HIGHER-ORDER MOMENT SHRINKAGE FOR RISK ASSESSMENT

Paul Ruelloux[1,2]**, Christian Bongiorno**[1]**, Damien Challet**[1]**,**
[1]Université Paris-Saclay, CentraleSupélec,
Mathématiques et Informatique pour la Complexité et les Systèmes,
91190, Gif-sur-Yvette, France

[2]Barclays Bank Ireland PLC
75008, Paris, France
{name}.{surname}@centralesupelec.fr

ABSTRACT

We study the denoising of higher-order moments of multivariate financial returns using co-moment tensors and show that exploiting these tensors leads to more accurate predictions of realized quantile and tail risk for general portfolios. These tensors exhibit super-symmetry, which enables a specialized Tucker decomposition and allows covariance-filtering techniques to extend naturally to this setting. We propose a cross-validated tensor-filtering heuristic that selects and aggregates core-tensor estimates, with the goal of empirically reducing Frobenius error relative to an out-of-sample reference tensor. For empirical applications, we develop a generalization of the Average Oracle to higher-order co-moment tensors. Using real data, we show that the resulting methods consistently outperform sample estimates in terms of Frobenius-norm error relative to the target tensor, and, in all specifications, yield more accurate quantile and tail risk measurements.

***K*eywords** Higher order moments · Tensor decomposition · Value-at-risk · Cross validation

1 Introduction

Quantitative risk management aims to characterize the distribution of portfolio losses, with particular emphasis on its lower tail, where rare but economically relevant events occur. Tail risk is commonly quantified by Value-at-Risk (VaR) [1, 2] and Conditional Value-at-Risk (CVaR)[3, 4], which capture, respectively, a quantile of losses and the expected loss exceeding that threshold. These quantities characterize rare but severe loss events and play a central role in portfolio construction, risk budgeting, and regulatory capital requirements [5].

In practice, computing VaR and CVaR for large portfolios rests on the modeling and estimation of the joint distribution of asset returns [6]. Ideally, one relies on a multivariate model flexible enough to capture nonlinear dependence, which is then reflected in higher-order co-moments beyond covariance. In large-dimensional multivariate data, however, estimating these higher-order quantities is data-intensive: the number of parameters grows rapidly with the number of assets, while sampling error decreases only slowly as the sample size increases [7, 8]. Extending the estimation window can reduce variance, but in financial markets, it may also introduce bias because return distributions evolve over time [9, 10]. This trade-off motivates moment-based approximations of tail risk that rely on low-order information to improve statistical stability and efficiency.

A common pragmatic compromise is indeed to approximate tail risk using only second-order dependence [11, 2]. Let $\boldsymbol{x} \in \mathbb{R}^n$ denote a vector of asset returns, and let $\boldsymbol{w} \in \mathbb{R}^n$ be the vector portfolio weights, so that the portfolio return is $\boldsymbol{w}^\top \boldsymbol{x}$. Under a joint Gaussian approximation for $\boldsymbol{x}$, the distribution of the portfolio's return is entirely characterized by its mean and variance, implying that VaR and CVaR at any confidence level α reduce to explicit functions of the portfolio volatility. Consequently, even within this baseline Gaussian framework, the accuracy of volatility-based tail-risk estimates is primarily determined by the quality of the covariance estimator and by the absence of non-Gaussian behavior in the data of interest.

When the number of assets is comparable to the available sample size, known as the high-dimensional case, the sample covariance matrix exhibits large estimation error, a phenomenon refered to as curse of dimensionality. As a consequence, naive sample-based estimators tend to produce overly noisy risk assessments and unstable portfolio allocations [12, 13].

To mitigate estimation noise in high dimensions, a standard strategy is to correct the spectrum of the sample covariance matrix, which assumes that the distribution of returns is rotationally invariant [14]. In the high-dimensional regime, random matrix theory describes how sampling distorts the eigenvalue distribution, motivating a spectral correction known as non-linear shrinkage [15]. While this correction is theoretically well-grounded under idealized assumptions about the data-generating process, those assumptions may only be approximately satisfied in empirical return data [16]. This motivates complementary procedures that retain the same rotationally invariant structure but rely more directly on the data.

A practical and widely used alternative is cross-validated spectral filtering based on sample splitting [17]. The basic idea is to use one part of the data to identify the main directions of co-movement in returns (the empirical risk factors), and a different part of the data to reassess how strong each of these directions is. In this way, the method keeps the factor directions learned in-sample (IS) but replaces their variances, which are strongly contaminated by high-dimensional noise, with out-of-sample (OOS) estimates of the variance explained along the same directions. Repeating the split-and-recalibrate step over multiple random partitions and averaging the resulting corrections further reduces sensitivity to any particular split, yielding a stable filtered covariance estimator. When the assumptions underlying random-matrix-based non-linear shrinkage are approximately satisfied, this cross-validated procedure recovers the same asymptotic spectral correction [17], while remaining applicable even when those assumptions are not strictly met.

Several variants of cross-validated spectral filtering differ in whether the split is performed at random or in a time-ordered (walk-forward) way, and in whether the calibration is performed on the same asset universe or in an asset-agnostic manner by pooling information across instruments. In non-stationary financial markets, overfitting transient features can become a dominant source of error, thus enforcing causality through chronological splits and increasing the effective calibration sample by borrowing cross-sectional information both improve robustness. The Average Oracle [18] follows the same core logic as walk-forward cross-validation, but it does so over many past time windows and averages the resulting "oracle" eigenvalues that can be used in the current window and may be periodically refreshed. This additional averaging acts as a regularizer against regime-specific fluctuations, and in empirical evaluations, yields systematically more accurate estimates of the covariance matrix than methods based on the most recent data only [19].

Our objective is to mitigate the curse of dimensionality for higher-order dependence through co-moments (e.g., co-skewness and co-kurtosis), which matter for tail risk, but whose naive estimation is extremely noisy in high dimensions and unreliable in finance because of non-stationarity. To make co-moments usable in practice, we extend covariance-cleaning methods to higher-order tensors. While a fully optimal theoretical analogue of non-linear shrinkage for tensors is technically demanding, numerical methods such as cross-validation and Average Oracle methods admit direct generalizations, providing robust performance without strong asymptotic assumptions.

A first conceptual difficulty in extending covariance cleaning to higher-order co-moments is that (co-moments) tensors do not possess a spectral decomposition analogous to the matrix case. Indeed, for covariance matrices, a single orthogonal change of basis yields a diagonal representation in which the relevant information is concentrated in a spectrum: the eigenvalues quantify the empirical strength of each risk direction, and most cleaning schemes act by regularizing these eigenvalues while leaving the eigenvectors unchanged. For a co-moment tensor of order $d \geq 3$, there is no comparably stable and essentially unique notion of eigenvalues and eigenvector basis that would diagonalize the object and provide a complete set of scalar coefficients capturing its content. A natural substitute is therefore to rely on multilinear factorizations such as the Tucker decomposition (or the higher-order SVD), which identifies orthogonal directions along each mode and represents the tensor as a rotated core tensor [20]. This mirrors the role of eigenvectors in the matrix setting by providing interpretable co-movement directions, but it also makes clear why a direct spectral-style shrinkage is not available: unlike matrices, generic higher-order tensors cannot always be rotated into a diagonal form [21]. Importantly, this framework remains consistent with the classical case: when applied to a matrix, the discussed tensor decomposition reduces to the usual matrix singular value decomposition [20, 22]. Another tensor decomposition is the ODECO of Ref. [23], for which overly strong constraints are required and appears unsable in the context of this paper.

The remainder of the paper is organized as follows. Section 2 introduces quantile risk measures and the related assessment methodology, higher-order co-moment tensors and tensor decompositions. Section 3 presents the proposed cross-validation and Higher-Order Average Oracle estimators of co-moments. Section 4 reports the empirical results on U.S. equity data, assessing both tensor estimation accuracy and the resulting VaR and CVaR performance, leading to the concluding discussion of Section 5.

2 Theoretical Background

2.1 Higher order co-moment tensors

Let $\boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$ be a random vector with mean $\boldsymbol{\mu} := \mathbb{E}[\boldsymbol{x}]$, and assume that the k-th centered moments exist. The centered co-moment tensor of order k is defined by

$$\boldsymbol{\mathcal{M}}^{(k)} := \mathbb{E}\big[(\boldsymbol{x} - \boldsymbol{\mu})^{\otimes k}\big] \in (\mathbb{R}^n)^{\otimes k}. \tag{1}$$

Here $(\boldsymbol{x} - \boldsymbol{\mu})^{\otimes k}$ denotes the k-fold tensor (outer) product of $\boldsymbol{x} - \boldsymbol{\mu}$ with itself, i.e., the k-way array whose entries are products of centered components:

$$\forall\, (i_1, \ldots, i_k) \in [\![1, n]\!]^k, \qquad \mathcal{M}^{(k)}_{i_1,\ldots,i_k} = \mathbb{E}[(x_{i_1} - \mu_{i_1}) \cdots (x_{i_k} - \mu_{i_k})]\,. \tag{2}$$

This implies that, $\boldsymbol{\mathcal{M}}^{(k)}$ has n^k entries and is supersymmetric, i.e., its entries do not change under any permutation of its indices. For $k = 2$ this recovers the covariance matrix $\boldsymbol{\Sigma} := \boldsymbol{\mathcal{M}}^{(2)}$; for $k = 3$ and $k = 4$ one obtains the coskewness tensor $\boldsymbol{\mathcal{S}} := \boldsymbol{\mathcal{M}}^{(3)}$ and the cokurtosis tensor $\boldsymbol{\mathcal{K}} := \boldsymbol{\mathcal{M}}^{(4)}$, respectively. Higher orders encode increasingly complex multi-asset co-movements and provide moment-based information about non-Gaussian shape features that are relevant for tail-risk assessment.

In practice, $\boldsymbol{\mathcal{M}}^{(k)}$ must be estimated from data. Given Δt independent observations $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\Delta t}\} \in \mathbb{R}^{n \times \Delta t}$, we use the maximum likelihood estimators

$$\hat{\boldsymbol{\mu}} = \frac{1}{\Delta t} \sum_{t=1}^{\Delta t} \boldsymbol{x}_t, \qquad \widehat{\boldsymbol{\Sigma}} = \frac{1}{\Delta t} \sum_{t=1}^{\Delta t} (\boldsymbol{x}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{x}_t - \hat{\boldsymbol{\mu}})^\top, \tag{3}$$

and, more generally, the sample centered co-moment tensor

$$\widehat{\boldsymbol{\mathcal{M}}}^{(k)} = \frac{1}{\Delta t} \sum_{t=1}^{\Delta t} (\boldsymbol{x}_t - \hat{\boldsymbol{\mu}})^{\otimes k}, \tag{4}$$

which is consistent for $\boldsymbol{\mathcal{M}}^{(k)}$ under standard conditions [24]. As k and n increase, the number of entries n^k grows rapidly, so finite-sample estimation noise becomes severe; moreover, higher-order sample moments are biased in finite samples, with biases depending on Δt, n, and lower-order moments [25].

