A coskewness shrinkage approach for estimating the skewness of linear combinations of random variables

Kris Boudt
Department of Business, Vrije Universiteit Brussel
Faculty of Economics and Business, Vrije Universiteit Amsterdam
and
Dries Cornilly*
Department of Business, Vrije Universiteit Brussel
Department of Mathematics, KU Leuven
and
Tim Verdonck
Department of Mathematics, KU Leuven

Abstract

Decision making in business and economics often requires an accurate estimate of the coskewness matrix to optimize the allocation to random variables with asymmetric distributions. The classical sample estimator of the coskewness matrix performs poorly in terms of mean squared error (MSE) when the sample size is small. A solution is to use shrinkage estimators, defined as the convex combination between the sample coskewness matrix and a target matrix, with the aim of minimizing the MSE. In this paper, we propose unbiased consistent estimators for the MSE loss function and include the possibility of having multiple target matrices. Simulations show that these improvements lead to a substantial reduction in the MSE when estimating the third order comoment matrix of asymmetric distributions, as well as for the estimation of the skewness of a linear combination of random variables. In a financial portfolio application, we find that the proposed shrinkage coskewness estimators are effective in determining the linear combination with the highest expected utility.

Keywords: Coskewness; MSE; multiple targets; portfolio optimization; shrinkage.

*Corresponding author: Vrije Universiteit Brussel, Pleinlaan 2, 1050 Bruxelles, Belgium; KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium. (email: dries.cornilly@kuleuven.be)

1 Introduction

Accurate estimation of the skewness of linear combinations of p random variables is a key concern in business and economics. The estimation and management of the third order interactions between financial returns is becoming increasingly important in financial applications such as asset allocation, portfolio management and risk analysis. The traditional approach to solving the problem is to estimate the skewness directly from the sample values of that linear combination (Hosking (1990)). An alternative, which is especially popular in finance, is to obtain the skewness statistic using an estimate of the $p \times p^2$ third order comoment matrix measuring the third-order interactions between the random variables. The estimation of such a coskewness matrix faces a similar curse of dimensionality as the estimation of the covariance matrix.

Though the higher order moments are becoming more important, the number of available estimation methods is limited, compared to the burgeoning literature on covariance estimation. In the seminal paper by Martellini & Ziemann (2010), the work on linear shrinkage of covariance matrices by Ledoit & Wolf (2003, 2004) is generalized to the use of shrinkage for the estimation of the coskewness matrix. Their extension includes proposing an estimator for the optimal shrinkage intensity in terms of Mean Squared Error (MSE). However, the proposed estimator is not consistent and uses a biased estimator for the MSE loss function. Moreover, their approach is limited to shrinking towards a single target matrix.

In this paper we contribute to the literature on skewness shrinkage estimation in four ways. Firstly, we propose unbiased consistent estimators for the MSE loss function, which leads to an improved estimation of the shrinkage intensity. Under the MSE loss function, we derive optimal targets for a given coskewness structure. Thirdly, we extend the methodology to include multiple targets simultaneously, eliminating the need to choose a single target matrix. Fourthly, we show on simulated and real-life return data that the computational complexity of estimating a $p \times p^2$ coskewness matrix pays off in terms of a more precise estimate of the skewness of a linear combination of skewed variables. This result is of direct interest for decision making based on higher order approximations of the expected utility function (see e.g. Jondeau & Rockinger (2006) and Martellini & Ziemann (2010)) or the density function (see e.g. Boudt et al. (2008), Del Brio et al. (2009) and Stoyanov et al. (2013)).

We illustrate our methodology on hedge fund return data for which accurate estimates of the coskewness matrix are needed in portfolio optimization with a maximum expected utility objective function. Rolling estimation windows are used to account for the transient nature of skewness of financial returns (Beedles (1979), Singleton & Wingender (1986)). When there are more assets than observations, this leads to a misspecified sample coskewness matrix. The out-of-sample evaluation of investment performance shows that the proposed MSE optimized shrinkage estimators not only solve the issue from a statistical perspective, but also lead to substantial economic gains for the fund of hedge funds investor with CRRA preferences.

The remainder of the paper is organized as follows. Section 2 introduces the notation and the traditional coskewness estimators. Section 3 describes the model of single-target linear shrinkage estimation of the coskewness matrix and extends the single-target framework to include the possibility of multiple targets. Section 4 provides both the plug-in estimators and our proposed unbiased consistent estimators. The good performance of the proposed estimators is illustrated in an extensive simulation study in Section 5. Finally, we document the usefulness of the shrinkage estimators for optimizing portfolios of hedge funds in Section 6. We end with a conclusion and some suggestions for further research.

A supplementary appendix contains additional simulation results and robustness checks for the empirical application. We also show how to include several other coskewness matrices from the literature into the multi-target shrinkage framework introduced in this paper. Finally, we demonstrate the `R` code for our estimators, which is publicly available in the `PerformanceAnalytics` package of Peterson & Carl (2016).

2 Coskewness Estimation

In this section we first introduce the notation for the remainder of the paper. We then present the sample estimator and the structured estimation approach, which will serve as the basis of the shrinkage estimator proposed in Section 3.

2.1 Notation

Let $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ with $\boldsymbol{x}_i \in \mathbb{R}^p$ be a sample of n independent and identically distributed p-dimensional vectors drawn from the distribution of a random variable $\boldsymbol{X}$ with mean $\boldsymbol{\mu}$ and coskewness matrix $\boldsymbol{\Phi}$. The matrix $\boldsymbol{\Phi}$ consists of the third order central moments

$$\phi_{ijk} = \mathbb{E}\left[(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)\right], \qquad i, j, k = 1, \ldots, p, \tag{1}$$

where X_i and μ_i are respectively the i-th component of the random variable $\boldsymbol{X}$ and the mean $\boldsymbol{\mu}$. The coskewness elements are stacked in the coskewness matrix as

$$\boldsymbol{\Phi} = \mathbb{E}\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})' \otimes (\boldsymbol{X} - \boldsymbol{\mu})'\right], \tag{2}$$

where $\otimes$ denotes the Kronecker product. Equivalently, we have that the coskewness matrix can be written as

$$\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1 | \cdots | \boldsymbol{\Phi}_p), \tag{3}$$

with

$$\boldsymbol{\Phi}_k = \begin{pmatrix} \phi_{k11} & \cdots & \phi_{k1p} \\ \vdots & \ddots & \vdots \\ \phi_{kp1} & \cdots & \phi_{kpp} \end{pmatrix}. \tag{4}$$

Note that ϕ_{ijk} has the same value for each permutation of the indices i, j and k. This property is called supersymmetry of the matrix $\boldsymbol{\Phi}$. The coskewness matrix $\boldsymbol{\Phi}$ is of dimension $p \times p^2$, but due to the property of supersymmetry, only $p(p+1)(p+2)/6$ elements are unique.

For a linear combination $\boldsymbol{v}'\boldsymbol{X}$ of $\boldsymbol{X}$, define the univariate skewness $\phi_{\boldsymbol{v}}$ as

$$\phi_{\boldsymbol{v}} = \mathbb{E}\left[(\boldsymbol{v}'\boldsymbol{X} - \boldsymbol{v}'\boldsymbol{\mu})^3\right] = \boldsymbol{v}'\boldsymbol{\Phi}(\boldsymbol{v} \otimes \boldsymbol{v}). \tag{5}$$

This paper uses boldface letters to denote matrices and vectors. For matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, the inner product is defined by $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \mathrm{trace}(\boldsymbol{A}'\boldsymbol{B})$. The Frobenius norm is denoted by $\| \cdot \|$ and it holds that $\langle \boldsymbol{A}, \boldsymbol{A} \rangle = \|\boldsymbol{A}\|^2$. Additionally, $\xrightarrow{a.s.}$ and $\xrightarrow{p}$ denote convergence almost surely and convergence in probability respectively, always for $n \to \infty$. When $\to$ is used, convergence of a sequence of real numbers is meant. Finally, $f(n) = O(g(n))$ if and only if there exists $n_0 \in \mathbb{N}$ and $M \in \mathbb{R}^+$ such that $|f(n)| \leq M|g(n)|$ for all $n \geq n_0$.

2.2 Sample Estimator

Probably the most intuitive way to estimate $\mathbf{\Phi}$ is by the plug-in method, where expectations are replaced by sample averages. First, the mean $\boldsymbol{\mu}$ is estimated by the sample average defined as

$$\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{l=1}^{n}\boldsymbol{x}_l. \tag{6}$$

By replacing the mean in (1) with the sample mean and the expectation with a sample average, we obtain the plug-in estimator used in Martellini & Ziemann (2010) for a coskewness element

$$\widehat{\phi}^{\text{pl}}_{ijk} = \frac{1}{n}\sum_{l=1}^{n}(x_{li}-\overline{x}_i)(x_{lj}-\overline{x}_j)(x_{lk}-\overline{x}_k). \tag{7}$$

The above estimator is biased. As an alternative, we recommend to use the unbiased estimator for ϕ_{ijk}, see e.g. Fisher (1929), given by

$$\widehat{\phi}_{ijk} = \frac{n}{(n-1)(n-2)}\sum_{l=1}^{n}(x_{li}-\overline{x}_i)(x_{lj}-\overline{x}_j)(x_{lk}-\overline{x}_k) \tag{8}$$

and define the sample coskewness estimator $\widehat{\mathbf{\Phi}}$ with elements $\widehat{\phi}_{ijk}$. Note that the constant $n/((n-1)(n-2))$ is in essence a small sample correction, with a larger impact than the factor $1/(n-1)$ for unbiased covariance estimation.

In Appendix A, we use classical convergence results (see e.g. Serfling (2009)) to prove that the sample estimator ϕ_{ijk} converges almost surely to the true coskewness value when $n \to \infty$. Proofs for all other properties in this paper are also given in Appendix A.

Property 2.1. *Assume that $\boldsymbol{X}$ has finite third order moments. Then for any sample size n, $\mathbb{E}[\widehat{\phi}_{ijk}] = \phi_{ijk}$, and when $n \to \infty$, it holds that $\widehat{\phi}_{ijk} \xrightarrow{a.s.} \phi_{ijk}$.*

The unbiased sample estimator of the univariate skewness of $\boldsymbol{v}'\boldsymbol{X}$ in (5) equals

$$\widehat{\phi}_{\boldsymbol{v}} = \frac{n}{(n-1)(n-2)}\sum_{l=1}^{n}(\boldsymbol{v}'\boldsymbol{x}_i - \boldsymbol{v}'\overline{\boldsymbol{x}})^3. \tag{9}$$

Computing the skewness of a linear combination of random variables by using the univariate sample skewness is equivalent to first computing the sample coskewness estimator and then transforming the result to obtain the univariate estimate.

