PORTFOLIO SELECTION WITH PARSIMONIOUS HIGHER COMOMENTS ESTIMATION

Nathan Lassance, Frédéric Vrins

REPRINT • 2021 / 05

UCLouvain

LFIN | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

LFIN

Voie du Roman Pays 34, L1.03.01 B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/lfin/publications.html

Journal of Banking and Finance (2021), 126(9), 106–115:
https://doi.org/10.1016/j.jbankfin.2021.106115

Portfolio selection with parsimonious higher comoments estimation*

Nathan Lassance[a,*], Frédéric Vrins[b]

[a] *UCLouvain, Louvain Institute of Data Analysis and Modeling, Louvain Finance, Chaussée de Binche 151, 7000 Mons, Belgium*
[b] *UCLouvain, Louvain Institute of Data Analysis and Modeling, Louvain Finance, Voie du Roman Pays 34, 1348 Louvain-la-Neuve, Belgium*

Abstract

Large investment universes are usually fatal to portfolio strategies optimizing higher moments because of computational and estimation issues resulting from the number of parameters involved. In this paper, we introduce a parsimonious method to estimate higher moments that consists of projecting asset returns onto a small set of maximally independent factors found via independent component analysis (ICA). In contrast to principal component analysis (PCA), we show that ICA resolves the curse of dimensionality affecting the comoment tensors of asset returns. The method is easy to implement, computationally efficient, and makes portfolio strategies optimizing higher moments appealing in large investment universes. Considering the value-at-risk as a risk measure, an investment universe of up to 500 stocks and adjusting for transaction costs, we show that our ICA approach leads to superior out-of-sample risk-adjusted performance compared with PCA, equally weighted, and minimum-variance portfolios.

Keywords: portfolio selection, estimation risk, independent component analysis, principal component analysis, higher moments

1. Introduction

It is widely acknowledged that investors care about the higher moments of their portfolio returns (Scott and Horvath, 1980; Harvey and Siddique, 2000; Ang et al., 2006). As a consequence, accounting for the non-Gaussianity of asset returns has been an important research axis since the seminal work of Markowitz (1952). When confronting skewness and kurtosis, Taylor expansion of the expected utility (Jondeau and Rockinger, 2006; Guidolin and Timmermann, 2008; Zakamouline and Koekebakker, 2009; Martellini and Ziemann, 2010), Gram-Charlier expansion of downside risk measures (Favre and Galeano, 2002; Leòn and Moreno, 2017; Zoia et al., 2018; Lassance and Vrins, 2019), and the shortage function of Briec et al. (2007, 2013) are used in the literature.

In practice, however, optimal portfolio strategies might perform poorly relative to suboptimal approaches, such as equally weighted (EW) or minimum-variance (MV) portfolios (DeMiguel et al., 2009b). This phenomenon can be explained by estimation errors and is further exacerbated when considering *higher-moment strategies* (i.e., portfolio strategies optimizing higher moments), for two reasons. First, accurately estimating higher moments is a difficult task. Second, if no specific model is assumed, the number of asset-return comoments to estimate grows following the power law N^k, where N is the number of assets and k is the moment order. As noted by Brandt et al. (2009,

*Part of this work was conducted while N. Lassance was supported by the research grant ASP FC.17775 of the Fonds de la Recherche Scientifique (F.R.S.-FNRS). The work of F. Vrins is supported by the research grant ARC 18-23/089 of the Belgian Federal Science Policy Office. We thank Kris Boudt, Victor DeMiguel, and Guofu Zhou for useful comments and stimulating discussions about the proposed approach. We also thank Floris Laly for his proofreading. Finally, we are grateful to the anonymous referees and G. Bekaert (managing editor) for suggesting relevant improvements.

*Corresponding author
Email addresses: nathan.lassance@uclouvain.be (Nathan Lassance), frederic.vrins@uclouvain.be (Frédéric Vrins)

p.3418), this is a serious issue: "[...] extending the traditional approach beyond first and second moments [...] 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." Hence, higher-moment strategies face a fatal *curse of dimensionality*, which significantly narrows their application scope. For instance, most corresponding empirical studies consider small investment universes relative to modern financial markets (typically, $N \leq 30$ when $k \geq 3$).

Several extensions of solutions initially derived in a mean-variance context exist to address the first problem, such as Bayesian estimation of coskewness (Harvey et al., 2010) and shrinkage estimators of coskewness and cokurtosis tensors (Martellini and Ziemann, 2010; Boudt et al., 2020a). In this paper, we focus on resolving the second problem via dimension reduction. This is the main purpose of *factor models*, which describe the dependence structure of the N asset returns via $K \leq N$ factors. Such models have been widely used for covariance matrix estimation; see De Nard et al. (2019) for a review. Recently, factor models have also been used to estimate higher-comoment tensors; see, for example, Boudt et al. (2015) and Jondeau et al. (2018). However, the effect of dimension reduction on the curse of dimensionality is limited. In particular, the growth in the number of comoments to estimate remains of order K^k in both approaches.

In this paper, we address the curse of dimensionality impacting higher-moment strategies by considering a factor model associated with a parsimonious dependence structure. Specifically, we first perform dimension reduction by projecting the asset returns onto the first K principal components (PCs), where K is chosen following the method of Bai and Ng (2002), which is consistent for high-dimensional data (i.e., for a large number N of assets). Next, we consider as factors the *independent components* (ICs) defined as the rotation of the first K PCs that minimize the mutual dependence. This transformation is obtained via independent component analysis (ICA), a well-known technique in signal processing (Hyvärinen et al., 2001). The combination of a factor model with the near independence of the ICs substantially reduces the number of comoments that must be estimated. Indeed, the higher-comoment tensors of asset returns are fully explained by the PCA projection matrix, the ICA rotation matrix, the marginal moments of the ICs and those of the errors.

Independent factor models have previously been used in the field of portfolio selection. Chen et al. (2007) rely on ICA to forecast portfolio value-at-risk (VaR) and find that it is more accurate than standard methods. Hitaj et al. (2015) rely on ICA to estimate the maximum-CARA-utility portfolio. However, these two papers involve parametric estimation of the densities of ICs and do not address comoment estimation. ICA and parametric estimation of factor densities are also used in Ghalanos et al. (2015) to estimate the portfolio-return characteristic function in a dynamic GARCH framework. In contrast to the previous two papers, they also consider the Taylor expansion of the CARA utility and rely on the formula that links the asset-return comoment tensors to those of the ICs. However, no connection is made with the curse of dimensionality affecting higher-moment strategies. In particular, no dimension reduction is performed ($K = N$), and they focus on a small dataset of $N = 14$ MSCI indices. Similar to our study, Boudt et al. (2020b) consider an independent factor model to improve the estimation of higher-comoment tensors and apply their model to various higher-moment strategies. They find improvements relative to the EW portfolio and portfolios estimated via empirical moments. However, their method still suffers from the curse of dimensionality: given the computational burden, the authors acknowledge that $N = 10$ assets is a reasonable limit. Finally, Lassance et al. (2020) exploit ICA to improve risk-parity strategies. They show that considering the ICs as risk factors provides better diversification and tail risk performance than considering the PCs. They note in Section 5.1 that the near independence between the ICs simplifies the estimation of the risk measure, but their risk-parity objective is fundamentally different from the approach considered in the present paper.

Our contributions are as follows. First, we propose a factor model that resolves the curse of dimensionality that affects higher-moment strategies. We demonstrate that working with K independent factors yields a parsimonious representation of the asset-return dependence structure: the number of parameters to estimate in a comoment tensor of order k is reduced and becomes linear in K. This parsimony substantially broadens the scope of investment universes that higher-moment strategies can accommodate. We also derive formulas for higher portfolio-return moments that avoid the need to plug in very large tensors.