2.2 Portfolio moments

Let $\boldsymbol{w} \in \mathbb{R}^n$ denote portfolio weights and let the (central) co-moments of asset returns up to order four be given by $\hat{\mu}_i$, $\widehat{\Sigma}_{ij}$, $\widehat{\mathcal{S}}_{ijk}$, and $\widehat{\mathcal{K}}_{ijkl}$. The corresponding portfolio moments are obtained by contracting these objects with the weight vector along all their indices. Concretely, if $\hat{x} = \sum_{i=1}^n w_i \hat{x}_i$ denotes the portfolio return, then the portfolio mean and variance follow from the familiar linear and quadratic forms, while portfolio skewness and kurtosis are induced by the third- and fourth-order co-moments:

$$\hat{\mu} = \sum_{i=1}^n \hat{\mu}_i w_i, \qquad \hat{\sigma}^2 = \sum_{i,j=1}^n \widehat{\Sigma}_{ij}\, w_i w_j, \qquad \hat{\mathfrak{s}} = \sum_{i,j,k=1}^n \frac{\hat{\mathcal{S}}_{ijk}\, w_i w_j w_k}{\hat{\sigma}^3}, \qquad \hat{\kappa} = \sum_{i,j,k,l=1}^n \frac{\widehat{\mathcal{K}}_{ijkl}\, w_i w_j w_k w_l}{\hat{\sigma}^4}. \tag{5}$$

The normalization by $\hat{\sigma}^3$ and $\hat{\sigma}^4$ yields the dimensionless standardized skewness and kurtosis, respectively, making these quantities comparable across portfolios and time windows.

The contraction operation can be interpreted as a multilinear extension of the quadratic form $\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$. For a matrix, contraction aggregates pairwise dependencies into the variance of the linear combination. For the third-order coskewness tensor $\widehat{\mathcal{S}}_{ijk}$, contracting with $w_i w_j w_k$ aggregates triple co-movements into the portfolio third central moment; analogously, contracting $\widehat{\mathcal{K}}_{ijkl}$ with $w_i w_j w_k w_l$ aggregates fourth-order co-movements into the portfolio fourth central moment. This viewpoint is useful because it clarifies how higher-order dependence structures at the asset level propagate to portfolio-level tail-shape descriptors beyond variance.

Two practical remarks are important in applications. First, the denominators in $\hat{\mathfrak{s}}$ and $\hat{\kappa}$ require $\hat{\sigma} > 0$; in near-degenerate cases (e.g., highly concentrated weights or nearly collinear assets over short windows), the standardized moments may become numerically unstable, and regularization of $\widehat{\Sigma}_{ij}$ is typically required before computing $\hat{\sigma}$. Second, higher-order empirical co-moments are substantially more noise-sensitive than $\widehat{\Sigma}_{ij}$: their estimation error grows quickly with n and

they react strongly to outliers, which motivates the filtering procedures developed in the remainder of the paper. In particular, while $\hat{\sigma}^2$ depends on $O(n^2)$ entries, the raw tensors $\widehat{\boldsymbol{\mathcal{S}}}$ and $\widehat{\boldsymbol{\mathcal{K}}}$ involve $O(n^3)$ and $O(n^4)$ coefficients, thus even moderate-dimensional portfolios can render naive plug-in estimates unreliable unless the effective dimension is reduced or the tensors are denoised.

2.3 Quantile Risk

Value-at-Risk (VaR) and Conditional Value-at-Risk / Expected Shortfalls (CVaR / ES) are standard risk measures dedicated to estimate potential loss induced by rare events [6]. Let x be the random variable associated with the portfolio returns of fixed time horizon, with continuous Cumulative Distribution Function (CDF) F_{x}. Downside risk therefore corresponds to the integral of left tail of the distribution of x. Let $\alpha \in [0, 1]$ be the considered confidence level. The VaR of x is defined as the left-tail quantile of level α:

$$\mathrm{VaR}_{\mathrm{x}}(\alpha) := \inf\{z \in \mathbb{R} | F_{\mathrm{x}}(z) > \alpha\} = F_{\mathrm{x}}^{-1}(\alpha) \tag{6}$$

However, VaR is not a coherent risk measure [5]. While $\mathrm{VaR}_x(\alpha)$ summarizes a single tail quantile of the return distribution, it does not quantify the severity of outcomes beyond that quantile. In particular, two return distributions can share the same $\mathrm{VaR}_x(\alpha)$ but exhibit materially different tail behavior, because VaR is insensitive to the magnitude of losses once the threshold is crossed.

CVaR was introduced to address this limitation by averaging losses in the α-tail [26]. Formally, for $\alpha \in [0, 1]$, CVaR is defined as

$$\mathrm{CVaR}_{\mathrm{x}}(\alpha) := \mathbb{E}[x | x \leq \mathrm{VaR}_{\mathrm{x}}(\alpha)] = \frac{1}{\alpha} \int_0^{\alpha} \mathrm{VaR}_{\mathrm{x}}(\beta)\, d\beta. \tag{7}$$

This tail-conditional expectation yields a coherent risk measure and provides a more informative summary of downside tail severity than VaR [27].

Because of the non-stationarity of financial returns [10], the return distribution is rarely well-approximated by a single time-invariant law over long horizons. To make the dependence on the estimation period explicit, we may write the return distribution at time t as a time-indexed CDF F_t.

In practice, VaR and CVaR are typically used either as instantaneous risk measures or as forecasted risk measures for a future horizon. Accordingly, an estimator built from an in-sample window $\mathcal{W}_{\mathrm{in}}(t)$ is best interpreted as producing a forecast $\widehat{\mathrm{VaR}}_t(\alpha)$ and $\widehat{\mathrm{CVaR}}_t(\alpha)$ for future returns. A common evaluation approach is to analyze VaR breach rates (coverage tests), which implicitly assumes that the conditional distribution of future returns is correctly captured by the distribution estimated from past data. Under non-stationarity, however, this assumption is fragile: even a well-regularized estimator can fail coverage if the underlying return distribution shifts between the in-sample and out-of-sample periods. Moreover, in realistic settings there is no observable ground-truth CDF for the next return; only realized samples from a future window are available.

For these reasons, we adopt an explicitly forecasting viewpoint and evaluate quantile and tail-risk estimators by comparing their forecasts to an OOS reference distribution constructed from a subsequent window of observations. Concretely, let $\mathcal{W}_{\mathrm{out}}(t)$ denote a future evaluation window following $\mathcal{W}_{\mathrm{in}}(t)$, and let F_t^{out} be the estimated CDF computed on $\mathcal{W}_{\mathrm{out}}(t)$. We then assess the quality of an estimator through forecasting losses comparing $(\widehat{\mathrm{VaR}}_t(\alpha), \widehat{\mathrm{CVaR}}_t(\alpha))$ to $(\widehat{\mathrm{VaR}}_t^{\mathrm{out}}(\alpha), \widehat{\mathrm{CVaR}}_t^{\mathrm{out}}(\alpha))$. This evaluation choice makes explicit that (i) a true distribution is not directly observable, and (ii) when the distribution evolves over time, the relevant benchmark for a forecast is the distribution realized in the future evaluation window, rather than an expectation of exact long-run coverage under stationarity.

2.4 Parametric Risk Estimation: Fernández and Steel Skew-t Distribution

To obtain parametric estimates of VaR and CVaR, one possibility is to model the portfolio returns using the skewed-t distribution, as introduced by Fernández and Steel (FS) [28]. While the Student-t distribution is already widely used for financial returns, primarily because it captures heavy tails through a single paramter ν, the FS skewed-t distribution has recently attracted attention as a flexible extension [29]. In addition to preserving heavy-tailed behavior, the FS distribution allows for asymmetry via a skewness parameter ($\xi > 0$). Technically, its density is obtained by applying an asymmetric scale transformation to an underlying symmetric Student-t distribution.

The FS probability density function with location $\mu \in \mathbb{R}$, scale $\sigma > 0$, skewness parameter $\xi > 0$, and degrees of freedom $\nu > 0$ is given by

$$P_{\mathrm{FS}}(x|\mu,\sigma,\xi,\nu) = \frac{2}{\sigma(\xi+\xi^{-1})} \begin{cases} P_{\mathrm{t}}\left(\dfrac{\xi(x-\mu)}{\sigma}|\nu\right), & x < \mu, \\ P_{\mathrm{t}}\left(\dfrac{x-\mu}{\sigma\,\xi}|\nu\right), & x \geq \mu, \end{cases} \tag{8}$$

where P_{t} denotes the standardardized Student-t density with ν degrees of freedom.

Given sample estimates of the univariate centered moments estimated for a given portfolio $(\hat{\mu}, \hat{\sigma}, \hat{\mathfrak{s}}, \hat{\kappa})$, we estimate (ν, ξ) via nonlinear moment matching. Let $\mathfrak{s}_{\mathrm{FS}}(\nu,\xi)$ and $\kappa_{\mathrm{FS}}(\nu,\xi)$ denote the model-implied skewness and kurtosis of the FS skew-t distribution (available in closed form [28]). We then compute

$$(\hat{\nu}, \hat{\xi}) = \arg\min_{\nu>4,\ \xi>0} \left[(\mathfrak{s}_{\mathrm{FS}}(\nu,\xi) - \hat{\mathfrak{s}})^2 + (\kappa_{\mathrm{FS}}(\nu,\xi) - \hat{\kappa})^2\right]. \tag{9}$$

The constraint $\nu > 4$ guarantees finite fourth moment.

Once $(\hat{\nu}, \hat{\xi})$ are obtained, the model-implied mean $\mu_{\mathrm{FS}}(\hat{\nu}, \hat{\xi})$ and standard deviation $\sigma_{\mathrm{FS}}(\hat{\nu}, \hat{\xi})$ need not coincide with the empirical $(\hat{\mu}, \hat{\sigma})$. To retain consistency with the observed scale and location of returns, the fitted skew-t quantiles are therefore standardized and rescaled:

$$\mathrm{VaR}_{\mathrm{FS}}(\alpha) := \hat{\mu} + \hat{\sigma}\,\frac{q_{\mathrm{FS}}(\alpha|\hat{\nu},\hat{\xi}) - \mu_{\mathrm{FS}}(\hat{\nu},\hat{\xi})}{\sigma_{\mathrm{FS}}(\hat{\nu},\hat{\xi})}, \tag{10}$$

where $q_{\mathrm{FS}}(\alpha|\nu,\xi)$ denotes the centered and standardized FS skew-t quantile.

2.5 Semi-parametric Risk Estimation: the Cornish-Fisher Expansion

Another viable approach is to approximate VaR and CVaR via the Cornish–Fisher (CF) expansion, a moment-based semi-parametric correction of Gaussian tail risk that is also referred to as the modified VaR in the literature [30]. The key idea is to approximate the lower-tail quantile function of portfolio returns by a truncated asymptotic expansion around the Gaussian distribution; in other words, instead of committing to a fully specified parametric family for the return distribution, CF uses only a finite set of standardized cumulants to adjust the normal quantile map [31, 32].

