Robust portfolio selection using sparse estimation of comoment tensors

Nathan Lassance and Frédéric Vrins
Robust portfolio selection using sparse estimation of comoment tensors

Nathan Lassance
UCLouvain (BE), Louvain Finance
nathan.lassance@uclouvain.be

Frédéric Vrins
UCLouvain (BE), Louvain Finance, Center for Operations Research and Econometrics
frederic.vrins@uclouvain.be

November 2019

Abstract
It is well known that estimation issues severely impact the performances of investment strategies. This becomes even more problematic when accounting for higher moments as the number of parameters to be estimated quickly explodes with the number of assets. In this paper, we address this issue by relying on specific factor models. Although useful to reduce the dimension of the problem, principal component analysis (PCA) is only a partial solution. In particular, it does not break the exponential law that links the number of parameters to the moment order. This issue is tackled by using a new robust portfolio-selection technique that relies on independent component analysis (ICA). By linearly projecting the asset returns onto a small set of maximally independent factors, we obtain a sparse approximation of the comoment tensors of asset returns. This drastically decreases the dimensionality of the problem and, as expected, leads to well-performing, robust and low-turnover investment strategies.

Keywords
Finance, portfolio selection, robustness, independent component analysis, higher moments.

1. Introduction

Robustness is one of the main concerns when it comes to the realized performance of optimal portfolio strategies. Portfolios that are too sensitive to the provided asset-return inputs are de facto disappointing in terms of out-of-sample performance, and require intensive rebalancing transactions. Robustness is particularly needed in high dimensions, because the number of (co)moments to estimate is considerable compared to the sample sizes generally available.
Various techniques have been introduced to improve the robustness of mean-variance portfolios; see Fabozzi et al. (2010), Section 3 of Kolm et al. (2014) and Levy and Levy (2014) for reviews. Beyond mean and variance, investors also display preferences for higher moments (Scott and Horvath 1980). In particular, Harvey and Siddique (2000) and Ang et al. (2006) show that the cross-section of stock returns reflects a large premium for downside risk. As a result, higher-moment portfolio strategies receive increasing attention in the literature; see Briec et al. (2013) and Section 3.3.5 of Kolm et al. (2014) for reviews. However, very few studies are devoted to the analysis of robust higher-moment portfolios. One possible explanation for this can be found in Brandt et al. (2009, p.3418): “[...] extending the traditional approach beyond first and second moments, when the investor’s utility function is not quadratic, is practically impossible because it requires modeling not only the conditional skewness and kurtosis of each stock but also the numerous high-order cross-moments.” Nonetheless, some authors have tackled this problem via Bayesian estimation (Harvey et al. 2010), shrinkage estimation (Martellini and Ziemann 2010, Boudt et al. 2018) and factor models (Boudt et al. 2015, 2019).

In this paper, we introduce a robust portfolio-selection strategy that combines dimension reduction with a specific factor model. The idea consists in exploiting the natural sparsity of the comoment tensors associated with independent factors. The method is intuitive, computationally efficient and suited for any portfolio strategy based on moments of any order. In a first step, we show that reducing dimension by projecting the $N$ asset returns onto the basis formed by the first $K$ principal components (PCs) helps to reduce the number of parameters in comoment tensors. However, the PCs are simply uncorrelated; there is no reason why they would be independent, too. Therefore, many higher comoments remain to estimate; the number of parameters in the $k$th higher-comoment tensor of PCs is of the order $K^k$. To address this problem, we propose to rely in a second step on independent component analysis (ICA), a statistical technique widely used in machine learning and data processing (Hyvärinen et al. 2001). We assume a factor model in which the asset returns are a linear mixture of independent components (ICs). The latter simply correspond to the rotation of PCs that are maximally independent. By neglecting the remaining higher-moment dependence between ICs, we obtain sparse approximations of the comoment tensors. In particular, the number of parameters in the $k$th higher-comoment tensor of ICs becomes linear in $K$ as only the marginal moments matter. An empirical analysis on Fama-French industry datasets shows the usefulness of our approach in terms of out-of-sample performance and turnover, especially in high dimensions. In particular, our results put some of the conclusions of Martellini and
Ziemann (2010) into perspective. Using sample, sparse and shrinkage comoment-tensor estimates, they conclude that, for monthly returns, there are no benefits in choosing a higher-moment portfolio strategy over the minimum-variance portfolio. In contrast, using monthly returns as well, we observe that our sparse low-dimensional estimation of tensors based on ICA succeeds in dominating the minimum-variance portfolio.

The proposed portfolio-selection strategy exploits the idea introduced in Section 5.1 of Lassance et al. (2019). In the aforementioned paper, the authors use ICA as a way to resolve the complete indeterminacy that affects the factor-risk-parity approach when relying solely on a decorrelation criterion for the choice of factors. In the special case where the risk measure used for factor-risk parity is impacted by higher moments, they notice that working with the ICs as factors triggers an additional computational advantage. Indeed, the latter being maximally independent, one can make the economy of their comoments estimation. In this paper, we use the same trick, but in a different context, and for different purposes. Instead of considering risk-parity strategies, we deal with traditional minimum-risk portfolios. Our goal is not to diversify a given risk measure across ICs but is to enhance the robustness of minimum-risk portfolios, in the same vein as shrinkage techniques. To do so, we first decompose the portfolio return as a weighted sum of ICs, neglect the comoments between the latter, and only then solve the optimization problem. Compared to the traditional approach, the resulting objective function features a much smaller number of parameters, especially when working in high-dimensional spaces. In the context of this paper, working with the ICs is only relevant when dealing with higher-moment risk measures. This stresses the difference with Lassance et al. (2019) who find that working with the ICs in factor-risk-parity strategies remains beneficial in terms of portfolio-return tail risk even when using variance as risk measure.

Our work is related to other approaches in the literature dealing with robust estimation of the higher comoments of portfolio returns. The most popular framework is the use of shrinkage estimators of comoment tensors, which are a combination between the sample estimate and a sparse target estimate. It has mainly been used for the covariance matrix; see Ledoit and Wolf (2004, 2017). Analogs for coskewness and cokurtosis tensors have been developed by Martellini and Ziemann (2010) and Boudt et al. (2018). The goal of shrinkage estimators is similar to that of the sparse estimators introduced in this paper. However, shrinkage still requires the sample estimator as input, which is made of many parameters, and crucially depends on the non-trivial choice of target estimator. In contrast, our sparse tensor estimates only require to determine the number $K$ of retained PCs. Several standard rules are available for this purpose.
More closely related to our work, factor models are also used to estimate comoment tensors in portfolio optimization. The vast majority of those studies concentrates on the estimation of the covariance matrix; see De Nard et al. (2019) for a recent review. Concerning higher-comoment tensors, Boudt et al. (2015) show that projecting the asset returns on \( K \) factors helps to reduce the number of parameters but, just like for the PCs above, the growth in the number of parameters remains of order \( K^k \). Jondeau et al. (2018) introduce moment component analysis, an extension of PCA to higher moments. The moment components are the directions accounting for most of the variations in selected comoment tensors, and can be used to obtain improved tensor estimates by ignoring irrelevant directions. The factors considered in the above papers do not exhibit a special dependence structure; they are not uncorrelated or independent, in general. As a result, many higher comoments remain to estimate. Instead, our approach precisely aims at working in a specific system of coordinates, in order to fully exploit the natural sparsity of the comoment tensors associated with independent factors. By doing so, we make the economy of the estimation of many comoments.