Property 2.2. *For any* $\boldsymbol{v} \in \mathbb{R}^p$*, it holds that*

$$\widehat{\phi}_{\boldsymbol{v}} = \boldsymbol{v}'\widehat{\boldsymbol{\Phi}}(\boldsymbol{v} \otimes \boldsymbol{v}). \tag{10}$$

When $p > n$, there are infinitely many vectors $\boldsymbol{v}$ such that $\boldsymbol{v}'\boldsymbol{x}_i = c, i = 1, \ldots, n$ and hence, $\widehat{\phi}_{\boldsymbol{v}} = 0$. In this case, we call the sample coskewness matrix misspecified. A similar observation is made by Engle (2009) in the case of the sample covariance matrix.

2.3 Structured Coskewness Estimation

The sample estimator faces a curse of dimensionality due to its many different elements. This can be avoided by regularizing the coskewness estimator by imposing some restrictions on its structure. In particular, when sample coskewness elements can be expected to be similar, it may be possible to reduce the mean squared error by replacing the corresponding sample estimates by the sample average of these elements. This is a similar intuition as for the diagonal covariance matrix in Ledoit & Wolf (2004) or the block diagonal structure in Devijver & Gallopin (2016).

Formally, let $\boldsymbol{E}$ be a supersymmetric matrix of dimension $p \times p^2$ containing only entries zero or one. The entry $\boldsymbol{E}_{ijk}$ equals one when the estimate $\widehat{\phi}_{ijk}$ is assumed to be close to the other sample estimates in $\boldsymbol{\Phi}$ on the position where $\boldsymbol{E}$ equals one. Combining several such selection matrix, we propose to use the coskewness estimator $\widehat{\boldsymbol{T}}^{\star}$ defined by

$$\widehat{\boldsymbol{T}}^{\star} = \sum_{q=1}^{Q} \frac{\langle \boldsymbol{E}_q, \widehat{\boldsymbol{\Phi}} \rangle}{\|\boldsymbol{E}_q\|^2} \boldsymbol{E}_q, \tag{11}$$

where $\langle \boldsymbol{E}_r, \boldsymbol{E}_s \rangle = 0$ when $r \neq s$, i.e. the matrices $\boldsymbol{E}_q$ do not have common 1-entries. Note that the coefficients in $\widehat{\boldsymbol{T}}^{\star}$ are sample averages of the sample estimates $\widehat{\boldsymbol{\phi}}_{ijk}$ corresponding to the positions where $\boldsymbol{E}_{q,ijk} = 1$. In the supplementary appendix, we provide a 2-dimensional example. When $Q = p(p+1)(p+2)/6$, this corresponds to assuming that each coskewness element is unique. Hence, $\widehat{\boldsymbol{T}}^{\star}$ equals the sample coskewness matrix $\widehat{\boldsymbol{\Phi}}$. A low choice of Q means that a lot of different coskewness elements are assumed to roughly equal, or zero. This might introduce an estimation bias, but possibly reduces the estimation variance drastically.

The structured coskewness matrices in (11) are unbiased and consistent for the coskewness matrix $\boldsymbol{T}^\star$ defined by

$$\boldsymbol{T}^\star = \sum_{q=1}^{Q} \frac{\langle \boldsymbol{E}_q, \boldsymbol{\Phi} \rangle}{\|\boldsymbol{E}_q\|^2} \boldsymbol{E}_q. \tag{12}$$

Property 2.3. *Assume that $\boldsymbol{X}$ has finite third order moments. For any sample size it holds that $\mathbb{E}[\widehat{\boldsymbol{T}}^\star] = \boldsymbol{T}^\star$ and as $n \to \infty$, $\widehat{\boldsymbol{T}}^\star \xrightarrow{a.s.} \boldsymbol{T}^\star$.*

We proceed by giving three examples of structured coskewness matrices derived under this approach. The matrices are numbered because they will also be used in the simulation study in Section 5 and the empirical application in Section 6.

The simplest structured coskewness matrix is the deterministic matrix of dimension $p \times p^2$ containing only zeros, denoted by $\boldsymbol{T}_1^\star$. This is the coskewness matrix arising from a random variable $\boldsymbol{X} \in \mathbb{R}^p$ that is central-symmetric about a certain $\boldsymbol{\theta} \in \mathbb{R}^p$. In particular this holds for any elliptical distribution.

Coskewness matrices $\widehat{\boldsymbol{T}}_2^\star$ and $\widehat{\boldsymbol{T}}_3^\star$ assume independence. Structured coskewness matrix $\widehat{\boldsymbol{T}}^\star$ also assumes that the marginals have a common third order central moment. They are given by

$$\widehat{\boldsymbol{T}}_2^\star = \frac{\sum_{l=1}^{p} \widehat{\phi}_{lll}}{p} \boldsymbol{E}_2, \qquad \text{and} \qquad \widehat{\boldsymbol{T}}_3^\star = \sum_{q=1}^{p} \widehat{\phi}_{qqq} \boldsymbol{E}_{3,q}, \tag{13}$$

with $\boldsymbol{E}_2$ the $p \times p^2$ matrix for which all elements are equal to zero, except for the marginal third order central moments, i.e. the positions of $\phi_{iii}, i = 1, \dots, p$ have as entry the value 1 and $\boldsymbol{E}_{3,q}, q = 1, \dots, p$ are the matrices with the only non-zero element at position ϕ_{qqq}.

Other structured coskewness matrices available in the literature include the coskewness matrix under the latent single-factor model of Simaan (1993), the single-factor and constant correlation coskewness matrix of Martellini & Ziemann (2010) and the coskewness matrix under the multi-factor model of Boudt et al. (2015).

3 Shrinkage Estimation

In this section we present the single-target shrinkage estimator as a convex combination which combines the sample coskewness estimator $\widehat{\boldsymbol{\Phi}}$ and an alternative estimator $\widehat{\boldsymbol{T}}$ into a shrinkage estimator of $\boldsymbol{\Phi}$. The sample estimator $\widehat{\boldsymbol{\Phi}}$ of the coskewness matrix is an unbiased estimator, but typically has a large estimation variance depending on moments up to the sixth order of the distribution. The target matrix $\widehat{\boldsymbol{T}}$ can be any $p \times p^2$ supersymmetric matrix for which $\widehat{\boldsymbol{T}} \neq \widehat{\boldsymbol{\Phi}}$. Typical choices for $\widehat{\boldsymbol{T}}$ are the structured estimators in (11), the coskewness matrix based on latent single-factor model (Simaan (1993)), the constant-correlation correlation coskewness matrix or coskewness matrix based on an observed single-factor model (Martellini & Ziemann (2010)), or the multi-factor coskewness matrix of Boudt et al. (2015). The estimator $\widehat{\boldsymbol{T}}$ of a structured coskewness matrix $\boldsymbol{T}$ usually has a lower estimation variance, but possibly is not consistent for the true coskewness matrix. A natural improvement is to combine both estimates into a single estimate by minimizing the mean squared error (MSE).

Formally, we define the single-target shrinkage estimator as

$$\widehat{\boldsymbol{\Phi}}^{\mathrm{ST}}(\lambda) = (1-\lambda)\widehat{\boldsymbol{\Phi}} + \lambda\widehat{\boldsymbol{T}}, \tag{14}$$

where the shrinkage intensity $\lambda \in [0, 1]$. Any coskewness matrix $\widehat{\boldsymbol{T}} \neq \widehat{\boldsymbol{\Phi}}$ can be used in shrinkage estimation and in the shrinkage setting we refer to the structured coskewness matrix as the target matrix. We focus on target matrices as in (11), while in the supplementary appendix we work out the details for the other available coskewness estimators.

Single-target shrinkage for the coskewness matrix was proposed in Martellini & Ziemann (2010) for two particular structured coskewness matrices. We generalize their framework by introducing other (multiple) target matrices and proposing an unbiased and consistent estimator of the MSE loss function, as well as deriving the theoretical properties of the proposed estimators.

3.1 Optimization of the Shrinkage Intensity

In Martellini & Ziemann (2010), it is recommended to optimize the shrinkage intensity parameter λ in terms of the MSE loss function, given by

$$L(\lambda) = \mathbb{E}\left[\|(1-\lambda)\widehat{\boldsymbol{\Phi}} + \lambda\widehat{\boldsymbol{T}} - \boldsymbol{\Phi}\|^2\right], \tag{15}$$

which is equivalent to

$$L(\lambda) = A\lambda^2 - 2b\lambda + \mathbb{E}\left[\|\widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}\|^2\right], \tag{16}$$

where

$$A = \mathbb{E}\left[\|\widehat{\boldsymbol{T}} - \widehat{\boldsymbol{\Phi}}\|^2\right] \qquad \text{and} \qquad b = \mathbb{E}\left[\langle \widehat{\boldsymbol{\Phi}} - \widehat{\boldsymbol{T}}, \widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}\rangle\right]. \tag{17}$$

Clearly, the approach of optimizing the MSE only makes sense if the MSE in (15) is finite. We therefore require that $\boldsymbol{X}$ has finite sixth order moments and the structured coskewness matrix $\widehat{\boldsymbol{T}}$ has a finite MSE as well. Then the optimal shrinkage intensity, minimizing the MSE in (15), is given by

$$\lambda^\star = \frac{b}{A}, \tag{18}$$

which implicitly depends on the sample size through both A and b.

Given the optimal shrinkage intensity $\lambda^\star$, the single-target shrinkage estimator $\widehat{\boldsymbol{\Phi}}^{\mathrm{ST}}(\lambda^\star)$ is consistent for $\boldsymbol{\Phi}$.

Property 3.1. *Assume that $\boldsymbol{X}$ has finite sixth order moments and $\widehat{\boldsymbol{T}} \xrightarrow{p} \boldsymbol{U}$ for some $\boldsymbol{U}$ such that $\mathrm{Var}(\widehat{t}_{ijk}) = O(n^{-1})$, then $\widehat{\boldsymbol{\Phi}}^{ST}(\lambda^\star) \xrightarrow{p} \boldsymbol{\Phi}$ as $n \to \infty$.*

The second condition in Property 3.1 implies that $b = O(n^{-1})$. Hence, when $\boldsymbol{U} \neq \boldsymbol{\Phi}$, it holds that $\lambda = O(n^{-1})$. These conditions are weak. For target matrices $\widehat{\boldsymbol{T}}^\star$ in (11), the conditions are satisfied when $\boldsymbol{X}$ has finite sixth order moments. Hence, for the the target coskewness matrices used in this paper, the only condition is the existence of sixth order moments of $\boldsymbol{X}$. The structured coskewness matrices in the supplementary appendix also satisfy the assumptions of Property 3.1 under mild assumptions on $\boldsymbol{X}$.

Note that $\lambda^\star$ depends on the unknown A and b. Hence, the optimal shrinkage intensity needs to be estimated through estimation of A and b, which we propose in Section 4. If the shrinkage intensity $\lambda^\star$ is consistently estimated, it follows that the estimator $\widehat{\boldsymbol{\Phi}}^{\mathrm{ST}}(\widehat{\lambda}^\star)$ is consistent as well.