Let x be the random variable associated with the portfolio returns, with first four estimated central moments $(\hat{\mu}, \hat{\sigma}, \hat{\mathfrak{s}}, \hat{\kappa})$. Specifically, CF postulates that the standardized return $(\mathrm{x} - \hat{\mu})/\hat{\sigma}$ can be approximated as a polynomial perturbation of a standard normal variable $z \sim \mathcal{N}(0,1)$:

$$\frac{\mathrm{x}-\hat{\mu}}{\hat{\sigma}} \approx z + \frac{s}{6}(z^2-1) + \frac{k}{24}(z^3-3z) - \frac{s^2}{36}(2z^3-5z), \tag{11}$$

where s and k are expansion coefficients inferred from the target skewness $\hat{\mathfrak{s}}$ and kurtosis $\hat{\kappa}$ as detailed in Ref. [33]. Importantly, because the mapping is a truncated approximation, these coefficients should be viewed as effective parameters of the expansion and need not coincide exactly with the true standardized cumulants [33, 34]. Since CF is obtained by truncating an asymptotic quantile expansion around the Gaussian distribution [31, 32], it is most accurate when departures from normality are moderate at the tail level of interest, and it can deteriorate for very extreme quantiles or pronounced skewness/kurtosis [35, 36]. Moreover, truncation does not guarantee that the resulting mapping defines a globally valid (monotone) quantile function [37].

For tail level $\alpha \in (0,1)$, let $q_\alpha := \mathrm{VaR}_z(\alpha) = \Phi^{-1}(\alpha)$ be the standard normal quantile. The standardized CF quantile is

$$\mathrm{VaR}_{z^\star}(\alpha) \approx q_\alpha + \frac{s}{6}(q_\alpha^2-1) + \frac{k}{24}(q_\alpha^3-3q_\alpha) - \frac{s^2}{36}(2q_\alpha^3-5q_\alpha), \tag{12}$$

which yields the CF approximation of the wanted quantile:

$$\mathrm{VaR}_{\mathrm{x}}(\alpha) = \mu_p + \sigma_p\,\mathrm{VaR}_{z^\star}(\alpha). \tag{13}$$

The computation of $\mathrm{CVaR}_{\mathrm{x}}(\alpha)$ uses the Gaussian tail factor $y_\alpha := \mathrm{CVaR}_z(\alpha)$. A convenient implementation is

$$\mathrm{CVaR}_{z^\star}(\alpha) \approx y_\alpha[1 + \frac{q_\alpha}{6}s + \frac{1-2q_\alpha^2}{36}s^2 + \frac{q_\alpha^2-1}{24}k], \tag{14}$$

finally:

$$\mathrm{CVaR}_{\mathrm{x}}(\alpha) = \mu_p + \sigma_p \, \mathrm{CVaR}_{z^\star}(\alpha) \tag{15}$$

Operationally, the procedure proceeds as follows. Portfolio moments are first obtained by contracting the estimated co-moment tensors with the portfolio weights over all indices, yielding the portfolio mean, volatility, skewness, and kurtosis (see Eq. 5). These empirical moments are then used to infer the effective CF parameters s and k, and the VaR and CVaR are then calculated using equations 13 and 15.

2.6 Tensor decomposition

Tensor decompositions generalize the matrix Singular Value Decomposition (SVD) to higher-order arrays. Among the most widely used are the Canonical Polyadic decomposition and the Higher-Order Singular Value Decomposition (HOSVD). Both reduce to the classical SVD in the matrix (second-order tensor) case: the CP decomposition extends the rank-one factorization of matrices, while the Tucker decomposition extends the rotation–scaling–rotation structure of the SVD. However, there is in general no single decomposition unifying both forms for generic tensors [21].

To connect tensor operations with familiar linear-algebraic ones, it is convenient to unfold (or matricize) the tensor along a chosen mode. This operations reshape tensor into a matrix by selecting one axis to play the role of "rows" and merging all remaining axes into a single "column" axis.

We denote the mode-r unfolding as

$$\mathrm{unfold}_r \boldsymbol{\mathcal{T}} := \boldsymbol{T}_r \in \mathbb{R}^{n_r \times p_r}, \quad p_r := \prod_{j \neq r}^{k} n_j \tag{16}$$

The mode-r product of $\boldsymbol{\mathcal{T}}$ by a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n_r}$ is denoted $\boldsymbol{\mathcal{T}} \times_r \boldsymbol{A} \in \mathbb{R}^{n_1 \times n_2 \cdots n_{r-1} \times m \times n_{r+1} \cdots \times n_k}$ and defined element-wise by

$$(\boldsymbol{\mathcal{T}} \times_r \boldsymbol{A})_{i_1 \cdots i_{r-1} j_r i_{r+1} \cdots i_k} := \sum_{i_r=1}^{n_r} \mathcal{T}_{i_1 \cdots i_r \cdots i_k} \, A_{j_r i_r}. \tag{17}$$

The HOSVD, proposed by De Lathauwer *et al* [20], is a generalization of the matrix SVD by expressing an order-k tensor $\boldsymbol{\mathcal{T}}$ in terms of k mode products. In our specific case, we are interested in decomposing the co-moments tensor $\boldsymbol{\mathcal{M}}^{(k)}$, which is super-symmetric (see Eq. (2)), therefore the decomposition can be written as

$$\boldsymbol{\mathcal{M}}^{(k)} = \boldsymbol{\mathcal{G}}^{(k)} \times_1 \boldsymbol{U}^\top \times_2 \boldsymbol{U}^\top \cdots \times_k \boldsymbol{U}^\top. \tag{18}$$

where

$$\boldsymbol{\mathcal{G}}^{(k)} \in \mathbb{R}^{n^{\otimes k}}, \quad \boldsymbol{U} \in O(n).$$

The core tensor $\boldsymbol{\mathcal{G}}^{(k)}$ has the same shape as the original tensor $\boldsymbol{\mathcal{M}}^{(k)}$ and encodes the interactions between the different modes, playing a role analogous to that of the singular values in the standard matrix SVD, which is recovered for $k = 2$.

Operationally, one may compute $\boldsymbol{U}$ from any unfolding

$$\boldsymbol{T}_1 \boldsymbol{T}_1^\top = \boldsymbol{U} \, \boldsymbol{\Lambda} \, \boldsymbol{U}^\top, \tag{19}$$

then the core tensor can be obtained by

$$\boldsymbol{\mathcal{G}}^{(k)} = \boldsymbol{\mathcal{M}}^{(k)} \times_1 \boldsymbol{U}^\top \times_2 \boldsymbol{U}^\top \cdots \times_k \boldsymbol{U}^\top. \tag{20}$$

Since the present work aims to regularize the core tensor, it is important to stress that a suitable core $\boldsymbol{\mathcal{G}}^{(k)}$ must satisfy the all-orthogonality condition [20], which is

$$\boldsymbol{\mathcal{G}}^{(k)} \in \mathbb{A}_k := \{\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n \times \cdots \times n} \mid \forall \, r \leq k, \; \forall \, \alpha, \; \beta \leq n, \; \left\langle \boldsymbol{\mathcal{A}}_{\, i_r = \alpha}, \boldsymbol{\mathcal{A}}_{\, i_r = \beta} \right\rangle = 0 \text{ if } \alpha \neq \beta\}, \tag{21}$$

where $\left\langle \bullet, \bullet \right\rangle$ denotes the canonical inner-product on tensor space, that is $\left\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \right\rangle := \sum_{i_1, \ldots, i_k} \mathcal{A}_{i_1 \ldots i_k} \mathcal{B}_{i_1 \ldots i_k}$. In plain words, any family of slices of a given mode are mutually orthogonal (see appendix A.1 for a discussion).

2.7 Oracle for co-moment tensors

In analogy with covariance filtering, we consider co-moment tensor cleaning in a fixed multilinear basis, where regularization acts only on the core. We define the oracle core as the Frobenius-optimal set of core coefficients that would be obtained if population (equivalently, strictly out-of-sample) information were available for calibration. While infeasible in applications, this oracle provides a natural lower bound for any implementable fixed-basis cleaning rule and identifies the core as the sole object to be filtered.

These points are formalized by two propositions. Proposition 1 establishes the orthogonal transformation properties of co-moment tensors and the structural implications of distributional symmetries. Proposition 2 then characterizes the unconstrained oracle core associated with a prescribed basis; under the symmetry assumptions of Proposition 1, this oracle core is compatible with the all-orthogonality structure of HOSVD-type representations. Finally, we show how the oracle principle can be approximated without data leakage by calibrating the core using information independent of the basis estimation (e.g., via sample splitting or cross-fitting).

Proposition 1. *Let $\mathbf{x} \in \mathbb{R}^n$ a random vector, and let $k \geq 2$. Assume that $\mathbf{x}$ is rotationally invariant, so that for any orthogonal matrix $\boldsymbol{U} \in O(n)$ we have in distribution:*

$$\mathbf{x} \overset{d}{=} \boldsymbol{U}\mathbf{x}. \tag{22}$$

Then, its co-moment tensor $\boldsymbol{\mathcal{M}}$ of order k verifies

$$\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{M}} \times_1 \boldsymbol{U} \cdots \times_k \boldsymbol{U}. \tag{23}$$

In words, $\boldsymbol{\mathcal{M}}$ is invariant under the simultaneous orthogonal action of $\boldsymbol{U}$ on all modes. A proof is presented in appendix A.2.

If $\mathbf{x}$ is rotationally invariant, then in particular for $\boldsymbol{U} = -\mathbb{I}_n$ we have

$$\begin{aligned}\boldsymbol{\mathcal{M}}^{(k)} &= \mathbb{E}[x^{\otimes k}] \\ &= \mathbb{E}[(-x)^{\otimes k}] \\ &= (-1)^k \mathbb{E}[x^{\otimes k}] \\ &= (-1)^k \boldsymbol{\mathcal{M}}^{(k)}.\end{aligned}$$

Therefore, any odd-order co-moment must vanish:

$$\boldsymbol{\mathcal{M}}^{(2k+1)} = \mathbf{0}, \tag{24}$$

including the coskewness tensor. Observing near-0 odd-order co-moments is therefore compatible with the assumption of rotationally invariant financial returns, as it will be the case in our empirical analysis of Sec. 4.

Proposition 2. *Let $\mathbf{x}$ be a random (vector) variable associated with returns. Let $k \geq 2$, and let $\boldsymbol{\mathcal{M}}_{\textit{population}}$ be the population co-moment tensor of order k. Consider realizations of $\mathbf{x}$ from which one estimates the co-moment tensor $\widehat{\boldsymbol{\mathcal{M}}}$ of order k. Let $\widehat{\boldsymbol{U}} \in O(n)$ be the mode basis of the HOSVD of $\widehat{\boldsymbol{\mathcal{M}}}$ (see Eq. (18)).*

Define the map

$$\widehat{\boldsymbol{\mathcal{M}}}(\boldsymbol{\mathcal{G}}) \;:=\; \boldsymbol{\mathcal{G}} \times_1 \widehat{\boldsymbol{U}} \cdots \times_k \widehat{\boldsymbol{U}}, \qquad \boldsymbol{\mathcal{G}} \in \mathbb{R}^{n \times \cdots \times n}. \tag{25}$$

Then the following unconstrained minimization problem:

$$\boldsymbol{\mathcal{G}}^{\star} = \arg\min_{\boldsymbol{\mathcal{G}} \in \mathbb{R}^{n \times \cdots \times n}} \|\widehat{\boldsymbol{\mathcal{M}}}(\boldsymbol{\mathcal{G}}) - \boldsymbol{\mathcal{M}}_{\textit{population}}\|_F^2 \tag{26}$$

has the unique solution

$$\boldsymbol{\mathcal{G}}^{\star} = \boldsymbol{\mathcal{M}}_{\textit{population}} \times_1 \widehat{\boldsymbol{U}}^{\top} \cdots \times_k \widehat{\boldsymbol{U}}^{\top}. \tag{27}$$

In words, given a fixed orthogonal basis $\widehat{\boldsymbol{U}}$, the best unconstrained core (in Frobenius sense) is obtained by inverse rotation of the population co-moment (see an algebraic proof in appendix A.3). Observe that $\boldsymbol{G}^{\star} \notin \mathbb{A}_k$ means that $\boldsymbol{\mathcal{G}}^{\star} \times_1 \widehat{\boldsymbol{U}} \cdots \times_k \widehat{\boldsymbol{U}}$ is not the HOSVD of a co-moment tensor even if it solves the considered least-squares problem.