Second, Monte Carlo simulations featuring non-Gaussian returns demonstrate the benefits of our approach for the portfolio minimizing the modified VaR (MVaR), which corresponds to the fourth-order Gram-Charlier expansion of the VaR (Favre and Galeano, 2002; Eling and Schuhmacher, 2007; Boudt et al., 2008). We consider up to $N = 500$ assets and find that out-of-sample risk is significantly reduced when exploiting the near independence of ICs. In addition, the portfolio weights are less extreme relative to using PCs.

Third, we analyze the performance of our approach for real financial data. Unlike previous related studies, we consider high-dimensional datasets of up to $N = 500$ stocks from CRSP and adjust the returns for transaction costs. In such a setting, the out-of-sample performance of higher-moment strategies is particularly prone to estimation errors. Nonetheless, our factor model based on the ICs systematically yields the best out-of-sample risk-adjusted performance and dominates the EW portfolio, the shrinkage estimator of the MV portfolio of Ledoit and Wolf (2004), and the PCA factor model. Additional experiments confirm the superior performance of our approach when considering the lower partial moment (LPM) instead of the VaR as a risk measure. Hence, our parsimonious estimation of comoment tensors based on ICA makes higher-moment strategies competitive in large investment universes. Our results also mitigate conclusions in Simaan (2014) and Khashanah et al. (2020), who find, for several utility functions, that the mean-variance estimate of the optimal portfolio outperforms the fully optimized portfolio that takes all moments into account. We find instead that the fully optimized portfolio can outperform the minimum-variance portfolio, provided it is estimated via our parsimonious approach.

This paper is structured as follows. Section 2 covers dimension reduction via PCA. Our parsimonious comoment estimation that relies on ICA is introduced in Section 3. Section 4 addresses the Monte Carlo study. The empirical analyses are performed in Section 5. Section 6 concludes the paper. Proofs and additional results are relegated to the Online Appendix.

2. Dimension reduction via PCA

Let $\boldsymbol{X} \in \mathbb{R}^N$ be the column vector of asset returns with mean $\boldsymbol{\mu}$ assumed to be zero and covariance matrix $\boldsymbol{\Sigma}$. We denote by $\boldsymbol{w}$ the vector of portfolio weights that belongs to a set of constraints $\mathcal{W}$, including the normalization $\mathbf{1}'\boldsymbol{w} = 1$, and $P(\boldsymbol{w}) := \boldsymbol{w}'\boldsymbol{X}$ is the associated portfolio return.

2.1. The notion of comoment tensors

The key objects we consider are *comoment tensors*, a generalization of the covariance matrix to higher comoments. The comoment tensor of order two is the covariance matrix $\boldsymbol{\Sigma}$, which is a two-dimensional array composed of N^2 entries $\boldsymbol{\Sigma}_{ij} = \mathbb{E}(X_i X_j)$ (assuming zero means). Similarly, the coskewness and cokurtosis tensors, denoted $\boldsymbol{M}_3(\boldsymbol{X})$ and $\boldsymbol{M}_4(\boldsymbol{X})$, are three- and four-dimensional arrays composed of N^3 and N^4 entries $(\boldsymbol{M}_3(\boldsymbol{X}))_{ijk} = \mathbb{E}(X_i X_j X_k)$ and $(\boldsymbol{M}_4(\boldsymbol{X}))_{ijkl} = \mathbb{E}(X_i X_j X_k X_l)$.

Comoment tensors are useful to compute the higher moments of portfolio returns. If the kth comoment tensor $\boldsymbol{M}_k(\boldsymbol{X})$ is rearranged by stacking columnwise the different dimensions in a $N \times N^{k-1}$ matrix (see Section 1 in Martellini and Ziemann (2010) for a good explanation), then the kth portfolio-return central moment $m_k(P) := \mathbb{E}((P - \mathbb{E}(P))^k)$ depends on $\boldsymbol{w}$ and $\boldsymbol{M}_k(\boldsymbol{X})$ as

$$m_k(P) = \boldsymbol{w}'\boldsymbol{M}_k(\boldsymbol{X}) \bigotimes_{i=1}^{k-1} \boldsymbol{w}, \tag{1}$$

where $\bigotimes_{i=1}^{k} \boldsymbol{w} := \boldsymbol{w} \otimes \cdots \otimes \boldsymbol{w}$ is the Kronecker product of k $\boldsymbol{w}$.

2.2. Curse of dimensionality

In this paper, we focus on portfolio strategies minimizing a function of moments,

$$\boldsymbol{w}_\psi := \operatorname*{argmin}_{\boldsymbol{w} \in \mathcal{W}} \psi\Big(m_j(P(\boldsymbol{w})), \ldots, m_k(P(\boldsymbol{w})), \ldots, m_l(P(\boldsymbol{w}))\Big), \tag{2}$$

where $j \geq 2$ is the lowest moment order considered, and $l \geq k \geq j$ the highest.

Remark 1. Asset managers typically have some views about mean returns $\boldsymbol{\mu}$ given the hundreds of cross-sectional anomalies identified in the literature (Green et al., 2013). Our framework can accommodate such views by extending ψ in (2) as $\psi\big(\mu(P), m_j(P), \ldots, m_k(P), \ldots, m_l(P)\big)$ with $\mu(P) := \boldsymbol{w}'\boldsymbol{\mu}$ estimated via the asset manager's views on $\boldsymbol{\mu}$. Because out-of-sample performance often deteriorates when using mean returns (Merton, 1980; DeMiguel et al., 2009a,b), we exclude $\mu(P)$ from (2) and set $\boldsymbol{\mu}$ equal to zero.

As clear from (1), the kth moment $m_k(P)$ depends on the kth comoment tensor $\boldsymbol{M}_k(\boldsymbol{X})$. Without further assumptions on $\boldsymbol{X}$, the cardinality (i.e., the number of distinct parameters) of $\boldsymbol{M}_k(\boldsymbol{X})$ is

$$\sharp\left(\boldsymbol{M}_k(\boldsymbol{X})\right) = \binom{N+k-1}{k} = O\left(N^k\right). \tag{3}$$

Thus, the number of parameters impacting the kth moment of P quickly becomes unmanageable when considering a large number N of assets and/or higher moments $k > 2$. This severely impacts the sensitivity to estimation error and out-of-sample performance of the portfolio $\boldsymbol{w}_\psi$. We refer to this problem as the *curse of dimensionality*.

2.3. Reducing dimensionality via principal components

As commonly done in the literature, we address the curse of dimensionality via dimension reduction. We approximate the asset returns $\boldsymbol{X}$ via $\hat{\boldsymbol{X}}$ constructed from a K-factor model,

$$\hat{\boldsymbol{X}} = \boldsymbol{A}_K \boldsymbol{Y}_K + \boldsymbol{\epsilon}, \tag{4}$$

where $\boldsymbol{A}_K$ is the $N \times K$ loading matrix, $\boldsymbol{Y}_K$ is the vector of K uncorrelated and unit-variance factors, and $\boldsymbol{\epsilon}$ is the vector of errors. As in the literature, we assume the errors $\boldsymbol{\epsilon}$ are mutually independent and independent from the factors $\boldsymbol{Y}_K$ (Boudt et al., 2015, 2020; DeNard et al., 2019).

