Improved Estimates of Higher-Order Comoments and Implications for Portfolio Selection

February 2010

Lionel Martellini
Professor of Finance, EDHEC Business School

Volker Ziemann
French Ministry of Economy, Industry and Employment
Abstract
In the presence of non-normally distributed asset returns, optimal portfolio selection techniques require estimates for variance-covariance parameters, along with estimates for higher-order moments and comoments of the return distribution. This is a formidable challenge that severely exacerbates the dimensionality problem already present with mean-variance analysis. This paper extends the existing literature, which has mostly focused on the covariance matrix, by introducing improved estimators for the coskewness and cokurtosis parameters. We find that the use of these enhanced estimates generates a significant improvement in investors’ welfare. We also find that it is only when improved estimators are used that portfolio selection with higher-order moments dominates mean-variance analysis from an out-of-sample perspective.

We would like to thank the editor, Joel Hasbrouck, as well as an anonymous referee for their very useful comments. We also thank René Garcia, Patrick Rousseau, Pierre Batteau, Patrick Navatte, and seminar participants at EDHEC Business school and University of Marseille-Aix III for very helpful suggestions. We further acknowledge financial support from Newedge Prime Brokerage, the sponsor of the Research Chair in “Advanced Modelling for Alternative Investments” at EDHEC Risk and Asset Management Research Center. Please send any correspondence to Lionel Martellini, 393-400 Promenade des Anglais, BP 3116, 06202 Nice Cedex 3. Tel. +33 (0)4 93 18 78 24. E-mail addresses: lionel.martellini@edhec.edu; volker.ziemann@dgtpe.fr.

EDHEC is one of the top five business schools in France. Its reputation is built on the high quality of its faculty and the privileged relationship with professionals that the school has cultivated since its establishment in 1906. EDHEC Business School has decided to draw on its extensive knowledge of the professional environment and has therefore focused its research on themes that satisfy the needs of professionals.

EDHEC pursues an active research policy in the field of finance. The EDHEC-Risk Institute carries out numerous research programmes in the areas of asset allocation and risk management in both the traditional and alternative investment universes.
Past research has documented 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, for example, Hong, Tu, and Zhou (2007) for a recent reference]. Under relatively weak assumptions regarding the shape of the utility function, Horvath and Scott (1980) and Kimball (1993) show that typical investors exhibit non-trivial preferences with respect to portfolio higher-order moments, in addition to mean and variance. In particular, it has been shown that investors are willing to accept lower expected return and higher volatility compared to the mean-variance benchmark in exchange for higher skewness and lower kurtosis [see Harvey and Siddique (2000), Dittmar (2002), or Mitton and Vorkink (2007)].

In the presence of non-normally distributed asset returns, optimal portfolio selection techniques require estimates for variance-covariance parameters, along with estimates for higher-order moments and comoments of the return distribution. The need to estimate coskewness and cokurtosis parameters severely increases, however, the dimensionality problem, already a serious concern in the context of covariance matrix estimation. For example, optimizing a portfolio having 20 stocks would require the estimation of 210 variance-covariance parameters, while it would require estimating 1,540 skewness-coskewness parameters and 8,855 kurtosis-cokurtosis parameters. In this context, given the dramatic increase in dimensionality involved, one might wonder whether portfolio selection techniques that rely on higher-order moments can efficiently be implemented at all in realistic situations. This concern has been emphasized, for example, by Brandt, Santa-Clara, and Valkanov (2009), who argue that "extending the traditional approach beyond first and second moments, when the investor's utility function is not quadratic, is practically impossible because it requires modeling [...] the numerous higher-order cross-moments." The authors propose an alternative approach that models the portfolio weights as functions of the assets' characteristics. Our paper attempts to show that implementing portfolio selection with higher-order moments can in fact be achieved through suitable extensions of various statistical techniques that have proven very useful in the context of covariance matrix estimation to the context of higher-order moments. Broadly speaking, these techniques consist of allowing for a significant decrease in the parameter space dimension. Indeed, in the mean-variance context, it has long been recognized 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 covariance parameter 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 notably the constant correlation estimation (Elton and Gruber, 1973), the single factor estimation (Sharpe, 1963) and the multi-factor estimation (Chan, Karceski, and Lakonishok, 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 error and specification error 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 involves a shrinkage effect and generates comparable results in terms of out-of-sample standard deviation.

While a variety of methods have been introduced to reduce the number of covariance parameters in a portfolio selection context, very little, if anything, is known about improved estimates for higher-order comoments of asset return distributions. Previous literature has mostly focused on asset pricing implications of non-trivial preferences for higher-order 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) and Ang, Chen, and Xing (2006)]. Very few papers have in fact focused on the problem of improving higher-order moment estimates. Kim

1 - See Table 1 for more details.
and White (2004) have introduced robust estimates for skewness and kurtosis parameters based on inter-quartile differences but they merely focus on univariate analyses, with no effort on improving higher-order comoments, which is obviously a major concern for portfolio selection techniques. There is also a related paper by Harvey, Liechty, Liechty, and Müller (2004), who introduce a semi-parametric approach to model skewness in the multivariate portfolio allocation process and address the problem of parameter uncertainty in a skew-normal framework. Finally, Jondeau and Rockinger (2003) and Patton (2004) have considered asset allocation models based on conditional models for higher-order moments of asset return distributions. Their analysis, however, focuses on a different key aspect in parameter estimation, namely the issue of time-varying conditional moments, as opposed to sample risk.

This paper extends the existing literature on improved estimates for the variance-covariance matrix by introducing enhanced estimates for coskewness and cokurtosis parameters. More specifically, our contribution from a theoretical standpoint is to introduce suitable extensions to higher-order comoments of several estimators that have been found useful in improving covariance matrix estimates, namely the constant correlation estimator, the factor-based estimator, as well as statistical shrinkage estimators. The empirical performance of these estimators is subsequently assessed in the context of an empirical analysis similar to those conducted in Chan, Karceski, and Lakonishok (1999) and Jagannathan and Ma (2003). We find that the use of these improved estimators leads to welfare gains that are comparable to what has been found in the context of portfolio construction techniques aiming at capturing the benefits of international diversification (Ang and Bekaert, 2002) or return predictability (Kandel and Stambaugh, 1996). Additionally, we find that the use of these improved estimators significantly enhances the stability of the constructed portfolios; as a consequence, utility gains are even higher when transaction costs are accounted for. We also find that improved estimators greatly enhance the probability that higher-order moment portfolio selection dominates mean-variance analysis from an out-of-sample perspective even for small sample sizes.

The rest of the paper is organized as follows. In Section 1, we motivate the introduction of higher moments in optimal portfolio selection techniques. 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. We also extend Ledoit and Wolf (2003, 2004) by deriving explicit expressions for optimal shrinkage intensities for coskewness and cokurtosis matrices. In Section 3, we assess the empirical performance of these improved estimators. Finally, Section 4 concludes the paper. The Appendix provides the technical details.

1. Standard Risk Aversion and Higher-Order Moment Tensors

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

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

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

$$U(W) \approx U(E(W)) + \sum_{k=1}^{4} \frac{U^{(k)}(E(W))}{k!} (W - E(W))^k.$$
Applying the expectation operator to both sides of Equation (2), we can approximate expected utility as:

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

(3)

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

$$
\mu^{(n)} = E ((W - E(W))^n).
$$

(4)

From (3), we see that expected utility derived from an investment in risky assets may be approximated by the derivatives of the utility function and the first four moments of the portfolio 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 centered moments of asset returns. As a result, the moments of portfolio returns in (2) are a function of portfolio weights, as well as first, second, third, and fourth-order moments and comoments of individual asset return distributions. In order to make this relationship explicit, we introduce so-called higher-order moment tensors (Jondeau and Rockinger, 2003) for the set of individual assets under consideration. The second moment tensor $M_2$ is the standard variance-covariance matrix. As far as the higher-order moment tensors $M_3$ and $M_4$ are concerned, we adopt the notation from Harvey, Liechty, Liechty, and Müller (2004) and Jondeau and Rockinger (2006) and stack the sub-comoment matrices column-wise. We define:

$$
s_{ijk} = E [(R_i - \mu_i)(R_j - \mu_j)(R_k - \mu_k)] \quad \forall \; i, j, k = 1 \ldots n
$$

$$
k_{ijkl} = E [(R_i - \mu_i)(R_j - \mu_j)(R_k - \mu_k)(R_l - \mu_l)] \quad \forall \; i, j, k, l = 1 \ldots n.
$$

(5)

where $R_s$ denotes the return on asset $s$.

For the sake of illustration, we write the higher-order moment tensors expressions for $N = 3$ assets:

$$
M_3 = \begin{bmatrix} S_1 & S_2 & S_3 \end{bmatrix}
$$

$$
M_4 = \begin{bmatrix} K_{11} & K_{12} & K_{13} & K_{21} & K_{22} & K_{23} & K_{31} & K_{32} & K_{33} \end{bmatrix}
$$

(6)

with

$$
S_p = \begin{bmatrix} s_{p11} & s_{p12} & s_{p13} \\
 s_{p21} & s_{p22} & s_{p23} \\
 s_{p31} & s_{p32} & s_{p33} \end{bmatrix} \quad K_{pq} = \begin{bmatrix} k_{pq11} & k_{pq12} & k_{pq13} \\
 k_{pq21} & k_{pq22} & k_{pq23} \\
 k_{pq31} & k_{pq32} & k_{pq33} \end{bmatrix}.
$$

(7)

This notation allows us to remain in the matrix space when analyzing high dimensional tensors. Note that $S_p$ and $K_{pq}$ are $N \times N$ matrices. Using the Kronecker product, we obtain:

$$
M_2 = E [(R - E(R))(R - E(R))']
$$

$$
M_3 = E [(R - E(R))(R - E(R))'] \otimes (R - E(R))'
$$

$$
M_4 = E [(R - E(R))(R - E(R))' \otimes (R - E(R))' \otimes (R - E(R))'].
$$

(8)
Finally, it can easily be shown that the portfolio moments in (2) are polynomial functions in the portfolio weight vector, parameterized by the corresponding higher-order moment tensors:

\[
\begin{align*}
\mu^{(2)} &= \omega' M_2 \omega \\
\mu^{(3)} &= \omega' M_3 (\omega \otimes \omega) \\
\mu^{(4)} &= \omega' M_4 (\omega \otimes \omega \otimes \omega)
\end{align*}
\]  

(9)

with \( \omega \), the \( N \times 1 \) vector containing the portfolio weights corresponding to the \( N \) available assets.

Assuming that the wealth \( W \) in (2) is entirely determined by the portfolio outcome and taking the initial wealth as a numeraire, we can rewrite the investor’s optimization problem as a function of asset return moment tensors and the vector of portfolio weights:

\[
\max_{\omega} \left[ U(\mathbb{E}(\mu'w)) + \frac{U^{(2)}(\mu'w)}{2} \omega' M_2 \omega + \frac{U^{(3)}(\mu'w)}{6} \omega' M_3 (\omega \otimes \omega) + \frac{U^{(4)}(\mu'w)}{24} \omega' M_4 (\omega \otimes \omega \otimes \omega) \right]
\]  

(10)

where \( \mu \) denotes the \( N \times 1 \) vector of mean gross asset returns. In contrast to bivariate asset pricing models, where only marginal co-movements with the market are of interest [e.g., Dittmar (2002) and Ang, Chen, and Xing (2006)], portfolio allocation models involves all possible permutations in the comoment space.

2. Estimators for Higher-Order Moment Tensors

The number of parameters involved when higher-order moments are taken into account increases exponentially with the number of risky assets in the portfolio. Given the total numbers of parameters to estimate, the higher-order moment tensor matrices will be rank-deficient even for relatively small portfolios. For 20 assets, for instance, one would need 45 years of monthly data so as to ensure that the number of observations exceeds the maximal possible rank of the corresponding moment tensors. Following the literature on improved estimates for the variance-covariance matrix, we next impose some structure on the higher-order moment tensor matrices so as to reduce the number of parameters involved.

2.1 Constant correlation estimator

The constant correlation estimator was 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 (1973) argue that imposing the assumption of a constant correlation across assets, while obviously involving significant specification error, leads to improved out-of-sample portfolio performance. An unbiased estimator for the constant correlation parameter is given by the average over all sample correlation parameters:

\[
\hat{\rho} = \frac{2}{N(N-1)} \sum_{i<j}^{N} \hat{r}_{ij} \quad \text{with} \quad \hat{r}_{ij} = \frac{s_{ij}}{\sqrt{s_{ii} s_{jj}}},
\]

(11)

where \( s \) denotes sample variances and covariances. Following this approach, the covariance (\( \sigma_{ij} \)) parameters can be estimated as a function of the constant correlation parameter and the asset volatility parameters \( \hat{\sigma}_{ij} = \hat{\rho} \sqrt{s_{ii} s_{jj}} \), thus allowing for a dramatic reduction in the number of parameters. Following the pioneering work by Elton and Gruber (1973), empirical research has assessed the out-of-sample properties of portfolios constructed employing the constant correlation estimator. In particular, Chan, Karceski, and Lakonishok (1999), Jagannathan and Ma (2003) and Ledoit and Wolf (2003), among others, have shown that the realized volatility for minimum variance

---

2 - Indeed, as shown in Table 1, the total number of required parameters is 10,605 in this case, calling for 45 years of monthly data for the 20 assets so as to ensure that the number of observations \( (N \times T \times 12) \) merely exceeds the number of parameters.
portfolios constructed with the sample estimator of the covariance matrix significantly exceeded the realized volatility of portfolios constructed with the constant correlation estimator.