However, under the assumption of rotational invariance of $\mathbf{x}$, we have $\boldsymbol{\mathcal{G}}^{\star} \in \mathbb{A}_k$ as shown in appendix A.4. Estimating $\widehat{\boldsymbol{\mathcal{G}}}^{\star}$ from realized returns can be assumed, under rotational invariance, to provide $\widehat{\boldsymbol{\mathcal{G}}}^{\star} \in \mathbb{A}_k$ in expectation.

3 Co-Moment Shrinkage

3.1 Cross-Validation estimators

We now describe the proposed shrinkage of sample co-moments based on the tensor decomposition of Sec. 2.6 using K-fold cross-validation. The goal is to construct an estimator $\widehat{\boldsymbol{\mathcal{M}}}^{(k)}$ whose core tensor $\widehat{\boldsymbol{\mathcal{G}}}^{(k)}$ is estimated in hold-out way, to empirically approximate the Oracle of Eq. (27) from realized returns.

The idea is to split the IS data into train and test subsets by randomly assigning the cross-sectional day block either to the train or to the test. We estimate the co-moments $\widehat{\mathcal{M}}_{\text{train}}$ and $\widehat{\mathcal{M}}_{\text{test}}$ from the training and testing set respectively. We then decompose $\widehat{\mathcal{M}}_{\text{train}}$ by HOSVD, yielding a training rotation matrix $\widehat{\boldsymbol{U}}_{\text{train}}$ and core tensor. We then build $\widehat{\boldsymbol{\mathcal{G}}}^{(k)}$ using $\widehat{\mathcal{M}}_{\text{test}}$ and $\widehat{\boldsymbol{U}}_{\text{train}}$ in Eq. (18). Repeating this procedure and averaging the newly formed tensors yields a cross-validated estimator of the core tensor.

Let the dataset be split into K folds of training/test (IS/OOS) splits. For a given fold b, denote $\widehat{\boldsymbol{\mathcal{M}}}^{(k)}_{\text{train}(b)}$ the co-moment tensor of order k computed from the train data, $\widehat{\boldsymbol{\mathcal{M}}}^{(k)}_{\text{test}(b)}$ the co-moment tensor from the test data. The proposed CV procedure consists in performing the HOSVD of the train co-moment tensor:

$$\widehat{\boldsymbol{\mathcal{M}}}^{(k)}_{\text{train}(b)} = \widehat{\boldsymbol{\mathcal{G}}}_{\text{train}(b)} \times_1 \widehat{\boldsymbol{U}}^{(b)}_{\text{train}} \cdots \times_k \widehat{\boldsymbol{U}}^{(b)}_{\text{train}} \tag{28}$$

Given $U^{(b)}_{\text{train}}$, the test tensor is projected onto the train multilinear basis:

$$\widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{(b)} := \widehat{\boldsymbol{\mathcal{M}}}^{(k)}_{\text{test}(b)} \times_1 \widehat{\boldsymbol{U}}^{(b)^\top}_{\text{train}} \cdots \times_k \widehat{\boldsymbol{U}}^{(b)^\top}_{\text{train}}. \tag{29}$$

This ensures that $\widehat{\boldsymbol{\mathcal{G}}}_{(b)}$ is the optimal core to estimate $\widehat{\boldsymbol{\mathcal{M}}}_{\text{test}}$ in the basis $\widehat{\boldsymbol{U}}^{(b)}_{\text{train}}$, as shown in Sec. 2.

The CV core estimator $\widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{\text{CV}}$ is then obtained by averaging over folds:

$$\widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{\text{CV}} := \frac{1}{K} \sum_{b=1}^{K} \widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{(b)}. \tag{30}$$

Once the CV core is obtained, the final estimator of the co-moment tensor is

$$\widehat{\boldsymbol{\mathcal{M}}}^{(k)}_{\text{CV}} := \widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{\text{CV}} \times_1 \widehat{\boldsymbol{U}} \cdots \times_k \widehat{\boldsymbol{U}}, \tag{31}$$

where $\widehat{\boldsymbol{U}}$ is typically chosen as the rotation basis computed from the HOSVD on the full dataset. As in the covariance setting, the estimator trades bias in the tensor core to lower the variance of the estimated co-moment. We explicitely assume $\widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{\text{CV}} \in \mathbb{A}_k$, which can be expected to hold for rotationally invariant returns, as shown in Appendix A.4.

3.2 Higher order Average Oracle estimator

An alternative to K-fold cross-validation is the Average Oracle approach, which preserves the temporal ordering of the data and is designed for settings where a long calibration window is available. Instead of repeatedly shuffling and splitting the sample of a single IS window, we slide a fixed-length IS window forward in time and use the subsequent OOS observations as a causal validation block.

More precisely, let the sample be $\{x_t\}_{t=1}^{\Delta t}$ and fix two calibration subwindows of length $\Delta t_{\text{train}}, \Delta t_{\text{test}} \ll \Delta t$. For each time $\tau = \Delta t_{\text{train}}, \Delta t_{\text{train}} + 1, \ldots, \Delta t - \Delta t_{\text{test}}$, let $\widehat{\boldsymbol{\mathcal{M}}}^{(k)}_{\text{train}(\tau)}$ be the sample k-order co-moment computed on the training set $\{x_{\tau - \Delta t_{\text{train}}+1}, \ldots, x_\tau\}$, and $\widehat{\boldsymbol{\mathcal{M}}}^{(k)}_{\text{test}(\tau)}$ the sample k-order co-moment computed on the testing set $\{x_\tau, \ldots, x_{\tau+\Delta t_{\text{test}}}\}$.

The proposed procedure consists in computing the HOSVD of the training co-moment tensor:

$$\widehat{\boldsymbol{\mathcal{M}}}^{(k)}_{\text{train}(\tau)} = \widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{\text{train}(\tau)} \times_1 \widehat{\boldsymbol{U}}^{(\tau)} \cdots \times_k \widehat{\boldsymbol{U}}^{(\tau)}, \tag{32}$$

followed by projecting the testing co-moment tensor onto the training basis:

$$\widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{\text{Oracle}(\tau)} := \widehat{\boldsymbol{\mathcal{M}}}^{(k)}_{\text{test}(\tau)} \times_1 \widehat{\boldsymbol{U}}^{(\tau)^\top} \cdots \times_k \widehat{\boldsymbol{U}}^{(\tau)^\top}. \tag{33}$$

The (Higher Order) Average Oracle core $\widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{\text{HOAO}}$ is obtained by an arithmetic mean of the calibrated Oracle cores:

$$\widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{\text{HOAO}} := \frac{1}{\Delta t - \Delta t_{\text{train}} - \Delta t_{\text{test}}} \sum_{\tau=\Delta t_{\text{train}}}^{\Delta t - \Delta t_{\text{test}}} \widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{\text{Oracle}(\tau)}. \tag{34}$$

Finally, the corresponding tensor estimator is

$$\widehat{\boldsymbol{\mathcal{M}}}^{(k)}_{\text{HOAO}} := \widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{\text{HOAO}} \times_1 \widehat{\boldsymbol{U}} \times_2 \cdots \times_k \widehat{\boldsymbol{U}}, \tag{35}$$

where $\widehat{\boldsymbol{U}}$ is chosen as the HOSVD basis matrix of the sample co-moment to be filtered. We call this estimator the Higher Order Average Oracle (HOAO). When $k = 2$, one recovers the shrinkage recipe of Ref. [18].

The HOAO method mirrors the logic of CV but enforces the chronological structure of the data, making it particularly suitable for financial or econometric applications where shuffling the time indices breaks the dependence structure. As explained in Sec. 4, $\widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{\text{HOAO}}$ can be calibrated on other instruments' time-series and applied on the desired sample co-moment tensor by replacing the noisy core with the HOAO one. This bootstrapping approach, which empirically works on equity data, allows one to calibrate a large number of past Oracles on a temporally limited dataset [18].

Furthermore, we assume explicitely that the expectation over a large number of Oracles in Eq. (34) implies $\widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{\text{HOAO}} \in \mathbb{A}_k$, so that Eq. (35) is the HOSVD of $\widehat{\boldsymbol{\mathcal{M}}}^{(k)}_{\text{HOAO}}$.

4 Empirical Study

We now turn to the empirical analysis, for which we use 30 years (1995-2024) of unbiased US equity close-to-close daily returns. The universe consists of common stocks listed on NYSE, AMEX, and NASDAQ.

In this context, population quantities are not accessible, and as explained in Sec. 2.3, we adopt an OOS evaluation perspective throughout. For tensor estimation, target quantities are defined as realized OOS co-moment tensors. For tail risk evaluation, we use the OOS realized portfolio moments, which are then employed to fit a FS skew-t distribution following the methodology described in Sec. 2.4, and subsequently compute the corresponding RMSE of the estimated risk measures.

We argue that this procedure provides a sound target : calibrating a skew-t distribution on OOS realized moments empirically ensures exact coverage of OOS realized percentiles and, equivalently, correct breach rates for the estimated quantiles in the test period. As such, it constitutes a coherent and internally consistent reference for assessing the performance of competing estimators in the absence of observable population quantities.

For computational reasons, we focus on portfolios of dimension $n = 50$, which allows for robust empirical analysis while keeping the manipulation of higher-order tensors tractable. The computational cost of tensor-based estimators grows rapidly with dimension, but the results appear to remain valid for dimensions $n \geq 100$.

Tensor accuracy is measured through Frobenius-norm errors to realized OOS co-moment tensors, for orders $k \in \{2, 3, 4\}$. Tail risk accuracy is assessed through the RMSE of VaR and CVaR estimates over a dense grid of percentiles $\{0.01, 0.02, \ldots, 0.50\}$, as the analysis focuses on downside risk.

In each IS window, we compute three tensor estimators: (i) the sample (MLE) co-moment tensors (Eq. (1)); (ii) the CV filtered estimator with 10 folds with a train/test split of 65% (Sec. 3.1); and (iii) the HOAO estimator (Sec. 3.2). Volatility is handled by first standardizing each marginal using the IS univariate standard deviations, and subsequently rescaling after estimation. The same procedure is applied to the oracle core tensors, which are computed using standardized returns.