The standard practice in factor models is to rely on PCA; see Section 6.4 of Campbell et al. (1997). Denoting by $\boldsymbol{V}_K \in \mathbb{R}^{N\times K}$ and $\boldsymbol{\Lambda}_K \in \mathbb{R}^{K\times K}$ the eigenvectors and eigenvalues matrices associated with the K largest eigenvalues of $\boldsymbol{\Sigma}$, the principal components (PCs) and the loading matrix are

$$\boldsymbol{Y}_K := \boldsymbol{\Lambda}_K^{-1/2} \boldsymbol{V}_K' \boldsymbol{X} \quad \text{and} \quad \boldsymbol{A}_K := \boldsymbol{V}_K \boldsymbol{\Lambda}_K^{1/2}. \tag{5}$$

Note that we use a statistical factor model, where $\boldsymbol{A}_K\boldsymbol{Y}_K$ in (4) is observed and computed directly from the asset returns $\boldsymbol{X}$. We assume that $\text{rank}(\boldsymbol{\Sigma}) \geq K$ to ensure that $\boldsymbol{\Lambda}_K$ is invertible. The approximation $\hat{P}$ of the portfolio return P is then obtained as

$$\hat{P} := \boldsymbol{w}'\hat{\boldsymbol{X}} = \boldsymbol{w}'\left(\boldsymbol{A}_K\boldsymbol{Y}_K + \boldsymbol{\epsilon}\right). \tag{6}$$

Unless $K = N$, $\hat{P} = \boldsymbol{w}'\hat{\boldsymbol{X}}$ is different from $P = \boldsymbol{w}'\boldsymbol{X}$ because, in general, the errors $\boldsymbol{\epsilon}$ are not mutually independent and not independent from the factors. Thus, using a K-factor model introduces a misspecification error. On the other hand, we show in Section 2.4 that using $K < N$ substantially reduces the number of parameters needed to estimate the higher moments of portfolio returns because they become fully determined by the marginal moments of the errors $\boldsymbol{\epsilon}$ and the comoments of the K factors $\boldsymbol{Y}_K$. Thus, the tradeoff between the specification and estimation error is controlled by K. In the empirical analysis in Section 5, we calibrate K using the method from Bai and Ng (2002), which is consistent as the number of assets and the sample size tend together to infinity, and find that it leads to good out-of-sample performance for large investment universes.

We can now estimate the portfolio $\boldsymbol{w}_\psi$ by optimizing the moments of $\hat{P}$ instead of P.

Definition 1. The PC estimate $\hat{\boldsymbol{w}}_{\psi,K}$ of the portfolio $\boldsymbol{w}_\psi$ is obtained by replacing P with $\hat{P}$ in (2):

$$\hat{\boldsymbol{w}}_{\psi,K} := \underset{\boldsymbol{w}\in\mathcal{W}}{\text{argmin}}\, \psi\Big(m_j(\hat{P}(\boldsymbol{w})), \ldots, m_k(\hat{P}(\boldsymbol{w})), \ldots, m_l(\hat{P}(\boldsymbol{w}))\Big). \tag{7}$$

We also derive formulas for the moments of $\hat{P}$ of order up to four, which are most often used in practice.

Proposition 1. *The moments of orders 2, 3, and 4 of $\hat{P}$ admit the following form:*

$$m_2(\hat{P}) = \boldsymbol{w}'\boldsymbol{V}_K\boldsymbol{\Lambda}_K\boldsymbol{V}_K'\boldsymbol{w} + \sum_{i=1}^{N} w_i^2 m_2(\epsilon_i), \tag{8}$$

$$m_3(\hat{P}) = m_3(\boldsymbol{w}'\boldsymbol{A}_K\boldsymbol{Y}_K) + \sum_{i=1}^{N} w_i^3 m_3(\epsilon_i), \tag{9}$$

$$m_4(\hat{P}) = m_4(\boldsymbol{w}'\boldsymbol{A}_K\boldsymbol{Y}_K) + \sum_{i=1}^{N} w_i^4 m_4(\epsilon_i) + 3\sum_{i=1}^{N}\sum_{j\neq i}^{N} w_i^2 w_j^2 m_2(\epsilon_i)m_2(\epsilon_j) + 6\boldsymbol{w}'\boldsymbol{V}_K\boldsymbol{\Lambda}_K\boldsymbol{V}_K'\boldsymbol{w}\sum_{i=1}^{N} w_i^2 m_2(\epsilon_i). \tag{10}$$

Note that only the marginal moments of errors $\boldsymbol{\epsilon}$ matter, and we estimate them from $\boldsymbol{X} - \boldsymbol{A}_K\boldsymbol{Y}_K$.

2.4. Cardinality of comoment tensors

As shown in the next proposition, the PCA factor model reduces the number of free parameters in the comoment tensors of asset returns.

Proposition 2. *The cardinality of the kth comoment tensor of the projected asset returns* $\hat{\boldsymbol{X}}$ *is*

$$\sharp(\boldsymbol{M}_k(\hat{\boldsymbol{X}})) = \sharp(\boldsymbol{A}_K) + \sharp(\boldsymbol{M}_k(\boldsymbol{Y}_K)) + \sharp(\boldsymbol{M}_k(\boldsymbol{\epsilon})), \tag{11}$$

where

$$\sharp(\boldsymbol{A}_K) = KN - K(K-1)/2, \tag{12}$$

$$\sharp(\boldsymbol{M}_2(\boldsymbol{Y}_K)) = 0, \quad \sharp(\boldsymbol{M}_k(\boldsymbol{Y}_K)) = \binom{K+k-1}{k} = O(K^k) \text{ for } k \geq 3, \tag{13}$$

$$\sharp(\boldsymbol{M}_k(\boldsymbol{\epsilon})) = N(k - 2 + \mathbb{1}_{k=2}), \tag{14}$$

where $\mathbb{1}_A$ *is the indicator function.*

Investment strategies focusing on the variance of portfolio returns rely on only the covariance matrix $\widehat{\boldsymbol{\Sigma}}$ of $\hat{\boldsymbol{X}}$,

$$\widehat{\boldsymbol{\Sigma}} := \boldsymbol{V}_K \boldsymbol{\Lambda}_K \boldsymbol{V}_K' + \boldsymbol{M}_2(\boldsymbol{\epsilon}), \tag{15}$$

where $\boldsymbol{M}_2(\boldsymbol{\epsilon})$ is diagonal. Such robust strategies ignore estimates of the mean return and higher moments and, consequently, often exhibit appealing out-of-sample performance (DeMiguel et al., 2009a; Behr et al., 2013). Covariance-matrix estimates as in (15) are commonly used in factor modeling (De Nard et al., 2019). The cardinality of $\widehat{\boldsymbol{\Sigma}}$ is equal to $\sharp\left(\widehat{\boldsymbol{\Sigma}}\right) = N(K+1) - K(K-1)/2$; thus, dimension reduction helps reduce the number of parameters because $\sharp(\widehat{\boldsymbol{\Sigma}})$ is smaller than $\sharp(\boldsymbol{\Sigma}) = N(N+1)/2$ for $K < N$.

Proposition 2 explains why the independence of the errors $\boldsymbol{\epsilon}$ helps reduce the number of parameters to estimate in higher-comoment tensors. Unfortunately, reducing the dimension from N to K does not resolve the curse of dimensionality. Indeed, because of the higher-moment dependence between PCs, $\sharp(\boldsymbol{M}_k(\boldsymbol{Y}_K))$ in (13) is similar to $\sharp(\boldsymbol{M}_k(\boldsymbol{X}))$ in (3), with N replaced by K. Specifically, $\sharp(\boldsymbol{M}_k(\boldsymbol{Y}_K))$ grows according to the power law K^k and, thus, can quickly become large. For example, when $K = 10$, PCs have 220 and 715 coskewness and cokurtosis coefficients, respectively, which is problematic because higher comoments are notoriously difficult to estimate (Martellini and Ziemann, 2010).