Other authors rely, just like us, on independent factor models in portfolio optimization. Chen et al. (2007) rely on ICA for forecasting portfolio Value-at-Risk (VaR). Ghalanos et al. (2015) introduce a dynamic autoregressive independent factor model for the same purpose. Both papers focus on the parametric density estimation of the independent factors via normal inverse Gaussian and generalized hyperbolic distributions, respectively. They both obtain more accurate VaR forecasts than via classical methods such as GARCH. In contrast, our focus in this paper is not on VaR forecasting and density estimation, but on robust moment-based portfolios via sparse comoment tensors obtained with ICA. Moreover, both papers do not apply dimension reduction (\( K = N \)). Hitaj et al. (2015) rely on ICA to estimate the maximum-CRRA-utility portfolio. Again, their focus is on the density estimation of the ICs via several parametric distributions.

The closest paper to ours is the recent work of Boudt et al. (2019). They consider a latent independent factor model and, in doing so, obtain improved estimates of higher-comoment tensors called nearest comoment estimates. They apply their estimates to different portfolio strategies—minimum VaR, maximum CRRA expected utility and the shortage-function approach of Briec et al. (2007)—and find improvements compared to sample estimators. The main benefit of our approach compared to theirs is that ours is better suited to high-dimensional settings, where robustness is particularly important. Indeed, the estimation methodology in Boudt et al. (2019) is computationally intensive. As they explain (\( p \) is the number of assets): “A computational disadvantage of the proposed estimator is the use of a weight matrix of dimension equal to the number of sample mo-
ments, thus growing as $O(p^4)$. In the simulations and empirical applications, we therefore consider only settings of up to $p = 10$, which is still highly relevant to asset allocation [...]”. In contrast, our approach based on ICA is computationally efficient even in large dimensions, for two reasons. First, all inputs needed to obtain the tensor estimates are easy to compute even for large $N$: we only need the eigendecomposition of the covariance matrix, the rotation matrix that deliver the ICs (very efficient numerical procedures are available to this end such as FastICA) and, finally, the marginal moments of the ICs. Second, as we show, dimension reduction also helps to reduce the dimension of the optimization problem: in our framework, the problem of finding the $N$ optimal portfolio weights simplifies to that of finding the $K$ optimal factor exposures.

The article is organized as follows. In Section 2, we show how dimension reduction via PCs helps to reduce the dimensionality of comoment tensors. In Section 3, we introduce our sparse estimation procedure based on ICA. In Section 4, we evaluate the out-of-sample performance of the proposed portfolio strategies. Section 5 concludes. The proofs of all results are relegated to the appendix.

2. Dimension reduction

Let $X \in \mathbb{R}^N$ be the column vector of asset returns with mean $\mu$ and covariance matrix $\Sigma$, $w \in \mathcal{W}$ the portfolio weights with

$$\mathcal{W} := \{v \in \mathbb{R}^N \mid 1'v = 1\},$$

and $P = P(w) := w'X$ the associated portfolio return. In the sequel, we rely on centered returns $X \leftarrow X - \mu$ because modeling the asset returns $X$ as a linear combination of $K$ factors $Y_K$ as $\mu + AY_K$ does not help to improve the estimates of asset mean returns. Moreover, including $\mu$ in the objective function often deteriorates the out-of-sample performance of optimal portfolios (Merton 1980, DeMiguel et al. 2009, 2009a).

2.1. Curse of dimensionality

We consider any portfolio strategy that minimizes a function of (central) moments:

$$w_\psi := \arg\min_{w \in \mathcal{W}} \psi \left( m_q(P(w)), m_r(P(w)), \ldots, m_s(P(w)) \right),$$

where $2 \leq q \leq r \leq s$ are integers, $m_k(P) := \mathbb{E}[\{P - \mathbb{E}[P]\}^k]$, and $s$ is the maximum moment order considered. Notice that this framework is very general. For instance, it encompasses maximum-
expected-utility portfolios as a special case, via a truncated Taylor series expansion as in Jondeau and Rockinger (2006) and Martellini and Ziemann (2010) for example. The $k$th moment $m_k(P)$ can be decomposed as

$$m_k(P) = w'M_k(X) \bigotimes_{i=1}^{k-1} w,$$

where $\bigotimes_{i=1}^{k} w := w \otimes \cdots \otimes w$ is the tensor product of $k$ $w$, and $M_k(X)$ is the $k$th comoment tensor of $X$:

$$M_k(X) := \mathbb{E} \left[ X \bigotimes_{i=1}^{k-1} X' \right].$$

Without specific assumptions on $X$, the cardinality (i.e. the number of distinct parameters to estimate) of the $N^k$-dimensional tensor $M_k(X)$ is

$$\#(M_k(X)) = \binom{N + k - 1}{k} = O(N^k).$$

Thus, for portfolio strategies based on a large number of assets $N$ and/or on higher moments $k > 2$, the number of parameters impacting the $k$th moment of $P(w)$ quickly becomes unmanageable. This severely affects the robustness and out-of-sample performance of the portfolio $w_\psi$. In the sequel, we refer to this problem as the curse of dimensionality.

### 2.2. Reducing dimensionality via principal components

A first common approach to address the curse of dimensionality is to reduce dimension (Xu et al. 2016). More precisely, we approximate the asset returns $X$ by $\hat{X}$, a well-chosen linear mixture of $K \leq N$ standardized factors. Mathematically, we seek for an approximation $\hat{X}$ of the form

$$\hat{X} = A_K Y_K,$$

where $A_K$ is the $N \times K$ loading matrix and the $K$ factors $Y_K$ are uncorrelated and have unit variance, $M_2(Y_K) = I_K$. It is well known that the best approximation $\hat{X}$ of $X$ in the least-squares sense is given by Principal Component Analysis (PCA). Defining $V_K \in \mathbb{R}^{N \times K}$ and $A_K \in \mathbb{R}^{K \times K}$ the eigenvectors and eigenvalues matrices associated with the $K$ largest eigenvalues of $\Sigma$, the optimal

\footnote{For instance, the cardinality of an $N \times N$ covariance matrix is $\binom{N+1}{2} = \frac{(N+1)!}{(N+1-2)!2!} = N(N + 1)/2.$
factors and loading matrix are given by
\[ Y_K := \Lambda_K^{-1/2} V_K' X \quad \text{and} \quad A_K := V_K \Lambda_K^{1/2}. \] (7)

Our lower-dimensional representation of the asset returns simply becomes \( \hat{X} = V_K V_K' X \). We assume that \( \text{rank}(\Sigma) \geq K \), so that \( \Lambda_K \) is invertible. The approximation \( \hat{P} \) of the portfolio return \( P \) resulting from ignoring the last \( N - K \) PCs is then obtained as
\[ \hat{P} := w' \hat{X} = w' V_K \Lambda_K^{1/2} Y_K = \tilde{w}' Y_K; \] (8)
where the factor exposures \( \tilde{w} \) and portfolio weights \( w \) are related as follows:
\[ w = V_K \Lambda_K^{-1/2} \tilde{w}. \] (9)

Since the rank of \( \Sigma \) is assumed to be larger than \( K \), there is a loss of information resulting from the projection step. Thus, \( \hat{P} = w' \hat{X} \) is different from \( P = w' X \) unless \( K = N \). We can now estimate the portfolio \( w_\psi \) by finding the factor exposures \( \tilde{w} \) optimizing the moments of \( \hat{P} \), the best approximation of \( P \) in the least-squares sense.