By definition, the MSE of the shrinkage estimator is equal or lower than the MSE of the sample estimator. This decrease comes at the cost of a bias, introduced by the target matrix. However, the shrinkage estimator optimally balances the trade-off between estimation variance and bias by selecting the shrinkage intensity resulting in the lowest MSE.

Optimal coskewness shrinkage estimation is not only a matter of balancing the sample estimator and the target, but also of selecting an appropriate target matrix. In this regard, the structured coskewness matrices in (11) can be seen as the solution minimizing the MSE with a more generic structured coskewness matrix.

Property 3.2. *Assume that $\boldsymbol{X}$ has finite sixth order moments and a structured coskewness matrix $\boldsymbol{T} = \sum_{q=1}^{Q} \nu_q \boldsymbol{E}_q$, with ν_q scalars and the matrices $\boldsymbol{E}_q$ satisfying the constraints as for* (11)*. Then for any $\lambda \in (0,1]$, the values of ν_q minimizing the MSE*

$$L(\nu_1, \ldots, \nu_q) = \mathbb{E}\left[\|(1-\lambda)\widehat{\boldsymbol{\Phi}} + \lambda \sum_{q=1}^{Q} \nu_q \boldsymbol{E}_q - \boldsymbol{\Phi}\|^2\right]. \tag{19}$$

are $\nu_q = \langle \boldsymbol{E}_q, \boldsymbol{\Phi}\rangle / \|\boldsymbol{E}_q\|^2$ and thus $\boldsymbol{T}^\star$ is optimal for the structure given by the $\boldsymbol{E}$-matrices.

3.2 Extension to Multi-Target Shrinkage

Often several plausible target coskewness estimators exist. We then recommend to use shrinkage estimation with multiple targets. As for the multi-target shrinkage covariance estimator of Bartz et al. (2014) and Lancewicki & Aladjem (2014), the target weights can be est in a data-driven manner by minimizing the MSE. This section introduces the multi-target shrinkage estimator for the coskewness matrix.

The proposed multi-target shrinkage coskewness estimator is given by

$$\widehat{\boldsymbol{\Phi}}^{\mathrm{MT}}(\boldsymbol{\lambda}) = \left(1 - \sum_{m=1}^{t} \lambda_m\right) \widehat{\boldsymbol{\Phi}} + \sum_{m=1}^{t} \lambda_m \widehat{\boldsymbol{T}}_m, \tag{20}$$

where t is the number of targets and $\widehat{\boldsymbol{T}}_m$ are the target matrices ($m = 1, \ldots, t$) of dimension $p \times p^2$ and the shrinkage intensity is constrained by

$$\sum_{m=1}^{t} \lambda_m \leq 1 \quad \text{and} \quad \lambda_m \geq 0, \quad m = 1, \ldots, t. \tag{21}$$

Assume that none of the targets equals the sample coskewness estimator and $\widehat{\boldsymbol{\Phi}} - \widehat{\boldsymbol{T}}_1, \ldots, \widehat{\boldsymbol{\Phi}} - \widehat{\boldsymbol{T}}_t$ are linearly independent.

As in the single-target shrinkage case, we seek to find $\boldsymbol{\lambda}$ such that the MSE loss function

$$L(\boldsymbol{\lambda}) = \mathbb{E}\left[\|\widehat{\boldsymbol{\Phi}}^{\mathrm{MT}}(\boldsymbol{\lambda}) - \boldsymbol{\Phi}\|^2\right] \tag{22}$$

is minimal, under the linear constraints in (21). This optimization problem is a linearly constrained convex quadratic program. To show this, we first rewrite the loss function as

$$L(\boldsymbol{\lambda}) = \boldsymbol{\lambda}' \boldsymbol{A} \boldsymbol{\lambda} - 2\boldsymbol{b}'\boldsymbol{\lambda} + \mathbb{E}\left[\|\widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}\|^2\right], \tag{23}$$

where $\boldsymbol{A} \in \mathbb{R}^{t\times t}, \boldsymbol{b} \in \mathbb{R}^t$ are defined as

$$A_{ij} = \mathbb{E}\left[\langle \widehat{\boldsymbol{T}}_i - \widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{T}}_j - \widehat{\boldsymbol{\Phi}}\rangle\right] \quad \text{and} \quad b_i = V(\widehat{\boldsymbol{\Phi}}) - C(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{T}}_i), \qquad i,j = 1,\dots,t, \tag{24}$$

with

$$V(\widehat{\boldsymbol{\Phi}}) = \mathbb{E}\left[\|\widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}\|^2\right] \quad \text{and} \quad C(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{T}}_i) = \mathbb{E}\left[\langle \widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}, \widehat{\boldsymbol{T}}_i - \mathbb{E}[\widehat{\boldsymbol{T}}_i]\rangle\right]. \tag{25}$$

Note that the matrix $\boldsymbol{A}$ is a Gramian matrix and hence is positive definite if and only if $\widehat{\boldsymbol{\Phi}} - \widehat{\boldsymbol{T}}_1, \dots, \widehat{\boldsymbol{\Phi}} - \widehat{\boldsymbol{T}}_t$ are linearly independent. Thus, for sufficiently different target matrices, the quadratic program is strictly convex with linear constraints and yields the solution $\boldsymbol{\lambda}^\star$. The vector of shrinkage intensities nests as a special case the single-target shrinkage intensity in (18).

The consistency result of the shrinkage estimator also holds in the multi-target setting.

Property 3.3. *Under the same conditions as Property 3.1 (for each of the targets) and the assumption that* $\widehat{\boldsymbol{\Phi}} - \widehat{\boldsymbol{T}}_1, \dots, \widehat{\boldsymbol{\Phi}} - \widehat{\boldsymbol{T}}_t$ *are linearly independent,* $\widehat{\boldsymbol{\Phi}}^{MT}(\boldsymbol{\lambda}^\star) \xrightarrow{p} \boldsymbol{\Phi}$ *as* $n \to \infty$*.*

It should be noted that the optimal shrinkage intensity, with respect to the MSE, depends on the quantities $\boldsymbol{b}$ and $\boldsymbol{A}$, which are unknown in practice.

4 Estimation of the Shrinkage Intensity

The optimal shrinkage coefficient $\boldsymbol{\lambda}^\star$, that minimizes the MSE loss function $L(\boldsymbol{\lambda})$ in (23), subject to the constraints (21), depends on the unknown quantities $\boldsymbol{A}$ and $\boldsymbol{b}$ in (24). An estimate of the optimal shrinkage intensity is obtained by minimizing the quadratic function

$$\widehat{L}(\boldsymbol{\lambda}) = \boldsymbol{\lambda}'\widehat{\boldsymbol{A}}\boldsymbol{\lambda} - 2\widehat{\boldsymbol{b}}'\boldsymbol{\lambda}, \tag{26}$$

subject to the constraints (21). The natural approach to estimate the matrix $\boldsymbol{A}$ is to use its sample version, i.e.

$$\widehat{A}_{ij} = \langle \widehat{\boldsymbol{T}}_i - \widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{T}}_j - \widehat{\boldsymbol{\Phi}}\rangle, \qquad i,j = 1,\dots,p. \tag{27}$$

Before providing the properties of this estimator, we note that for A_{ij} in (24),

$$A_{ij} \to \langle \boldsymbol{T}_i - \boldsymbol{\Phi}, \boldsymbol{T}_j - \boldsymbol{\Phi}\rangle, \tag{28}$$

when $n \to \infty$. Define the limit matrix to be $\boldsymbol{A}_0$.

Property 4.1. *Under the assumptions of Property 3.3, it holds that* $\mathbb{E}[\widehat{A}_{ij}] = A_{ij}$ *at any sample size and* $\widehat{A}_{ij} \xrightarrow{p} A_{0,ij}$ *as* $n \to \infty$.

The estimation of $\boldsymbol{b}$ is more complicated since $V(\widehat{\boldsymbol{\Phi}})$ and $C(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{T}}_m)$ cannot simply be replaced by their sample realisations. An entry b_m of $\boldsymbol{b}$ can be written as

$$b_m = \sum_{i,j,k=1}^{p} \left(\mathrm{Var}(\widehat{\phi}_{ijk}) - \mathrm{Cov}(\widehat{\phi}_{ijk}, \widehat{t}_{m,ijk}) \right), \tag{29}$$

and hence, under the assumptions of Property 3.1 on $\widehat{\boldsymbol{T}}_m$, $n\boldsymbol{b} \to \boldsymbol{c}_0$ as $n \to \infty$, where

$$c_{0,m} = \sum_{i,j,k=1}^{p} \left(\mathrm{AVar}(\sqrt{n}\widehat{\phi}_{ijk}) - \mathrm{ACov}(\sqrt{n}\widehat{\phi}_{ijk}, \sqrt{n}\widehat{t}_{m,ijk}) \right), \tag{30}$$

with AVar and ACov denoting the asymptotic variance and asymptotic covariance. Hence, $b_m = O(n^{-1})$ and a consistent estimator for $\boldsymbol{c}_0$ is required instead of any sequence of order $O(n^{-1})$ to estimate b_m.

In Section 4.1 we first describe the traditional approach of using plug-in estimators. Since this approach yields biased and non-consistent estimators for $\boldsymbol{c}_0$ we recommend as an alternative to use estimators based on k-statistics and polykays, which we propose in Section 4.2.

4.1 Plug-in Estimation of $\boldsymbol{b}$

Plug-in estimation of the required quantities is the standard in the shrinkage literature. The principle is to replace each expectation by a sample average. An intuitive estimator for $V(\widehat{\boldsymbol{\Phi}})$ is given in Martellini & Ziemann (2010):

$$\widehat{V}^{\mathrm{pl}}(\widehat{\boldsymbol{\Phi}}) = \frac{1}{n^2} \sum_{l=1}^{n} \left\| \widehat{\boldsymbol{\Phi}}_l - \widehat{\boldsymbol{\Phi}}^{\mathrm{pl}} \right\|^2, \tag{31}$$

where $\widehat{\boldsymbol{\Phi}}_l, l = 1, \ldots, n$, is defined by

$$\widehat{\boldsymbol{\Phi}}_l = (\boldsymbol{x}_l - \overline{\boldsymbol{x}})(\boldsymbol{x}_l - \overline{\boldsymbol{x}})' \otimes (\boldsymbol{x}_l - \overline{\boldsymbol{x}})'. \tag{32}$$

This estimator is intuitive because it resembles the sample variance estimator.