3. Parsimonious higher-comoment tensors via ICA

In this section, we introduce a factor model whose special dependence structure renders the higher-comoment tensors naturally *parsimonious*. Specifically, we show that considering as factors the maximally independent linear transformation of the PCs avoids the power-law growth discussed in Section 2.4, thus resolving the curse of dimensionality.

3.1. Independent factor model

Let us assume that the projected asset returns $\hat{\boldsymbol{X}}$ can be expressed via the factor model

$$\hat{\boldsymbol{X}} = \boldsymbol{A}_K^{\perp} \boldsymbol{S}_K + \boldsymbol{\epsilon}, \tag{16}$$

where the loading matrix $\boldsymbol{A}_K^{\perp}$ is of size $N \times K$, and the standardized factors $\boldsymbol{S}_K$ are actually *independent*. A key feature of this model is that the higher-comoment tensors of $\hat{\boldsymbol{X}}$ are completely determined by the loading matrix and *marginal* moments only. As a consequence, the independence of the factors makes the comoment tensors of asset returns more parsimonious than those with the merely uncorrelated PCs. In particular, we have the following proposition.

Proposition 3. *Under the independent factor model* (16)*, the cardinality of the kth comoment tensor of* $\hat{\boldsymbol{X}}$ *is*

$$\boldsymbol{M}_k(\hat{\boldsymbol{X}}) = \sharp(\boldsymbol{A}_K^{\perp}) + \sharp(\boldsymbol{M}_k(\boldsymbol{S}_K)) + \sharp(\boldsymbol{M}_k(\boldsymbol{\epsilon})), \tag{17}$$

and, for $k \geq 3$*,*

$$\sharp(\boldsymbol{M}_k(\boldsymbol{S}_K)) = K(k - 3 + \mathbb{1}_{k=3}). \tag{18}$$

This result has a fundamental consequence: conditional on $\boldsymbol{A}_K^{\perp}$ and $\boldsymbol{\epsilon}$, the cardinality of the tensor $\boldsymbol{M}_k(\hat{\boldsymbol{X}})$ grows only *linearly* with K because only the marginal moments of the K factors $\boldsymbol{S}_K$ matter. This scenario contrasts with the PCs for which $\sharp(\boldsymbol{M}_k(\boldsymbol{Y}_K)) = O(K^k)$ in (13) grows according to a power law.

3.2. Approximation via independent component analysis

The PCA dimension reduction in Section 2 leads to the approximation $\boldsymbol{M}_k(\boldsymbol{X}) \approx \boldsymbol{M}_k(\hat{\boldsymbol{X}})$, where $\hat{\boldsymbol{X}} = \boldsymbol{A}_K\boldsymbol{Y}_K + \boldsymbol{\epsilon}$ is an approximation of $\boldsymbol{X} = \boldsymbol{A}_N\boldsymbol{Y}_N$ based on $K \leq N$ factors together with K independent errors. Note that the PCs $\boldsymbol{Y}_K$ form just one set of uncorrelated factors out of infinitely many others. Indeed, any rotation $\boldsymbol{R}_K$ of the PCs yields a new set of factors, $\boldsymbol{R}_K\boldsymbol{Y}_K$, with an identity covariance matrix and spanning the same subspace. Combining those factors according to the loading matrix $\boldsymbol{A}_K\boldsymbol{R}_K'$ yields $\hat{\boldsymbol{X}}$:

$$(\boldsymbol{A}_K\boldsymbol{R}_K')(\boldsymbol{R}_K\boldsymbol{Y}_K) + \boldsymbol{\epsilon} = \boldsymbol{A}_K\boldsymbol{Y}_K + \boldsymbol{\epsilon} = \hat{\boldsymbol{X}}.$$

We consider the specific rotation of the PCs called *independent components* (ICs); see Hyvärinen et al. (2001) for details.

Definition 2. The ICs are the rotation $\boldsymbol{R}_K$ of the PCs $\boldsymbol{Y}_K$ that have the least dependence:

$$\boldsymbol{Y}_K^{\perp} := \boldsymbol{R}_K^{\perp}\boldsymbol{Y}_K \quad \text{where} \quad \boldsymbol{R}_K^{\perp} \in \left\{ \underset{\boldsymbol{R}_K}{\operatorname{argmin}}\ I(\boldsymbol{R}_K\boldsymbol{Y}_K) \right\}, \tag{19}$$

where I stands for the mutual information operator defined as

$$I(\boldsymbol{Y}) := \mathbb{E}\left[\ln \frac{f_{\boldsymbol{Y}}(\boldsymbol{Y})}{\prod_{i=1}^{K} f_{Y_i}(Y_i)} \right], \tag{20}$$

with f_Y being the density of Y.

Mutual information is a well-known criterion in information theory for quantifying dependence. It is nonnegative and vanishes if and only if the components of $\boldsymbol{Y}$ are mutually independent (Cover and Thomas, 2006). The ICs are maximally independent and are not guaranteed to be exactly independent in general. The particular representation

$$\hat{\boldsymbol{X}} = \boldsymbol{A}_K^{\perp}\boldsymbol{Y}_K^{\perp} + \boldsymbol{\epsilon} \quad \text{with} \quad \boldsymbol{A}_K^{\perp} := \boldsymbol{A}_K(\boldsymbol{R}_K^{\perp})' \tag{21}$$

is useful because, relative to the PCs $\boldsymbol{Y}_K$, the ICs $\boldsymbol{Y}_K^{\perp}$ are uncorrelated factors that feature the least dependence. Therefore, we can assume that $\boldsymbol{M}_k(\boldsymbol{Y}_K^{\perp}) \approx \widehat{\boldsymbol{M}}_k(\boldsymbol{Y}_K^{\perp})$, where $\widehat{\boldsymbol{M}}_k(\boldsymbol{Y}_K^{\perp})$ is defined as the tensor for which the diagonal elements correspond to those of $\boldsymbol{M}_k(\boldsymbol{Y}_K^{\perp})$, but the off-diagonal elements are replaced by their independent counterparts defined as the product of the corresponding marginal moments.[1] This, in turn, leads to the approximation $\boldsymbol{M}_k(\hat{\boldsymbol{X}}) \approx \widehat{\boldsymbol{M}}_k(\hat{\boldsymbol{X}})$, where the cardinality of the latter grows linearly with K, as shown in Proposition 3. Specifically,

$$\sharp\left(\widehat{\boldsymbol{M}}_k(\hat{\boldsymbol{X}})\right) = \sharp(\boldsymbol{A}_K^{\perp}) + \sharp\left(\widehat{\boldsymbol{M}}_k(\boldsymbol{Y}_K^{\perp})\right) + \sharp(\boldsymbol{M}_k(\boldsymbol{\epsilon})), \tag{22}$$

where $\sharp(\boldsymbol{A}_K^{\perp}) = KN$, $\sharp(\boldsymbol{M}_k(\boldsymbol{\epsilon}))$ in (14), and $\sharp\left(\widehat{\boldsymbol{M}}_k(\boldsymbol{Y}_K^{\perp})\right)$ in (18). This approach is fairly inexpensive: the only additional cost compared with that of PCA is the estimation of the rotation matrix $\boldsymbol{R}_K^{\perp}$ that features only $K(K-1)/2$ free parameters and leads to a substantial reduction in the number of factor comoments to estimate.