We now derive the counterparts of correlation coefficients for higher-order comoments and extend the concept of a constant correlation to the context of higher-order moment tensors. We include all possible combinations of higher-order comoments according to the definition of the matrices $M_3$ and $M_4$ as in (5). Denoting $\overline{X}$ as the centered version of the variable of $X$, we have:

\[ s_{ijk} = E(\overline{R}_i \overline{R}_j \overline{R}_k) \]
\[ k_{ijkl} = E(\overline{R}_i \overline{R}_j \overline{R}_k \overline{R}_l) \]  \hspace{1cm} (12)

where $\overline{R}$ denotes centered returns, that is, $\overline{R} = R - \mu$.

According to several permutations in (12), we introduce the following extended correlation coefficients, where $r^{(1)}$ is the standard correlation coefficient:

\[
\begin{align*}
    r^{(1)}_{ij} &= \frac{E(R_i R_j)}{\sqrt{\mu_i^{(2)} \mu_j^{(2)}}} \\
    r^{(2)}_{ij} &= \frac{E(R_i^2 R_j)}{\sqrt{\mu_i^{(4)} \mu_j^{(2)}}} \\
    r^{(3)}_{ij} &= \frac{E(R_i R_j^2)}{\sqrt{\mu_i^{(2)} \mu_j^{(4)}}} \\
    r^{(4)}_{ijk} &= \frac{E(R_i R_j R_k)}{\sqrt{\mu_i^{(2)} \mu_j^{(2)} \mu_k^{(2)}}} \\
    r^{(5)}_{ijk} &= \frac{E(R_i R_j R_k R_l)}{\sqrt{\mu_i^{(4)} \mu_j^{(2)} \mu_k^{(2)}}} \\
    r^{(7)}_{ijkl} &= \frac{E(R_i R_j R_k R_l)}{\sqrt{\mu_i^{(2)} \mu_j^{(2)} \mu_k^{(2)} \mu_l^{(2)}}}
\end{align*}
\]  \hspace{1cm} (13)

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

\[ r^{(n)} \in [-1, 1] \hspace{1cm} n = 1...7. \]  \hspace{1cm} (14)

Consistent with Elton and Gruber (1973), we assume that the parameters $r^{(1)} - r^{(7)}$ are constant across the assets. In other words, we use the following estimators, corresponding to sample average correlations across assets:

\[
\hat{r}^{(1)} = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{\sum_{t=1}^{T} (\overline{R}_i \overline{R}_j) / T}{\sqrt{m_i^{(2)} m_j^{(2)}}}
\]

\[
\hat{r}^{(7)} = \frac{24}{N(N-1)(N-2)(N-3)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=j+1}^{N} \sum_{l=k+1}^{N} \frac{\sum_{t=1}^{T} (\overline{R}_i \overline{R}_j \overline{R}_k \overline{R}_l) / T}{\sqrt{\hat{r}^{(5)} m_i^{(4)} m_j^{(2)} m_k^{(2)} m_l^{(2)}}}
\]  \hspace{1cm} (15)
where \( \overline{R}_{it} \) denotes the centered return of asset \( i \) at time \( t \). \( m_i^{(n)} \) is the \( n \)-th centered sample moment of asset \( i \) given by:

\[
m_i^{(n)} = \frac{1}{T} \sum_{t=1}^{T} (\overline{R}_{it})^n
\]

Replacing individual sample correlations by those seven constant coefficients allows us to determine all the elements in \( M_2 \), \( M_3 \), and \( M_4 \) by simply using sample estimates for the first, second, third, fourth, and sixth central moments:

\[
\begin{align*}
\hat{s}_{ij} &= \hat{\rho}^{(2)} \sqrt{m_i^{(4)} m_j^{(2)}} \\
\hat{s}_{ij} &= \hat{\rho}^{(4)} \sqrt{m_i^{(2)} m_j^{(5)}} \sqrt{m_i^{(4)} m_j^{(4)}}
\end{align*}
\]

\[
\begin{align*}
\hat{k}_{iij} &= \hat{\rho}^{(3)} \sqrt{m_i^{(6)} m_j^{(2)}} \\
\hat{k}_{iij} &= \hat{\rho}^{(5)} \sqrt{m_i^{(4)} m_j^{(4)}} \\
\hat{k}_{iij} &= \hat{\rho}^{(6)} \sqrt{m_i^{(4)} m_j^{(5)}} \sqrt{m_i^{(4)} m_j^{(4)}} \\
\hat{k}_{ijkl} &= \hat{\rho}^{(7)} \sqrt{m_i^{(4)} m_j^{(5)}} \sqrt{m_j^{(5)} m_k^{(4)}} \sqrt{m_i^{(4)} m_j^{(4)}}
\end{align*}
\]

This approach allows us to substantially reduce the number of parameters compared to the sample estimators (Table 1).

2.2 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:

\[
R_{it} = c + \beta_i F_t + \epsilon_{it}, \quad (18)
\]

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:

\[
\epsilon \sim (0, \Psi). \quad (19)
\]

According to the assumptions, all off-diagonal elements in \( \Psi \) are zero. The idiosyncratic risks 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:

\[
M_2 = \beta \beta' \mu_0^{(2)} + \Psi, \quad (20)
\]

where \( \beta \) indicates the \( N \times 1 \) vector of the regression coefficients in (18) and \( \mu_0^{(2)} \) the second centered moment, that is the variance, of the single factor. Next, we derive the corresponding single-factor estimators for the moment tensors \( M_3 \) and \( M_4 \).\(^3\) We substitute (18) in (8) to obtain:

\[
\begin{align*}
M_3 &= \mathbb{E} \left[ (\beta \overline{F} + \epsilon)(\beta \overline{F} + \epsilon)' \otimes (\beta \overline{F} + \epsilon)' \right] \\
M_4 &= \mathbb{E} \left[ (\beta \overline{F} + \epsilon)(\beta \overline{F} + \epsilon)' \otimes (\beta \overline{F} + \epsilon)' \otimes (\beta \overline{F} + \epsilon)' \right]
\end{align*}
\]

where \( \overline{F} \) is the \( T \times 1 \) vector of centered market returns \( (\overline{F} = F - \mu_0) \). We finally obtain the following decompositions:

\(^3\) 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 (e.g., Chan, Karceski, and Lakonishok [1999] and Connor and Korajczyk [1993]).
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 $\varepsilon_i$ and $\varepsilon_j$ ($i \neq j$) are independent. Additionally, due to the fundamental properties of the least-squares regression technique, we also have that the factor return process is independent from the residual return process.

The structure of the $N \times N$ covariance matrix $\Psi$ of residual returns is well-known to be:

$$\psi_{ii} = \mathbb{E}(\varepsilon_i^2), \text{ with a sample estimate defined as } \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it}^2$$

$$\psi_{ij} = 0 \quad \forall i \neq j.$$  

(24)

The structure of the $N \times N^2$ specific return coskewness matrix $\Phi$ is similar:

$$\phi_{iii} = \mathbb{E}(\varepsilon_i^3), \text{ with a sample estimate defined as } \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it}^3$$

$$\phi_{iij} = 0$$

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

(25)

On the superdiagonal, we find the expressions for idiosyncratic skewnesses proxied by the third-order sample moment of the residuals. For all other index permutations, that is, for all off-superdiagonal elements, either the single factor $F$ or the residual term $\varepsilon$ or both enter the product obtained by (21) with the power of one and thus with expectation zero. The independence assumption then explains why all off-superdiagonal elements are zero.

Obtaining the entries of the $N \times N^3$ matrix follows as:

$$v_{iii} = \mathbb{E}(\varepsilon_i^4), \text{ with a sample estimate defined as } \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it}^4$$

$$v_{iij} = 3\beta_i\beta_j\mu_0^{(2)}\psi_{ii}$$

$$v_{iij} = \beta_i^2\mu_0^{(2)}\psi_{jj} + \beta_j^2\mu_0^{(2)}\psi_{ii} + \psi_{ii}\psi_{jj}$$

$$v_{iik} = \beta_i\beta_k\mu_0^{(2)}\psi_{ii}$$

$$v_{ijkl} = 0 \quad \forall i \neq j \neq k \neq l.$$  

(26)

On the superdiagonal, we have the idiosyncratic kurtosises proxied by the fourth-order sample moment of the residuals. As far as off-superdiagonal elements are concerned, the only elements in (22) that are not zero are those where neither the single factor $F$ nor the residuals term $\varepsilon$ enter to the power of one. The results in (26) are easy to verify by developing the elements from (22). For the permutation case $iii$, for instance, we obtain:

$$M_4^{iii} = \mathbb{E}\left[(\beta_iF + \varepsilon_i)^3(\beta_jF + \varepsilon_j)\right]$$

$$= \mathbb{E}\left[\beta_i^3F^3 + 3\beta_i^2F^2\varepsilon_i + 3\beta_iF\varepsilon_i^2 + \varepsilon_i^3\right]$$

(27)

$$= \beta_i^3\beta_j\mu_0^{(4)} + 3\beta_i\beta_j\mu_0^{(2)}\psi_{ii}. $$

4 - Note that for non-Gaussian variables, the assumption of independence is stronger than absence of correlation.
This last line follows from the fact that all other elements host either \( \mathbb{E}(F) \) or \( \mathbb{E}(e_i) \) or \( \mathbb{E}(e_j) \), which are all zero, and all terms containing these elements are also zero due to the independence assumption. Note further, that the term \( \beta_i^2 \beta_j^2 \psi_{(4)} \) is an element of the \( (\beta \beta' \otimes \beta' \otimes \beta')P_{0}^{(4)} \) matrix in (23), and thus does not enter the \( \Upsilon \) matrix.

Finally, we report the number of parameters that enter the estimation process of the higher-order moment tensors according to the different estimation approaches in Table 1. The substantial reduction in dimensionality for structured estimators comes at the cost of specification error, related to the imposed structure that all cross-sectional residuals are independent (factor approach) or that joint distributions are driven by a constant higher-order correlation parameter (constant correlation approach).

**Table 1: Required number of parameters**

<table>
<thead>
<tr>
<th>Panel A: N = 5</th>
<th>M_2</th>
<th>M_3</th>
<th>M_4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>15</td>
<td>35</td>
<td>70</td>
<td>120</td>
</tr>
<tr>
<td>Constant Correlation</td>
<td>6</td>
<td>18</td>
<td>19</td>
<td>27</td>
</tr>
<tr>
<td>Single Factor</td>
<td>11</td>
<td>11</td>
<td>17</td>
<td>23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: N = 20</th>
<th>M_2</th>
<th>M_3</th>
<th>M_4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>210</td>
<td>1,540</td>
<td>8,855</td>
<td>10,605</td>
</tr>
<tr>
<td>Constant Correlation</td>
<td>21</td>
<td>63</td>
<td>64</td>
<td>87</td>
</tr>
<tr>
<td>Single Factor</td>
<td>41</td>
<td>41</td>
<td>62</td>
<td>83</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: N = 100</th>
<th>M_2</th>
<th>M_3</th>
<th>M_4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>5,050</td>
<td>171,700</td>
<td>4,421,275</td>
<td>4,598,025</td>
</tr>
<tr>
<td>Constant Correlation</td>
<td>101</td>
<td>303</td>
<td>304</td>
<td>407</td>
</tr>
<tr>
<td>Single Factor</td>
<td>201</td>
<td>201</td>
<td>302</td>
<td>403</td>
</tr>
</tbody>
</table>

The number of required parameters for each tensor and each estimation method is reported. \( N \) denotes the number of assets in the portfolio. The constant correlation estimator requires a total of \( 4N+7 \) parameters corresponding to the second, third, fourth, and sixth moments of each asset and the seven constant correlation coefficients. The single factor approach involves \( 4N+3 \) parameters, namely the second, third, and fourth moments of each asset, the beta exposure of each asset with the common factor and the second, third, and fourth moments of the common factor. The numbers for the sample estimator stem from the binomial coefficient \( \binom{N}{k} \) that gives the number of multi-sets of length \( k \) on \( N \) elements (e.g., \( k=2 \) for \( M_2 \)). The total number is thus obtained as \( \frac{1}{2}N(N+1) + \frac{1}{6}N(N+1)(N+2) + \frac{1}{24}N(N+1)(N+2)(N+3) \).

As can be seen from Table 1, the number of parameters increases exponentially in the number of assets for the sample estimator while the number of parameters is close to proportional to the number of assets for both structured estimators.

### 2.3 Shrinkage estimators for higher-order moment matrices

In the previous sub-sections, we introduced structured estimators for the higher-order moment matrices in an attempt to reduce the number of required parameter estimates compared to the sample estimator. On the one hand, these estimators involve lower estimation risk due to the imposed structure. On the other hand, they involve some misspecification inherent in the artificial structure imposed either by the constant correlation or single-factor model. In an attempt to find the optimal trade-off between sample risk and specification error, Ledoit and Wolf (2003) introduced 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 the combined estimator as:

\[
L(\alpha) = \| \alpha \Lambda + (1-\alpha)S - \Omega \|_F,
\]

with \( \Omega \) the true (unobservable) covariance matrix, \( S \) the sample estimator, \( \Omega \) the shrinkage target, that is, a structured estimator for \( \Omega \) (constant correlation or single-factor estimator).
Additionally, \( \| \cdot \|_F \) denotes the Frobenius norm of a matrix, defined as the square root of the sum of its squared entries. Minimizing the expected value of this function yields the optimal linear combination of the two estimators as parameterized by the shrinkage intensity \( \alpha \). Ledoit and Wolf (2003) derive an asymptotic estimator for this quantity as:

\[
\alpha^* = \frac{1}{T} \frac{\pi - \rho}{\gamma}
\]

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.