HOAO is calibrated on the $1995 - 2010$ period. During calibration, we repeatedly draw (with replacement) random subsets of n stocks and create causal decompositions of the data into independent training and testing windows of length $\Delta t_{\text{train}} = \Delta t_{\text{test}} = 100$ trading days each. For each split and each order $k \in \{2, 3, 4\}$, we compute the IS sample co-moment tensor and perform its HOSVD, producing factor matrices and a core tensor. Core tensors are accumulated across all calibration draws and averaged order-by-order. In the testing period, the data are processed in rolling-window fashion: each evaluation uses an IS window of $\Delta t_{\text{IS}} = 100$ days to estimate tensors and risk models, followed by a non-overlapping OOS window of $\Delta t_{\text{OOS}} = 100$ days used for realized evaluation. Additionally, every two years we update the oracle by incorporating $2{,}000$ new calibrated oracle cores from newly available data. We conduct 10,000 experiments over the 2010–2024 period. In each experiment, n stocks are randomly selected and two causal, non-overlapping windows of 100 trading days each are used for IS estimation and OOS evaluation. Sample

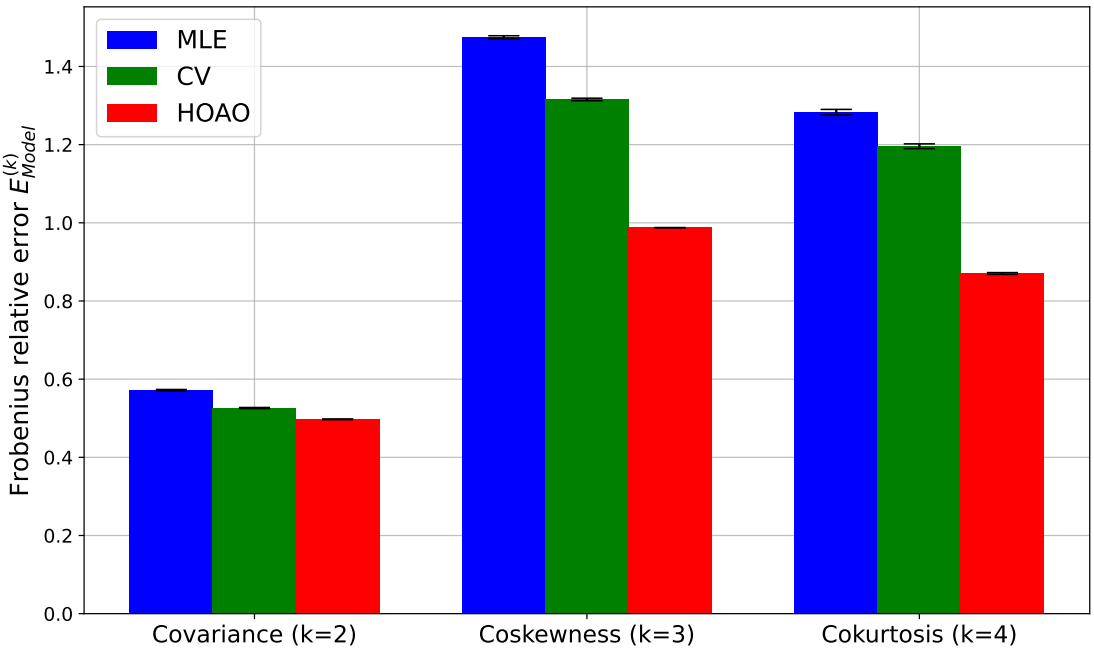

Figure 1: Frobenius relative errors $E^{(k)}_{\text{model}} = \frac{\|\boldsymbol{\mathcal{M}}^{(k)}_{\text{model}} - \boldsymbol{\mathcal{M}}^{(k)}_{\text{OOS}} r\|_F}{\|\boldsymbol{\mathcal{M}}^{(k)}_{\text{OOS}} r\|_F}$ for co-moment tensors of orders $k \in \{2,3,4\}$, compared across MLE (sample), CV, and HOAO estimators, along confidence bands. Results are based on US equity close-to-close daily returns with portfolio dimension $n = 50$, in-sample calibration window of size $\Delta t_{\text{IS}} = 100$ and out-of-sample test window of size of $\Delta t_{\text{OOS}} = 100$, with 10,000 experiments.

(MLE), CV, and HOAO co-moment estimators are computed, after which 100 portfolios are drawn from a Dirichlet distribution. VaR and CVaR are then evaluated for each portfolio, yielding a total of one million portfolio-level risk estimation experiments.

For each order $k \in \{2,3,4\}$, let $\boldsymbol{\mathcal{M}}^{(k)}_{\text{OOS}}$ be the realized OOS co-moment tensor of order k computed from the entire 100-day OOS window, and let $\boldsymbol{\mathcal{M}}^{(k)}_{\text{model}}$ denote the corresponding IS estimator (MLE/sample, CV, or HOAO). We measure tensor accuracy using the Frobenius relative error

$$E^{(k)}_{\text{model}} := \frac{\|\boldsymbol{\mathcal{M}}^{(k)}_{\text{model}} - \boldsymbol{\mathcal{M}}^{(k)}_{\text{OOS}} r\|_F}{\|\boldsymbol{\mathcal{M}}^{(k)}_{\text{OOS}} r\|_F}, \qquad \text{model} \in \{\text{MLE}, \text{CV}, \text{HOAO}\}. \tag{36}$$

This definition places all tensor orders on a common scale, and is applied at the de-volatilized level, isolating the accuracy of the covariate structure. Results are reported in Figure 1.

The Frobenius-norm results exhibit a clear and stable ordering that mirrors what is known for covariance filtering and extends to higher orders: the sample estimator is dominated by CV, and HOAO improves further, i.e. $\text{MLE} \succ \text{CV} \succ \text{HOAO}$ in terms of error magnitude for all tensor orders considered. The HOAO coskewness tensor shrinks strongly toward zero, but not exactly: its residual magnitude remains statistically distinguishable from zero, with a Frobenius norm significantly below 1 under a one-sample t-test, indicating that HOAO removes most spurious third-order structure while preserving a small but detectable component. Furthermore, this observation conforts the discussion of Sec. 1, as we expect such RIE to shrink odd-order co-moment towards 0.

We further observe that, for $k \in \{3,4\}$, $\widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{\text{HOAO}}$ is almost in set $\mathbb{A}_k$, while $\widehat{\boldsymbol{\mathcal{G}}}^{(k)}_{\text{CV}}$ exhibit non-negligeable violation of the all-orthogonality constraint. Specifically, our initial assumption holds well for HOAO and much less for CV as discussed in App. B.1.

For each IS window and fixed portfolio, we compute sample, CV and HOAO portfolio moments $(\hat{\mu}, \hat{\sigma}, \hat{\mathfrak{s}}, \hat{\kappa})$ and construct previoulsy described VaR/CVaR estimators, see sections 2.4 and 2.5. We also consider a Gaussian parametric estimator using only $(\hat{\mu}, \hat{\sigma})$ and a historical estimator using the empirical quantile of the 100 IS datapoints.

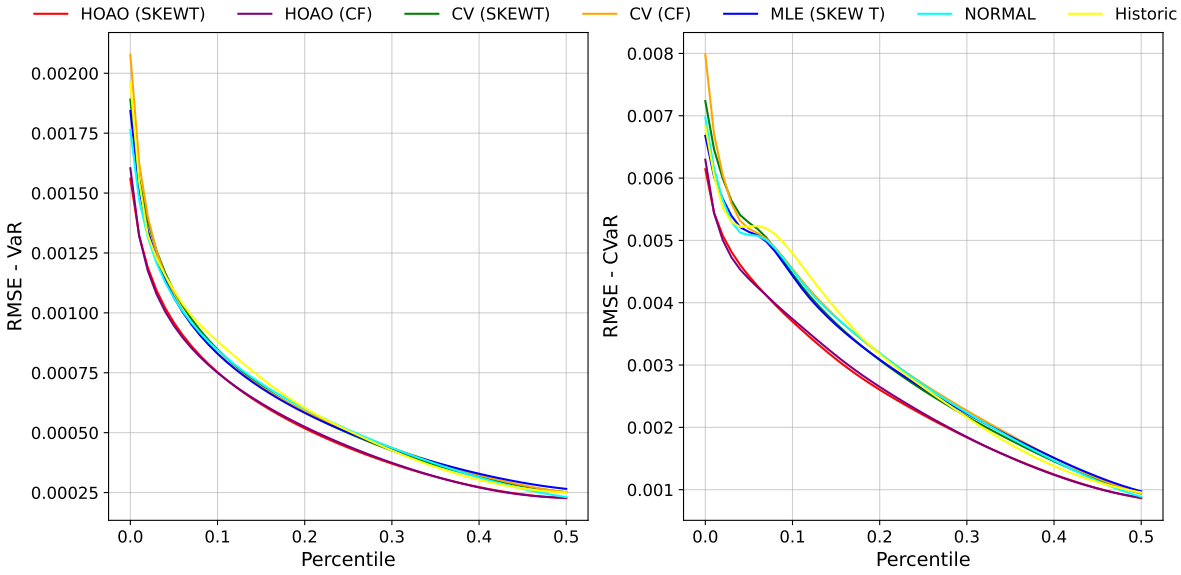

Figure 2: Out-of-sample accuracy of VaR and CVaR estimators across confidence levels. Left panel: RMSE of VaR estimates as a function of the confidence level $\alpha \in \{0.01, \cdots, 0.50\}$ for the different estimators considered. Right panel: RMSE of CVaR estimates over the same range of confidence levels. The target VaR and CVaR are computed from a Fernández–Steel skew-t model fitted to realized out-of-sample portfolio moments. Results are based on US equity close-to-close daily returns with portfolio dimension $n = 50$, in-sample calibration window of size $\Delta t_{\text{IS}} = 100$ and out-of-sample test window of size of $\Delta t_{\text{OOS}} = 100$, with 1,000,000 experiments: we performed 10,000 bootstrapped selections, and for each selection we evaluate 100 portfolios drawn from a Dirichlet distribution.

Results of the RMSE against the percentile level are reported in Figure 2. To furthur quantify robustness of performance differences across estimators and level percentile, we compute a MCS at test size 0.001 from RMSE values. The MCS identifies the subset of estimators statistically indistinguishable from the best-performing method [38], see App. B.2.

The tail-risk results corroborate the tensor-level findings. Across all percentiles below 50%, the MCS analysis at 0.001 confidence level excludes the Gaussian, historical, and MLE-based skew-t and CV estimators, leaving only the HOAO-based procedures as statistically competitive. Both HOAO parametric (skew-t) and HOAO semi-parametric (corrected Cornish–Fisher) variants remain alternatively in the confidence set, indicating that the performance gains translate into robust improvements in VaR/CVaR estimation. The consistent under-performance of the historical estimator for relatively large percentile is a sign of the non-stationary nature of financial returns distributions, while the observed superiority of the normal approach over the parametric MLE indicates that the use of unfiltered higher-order moments may be detrimental rather than beneficial for risk assessment.

Aggregating the RMSE across all α percentiles, rather than analyzing each percentile separately, yields the summary statistics reported in Table 1. The table summarizes VaR and CVaR performance aggregated over all experiments, and reports the correlation between estimated risk measures and realized OOS risk. The results show that parametric risk estimation based on HOAO-filtered moments consistently overperforms the competing approaches. Best-performing RMSE values are reported in bold when they are statistically significant at the 1% level according to a one-sided paired Wilcoxon signed-rank test.

5 Conclusion

Across both tensor-level and risk-level criteria, the proposed higher-order filtering procedures deliver systematic gains over naive plug-in estimation. On real U.S. equity data in a rolling out-of-sample design, the ranking of performance of tensor estimators is stable across tensor orders: sample estimates are dominated by cross-validation, and HOAO further lowers relative Frobenius errors for covariance, coskewness, and cokurtosis. These tensor improvements translate into materially more accurate tail-risk measurements, as evidenced by lower RMSE for VaR and CVaR across confidence levels and by decisive MCS tests that retain HOAO-based estimators as the only consistently competitive procedures over the relevant lower-tail region.