Remark 2. If the K-factor model does not hold exactly ($\hat{\boldsymbol{X}} \neq \boldsymbol{X}$) or if the ICs are not exactly independent ($\widehat{\boldsymbol{M}}_k(\hat{\boldsymbol{X}}) \neq \boldsymbol{M}_k(\hat{\boldsymbol{X}})$), then the ICA-based tensor estimate $\widehat{\boldsymbol{M}}_k(\hat{\boldsymbol{X}})$ is not consistent. One way to address this issue is to shrink $\widehat{\boldsymbol{M}}_k(\hat{\boldsymbol{X}})$ toward the unbiased sample estimate of $\boldsymbol{M}_k(\boldsymbol{X})$, as in Martellini and Ziemann (2010) and Boudt et al. (2020a), but doing so might not provide a substantial gain in performance for two reasons. First, the shrinkage intensity must be estimated, which introduces additional estimation risk. Second, we expect the optimal shrinkage intensity to be quite high; thus, setting it to 100%, as we do implicitly, is a fair approximation. The reason is that Martellini and Ziemann (2010) consider shrinkage targets based on a constant correlation or a single-factor model and a number of assets N between 10 and 30 but still find that the shrinkage intensities for the coskewness and cokurtosis tensors are very high. In our empirical setting, we use a much larger $N \in \{100, 300, 500\}$, and our tensor estimate $\widehat{\boldsymbol{M}}_k(\hat{\boldsymbol{X}})$ requires weaker assumptions.

[1] For $k = 3$, this corresponds to setting the (i,i,i) entries of $\widehat{\boldsymbol{M}}_3(\boldsymbol{Y}_K^{\perp})$ to those in $\boldsymbol{M}_3(\boldsymbol{Y}_K^{\perp})$ and all other entries to 0. For $k = 4$, this corresponds to setting the (i,i,i,i) entries of $\widehat{\boldsymbol{M}}_4(\boldsymbol{Y}_K^{\perp})$ to those in $\boldsymbol{M}_4(\boldsymbol{Y}_K^{\perp})$, the (i,i,j,j), (i,j,i,j), (i,j,j,i) entries to $m_2(Y_i^{\perp})m_2(Y_j^{\perp}) = 1$, and all other entries to 0.

3.3. Parsimonious estimate of the optimal portfolio

To estimate the portfolio $\boldsymbol{w}_\psi$ based on the parsimonious ICA factor model, we need to compute the rotation matrix $\boldsymbol{R}_K^\perp$ in (19). No closed-form expression exists for this matrix, but various algorithms can be used to compute it numerically. In this paper, we use the popular and computationally efficient *FastICA* algorithm of Hyvärinen (1999).

Definition 3. The IC estimate $\hat{\boldsymbol{w}}_{\psi,K}^\perp$ of the portfolio $\boldsymbol{w}_\psi$ is given by replacing $m_k(\cdot)$ with $\hat{m}_k(\cdot)$ in (7):

$$\hat{\boldsymbol{w}}_{\psi,K}^\perp := \underset{\boldsymbol{w}\in\mathcal{W}}{\operatorname{argmin}}\, \psi\Big(\hat{m}_j(\hat{P}(\boldsymbol{w})), \ldots, \hat{m}_k(\hat{P}(\boldsymbol{w})), \ldots, \hat{m}_l(\hat{P}(\boldsymbol{w}))\Big), \tag{23}$$

where $\hat{m}_k(\hat{P})$ is the approximation of $m_k(\hat{P})$ obtained when replacing $\boldsymbol{M}_k(\hat{\boldsymbol{X}})$ by $\widehat{\boldsymbol{M}}_k(\hat{\boldsymbol{X}})$.

Remark 3. A simple variant of the IC estimate above consists of completely discarding the errors and considering $\hat{\boldsymbol{X}} = \boldsymbol{A}_K^\perp \boldsymbol{Y}_K^\perp$ instead of (21). In fact, many financial applications of factor models include errors and choose a low value of K, such as between three and five. However, for the individual stock data considered in the empirical analysis, the method of Bai and Ng (2002) usually leads to large values for K (see Figure 4), such that the impact of the errors might become less relevant. Moreover, discarding the errors leads to an optimization program in K variables instead of N, which substantially accelerates the optimization in large investment universes. This method, called *fast IC estimate*, is detailed in Online Appendix C. We observe, however, that completely discarding the errors comes at a cost: the fast IC estimate of the optimal portfolio has a larger turnover and worse out-of-sample performance.

In the next proposition, we derive formulas for the moments $\hat{m}_k(\hat{P})$ of order up to four, as in Proposition 1, which avoids having to plug in very large coskewness and cokurtosis tensors.

Proposition 4. *The approximation of the moments of $\hat{P}$ of orders 2, 3, and 4 in* (23) *takes the same form as in Proposition 1, with $m_3(\boldsymbol{w}'\boldsymbol{A}_K\boldsymbol{Y}_K)$ and $m_4(\boldsymbol{w}'\boldsymbol{A}_K\boldsymbol{Y}_K)$, respectively, replaced by*

$$\hat{m}_3\left(\boldsymbol{w}'\boldsymbol{A}_K^\perp\boldsymbol{Y}_K^\perp\right) = \sum_{i=1}^{K}\left(\boldsymbol{w}'\boldsymbol{A}_K^\perp\right)_i^3 m_3\left(\boldsymbol{Y}_{K,i}^\perp\right), \tag{24}$$

$$\hat{m}_4\left(\boldsymbol{w}'\boldsymbol{A}_K^\perp\boldsymbol{Y}_K^\perp\right) = \sum_{i=1}^{K}\left(\boldsymbol{w}'\boldsymbol{A}_K^\perp\right)_i^4 m_4\left(\boldsymbol{Y}_{K,i}^\perp\right) + 3\sum_{i=1}^{K}\sum_{j\neq i}^{K}\left(\boldsymbol{w}'\boldsymbol{A}_K^\perp\right)_i^2\left(\boldsymbol{w}'\boldsymbol{A}_K^\perp\right)_j^2. \tag{25}$$

When higher moments are considered in the objective function ψ, the PC and IC estimates $\hat{\boldsymbol{w}}_{\psi,K}$ and $\hat{\boldsymbol{w}}_{\psi,K}^\perp$ differ if $I(\boldsymbol{Y}_K^\perp) > 0$. We quantify below the difference in the number of parameters to be estimated when using the PCA factor model instead of the ICA factor model.

Proposition 5. *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*

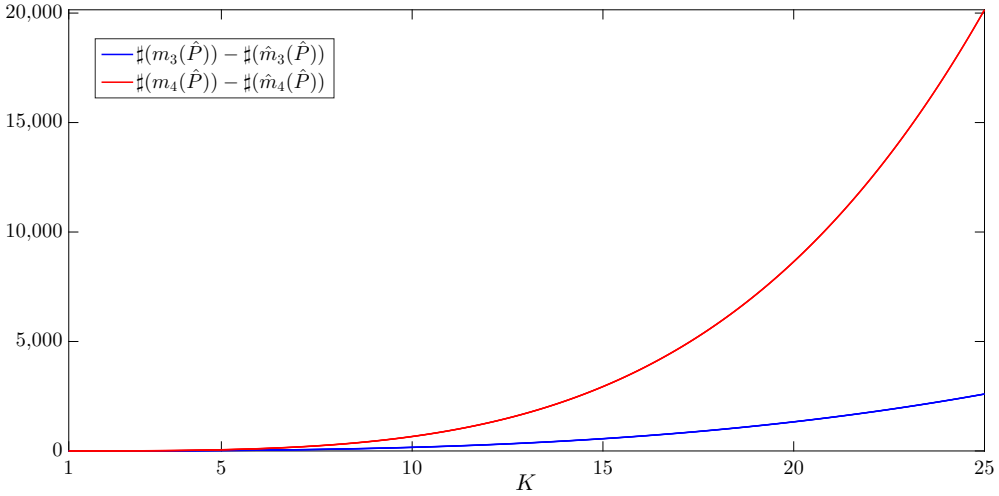
$$\sharp(m_k(\hat{P})) - \sharp(\hat{m}_k(\hat{P})) = \binom{K+k-1}{k} - K\left(\frac{K-1}{2} + k - 3 + \mathbb{1}_{k=3}\right). \tag{26}$$

Figure 1 depicts the difference in (26) for the third and fourth moments and shows that the gain in parameters via our parsimonious approximation is substantial and increases quickly with K.