The authors further show that \( \pi_{ij} \) represents the asymptotic variance of the sample estimator, with a consistent estimator given by:

\[
\hat{\pi}_{ij} = \frac{1}{T} \sum_{t=1}^{T} \left[ s_{ijt} - s_{ij} \right]^2,
\]

where \( s_{ij} = \frac{1}{T} \sum_{t=1}^{T} s_{ijt} \).

Similarly, the \( \rho_{ij} \) parameters represent the asymptotic covariances between the sample and the structured estimator. Ledoit and Wolf (2003, 2004) derive explicit formulas for consistent estimators of \( \rho_{ij} (b_{ij}) \) parameters when \( \Omega \) is the covariance matrix and the single factor or the constant correlation estimator respectively (see below). Finally, \( \gamma_{ij} \) denotes the squared error of the structured estimator, with a consistent estimator given in Ledoit and Wolf (2003, 2004).

As far as higher-order moment tensors are concerned, we can directly obtain consistent estimators for the asymptotic variances of the sample estimator (\( \pi \)) and the misspecification of the structured estimator (\( \gamma \)). Indeed, since the Frobenius norm is not restricted to squared matrices, we obtain in our setting:

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

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

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

and

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

\[
\hat{\gamma}_{ijk} = \left( \hat{\lambda}_{ijk} - M_3^{(ijk)} \right)^2
\]

\[
\hat{\gamma}_{ijkl} = \left( \hat{\lambda}_{ijkl} - M_4^{(ijkl)} \right)^2
\]

where \( m_i \) denotes the sample mean of asset \( i \) and \( M_k^{(i)} \) are the sample estimates for the corresponding tensor entries, while \( \hat{\lambda}_{i1j} \) are structured estimates for 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 [e.g., Greene (2003, p. 913)]:

$$\sqrt{T} \left( g(\hat{\theta}) - g(\theta) \right) \xrightarrow{d} \mathcal{N} \left( 0_L, \nabla g^T \Sigma \nabla g \right), \tag{34}$$

where $\hat{\theta}$ is an unbiased estimator for the unknown $N \times 1$ parameter vector $\theta$ with finite asymptotic covariance matrix $\Sigma$, and the function $g$ is a real function with $g : \mathbb{R}^N \rightarrow \mathbb{R}^L$ and $\nabla_{\theta} g$ its Jacobian matrix with respect to $\theta$.

Accordingly, we need to define suitable functional transformations $g$ and parameter vectors $\theta$ in order to obtain explicit expressions for the asymptotic covariances ($\rho$) of the estimators. We denote $\rho_{ij}^{CC}$ as 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, $\rho_{ij}^{SF}$ 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 $\mu_i$ by $m_i$.

Obviously, $\rho_{ij}^{CC}$ and $\rho_{ij}^{SF}$ are identical and equal to $\pi_{ij}$. Next, we have in a straightforward manner:

$$\hat{\rho}_{ij}^{CC} = \text{AsyCov} \left( \sqrt{T} \hat{\beta}_1, \sqrt{T} \hat{s}_{ij} \right)$$

$$\hat{\rho}_{ij}^{SF} = \text{AsyCov} \left( \sqrt{T} \beta_1 \beta_2 m_0^{(2)}, \sqrt{T} \hat{s}_{ij} \right) = \text{AsyCov} \left( \sqrt{T} s_{ij} m_0^{(2)}, \sqrt{T} \hat{s}_{ij} \right) \tag{35}$$

where $s_{ij}$ denotes the sample covariance of assets $i$ and $j$ and lower index 0 indicates the single factor. We apply the delta-method and customize the vector $\theta$ and the function $g$ in order to decompose the asymptotic covariances in (35) and (35):

$$\theta_2^{CC} = (s_{ii}, s_{jj}, s_{ij})$$

$$\theta_2^{CC} (\theta_2^{CC}) = \begin{pmatrix} \hat{\beta}_1 \\ \sqrt{s_{ii} s_{jj}} \\ s_{ij} \end{pmatrix} \tag{36}$$

and

$$\theta_2^{SF} = (s_{i0}, s_{j0}, m_0^{(2)}, s_{ij})$$

$$\theta_2^{SF} (\theta_2^{SF}) = \begin{pmatrix} s_{i0} s_{j0}^{(2)} \\ m_0^{(2)} \\ s_{ij} \end{pmatrix} \tag{37}$$

We further note that:

$$\Sigma^{CC} = \begin{pmatrix} \text{AsyCov} \left( \sqrt{T} s_{ii}, \sqrt{T} s_{ii} \right) & \text{AsyCov} \left( \sqrt{T} s_{ii}, \sqrt{T} s_{jj} \right) & \text{AsyCov} \left( \sqrt{T} s_{ii}, \sqrt{T} s_{ij} \right) \\ \text{AsyCov} \left( \sqrt{T} s_{jj}, \sqrt{T} s_{ii} \right) & \text{AsyCov} \left( \sqrt{T} s_{jj}, \sqrt{T} s_{jj} \right) & \text{AsyCov} \left( \sqrt{T} s_{jj}, \sqrt{T} s_{ij} \right) \\ \text{AsyCov} \left( \sqrt{T} s_{ij}, \sqrt{T} s_{ii} \right) & \text{AsyCov} \left( \sqrt{T} s_{ij}, \sqrt{T} s_{jj} \right) & \text{AsyCov} \left( \sqrt{T} s_{ij}, \sqrt{T} s_{ij} \right) \end{pmatrix}$$

and

$$\Sigma^{SF} = \begin{pmatrix} \text{AsyCov} \left( \sqrt{T} s_{i0}, \sqrt{T} s_{i0} \right) & \text{AsyCov} \left( \sqrt{T} s_{i0}, \sqrt{T} s_{j0} \right) & \text{AsyCov} \left( \sqrt{T} s_{i0}, \sqrt{T} m_0^{(2)} \right) & \text{AsyCov} \left( \sqrt{T} s_{i0}, \sqrt{T} s_{ij} \right) & \text{AsyCov} \left( \sqrt{T} s_{i0}, \sqrt{T} s_{ij} \right) \\ \text{AsyCov} \left( \sqrt{T} s_{j0}, \sqrt{T} s_{i0} \right) & \text{AsyCov} \left( \sqrt{T} s_{j0}, \sqrt{T} s_{j0} \right) & \text{AsyCov} \left( \sqrt{T} s_{j0}, \sqrt{T} m_0^{(2)} \right) & \text{AsyCov} \left( \sqrt{T} s_{j0}, \sqrt{T} s_{ij} \right) & \text{AsyCov} \left( \sqrt{T} s_{j0}, \sqrt{T} s_{ij} \right) \\ \text{AsyCov} \left( \sqrt{T} m_0^{(2)}, \sqrt{T} s_{i0} \right) & \text{AsyCov} \left( \sqrt{T} m_0^{(2)}, \sqrt{T} s_{j0} \right) & \text{AsyCov} \left( \sqrt{T} m_0^{(2)}, \sqrt{T} m_0^{(2)} \right) & \text{AsyCov} \left( \sqrt{T} m_0^{(2)}, \sqrt{T} s_{ij} \right) & \text{AsyCov} \left( \sqrt{T} m_0^{(2)}, \sqrt{T} s_{ij} \right) \\ \text{AsyCov} \left( \sqrt{T} s_{ij}, \sqrt{T} s_{i0} \right) & \text{AsyCov} \left( \sqrt{T} s_{ij}, \sqrt{T} s_{j0} \right) & \text{AsyCov} \left( \sqrt{T} s_{ij}, \sqrt{T} m_0^{(2)} \right) & \text{AsyCov} \left( \sqrt{T} s_{ij}, \sqrt{T} s_{ij} \right) & \text{AsyCov} \left( \sqrt{T} s_{ij}, \sqrt{T} s_{ij} \right) \end{pmatrix}.$$
Next we derive the Jacobian matrix of $g_{CC}$ and $g_{SF}$ and use (34) to obtain:

\[
\hat{\rho}_{ij}^{CC} = \frac{\bar{\tau}_{ij}}{2} \left[ \sqrt{\frac{s_{ij}}{s_{ii}}} \text{AsyCov} \left( \sqrt{T_{si}}, \sqrt{T_{sj}} \right) + \sqrt{\frac{s_{ii}}{s_{jj}}} \text{AsyCov} \left( \sqrt{T_{sj}}, \sqrt{T_{si}} \right) \right]
\]

\[
\hat{\rho}_{ij}^{SF} = \frac{s_{ij}}{m_{0}^{(2)}} \text{AsyCov} \left( \sqrt{T_{si}}, \sqrt{T_{sj}} \right) + \frac{s_{ij}}{m_{0}^{(2)}} \text{AsyCov} \left( \sqrt{T_{sj}}, \sqrt{T_{si}} \right)
\]

\[
- \frac{s_{ij}^{2} (2m_{0})^{2}}{m_{0}^{(2)}} \text{AsyCov} \left( \sqrt{T_{m_{i}^{(2)}}, \sqrt{T_{m_{j}^{(2)}}}} \right)
\]

(38)

where $\hat{\rho}_{ij}^{CC}$ and $\hat{\rho}_{ij}^{SF}$ are the off-diagonal elements of the 2x2 matrices $\nabla_{g}^{T} \Sigma_{g}^{CC} \nabla_{g}$ and $\nabla_{g}^{T} \Sigma_{g}^{SF} \nabla_{g}$, respectively. Referring to Ledoit and Wolf (2003), we introduce consistent estimators for the asymptotic covariances in (38):

\[
\text{AsyCov} \left( \sqrt{T_{si}}, \sqrt{T_{sj}} \right) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{2} \left[ (R_{ii} - m_{i})^{2} - s_{ii} \right] \left[ (R_{ii} - m_{i})(R_{jj} - m_{j}) - s_{ij} \right]
\]

\[
\text{AsyCov} \left( \sqrt{T_{si}}, \sqrt{T_{sj}} \right) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{2} \left[ (R_{kt} - m_{k})(R_{ft} - m_{f}) - s_{kt} \right] \left[ (R_{kt} - m_{k})(R_{ft} - m_{f}) - s_{kt} \right]
\]

(39)

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 (34) appropriately. In what follows, we explain how to use this approach for the third moment tensor, and refer to the Appendix for more details on how to apply the procedure for the fourth moment tensor.

First, note that the $\hat{\rho}_{iii}$ are trivially given by $\hat{\tau}_{iii}$. As far as the constant correlation approach is concerned, the extension using the delta-method is straightforward. As in (36) for the covariance matrix, we customize the parameter vector $\theta$ and the transformation function $g$ consistently with the higher-order extension of the constant correlation coefficients defined in (17). Following (17), we distinguish two different index combinations for the third-order moment matrix.

Using the delta-method, we find the asymptotic covariances between the constant correlation estimates and the sample estimates for the third-order moment tensor matrix entries as the off-diagonal elements of the 2x2 matrix $\nabla_{g}^{T} \Sigma_{g} \nabla_{g}$ (cf. 34):

\[
\hat{\rho}_{ij}^{CC} = \frac{\bar{\tau}_{ij}}{2} \left[ \sqrt{\frac{m_{i}^{(4)}}{m_{i}^{(2)}}} \text{AsyCov} \left( \sqrt{T_{m_{i}^{(4)}}, \sqrt{T_{si_{j}}}} \right) + \sqrt{\frac{m_{j}^{(4)}}{m_{j}^{(2)}}} \text{AsyCov} \left( \sqrt{T_{m_{j}^{(4)}}, \sqrt{T_{sj_{i}}}} \right) \right]
\]

\[
\hat{\rho}_{ijk}^{CC} = \left( \frac{\bar{\tau}_{ij}}{4} \sqrt{\frac{m_{k}^{(2)}}{m_{i}^{(4)} m_{j}^{(4)}}} \right) \text{AsyCov} \left( \sqrt{T_{m_{i}^{(4)}}, \sqrt{T_{s_{j},k}}} \right)
\]

\[
+ \left( \frac{\bar{\tau}_{ij}}{4} \sqrt{\frac{m_{k}^{(2)}}{m_{j}^{(4)} m_{i}^{(4)}}} \right) \text{AsyCov} \left( \sqrt{T_{m_{j}^{(4)}}, \sqrt{T_{s_{i},k}}} \right)
\]

\[
+ \left( \frac{\bar{\tau}_{ij}}{2} \sqrt{\frac{m_{k}^{(2)}}{m_{i}^{(4)} m_{j}^{(4)}}} \right) \text{AsyCov} \left( \sqrt{T_{m_{k}^{(4)}}, \sqrt{T_{s_{i},j}}} \right)
\]

(40)
As in equation (39), consistent estimators for the above asymptotic covariance terms are given by:

\[
\text{AsyCov} \left( \sqrt{T} m_t^{(n)}, \sqrt{T} s_{ijkl} \right) = \frac{1}{T} \sum_{t=1}^{T} \left[ \left( R_{it} - m_t \right)^n - m_t^{(n)} \right] \left[ \left( R_{jt} - m_j \right) \left( R_{kt} - m_k \right) (R_{kt} - m_k) - s_{ijkl} \right]
\]

\[
\text{AsyCov} \left( \sqrt{T} m_t^{(n)}, \sqrt{T} s_{ijkl} \right) = \frac{1}{T} \sum_{t=1}^{T} \left[ \left( R_{it} - m_t \right)^n - m_t^{(n)} \right] \left[ \left( R_{jt} - m_j \right) \left( R_{kt} - m_k \right) (R_{kt} - m_k) - s_{ijkl} \right]
\] (41)