Property 4.2. *When $\boldsymbol{X}$ has finite sixth order moments, it holds that*

$$\lim_{n\to\infty} nV^{pl}(\widehat{\boldsymbol{\Phi}}) = \sum_{i,j,k=1}^{p} \mathrm{AVar}(\sqrt{n}\widehat{\phi}_{ijk}), \quad i,j,k = 1,\dots,p, \tag{33}$$

where

$$\begin{aligned}\mathrm{AVar}(\sqrt{n}\widehat{\phi}_{ijk}) = m_{2,2,2} &- m_{1,1,1}^2 - 2m_{2,1,1}m_{0,1,1} - 2m_{1,2,1}m_{1,0,1} - 2m_{1,1,2}m_{1,1,0} \\ &+ m_{2,0,0}m_{0,1,1}^2 + m_{0,2,0}m_{1,0,1}^2 + m_{0,0,2}m_{1,1,0}^2 + 6m_{0,1,1}m_{1,0,1}m_{1,1,0},\end{aligned} \tag{34}$$

and $m_{u,v,w}$ denotes the central moment, $m_{u,v,w} = \mathbb{E}[(X_i-\mu_i)^u(X_j-\mu_j)^v(X_k-\mu_k)^w]$.

Since $n\widehat{V}^{\mathrm{pl}}(\widehat{\boldsymbol{\Phi}}) \xrightarrow{a.s.} m_{2,2,2} - m_{1,1,1}^2$, as $n \to \infty$ and the extra terms in (34) are not zero in general, the plug-in estimator is not consistent.

For a target $\widehat{\boldsymbol{T}}^\star$ as in (11), it can be shown that

$$\mathbb{E}\left[\langle \widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}, \widehat{\boldsymbol{T}}^\star - \mathbb{E}[\widehat{\boldsymbol{T}}^\star]\rangle\right] = \mathbb{E}\left[\|\widehat{\boldsymbol{T}}^\star - \mathbb{E}[\widehat{\boldsymbol{T}}^\star]\|^2\right] = V(\widehat{\boldsymbol{T}}^\star), \tag{35}$$

The corresponding plug-in estimator, obtained by replacing expectations with sample averages, is given by

$$\widehat{V}^{\mathrm{pl}}(\widehat{\boldsymbol{T}}^\star) = \frac{1}{n^2}\sum_{l=1}^{n}\left\|\widehat{\boldsymbol{T}}_l^\star - \frac{(n-1)(n-2)}{n^2}\widehat{\boldsymbol{T}}^\star\right\|^2, \tag{36}$$

where $\widehat{\boldsymbol{T}}_l^\star$ is given by

$$\widehat{\boldsymbol{T}}_l^\star = \sum_{q=1}^{Q}\frac{\langle \boldsymbol{E}_q, \widehat{\boldsymbol{\Phi}}_l\rangle}{\|\boldsymbol{E}_q\|^2}\boldsymbol{E}_q. \tag{37}$$

Combining estimators (31) and (36) yields an estimate for b_m whenever the m-th target is of form (11).

4.2 Unbiased Estimation of $\boldsymbol{b}$

4.2.1 Estimation of $V(\widehat{\boldsymbol{\Phi}})$

Unbiased estimators for both $V(\widehat{\boldsymbol{\Phi}})$ and $C(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{T}}_m^\star)$ are obtained by element-wise estimation of $\mathrm{Var}(\widehat{\phi}_{ijk})$ and $\mathrm{Cov}(\widehat{\phi}_{ijk}, \widehat{t}_{m,ijk})$ as in (29). For $\widehat{\phi}_{iii}, i = 1,\dots,p$, the variance of the estimator is given by

$$\mathrm{Var}(\widehat{\phi}_{iii}) = \frac{1}{n}\kappa_6 + \frac{1}{n-1}\left(9\kappa_4\kappa_2 + 9\kappa_3^2 + \frac{6n}{n-2}\kappa_2^3\right), \tag{38}$$

where κ_m denotes the m-th cumulant of the distribution of $\boldsymbol{X}_i$. For ease of notation we omit the reference to i. Each of the terms in the right hand side of (38) is estimated by the corresponding unbiased estimator, as presented in Di Nardo et al. (2008, 2009). Combining k-statistics and polykays, the unbiased estimator for (38) is

$$\widehat{\mathrm{Var}}(\widehat{\phi}_{iii}) = c_{1,1}S_6 + c_{1,2}S_4S_2 + c_{1,3}S_3^2 + c_{1,4}S_2^3, \quad \text{where} \quad S_m = \sum_{l=1}^{n}(x_{li} - \overline{\boldsymbol{x}}_i)^m. \tag{39}$$

Again, the reference to component i in S_m is omitted for ease of notation. The constants $c_{1,1}, \ldots, c_{1,4}$ can be derived from the formula for the estimator $\widehat{\mathrm{Var}}(\widehat{\phi}_{ijk})$, presented in Appendix B. The resulting unbiased estimator for $V(\widehat{\boldsymbol{\Phi}})$ is

$$\widehat{V}(\widehat{\boldsymbol{\Phi}}) = \sum_{i,j,k=1}^{p} \widehat{\mathrm{Var}}(\widehat{\phi}_{ijk}). \tag{40}$$

In Section 5 we show empirically how this estimator is an improvement over the biased and inconsistent plug-in estimator given in the previous section.

4.2.2 Estimation of $C(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{T}}^{\star})$

Estimation of $C(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{T}})$ has to be considered for each target coskewness matrix individually. In this section we provide unbiased estimators when the target is as in (11). The supplementary appendix contains consistent estimators for the other structured coskewness matrices mentioned in Section 2.3. There we also correct the estimators given in Martellini & Ziemann (2010) for the single-factor and constant correlation coskewness matrices.

For any target $\widehat{\boldsymbol{T}}^{\star}$, an unbiased estimator for $C(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{T}}^{\star})$ can be constructed using multivariate k-statistics and polykays. Since $\boldsymbol{T}_1^{\star}$ is deterministic, $C(\widehat{\boldsymbol{\Phi}}, \boldsymbol{T}_1^{\star}) = 0$. For target $\widehat{\boldsymbol{T}}_2^{\star}$, with common third order central moments it holds that

$$C(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{T}}_2^{\star}) = \frac{1}{p}\left(\sum_{i=1}^{p}\mathrm{Var}(\widehat{\phi}_{iii}) + \sum_{i=1}^{p}\sum_{j\neq i}^{p}\mathrm{Cov}(\widehat{\phi}_{iii}, \widehat{\phi}_{jjj})\right) \tag{41}$$

and for $\widehat{\boldsymbol{T}}_3^{\star}$ the expression is

$$C(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{T}}_3^{\star}) = \sum_{i=1}^{p}\mathrm{Var}(\widehat{\phi}_{iii}). \tag{42}$$

An estimator for the terms $\mathrm{Cov}(\widehat{\phi}_{iii}, \widehat{\phi}_{jjj})$ using k-statistics and polykays is provided in Appendix B.

Combining the estimators $\widehat{\mathrm{Var}}(\widehat{\phi}_{ijk})$ and $\widehat{\mathrm{Cov}}(\widehat{\phi}_{ijk}, \widehat{t}_{m,ijk})$, an unbiased and consistent estimator $\widehat{\boldsymbol{b}}$ is obtained.

Property 4.3. *For target matrices as in* (11) *it holds that* $n\widehat{\boldsymbol{b}} \xrightarrow{a.s.} \boldsymbol{c}_0$ *as* $n \to \infty$ *and, at any sample size* n*, it holds that* $\mathbb{E}[\widehat{\boldsymbol{b}}] = \boldsymbol{b}$.

Hence, due to Properties 4.1 and 4.3 it holds that the MSE loss function determining $\widehat{\boldsymbol{\lambda}}^\star$ is consistently estimated. Thus $\widehat{\boldsymbol{\Phi}}^{\mathrm{MT}}(\widehat{\boldsymbol{\lambda}}^\star) \xrightarrow{p} \boldsymbol{\Phi}$, as $n \to \infty$, which includes the single-target shrinkage estimator as a special case.

5 Simulation Study

5.1 Set-up

The simulation set-up is chosen such that the simulated data shows the same characteristics for variance, skewness and kurtosis as observed in the dataset with monthly hedge fund returns used in Section 6. In a similar fashion as Fan et al. (2008), we estimate the multi-factor model

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{F} + \boldsymbol{\varepsilon} \tag{43}$$

on the 100 funds of strategies Equity Hedge, Macro, Relative Value and Event-Driven according to Hedge Fund Research with most Assets Under Management (AUM) in December 2010. As factors, the corresponding HFRI indices (available on www.hedgefundresearch.com) are used.

As in Jondeau & Rockinger (2012), the factors are first standardized and then fitted by a skew-t distribution with density

$$f(x;\nu,\xi) = \frac{2b}{\xi + \frac{1}{\xi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{\kappa^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \tag{44}$$

where $\kappa = (\beta x + \alpha)\xi$ if $\beta x + \alpha < 0$ and $(\beta x + \alpha)/\xi$ if $\beta x + \alpha \geq 0$. The parameters α and β are such that the distribution has mean zero and unit variance, namely

$$\alpha = \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(\xi + \frac{1}{\xi}\right) \qquad \text{and} \qquad \beta = \sqrt{\xi^2 + \frac{1}{\xi^2} - 1 - \alpha^2}. \tag{45}$$

Moments up to order 6 exist when $\nu > 6$, hence when fitting the distribution to the factors we use a lower bound of $\nu = 7$. Factor loadings and idiosyncratic terms are obtained by least squares regression based on the 60 observations from January 2006 to December 2010. Each of the idiosyncratic terms is also modelled by a skew-t distribution.

We consider the dimensions $p = 5, 10, 20, 30, 50, 100$ and sample sizes $n = 10, 20, 30, 50, 100, 250, 500, 1000$. For p smaller than 100, we subset the data generating process to include only the p funds with largest AUM. For each sample size and dimension, 10 000 samples are generated.

5.2 Results

The focus of our simulation is to measure the accuracy of estimating the skewness of a linear combination of random variables. In our simulations, we take the the sum ($\boldsymbol{v} = \mathbf{1}_p$). Denote by $\widehat{\phi}_{\boldsymbol{v}}^{(\cdot)}$ the estimator of $\phi_{\boldsymbol{v}}^{(\cdot)}$ obtained by replacing in (5) the coskewness matrix $\boldsymbol{\Phi}$ by an estimator $\widehat{\phi}_{\boldsymbol{v}}^{(\cdot)}$. Due to Property 2.2, the multivariate sample estimator yields the same MSE when estimating the skewness of the linearly transformed variable. For the shrinkage estimators this equivalence does not hold.