4. Monte Carlo simulations

We now use simulated data to assess the out-of-sample benefit of using the IC versus the PC portfolio estimate. Simulated data allow us to control the parameters of interest, particularly the number of assets N and the sample size T. In addition, we can average the performance across various scenarios, in contrast to real data for which a single history exists. We discuss how we simulate the data in Section 4.1, the objective function considered in Section 4.2, and the results are discussed in Section 4.3.

Figure 1: Difference in the number of parameters needed to estimate the portfolio-return third and fourth central moments when using principal versus independent components



Notes. The figure depicts the difference between the number of parameters that need to be estimated to estimate the third and fourth central moments of the portfolio return $\hat{P}$, which is based on a K-factor model and defined in (6). We depict the difference between the number of parameters when the principal components are used as factors $(m_3(\hat{P}), m_4(\hat{P}))$ versus when the independent components are used as factors $(\hat{m}_3(\hat{P}), \hat{m}_4(\hat{P}))$. The blue line depicts the difference for the third central moment and the red line for the fourth central moment. The general expression for the difference is given in Proposition 5, Equation (26).

4.1. Data generation process

The asset-return data are generated in a realistic manner given what is known about observed returns, but the benefit here is that we have control over the simulation setting.

We simulate $T + \tau$ asset returns $\boldsymbol{X}$, where T is the sample size used to estimate the portfolios, and τ is the size of the out-of-sample period. We consider returns at daily and weekly frequencies, and we consider 252 days and 52 weeks per year. We fix τ to be 20 years of returns while varying T. As in the numerical settings of MacKinlay and Pastor (2000), DeMiguel et al. (2009b) and Chen and Yuan (2016), we simulate the asset returns $\boldsymbol{X}$ from a factor model. However, in contrast to those papers, we consider non-Gaussian returns. We set

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{Y} + \boldsymbol{\epsilon}, \tag{27}$$

where the loading matrix $\boldsymbol{B}$ is of size $N \times K$, the factors $\boldsymbol{Y}$ are of size $K \times (T + \tau)$, and the errors $\boldsymbol{\epsilon}$ are of size $N \times (T + \tau)$. A detailed description of the model is available in Online Appendix B. In summary, the factors $\boldsymbol{Y}$ have marginal Student's t distributions with a dependence controlled by a Clayton copula, the errors $\boldsymbol{\epsilon}$ are Gaussian, and the loading matrix $\boldsymbol{B}$ is simulated such that more variance comes from the factors than the errors.

4.2. Objective function

Once the $T+\tau$ asset returns $\boldsymbol{X}$ are randomly generated, we compute the PC and IC estimates of the optimal portfolio $\boldsymbol{w}_\psi$ (see Definitions 1 and 3) on the returns from 1 to T and assess their out-of-sample performance on the returns from $T+1$ to $T+\tau$. Unreported simulations show that estimation errors in the sample estimate of the factor-model covariance matrix $\widehat{\boldsymbol{\Sigma}}$ in (15) increase quickly with the dimension N. To alleviate this issue, the eigendecomposition $(\boldsymbol{V}_K, \boldsymbol{\Lambda}_K)$ and the variance of errors $\boldsymbol{\epsilon}$ are estimated via the shrinkage estimator of the covariance matrix of Ledoit and Wolf (2004). Higher comoments are estimated via sample averages.

As a risk measure ψ, we consider the *value-at-risk* (VaR), which is one of the most popular risk measures; see, for instance, Campbell et al. (2001), Alexander and Baptista (2004), and Lwin et al. (2017). Specifically, we consider the modified VaR (MVaR), which corresponds to the four-moment Gram-Charlier expansion of the VaR. The MVaR is commonly used in the portfolio-selection literature because it focuses on moments that have an intuitive interpretation

and, thus, that matter for investors; see Favre and Galeano (2002), Eling and Schuhmacher (2007), and Boudt et al. (2008). The MVaR at confidence level α corresponds to the objective function

$$\psi\Big(m_2(P), m_3(P), m_4(P)\Big) = \text{MVaR}_\alpha(P) = \sqrt{m_2(P)}\left((2z_\alpha^3 - 5z_\alpha)\frac{\zeta(P)^2}{36} - (z_\alpha^2 - 1)\frac{\zeta(P)}{6} - (z_\alpha^3 - 3z_\alpha)\frac{\kappa(P)}{24} - z_\alpha\right), \quad (28)$$

where

$$\zeta(P) := m_3(P)/m_2(P)^{3/2} \quad \text{and} \quad \kappa(P) := m_4(P)/m_2(P)^2 - 3 \quad (29)$$

are the skewness and excess kurtosis, respectively, and z_α is the standard Gaussian quantile at confidence level α. We fix $\alpha = 1\%$, as in Bali et al. (2013). Note that standardized higher moments appear in (28) but are multiplied by the standard deviation $\sqrt{m_2(P)}$, which accounts for the magnitude of the losses and positively impacts the MVaR. Moreover, the MVaR formula in (28) is relevant in practice because it is imposed by European Supervisory Authorities to assess market risk in the packaged retail and insurance-based investment products (PRIIPs) regulations.

Although we focus our analysis on $\text{MVaR}_\alpha(P)$ as a risk measure for conciseness, our methodology is suited for any objective function ψ based on moments. Additional analyses demonstrate that our approach continues to work well when considering other objective functions, such as the lower partial moment (LPM) or the entropy. The empirical results for the LPM are discussed in Section 5.4.