**Definition 1.** The PC-estimate of the portfolio \( w_\psi \) in (2) is
\[ \tilde{w}_{\psi,K} := V_K \Lambda_K^{-1/2} \tilde{w}_{\psi,K}, \] (10)
where
\[ \tilde{w}_{\psi,K} := \arg\min_{\tilde{w} \in \tilde{W}} \psi(m_q(\hat{P}), m_r(\hat{P}), \ldots, m_s(\hat{P})) \] (11)
with \( \hat{P} = \tilde{w}' Y_K \) and
\[ \tilde{W} := \left\{ v \in \mathbb{R}^K \mid V_K \Lambda_K^{-1/2} v \in W \right\}. \] (12)

**Remark 1.** When \( K = 1 \), the PC-estimate reduces to \( \tilde{w}_{\psi,1} = V_1 / V_1' V_1 \) and is independent of \( \psi \). Thus, we restrict to \( K \geq 2 \) in the sequel.

The above problem can also be formulated as an optimization with respect to \( w \in W \) directly since we also have \( \hat{P} = w' \hat{X} \). However, problem (11) is computationally much more efficient as only \( K \) variables are optimized instead of \( N \). Thus, in contrast with classical methods used to obtain improved comoment-tensor estimates (see Section 1), we do not need to plug high-dimensional asset-
return comoment tensors to compute the optimal portfolio. Instead, only the comoment tensors of the $K$ PCs $Y_K$ are needed in (11) to compute the moments of $\hat{P}$, which substantially speeds up the optimization.

It is worth noting that the PC-estimate $\hat{w}_{\psi,K}$ focuses on the $K$ directions with the largest eigenvalues. These are the most relevant and, moreover, are most precisely estimated, which is expected to improve the robustness of the portfolio. This will be shown empirically in Section 4: the portfolio with dimension reduction $\hat{w}_{\psi,K}$ performs better than $w_\psi$, and requires less turnover.

2.3. Approximation of comoment tensors

By disregarding the contribution of the $N - K$ least relevant dimensions, the projection step allows us to reduce the dimensionality of the problem. In the next proposition, we formalize this result by showing how the comoment tensors of asset returns simplify when assuming the dimension-reduction factor model based on PCs in (6).

**Proposition 1.** The $k$th comoment tensor of the projected asset returns $\hat{X} = A_K Y_K$ is given by

$$M_k(\hat{X}) = A_K M_k(Y_K) \bigotimes_{i=1}^{k-1} A'_K,$$

and has the cardinality $\#(M_k(\hat{X})) = \#(A_K) + \#(M_k(Y_K))$, where

$$\#(A_K) = KN - \frac{K(K-1)}{2},$$

$\#(M_2(Y_K)) = 0$ and, for $k \geq 3$,

$$\#(M_k(Y_K)) = \binom{K + k - 1}{k} = O(K^k).$$

In the following corollary, we consider the minimum-variance portfolio that is known to work very well in practice, precisely due to its remarkable robustness (DeMiguel et al. 2009, Bodnar et al. 2017). It corresponds to solving (2) with $\psi(m_2(P)) = m_2(P) = w' \Sigma w$.

**Corollary 1.** The PC-estimate of the minimum-variance portfolio $w_{MV}$ is given by

$$\hat{w}_{MV,K} := \frac{\hat{\Sigma}^{-1}1}{1'\hat{\Sigma}^{-1}1},$$

where $\hat{\Sigma} := V_K \Lambda K V_K'$, and $\#(\hat{\Sigma}) = \#(A_K)$ in (14).
The minimum-variance portfolio $\mathbf{w}_{MV}$ corresponds to $K = N$, that is, amounts to replace $\hat{\Sigma}$ by $\Sigma = \mathbf{V}_N \Lambda_N \mathbf{V}_N'$ in (16). The above corollary shows that even when focusing on covariance-based strategies, dimension reduction helps to reduce the number of parameters needed to estimate $\Sigma$ as $\sharp(\hat{\Sigma})$ is much lower than $\sharp(\Sigma) = N(N+1)/2$ for $K$ small relative to $N$.

Estimates of the covariance matrix via PCs of the form $\hat{\Sigma}$ are quite standard in factor modeling (De Nard et al. 2019), although a notable difference is that, in our approach, we do not try to estimate the covariance matrix of the idiosyncratic errors $\varepsilon = \mathbf{X} - \hat{\mathbf{X}}$. However, in contrast with most of the literature that concentrates on the covariance matrix, Proposition 1 is much more general: it explains how linear dimension reduction helps with respect to higher-comoment tensors. Unfortunately, reducing the number of factors from $N$ to $K$ does not resolve the curse of dimensionality. Indeed, the $k$th moment of $\hat{\mathbf{P}}$ is

$$m_k(\hat{\mathbf{P}}) = \hat{\mathbf{w}}' M_k(Y_K) \bigotimes_{i=1}^{k-1} \hat{\mathbf{w}},$$

and, for $k \geq 3$, $\sharp(M_k(Y_K))$ is similar to $\sharp(M_k(X))$ in (5) with $N \leftarrow K$. In particular, it still grows exponentially with $k$, and thus can be large even for small $K$.

**Example 1.** Let $N = 30$ and $K = 5$. The PC-estimate of the minimum-variance portfolio $\hat{\mathbf{w}}_{MV,K}$ relies on the estimation of $\sharp(\hat{\Sigma}) = 140$ parameters, which is much smaller than the 465 associated with $\mathbf{w}_{MV}$. Yet, a strategy that would further feature the third and fourth moments requires the estimation of an extra $\sharp(M_3(Y_K)) + \sharp(M_4(Y_K)) = 35 + 70 = 105$ parameters that, moreover, are notoriously difficult to estimate (Martellini and Ziemann 2010).

3. Sparse higher-comoment tensors

As explained above, dimension reduction helps to decrease the number of parameters to be estimated in a given portfolio strategy but, in general, does not solve the curse of dimensionality. In particular, the number of parameters still grows with the number $K$ of factors to the power $s$, the highest moment order involved in the portfolio strategy. Interestingly however, exploiting a specific dependence structure among factors can help sparsifying the higher-comoment tensors. In this section, we show that working with a specific linear transform of the PCs—namely, the rotation of PCs that features the lowest dependence—allows one to further lower in a substantial way the dimensionality underlying the portfolio-optimization problem, avoiding exponential growth.
3.1. Independent factor model

Let us assume that the centered asset returns can be explained by the following linear factor model:

\[ X = A_K^\perp S_K, \]  

(18)

where the loading matrix \( A_K^\perp \) is of size \( N \times K \) the factors \( S \) are standard and independent (i.e. they have zero mean, covariance matrix \( M_2(S_K) = I_K \), and satisfy \( S_i \perp S_j \) for all \( i \neq j \)). Notice that this is an independent factor model: we do not merely require \( S_K \) to be uncorrelated, but to be independent. This requirement implies that only marginal moments must be estimated in \( M_k(S) \), resulting in much sparser estimates of the comoment tensors of asset returns than with the PCs, which are merely uncorrelated. In particular, we have the following proposition.

**Proposition 2.** Under the independent factor model (18), the \( k \)th comoment tensor of \( X \) is

\[ M_k(X) = A_K^\perp M_k(S_K) \bigotimes_{i=1}^{k-1} (A_K^\perp)' , \]  

(19)

and, for \( k \geq 3 \), \( \sharp(M_k(X)) = \sharp(A_K^\perp) + \sharp(M_k(S_K)) \) with

\[ \sharp(M_k(S_K)) = K(k - 3 + 1_{k=3}). \]  

(20)

This result has a fundamental consequence: conditional upon \( A_K^\perp \), the cardinality of the tensor \( M_k(X) \) grows only linearly with \( K \). This is intuitive since under our independence factor assumption, the \( k \)th order comoment tensor of \( X \) solely depends on the marginal moments of the \( K \) factors \( S_K \) up to order \( k \).