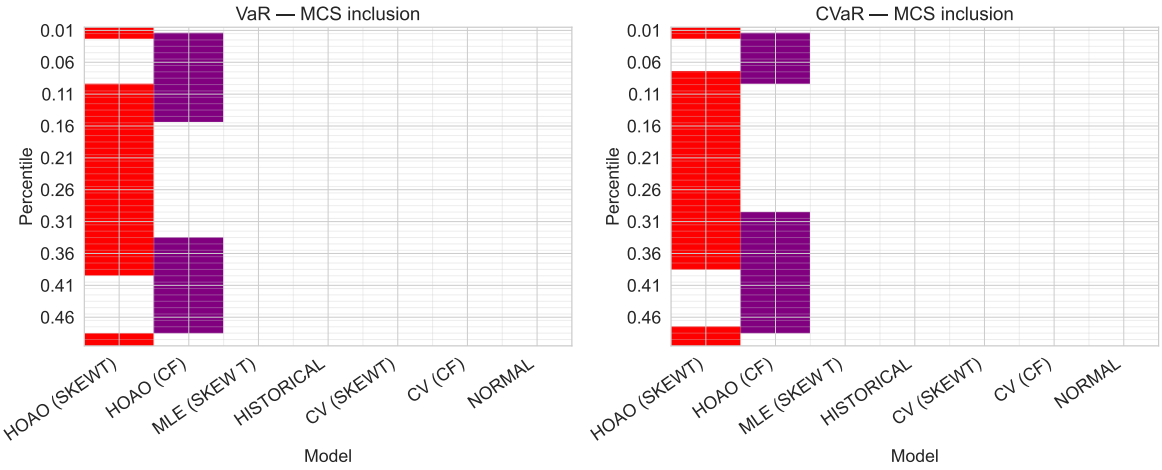

Figure 3: Model Confidence Set (MCS) inclusion results at test size 0.001 with block bootstrap, for percentiles in $\{0.01, \cdots 0.50\}$. Inclusion in the MCS is marked by a colored band percentile-wise. Left panel: VaR best performing models from left to right. Right panel: CVaR best performing models from left to right. The MCS identifies the subset of estimators statistically indistinguishable from the best-performing method in each setting. Results are based on US equity close-to-close daily returns with portfolio dimension $n = 50$, in-sample calibration window of size $\Delta t_{\text{IS}} = 100$ and out-of-sample test window of size of $\Delta t_{\text{OOS}} = 100$, with 1,000,000 experiments: we performed 10,000 bootstrapped selections, and for each selection we evaluate 100 portfolios drawn from a Dirichlet distribution.

	VaR		CVaR	
Model	RMSE (10^{-4})	Correlation	RMSE (10^{-3})	Correlation
HOAO (SKEWT)	**5.067**	0.951	**1.792**	0.961
HOAO (CF)	5.228	0.948	**1.839**	0.960
CV (SKEWT)	5.677	0.935	2.077	0.941
Normal	5.746	0.938	2.080	0.943
MLE (SKEW T)	5.786	0.936	2.068	0.943
Historical	5.866	0.933	2.104	0.940
CV (CF)	5.956	0.933	2.172	0.938

Table 1: Out-of-sample RMSE and correlation for VaR and CVaR estimation, aggregated across all confidence levels. Best-performing RMSE values are shown in bold when statistically significant at the 1% level according to a paired Wilcoxon signed-rank test. Results are based on US equity close-to-close daily returns with portfolio dimension $n = 50$, in-sample calibration window of size $\Delta t_{\text{IS}} = 100$ and out-of-sample test window of size of $\Delta t_{\text{OOS}} = 100$, with 1,000,000 experiments: we performed 10,000 bootstrapped selections, and for each selection we evaluate 100 portfolios drawn from a Dirichlet distribution.

When moving from co-moment estimation to tail-risk evaluation, the choice of the moment-to-tail map becomes an essential component of the methodology. It is particularly relevant for optimized portfolios, where the weight vector is itself a function of the estimated moments and therefore amplifies small distortions in higher-order inputs. In our empirical analysis, we observe that stabilizing higher-order tensors is necessary but not sufficient to guarantee accurate VaR/CVaR for such portfolios: the link function must be consistent with the distributional shape implied by the estimated moments, and robust to residual estimation noise. Under these constraints, the proposed pipeline yields the greatest and most stable improvements when the tail-risk functional is computed from a parametric or corrected approximation calibrated to the filtered moments, whereas purely nonparametric quantile estimation remains comparatively sensitive to window choice and regime shifts. Overall, the evidence indicates that the dominant bottleneck in practice lies upstream—in the estimation of higher-order dependence—but that translating those gains into tail-risk accuracy requires an explicit, carefully chosen mapping from moments to tail quantities.

Acknowledgments

P.R. acknowledges funding from ANRT, under the CIFRE contract nr 2025/0279.

This publication used HPC resources from the "Mésocentre" computing center of CentraleSupélec and École Normale Supérieure Paris-Saclay supported by CNRS and Région Île-de-France.

References

[1] Darrell Duffie and Jun Pan. An overview of value at risk. *Journal of derivatives*, 4(3):7–49, 1997.

[2] Evert Wipplinger. Philippe jorion: value at risk-the new benchmark for managing financial risk. *Financial Markets and Portfolio Management*, 21(3):397, 2007.

[3] R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471, 2002.

[4] Dirk Tasche. Expected shortfall and beyond. *Journal of Banking & Finance*, 26(7):1519–1533, 2002.

[5] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

[6] Carlo Acerbi, Claudio Nordio, and Carlo Sirtori. Expected shortfall as a tool for financial risk management. *arXiv preprint cond-mat/0102304*, 2001.

[7] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411, 2004.

[8] Lionel Martellini and Volker Ziemann. Improved estimates of higher-order comoments and implications for portfolio selection. *The Review of Financial Studies*, 23(4):1467–1502, 2010.

[9] Cătălin Stărică and Clive Granger. Nonstationarities in stock returns. *Review of economics and statistics*, 87(3):503–522, 2005.

[10] Rama Cont and Lakshithe Wagalath. Running for the exit: distressed selling and endogenous correlation in financial markets. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 23(4):718–741, 2013.

[11] JP Morgan et al. Creditmetrics-technical document. *JP Morgan, New York*, 1:102–127, 1997.

[12] Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters. Noise dressing of financial correlation matrices. *Physical review letters*, 83(7):1467, 1999.

[13] Imre Kondor, Szilárd Pafka, and Gábor Nagy. Noise sensitivity of portfolio selection under various risk measures. *Journal of Banking & Finance*, 31(5):1545–1573, 2007.

[14] Joël Bun, Jean-Philippe Bouchaud, and Marc Potters. Cleaning large correlation matrices: tools from random matrix theory. *Physics Reports*, 666:1–109, 2017.

[15] Olivier Ledoit and Sandrine Péché. Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields*, 151(1):233–264, 2011.

[16] Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621, 2003.

[17] Olivier Ledoit and Michael Wolf. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*.

[18] Christian Bongiorno, Damien Challet, and Grégoire Loeper. Filtering time-dependent covariance matrices using time-independent eigenvalues. *Journal of Statistical Mechanics: Theory and Experiment*, 2023(2):023402, 2023.

[19] Christian Bongiorno and Damien Challet. Covariance matrix filtering and portfolio optimisation: the average oracle vs non-linear shrinkage and all the variants of DCC-NLS. *Quantitative Finance*, 24(9):1227–1234, 2024.

[20] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

[21] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[22] Andrzej Cichocki, Danilo Mandic, Lieven De Lathauwer, Guoxu Zhou, Qibin Zhao, Cesar Caiafa, and Huy Anh Phan. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE signal processing magazine*, 32(2):145–163, 2015.

[23] Animashree Anandkumar, Rong Ge, Daniel J Hsu, Sham M Kakade, Matus Telgarsky, et al. Tensor decompositions for learning latent variable models. *J. Mach. Learn. Res.*, 15(1):2773–2832, 2014.

[24] Georgios Afendras, Nickos Papadatos, and Violetta E Piperigou. On the limiting distribution of sample central moments. *Annals of the Institute of Statistical Mathematics*, 72(2):399–425, 2020.

[25] Derrick N Joanes and Christine A Gill. Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1):183–189, 1998.

[26] R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[27] Carlo Acerbi and Dirk Tasche. On the coherence of expected shortfall. *Journal of banking & finance*, 26(7):1487–1503, 2002.

[28] Carmen Fernández and Mark FJ Steel. On bayesian modeling of fat tails and skewness. *Journal of the american statistical association*, 93(441):359–371, 1998.

[29] Kris Boudt, Brian G Peterson, and Christophe Croux. Estimation and decomposition of downside risk for portfolios with non-normal returns. *Journal of risk*, 11(2):79–103, 2008.

[30] Guy Hadash, Einat Kermany, Boaz Carmeli, Ofer Lavi, George Kour, and Alon Jacovi. Estimate and replace: A novel approach to integrating deep neural networks with existing applications. *arXiv preprint arXiv:1804.09028*, 2018.

[31] E. A. Cornish and R. A. Fisher. Moments and cumulants in the specification of distributions. *Review of the International Statistical Institute*, 5(4):307–322, 1937.

[32] Christopher S. Withers. Asymptotic expansions for distributions and quantiles with power series cumulants. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(3):389–396, 1984.

[33] Charles-Olivier Amédée-Manesme, Fabrice Barthélémy, and Didier Maillard. Computation of the corrected Cornish–Fisher expansion using the response surface methodology: application to VaR and CVaR. *Annals of Operations Research*, 281, 10 2019.

[34] Didier Maillard. A user's guide to the cornish fisher expansion. *SSRN Electronic Journal*, 02 2012.

[35] Kris Boudt, Brian G. Peterson, and Christophe Croux. Estimation and decomposition of downside risk for portfolios with non-normal returns. *The Journal of Risk*, 11(2):79–103, 2008.

[36] Stefan R. Jaschke. The cornish–fisher-expansion in the context of delta–gamma–normal approximations. Technical report, SFB 373 Discussion Paper No. 2001,54, Humboldt University of Berlin, 2001.

[37] Victor Chernozhukov, Iván Fernández-Val, and Alfred Galichon. Rearranging edgeworth–cornish–fisher expansions. *Economic Theory*, 42:419–435, 2010.

[38] Peter R. Hansen, Asger Lunde, and James M. Nason. The model confidence set. *Econometrica*, 79(2):453–497, 2011.

[39] Lieven De Lathauwer and Bart De Moor. From matrix to tensor: Multilinear algebra and signal processing. In *Institute of mathematics and its applications conference series*, volume 67, pages 1–16. Oxford University Press, 1998.