As far as the single factor approach is concerned, the comoment elements in \( \Psi \) and \( \Phi \) are by construction asymptotically uncorrelated with \( s_{ijk} \). On the other hand, the situation is less straightforward for the elements of \( \Upsilon \) [see Equations (24)-(26)]. As for the constant correlation approach, we explain how to proceed for the third moment matrix, and refer the reader to the Appendix for details concerning the fourth moment matrix. First, we can derive similar results to (38) for the asymptotic covariance between the single factor estimate of the third-order moment matrix and the corresponding sample estimate. We set:

\[
\theta_{3}^{SF} = \left( \begin{array}{cccc}
{s_{i0}} & {s_{j0}} & {s_{k0}} & {m_0^{(2)}} & {m_0^{(3)}} & {s_{ijkl}}
\end{array} \right)^T
\]

\[
g_{3}^{SF} (\theta_{3}^{SF}) = \left( \begin{array}{c}
{s_{i0}s_{j0}s_{k0}m_0^{(3)}} \\
{\left( m_0^{(2)} \right)^3 m_0^{(3)}} \\
{s_{ijkl}}
\end{array} \right) = \left( \begin{array}{c}
{\beta_i\beta_j\beta_km_0^{(3)}} \\
{s_{ijkl}}
\end{array} \right)
\] (42)

and obtain:

\[
\tilde{\rho}_{ijkl} = \frac{s_{j0} s_{k0} m_0^{(3)}}{\left( m_0^{(2)} \right)^3} \text{AsyCov} \left( \sqrt{T} s_{i0}, \sqrt{T} s_{ijkl} \right) + \frac{s_{i0} s_{k0} m_0^{(3)}}{\left( m_0^{(2)} \right)^3} \text{AsyCov} \left( \sqrt{T} s_{j0}, \sqrt{T} s_{ijkl} \right)
\]

\[
+ \frac{s_{i0} s_{j0} m_0^{(3)}}{\left( m_0^{(2)} \right)^3} \text{AsyCov} \left( \sqrt{T} s_{k0}, \sqrt{T} s_{ijkl} \right) - \frac{3s_{i0} s_{j0} s_{k0} m_0^{(3)}}{\left( m_0^{(2)} \right)^4} \text{AsyCov} \left( \sqrt{T} m_0^{(2)}, \sqrt{T} s_{ijkl} \right)
\]

\[
+ \frac{s_{i0} s_{j0} s_{k0} m_0^{(3)}}{\left( m_0^{(2)} \right)^3} \text{AsyCov} \left( \sqrt{T} m_0^{(3)}, \sqrt{T} s_{ijkl} \right).
\] (43)

All asymptotic covariances in (43) are, similarly to (39), consistently estimated by:

\[
\text{AsyCov} \left( \sqrt{T} s_{i0}, \sqrt{T} s_{ijkl} \right) = \frac{1}{T} \sum_{t=1}^{T} \left[ \left( R_{it} - m_t \right) \left( R_{i0} - m_0 \right) - s_{i0} \right] \left[ \left( R_{jt} - m_j \right) \left( R_{kt} - m_k \right) (R_{kt} - m_k) - s_{ijkl} \right]
\]

\[
\text{AsyCov} \left( \sqrt{T} \psi_{it}, \sqrt{T} s_{ijkl} \right) = \frac{1}{T} \sum_{t=1}^{T} \left[ \varepsilon_{it}^2 - \tilde{\psi}_{it} \right] \left[ \left( R_{it} - m_t \right) \left( R_{jt} - m_j \right) \left( R_{kt} - m_k \right) - s_{ijkl} \right]
\] (44)

where:

\[
s_{i0} = \frac{1}{T} \sum_{t=1}^{T} \left( R_{it} - m_t \right) \left( R_{i0} - m_0 \right)
\]

\[
\tilde{\psi}_{it} = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it}^2
\] (45)
As evidenced by (26), and explained in the Appendix, the situation is far more complex for the fourth-order moment matrix estimator, where we need to distinguish four cases according to the index permutations of \( \hat{\rho}_{ijkl} \).

Finally, as in (30), we sum up over all permutations and obtain the asymptotically optimal linear combination of the constant correlation (single factor) estimate and the sample estimate for \( M_2, M_3, \) and \( M_4 \). In terms of optimal shrinkage intensities (cf. (29)), we obtain:

\[
\begin{align*}
\hat{\alpha}^{CC}_2 &= \frac{1}{T} \frac{\hat{\pi}_2 - \hat{\rho}^{CC}_2}{\hat{\rho}^{CC}_2} \\
\hat{\alpha}^{SF}_2 &= \frac{1}{T} \frac{\hat{\pi}_2 - \hat{\rho}^{SF}_2}{\hat{\rho}^{SF}_2} \\
\hat{\alpha}^{CC}_3 &= \frac{1}{T} \frac{\hat{\pi}_3 - \hat{\rho}^{CC}_3}{\hat{\rho}^{CC}_3} \\
\hat{\alpha}^{SF}_3 &= \frac{1}{T} \frac{\hat{\pi}_3 - \hat{\rho}^{SF}_3}{\hat{\rho}^{SF}_3} \\
\hat{\alpha}^{CC}_4 &= \frac{1}{T} \frac{\hat{\pi}_4 - \hat{\rho}^{CC}_4}{\hat{\rho}^{CC}_4} \\
\hat{\alpha}^{SF}_4 &= \frac{1}{T} \frac{\hat{\pi}_4 - \hat{\rho}^{SF}_4}{\hat{\rho}^{SF}_4}
\end{align*}
\]

Consequently, we can define six 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.

3. Methodology

For our empirical testing of the out-of-sample performance of the improved estimators for higher-order moments and comoments, we mostly follow the empirical methodology adopted by Chan, Karceski, and Lakonishok (1999) and Jagannathan and Ma (2003).

Our methodology deviates, however, from the aforementioned papers on two important dimensions. First, instead of considering a single out-of-sample portfolio based on a randomly chosen basket of stocks, we extracted 100 randomly chosen baskets of stocks from the CRSP data-base. Secondly, for each portfolio, we hold constant the universe of stocks as opposed to re-sampling every year (Jagannathan and Ma, 2003). This procedure allows us to measure the stability of each of the constructed portfolios, a key indicator of the estimation quality of the underlying estimators that would not be available when the menu of underlying stocks changes through time. Also, and perhaps more importantly, our methodology enables us to perform an ex-post in-sample estimation analysis that can serve as a useful comparison benchmark. On the other hand, this selection constraint involves the introduction of some survivorship bias, the impact of which will be the subject of a dedicated robustness check analysis (see Section 4.1).

We collect monthly returns on common stocks listed on the NYSE and NYSE Amex Equities (formerly known as the AMEX) stock exchanges, with a sample period ranging from May 1968 through April 2006, and apply the same normalizations regarding penny stocks and market capitalization. In fact, we exclude small stocks from the sample, that is stocks with a market capitalization less than the 20% percentile of market capitalization, and penny stocks characterized by listed prices smaller than $5.\textsuperscript{6} We thus obtain valid return series for 255 stocks from the CRSP data-base. At the end of April each year, the various competing estimators for higher-order moment matrices \( M_2, M_3, \) and \( M_4 \) are obtained based on the previous 60 months of sample observations. As in Chan, Karceski, and Lakonishok (1999) or Jagannathan and Ma (2003), the optimized portfolios are then held throughout the subsequent 12 months until a new allocation decision takes place. As a result, we eventually obtain 33 years of monthly out-of-sample data.\textsuperscript{7}

The introduction of randomly selected baskets of stocks allows us to report the distribution of the various statistical indicators across the 100 portfolios. The goal here is to alleviate the concern over results that might be driven by a specific outcome of the stock selection process at a given point in time. Next, holding constant the sub-universe of stocks for each portfolio across time

---

\textsuperscript{6} - Since we only consider stocks that have valid returns over the whole sample period from May 1968 through April 2006, we rely on average market capitalizations and average prices in order to filter the stocks.

\textsuperscript{7} - 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 significantly differ across competing estimators. The situation would be different in a fixed-mix analysis.
allows us to measure the stability of each constructed portfolios, a key indicator of the estimation quality of the underlying estimators. Again, this would not be possible when the menu of assets changes through time.

Obviously, true moments and comoments parameter population values are unknown. However, we may proxy them with in-sample estimates obtained over the entire sample covering the period from May 1968 through April 2006, using the property that the sample estimator converges towards the true long-term parameter value for which it is a consistent estimator.\(^8\) As a first set of results, we thus focus on estimation errors for the competing estimators of higher-order moment matrices, using ex post in-sample estimates as imperfect proxies for the true parameter values. Our measure of estimation error is based on the Frobenius norm of the differences between an estimated comoment matrix and its corresponding benchmark. For each of the 100 portfolios and for each comoment matrix, we obtain 33 of such estimation error measures, one for each rebalancing step at the end of April each year starting in 1973. We then average these values across time for each portfolio and each estimator and report the distribution of this average estimation error in Table 2.

Table 2: Mean squared estimation error

<table>
<thead>
<tr>
<th>Panel A: Co-variance</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>0.0333</td>
<td>0.0056</td>
<td>0.0233</td>
<td>0.0260</td>
<td>0.0329</td>
<td>0.0438</td>
<td>0.0496</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>0.0304</td>
<td>0.0055</td>
<td>0.0206</td>
<td>0.0233</td>
<td>0.0296</td>
<td>0.0416</td>
<td>0.0462</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.0310</td>
<td>0.0052</td>
<td>0.0227</td>
<td>0.0241</td>
<td>0.0305</td>
<td>0.0412</td>
<td>0.0461</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.0296</td>
<td>0.0056</td>
<td>0.0201</td>
<td>0.0227</td>
<td>0.0287</td>
<td>0.0406</td>
<td>0.0457</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.0306</td>
<td>0.0054</td>
<td>0.0219</td>
<td>0.0238</td>
<td>0.0300</td>
<td>0.0410</td>
<td>0.0464</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Co-skewness</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>0.0171</td>
<td>0.0044</td>
<td>0.0096</td>
<td>0.0115</td>
<td>0.0165</td>
<td>0.0284</td>
<td>0.0309</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>0.0122</td>
<td>0.0044</td>
<td>0.0050</td>
<td>0.0073</td>
<td>0.0110</td>
<td>0.0237</td>
<td>0.0261</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.0108</td>
<td>0.0039</td>
<td>0.0058</td>
<td>0.0067</td>
<td>0.0099</td>
<td>0.0219</td>
<td>0.0229</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.0121</td>
<td>0.0044</td>
<td>0.0050</td>
<td>0.0073</td>
<td>0.0110</td>
<td>0.0237</td>
<td>0.0261</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.0108</td>
<td>0.0039</td>
<td>0.0058</td>
<td>0.0067</td>
<td>0.0099</td>
<td>0.0219</td>
<td>0.0229</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Co-kurtosis</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>0.0202</td>
<td>0.0074</td>
<td>0.0098</td>
<td>0.0125</td>
<td>0.0184</td>
<td>0.0369</td>
<td>0.0432</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>0.0190</td>
<td>0.0083</td>
<td>0.0077</td>
<td>0.0101</td>
<td>0.0161</td>
<td>0.0402</td>
<td>0.0439</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.0146</td>
<td>0.0060</td>
<td>0.0065</td>
<td>0.0092</td>
<td>0.0128</td>
<td>0.0316</td>
<td>0.0337</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.0177</td>
<td>0.0074</td>
<td>0.0075</td>
<td>0.0096</td>
<td>0.0150</td>
<td>0.0371</td>
<td>0.0400</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.0147</td>
<td>0.0061</td>
<td>0.0064</td>
<td>0.0092</td>
<td>0.0128</td>
<td>0.0315</td>
<td>0.0341</td>
</tr>
</tbody>
</table>

At the end of April of each year, higher-order moment tensor estimates based on the prior 60 months are compared to the estimated parameters using the full sample (May 1968 - April 2006). Frobenius norms of the matrix differences are calculated and averaged over time for each portfolio and each estimator. Descriptive statistics for the 100 average Frobenius norms are reported. Std denotes the standard deviation and Med the median. The period considered is May 1973 through April 2006.

First, we note that estimation errors concerning the variance-covariance matrix \(M_2\) are in line with previous findings in Chan, Karceski, and Lakonishok (1999). Indeed, the sample estimator yields the greatest deviations from the long-term in-sample parameter values and is clearly dominated by structured estimators. The reduction in estimation error amounts to roughly 9% for the constant correlation estimator and 7% for the single factor estimator, respectively. Secondly, we observe that shrinkage estimators lead to slightly higher gains in estimation quality, which is also consistent with findings from existing literature.

As far as higher-order moment tensors are concerned, we find that both structured estimators and their shrinkage counterparts also reduce estimation errors. A second result we obtain is that the single-factor approach substantially outperforms the constant correlation approach. Finally, our results strongly suggest that the benefits of structured and shrinkage estimators in terms of

\(^8\) - Note that the full sample estimate may be closer to the true parameter value but it is still only an estimate. Indeed, our reasoning abstracts away from structural breaks and time-variation in the true underlying parameters.
estimation quality are more pronounced in the case of coskewness estimates ($M_3$). On average, the constant correlation (single-factor) approach leads to reductions in estimating errors that reach as much as 29% (37%) for the third moment tensor, while cokurtosis estimates ($M_4$) display relative average benefits of 6% (28%).