**Remark 2.** Equation (20) results from the fact that, because \( S_K \) is mutually independent, the tensor \( M_k(S_K) \) for \( k \geq 3 \) requires all marginal moments of order 3 to \( k \) (those of order 2 are all equal to one), except those of order \( k-1 \) as they appear in the terms \( \mathbb{E}[S_i S_i \ldots S_i S_j] = m_{k-1}(S_i)\mathbb{E}[S_j] = 0 \). In practical portfolio settings, the tensors are generally used in successive order, and estimating all tensors \( M_k(X) \) from order 2 to \( s \) entails \( \sharp(A_K^\perp) + \sharp(M_3(S_K), \ldots, M_s(S_K)) = K(s - 2) \) parameters as the latter merely requires all marginal moments \( m_3(S_i), \ldots, m_s(S_i), i = 1, \ldots, K \).
3.2. Approximation via independent component analysis

The dimension-reduction step introduced in Section 2 yields $M_k(X) \approx M_k(\hat{X})$ where $\hat{X} = A_K Y_K$ is an approximation of $X = A_N Y_N$ based on only $K \leq N$ factors. Of course, the PCs $Y_K$ form just one set of uncorrelated factors, out of infinitely many others. Indeed, any rotation $R_K$ of the latter yields a new set of factors, $R_K Y_K$, with identity covariance matrix and spanning the same subspace. Combining those factors according to the loading matrix $A_K R_K'$ yields $\hat{X}$:

$$(A_K R_K')(R_K Y_K) = A_K Y_K = \hat{X}.$$  

An interesting choice is to consider a specific rotation of the PCs called independent components (ICs); see Hyvärinen et al. (2001) and Lassance et al. (2019) for details. They are given by choosing the rotation $R_K$ minimizing the dependence between them:

$$Y_{K}^\perp := R_K^\perp Y_K \quad \text{with} \quad R_K^\perp \in \left\{ \arg\min_{R_K} I(R_K Y_K) \right\}. \quad (21)$$

In this expression, $I$ stands for the mutual information, a well-known criterion in information theory quantifying the factor dependence:

$$I(Y) := \mathbb{E} \left[ \ln \frac{f_Y(Y)}{\prod_{i=1}^K f_{Y_i}(Y_i)} \right], \quad (22)$$

where $f_Y$ stands for the density of $Y$. Mutual information is a non-negative operator, and vanishes if and only if the components of $Y$ are mutually independent (Cover and Thomas 2006). Note that the ICs only lead to the minimum mutual information; there is no guarantee that $I(Y_{K}^\perp) = 0$.

However, a well-known identification theorem from signal processing (Comon 1994) states that if the factor model in (18) holds with at most one $S_i$ being Gaussian, we have that $Y_{K}^\perp$ found using (21) agrees with $S_K$, up to an irrelevant change of sign and permutation. In other words, if $X$ can be written as a linear invertible transform of truly independent factors of which at most one is Gaussian, then the ICs will agree with those and therefore, will be independent as well.

Otherwise, the ICs correspond to the maximally independent linear mixture of $X$. The particular representation

$$\hat{X} = A_K^\perp Y_{K}^\perp \quad \text{with} \quad A_K^\perp = A_K (R_K^\perp)'$$  

is useful because, compared to the PCs $Y_K$, the ICs $Y_{K}^\perp$ have a special feature: they correspond
to the least dependent uncorrelated factors. Therefore, we can assume that $M_k(Y_K^\perp) \approx \hat{M}_k(Y_K^\perp)$ where $\hat{M}_k(Y_K^\perp)$ is defined as the tensor whose diagonal elements are given by those of $M_k(Y_K^\perp)$ but the off-diagonal elements are replaced by their independent counterparts.\footnote{For $k = 3$, this corresponds to setting the $(i, i, i)$ entries of $\hat{M}_3(Y_K^\perp)$ to the ones in $M_3(Y_K^\perp)$, and all other entries to 0. For $k = 4$, this corresponds to setting the $(i, i, i, i)$ entries of $\hat{M}_4(Y_K^\perp)$ to the ones in $M_4(Y_K^\perp)$, the $(i, i, j, j)$ entries to 1, and all other entries to 0.} This is not unrealistic as $M_2(Y_K^\perp)$ is already diagonal and the ICs $Y_K^\perp$ are close to mutual independence. Therefore, we obtain the approximation

$$\hat{M}_k(\hat{X}) := A_K^\perp \hat{M}_k(Y_K^\perp) \bigotimes (A_K^\perp)' \bigotimes_{i=1}^{k-1}, \quad (24)$$

which enjoys the sparse cardinality of Proposition 2 with $\sharp(A_K^\perp) = KN$ and $\sharp(\hat{M}_k(Y_K^\perp))$ in (20).

**Remark 3.** Obviously, $\hat{M}_k(\hat{X}) = M_k(\hat{X})$ whenever $I(Y_K^\perp) = 0$: when the independence assumption perfectly holds, the tensors $M_k(\hat{X})$ are naturally sparse and agree with $\hat{M}_k(\hat{X})$. In practice, the ICs may not be perfectly independent, which leads to the approximation $\hat{M}_k(\hat{X}) \approx M_k(\hat{X})$.

However, we do not assume that independence holds for an arbitrary set of uncorrelated factors such as the PCs; this would be arguably dubious. Indeed, the PCs are given by eigendecomposition of the covariance matrix of the asset returns, so that really nothing can be said \textit{a priori} regarding their higher-moment dependence. Instead, we precisely make the independence assumption when working with the very specific set of factors that \textit{best matches} this hypothesis. This trick is pretty cheap (the only additional cost is the estimation of the rotation matrix $R_K^\perp$ that features only $K(K-1)/2$ free parameters), and leads to a substantial reduction of the number of free parameters.

### 3.3. Sparse estimate of optimal portfolio

Our portfolio strategy is thus made of four steps. First, a PCA-based projection yields the factors $Y_K$. Second, we compute the matrix $R_K^\perp$. Unfortunately, there is no closed form expression for $R_K^\perp$ in (21), but it can be found using Independent Component Analysis (ICA). We use the well-known \textit{FastICA} algorithm, an efficient numerical procedure developed for this purpose. The latter is proven to perform very well on a wide range of applications; see Hyvärinen (1999) for details. Third, we approximate the ICs as being actually independent: $M_k(Y_K^\perp) \approx \hat{M}_k(Y_K^\perp)$. Fourth, we compute the IC-estimate of the portfolio $w_\psi$ according to the following definition.

**Definition 2.** The IC-estimate of the portfolio $w_\psi$ in (2) is

$$\hat{w}_{\psi,K}^\perp := V_K A_K^{-1/2} (R_K^\perp)'^\perp \hat{w}_{\psi,K}, \quad (25)$$
where
\[ \tilde{\mathbf{w}}_{\psi,K}^\perp := \arg\min_{\tilde{\mathbf{w}} \in \tilde{\mathbf{W}}^\perp} \psi(\tilde{\mathbf{m}}_q(\hat{P}), \tilde{\mathbf{m}}_r(\hat{P}), \ldots, \tilde{\mathbf{m}}_s(\hat{P})) \]  
(26)
with
\[ \tilde{\mathbf{W}}^\perp := \left\{ \mathbf{v} \in \mathbb{R}^K \mid \mathbf{V}_K \mathbf{A}_K^{-1/2} (\mathbf{R}_K^\perp)' \mathbf{v} \in \mathcal{W} \right\} \]  
(27)
and \( \hat{m}_k(\hat{P}) \) the approximation of \( m_k(\hat{P}) \) obtained by replacing \( M_k(Y_K^\perp) \) by \( \hat{M}_k(Y_K^\perp) \) in (17).