In Table 1 we measure the improvement of the shrinkage estimators over the sample estimator. This is done using the percentage relative improvement in average loss (PRIAL) frequently used in the shrinkage literature, see e.g. Ledoit & Wolf (2003). This measure shows to which extent the shrinkage estimators outperform the sample coskewness matrix $\widehat{\boldsymbol{\Phi}}$ in terms of MSE, measured as a percentage. The PRIAL of an estimator $\widehat{\phi}_{\boldsymbol{v}}^{(\cdot)}$ for $\phi_{\boldsymbol{v}}$, compared to the sample estimator $\widehat{\phi}_{\boldsymbol{v}}$, is defined by

$$\mathrm{PRIAL}\left(\widehat{\phi}_{\boldsymbol{v}}^{(\cdot)}\right) = \left(\frac{\mathbb{E}\left[\|\widehat{\phi}_{\boldsymbol{v}} - \phi_{\boldsymbol{v}}\|^2\right] - \mathbb{E}\left[\|\widehat{\phi}_{\boldsymbol{v}}^{(\cdot)} - \phi_{\boldsymbol{v}}\|^2\right]}{\mathbb{E}\left[\|\widehat{\phi}_{\boldsymbol{v}} - \phi_{\boldsymbol{v}}\|^2\right]}\right) \times 100\%. \tag{46}$$

Note that the PRIAL of the sample skewness estimator $\widehat{\phi}_{\boldsymbol{v}}$ for $\phi_{\boldsymbol{v}}$ is zero by definition and the PRIAL cannot exceed 100%. A negative value indicates a larger MSE than the MSE of the sample estimator.

Table 1 shows remarkable improvements in the univariate skewness estimate when shrinkage estimators are used. The proposed shrinkage estimators outperform the plug-in versions consistently over a range of dimensions and sample sizes, as indicated by the

Table 1: PRIAL values for univariate skewness estimates using the shrinkage estimators.

		$p=5$		$p=10$		$p=20$		$p=30$		$p=50$		$p=100$	
	n	Pl.	Unb.	Pl.	Unb.	Pl.	Unb.	Pl.	Unb.	Pl.	Unb.	Pl.	Unb.
$\widehat{\mathbf{\Phi}}^{\text{ST1}}$	10	74.91	**97.30**	74.57	**97.38**	73.77	**97.35**	73.91	**97.71**	73.71	**97.66**	73.58	**97.63**
	20	88.87	**94.84**	88.20	**94.38**	87.65	**94.19**	88.02	**94.51**	87.58	**94.18**	87.21	**93.77**
	30	83.53	**84.46**	83.85	**85.52**	82.87	**85.34**	83.47	**85.83**	83.60	**86.31**	83.48	**86.39**
	50	83.10	**84.54**	82.92	**84.68**	80.84	**83.31**	80.79	**83.27**	80.93	**83.48**	80.76	**83.46**
	100	72.64	**76.43**	71.84	**76.00**	69.25	**74.43**	68.85	**74.05**	70.18	**74.98**	70.10	**75.05**
	250	47.47	**56.45**	47.20	**55.87**	46.65	**55.94**	44.12	**54.49**	50.46	**58.13**	52.46	**59.60**
	500	36.29	**45.05**	36.17	**44.38**	36.81	**45.21**	32.00	**41.80**	41.15	**47.71**	42.91	**48.91**
	1000	45.51	**50.06**	46.05	**50.33**	49.78	**53.75**	45.88	**50.89**	52.13	**55.16**	54.52	**57.18**
$\widehat{\mathbf{\Phi}}^{\text{ST2}}$	10	68.20	**95.76**	72.83	**97.30**	73.39	**97.36**	73.62	**97.69**	73.61	**97.66**	73.56	**97.63**
	20	84.26	**93.74**	87.05	**94.34**	87.43	**94.20**	87.77	**94.44**	87.50	**94.18**	87.19	**93.77**
	30	81.22	**83.43**	83.34	**85.53**	82.80	**85.38**	83.21	**85.68**	83.55	**86.31**	83.47	**86.39**
	50	82.30	**83.71**	82.85	**84.67**	80.87	**83.35**	80.55	**83.08**	80.91	**83.47**	80.76	**83.46**
	100	73.97	**76.04**	72.29	**76.03**	69.41	**74.47**	68.54	**73.77**	70.19	**74.97**	70.11	**75.05**
	250	51.20	**56.59**	48.27	**56.01**	46.92	**55.99**	43.66	**54.08**	50.49	**58.13**	52.48	**59.60**
	500	38.90	**44.45**	36.97	**44.32**	36.99	**45.20**	31.59	**41.44**	41.17	**47.70**	42.92	**48.91**
	1000	45.12	**48.07**	46.04	**49.88**	49.77	**53.66**	45.61	**50.63**	52.11	**55.12**	54.52	**57.17**
$\widehat{\mathbf{\Phi}}^{\text{ST3}}$	10	67.87	**95.70**	72.75	**97.28**	73.37	**97.36**	73.55	**97.67**	73.60	**97.65**	73.55	**97.63**
	20	84.02	**93.68**	87.02	**94.33**	87.43	**94.20**	87.77	**94.43**	87.50	**94.18**	87.19	**93.77**
	30	81.05	**83.33**	83.41	**85.54**	82.84	**85.38**	83.36	**85.68**	83.58	**86.31**	83.48	**86.40**
	50	82.32	**83.64**	82.97	**84.68**	80.94	**83.36**	80.80	**83.02**	80.96	**83.47**	80.77	**83.46**
	100	74.62	**76.15**	72.67	**76.08**	69.59	**74.52**	68.96	**73.42**	70.29	**74.96**	70.14	**75.06**
	250	53.37	**57.02**	49.19	**56.16**	47.30	**56.10**	43.68	**53.21**	50.65	**58.16**	52.53	**59.63**
	500	41.31	**44.96**	37.88	**44.50**	37.43	**45.36**	31.81	**41.13**	41.37	**47.80**	42.96	**48.93**
	1000	46.38	**48.27**	46.54	**49.99**	50.01	**53.76**	46.55	**51.16**	52.30	**55.25**	54.55	**57.19**
$\widehat{\mathbf{\Phi}}^{\text{MT}}$	10	74.93	**97.38**	74.57	**97.39**	73.77	**97.36**	73.91	**97.71**	73.71	**97.66**	73.58	**97.63**
	20	88.91	**94.96**	88.21	**94.39**	87.65	**94.19**	88.01	**94.46**	87.57	**94.17**	87.21	**93.76**
	30	83.67	**84.78**	83.87	**85.55**	82.87	**85.34**	83.41	**85.70**	83.59	**86.29**	83.48	**86.39**
	50	83.25	**84.80**	82.94	**84.71**	80.84	**83.31**	80.67	**82.99**	80.91	**83.45**	80.76	**83.46**
	100	72.89	**76.74**	71.86	**76.01**	69.24	**74.41**	68.41	**73.30**	70.15	**74.91**	70.10	**75.04**
	250	47.87	**56.84**	47.23	**55.89**	46.62	**55.92**	42.33	**52.74**	50.37	**58.03**	52.46	**59.59**
	500	36.56	**45.38**	36.18	**44.40**	36.81	**45.21**	30.36	**40.43**	41.10	**47.66**	42.91	**48.91**
	1000	45.63	**50.23**	46.05	**50.35**	49.78	**53.76**	45.38	**50.43**	52.12	**55.14**	54.52	**57.18**

Note: PRIAL values over the sample estimator of the skewness of the sum of $\boldsymbol{X}$ are given for the shrinkage estimators when the data is generated according to a factor model (Section 5.1). Three different single-target estimators are used, as well as the multi-target estimator that uses all three targets. PRIAL values for the plug-in (Pl.) and unbiased (Unb.) estimators are given for sample sizes ranging from $n = 10$ to 1000 and dimensions ranging from $p = 5$ to 100. The PRIAL values are reported in percentage points. The bold indicates the highest PRIAL between the plug-in and proposed estimator, while the background color ranks the PRIAL values for each combination of n and p; the darker, the higher the PRIAL.

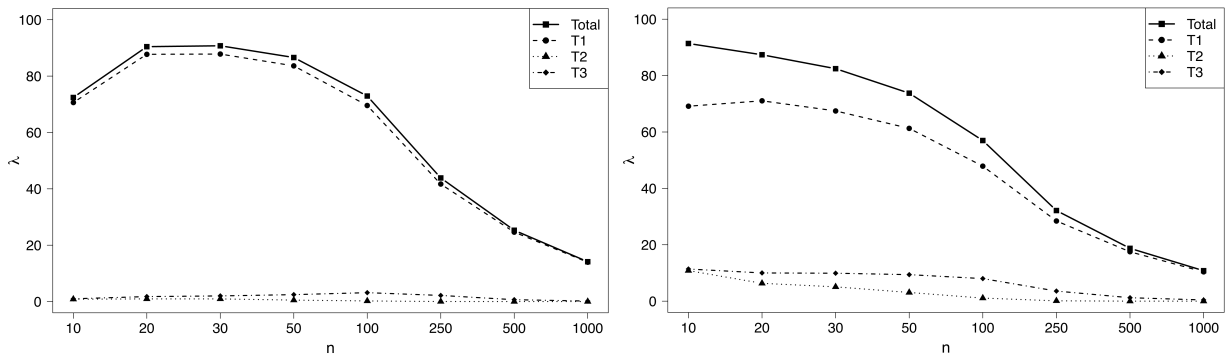
bold highlighting. The improvements are as large as 25 percentage points. Finite sample properties result in reliable estimates when the sample size is small. Because the sample estimator is consistent, it is important to correctly estimate the size of the shrinkage intensity for large sample sizes in order to have a good PRIAL. The proposed shrinkage estimators still offer over 50% reduction in MSE compared to the sample estimator at a sample size of $n = 1000$. PRIAL values measured on the full coskewness matrix instead of on the linear combination show similar results and are given in the supplementary appendix. We remark that when $p > n$, the sample coskewness matrix is misspecified, while the shrinkage estimators are not. Consequently, it is clear that the shrinkage estimators offer remarkable advantages, both in terms of estimation accuracy, as in the lack of misspecification.

A key question is whether the multi-target shrinkage estimator offers additional benefits compared to single-target shrinkage. Table 1 shows that for $p = 5$ and $p = 10$ the PRIAL of the multi-target shrinkage estimator is higher than for each of the single-target shrinkage estimators, indicated by the darker shade of gray. For higher dimensions the PRIAL is either highest or within one percentage point of the highest PRIAL of the individual single-target estimators. This behaviour is to be expected since only p out of p^3 coskewness elements are different between the targets. Hence, for larger dimensions, this results in almost identical targets. Even in this case, multi-target shrinkage offers advantages as it is able to select the 'best' target. The best target is not necessarily the one closest to the true data generating process, but the one that offers the largest reduction in MSE when combined with the sample estimator. The gray shadings in Table 1 indicate that in this simulation setting, target $\boldsymbol{T}_1^\star$ offers the largest reduction in MSE when the sample sizes are small. For larger sample sizes, $\widehat{\boldsymbol{T}}_3^\star$ becomes better. The results based on the full coskewness matrix are even more pronounced in showing the ability of the multi-target estimator to select the best target and realize a high PRIAL.

In Figure 1, the multi-target shrinkage intensity is shown when $p = 50$ for a range of sample sizes. The decomposition into the contribution of each target confirms the findings in PRIAL values that $\boldsymbol{T}_1^\star$ provides a large reduction in MSE when the sample size is small. Note that its influence diminishes rather quickly when sample size increases. This is in contrast with $\widehat{\boldsymbol{T}}_3^\star$ for which the influence decreases more slowly or even stays relatively

Figure 1: Mean shrinkage intensities of the multi-target estimators ($p = 50$).

