sample estimate for $\mu^{(k)}$ by $m^{(k)}$.

Obviously, $\rho^{cc}_y$ and $\rho^{sf}_y$ are identical and equal to $\pi_{ij}$. Next, we have in a straightforward manner:

\[ \hat{\rho}^{cc}_y = \text{AsyCov}\left(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}\right) \]  \hfill (45)

\[ \hat{\rho}^{sf}_y = \text{AsyCov}\left(\sqrt{T}\beta_{ij}m^{(2)}, \sqrt{T}s_{ij}\right) = \text{AsyCov}\left(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}\right) \]  \hfill (46)

where $s_{ij}$ denotes the sample covariance of assets $i$ and $j$ and lower index $\theta$ indicates the single factor. In order to make use of corollary 1, we customise the vector $\theta$ and the function $g$ so as to decompose the asymptotic covariances in (45) and (46):

\[ \theta^{cc}_2 = (s_{ij}, s_{ij}, s_{ij})^\prime \]

\[ g^{cc}_2(\theta^{cc}_2) = \begin{pmatrix} \sigma_{ij}^{(2)} s_{ij} \\ \sigma_{ij}^{(2)} s_{ij} \\ s_{ij} \end{pmatrix} \]  \hfill (47)

and

\[ \theta^{sf}_2 = (s_{ij}, s_{ij}, m^{(2)}_{ij}, s_{ij})^\prime \]

\[ g^{sf}_2(\theta^{sf}_2) = \begin{pmatrix} \sigma_{ij}^{(2)} s_{ij} \\ m^{(2)}_{ij} \\ s_{ij} \end{pmatrix} \]  \hfill (48)

From (36) we have:

\[ \Sigma^{cc} = \begin{bmatrix} \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) \\ \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) \\ \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) \end{bmatrix} \]  \hfill (49)

and

\[ \Sigma^{sf} = \begin{bmatrix} \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) \\ \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) \\ \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) & \text{AsyCov}(\sqrt{T}s_{ij}, \sqrt{T}s_{ij}) \end{bmatrix} \]  \hfill (50)

- Consistently throughout this appendix, the index $cc$ refers to the constant correlation estimator and $sf$ to the single factor estimator respectively.
Computing the Jacobian of $g^{CC}$ and $g^{SF}$, (36) yields:

\[
\rho_{ij}^{CC} = \frac{e_{(i)}^{(j)}}{2} \left[ s_{ij} \text{AsyCov}(\sqrt{T}S_{ij}, \sqrt{T}S_{ij}) + \frac{s_{ij}}{\sqrt{m_{ij}}} \text{AsyCov}(\sqrt{T}S_{ij}, \sqrt{T}S_{ij}) \right] \\
\rho_{ij}^{SF} = \frac{s_{ij}^{(2)}}{m_{ij}^{(2)}} \text{AsyCov}(\sqrt{T}S_{ij}, \sqrt{T}S_{ij}) + \frac{s_{ij}^{(0)}}{m_{ij}^{(2)}} \text{AsyCov}(\sqrt{T}S_{ij}, \sqrt{T}S_{ij})
\]

where $\rho_{ij}^{CC}$ and $\rho_{ij}^{SF}$ are the off-diagonal elements of the 2x2 matrices $V_{g}^{\Sigma^{CC}}V_{g}$ and $V_{g}^{\Sigma^{SF}}V_{g}$ respectively. Referring to Ledoit and Wolf (2003), we introduce consistent estimators for the asymptotic covariances in (51) and (52):

\[
\text{AsyCov}(\sqrt{T}S_{ij}, \sqrt{T}S_{ij}) = \\
\frac{1}{T} \sum_{t=1}^{T} [(R_{ij} - m_{ij})^{2} - s_{ij}] [(R_{ij} - m_{ij})(R_{ij} - m_{ij}) - s_{ij}]
\]