Table 3: Optimal shrinkage intensities

<p>| Panel A: Shrinkage towards constant correlation |</p>
<table>
<thead>
<tr>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2</td>
<td>0.7644</td>
<td>0.0690</td>
<td>0.5465</td>
<td>0.6313</td>
<td>0.7695</td>
<td>0.8657</td>
</tr>
<tr>
<td>M3</td>
<td>0.9975</td>
<td>0.0055</td>
<td>0.9754</td>
<td>0.9825</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>M4</td>
<td>0.7931</td>
<td>0.0682</td>
<td>0.6342</td>
<td>0.6775</td>
<td>0.7992</td>
<td>0.9081</td>
</tr>
</tbody>
</table>

<p>| Panel B: Shrinkage towards single factor |</p>
<table>
<thead>
<tr>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2</td>
<td>0.6145</td>
<td>0.0574</td>
<td>0.4681</td>
<td>0.5194</td>
<td>0.6174</td>
<td>0.7127</td>
</tr>
<tr>
<td>M3</td>
<td>0.9755</td>
<td>0.0218</td>
<td>0.8777</td>
<td>0.9321</td>
<td>0.9820</td>
<td>0.9975</td>
</tr>
<tr>
<td>M4</td>
<td>0.8805</td>
<td>0.0525</td>
<td>0.6941</td>
<td>0.7749</td>
<td>0.8943</td>
<td>0.9483</td>
</tr>
</tbody>
</table>

At the end of April of each year, optimal shrinkage intensities are obtained based on the prior 60 months and averaged across time for each of the 100 portfolios. Descriptive statistics for the 100 average shrinkage intensities are reported. Std denotes the standard deviation and Med the median. The period considered is May 1973 through April 2006.

Table 3 reports average optimal shrinkage intensities for the various approaches. The results are highly consistent with the conclusions drawn from the estimation quality measures. In fact, due to the presence of an extremely high estimation risk for $M_3$, optimal shrinkage intensities tend to be very high (above 97% on average). On the other hand, shrinkage intensities for the fourth moment tensor are comparable to those of the covariance matrix and only slightly higher for the single factor approach, which may be explained by its superior estimation quality reported in Table 2. We now turn to the portfolio construction analysis. It has long been recognized that the sample mean is a poor estimator for the true population mean [e.g., 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 tangency portfolios that require expected return estimates [e.g., Brandt (2005, p.34)]. Hence, we follow Chan, Karceski, and Lakonishok (1999) and Jagannathan and Ma (2003), who focus on minimum variance portfolio, and generate global minimum risk portfolios, here represented by tangency portfolios in a 3-dimensional risk space. To do so, we have simply neutralized the impact of the expected return parameter by assuming $\mu=0$ (or some other constant value) for all assets. We find optimal allocations by assuming CRRA preferences in the objective function (10), which then can be written as:

$$
\min_w \left[ \frac{\gamma}{2} w'M_2 w - \frac{\gamma (\gamma + 1)}{6} w'M_3 (w \otimes w) + \frac{\gamma (\gamma + 1) (\gamma + 2)}{24} w'M_4 (w \otimes w \otimes w) \right],
$$

s.t.: $w' I_n = 1.$

(47)

The relative risk aversion coefficient $\gamma$, which defines the risk-trade-off between variance, skewness, and kurtosis, is taken equal to 10 in our base case, and we also try $\gamma=1$, $\gamma=5$, and $\gamma=15$ in subsequent comparative static analyses.

In order to analyze the out-of-sample performance of the constructed portfolios, we compare certainty equivalents for an investment in the various competing portfolios. We follow Ang and Bekaert (2002) and define the economic loss of investing in the portfolio built from sample estimates as the monetary payment the investor requires to be indifferent between sticking to this portfolio and switching to a portfolio based upon structured or shrinkage estimates. Accordingly, we solve the following equation for $\bar{W}$:
where index $i$ signals the use of a structured or a shrinkage estimator approach and index 0 the use of the sample estimator approach, respectively. The annual required payment [monetary utility gain (MUG)] is then given as $MUG = \sqrt{1 - 12 \cdot \bar{W}}$. Note that this measure is closely related to the difference in certainty equivalents used in Kandel and Stambaugh (1996), Campbell and Viceira (1999), and Hong, Tu, and Zhou (2007).\footnote{In fact, our MUGs are normalized certainty equivalent returns (CER): $\text{MUG} = \frac{\text{CER}}{1 + r(0)}$, with the scale factor $1 + r(0)$ close to 1, which rationalizes the quantitative comparison of the results.}

Certainty equivalents and monetary utility gains are reminiscent of the realized standard deviation comparison reported in Chan, Karceski, and Lakonishok (1999), Ledoit and Wolf (2003), and Jagannathan and Ma (2003), where preferences for higher-order moments are not taken into account.

4. Empirical Results

In the base case, we take the risk aversion parameter to be equal to $\gamma = 10$ and we assume a 60-month rolling calibration period. Table 4 reports average annualized MUG for different portfolio sizes.

<table>
<thead>
<tr>
<th>Panel A: N = 10</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const. Correlation</td>
<td>1.06</td>
<td>1.73</td>
<td>-7.03</td>
<td>-1.25</td>
<td>0.95</td>
<td>3.81</td>
<td>6.40</td>
</tr>
<tr>
<td>Single Factor</td>
<td>1.78</td>
<td>1.49</td>
<td>-0.89</td>
<td>-0.19</td>
<td>1.60</td>
<td>4.95</td>
<td>7.56</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>1.34</td>
<td>1.31</td>
<td>-1.37</td>
<td>-0.52</td>
<td>1.14</td>
<td>3.74</td>
<td>6.30</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>1.66</td>
<td>1.22</td>
<td>-0.27</td>
<td>0.10</td>
<td>1.45</td>
<td>4.35</td>
<td>7.17</td>
</tr>
<tr>
<td>Ex post optimal</td>
<td>3.96</td>
<td>2.16</td>
<td>-0.33</td>
<td>0.65</td>
<td>3.61</td>
<td>8.77</td>
<td>12.68</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: N=20</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const. Correlation</td>
<td>2.86</td>
<td>3.09</td>
<td>-3.24</td>
<td>-1.99</td>
<td>2.68</td>
<td>7.63</td>
<td>15.91</td>
</tr>
<tr>
<td>Single Factor</td>
<td>4.32</td>
<td>2.90</td>
<td>-1.08</td>
<td>0.33</td>
<td>3.75</td>
<td>9.50</td>
<td>17.09</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>3.56</td>
<td>2.81</td>
<td>-2.27</td>
<td>-0.55</td>
<td>3.05</td>
<td>7.97</td>
<td>16.67</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>4.41</td>
<td>2.57</td>
<td>-0.28</td>
<td>1.11</td>
<td>4.00</td>
<td>9.05</td>
<td>17.35</td>
</tr>
<tr>
<td>Ex post optimal</td>
<td>7.44</td>
<td>3.39</td>
<td>0.39</td>
<td>2.65</td>
<td>7.15</td>
<td>13.11</td>
<td>22.29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: N=30</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const. Correlation</td>
<td>5.35</td>
<td>4.13</td>
<td>-4.11</td>
<td>-1.04</td>
<td>4.98</td>
<td>10.77</td>
<td>21.75</td>
</tr>
<tr>
<td>Single Factor</td>
<td>7.75</td>
<td>3.92</td>
<td>0.01</td>
<td>2.84</td>
<td>7.06</td>
<td>13.04</td>
<td>25.49</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>6.58</td>
<td>3.77</td>
<td>-2.25</td>
<td>1.33</td>
<td>5.94</td>
<td>12.34</td>
<td>21.86</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>8.03</td>
<td>3.50</td>
<td>1.27</td>
<td>3.66</td>
<td>7.33</td>
<td>12.90</td>
<td>24.05</td>
</tr>
<tr>
<td>Ex post optimal</td>
<td>11.56</td>
<td>4.08</td>
<td>1.14</td>
<td>6.58</td>
<td>11.05</td>
<td>17.78</td>
<td>28.01</td>
</tr>
</tbody>
</table>

At the end of April of each year, optimal portfolio weights are obtained for a CRRA investor with risk aversion parameter $\gamma = 10$ according to higher-order moment tensor estimates based on the prior 60 months. These weights are then applied to form portfolios that are held until the next estimation (end of April of next year). The MUG is defined as the annual payment that an investor requires in order to be indifferent between a portfolio based upon the sample estimator and a portfolio based upon the corresponding structured or a shrinkage estimator. Descriptive statistics of the 100 MUG are reported. Std denotes the standard deviation and Med the median. Each portfolio consists of eligible stocks listed on the NYSE and the NYSE Amex Equities stock exchanges. Different portfolios sizes (10, 20, and 30) are considered. The historical data is from May 1973 through April 2006.

The results strongly suggest that portfolio construction incorporating higher-order moments generates substantial differences in out-of-sample returns depending on the chosen estimator. Indeed, for instance for $N = 20$, the estimated monetary utility gain for a CRRA investor ranges from 2.86% to 4.32% per annum when using structured instead of sample estimators. The use of shrinkage estimators leads to a slight increase in the monetary utility gains reaching from 3.56% to 4.41%. These numbers are highly economically significant. For instance, in a comparable study, and for same levels of risk aversion and investment horizon, Ang and Bekaert (2002) report economic losses between 0.12% and 1.26% when the regimeswitching nature of parameters is ignored.

The data in Table 4 further indicate that the benefits from building portfolios based on structured estimators increase with the number of assets in the portfolio. This is consistent with the argument.

\[ \frac{1}{T} \sum_{t=1}^{T} \left( 1 + \frac{r(t)}{1 - \gamma} \right)^{1-\gamma} = \frac{1}{T} \sum_{t=1}^{T} \left[ \frac{\bar{W}(1 + r(t))}{1 - \gamma} \right]^{1-\gamma}, \]
that estimation risk increases with the number of parameters to estimate. Moreover, the results suggest that economic significance as measured by monetary utility increases proportionally to the number of parameters to estimate, which is more than proportionally to portfolio size. While the MUG ranges from 1.06% to 1.66% for \( N = 10 \), it reaches values between 2.86% and 4.41% per year in the base case setting with \( N = 20 \) stocks per portfolio or 5.35% to 8.03% for \( N = 30 \). Interestingly, as the portfolio size increases, structured and shrinkage estimators come closer to benchmark maximal monetary utility gains obtained from ex-post in-sample estimates. This can be regarded as a direct consequence of the linear increase in the number of parameters as compared to the exponential increase for sample estimators (see Table 1).

Table 5 displays several indicators related to the obtained optimal weights for the base case portfolio size \( N = 20 \). All statistics refer to average numbers across time (the weights are re-balanced at the end of April of each year). As evidenced by the results reported in Panels A and B, portfolios based on the sample estimator generate a significantly higher amount of short positions and lead to fewer positively weighted assets compared to portfolios constructed using structured or shrinkage estimators. On average, the short interest induced by sample estimators is as high as 54.62%, while the average short interest for structured and shrinkage estimators ranges between 18.18% and 21.86%.

### Table 5: Weight analysis

<table>
<thead>
<tr>
<th>Panel A: Short interest</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>0.5462</td>
<td>0.0585</td>
<td>0.4293</td>
<td>0.4465</td>
<td>0.5380</td>
<td>0.6410</td>
<td>0.7152</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>0.2160</td>
<td>0.0249</td>
<td>0.1539</td>
<td>0.1797</td>
<td>0.2137</td>
<td>0.2624</td>
<td>0.2870</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.2041</td>
<td>0.0259</td>
<td>0.1613</td>
<td>0.1699</td>
<td>0.2003</td>
<td>0.2625</td>
<td>0.2745</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.2029</td>
<td>0.0288</td>
<td>0.1481</td>
<td>0.1621</td>
<td>0.1984</td>
<td>0.2572</td>
<td>0.3045</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.2186</td>
<td>0.0304</td>
<td>0.1606</td>
<td>0.1775</td>
<td>0.2157</td>
<td>0.2786</td>
<td>0.2955</td>
</tr>
<tr>
<td>Ex post optimal</td>
<td>0.1818</td>
<td>0.0542</td>
<td>0.0736</td>
<td>0.0981</td>
<td>0.1721</td>
<td>0.2685</td>
<td>0.3228</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Positive weights</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>0.6182</td>
<td>0.0181</td>
<td>0.5621</td>
<td>0.5848</td>
<td>0.6205</td>
<td>0.6508</td>
<td>0.6606</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>0.6058</td>
<td>0.0334</td>
<td>0.5197</td>
<td>0.5477</td>
<td>0.6061</td>
<td>0.6652</td>
<td>0.6864</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.6642</td>
<td>0.0253</td>
<td>0.6015</td>
<td>0.6212</td>
<td>0.6621</td>
<td>0.7045</td>
<td>0.7212</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.6331</td>
<td>0.0285</td>
<td>0.5621</td>
<td>0.5818</td>
<td>0.6303</td>
<td>0.6765</td>
<td>0.7015</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.6847</td>
<td>0.0232</td>
<td>0.6167</td>
<td>0.6439</td>
<td>0.6871</td>
<td>0.7152</td>
<td>0.7348</td>
</tr>
<tr>
<td>Ex post optimal</td>
<td>0.7045</td>
<td>0.0682</td>
<td>0.5000</td>
<td>0.6000</td>
<td>0.7000</td>
<td>0.8000</td>
<td>0.8500</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Min weight</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>-0.1727</td>
<td>0.0232</td>
<td>-0.2463</td>
<td>-0.2155</td>
<td>-0.1700</td>
<td>-0.1397</td>
<td>-0.1276</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>-0.0448</td>
<td>0.0052</td>
<td>-0.0557</td>
<td>-0.0540</td>
<td>-0.0442</td>
<td>-0.0360</td>
<td>-0.0334</td>
</tr>
<tr>
<td>Single Factor</td>
<td>-0.0659</td>
<td>0.0103</td>
<td>-0.0949</td>
<td>-0.0856</td>
<td>-0.0653</td>
<td>-0.0514</td>
<td>0.0485</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>-0.0483</td>
<td>0.0063</td>
<td>-0.0678</td>
<td>-0.0613</td>
<td>-0.0469</td>
<td>-0.0402</td>
<td>-0.0368</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>-0.0757</td>
<td>0.0120</td>
<td>-0.1151</td>
<td>-0.0965</td>
<td>-0.0750</td>
<td>-0.0584</td>
<td>-0.0546</td>
</tr>
<tr>
<td>Ex-post optimal</td>
<td>-0.0689</td>
<td>0.0241</td>
<td>-0.1439</td>
<td>-0.1212</td>
<td>-0.0661</td>
<td>-0.0382</td>
<td>-0.0241</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel D: Max weight</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>0.4033</td>
<td>0.0385</td>
<td>0.3397</td>
<td>0.3513</td>
<td>0.3971</td>
<td>0.4796</td>
<td>0.5040</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>0.3252</td>
<td>0.0439</td>
<td>0.2309</td>
<td>0.2648</td>
<td>0.3215</td>
<td>0.4042</td>
<td>0.5277</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.2845</td>
<td>0.0388</td>
<td>0.2131</td>
<td>0.2239</td>
<td>0.2833</td>
<td>0.3609</td>
<td>0.4689</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.3245</td>
<td>0.0428</td>
<td>0.2415</td>
<td>0.2639</td>
<td>0.320</td>
<td>0.4021</td>
<td>0.5145</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.2958</td>
<td>0.0353</td>
<td>0.2298</td>
<td>0.2448</td>
<td>0.2939</td>
<td>0.3633</td>
<td>0.4632</td>
</tr>
<tr>
<td>Ex post optimal</td>
<td>0.2797</td>
<td>0.0690</td>
<td>0.1357</td>
<td>0.1834</td>
<td>0.2743</td>
<td>0.3748</td>
<td>0.5571</td>
</tr>
</tbody>
</table>