4.3. Results

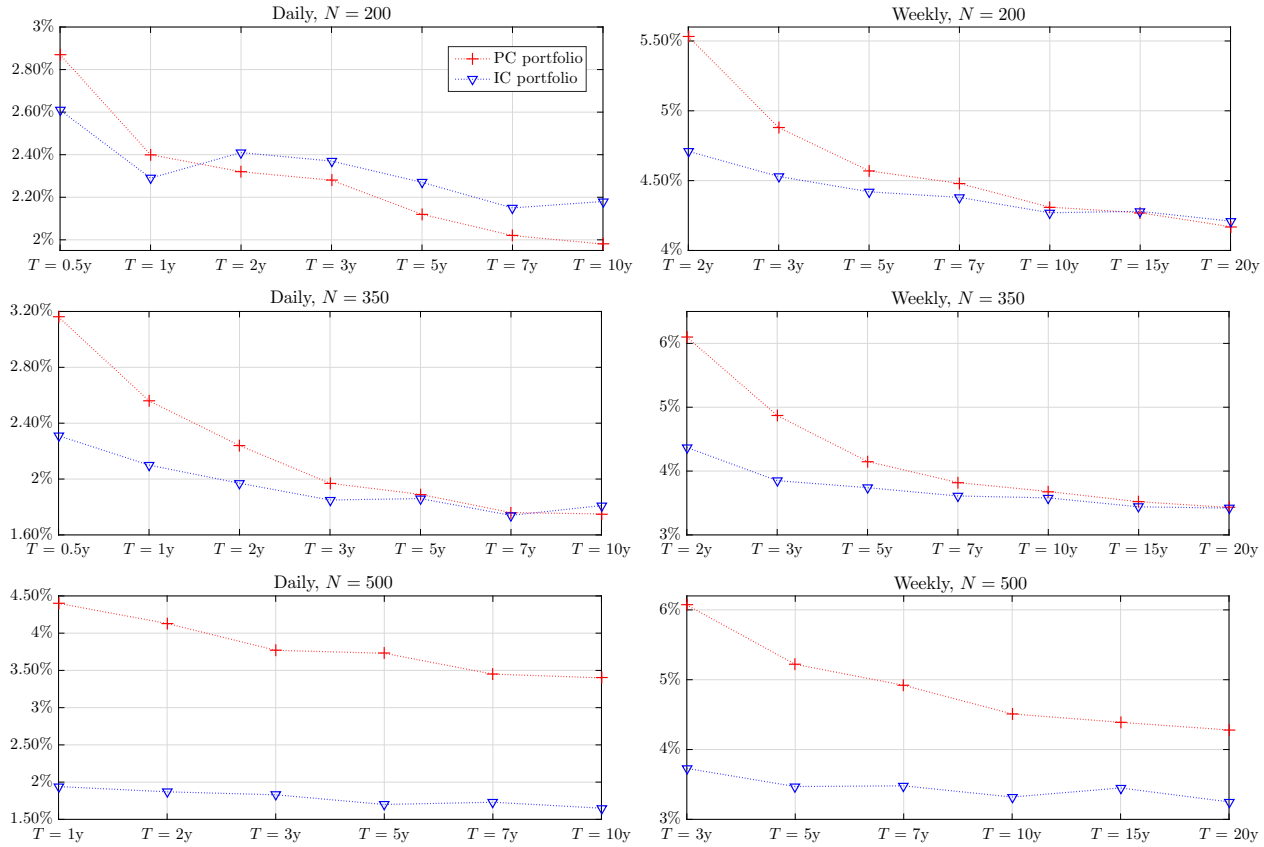
In Figure 2, we report the out-of-sample MVaR in (28) of the PC and IC estimates of the minimum-MVaR portfolio, averaged over 100 Monte Carlo simulations. We consider high-dimensional datasets of size $N = 200$, 350, and 500 with $K = N/10$. For daily returns, we consider sample sizes T from 0.5 to 10 years, and for weekly returns, we consider sample sizes T from 2 to 20 years. The only difference is for the case $N = 500$, where we begin at $T = 1$ year for daily returns and $T = 3$ years for weekly returns to obtain a sample size sufficiently large relative to N. Optimal sample portfolios might not exist when $N > T$ because the sample covariance matrix is not invertible. Therefore, for the few simulations in which the PC or IC portfolio does not converge numerically, we add no-short-selling constraints to ensure convergence.

We make several observations. First, the out-of-sample performance of both the PC and IC portfolios improves with the sample size T. This result is expected because the asset-return distribution is stationary in our simulation setting. Second, for a given T, the outperformance of the IC portfolio relative to the PC portfolio is more prominent for a large number of assets N and, thus, for a larger number of factors K. This result is also expected because Proposition 2 shows that the number of comoments of PCs quickly increases with K. Specifically, for daily returns, PC manages to outperform IC for $N = 200$ when the sample size is two years or more. However, for $N = 350$, PC outperforms only for $T = 10$ years or more. For $N = 500$, IC always performs much better. Even when the sample size is $T = 10$ years = 2520 days, IC has a daily out-of-sample MVaR of 1.65% on average over the 100 simulations versus 3.40% for PC. For weekly returns with smaller sample sizes, the results are even more in favor of the ICA parsimonious approach.

Finally, another important benefit of the parsimonious IC portfolio is that its weights are less extreme than those of the PC portfolio. To illustrate this characteristic, Figure 3 depicts the PC and IC portfolio weights sorted in descending order for the case with $T = 5$ years of weekly returns. The figure shows that the IC portfolio weights are closer to $1/N$ and, thus, are better diversified. Moreover, the IC portfolio is characterized by less short-selling. For example, for the case with $N = 200$ in Figure 3, PC has 84 negative weights versus 69 for IC. The less extreme weights of IC mean that when we report the out-of-sample volatility instead of MVaR in Figure 2, the returns of IC portfolio are *always* less volatile. Taken together, these characteristics are attractive to asset managers who are reluctant to implement strategies with portfolio compositions that are overly extreme.

To summarize, even in the ideal case of a very long estimation window of i.i.d. returns, the IC estimate of the optimal portfolio can outperform the PC estimate, particularly when considering large investment universes. This is because the benefit resulting from the sharp decrease in the number of parameters to estimate is greater than the loss associated with the assumption that the K ICs are independent.

Figure 2: Out-of-sample modified value-at-risk via Monte Carlo simulations



Notes. The figure depicts the out-of-sample 1% modified VaR (MVaR) of the PC estimate (red crosses) and IC estimate (blue triangles) of the minimum-MVaR portfolio following the Monte Carlo simulations of Section 4. The random sampling of asset returns is described in Section 4.1, and the computation of the portfolio strategies in Section 4.2. The portfolios are estimated on a window of T returns reported on the x-axis, and their MVaR on the y-axis is computed out of sample on a window of $\tau = 20$ years. The left plots are for daily returns, and the right plots are for weekly returns. We consider a number of assets $N = 200$, 350, and 500 with a number of factors $K = N/10$. The out-of-sample performance is averaged over 100 Monte Carlo simulations. One can see that the IC portfolio outperforms the PC portfolio unless the sample size T is large relative to N.

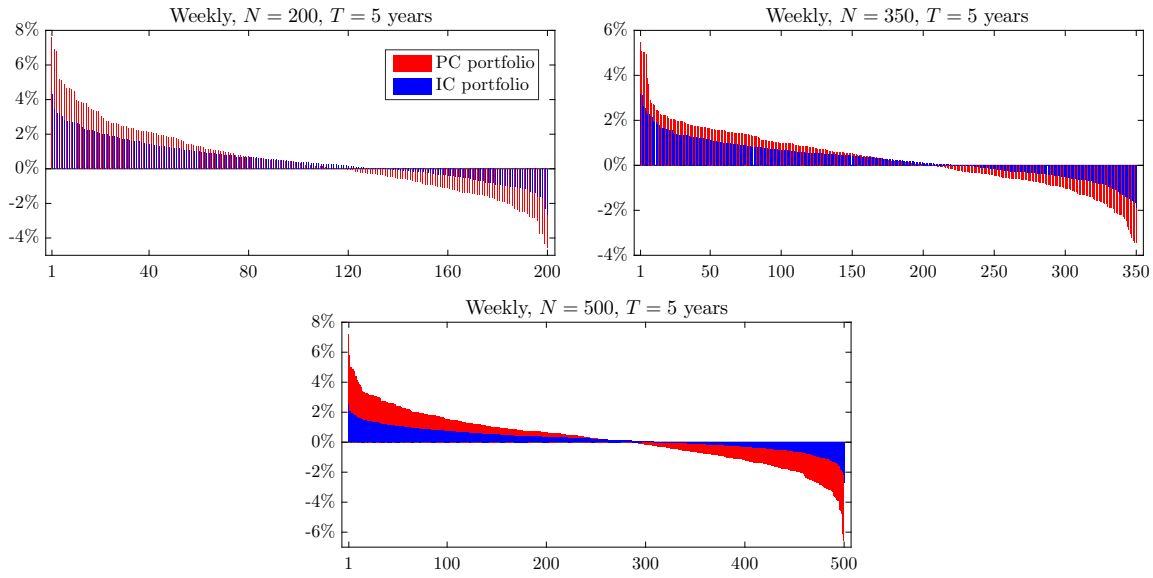
5. Empirical analysis

We now assess the out-of-sample empirical performance of the PC and IC portfolio estimates on high-dimensional datasets of individual stocks. Section 5.1 details the methodology, Section 5.2 compares the independence of PCs versus ICs, Section 5.3 discusses the results, and Section 5.4 closes with robustness tests.