This method can be extended to the context of higher-order moment tensor matrices by choosing the function $g$ and the parameter vector $\theta$ in (38) appropriately. First, $\hat{\rho}_{ii}$ and $\rho_{ii}$ are trivially given by $\hat{\rho}_{ii}$ and $\hat{\rho}_{ii}$. Next, it is straightforward to see that the expression for $g$ and $\theta$ follows from the relationships established in (23) for the constant correlation estimator and in (25) for the single-factor estimator. As far as the constant correlation approach is concerned, the extension using the central limit theorem (36) is also straightforward. Following (13), we distinguish two different index combinations for the third-order moment tensor matrix, and four different index combinations for the fourth-order moment tensor matrix. As in (47) for the second-order moment tensor, we customise the parameter vector $\theta$ and the transformation function $g$ consistently with the higher-order extension of the constant correlation coefficients defined in (13). Using the above corollary of the central limit theorem, we find the asymptotic covariances between the constant correlation estimates and the sample estimates of the third- and fourth-order moment tensor matrix entries as the off-diagonal elements of the 2x2 matrix $V_{g}^{\Sigma^{CC}}V_{g}$ (36):
Applying the same logic as in (51), consistent estimators for the above asymptotic covariance terms are given as:

\[
\hat{\rho}_{ij}^{CC} = \frac{r}{2} \left[ \frac{m_j^6}{m_i^2} \text{AsyCov} \left( \sqrt{T m_i^{(2)}}, \sqrt{T s_{ij}} \right) + \frac{m_i^2}{m_j^6} \text{AsyCov} \left( \sqrt{T m_j^{(6)}}, \sqrt{T s_{ij}} \right) \right]
\]

\[
\hat{\rho}_{ij}^{CC} = \frac{r}{2} \left[ \frac{m_i^4}{m_j^4} \text{AsyCov} \left( \sqrt{T m_i^{(4)}}, \sqrt{T s_{ij}} \right) + \frac{m_j^4}{m_i^4} \text{AsyCov} \left( \sqrt{T m_j^{(4)}}, \sqrt{T s_{ij}} \right) \right]
\]

\[
\hat{\rho}_{ijk}^{CC} = \left\{ \begin{array}{l}
\frac{r}{2} \left[ \frac{m_j^6}{m_i^2} \text{AsyCov} \left( \sqrt{T m_i^{(2)}}, \sqrt{T s_{jk}} \right) \\
+ \frac{m_i^2}{m_j^6} \text{AsyCov} \left( \sqrt{T m_j^{(6)}}, \sqrt{T s_{ijk}} \right) \right] \\
+ \frac{r}{4} \left[ \frac{m_j^4}{m_i^4} \text{AsyCov} \left( \sqrt{T m_i^{(4)}}, \sqrt{T s_{ijk}} \right) \\
+ \frac{m_i^4}{m_j^4} \text{AsyCov} \left( \sqrt{T m_j^{(4)}}, \sqrt{T s_{ijk}} \right) \right] \\
+ \frac{r}{4} \left[ \frac{m_j^4}{m_i^4} \text{AsyCov} \left( \sqrt{T m_i^{(4)}}, \sqrt{T s_{ijkl}} \right) \\
+ \frac{m_i^4}{m_j^4} \text{AsyCov} \left( \sqrt{T m_j^{(4)}}, \sqrt{T s_{ijkl}} \right) \right] \\
+ \frac{r}{4} \left[ \frac{m_j^4}{m_i^4} \text{AsyCov} \left( \sqrt{T m_i^{(4)}}, \sqrt{T s_{ijkl}} \right) \\
+ \frac{m_i^4}{m_j^4} \text{AsyCov} \left( \sqrt{T m_j^{(4)}}, \sqrt{T s_{ijkl}} \right) \right] \end{array} \right\}
\]

(54)

As far as the single-factor approach is concerned, by construction, the co-moment elements in \( \Psi \) and \( \Phi \) are asymptotically uncorrelated with \( s_{ij} \), whereas the situation is less straightforward for the elements of \( \Upsilon \) (see equations 27-29). Accordingly, we can derive results similar to (52) for the asymptotic covariance between the single-factor estimate of the third-order moment tensor matrix and the corresponding sample estimate. We set:
\[ \Theta_3^{SF} = (s_{i_0} s_{j_0} s_{k_0} m_0^{(2)} m_0^{(3)} s_{l_0})^T \]
\[ g_3^{SF}(\Theta_3^{SF}) = \begin{pmatrix}
\frac{s_{i_0}s_{j_0}s_{k_0}}{m_0^{(2)}} m_0^{(3)} \\
\frac{m_0^{(2)}}{s_{i_0} s_{j_0} s_{k_0}} s_{l_0}
\end{pmatrix} = \begin{pmatrix}
\beta, \beta, \beta, m_0^{(3)} \\
\beta, \beta, \beta, s_{l_0}
\end{pmatrix} \]