At the end of April of each year, optimal portfolio weights are obtained for a CRRA investor with risk aversion parameter \( \gamma = 10 \) according to higher-order moment tensor estimates based on the prior 60 months. Descriptive statistics of the portfolio weights across the 100 portfolios are reported. \( \text{Std} \) denotes the standard deviation and \( \text{Med} \) the median. Each portfolio contains 20 assets drawn from eligible stocks listed on the NYSE and the NYSEAmex Equities stock exchanges. The period considered is May 1973 through April 2006.

The use of sample estimators also leads to more extreme allocations, as is evidenced by the numbers in Panels C and D in Table 5. Indeed, differences between minimum and maximum weights are more pronounced for portfolios based on sample estimators compared to portfolios based on structured or
shrinkage estimators. Panel A in Table 6 reports the distribution of average turnovers for the competing estimators, again for portfolios of size $N = 20$. The results confirm the intuition that portfolios using sample estimates exhibit a higher deviation in the weight vector compared to portfolios using the competing estimators. On average, sample estimate portfolios require between 27% and 38% additional turnover compared to portfolios based on structured and shrinkage estimator portfolios.

Table 6: Turnover and optimal allocation

<table>
<thead>
<tr>
<th>Panel A: Realized turnover</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>2.00</td>
<td>0.12</td>
<td>1.75</td>
<td>1.80</td>
<td>2.00</td>
<td>2.20</td>
<td>2.38</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>1.47</td>
<td>0.09</td>
<td>1.24</td>
<td>1.33</td>
<td>1.47</td>
<td>1.60</td>
<td>1.70</td>
</tr>
<tr>
<td>Single Factor</td>
<td>1.37</td>
<td>0.07</td>
<td>1.20</td>
<td>1.26</td>
<td>1.37</td>
<td>1.49</td>
<td>1.55</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>1.41</td>
<td>0.09</td>
<td>1.22</td>
<td>1.27</td>
<td>1.41</td>
<td>1.56</td>
<td>1.72</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>1.35</td>
<td>0.08</td>
<td>1.18</td>
<td>1.22</td>
<td>1.35</td>
<td>1.51</td>
<td>1.57</td>
</tr>
<tr>
<td>Ex post optimal</td>
<td>1.25</td>
<td>0.15</td>
<td>0.83</td>
<td>1.00</td>
<td>1.26</td>
<td>1.48</td>
<td>1.61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Required turnover</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>1.75</td>
<td>0.16</td>
<td>1.42</td>
<td>1.52</td>
<td>1.74</td>
<td>2.03</td>
<td>2.12</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>1.14</td>
<td>0.10</td>
<td>0.95</td>
<td>0.98</td>
<td>1.13</td>
<td>1.33</td>
<td>1.41</td>
</tr>
<tr>
<td>Single Factor</td>
<td>1.09</td>
<td>0.10</td>
<td>0.91</td>
<td>0.95</td>
<td>1.09</td>
<td>1.28</td>
<td>1.30</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>1.10</td>
<td>0.09</td>
<td>0.94</td>
<td>0.96</td>
<td>1.10</td>
<td>1.25</td>
<td>1.38</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>1.11</td>
<td>0.09</td>
<td>0.91</td>
<td>0.99</td>
<td>1.10</td>
<td>1.24</td>
<td>1.33</td>
</tr>
</tbody>
</table>

At the end of April of each year, optimal portfolio weights are obtained for a CRRA investor with risk aversion parameter $\gamma = 10$ according to higher-order moment tensor estimates based on the prior 60 months. Each portfolio contains 20 assets drawn from eligible stocks listed on the NYSE and the NYSE Amex Equities stock exchanges. Realized turnover is the across-time average of the sum of absolute deviations induced by a re-balancing at the end of April of each year. Required turnover refers to the turnover that would be necessary in order to switch to the optimal portfolio determined by the estimations of the higher-order moment tensors using the full sample. The realized turnover measure is a clear indicator for the ex-post quality of the allocation decision. Differences among structured and shrinkage estimators are relatively small, suggesting that imposing some structure helps regardless of which particular kind of structure is imposed, a point emphasized in Jagannathan and Ma (2003). It should be noted at this stage that turnover can be induced not only by estimation error, in which case it is not desirable, but also by time-varying return distributions that would naturally result in volatile optimal portfolio weights. Disentangling the two effects is not straightforward, however, and we simply assume here that at least a part of the turnover rates does reflect estimation error.10

Panel B in Table 6 reports the average annual turnover induced by transactions theoretically needed to switch to the optimal portfolio determined by the full sample ex-post estimates for higher-order moment tensors. From these results, we confirm that portfolios employing structured or shrinkage estimators are closer to the optimal portfolio than those using sample estimators. This required turnover can be seen as an (in)efficiency measure in the 3-dimensional risk space of variance, skewness, and kurtosis.

4.1 Robustness checks

In this Section, we attempt to assess the sensitivity of the above results with respect to various changes in the experimental setting and/or parameter values. In Table 7, we analyze the impact of changes in the risk-aversion level. We find that increasing risk-aversion leads to increased benefits from using improved estimators of higher-order moment matrices, which is consistent with the fact that more weight is given to higher-order moments in the expected utility approximation as risk-aversion increases.

---

10 - It should be noted that the ex post optimal portfolio that has been constructed upon full sample estimates lead to the largest monetary utility gain and the lowest turnover rates. The turnover rate for the ex post optimal portfolio accounts for the performance effect during the holding period and is due to the buy-and-hold investment strategy, as opposed to a fixed mix strategy.
Table 7: MUGS: Risk aversion

<table>
<thead>
<tr>
<th>Panel A: $\gamma = 1$</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>0.23</td>
<td>1.55</td>
<td>−2.94</td>
<td>−2.27</td>
<td>0.13</td>
<td>2.99</td>
<td>3.98</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>1.31</td>
<td>1.48</td>
<td>−1.74</td>
<td>−1.24</td>
<td>1.26</td>
<td>4.35</td>
<td>5.50</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.28</td>
<td>1.41</td>
<td>−2.69</td>
<td>−2.00</td>
<td>0.11</td>
<td>2.86</td>
<td>3.46</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.91</td>
<td>1.14</td>
<td>−1.59</td>
<td>−0.88</td>
<td>0.84</td>
<td>2.85</td>
<td>4.03</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>2.08</td>
<td>1.74</td>
<td>−2.91</td>
<td>−0.89</td>
<td>2.09</td>
<td>4.95</td>
<td>5.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: $\gamma = 5$</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const. Correlation</td>
<td>1.27</td>
<td>1.92</td>
<td>−2.74</td>
<td>−2.03</td>
<td>1.20</td>
<td>4.63</td>
<td>5.58</td>
</tr>
<tr>
<td>Single Factor</td>
<td>2.49</td>
<td>1.80</td>
<td>−1.16</td>
<td>−0.56</td>
<td>2.27</td>
<td>5.72</td>
<td>7.10</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>1.60</td>
<td>1.73</td>
<td>−2.27</td>
<td>−1.26</td>
<td>1.66</td>
<td>4.72</td>
<td>5.74</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>2.31</td>
<td>1.46</td>
<td>−0.86</td>
<td>−0.09</td>
<td>2.18</td>
<td>4.87</td>
<td>5.98</td>
</tr>
<tr>
<td>Ex post optimal</td>
<td>4.22</td>
<td>2.13</td>
<td>−0.57</td>
<td>0.63</td>
<td>4.12</td>
<td>7.95</td>
<td>9.69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: $\gamma = 10$</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const. Correlation</td>
<td>2.86</td>
<td>3.09</td>
<td>−3.24</td>
<td>−1.99</td>
<td>2.68</td>
<td>7.63</td>
<td>15.91</td>
</tr>
<tr>
<td>Single Factor</td>
<td>4.32</td>
<td>2.90</td>
<td>−1.08</td>
<td>0.33</td>
<td>3.75</td>
<td>9.50</td>
<td>17.09</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>3.56</td>
<td>2.81</td>
<td>−2.27</td>
<td>−0.55</td>
<td>3.05</td>
<td>7.97</td>
<td>16.67</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>4.41</td>
<td>2.57</td>
<td>−0.28</td>
<td>1.11</td>
<td>4.00</td>
<td>9.05</td>
<td>17.35</td>
</tr>
<tr>
<td>Ex post optimal</td>
<td>7.44</td>
<td>3.39</td>
<td>0.39</td>
<td>2.66</td>
<td>7.15</td>
<td>13.11</td>
<td>22.29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel D: $\gamma = 15$</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const. Correlation</td>
<td>5.29</td>
<td>7.74</td>
<td>−5.34</td>
<td>−2.55</td>
<td>4.14</td>
<td>14.56</td>
<td>64.16</td>
</tr>
<tr>
<td>Single Factor</td>
<td>7.24</td>
<td>7.56</td>
<td>−1.90</td>
<td>1.03</td>
<td>5.62</td>
<td>16.44</td>
<td>66.13</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>6.45</td>
<td>7.49</td>
<td>−2.83</td>
<td>−0.16</td>
<td>4.92</td>
<td>14.92</td>
<td>65.43</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>7.62</td>
<td>7.32</td>
<td>−0.63</td>
<td>1.91</td>
<td>6.04</td>
<td>15.67</td>
<td>66.62</td>
</tr>
<tr>
<td>Ex post optimal</td>
<td>12.19</td>
<td>8.48</td>
<td>1.19</td>
<td>4.94</td>
<td>10.50</td>
<td>24.38</td>
<td>76.11</td>
</tr>
</tbody>
</table>

Methodology is as in the base case (see Table 4) and the case of N=20 assets per portfolio is considered for different risk aversions than in the base case ($\gamma=10$).

Table 8 analyzes the impact of shorter sample window sizes and the introduction of transaction costs on monetary utility gains. As far as the window size is concerned (Panel A), the results confirm the intuition that smaller estimation windows lead to increases in estimation risk, which makes the use of structured and shrinkage estimators even more attractive. The magnitude of the window size effect is remarkable: average annual MUG for 36-month rolling window estimations range between 9.77% and 11.34%. This window size effect is consistent with the cutoff point analysis in Jagannathan and Ma (2003), who estimate the window size or number of observations from which structured estimators are out-performed by the corresponding sample estimators.