Because \( \hat{M}_2(\hat{X}) = M_2(\hat{X}) = \hat{\Sigma} = \mathbf{A}_K \mathbf{A}_K' \), the IC-estimate of the minimum-variance portfolio coincides with the PC-estimate in Corollary 1. This is not surprising since the rotation transform does not impact the covariance matrix of the reconstructed asset returns. If higher moments are considered (\( s > 2 \)), then the PC and IC-estimates \( \tilde{\mathbf{w}}_{\psi,K} \) and \( \tilde{\mathbf{w}}_{\psi,K}^\perp \) differ as soon as \( I(Y_K^\perp) > 0 \). Specifically, we have the following proposition concerning the gain in parameters.

**Proposition 3.** Let \( k \geq 3 \). Then, the difference in the number of parameters needed to estimate \( m_k(\hat{P}) \) and \( \hat{m}_k(\hat{P}) \) is
\[ \left( \binom{K + k - 1}{k} \right) - K \left( \frac{K - 1}{2} + k - 3 + 1_{k=3} \right). \]  
(28)

**Example 2.** In Figure 1, we depict the difference in (28) for the third and fourth moments (\( k \in \{3, 4\} \)) as a function of \( K \). It shows that the gain in parameters allowed by our sparse approximation is substantial and raises quickly with \( K \). For example, the differences for the coskewness and cokurtosis tensors are respectively 20 and 55 for \( K = 5 \), versus 165 and 660 for \( K = 10 \).

In Table 1, we report a summarized comparison of the cardinality of the comoment tensors via i) full estimation \( M_k(X) \), ii) PC estimation \( M_k(\hat{X}) \), and iii) IC estimation \( \hat{M}_k(\hat{X}) \). The values of \( N \) and \( K \) reflect those used in the empirical analysis of Section 4.

4. **Empirical analysis**

We now assess the out-of-sample performance of our PC- and IC-estimates of portfolio strategies. Section 4.1 details the methodology, Section 4.2 compares the independence of PCs versus ICs, and Section 4.3 discusses the results.

4.1. **Methodology**

In order to assess the effect of the dimensionality on portfolio performances, we use value-weighted monthly arithmetic returns from July 1979 to June 2019 for the 5, 17 and 30 Fama-French industry
Figure 1  Gain of parameters for the estimation of higher moments via independent versus principal components

Notes. The figure depicts the difference between the number of parameters needed to estimate the third moment (solid blue) and fourth moment (dashed red) of $\hat{P}$ via principal components ($m_3(\hat{P}), m_4(\hat{P})$) versus independent components ($\hat{m}_3(P), \hat{m}_4(P)$). The general expression for the difference is given in Proposition 3, Equation (28).

portfolios ($5\text{Ind}, 17\text{Ind}, 30\text{Ind}$). These are standard benchmark datasets, commonly used in portfolio optimization; see, for example, DeMiguel et al. (2009, 2009a) and Levy and Levy (2014).

We consider two portfolio strategies $w_\psi$, corresponding to two risk measures $\psi$. First, let us take the variance, associated with the objective function $\psi(m_2(P)) = m_2(P)$. Indeed, recall from Corollary 1 that dimension reduction is helpful even in the minimum-variance setup. Second, to evaluate the performances of strategies featuring higher moments, we consider the Value-at-Risk (VaR) which is probably the most popular risk measure besides the variance; see Lwin et al. (2017) for a review of the use of VaR in portfolio optimization. To express VaR in terms of moments, we rely on the modified VaR (MVaR), which is a four-moment Cornish-Fisher expansion of VaR (Favre and Galeano 2002) corresponding to

$$
\psi(m_2(P), m_3(P), m_4(P)) := -\sqrt{m_2(P)} \times
\left( z_\alpha + \frac{1}{6}(z_\alpha^2 - 1)\frac{m_3(P)}{m_2(P)^{3/2}} + \frac{1}{24}(z_\alpha^3 - 3z_\alpha)\left( \frac{m_4(P)}{m_2(P)^2} - 3 \right) - \frac{1}{36}(2z_\alpha^3 - 5z_\alpha)\left( \frac{m_3(P)}{m_2(P)^{3/2}} \right)^2 \right),
$$

(29)

where $z_\alpha$ is the Gaussian quantile at confidence level $\alpha$. We fix $\alpha = 1\%$. To implement the minimum-variance and minimum-MVaR portfolios, we consider three estimation strategies, all computed using sample moment averages.
Table 1  Comparison of cardinality of comoment tensors

(a) General expressions

\[
\begin{align*}
\sharp(M_2) & = \frac{N(N+1)}{2} \\
\sharp(M_k \geq 3) & = \binom{N-k+1}{k}
\end{align*}
\]

<table>
<thead>
<tr>
<th></th>
<th>$\sharp(M_2)$</th>
<th>$\sharp(M_k \geq 3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full estimation</td>
<td>$N(N+1)/2$</td>
<td>$\binom{N-k+1}{k}$</td>
</tr>
<tr>
<td>PC estimation</td>
<td>$NK-K(K-1)/2$</td>
<td>$NK-K(K-1)/2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$+\binom{K-k+1}{k}$</td>
</tr>
<tr>
<td>IC estimation</td>
<td>$NK-K(K-1)/2$</td>
<td>$K(N+k-3+1_{k=3})$</td>
</tr>
</tbody>
</table>

(b) Particular cases

\[
\begin{align*}
\sharp(M_2) & \quad \sharp(M_3) & \quad \sharp(M_4) & \quad \sharp(M_2, M_3, M_4) \\
N = 5, K = 3 & \quad 15 & \quad 35 & \quad 70 & \quad 120 \\
\text{Full estimation} & \quad 15 & \quad 35 & \quad 70 & \quad 120 \\
\text{PC estimation} & \quad 12 & \quad 22 & \quad 27 & \quad 37 \\
\text{IC estimation} & \quad 12 & \quad 18 & \quad 18 & \quad 21 \\
N = 17, K = 6 & \quad 153 & \quad 969 & \quad 4845 & \quad 5967 \\
\text{Full estimation} & \quad 87 & \quad 143 & \quad 213 & \quad 269 \\
\text{PC estimation} & \quad 87 & \quad 108 & \quad 108 & \quad 114 \\
\text{IC estimation} & \quad 87 & \quad 108 & \quad 108 & \quad 114 \\
N = 30, K = 10 & \quad 465 & \quad 4960 & \quad 40920 & \quad 46345 \\
\text{Full estimation} & \quad 255 & \quad 475 & \quad 970 & \quad 1190 \\
\text{PC estimation} & \quad 255 & \quad 310 & \quad 310 & \quad 320 \\
\text{IC estimation} & \quad 255 & \quad 310 & \quad 310 & \quad 320
\end{align*}
\]

**Notes.** The table reports the cardinality of the comoment tensors of asset returns in three separate cases. First, the full estimation $M_k(X)$. Second, the PC estimation $\hat{M}_k(\hat{X})$. Third, the IC estimation $\hat{M}_k(\hat{X})$. Table (a) reports the general expressions, and Table (b) reports particular cases of $N$ of $K$. The values of $N$ correspond to the three datasets used in the empirical analysis. The values of $K$ correspond to the rounded average $K$ used over time for the three datasets; see Section 4.1 for details on the choice of $K$. The column $\sharp(M_2, M_3, M_4)$ in Table (b) is the cardinality implied by the estimation of the three tensors together, and is equal to or lower than the sum of the cardinality of each separate tensor because of redundancies.