(a) Plug-in estimator (b) Unbiased estimator



constant, indicating that its relative importance in the contribution of the multi-target shrinkage estimator increases. Hence, $\widehat{\boldsymbol{T}}_3^{\star}$ becomes more informative when sample size increases. Also note the more natural decay of the shrinkage intensity of the proposed estimator which is in strong contrast with the hump shape seen under plug-in estimation.

In the supplementary appendix, we confirm that our estimators for $\boldsymbol{b}$ are unbiased and provide the PRIAL values measured for the full coskewness estimators. These confirm the superiority of the proposed coskewness shrinkage estimators and validate the use of the multi-target estimator. In addition, we provide two other simulation settings. First, the setting of an independent component model with Pareto distributed components and additive noise serves as a second example of a skewed and heavy-tailed distribution for which the targets are misspecified. Second, a Gaussian data-generating process is used to confirm that the estimators perform well even when $\boldsymbol{X}$ is symmetric.

6 Empirical Application

The above simulations have confirmed the statistical gains in accuracy when using the proposed shrinkage estimators. In this section, we further analyse the usefulness of the proposed estimators for dynamic portfolio allocation. We focus on hedge fund returns because they are generally assumed to show more skewness than stock returns. The data consists of monthly returns of hedge funds with the four main strategies (Equity Hedge, Macro, Relative Value and Event-Driven) that have a history of at least 120 consecutive months over the period January 2000 until December 2013. At any point in time, the investment universe consists of the 50 funds with most Assets Under Management.

To account for the potential time-variation in the coskewness matrix, we follow the industry practice of using rolling five-year samples. In this case, shrinkage is clearly needed, as we have only 60 observations to estimate up to 22 100 unique parameters in the coskewness matrix.

Definition of the maximum expected utility portfolio We consider an investor with CRRA preferences with a risk aversion parameter γ. As common in the literature, the utility function is approximated using a third order Taylor expansion, see Harvey et al. (2010) and the supplementary appendix for details. At the end of each month, we construct a portfolio maximizing the expected utility according to the following optimization problem. Given estimates $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Phi}}^{(\cdot)}$ for the covariance and coskewness matrix, the maximum expected utility portfolio has weight vector $\boldsymbol{w} \in \mathbb{R}^{p\times 1}$ given by

$$\begin{aligned} &\underset{\boldsymbol{w}}{\text{maximize}} \quad -\frac{\gamma}{2}\boldsymbol{w}'\widehat{\boldsymbol{\Sigma}}\boldsymbol{w} + \frac{\gamma(\gamma+1)}{6}\boldsymbol{w}'\widehat{\boldsymbol{\Phi}}^{(\cdot)}(\boldsymbol{w}\otimes\boldsymbol{w}) \\ &\text{subject to} \quad 0 \leq w_i \leq 0.2, \quad i = 1,\dots,p, \quad \text{and} \quad \sum_{i=1}^{p} w_i = 1, \end{aligned} \tag{47}$$

where we impose a full-investment constraint combined with a no short selling constraint. The maximum weight of 20% is set to keep the portfolios somewhat diversified. We consider the impact of the choice of estimator $\widehat{\boldsymbol{\Phi}}^{(\cdot)}$, keeping $\widehat{\boldsymbol{\Sigma}}$ fixed to the shrinkage estimator of Ledoit & Wolf (2003) with a diagonal target covariance matrix. Details are discussed in the supplementary appendix.

Results We consider a risk averse investor with $\gamma = 15, 25, 50$. For the latter, the role of skewness in (47) is more pronounced. The equal-weighted portfolio and the maximum expected utility portfolio constructed using the sample coskewness estimator are used as a benchmark. They are compared to the maximum expected utility portfolios based on the proposed single-target shrinkage estimators with target coskewness matrices $\boldsymbol{T}_1^\star$, $\boldsymbol{T}_2^\star$ and $\boldsymbol{T}_3^\star$, and the multi-target estimator with targets $(\boldsymbol{T}_1^\star, \boldsymbol{T}_2^\star, \boldsymbol{T}_3^\star)$. The shrinkage intensities vary between 0.62 and 1, with a median of about 0.83, for all the coskewness shrinkage estimators.

Table 2: Out-of-sample performance of the portfolios.

	γ	EW	Sample	ST1	ST2	ST3	MT
MUG relative to EW (10^{-2})	15	0	1.868	1.896	**1.938**	1.888	1.892
	25	0	4.713	4.751	**4.836**	4.737	4.744
	50	0	13.375	16.350	**16.400**	16.312	16.337
MUG relative to S (10^{-2})	15	-1.834	0	0.027	**0.069**	0.019	0.023
	25	-4.501	0	0.036	**0.117**	0.023	0.029
	50	-11.797	0	2.624	**2.668**	2.590	2.613
Ann. Geom. Mean (10^{-2})	15	**5.775**	4.633	4.660	4.706	4.653	4.657
	25	**5.775**	4.625	4.658	4.753	4.646	4.653
	50	**5.775**	4.742	4.651	4.794	4.628	4.651
Ann. Standard Deviation (10^{-2})	15	6.395	2.315	**2.310**	2.318	2.313	2.313
	25	6.395	2.320	**2.311**	2.333	2.314	2.315
	50	6.395	3.843	**2.312**	2.396	2.319	2.319
Skewness (10^{-7})	15	-81.477	-5.132	-5.144	**−5.005**	-5.148	-5.152
	25	-81.477	-5.145	-5.148	**−5.048**	-5.155	-5.161
	50	-81.477	-11.746	**−5.160**	-5.286	-5.176	-5.185
Ann. Turnover	15	**0.813**	1.341	1.341	1.418	1.339	1.375
	25	**0.813**	1.346	1.340	1.407	1.338	1.381
	50	**0.813**	3.032	1.341	1.476	1.338	1.405
Mean Herfindahl Index	15	**0**	11.650	11.662	11.320	11.636	11.670
	25	**0**	11.637	11.660	11.167	11.622	11.662
	50	**0**	13.161	11.659	10.793	11.590	11.628

Note: the table presents following out-of-sample performance measures for the different portfolios: Monetary Utility Gain, annualized geometric mean and standard deviation, skewness (non-standardized), annualized turnover and the mean Herfindahl index (see (48)). The equal-weighted portfolio is denoted by EW while the maximum expected utility portfolios estimated using the sample coskewness matrix is denoted by Sample. The shrinkage-based portfolios are denoted by ST1, ST2 and ST3 and MT. Results are given for risk aversion parameters $\gamma = 15, 25, 50$. The portfolio doing best on each individual performance measure is highlighted in bold.

The relative performance of the portfolios is evaluated using the Monetary Utility Gain (MUG) as in Ang & Bekaert (2002) and Martellini & Ziemann (2010). The MUG of an alternative portfolio is equal to the monetary payment required by an investor in the benchmark portfolio so that she is indifferent to changing to the alternative portfolio.

Table 2 shows the MUG relative to the equal-weighted portfolio and the maximum expected utility portfolio estimated using the sample coskewness matrix. In addition, we show the first three realized moments, the annualized turnover and the mean Herfindahl index. The Herfindahl index measures the concentration of a portfolio and is defined as

$$\text{Mean Herfindahl Index} = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\sum_{j=1}^{p} w_{t,j}^2 - \frac{1}{p}}{1-\frac{1}{p}}\right)\times 100\%. \tag{48}$$

It is equal to zero for the equal-weighted portfolio, which is most diversified and is maximal at 100% when all capital is invested in a single fund.

Since all MUG values relative to the equal-weighted portfolio are positive, the maximum expected utility portfolios outperform the equal-weighted portfolio in terms of delivering a higher out-of-sample realized utility. If the coefficient of risk aversion γ is increased, then the investor attaches more importance to skewness and we observe a stronger preference for the optimized portfolios. The positive MUG values of the coskewness shrinkage based portfolios relative to the portfolio optimized using the sample coskewness estimator indicate that an investor prefers the more sophisticated approach of shrinkage to estimate the coskewness matrix. Her preferences become stronger when her coefficient of risk aversion increases.

The price paid for switching from the equal-weighted portfolio to the optimized portfolios is a decrease in annual return by 1%. For this, the investor receives a lower annual standard deviation and a higher skewness. In addition, the turnover of the optimized portfolios rises and the portfolio becomes more concentrated.

The detrimental effect of estimation error on the sample coskewness estimator is clearly visible when $\gamma = 50$; a very large negative skewness, a larger realized standard deviation, increased turnover and more concentrated portfolios. This is in strong contrast to the stability of the shrinkage based portfolios, even for large coefficients of risk aversion.

Overall, the portfolios optimized using the proposed shrinkage estimators for the coskewness matrix are preferable to the equal-weighted portfolio, as well as the portfolio optimized using the sample coskewness estimator. Robustness checks where we investigate the influence of the time-window and the maximal allocation to a single fund are discussed in the supplementary appendix. We show that a change in time-window or maximal weight does not affect the general conclusions presented in this section.

7 Conclusion

Many economic decisions involve the evaluation or optimization of a higher order approximation of the expected utility function or the density functions of skewed random variables. The quality of those decisions heavily depends on the accuracy of the estimates of the corresponding coskewness matrix.

The main message of this paper is that, for the estimation of the skewness of linear combinations of random variables, one should consider a multivariate approach using a shrinkage estimate of the coskewness matrix, where the MSE loss function is estimated unbiasedly. Our simulations show that the unbiased estimators for the MSE loss function improve the finite sample behaviour of the shrinkage estimator by up to 25 percentage points in terms of MSE. We further contribute by extending the methodology to accommodate the use of multiple targets and derive the optimal targets for a given coskewness structure.

In the empirical application, we show that there is an economic incentive for an investor with CRRA preferences to invest in the maximum expected utility portfolios constructed by estimating skewness using the proposed shrinkage coskewness estimators. This empirical evidence is in line with our simulation results about the gains in accuracy when estimating the skewness of linear combinations of random variables using a shrinkage-based estimate of the full coskewness matrix. In particular, a risk averse CRRA investor requires a payment of up to 2.7% annually in order to remain invested in the portfolio optimized using the sample estimator instead of applying the proposed shrinkage estimators

We think our paper may open up a number of interesting research avenues to further improve the coskewness estimates. A first outstanding research question is to use the multivariate approach to define robust estimates of skewness, as an alternative to the quantile and medcouple approaches in Kim & White (2004) and Brys et al. (2004) respectively. A second question is to generalize the linear shrinkage structure used in the coskewness estimation to a non-linear shrinkage methodology as in Ledoit & Wolf (2012). Finally, based on the evidence in Jondeau & Rockinger (2012) for time-variation in the higher order comoments, it would be interesting to develop a conditional approach to the shrinkage estimation of the coskewness matrix, as was done by Hafner & Reznikova (2012) in the case of the estimation of dynamic conditional covariances.