Table 8: MUGs: Robustness checks

<table>
<thead>
<tr>
<th>Panel A: 36 months in-sample</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const. Correlation</td>
<td>9.77</td>
<td>6.45</td>
<td>−1.54</td>
<td>0.16</td>
<td>8.98</td>
<td>20.06</td>
<td>45.83</td>
</tr>
<tr>
<td>Single Factor</td>
<td>11.72</td>
<td>6.57</td>
<td>−0.36</td>
<td>2.40</td>
<td>10.82</td>
<td>22.55</td>
<td>48.53</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>10.39</td>
<td>6.37</td>
<td>−1.53</td>
<td>0.48</td>
<td>9.63</td>
<td>20.34</td>
<td>45.27</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>11.34</td>
<td>6.36</td>
<td>−0.93</td>
<td>2.16</td>
<td>10.34</td>
<td>21.57</td>
<td>46.47</td>
</tr>
<tr>
<td>Ex post optimal</td>
<td>14.74</td>
<td>6.76</td>
<td>1.90</td>
<td>4.85</td>
<td>13.59</td>
<td>26.73</td>
<td>48.98</td>
</tr>
<tr>
<td>Panel B: TC=100 bps</td>
<td>Mean</td>
<td>Std</td>
<td>Min</td>
<td>5%</td>
<td>Med</td>
<td>95%</td>
<td>Max</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>3.43</td>
<td>3.09</td>
<td>−2.53</td>
<td>−1.37</td>
<td>3.24</td>
<td>8.12</td>
<td>16.40</td>
</tr>
<tr>
<td>Single Factor</td>
<td>5.02</td>
<td>2.90</td>
<td>−0.32</td>
<td>1.01</td>
<td>4.57</td>
<td>10.08</td>
<td>17.71</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>4.22</td>
<td>2.82</td>
<td>−1.94</td>
<td>0.23</td>
<td>3.92</td>
<td>8.52</td>
<td>17.33</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>5.16</td>
<td>2.58</td>
<td>0.18</td>
<td>1.80</td>
<td>4.77</td>
<td>9.71</td>
<td>18.12</td>
</tr>
<tr>
<td>Ex post optimal</td>
<td>8.32</td>
<td>3.44</td>
<td>1.32</td>
<td>3.43</td>
<td>7.95</td>
<td>14.12</td>
<td>23.50</td>
</tr>
</tbody>
</table>

Methodology is as in the base case (see Table 4) and several independent robustness tests are conducted with respect to the base case. While Panel A considers a shorter in-sample period of 36 months, Panel B introduces transaction costs of 100 basis points per round-trip. Each portfolio contains 20 assets drawn from eligible stocks listed on the NYSE and the NYSE Amex Equities stock exchanges and CARA utility with risk aversion parameter $\gamma=10$ is applied to obtain optimal portfolios.
We also relax the hypothesis of no transaction costs and analyze the impact of such frictions on the final wealth across competing estimators (Panel B of Table 8). We assume a total cost (including commissions, quoted and effective spreads) of 100 basis points per transaction, a value that is consistent with the findings in Battalio, Ellul, and Jennings (2007) and the parameters used in Jang, Koo, Liu, and Loewenstein (2007). The results suggest that, as expected, structured and shrinkage estimators benefit from the introduction of transaction costs when compared to sample estimators. This is directly related to the observed gain in allocation stability and the fact that improved estimators lead to less extreme allocations. Taking transaction costs into account enhances the utility gains for the constant correlation (single factor) estimator by 57 (70) basis points for structured estimators and 66 (75) basis points for shrinkage estimators.

We next turn to an analysis of the potential impact of survivorship bias induced by the restricted selection of stocks available over the entire sample period. In fact, while previous literature has widely documented that the survivorship bias has a significant impact on expected returns estimates [e.g., Brown, Goetzmann, Ibbotson, and Ross (1992)], little is known on the impact, if any, of survivorship bias on higher-order risk parameter estimates. To address this question, we relax the afore-mentioned constraint and follow the exact methodology employed in Jagannathan and Ma (2003). In other words, we re-balance the basket of stocks every year, which leaves us with a number of eligible assets ranging from 1,000 to 1,200 for the various dates. To be consistent with the previous results, however, we still look at 100 different baskets for each year. Table 9 reports the distribution of shrinkage intensities and monetary utility gains obtained across the 100 portfolios.

Table 9: Impact of survivorship bias

<table>
<thead>
<tr>
<th>Panel A: MUGs</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const. Correlation</td>
<td>2.63</td>
<td>2.55</td>
<td>−3.32</td>
<td>−1.17</td>
<td>2.46</td>
<td>6.50</td>
<td>11.56</td>
</tr>
<tr>
<td>Single Factor</td>
<td>4.62</td>
<td>2.14</td>
<td>1.02</td>
<td>1.61</td>
<td>4.61</td>
<td>7.65</td>
<td>14.50</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>2.99</td>
<td>2.38</td>
<td>−2.26</td>
<td>−0.52</td>
<td>2.85</td>
<td>6.24</td>
<td>12.27</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>5.03</td>
<td>2.38</td>
<td>0.95</td>
<td>1.23</td>
<td>4.88</td>
<td>8.63</td>
<td>15.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Shrinkage Intensities (CC)</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2</td>
<td>0.8295</td>
<td>0.0251</td>
<td>0.7667</td>
<td>0.7911</td>
<td>0.8284</td>
<td>0.8726</td>
<td>0.8927</td>
</tr>
<tr>
<td>M3</td>
<td>0.9965</td>
<td>0.0042</td>
<td>0.9788</td>
<td>0.9882</td>
<td>0.9982</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>M4</td>
<td>0.8073</td>
<td>0.0337</td>
<td>0.7310</td>
<td>0.7534</td>
<td>0.8063</td>
<td>0.8603</td>
<td>0.8776</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Shrinkage Intensities (SF)</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2</td>
<td>0.6931</td>
<td>0.0231</td>
<td>0.6313</td>
<td>0.6532</td>
<td>0.6934</td>
<td>0.7277</td>
<td>0.7629</td>
</tr>
<tr>
<td>M3</td>
<td>0.9777</td>
<td>0.0113</td>
<td>0.9452</td>
<td>0.9563</td>
<td>0.9786</td>
<td>0.9940</td>
<td>0.9972</td>
</tr>
<tr>
<td>M4</td>
<td>0.9038</td>
<td>0.0231</td>
<td>0.8264</td>
<td>0.8594</td>
<td>0.9064</td>
<td>0.9383</td>
<td>0.9502</td>
</tr>
</tbody>
</table>

The table presents monetary utility gains (Panel A) and shrinkage intensities (Panels B and C) when relaxing the constraint that only assets that have valid returns over the whole sample period (May 1968 through April 2006) are eligible for inclusion in the 100 baskets. In other words, we follow the methodology in Jagannathan and Ma (2003) and redraw the baskets of assets every year from the set of assets that have valid returns for the past 60 and the subsequent 12 months. The rest of the methodology is as in the base case (see Table 4).

Panel A of Table 9 suggests that monetary utility gains do not substantially differ from the base case methodology, where the asset universe in each basket remains constant over the whole sample period. In fact, MUGs slightly increase for the portfolios based on the single factor estimator, while they slightly decrease for portfolios based on the constant correlation estimator. On the other hand, we find that shrinkage intensities increase with respect to the base case results. This last finding might indicate that estimation risk increases when the asset universe is unrestricted. Overall, however, the observed differences with respect to our base case methodology are not substantial.
4.2 Comparison with mean-variance analysis

The previous analysis has unambiguously shown that an investor who focuses on maximizing expected utility of terminal wealth based on a fourth order approximation of the utility function would benefit from employing improved estimates, as opposed to naïve sample estimates, for higher-order comoments of asset returns. On the other hand, an outstanding question that remains unanswered by the previous analysis is whether using a fourth-order approximation of the utility function always leads to a better portfolio compared to using a second-order approximation. In fact, while it must be the case that using a fourth-order approximation of the utility function generates better results compared to using a second-order approximation from an ex post perspective (i.e., after removing all uncertainty regarding parameter estimates), it is not necessarily the case that the same holds true ex ante, especially for small sample sizes, because of the presence of a significantly greater estimation risk in the former situation compared to the latter situation.

To analyze this question, we have estimated the monetary utility gains (MUGs) (or monetary utility losses when negative) implied by moving away from mean-variance analysis based on naïve sample estimates, and using instead a fourth-order approximation (based either on sample or improved estimates). These results can be found in Panel A of Table 10.

Table 10: MUGs: Added value of improved higher-order comoment estimates

<table>
<thead>
<tr>
<th>Panel A: Versus the sample minimum variance portfolio</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>-0.23</td>
<td>0.52</td>
<td>-3.11</td>
<td>-0.99</td>
<td>-0.16</td>
<td>0.39</td>
<td>0.88</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>2.61</td>
<td>2.90</td>
<td>-3.31</td>
<td>-1.86</td>
<td>2.38</td>
<td>7.46</td>
<td>12.20</td>
</tr>
<tr>
<td>Single Factor</td>
<td>4.07</td>
<td>2.68</td>
<td>-1.16</td>
<td>0.14</td>
<td>3.70</td>
<td>8.72</td>
<td>13.33</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>3.31</td>
<td>2.62</td>
<td>-2.70</td>
<td>-0.77</td>
<td>2.98</td>
<td>7.70</td>
<td>12.92</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>4.16</td>
<td>2.36</td>
<td>-0.64</td>
<td>0.82</td>
<td>3.68</td>
<td>8.30</td>
<td>13.58</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Marginal contribution of 3rd and 4th order structured tensors</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const. Correlation</td>
<td>0.26</td>
<td>0.73</td>
<td>-2.29</td>
<td>-0.86</td>
<td>0.31</td>
<td>1.20</td>
<td>2.40</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.09</td>
<td>0.69</td>
<td>-4.14</td>
<td>-0.66</td>
<td>0.14</td>
<td>0.91</td>
<td>1.90</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.21</td>
<td>0.60</td>
<td>-3.51</td>
<td>-0.53</td>
<td>0.28</td>
<td>1.05</td>
<td>1.63</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.07</td>
<td>0.59</td>
<td>-4.42</td>
<td>-0.37</td>
<td>0.16</td>
<td>0.76</td>
<td>0.95</td>
</tr>
</tbody>
</table>

The table assesses the marginal MUGs of using improved estimates for higher-order comoment parameters. Panel A presents MUGs with respect to the sample minimum variance portfolios. Panel B presents marginal MUGs with respect to a situation where structure is imposed solely on $M_2$ and sample estimates are used for both $M_3$ and $M_4$. The rest of the methodology is as in the base case (see Table 4).

These results first confirm that mean-variance analysis in fact dominates portfolio selection with higher-order moments if one uses naïve sample estimates. This can be seen in the first line of Panel A of Table 10, which indicates that implementing a fourth-order approximation but relying on sample estimates leads to an average MUG of -0.23% as compared to using mean-variance analysis also based on sample estimates. Overall, this result confirms that one should prefer using mean-variance analysis even in the presence of non-Gaussian returns if one is not prepared to use improved risk parameter estimates. On the other hand, the results in Panel A also suggest that using improved parameter estimates would allow an investor who uses a fourth-order approximation of the utility function to dominate a mean-variance investor using sample estimates of the covariance matrix, with average monetary utility gains ranging from 2.61% to 4.16%.

In this context, one would next like to know what is the marginal contribution of improving $M_3$ and $M_4$ with respect to solely imposing structure on $M_2$. In order to answer this question, we consider in Panel B of Table 10 a benchmark investor who uses the fourth-order approximation of the utility function, imposes structure on $M_2$ (constant correlation or single factor, respectively), but relies on naïve sample estimates for $M_3$ and $M_4$, and we present the MUGs that arise from additionally improving $M_3$ and $M_4$ parameter estimates as well. The results in Panel B show on
the one hand that the marginal contribution of improving coskewness and cokurtosis parameter estimates is positive on average, but on the other hand that these marginal improvements are relatively small compared to the overall benefits reported in Panel A.

This last result suggests that the benefits obtained from using improved estimators for higher-order moment matrices are small compared to the benefits obtained from improving the estimate of the covariance matrix only. It also raises the question whether the better performance of improved estimators with respect to sample estimates for third-order and fourth-order comoment matrices justifies in itself the use of a fourth-order approximation, or whether one would instead be better off focusing on minimum variance portfolios and solely benefiting from improved estimators of the covariance matrix. To answer this question, we present in Table 11 the annualized MUGs obtained from using a fourth-order approximation instead of using a second-order approximation (minimum variance criterion) of the expected utility function. Each row of Table 11 corresponds to a situation where all three moment tensors are estimated with the same estimation methodology (sample estimates or improved estimates) and we provide in each case an estimate of the monetary utility gain (or loss) resulting from using a fourth-order as opposed to a second-order approximation.

Table 11: MUGs: Added value of fourth-order approximation

<table>
<thead>
<tr>
<th>Panel A: Monthly returns</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>−0.23</td>
<td>0.52</td>
<td>−3.11</td>
<td>−0.99</td>
<td>−0.16</td>
<td>0.39</td>
<td>0.88</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>−0.12</td>
<td>0.12</td>
<td>−0.51</td>
<td>−0.37</td>
<td>−0.11</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>Single Factor</td>
<td>−0.09</td>
<td>0.08</td>
<td>−0.46</td>
<td>−0.26</td>
<td>−0.08</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>−0.09</td>
<td>0.14</td>
<td>−0.51</td>
<td>−0.34</td>
<td>−0.08</td>
<td>0.07</td>
<td>0.41</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>−0.07</td>
<td>0.08</td>
<td>−0.35</td>
<td>−0.22</td>
<td>−0.06</td>
<td>0.04</td>
<td>0.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Weekly returns</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>−0.18</td>
<td>0.91</td>
<td>−2.36</td>
<td>−1.73</td>
<td>−0.31</td>
<td>1.39</td>
<td>2.00</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>0.28</td>
<td>1.15</td>
<td>−1.80</td>
<td>−1.43</td>
<td>0.24</td>
<td>2.03</td>
<td>3.88</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.23</td>
<td>0.91</td>
<td>−1.60</td>
<td>−1.35</td>
<td>0.30</td>
<td>1.65</td>
<td>2.13</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.45</td>
<td>0.82</td>
<td>−1.41</td>
<td>−1.03</td>
<td>0.49</td>
<td>1.65</td>
<td>2.03</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.49</td>
<td>0.86</td>
<td>−1.22</td>
<td>−0.76</td>
<td>0.44</td>
<td>1.92</td>
<td>2.74</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Daily returns</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>5%</th>
<th>Med</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>0.50</td>
<td>0.97</td>
<td>−1.96</td>
<td>−1.11</td>
<td>0.43</td>
<td>2.28</td>
<td>2.95</td>
</tr>
<tr>
<td>Const. Correlation</td>
<td>0.90</td>
<td>1.00</td>
<td>−1.55</td>
<td>−0.83</td>
<td>0.80</td>
<td>2.62</td>
<td>3.13</td>
</tr>
<tr>
<td>Single Factor</td>
<td>0.83</td>
<td>1.05</td>
<td>−1.60</td>
<td>−0.86</td>
<td>0.83</td>
<td>2.54</td>
<td>3.49</td>
</tr>
<tr>
<td>Shrinkage to CC</td>
<td>0.81</td>
<td>0.94</td>
<td>−1.31</td>
<td>−0.94</td>
<td>0.87</td>
<td>2.33</td>
<td>2.76</td>
</tr>
<tr>
<td>Shrinkage to SF</td>
<td>0.91</td>
<td>0.86</td>
<td>−1.32</td>
<td>−0.54</td>
<td>0.89</td>
<td>2.32</td>
<td>2.90</td>
</tr>
</tbody>
</table>

The table presents MUGs from using a fourth-order approximation instead of using a second-order approximation (minimum variance criterion) of the expected utility function [see Equation (48)]. Each line corresponds to a situation where all 3 moment tensors are estimated with the same estimation methodology (sample estimates or improved estimates) and we provide in each case an estimate of the monetary utility gain (or loss) resulting from using a fourth-order as opposed to a second-order approximation. Different return frequencies are used and annualized MUGs are reported. The rest of the methodology is as in the base case (see Table 4).