1) MV and MMVaR: estimation using full comoment tensors of asset returns.

2) MVPC and MMVaRPC: PC-estimates $\hat{w}_{\psi,K}$ introduced in Definition 1. MVPC is known in closed form from Corollary 1.

3) MMVaRIC: IC-estimate $\hat{w}_{\psi,K}^{\perp}$ introduced in Definition 2. We only consider MVaR for the IC-estimate because, as noted in Section 3.3, the PC and IC-estimates of the minimum-variance portfolio coincide.

We select $K \in \{2, \ldots, N\}$ as the lowest value of $K$ for which the first $K$ PCs retain 90% of the variance; that is, for which

\[
\frac{\sum_{i=1}^{K} \lambda_i}{\sum_{j=1}^{N} \lambda_j} > 90\%,
\]

\[
(30)
\]

Note. The figure depicts the number of principal components kept, $K$, over time for the three datasets. It is selected so that 90% of the variance is retained as in Equation (30). We use 60 rolling windows of ten years of monthly returns, rolled over by six months over time, covering the period July 1979 to June 2019.

where $\lambda_i$ is the $i$th eigenvalue. This approach, which consists of selecting $K$ via a threshold on the percentage of total variance explained by the selected factors, is pretty standard in many applications of data processing (Everitt and Dunn 2010). Figure 2 depicts the time evolution of $K$ for the three datasets. It shows that the datasets are well diversified as the number of factors needed to recover 90% of the variance is quite large. Following Proposition 3, this is precisely the kind of situations where our sparse ICA-based tensor approximations can bring a substantial robustness gain compared to the simple PCA-based reduction of dimension.

To evaluate the out-of-sample performance of the portfolios, we employ a classical rolling-horizon methodology as in DeMiguel et al. (2009), using an estimation window of ten years (a sample size of 120 months) and a rolling window of six months. Thus, the 30-years out-of-sample period goes from July 1989 to June 2019. The returns are computed net of transaction costs as in DeMiguel et al. (2009a, p.1930) with a proportional transaction cost of $c = 50$ basis points.

We use several performance criteria. For mean-variance tradeoff: annualized mean, volatility, and Sharpe ratio. For higher moments: monthly skewness ($m_3(P)/m_2(P)^{3/2}$), excess kurtosis ($m_4(P)/m_2(P)^2 - 3$), 1% MVaR in (29), and 1% modified Sharpe ratio (i.e. the ratio between mean and 1% MVaR). For stability: monthly portfolio turnover as in DeMiguel et al. (2009a, p.1929).

### 4.2. Independence of PCs versus ICs

The sparse representation of the asset-return comoment tensors via ICA is based on the assumption that the ICs are independent. For the coskewness tensor, this amounts to disregard the off-diagonal elements. Intuitively, the error resulting from this approximation is expected to be much smaller when considering the ICs as factors instead of the PCs. To illustrate this, let us compare the coskewness tensors of the $K = 2$ PCs ($M_3(Y_2)$) to that of the $K = 2$ ICs ($M_3(Y_2^\perp)$) on the full
The out-of-sample period for the 30Ind dataset:

\[
M_3(Y_2) = \begin{pmatrix}
-3.03 & -1.15 & -1.15 & -0.15 \\
-1.15 & -0.15 & -0.15 & -0.40 \\
\end{pmatrix}, \quad M_3(Y_2^\perp) = \begin{pmatrix}
-3.61 & 0.11 & 0.11 & 0.10 \\
0.11 & 0.10 & 0.10 & -0.46 \\
\end{pmatrix}.
\]

As expected, the off-diagonal elements of \(M_3(Y_2^\perp)\) are much closer to zero than those of \(M_3(Y_2)\); it is therefore more natural to disregard the former than the latter. This is precisely what we do by choosing to rely on the ICs. Similar observations are made for the other two datasets and the cokurtosis tensors. Second, for the three datasets, we report in Figure 3 the time evolution of the non-linear correlation between the \(K = 2\) PCs and ICs,

\[
\rho_G(Y_1,Y_2) := \frac{\text{Cov}(G(Y_1), G(Y_2))}{\sqrt{m_2(G(Y_1))m_2(G(Y_2))}},
\]

for \(G(x) = \ln \cosh x\). This choice is motivated by the fact that independence between two variables \((X,Y)\) amounts to decorrelation between any non-linear transforms of them, i.e. \(\text{Cov}(f(X), g(Y)) = 0\) for all functions \(f,g\). Obviously, linear functions only capture linear dependence. Here, we measure the residual dependence among the PCs and ICs by computing the correlation of their \(G\)-transform, because \(G\) is the non-linear mapping used in FastICA. We also depict the 95% confidence intervals estimated via bootstrap. For the PCs, \(\rho_G\) is nearly always positive, and \(\rho_G = 0\) is quite a few times outside the confidence interval. In contrast, for the ICs, \(\rho_G\) remains close to zero without up or down trends, and \(\rho_G = 0\) is nearly always in the confidence interval.

4.3. Comparison of portfolio performances

The out-of-sample results obtained with the methodology of Section 4.1 are reported in Table 2. We make the following observations.

First, dimension reduction is not useful in terms of performance for small size datasets (5Ind), but is very useful for moderate and large size datasets (17Ind and 30Ind). Indeed, in terms of Sharpe ratio and modified Sharpe ratio, MVPC and MMVaRPC are outperformed by MV and MMVaR for 5Ind, and the contrary for 17Ind and 30Ind. This is not surprising as, for \(N = 5\), there are not so many parameters to estimate, and thus dimension reduction brings no further benefit. For \(N = 17\) and 30 however, the benefits are very substantial (see Table 1). Also, as far as the

---

4 For \(K = 2\), the third comoment (coskewness) tensor is of dimension \(2 \times 2 \times 2\). For visualization purposes, we represent it in matrix form by stacking the two \(2 \times 2\) subcomoment matrices columnwise. The two diagonal elements \((1,1,1)\) and \((2,2,2)\) are now the elements \((1,1)\) and \((2,4)\) in bold.
Table 2  Out-of-sample performance of portfolio strategies