A Tensor algebra

A.1 Numerical estimation of mode matrices $\boldsymbol{U}$

Let $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{n\times\cdots\times n}$ be an order-k supersymmetric co-moment tensor. By super-symmetry of $\boldsymbol{\mathcal{M}}$, all of its unfoldings are equal: $\forall\ r \in [1,k],\ \boldsymbol{M}_r = \boldsymbol{M}_1$. However, in practice, their respective SVDs may yield different left singular vectors in case of degeneracy of the singular values and sign ambiguity. Instead of estimating k potentially different mode matrices, we enforce a shared orthogonal basis $\boldsymbol{U}$ from a single matrix unfolding of $\boldsymbol{\mathcal{M}}$. Concretely, we form the mode-1 unfolding

$$\boldsymbol{M}_{(1)} \in \mathbb{R}^{n\times n^{k-1}}, \qquad M_{(1);i,\,(j_2,\ldots,j_k)} := \mathcal{M}_{ij_2\ldots j_k}, \tag{37}$$

and compute its SVD:

$$\boldsymbol{M}_{(1)} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^\top. \tag{38}$$

We then define the core tensor by successive mode multiplications with $\boldsymbol{U}^\top$:

$$\boldsymbol{\mathcal{G}} := \boldsymbol{\mathcal{M}} \times_1 \boldsymbol{U}^\top \times_2 \boldsymbol{U}^\top \cdots \times_k \boldsymbol{U}^\top \in \mathbb{A}_k \subset \mathbb{R}^{n\times\cdots\times n}, \tag{39}$$

or expressed in coefficients,

$$\mathcal{G}_{a_1\dots a_k} = \sum_{i_1,\dots,i_k=1}^{n} U_{i_1 a_1}\cdots U_{i_k a_k}\,\mathcal{M}_{i_1\dots i_k}. \tag{40}$$

This construction is a constrained HOSVD: the equality of mode matrices is enforced by definition through the single mode matrix, rather than obtained as a property of the usual HOSVD algorithm of Ref. [20]. This allows us to reduce memory as only one orthogonal matrix is used per estimated co-moment, with substantially no loss in accuracy of tensor reconstruction. As a remark, notice that the HOSVD construction, if applied to a symmetric matrix $\boldsymbol{\Sigma}$ of EVD $\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top$ is exactly its diagonalisation. Indeed, as a matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1$ and thus the EVD of $\boldsymbol{\Sigma}$ is the SVD of $\boldsymbol{\Sigma}_1$ so that the proposed matrix core is $\boldsymbol{G} = \boldsymbol{\Sigma} \times_1 \boldsymbol{V}^\top \times_2 \boldsymbol{V}^\top = \boldsymbol{V}^\top\boldsymbol{\Sigma}\boldsymbol{V} = \boldsymbol{\Lambda}$.

A.2 Rotational invariance of co-moment tensors

Fix $\boldsymbol{U} \in O(n)$. Consider the order-k co-moment tensor of $\boldsymbol{U}\mathrm{x}$:

$$\boldsymbol{\mathcal{M}}^{(\boldsymbol{U})} \;:=\; \mathbb{E}[(\boldsymbol{U}\mathrm{x})^{\otimes k}], \qquad \mathcal{M}^{(\boldsymbol{U})}_{j_1\dots j_k} = \mathbb{E}[(\boldsymbol{U}\mathrm{x})_{j_1}\cdots(\boldsymbol{U}\mathrm{x})_{j_k}]\,.$$

Using $(\boldsymbol{U}\mathrm{x})_j = \sum_{i=1}^n U_{ji}x_i$ and multilinearity of the multilinear product, we compute componentwise:

$$\begin{aligned}
\mathcal{M}^{(\boldsymbol{U})}_{j_1\dots j_k} &= \mathbb{E}\left[\prod_{m=1}^{k}(\boldsymbol{U}\mathrm{x})_{j_m}\right] = \mathbb{E}\left[\prod_{m=1}^{k}\left(\sum_{i_m=1}^{n} U_{j_m i_m}\,x_{i_m}\right)\right] \\
&= \mathbb{E}\left[\sum_{i_1,\dots,i_k=1}^{n}\left(\prod_{m=1}^{k} U_{j_m i_m}\right) x_{i_1}\cdots x_{i_k}\right] \\
&= \sum_{i_1,\dots,i_k=1}^{n}\left(\prod_{m=1}^{k} U_{j_m i_m}\right)\mathbb{E}[x_{i_1}\cdots x_{i_k}] \\
&= \sum_{i_1,\dots,i_k=1}^{n} U_{j_1 i_1}\cdots U_{j_k i_k}\,\mathcal{M}_{i_1\dots i_k}.
\end{aligned}$$

Where the fourth equality stems from linearity of the expectation assuming $\mathbb{E}[|x_{i_1}\cdots x_{i_k}|] < \infty$.

We recognize the last expression, by definition, as the multilinear product with $\boldsymbol{U}$ applied on all modes of :

$$\boldsymbol{\mathcal{M}}^{(\boldsymbol{U})} = \boldsymbol{\mathcal{M}} \times_1 \boldsymbol{U}\cdots\times_k \boldsymbol{U}. \tag{41}$$

Assuming rotational invariance of the random variable associated to the vector of returns $\mathbf{x} = (x_1,\dots,x_n)$, meaning $\forall\, \boldsymbol{U} \in O(n),\ \mathbf{x} \overset{d}{=} \boldsymbol{U}\mathbf{x}$, we have $\mathbb{E}[f(\mathbf{x})] = \mathbb{E}[f(\boldsymbol{U}\mathbf{x})]$ for any measurable function $f : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbb{E}|f(\mathbf{x})| < \infty$. Applying this with the element-wise integrable map $f_{i_1\dots i_k}(z) := z_{i_1}\cdots z_{i_k}$ yields

$$\mathbb{E}[\mathbf{x}^{\otimes k}] = \mathbb{E}[(\boldsymbol{U}\mathbf{x})^{\otimes k}].$$

We can finally conclude by

$$\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{M}} \times_1 \boldsymbol{U}\cdots\times_k \boldsymbol{U}, \qquad \forall\, \boldsymbol{U} \in O(n). \tag{42}$$

This allows us to define Rotationally Invariant Estimators (RIE) of $\boldsymbol{\mathcal{M}}$ as an estimator $\widehat{\boldsymbol{\mathcal{M}}}$ under the assumption of rotational invariance of the random varibale $\mathbf{x}$, which boils down to finding an estimator of the core $\widehat{\boldsymbol{\mathcal{G}}}$ of the HOSVD of $\widehat{\boldsymbol{\mathcal{M}}}$, solution of the HOSVD problem.

A.3 Proof of Proposition 2 (Sec. 2)

Let $k \geq 2$ and let $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{n\times\cdots\times n}$. Fix mode rotation matrices $\boldsymbol{U} \in O(n)$ and define the multilinear map associated to the HOSVD under the assumption of a unique rotation, that is assuming no rotational ambiguity :

$$\widehat{\boldsymbol{\mathcal{M}}}(\boldsymbol{\mathcal{G}}) \;:=\; \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{U}\ldots\times_k \boldsymbol{U}, \qquad \boldsymbol{\mathcal{G}} \in \mathbb{A}_k. \tag{43}$$

We equip the considered tensor space $\mathbb{R}^{n^{\otimes^k}}$ with its canonical (Frobenius) inner product, so that for $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ tensors

$$\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle \;:=\; \sum_{i_1,\ldots,i_k} \mathcal{A}_{i_1\ldots i_k}\, \mathcal{B}_{i_1\ldots i_k}, \qquad \|\boldsymbol{\mathcal{A}}\|_F^2 := \langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{A}} \rangle.$$

For any core tensor $\boldsymbol{\mathcal{G}}$, the map in (43) preserves the Frobenius norm:

$$\|\widehat{\boldsymbol{\mathcal{M}}}(\boldsymbol{\mathcal{G}})\|_F = \|\boldsymbol{\mathcal{G}}\|_F. \tag{44}$$

Indeed, by writing the multilinear rotation entrywise:

$$\widehat{\mathcal{M}}(\mathcal{G})_{j_1\ldots j_k} = \sum_{a_1=1}^{n} \cdots \sum_{a_k=1}^{n} U_{j_1 a_1} \cdots U_{j_k a_k}\, \mathcal{G}_{a_1\ldots a_k}.$$

Then

$$\begin{aligned}\|\widehat{\boldsymbol{\mathcal{M}}}(\boldsymbol{\mathcal{G}})\|_F^2 &= \sum_{j_1,\ldots,j_k} \widehat{\mathcal{M}}(\mathcal{G})_{j_1\ldots j_k}\, \widehat{\mathcal{M}}(\mathcal{G})_{j_1\ldots j_k} \\ &= \sum_{j_1,\ldots,j_k} \sum_{a_1,\ldots,a_k} \sum_{b_1,\ldots,b_k} \left(\prod_{m=1}^{k} U_{j_m a_m}\, U_{j_m b_m} \right) \mathcal{G}_{a_1\ldots a_k}\, \mathcal{G}_{b_1\ldots b_k}.\end{aligned}$$

Separation of the sums yields

$$\|\widehat{\boldsymbol{\mathcal{M}}}(\boldsymbol{\mathcal{G}})\|_F^2 = \sum_{a_1,\ldots,a_k} \sum_{b_1,\ldots,b_k} \left(\prod_{m=1}^{k} \sum_{j_m} U_{j_m a_m} U_{j_m b_m} \right) \mathcal{G}_{a_1\ldots a_k}\, \mathcal{G}_{b_1\ldots b_k}.$$

Since $\sum_{j_m} U_{j_m a_m} U_{j_m b_m} = \delta_{a_m b_m}$, the inner product simplifies to

$$\|\widehat{\boldsymbol{\mathcal{M}}}(\boldsymbol{\mathcal{G}})\|_F^2 = \sum_{a_1,\ldots,a_k} \mathcal{G}_{a_1\ldots a_k}^2 = \|\boldsymbol{\mathcal{G}}\|_F^2.$$

We next define the (adjoint) map:

$$\widehat{\boldsymbol{\mathcal{M}}}^{\,*}(\boldsymbol{\mathcal{S}}) \;:=\; \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{U}^\top \cdots \times_k \boldsymbol{U}^\top, \qquad \boldsymbol{\mathcal{S}} \in \mathbb{R}^{n\times\cdots\times n}. \tag{45}$$

Then for all $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{n\times\cdots\times n}$ and all $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{n\times\cdots\times n}$,

$$\langle \widehat{\boldsymbol{\mathcal{M}}}(\boldsymbol{\mathcal{G}}),\, \boldsymbol{\mathcal{S}} \rangle \;=\; \langle \boldsymbol{\mathcal{G}},\, \widehat{\boldsymbol{\mathcal{M}}}^{\,*}(\boldsymbol{\mathcal{S}}) \rangle. \tag{46}$$

We derive the identity (46) using the expression of $\widehat{\boldsymbol{\mathcal{M}}}(\boldsymbol{\mathcal{G}})$ written coefficient-wise:

$$\begin{aligned}\langle \widehat{\boldsymbol{\mathcal{M}}}(\boldsymbol{\mathcal{G}}),\, \boldsymbol{\mathcal{S}} \rangle &= \sum_{j_1,\ldots,j_k} \widehat{\mathcal{M}}(\mathcal{G})_{j_1\ldots j_k}\, \mathcal{S}_{j_1\ldots j_k} \\ &= \sum_{j_1,\ldots,j_k} \sum_{a_1,\ldots,a_k} \left(\prod_{m=1}^{k} U_{j_m a_m} \right) \mathcal{G}_{a_1\ldots a_k}\, \mathcal{S}_{j_1\ldots j_k}.\end{aligned}$$

Interverting the two sums allows to recognize the adjoint map acting on $\boldsymbol{\mathcal{S}}$:

$$\langle \widehat{\boldsymbol{\mathcal{M}}}(\boldsymbol{\mathcal{G}}),\, \boldsymbol{\mathcal{S}} \rangle = \sum_{a_1,\ldots,a_k} \mathcal{G}_{a_1\ldots a_k} \underbrace{\sum_{j_1,\ldots,j_k} \left(\prod_{m=1}^{k} U_{j_m a_m} \right) \mathcal{S}_{j_1\ldots j_k}}_{=\,(\widehat{\mathcal{M}}^{\,*}(\mathcal{S}))_{a_1\ldots a_k}} = \langle \boldsymbol{\mathcal{G}},\, \widehat{\boldsymbol{\mathcal{M}}}^{\,*}(\boldsymbol{\mathcal{S}}) \rangle, \tag{47}$$

which is (46).