As can be seen from these results, for monthly returns (Panel A of Table 11), all estimators lead to monetary utility losses when implementing the fourth-order approximation instead of the second-order (minimum variance) objective function. However, when using improved estimators, the underperformance (measured in terms of MUG) of the fourth-order approximation is smaller than when the sample estimator is used for both cases. Indeed, the monetary utility losses from using the fourth-order approximation instead of the second-order one range from 0.07% to 0.12% when using improved estimators (for both, the second- and the fourth-order approximations) while these losses amount to 0.23% in the case of sample estimators.

In an attempt to analyze the conditions under which a fourth-order approximation would dominate a second-order approximation, we have subsequently used different return frequencies
such as weekly and daily returns in addition to monthly returns. We find that the situation changes dramatically when increasing the data frequency. For daily returns (Panel C of Table 11) all estimators lead to superior portfolios when using the fourth-order approximation compared to the second-order approximation. On the other hand, for weekly returns (Panel B of Table 11), the fourth-order approximation interestingly deteriorates the investor’s welfare when sample estimates are used, but increases the investor’s welfare when structured or shrinkage estimators are used.

The interpretation for these results is relatively straightforward. Increasing sample size (through increasing the return frequency) favors the use of a fourth-order approximation as opposed to the second-order approximation. Indeed, the second-order approximation can be interpreted as an approach that uses highly structured estimators for M3 and M4, setting them to zero (or to trivial constants). This approach, which implies significant specification error in case asset returns are not normally distributed, involves lower estimation risk and dominates any attempt to estimate the higher-order comoments for small sample sizes, even using improved estimators. When increasing the data frequency (which leads to increasing the number of observations for a constant rolling window size), sample risk embedded in any non-trivial higher-order risk parameter estimates decreases up to the point where the improvement in the utility function higher-order approximation eventually pays off. The results in Table 11 indicate that the sample size needed for this to happen is smaller for improved estimators (corresponding to the use of weekly data), as compared to sample estimators (corresponding to the use of daily data). It should be noted that changing the sampling frequency not only changes the number of observations (for a fixed window of time) but may also change the distribution of returns. Given that daily returns are notoriously exhibiting stronger deviations from normality compared to weekly and monthly returns, it might be that increasing the sampling frequency not only favors the higher-order approximation because more observations are available to estimate higher-order comoments, but also because the resulting return distribution shows a stronger departure from the Gaussian distribution.

5. Conclusion
This paper introduces improved estimators for higher-order moments and comoments 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.

We find that improved estimates strongly reduce the estimation error as measured by the mean squared error of the corresponding comoments. The reduction is particularly significant for the skewness-coskewness matrix, a result in line with the intuition that estimators for odd moments are more noisy than estimators for even moments of asset return distributions. Consequently, optimal shrinkage intensities for the third-order moment tensor tend to be significantly higher than those observed for second- and fourth-order moment tensors. Next, we follow related literature [cf. Jagannathan and Ma (2003) or Chan, Karceski, and Lakonishok (1999)] by performing an empirical analysis of the competing estimators, using rolling window optimizations in a buy-and-hold framework. Portfolios built on structured or shrinkage estimators are found to outperform portfolios built on sample estimators in terms of out-of-sample expected utility. The monetary utility gain ranges between 100 and 500 basis points per year depending on the parameter values. Monetary utility gains increase with the portfolio size and the level of risk aversion, and decrease with the size of the rolling estimation window. Finally, we find substantial reductions of turnover

---

11 - To confirm this finding, we have extended to the context of higher-order moment tensors the formal cut-off point analysis of Jagannathan and Ma (2003), which aims at estimating the sample size needed for specification error and estimation error to even out. Our results, which are omitted for the sake of brevity and may be obtained from the authors upon request, indicate that a successful attempt at implementing higher-order moment portfolio selection can be achieved for a smaller sample size when improved estimators are used, compared to a situation where naïve sample estimators are used.

12 - We would like to thank the referee for this observation.
rates when using improved estimators, as compared to sample estimators. We argue that this result confirms the superior estimation quality associated to improved estimators.

In general, the single-factor approach and the estimator shrunk towards the single-factor estimator outperform their constant correlation counterparts, which likely involve higher specification error. Overall, our results suggest that when an investor's objective incorporates higher-order moments, improved estimators generate significant welfare gains in an out-of-sample context, a result that extends previous similar findings about the covariance matrix.
Appendix

We now explain how to extend the statistical shrinkage approach of Ledoit and Wolf (2003) to the estimation of fourth-order comoment parameters by choosing the function $g$ and the parameter vector in Equation (34) appropriately.

The $\hat{\rho}_{iiij}$ are trivially given by $\hat{\sigma}_{iiij}$, respectively. Next, the expressions for $g$ and $\theta$ follow from the relationships established in (17) for the constant correlation estimator and in (23) for the single factor estimator. As far as the constant correlation approach is concerned, the extension using the Delta-method is straightforward. Following (13), we distinguish four different index combinations for the fourth-order moment tensor matrix. As in (36) for the second-order moment tensor, we customize 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 Delta-method, we find the asymptotic covariances between the constant correlation estimates and the sample estimates for the fourth-order moment tensor matrix entries as the off-diagonal elements of the $2 \times 2$ matrix $\nabla_g^T \Sigma \nabla_g$ (cf. 34):

\[
\begin{align*}
\hat{\rho}^{CC}_{ijij} & = \frac{\hat{\rho}^{(3)}_{ijij}}{2} \left[ \frac{m^{(2)}_i}{m^{(2)}_i} \text{AsyCov} \left( \sqrt{T m^{(2)}_i}, \sqrt{T s_{ijij}} \right) + \frac{m^{(2)}_j}{m^{(2)}_j} \text{AsyCov} \left( \sqrt{T m^{(2)}_j}, \sqrt{T s_{ijij}} \right) \right] \\
\hat{\rho}^{CC}_{iiij} & = \frac{\hat{\rho}^{(5)}_{iiij}}{2} \left[ \frac{m^{(4)}_i}{m^{(4)}_i} \text{AsyCov} \left( \sqrt{T m^{(4)}_i}, \sqrt{T s_{iiij}} \right) + \frac{m^{(4)}_j}{m^{(4)}_j} \text{AsyCov} \left( \sqrt{T m^{(4)}_j}, \sqrt{T s_{iiij}} \right) \right] \\
\hat{\rho}^{CC}_{ijjk} & = \frac{\hat{\rho}^{(6)}_{ijjk}}{2} \left[ \frac{m^{(4)}_i}{m^{(4)}_i} \text{AsyCov} \left( \sqrt{T m^{(4)}_i}, \sqrt{T s_{ijjk}} \right) \right] + \frac{\hat{\rho}^{(6)}_{ijjk}}{4} \left[ \frac{m^{(4)}_i}{m^{(4)}_i} \text{AsyCov} \left( \sqrt{T m^{(4)}_i}, \sqrt{T s_{ijjk}} \right) \right] \\
\hat{\rho}^{CC}_{ijkl} & = \frac{\hat{\rho}^{(7)}_{ijkl}}{4} \left[ \frac{m^{(4)}_i}{m^{(4)}_i} \text{AsyCov} \left( \sqrt{T m^{(4)}_i}, \sqrt{T s_{ijkl}} \right) \right] + \frac{\hat{\rho}^{(7)}_{ijkl}}{4} \left[ \frac{m^{(4)}_j}{m^{(4)}_j} \text{AsyCov} \left( \sqrt{T m^{(4)}_j}, \sqrt{T s_{ijkl}} \right) \right]
\end{align*}
\]

(A.1)

As far as the single factor approach is concerned, by construction, the comoment elements in $\Psi$ and $\Phi$ are asymptotically uncorrelated with $s_{ij}$ whereas the situation is less straightforward for the elements of $\Upsilon$ [Equations (24)-(26)]. Accordingly, we can derive similar results to (38) for the asymptotic covariance between the single factor estimate of the fourth-order
moment tensor matrix and the corresponding sample estimate. We set:

\[ \theta_4^{SF} = (s_{i0}, s_{j0}, s_{k0}, m_0^{(2)}, m_0^{(4)}, s_{ijkl})^T \]

\[ \gamma_4^{SF}(\theta_4^{SF}) = \left( \begin{array}{c} s_{i0}s_{j0}s_{k0}m_0^{(4)} \\ (m_0^{(2)})^2m_0^{(4)} \\ s_{ijkl} \end{array} \right) = \left( \begin{array}{c} \beta_1 \beta_2 \beta_3 \beta_4 m_0^{(4)} \\ s_{ijkl} \end{array} \right) \]

(A.2)

and obtain:

\[ \rho_{ijkl}^{SF} = \frac{s_{i0}s_{j0}s_{k0}m_0^{(4)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}s_{i0}, \sqrt{T}s_{ijkl} \right) + \frac{s_{i0}s_{j0}s_{k0}m_0^{(4)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}s_{j0}, \sqrt{T}s_{ijkl} \right) \]

\[ + \frac{s_{i0}s_{j0}s_{k0}m_0^{(4)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}s_{k0}, \sqrt{T}s_{ijkl} \right) + \frac{s_{i0}s_{j0}s_{k0}m_0^{(4)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}s_{l0}, \sqrt{T}s_{ijkl} \right) \]

\[ - \frac{4s_{i0}s_{j0}s_{k0}m_0^{(4)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}m_0^{(2)}, \sqrt{T}s_{ijkl} \right) + \frac{4s_{i0}s_{j0}s_{k0}m_0^{(4)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}m_0^{(4)}, \sqrt{T}s_{ijkl} \right) \]

\[ + r_{ijkl}^{*} \]

(A.3)

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 (26), we have:

\[ r_{ijij}^{*} = 3 \frac{s_{i0}s_{j0}m_0^{(2)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}s_{i0}, \sqrt{T}s_{ijij} \right) + \frac{s_{i0}s_{j0}m_0^{(2)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}s_{j0}, \sqrt{T}s_{ijij} \right) \]

\[ - \frac{s_{i0}s_{j0}m_0^{(2)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}m_0^{(2)}, \sqrt{T}s_{ijij} \right) + \frac{s_{i0}s_{j0}m_0^{(2)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}m_0^{(4)}, \sqrt{T}s_{ijij} \right) \]

\[ r_{iijj}^{*} = 2 \frac{s_{i0}s_{j0}m_0^{(2)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}s_{i0}, \sqrt{T}s_{iijj} \right) - \frac{\psi_{ij}^2}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}m_0^{(2)}, \sqrt{T}s_{iijj} \right) \]

\[ + \frac{s_{i0}^2m_0^{(2)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}s_{00}, \sqrt{T}s_{ijij} \right) + \frac{2s_{i0}m_0^{(2)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}s_{j0}, \sqrt{T}s_{ijij} \right) \]

\[ - \frac{s_{i0}^2m_0^{(2)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}m_0^{(2)}, \sqrt{T}s_{ijij} \right) + \frac{s_{i0}^2m_0^{(2)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}m_0^{(4)}, \sqrt{T}s_{ijij} \right) \]

\[ + \psi_{ij} \text{AsyCov} \left( \sqrt{T}s_{ijij}, \sqrt{T}s_{ijij} \right) + \psi_{ij} \text{AsyCov} \left( \sqrt{T}s_{ijij}, \sqrt{T}s_{ijij} \right) \]

\[ r_{ijkj}^{*} = \frac{s_{i0}s_{k0}m_0^{(2)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}s_{i0}, \sqrt{T}s_{ijkj} \right) + \frac{s_{j0}s_{k0}m_0^{(2)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}s_{j0}, \sqrt{T}s_{ijkj} \right) \]

\[ - \frac{s_{j0}s_{k0}m_0^{(2)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}m_0^{(2)}, \sqrt{T}s_{ijkj} \right) + \frac{s_{j0}s_{k0}m_0^{(2)}}{m_0^{(2)}} \text{AsyCov} \left( \sqrt{T}m_0^{(4)}, \sqrt{T}s_{ijkj} \right) \]

\[ r_{ijkl}^{*} = 0. \]

(A.4)

All asymptotic covariances in (A.3) and (A.4) are, similarly to (39), consistently estimated by:

\[ \text{AsyCov} \left( \sqrt{T}s_{i0}, \sqrt{T}s_{ijkl} \right) = \frac{1}{T} \sum_{t=1}^{T} \left[ (R_{it} - m_i)(R_{it} - m_i) - s_{ijkl} \right] \]

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

(A.5)
where,

\[
\begin{align*}
  s_{n0} &= \frac{1}{T} \sum_{t=1}^{T} (R_{nt} - m_n)(R_{0t} - m_0) \\
  \hat{\psi}_{it} &= \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it}^2.
\end{align*}
\]

(A.6)
References