<table>
<thead>
<tr>
<th></th>
<th>MV</th>
<th>MVPC</th>
<th>MMVaR</th>
<th>MMVaRPC</th>
<th>MMVaRIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>11.10%</td>
<td>11.02%</td>
<td>11.12%</td>
<td>11.57%</td>
<td>10.89%</td>
</tr>
<tr>
<td>Volatility</td>
<td>12.26%</td>
<td>13.53%</td>
<td>13.90%</td>
<td>14.72%</td>
<td>14.53%</td>
</tr>
<tr>
<td>Sharpe ratio</td>
<td>0.91</td>
<td>0.81</td>
<td>0.80</td>
<td>0.79</td>
<td>0.75</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.44</td>
<td>-0.64</td>
<td>-0.22</td>
<td>-0.34</td>
<td>-0.34</td>
</tr>
<tr>
<td>Excess kurtosis</td>
<td>0.43</td>
<td>1.39</td>
<td>0.06</td>
<td>0.28</td>
<td>0.78</td>
</tr>
<tr>
<td>Modified VaR</td>
<td>8.55%</td>
<td>10.67%</td>
<td>9.03%</td>
<td>10.08%</td>
<td>10.48%</td>
</tr>
<tr>
<td>Modified SR (×10)</td>
<td>1.08</td>
<td>0.86</td>
<td>1.03</td>
<td>0.96</td>
<td>0.87</td>
</tr>
<tr>
<td>Turnover</td>
<td>7.84%</td>
<td>5.27%</td>
<td>18.80%</td>
<td>16.08%</td>
<td>8.99%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MV</th>
<th>MVPC</th>
<th>MMVaR</th>
<th>MMVaRPC</th>
<th>MMVaRIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>7.47%</td>
<td>10.14%</td>
<td>8.65%</td>
<td>11.07%</td>
<td>10.98%</td>
</tr>
<tr>
<td>Volatility</td>
<td>11.66%</td>
<td>11.76%</td>
<td>17.18%</td>
<td>14.92%</td>
<td>12.31%</td>
</tr>
<tr>
<td>Sharpe ratio</td>
<td>0.64</td>
<td>0.86</td>
<td>0.50</td>
<td>0.74</td>
<td>0.89</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.46</td>
<td>-0.42</td>
<td>0.16</td>
<td>0.19</td>
<td>-0.29</td>
</tr>
<tr>
<td>Excess kurtosis</td>
<td>0.48</td>
<td>0.66</td>
<td>3.09</td>
<td>1.52</td>
<td>0.52</td>
</tr>
<tr>
<td>Modified VaR</td>
<td>8.46%</td>
<td>8.40%</td>
<td>13.79%</td>
<td>9.97%</td>
<td>8.44%</td>
</tr>
<tr>
<td>Modified SR (×10)</td>
<td>0.74</td>
<td>1.01</td>
<td>0.52</td>
<td>0.93</td>
<td>1.09</td>
</tr>
<tr>
<td>Turnover</td>
<td>17.59%</td>
<td>10.51%</td>
<td>48.42%</td>
<td>30.81%</td>
<td>16.81%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MV</th>
<th>MVPC</th>
<th>MMVaR</th>
<th>MMVaRPC</th>
<th>MMVaRIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>6.46%</td>
<td>8.95%</td>
<td>-2.03%</td>
<td>7.62%</td>
<td>9.87%</td>
</tr>
<tr>
<td>Volatility</td>
<td>12.66%</td>
<td>11.71%</td>
<td>28.29%</td>
<td>14.36%</td>
<td>12.00%</td>
</tr>
<tr>
<td>Sharpe ratio</td>
<td>0.51</td>
<td>0.76</td>
<td>-0.07</td>
<td>0.53</td>
<td>0.82</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.12</td>
<td>-0.41</td>
<td>0.15</td>
<td>-0.35</td>
<td>-0.28</td>
</tr>
<tr>
<td>Excess kurtosis</td>
<td>1.14</td>
<td>1.36</td>
<td>6.72</td>
<td>1.39</td>
<td>0.64</td>
</tr>
<tr>
<td>Modified VaR</td>
<td>9.24%</td>
<td>9.00%</td>
<td>31.06%</td>
<td>11.23%</td>
<td>8.35%</td>
</tr>
<tr>
<td>Modified SR (×10)</td>
<td>0.58</td>
<td>0.83</td>
<td>-0.05</td>
<td>0.57</td>
<td>0.98</td>
</tr>
<tr>
<td>Turnover</td>
<td>30.63%</td>
<td>14.99%</td>
<td>193.21%</td>
<td>35.23%</td>
<td>22.00%</td>
</tr>
</tbody>
</table>

Notes. The table reports the out-of-sample performance of the five considered portfolio policies on the three datasets following the methodology in Section 4.1. The portfolio mean return, volatility and Sharpe ratio are annualized, while all other performance criteria are reported in monthly terms. All performance criteria are reported net of transaction costs, assuming a proportional transaction cost of \( c = 50 \) basis points. The number of principal components, \( K \), is selected so that 90\% of the variance is retained. The out-of-sample period ranges from July 1989 to June 2019, and we use an estimation window of ten years and a rolling window of six months.
Figure 3  Non-linear correlation of principal versus independent components

Notes. The figure depicts the non-linear correlation of the $K = 2$ principal and independent components over time, for the 5Ind, 17Ind and 30Ind datasets. The non-linear correlation is computed by transforming the factors with the function $G(x) = \ln \cosh x$, and computing Pearson's correlation on the transformed factors; see Equation (31). The solid blue lines depict the point estimates, and the dotted blue lines the 95% confidence intervals computed via bootstrap using Matlab \texttt{bootci} function with 1000 bootstrap resamples. We use 60 rolling windows of ten years of monthly returns, rolled over by six months over time, covering the period July 1979 to June 2019. These windows are depicted on the $x$-axis.

objective functions are concerned, MMVaRPC achieves a lower modified VaR than MMVaR for 17Ind and 30Ind, and MVPC achieves a similar volatility than MV for 17Ind, and a lower one for 30Ind. Finally, it is particularly striking to observe that the out-of-sample performance of MMVaR completely collapses for 30Ind: although it aims to minimize MVaR, its out-of-sample MVaR is 31.06%, compared with 11.23% for MMVaRPC. This shows that the curse of dimensionality is prominent for strategies accounting for higher moments.

Second, dimension reduction is always very useful in terms of turnover. For example, for the 30Ind dataset, the monthly turnover of MV is 30.63% versus 14.99% for MVPC, and is 193.21% for MMVaR versus 35.23% for MMVaRPC. This does not come as a surprise since a higher sparsity is expected to be beneficial in terms of robustness.

Third, for MVaR as higher-moment risk measure, using the IC-estimate MMVaRIC rather than the PC-estimate MMVaRPC is not useful in terms of performance for small size datasets (5Ind), but is very useful for moderate and large size datasets (17Ind and 30Ind). Indeed, in terms of Sharpe ratio and modified Sharpe ratio, MMVaRIC is outperformed by MMVaRPC for 5Ind, and the contrary for 17Ind and 30Ind. As far as the goal of minimizing MVaR is concerned, MMVaRIC outperforms MMVaRPC: the MVaR is 8.44% versus 9.97% for 17Ind, and 8.35% versus 11.23% for 30Ind. This shows that considering (nearly) independent rather than merely uncorrelated factors
enhances robustness and performance.

Fourth, in addition to improving portfolio performance, our sparse IC estimates of comoment tensors also significantly reduce turnover; for all three datasets, MMVaRIC has a lower turnover than MMVaRPC. In Figure 4, we depict boxplots of portfolio weights of MMVaR, MMVaRPC and MMVaRIC for the $17\text{Ind}$ dataset. The figure shows that MMVaRIC is by far the most stable portfolio. Moreover, we have also observed that it entails less short positions over time.

Finally, our sparse tensor estimates based on ICA yield higher-moment portfolio strategies that outperform the minimum-variance strategy (except, again, for $5\text{Ind}$). For $17\text{Ind}$, MMVaRIC features a similar skewness, kurtosis and MVar than MV and MVPC, but a much higher modified Sharpe ratio. For $30\text{Ind}$, MMVaRIC is the strategy with the lowest kurtosis, lowest MVar and highest modified Sharpe ratio. In particular, its MVar is 8.35%, versus 9.24% for MV and 9.00% for MVPC. In contrast, the classical estimates (MMVaR) or PC estimates (MMVaRPC) yield portfolios that do not outperform the minimum-variance strategy: MMVaR and MMVaRPC feature a higher MVar than MV for all datasets. Thus, robust estimation procedures are needed to reap higher-moment benefits in portfolio optimization. This result is to be put in contrast with Martellini and Ziemann (2010) who find that, for monthly returns, there is no benefit in using a four-moment portfolio strategy over the minimum-variance portfolio, no matter whether one relies on sample,
sparse or shrinkage estimators of comoment tensors. Instead, even though we use monthly returns, we find that higher-moment strategies can outperform the minimum-variance portfolio when using our sparse low-dimensional estimation of the comoment tensors based on ICA.