Acknowledgements

The authors gratefully acknowledge support from the Research Foundation - Flanders (FWO research grant $G023815N$ and PhD fellowship $1114117N$), the Internal Funds KU Leuven (project $C16/15/068$) and Google Summer of Code 2017. We thank the Editor (Todd Clark), two anonymous referees, and participants at various conferences and seminars for their valuable comments. We are especially grateful to Christophe Croux, Siem Jan Koopman and Brian G. Peterson for stimulating discussions.

SUPPLEMENTARY MATERIAL

Supplementary appendix The appendix contains additional examples of structured estimators as well as extra simulation studies and a more in-depth discussion of the empirical application. At the end, we present a short tutorial for using the proposed shrinkage estimators in `R`. (.pdf file)

A Proofs

Proof of Property 2.1. Existence of third order moments ensures the existence of all variables in this proof. For the plug-in coskewness estimator $\widehat{\phi}^{\text{pl}}_{ijk}$ it holds that

$$\begin{aligned}\widehat{\phi}^{\text{pl}}_{ijk} = \frac{1}{n}\sum_{m=1}^{n}(x_{mi}-\mu_i)(x_{mj}-\mu_j)(x_{mk}-\mu_k) - (\overline{x}_i-\mu_i)\,\sigma_{jk} - (\overline{x}_j-\mu_j)\,\sigma_{ik} \\ - (\overline{x}_k-\mu_k)\,\sigma_{ij} - (\overline{x}_i-\mu_i)\,(\widehat{\sigma}_{jk}-\sigma_{jk}) - (\overline{x}_j-\mu_j)\,(\widehat{\sigma}_{ik}-\sigma_{ik}) \\ - (\overline{x}_k-\mu_k)\,(\widehat{\sigma}_{ij}-\sigma_{ij}) + 2\,(\overline{x}_i-\mu_i)\,(\overline{x}_j-\mu_j)\,(\overline{x}_k-\mu_k)\,,\end{aligned} \tag{49}$$

with $\sigma_{ij} = \text{Cov}(X_i, X_j)$ and $\widehat{\sigma}_{ij} = n^{-1}\sum_{m=1}^{n}(x_{mi}-\mu_i)(x_{mj}-\mu_j)$. By the Strong Law of Large Numbers and Application D of Corollary 1.7 in Serfling (2009), it follows that $\widehat{\phi}^{\text{pl}}_{ijk} \xrightarrow{a.s.} \phi_{ijk}$, as $n \to \infty$. Since $n^2/((n-1)(n-2)) \to 1$ when $n \to \infty$, the same result holds for the unbiased sample estimator. Taking the expected value of the right hand side of Equation (49), it holds that

$$\mathbb{E}[\widehat{\phi}^{\text{pl}}_{ijk}] = \phi_{ijk} - \frac{3}{n}\phi_{ijk} + \frac{2}{n}\phi_{ijk} = \frac{(n-1)(n-2)}{n^2}\phi_{ijk}, \tag{50}$$

and hence the sample estimator $\widehat{\phi}_{ijk}$ is unbiased.

Proof of Property 2.2. For any $\boldsymbol{v} \in \mathbb{R}^p$ we reformulate the expression of $\widehat{\phi}_{\boldsymbol{v}}$ in (9) to

$$\begin{aligned}
\widehat{\phi}_{\boldsymbol{v}} &= \frac{n}{(n-1)(n-2)} \sum_{l=1}^{n} \left(\sum_{i=1}^{p} (v_i x_{li} - v_i \overline{x}_i) \right)^3 \\
&= \frac{n}{(n-1)(n-2)} \sum_{l=1}^{n} \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} (v_i x_{li} - v_i \overline{x}_i)(v_j x_{lj} - v_j \overline{x}_j)(v_k x_{lk} - v_k \overline{x}_k) \qquad (51) \\
&= \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} v_i v_j v_k \widehat{\phi}_{ijk} = \boldsymbol{v}' \widehat{\boldsymbol{\Phi}} (\boldsymbol{v} \otimes \boldsymbol{v}).
\end{aligned}$$

Proof of Property 2.3. Observe that $\widehat{\boldsymbol{T}}^\star$ is a linear function of $\widehat{\boldsymbol{\Phi}}$. Hence, by Property (2.1) and Theorem 1.7 in Serfling (2009) it holds that $\widehat{\boldsymbol{T}}^\star \xrightarrow{a.s.} \boldsymbol{T}^\star$ as $n \to \infty$ and $\mathbb{E}[\widehat{\boldsymbol{T}}^\star] = \boldsymbol{T}^\star$.

Proof of Property 3.1. If $\boldsymbol{U} = \boldsymbol{\Phi}$, then for any sequence $\lambda^\star$, $\widehat{\boldsymbol{\Phi}}^{\mathrm{ST}}(\lambda) \xrightarrow{p} \boldsymbol{\Phi}$ as $n \to \infty$. Assume that $\boldsymbol{U} \neq \boldsymbol{\Phi}$, then $A \to A_0 > 0$. By the Cauchy-Schwartz inequality, $b = O(n^{-1})$, and thus $\lambda^\star = O(n^{-1})$. Hence, because $\widehat{\boldsymbol{\Phi}}$ is consistent it holds that $\widehat{\boldsymbol{\Phi}}^{\mathrm{ST}}(\lambda) \xrightarrow{p} \boldsymbol{\Phi}$ as $n \to \infty$.

Proof of Property 3.2. The gradient of the MSE loss function (19) with respect to $(\nu_1, \ldots, \nu_Q)'$ has entries

$$\frac{\partial}{\partial \nu_q} L = 2\lambda^2 \, \mathbb{E}\left[\langle \boldsymbol{E}_q, \nu_q \boldsymbol{E}_q - \widehat{\boldsymbol{\Phi}} \rangle \right], \qquad q = 1, \ldots, Q. \qquad (52)$$

Solving the first order conditions and checking the second order conditions yields the optimal solution $\nu_q = \langle \boldsymbol{E}_q, \boldsymbol{\Phi} \rangle / \|\boldsymbol{E}_q\|^2$, which are the coefficients in $\boldsymbol{T}^\star$.

Proof of Property 3.3. By definition of the quadratic program (23) and the restriction of $\boldsymbol{\lambda}^\star$ to the convex set (21) it holds that

$$\mathbb{E}\left[\|\widehat{\boldsymbol{\Phi}}^{\mathrm{MT}}(\boldsymbol{\lambda}^\star) - \boldsymbol{\Phi}\|^2 \right] \leq \min_{m=1,\ldots,t} \mathbb{E}\left[\|\widehat{\boldsymbol{\Phi}}^{\mathrm{ST}}(\lambda_m^\star) - \boldsymbol{\Phi}\|^2 \right], \qquad (53)$$

with $\widehat{\boldsymbol{\Phi}}^{\mathrm{ST}}(\lambda_m^\star)$ the single-target shrinkage estimator with target $\widehat{\boldsymbol{T}}_m$ and $\lambda_m^\star$ the optimal shrinkage intensity for this estimator. The proof of Property 3.1 implies that

$$\mathbb{E}\left[\|\widehat{\boldsymbol{\Phi}}^{\mathrm{ST}}(\lambda_m^\star) - \boldsymbol{\Phi}\|^2 \right] \to 0, \qquad \text{as } n \to \infty. \qquad (54)$$

Hence,

$$\mathbb{E}\left[\|\widehat{\boldsymbol{\Phi}}^{\mathrm{MT}}(\boldsymbol{\lambda}^\star) - \boldsymbol{\Phi}\|^2 \right] \to 0, \qquad \text{as } n \to \infty, \qquad (55)$$

implying that $\widehat{\boldsymbol{\Phi}}^{\mathrm{MT}}(\boldsymbol{\lambda}^\star) \xrightarrow{p} \boldsymbol{\Phi}$ as $n \to \infty$.

Proof of Property 4.1. Due to Property 2.1 and the consistency assumption for the estimators of the target matrices, it follows that, by Theorem 1.7 and Application D of Corollary 1.7 in Serfling (2009), $\widehat{A}_{ij} \xrightarrow{p} A_{0,ij}$, as $n \to \infty$. The estimator is unbiased by definition, since $\mathbb{E}[\widehat{A}_{ij}] = \mathbb{E}[\langle \widehat{\boldsymbol{T}}_i - \widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{T}}_j - \widehat{\boldsymbol{\Phi}} \rangle] = A_{ij}$.

Proof of Property 4.2. Multiplying equation (49) with $\sqrt{n}$ and applying Slutsky's lemma to the last four terms, it holds that

$$\begin{aligned}\sqrt{n}\widehat{\phi}_{ijk}^{\text{pl}} = \sqrt{n}\Bigg(\frac{1}{n}\sum_{m=1}^{n}(x_{mi}-\mu_i)(x_{mj}-\mu_j)(x_{mk}-\mu_k) \\ - (\overline{x}_i - \mu_i)\,\sigma_{jk} - (\overline{x}_j - \mu_j)\,\sigma_{ik} - (\overline{x}_k - \mu_k)\,\sigma_{ij}\Bigg) + o_P(1),\end{aligned} \tag{56}$$

with $o_P(1)$ a term converging in probability to zero. The right hand side is a mean over independent and identically distributed observations with a variance that is equal to (34).

Proof of Property 4.3. The estimators $\widehat{\text{Var}}(\widehat{\phi}_{ijk})$ and $\widehat{\text{Cov}}(\widehat{\phi}_{iii}, \widehat{\phi}_{jjj})$ needed to construct $\widehat{\boldsymbol{b}}$ can be found in Appendix B. Since $n^{-1}S_{u,v,w} \xrightarrow{a.s.} m_{u,v,w}$ as $n \to \infty$, it follows that

$$n\widehat{\text{Var}}(\widehat{\phi}_{ijk}) \xrightarrow{a.s.} \text{AVar}(\sqrt{n}\widehat{\phi}_{ijk}), \qquad \text{as} \quad n \to \infty, \tag{57}$$

with $\text{AVar}(\sqrt{n}\widehat{\phi}_{ijk})$ given in Equation (34). Analogously,

$$n\widehat{\text{Cov}}(\widehat{\phi}_{iii}, \widehat{\phi}_{jjj}) \xrightarrow{a.s.} \text{ACov}(\sqrt{n}\widehat{\phi}_{iii}, \sqrt{n}\widehat{\phi}_{jjj}), \qquad \text{as} \quad n \to \infty. \tag{58}$$

Hence, it holds that $n\widehat{\boldsymbol{b}} \xrightarrow{a.s.} \boldsymbol{c}_0$ as $n \to \infty$ and by the properties of polykays and k-statistics that at any sample size n, $\mathbb{E}[\widehat{\boldsymbol{b}}] = \boldsymbol{b}$.