Optimality of $\mathcal{G}^\star$

Consider the least-squares problem

$$\mathcal{G}^\star = \arg\min_{\mathcal{G}\in\mathbb{R}^{n\times\cdots\times n}} \|\widehat{\mathcal{M}}(\mathcal{G}) - \mathcal{M}\|_F^2. \tag{48}$$

In order to derive the result of equation 2, we start by expanding the Frobenius norm written as an inner product:

$$\|\widehat{\mathcal{M}}(\mathcal{G}) - \mathcal{M}\|_F^2 = \|\widehat{\mathcal{M}}(\mathcal{G})\|_F^2 - 2\langle\widehat{\mathcal{M}}(\mathcal{G}),\, \mathcal{M}\rangle + \|\mathcal{M}\|_F^2.$$

Using Eq. (44) and Eq. (46), we obtain

$$\begin{aligned}\|\widehat{\mathcal{M}}(\mathcal{G}) - \mathcal{M}\|_F^2 &= \|\widehat{\mathcal{M}}(\mathcal{G})\|_F^2 - 2\langle\widehat{\mathcal{M}}(\mathcal{G}),\, \mathcal{M}\rangle + \|\mathcal{M}\|_F^2 \\ &= \|\mathcal{G}\|_F^2 - 2\langle\mathcal{G},\, \widehat{\mathcal{M}}^{\,*}(\mathcal{M})\rangle + \|\mathcal{M}\|_F^2 \\ &= \|\mathcal{G} - \widehat{\mathcal{M}}^{\,*}(\mathcal{M})\|_F^2 + \|\mathcal{M}\|_F^2 - \|\widehat{\mathcal{M}}^{\,*}(\mathcal{M})\|_F^2.\end{aligned}$$

Since the last two terms are constant in $\mathcal{G}$, the unique minimizer is $\mathcal{G}^\star = \widehat{\mathcal{M}}^{\,*}(\mathcal{M})$, which yields equation (27). The result is not surprising as $||\widehat{\mathcal{M}}(\mathcal{G}^\star) - \mathcal{M}\|_F = 0$.

This solution also provides an optimal core for the HOSVD of an estimator of a co-moment tensor of order $k \geq 2$ of a random vector under the assumption of its rotational invariance, yielding an optimal RIE in the sense of enforcing a mode basis and replacing the core as to minimize the Frobenius distance to a target tensor.

A.4 Oracle under rotational invariance

The tensor $\mathcal{G}^\star$ of Eq. 27 is in general not an element of the set $\mathbb{A}_k := \{\mathcal{A} \in \mathbb{R}^{n_1\times\cdots\times n_k} \mid \forall\, r \leq k,\ \forall\, \alpha,\ \beta \leq n_r,\ \left\langle \mathcal{A}_{\,i_r=\alpha}, \mathcal{A}_{\,i_r=\beta} \right\rangle = 0 \text{ if } \alpha \neq \beta\}$, that is to say the set of admissible HOSVD core tensors. Note that $\mathcal{G}^\star$ is however supersymmetric by construction from the supersymmetric population co-moment.

However, the HOSVD of the unconstrained, supersymmetric, Oracle core reads $\mathcal{G}^\star = \mathcal{G}_{\mathcal{G}^\star} \times_1 \boldsymbol{U}_{\mathcal{G}^\star} \ldots \times_k \boldsymbol{U}_{\mathcal{G}^\star}$ where we do have $\mathcal{G}_{\mathcal{G}^\star} \in \mathbb{A}_k$. Under the assumption of rotational invariance, then we know that $\mathcal{M}_{\text{population}}$ is preserved by multilinear rotation, thus $\mathcal{M}_{\text{population}} = \mathcal{G}^\star = \mathcal{G}_{\mathcal{G}^\star} \in \mathbb{A}_k$.

Therefore, when using Eq. (27) to estimate $\widehat{\mathcal{G}}^\star$ from finite samples of realized returns, we can assume that on average, if the assumption of rotational invariance holds, then the estimated core $\widehat{\mathcal{G}}$ belongs to $\mathbb{A}_k$. In other words, we assume that the averaging procedure of Sec. 3.1 and 3.2 makes the estimated tensor cores admissible HOSVD cores. This assumption seems to hold relatively well considering the empirical overperformance of proposed estimators, and the observed shrinkage to the 0-tensor of the estimated coskewness, which is to be expected for rotationally invariant returns. A more detailed analysis of this question if given in appendix B.1.

Let us finally remark that, ideally, the tensor Oracle solves:

$$\mathcal{G}^{\text{Oracle}} = \arg\min_{\mathcal{G}\in\mathbb{A}} \|\widehat{\mathcal{M}}(\mathcal{G}) - \mathcal{M}\|_F^2, \tag{49}$$

for which we did not find any closed form solution. To be more precise, we lack the orthogonal projector, in the Frobenius sense, $\Pi_{\mathbb{A}_k} : \mathbb{R}^{n^{\otimes^k}} \to \mathbb{A}_k$, such that

$$\mathcal{G}^{\text{Oracle}} = \Pi_{\mathbb{A}}(\mathcal{G}^\star). \tag{50}$$

In the matrix case, we know that $\Pi_{\mathbb{A}}$ is precisely $\mathrm{Diag}()$, which we therefore use in the empirical analysis to stricly enforce all-orthogonality of the filtered covariance eigenvalues $\widehat{\mathcal{G}}^{(2)}$.

B Empirical Analysis

B.1 Do $\widehat{\mathcal{G}}^{(k)}_{\text{CV/HOAO}} \in \mathbb{A}_k$ for $k \in \{2,3,4\}$?

Following the calculations of [39], membership in the structured set $\mathbb{A}_k$ can be characterized through the diagonality of a Gram matrix built from an unfolding of the estimated core tensor. More precisely, for each moment order $k \in \{3,4\}$

Model m	$\mathbb{E}[\|\mathrm{Gram}_m^{(3)}\|_F]$	$\mathbb{E}[\delta_m^{(3)}]$	$\mathbb{E}[\|\mathrm{Gram}_m^{(4)}\|_F]$	$\mathbb{E}[\delta_m^{(4)}]$
CV	25.9	0.40	1380	0.22
HOAO	1.8	0.38	1500	0.013

Table 2: Mean relative Gram distance to diagonality $\mathbb{E}[\delta_m^{(k)}]$ (in (52)) over experiments, for $k \in \{3,4\}$ and $m \in \{\mathrm{CV}, \mathrm{HOAO}\}$.

and each model $m \in \{\mathrm{CV}, \mathrm{HOAO}\}$, let $\widehat{\boldsymbol{\mathcal{G}}}_m^{(k)}$ denote the estimated (super-symmetric) core tensor. Let $\big(\widehat{\boldsymbol{\mathcal{G}}}_m^{(k)}\big)_{(1)}$ be its mode-1 unfolding (the choice of mode is without loss of generality under super-symmetry), and define the associated Gram matrix by

$$\mathrm{Gram}_m^{(k)} \;:=\; \big(\widehat{\boldsymbol{\mathcal{G}}}_m^{(k)}\big)_{(1)}\big(\widehat{\boldsymbol{\mathcal{G}}}_m^{(k)}\big)_{(1)}^{\top}. \tag{51}$$

In this setting, $\widehat{\boldsymbol{\mathcal{G}}}_m^{(k)} \in \mathbb{A}_k$ if and only if $\mathrm{Gram}_m^{(k)}$ is diagonal. Therefore, we quantify the deviation from $\mathbb{A}_k$ through the relative off-diagonal norm

$$\delta_m^{(k)} \;:=\; \frac{\big\|\mathrm{Gram}_m^{(k)} - \mathrm{Diag}\big(\mathrm{Gram}_m^{(k)}\big)\big\|_F}{\big\|\mathrm{Gram}_m^{(k)}\big\|_F} \;\in\; [0,1], \tag{52}$$

where $\mathrm{Diag}(\cdot)$ denotes the diagonal matrix obtained by zeroing out off-diagonal entries.

Since $\mathrm{Diag}(\cdot)$ is the orthogonal projector onto the subspace of diagonal matrices for the Frobenius inner product, the numerator in (52) is precisely the residual of the Euclidean (Frobenius) projection, which makes the Frobenius loss the natural criterion to minimize when projecting onto $\mathbb{A}_k$. In the worst case, when the diagonal part vanishes, the residual equals the full norm and $\delta_m^{(k)} = 1$. Importantly, in the case $\widehat{\boldsymbol{\mathcal{G}}}_m^{(k)} = \mathbf{0} \in \mathbb{A}$ the proposed error metric becomes undefined because of the scaling factor, and caution is advised when investigating $\widehat{\boldsymbol{\mathcal{G}}}_m^{(k)} \approx 0$.

We do not report this diagnostic for $k = 2$, since in practice we enforce full diagonality

$$\widehat{\boldsymbol{\mathcal{G}}}_{\mathrm{CV/HOAO}}^{(2)} \leftarrow \mathrm{Diag}\big(\widehat{\boldsymbol{\mathcal{G}}}_{\mathrm{CV/HOAO}}^{(2)}\big).$$

Table 2 reports the measured mean value of $\delta_m^{(k)}$ for $k \in \{3,4\}$ along with $\big\|\mathrm{Gram}_m^{(k)}\big\|_F$. While the relative deviation is non-negligible for HOAO coskewness ($k = 3$), it should be interpreted with care as the residual $\big\|\mathrm{Gram}_m^{(k)}\big\|_F$ is small for a tensor of 50^3 entries, and may be dominated with noise.

B.2 Model Confidence Set

To compare multiple estimators while accounting for data dependence, we compute a Model Confidence Set (MCS) following Hansen et al. (2011). For each method m, we form the loss series $\{\ell_{m,t}(\alpha)\}_{t=1}^{\Delta t}$ (here, RMSE for VaR/CVaR at percentile α, aggregated into RMSE). The MCS procedure tests the null hypothesis of Equal Predictive Ability across the candidate set using loss differentials

$$d_{i,j,t}(\alpha) \;=\; \ell_{i,t}(\alpha) \;-\; \ell_{j,t}(\alpha). \tag{53}$$

The distribution of the MCS test statistic is approximated via a moving block bootstrap over t with block length $L = 1000$. At significance level $\delta = 0.001$, the MCS returns the subset of methods that cannot be statistically distinguished from the best performer under the chosen RMSE loss for VaR and CVaR accuracy estimates.

The reported results in Fig. 3 establish clearly that, in the considered setting, HOAO-based risk models outperform other proposed models with high degree of significance. The same MCS is obtained for different dimensions, ranging from $n = 10$ to $n = 100$ in our experiments.