5. Conclusion

Although it is clear that investors care about higher-moment features such as asymmetry and fat tails, these are often dismissed in portfolio strategies. The reason has to be found in the dimensionality of the problem: in practice, the numbers of parameters to be estimated quickly explodes with the dimension of the investment universe and the order of the moments. This, in turn, deeply impacts the robustness of the corresponding strategies, and leads to disappointing out-of-sample performances. Several techniques have been proposed in the literature to deal with this problem, for example by using shrinkage or imposing a factor structure. Although successfully applied earlier to derive robust estimators of the covariance matrix, their extension to higher-comoment tensors is complex and often struggles to outperform simpler strategies ignoring higher moments for moderate sample sizes. By contrast, we derive a simple and elegant framework whose purpose is precisely to tackle the curse of dimensionality: we seek to enhance robustness by proposing a much sparser representation of the comoment tensors of asset returns. To achieve this goal, the asset returns are projected in a linear way onto a small set of maximally independent factors, found via independent component analysis. This procedure is proven to solve the curse of dimensionality impacting the classical estimates: by exploiting the near-dependence structure of these specific factors, we drastically reduce the number of free parameters in the comoment tensors. In particular, the latter is no longer given by a power law, but evolves linearly with the higher moments involved in the objective function. This makes our strategy very scalable and efficient, even when applied in high-dimensional investment universes. Using Value-at-Risk as higher-moment risk measure, we find empirically that the resulting portfolio enjoys remarkable out-of-sample performance and turnover. In particular, it outperforms the classical non-sparse portfolio estimates when the number of assets is moderate or large. Moreover, it allows for improved higher moments compared to the minimum-variance portfolio, contrary to non-sparse estimates.
Acknowledgements

Nathan Lassance’s research is funded by the Fonds de la Recherche Scientifique (F.R.S.-FNRS) under grant ASP [grant number FC 17775]. We thank Kris Boudt, Victor DeMiguel and Guofu Zhou for comments and discussions.

Appendix: Proofs of results

A useful property for the proofs below is that, given an $N$-dimensional random vector $Z$, the $k$th comoment tensor of $AZ$ is

$$M_k(AZ) = AM_k(Z) \otimes A'.$$  \hfill (32)

This is straightforward to derive from the definition of comoment tensor in (4) and the fact that the tensor product $\otimes$ is associative:

$$M_k(AZ) = E \left[ A Z \otimes (Z' A') \right] = A E \left[ Z \otimes Z' \otimes A' \right] = AM_k(Z) \otimes A'.$$

A. Proposition 1

Proof. Equation (13) is a direct application of (32). The cardinality of $A_K = V_K \Lambda_K^{1/2}$ is $\sharp(A_K) = \sharp(V_K) + \sharp(\Lambda_K)$. Clearly, $\sharp(\Lambda_K) = K$. The matrix $V_K$ is made of $NK$ entries, of which we must withdraw $K$ normalization constraints on the columns and $\binom{K}{2} = K(K-1)/2$ orthogonalization constraints (one for each pair of columns). This results in $\sharp(A_K)$ given by (14). The cardinality of $M_K(Y_K)$ is obtained by plugging $N \leftarrow K$ in (5).

B. Corollary 1

Proof. Given that $\hat{P} = w' \hat{X}$, the PC-estimate of the portfolio $w_\psi$ in Definition 1 is equivalently given by

$$\hat{w}_{\psi,K} = \text{argmin}_{w \in W} \psi\left( m_q(\hat{P}), m_r(\hat{P}), \ldots, m_s(\hat{P}) \right),$$ \hfill (33)

which, for the special case $\psi(m_2(\hat{P})) = m_2(\hat{P}) = w' M_2(\hat{X}) w$ has the well-known solution

$$\hat{w}_{\psi,K} = \frac{M_2(\hat{X})^{-1} 1}{1'M_2(\hat{X})^{-1} 1}.$$  \hfill (34)

22
Finally, we have from (13) that $M_2(\hat{X}) = A_K A'_K = \Sigma$.

\[ \square \]

C. Proposition 2

\textit{Proof.} Equation (19) is an immediate consequence of (32). To count the number of distinct parameters in each higher-comoment tensor $M_k(S_K)$ for $k \geq 3$ knowing that the factors $S_K$ are mutually independent, let us start with some specific values of $K$. We denote $S_i$ the $i$th entry of $S_K$. Note that, because of independence, the expectation of a product of several $S_i$‘s simplifies to the product of the corresponding moments. For $K = 3$, the entries $(i,j,k)$ and $(i,i,j)$ are all zero because the mean of each factor is zero. Thus, only the $K$ entries $(i,i,i)$, equal to $m_3(S_i)$, must be counted. For $K = 4$, the entries $(i,j,k,l)$ and $(i,i,j,j)$ are zero and the entries $(i,i,j)$ equal $m_2(S_i)m_2(S_j) = 1$.

Thus, only the $K$ moments $m_4(S_i)$ must be counted. For $K = 5$ however, one must count not only the $K$ moments $m_5(S_i)$, but also the entries $(i,i,j,j)$ that equal $m_3(S_i)m_2(S_j) = m_3(S_i)$. Thus, one must add the $K$ moments $m_4(S_i)$ not counted because they appear in the terms $(i,i,j,j)$ that equal zero. By generalizing this, we therefore have that $\sharp(M_k(S_K)) = K$ and, for $k \geq 4$, $M_k(S_K)$ requires all marginal moments of order 3 to $k$, except those of order $k-1$. Thus, $\sharp(M_k(S_K)) = K(K-3)$ for $k \geq 4$, proving Equation (20).

\[ \square \]

D. Proposition 3

\textit{Proof.} We have that $m_k(\hat{P}) = w'M_k(\hat{X}) \otimes_{i=1}^{k-1} w$ and $\hat{m}_k(\hat{P}) = w'\hat{M}_k(\hat{X}) \otimes_{i=1}^{k-1} w$. Thus, the difference in the number of parameters is $\sharp(M_k(\hat{X})) - \sharp(\hat{M}_k(\hat{X}))$. From Proposition 1, the cardinality $\sharp(M_k(\hat{X})) = \sharp(A_K) + \sharp(M_k(Y_K)) = \sharp(V_K) + \sharp(A_K) + \sharp(M_k(Y_K))$. Similarly, from Proposition 2, the cardinality $\sharp(\hat{M}_k(\hat{X})) = \sharp(A_K) + \sharp(\hat{M}_k(Y_K)) = \sharp(V_K) + \sharp(A_K) + \sharp(R_K) + \sharp(\hat{M}_k(Y_K))$. Thus, the difference becomes $\sharp(M_k(Y_K)) - \big(\sharp(R_K) + \sharp(\hat{M}_k(Y_K))\big)$, and simplifies to (28) from $\sharp(M_k(Y_K))$ in (15), $\sharp(\hat{M}_k(Y_K))$ in (20), and $\sharp(R_K) = K(K-1)/2$ for any $K \times K$ rotation matrix.

\[ \square \]

References


Quantitative Finance, 10(5):469–485.