and obtain:
\[ \hat{\rho}_{ijk} = \frac{s_{i_0}s_{j_0}s_{k_0}m_0^{(3)}}{m_0^{(2)}}, \text{AsyCov} \left( \sqrt{T_{s_{i_0}}, \sqrt{T_{s_{j_0}}}} \right), \quad \frac{s_{i_0}s_{j_0}s_{k_0}m_0^{(3)}}{m_0^{(2)}}, \text{AsyCov} \left( \sqrt{T_{s_{j_0}}, \sqrt{T_{s_{k_0}}}} \right) \]
\[ + \frac{s_{i_0}s_{j_0}s_{k_0}m_0^{(3)}}{m_0^{(2)}}, \text{AsyCov} \left( \sqrt{T_{s_{k_0}}, \sqrt{T_{s_{j_0}}}} \right), \quad \frac{3s_{i_0}s_{j_0}s_{k_0}m_0^{(3)}}{m_0^{(2)}}, \text{AsyCov} \left( \sqrt{T_{m_0^{(2)}}, \sqrt{T_{s_{j_0}}}} \right) \]
\[ + \frac{s_{i_0}s_{j_0}s_{k_0}m_0^{(3)}}{m_0^{(2)}}, \text{AsyCov} \left( \sqrt{T_{m_0^{(3)}}, \sqrt{T_{s_{j_0}}}} \right) \]

As evidenced by (29), the situation is more complex for the fourth-order moment tensor matrix estimator, where we need to distinguish four cases according to the index permutations of \( \hat{\rho}_{ijkl} \). Let us first analyse the component that is common to all index combinations (similar to (57)). Accordingly, we define:

\[ \Theta_4^{SF} = (s_{i_0} s_{j_0} s_{k_0} s_{l_0} m_0^{(2)} m_0^{(4)} s_{l_0})^T \]
\[ g_4^{SF}(\Theta_4^{SF}) = \begin{pmatrix}
\frac{s_{i_0}s_{j_0}s_{k_0}s_{l_0}}{m_0^{(2)}} m_0^{(4)} \\
\frac{m_0^{(2)}}{s_{i_0} s_{j_0} s_{k_0} s_{l_0}} s_{l_0}
\end{pmatrix} = \begin{pmatrix}
\beta, \beta, \beta, \beta, m_0^{(4)} \\
\beta, \beta, \beta, \beta, s_{l_0}
\end{pmatrix} \]

and obtain:
\[ \hat{\rho}_{ijkl} = \frac{s_{i_0}s_{j_0}s_{k_0}s_{l_0}m_0^{(4)}}{m_0^{(2)}}, \text{AsyCov} \left( \sqrt{T_{s_{i_0}}, \sqrt{T_{s_{j_0}}}} \right), \quad \frac{s_{i_0}s_{j_0}s_{k_0}s_{l_0}m_0^{(4)}}{m_0^{(2)}}, \text{AsyCov} \left( \sqrt{T_{s_{j_0}}, \sqrt{T_{s_{k_0}}}} \right) \]
\[ + \frac{s_{i_0}s_{j_0}s_{k_0}s_{l_0}m_0^{(4)}}{m_0^{(2)}}, \text{AsyCov} \left( \sqrt{T_{s_{k_0}}, \sqrt{T_{s_{j_0}}}} \right), \quad \frac{s_{i_0}s_{j_0}s_{k_0}s_{l_0}m_0^{(4)}}{m_0^{(2)}}, \text{AsyCov} \left( \sqrt{T_{m_0^{(2)}}, \sqrt{T_{s_{k_0}}}} \right) \]
\[ + \frac{s_{i_0}s_{j_0}s_{k_0}s_{l_0}m_0^{(4)}}{m_0^{(2)}}, \text{AsyCov} \left( \sqrt{T_{m_0^{(3)}}, \sqrt{T_{s_{l_0}}}} \right) \]
\[ + r_{ijkl}^{*}. \]

where \( r_{ijkl}^{*} \) denotes the term that is conditional on the index combination of the several moment tensor elements. Applying the same algorithm and consistent with (29), we have:
Finally, all asymptotic covariances in (57), (59) and (60) are, similarly to (53), consistently estimated by:

\[ \text{AsyCov}\left(\sqrt{T}s_{o_i}, \sqrt{T}s_{ij} \right) = \frac{1}{T} \sum_{t=1}^{T} \left[ (R_{o_i} - m_{o_i})(R_{ij} - m_{ij}) - s_{o_i} \right] \left[ (R_{o_i} - m_{o_i})(R_{ij} - m_{ij}) - s_{o_i} \right] \]