B Unbiased Estimators

Here we provide complete formulas for the $\widehat{\text{Var}}(\widehat{\phi}_{ijk})$ and $\widehat{\text{Cov}}(\widehat{\phi}_{iii}, \widehat{\phi}_{jjj})$ needed to estimate $n\boldsymbol{b}$ consistently and unbiasedly. The formulas for the variances of the sample coskewness estimators in terms of multivariate cumulants can be found in Stuart & Ord (1994). The framework developed in Di Nardo et al. (2008, 2009) is used to construct the unbiased estimators.

The variance of the sample coskewness

$$\begin{aligned}\operatorname{Var}(\widehat{\phi}_{ijk}) = \frac{1}{n}\kappa_{2,2,2} + \frac{1}{n-1}\Bigg(&\kappa_{2,0,0}\kappa_{0,2,2} + \kappa_{0,2,0}\kappa_{2,0,2} + \kappa_{0,0,2}\kappa_{2,2,0} + 2\kappa_{1,1,0}\kappa_{1,1,2} \\ &+ 2\kappa_{1,0,1}\kappa_{1,2,1} + 2\kappa_{0,1,1}\kappa_{2,1,1} + 3\kappa_{1,1,1}^2 + 2\kappa_{2,1,0}\kappa_{0,1,2} + 2\kappa_{1,2,0}\kappa_{1,0,2} + 2\kappa_{2,0,1}\kappa_{0,2,1} \\ &+ \frac{n}{n-2}\big(\kappa_{2,0,0}\kappa_{0,2,0}\kappa_{0,0,2} + \kappa_{2,0,0}\kappa_{0,1,1}^2 + \kappa_{1,1,0}^2\kappa_{0,0,2} + \kappa_{1,0,1}^2\kappa_{0,2,0} + 2\kappa_{1,1,0}\kappa_{0,1,1}\kappa_{1,0,1}\big)\Bigg),\end{aligned} \tag{59}$$

where $\kappa_{u,v,w}$ denotes the three-dimensional cumulant of order (u, v, w) of the random vector $(\boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_k)$. An unbiased estimator is given by

$$\begin{aligned}\widehat{\operatorname{Var}}(\widehat{\phi}_{ijk}) = {}& c_1 S_{2,2,2} + c_2(S_{2,2,0}S_{0,0,2} + S_{2,0,2}S_{0,2,0} + S_{0,2,2}S_{2,0,0}) + c_3(S_{2,1,1}S_{0,1,1} \\ &+ S_{1,2,1}S_{1,0,1} + S_{1,1,2}S_{1,1,0}) + c_4(S_{0,2,1}S_{2,0,1} + S_{0,1,2}S_{2,1,0} + S_{1,0,2}S_{1,2,0}) + c_5 S_{1,1,1}^2 \\ &+ c_6 S_{0,0,2}S_{0,2,0}S_{2,0,0} + c_7(S_{2,0,0}S_{0,1,1}^2 + S_{0,2,0}S_{1,0,1}^2 + S_{0,0,2}S_{1,1,0}^2) + c_8 S_{0,1,1}S_{1,0,1}S_{1,1,0},\end{aligned} \tag{60}$$

where $S_{u,v,w} = \sum_{l=1}^{n}(x_{li} - \overline{\boldsymbol{x}}_i)^u(x_{lj} - \overline{\boldsymbol{x}}_j)^v(x_{lk} - \overline{\boldsymbol{x}}_k)^w$ and
$c_1 = \alpha(n^6 - 5n^5 + 13n^4 - 23n^3 + 22n^2 - 8n)$, $c_2 = \alpha(-n^4 + 4n^3 - 9n^2 + 14n - 8)$,
$c_3 = \alpha(-2n^5 + 12n^4 - 18n^3 - 16n^2 + 56n - 32)$, $c_4 = \alpha(-2n^4 + 8n^3 - 2n^2 - 20n + 16)$,
$c_5 = \alpha(-n^5 + 2n^4 + 17n^3 - 34n^2 - 40n + 32)$, $c_6 = \alpha(4n^2 - 12n + 8)$,
$c_7 = \alpha(n^4 - 8n^3 + 25n^2 - 34n + 16)$ and $c_8 = \alpha(6n^4 - 48n^3 + 134n^2 - 156n + 64)$,
with $\alpha = (n(n-1)^2(n-2)^2(n-3)(n-4)(n-5))^{-1}$.

The covariance between two estimates of marginal skewness equals

$$\operatorname{Cov}(\widehat{\phi}_{iii}, \widehat{\phi}_{jjj}) = \frac{1}{n}\kappa_{3,3} + \frac{1}{n-1}\left(9\kappa_{2,2}\kappa_{1,1} + 9\kappa_{1,2}\kappa_{2,1} + \frac{6n}{n-2}\kappa_{1,1}^3\right), \tag{61}$$

for which an unbiased estimator is given by

$$\begin{aligned}\widehat{\operatorname{Cov}}(\widehat{\phi}_{iii}, \widehat{\phi}_{jjj}) = {}& c_1 S_{3,3} + c_9 S_{3,0}S_{0,3} + c_{10}S_{2,1}S_{1,2} + c_{11}S_{3,1}S_{0,2} \\ &+ c_{11}S_{1,3}S_{2,0} + c_{12}S_{2,2}S_{1,1} + c_{13}S_{2,0}S_{0,2}S_{1,1} + c_{14}S_{1,1}^3,\end{aligned} \tag{62}$$

where $c_9 = \alpha(-n^5 + 5n^4 + 5n^3 - 31n^2 - 10n + 8)$, $c_{10} = \alpha(-9n^4 + 36n^3 - 9n^2 - 90n + 72)$,
$c_{11} = \alpha(-3n^5 + 21n^4 - 39n^3 + 3n^2 + 42n - 24)$, $c_{12} = \alpha(-9n^4 + 36n^3 - 81n^2 + 126n - 72)$,
$c_{13} = \alpha(9n^4 - 72n^3 + 189n^2 - 198n + 72)$ and $c_{14} = \alpha(24n^2 - 72n + 48)$.

References

Ang, A. & Bekaert, G. (2002), 'International asset allocation with regime shifts', *Review of Financial studies* **15**(4), 1137–1187.

Bartz, D., Höhne, J. & Müller, K.-R. (2014), 'Multi-target shrinkage', *arXiv preprint arXiv:1412.2041* .

Beedles, W. L. (1979), 'On the asymmetry of market returns', *Journal of Financial and Quantitative Analysis* **14**(3), 653–660.

Boudt, K., Lu, W. & Peeters, B. (2015), 'Higher order comoments of multifactor models and asset allocation', *Finance Research Letters* **13**, 225–233.

Boudt, K., Peterson, B. G. & Croux, C. (2008), 'Estimation and decomposition of downside risk for portfolios with non-normal returns', *Journal of Risk* **11**(2), 79–103.

Brys, G., Hubert, M. & Struyf, A. (2004), 'A robust measure of skewness', *Journal of Computational and Graphical Statistics* **13**(4), 996–1017.

Del Brio, E. B., Niguez, T.-M. & Perote, J. (2009), 'Gram–charlier densities: a multivariate approach', *Quantitative Finance* **9**(7), 855–868.

Devijver, E. & Gallopin, M. (2016), 'Block-diagonal covariance selection for high-dimensional gaussian graphical models', *Journal of the American Statistical Association* (forthcoming).

Di Nardo, E., Guarino, G. & Senato, D. (2008), 'A unifying framework for k-statistics, polykays and their multivariate generalizations', *Bernoulli* **14**(2), 440–468.

Di Nardo, E., Guarino, G. & Senato, D. (2009), 'A new method for fast computing unbiased estimators of cumulants', *Statistics and Computing* **19**(2), 155–165.

Engle, R. (2009), *Anticipating correlations: a new paradigm for risk management*, Princeton University Press.

Fan, J., Fan, Y. & Lv, J. (2008), 'High dimensional covariance matrix estimation using a factor model', *Journal of Econometrics* **147**(1), 186–197.

Fisher, R. A. (1929), 'Moments and product moments of sampling distributions', *Proceedings of the London Mathematical Society* **30**(2), 199–238.

Hafner, C. M. & Reznikova, O. (2012), 'On the estimation of dynamic conditional correlation models', *Computational Statistics & Data Analysis* **56**(11), 3533–3545.

Harvey, C. R., Liechty, J. C., Liechty, M. W. & Müller, P. (2010), 'Portfolio selection with higher moments', *Quantitative Finance* **10**(5), 469–485.

Hosking, J. R. (1990), 'L-moments: analysis and estimation of distributions using linear combinations of order statistics', *Journal of the Royal Statistical Society. Series B (Methodological)* **52**(1), 105–124.

Jondeau, E. & Rockinger, M. (2006), 'Optimal portfolio allocation under higher moments', *European Financial Management* **12**(1), 29–55.

Jondeau, E. & Rockinger, M. (2012), 'On the importance of time variability in higher moments for asset allocation', *Journal of Financial Econometrics* **10**(1), 84–123.

Kim, T.-H. & White, H. (2004), 'On more robust estimation of skewness and kurtosis', *Finance Research Letters* **1**(1), 56–73.

Lancewicki, T. & Aladjem, M. (2014), 'Multi-target shrinkage estimation for covariance matrices', *IEEE Transactions on Signal Processing* **62**(24), 6380–6390.

Ledoit, O. & Wolf, M. (2003), 'Improved estimation of the covariance matrix of stock returns with an application to portfolio selection', *Journal of Empirical Finance* **10**(5), 603–621.

Ledoit, O. & Wolf, M. (2004), 'A well-conditioned estimator for large-dimensional covariance matrices', *Journal of Multivariate Analysis* **88**(2), 365–411.

Ledoit, O. & Wolf, M. (2012), 'Nonlinear shrinkage estimation of large-dimensional covariance matrices', *Annals of Statistics* **40**(2), 1024–1060.

Martellini, L. & Ziemann, V. (2010), 'Improved estimates of higher-order comoments and implications for portfolio selection', *Review of Financial Studies* **23**(4), 1467–1502.

Peterson, B. G. & Carl, P. (2016), *PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis*. R package version 1.5.1.
URL: *https://github.com/braverock/PerformanceAnalytics*

Serfling, R. J. (2009), *Approximation theorems of mathematical statistics*, Vol. 162, John Wiley & Sons.

Simaan, Y. (1993), 'Portfolio selection and asset pricing-three parameter framework', *Management Science* **39**(5), 568–577.

Singleton, J. C. & Wingender, J. (1986), 'Skewness persistence in common stock returns', *Journal of Financial and Quantitative Analysis* **21**(3), 335–341.

Stoyanov, S. V., Rachev, S. T. & Fabozzi, F. J. (2013), 'Sensitivity of portfolio VaR and CVaR to portfolio return characteristics', *Annals of Operations Research* **205**(1), 169–187.

Stuart, A. & Ord, J. K. (1994), *Kendall's Advanced Theory of Statistics*, Edward Arnold.