\[ \text{AsyCov}\left(\sqrt{T}S_{o_i}, \sqrt{T}s_{ij} \right) = \frac{1}{T} \sum_{t=1}^{T} \left[ (R_{o_i} - m_{o_i})(R_{ij} - m_{ij}) - s_{o_i} \right] \left[ (R_{o_i} - m_{o_i})(R_{ij} - m_{ij}) - s_{o_i} \right] \]

for the third-order components, and by:

\[ \text{AsyCov}\left(\sqrt{T}s_{o_i}, \sqrt{T}s_{ij} \right) = \frac{1}{T} \sum_{t=1}^{T} \left[ (R_{o_i} - m_{o_i})(R_{ij} - m_{ij}) - s_{o_i} \right] \left[ (R_{o_i} - m_{o_i})(R_{ij} - m_{ij}) - s_{o_i} \right] \]

\[ \text{AsyCov}\left(\sqrt{T}S_{o_i}, \sqrt{T}s_{ij} \right) = \frac{1}{T} \sum_{t=1}^{T} \left[ (R_{o_i} - m_{o_i})(R_{ij} - m_{ij}) - s_{o_i} \right] \left[ (R_{o_i} - m_{o_i})(R_{ij} - m_{ij}) - s_{o_i} \right] \]

for the fourth-order components, where
Overall, equations (31), (33), (34) and (54)-(56) for the constant correlation estimator, as well as (57)-(62) for the single-factor estimator, allow us to obtain the asymptotically optimal linear combination of the constant correlation (single factor) estimate and the sample estimate of $M_2$, $M_3$, and $M_4$.

For the optimal shrinkage intensities (31), we obtain:

$$s_{x0} = \frac{1}{T} \sum_{t=1}^{T} (\hat{R}_{yt} - m_y)(\hat{R}_{yt} - m_y)$$

$$\hat{\psi}_t = \frac{1}{T} \sum_{t=1}^{T} e_t^2.$$

Overall, equations (31), (33), (34) and (54)-(56) for the constant correlation estimator, as well as (57)-(62) for the single-factor estimator, allow us to obtain the asymptotically optimal linear combination of the constant correlation (single factor) estimate and the sample estimate of $M_2$, $M_3$, and $M_4$.

For the optimal shrinkage intensities (31), we obtain:

$$\hat{\alpha}_2^C = \frac{1}{\gamma^C_2} \hat{\alpha}_2^C$$

$$\hat{\alpha}_3^C = \frac{1}{\gamma^C_3} \hat{\alpha}_3^C$$

$$\hat{\alpha}_4^C = \frac{1}{\gamma^C_4} \hat{\alpha}_4^C$$

$$\hat{\alpha}_2^F = \frac{1}{\gamma^F_2} \hat{\alpha}_2^F$$

$$\hat{\alpha}_3^F = \frac{1}{\gamma^F_3} \hat{\alpha}_3^F$$

$$\hat{\alpha}_4^F = \frac{1}{\gamma^F_4} \hat{\alpha}_4^F$$

with

$$\hat{\pi}_2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\pi}_{ij}$$

$$\hat{\pi}_3 = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \hat{\pi}_{ijk}$$

$$\hat{\pi}_4 = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} \hat{\pi}_{ijkl}$$

and

$$\gamma_2^C = \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_{ij}^C \gamma_2^F = \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_{ij}^F$$

$$\gamma_3^C = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \gamma_{ijk}^C \gamma_3^F = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \gamma_{ijk}^F$$

$$\gamma_4^C = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} \gamma_{ijkl}^C \gamma_4^F = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} \gamma_{ijkl}^F$$

where $\hat{\gamma}_{ij}^C$ ($\hat{\gamma}_{ij}^F$) indicates that the constant correlation (single-factor) estimates have been used for the $\lambda_{ij}$ in (34). Finally,

$$\hat{\rho}_2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\rho}_{ij}$$

$$\hat{\rho}_3 = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \hat{\rho}_{ijk}$$

$$\hat{\rho}_4 = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} \hat{\rho}_{ijkl}$$

Consequently, we can define our 6 shrinkage estimators, that is, for each moment tensor ($M_2$, $M_3$, and $M_4$) one estimator shrunk towards the constant correlation estimate and one estimator shrunk towards the single-factor estimate.
References