5.1. Methodology

Data. We rely on datasets of individual stocks using the methodology in Jagannathan and Ma (2003), DeMiguel et al. (2009a), and Behr et al. (2013), among others, which avoids survivorship bias. In particular, we download daily returns for all available stocks from the Center for Research in Security Prices (CRSP) for the period from January 1980 to December 2019. Starting from January 1990 and then every six months, we identify the subset of CRSP stocks that have historical returns for the ten preceding years and the next six months. Following the methodology adopted in the aforementioned papers, we remove microcap and penny stocks, which we define as stocks that, in the preceding five years, were below the 20th percentile of average market capitalization and had an average price of less than 5 dollars. Finally, from all remaining stocks, we randomly select $N = 500$ stocks to form our investment universe for the next six months. We repeat this procedure every six months until the end of the sample is reached. We also consider datasets of size $N = 100$ and 300, which we randomly select from the entire dataset of size $N = 500$, enabling us to

Figure 3: Portfolio weights of PC and IC estimates of the minimum-MVaR portfolio



Notes. The figure depicts the portfolio weights of the PC and IC estimates of the minimum-1%-modified-VaR (MVaR) portfolio following the Monte Carlo simulations of Section 4. The random sampling of asset returns is described in Section 4.1, and the computation of the portfolio strategies in Section 4.2. The plots are generated for the case in which the portfolios are estimated on a window of $T = 5$ years of weekly returns. The portfolio weights are sorted in descending order. We consider a number of assets $N = 200$, 350, and 500 with a number of factors $K = N/10$. One can see that the IC estimate of the minimum-MVaR portfolio leads to less extreme weights than the PC estimate.

evaluate how the size of the investment universe affects out-of-sample performance. Notably, we consider significantly larger datasets than those used in previous studies on higher-moment portfolio selection.[2]

We consider daily returns because they exhibit stronger deviations from normality than do lower-frequency returns (Martellini and Ziemann, 2010), a situation in which higher-moment strategies are particularly relevant. Daily returns also improve estimation accuracy, which is important for portfolios optimizing higher moments.

In Online Appendix D, we compare the ICs with the five Fama-French factors, one of the most popular ways to explain the cross-section of stock returns.

Portfolio strategies. We consider two robust benchmarks: the EW portfolio and the MV portfolio using the shrinkage estimator of the covariance matrix of Ledoit and Wolf (2004). The MV portfolio is a sensible benchmark because it has reduced sensitivity to estimation error (DeMiguel et al., 2009a; Behr et al., 2013) and has been shown by Martellini and Ziemann (2010) to be difficult to outperform out of sample, even in terms of higher-moment criteria.

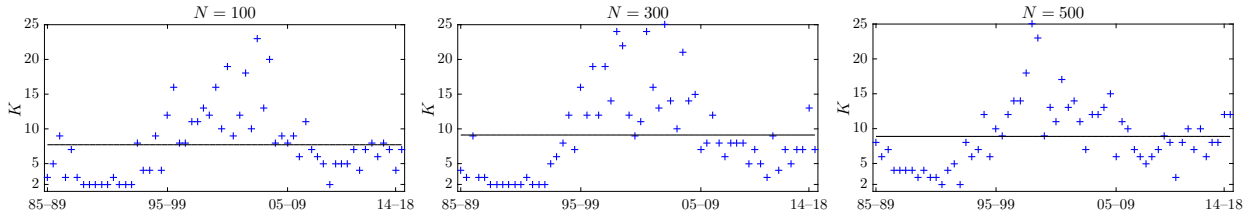
Next, as in the Monte Carlo simulations in Section 4, we consider the PC and IC estimates of the portfolio $\boldsymbol{w}_\psi$ using as risk measure ψ the MVaR in (28). To be consistent with the MV benchmark, the eigendecomposition $(\boldsymbol{V}_K, \boldsymbol{\Lambda}_K)$ and the variance of errors $\boldsymbol{\epsilon}$ are estimated via the covariance matrix of Ledoit and Wolf (2004). Higher comoments are estimated via sample averages.

Finally, all portfolios are computed under short-selling constraints, $w_i \geq 0$ for all i. As we show in Section 5.4, this approach improves the out-of-sample performance relative to that when short-selling is allowed.

Rolling-window methodology. To evaluate the out-of-sample performance of the portfolios, we employ a classical rolling-window methodology, as in DeMiguel et al. (2009a). We use an estimation window of T daily returns to

[2]The maximum number of assets in other papers considering higher-moment strategies under estimation risk is as follows: $N = 7$ for Harvey et al. (2010), $N = 30$ for Martellini and Ziemann (2010), $N = 17$ for Boudt et al. (2015), $N = 14$ for Ghalanos et al. (2015), $N = 18$ for Hitaj et al. (2015), and $N = 10$ for Boudt et al. (2020b). A notable difference is Boudt et al. (2020a), who consider up to $N = 100$. In contrast, we consider up to $N = 500$.

Figure 4: Time evolution of the number of factors K



Notes. The figure depicts the number K of factors retained for the three datasets of individual stocks and a sample size of $T = 5$ years of daily returns. It is selected via the method of Bai and Ng (2002) in Equation (30). The horizontal black line depicts the average value of K. The five-year estimation window is rolled by six months over time and covers the period from January 1985 to April 2019. The corresponding years are reported on the x-axis.

estimate the portfolios, where T is taken as 3, 5, and 10 years to assess the impact of the sample size. The portfolio returns are then computed out of sample for the next six months, and the estimation window is rolled forward by six months. The procedure is repeated until the end of the sample period is reached. Regardless of the sample size T used, the out-of-sample period spans January 1990 to December 2019.

Calibration of number of factors K. We follow the method of Bai and Ng (2002), which is well suited for large investment universes because it is consistent as both N and T tend together to infinity, in contrast to the popular AIC and BIC criteria (Bai and Ng, 2002). More explicitly, K is chosen as

$$K = \underset{K\in[1,N]}{\operatorname{argmin}} \ \ln\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}^2\right) + K\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right), \tag{30}$$

where ϵ_{it} is the error of asset i at time t when K factors are retained. Chen and Yuan (2016) introduce a method based on PCA to improve the estimation of mean-variance portfolios in large investment universes, and use the same method to fix the number of factors. They find, on the basis of simulation studies, that it often yields the true number of factors. Figure 4 depicts the time evolution of K for the three datasets for the case where $T = 5$ years. The figure shows that K remains reasonably small with respect to N but can nevertheless reach values up to 25, in which case the reduction in the number of parameters to estimate achieved by our ICA methodology is substantial, as shown in Figure 1.

Performance criterion, transaction costs, and p-values. As a performance criterion, we first report the modified Sharpe ratio (Gregoriou and Gueyie, 2003; Eling and Schuhmacher, 2007), which is defined as the annualized ratio between the mean return and MVaR:

$$\sqrt{252}\times\mu(P)/\mathrm{MVaR}_\alpha(P), \tag{31}$$

where $\mathrm{MVaR}_\alpha(P)$ is the risk measure in (28) that the PC and IC portfolios attempt to minimize. The modified Sharpe ratio measures risk-adjusted performance using MVaR as a risk measure. Such mean-VaR ratios are popular in the literature to assess hedge-fund performance (Bali et al., 2013). We also consider risk-based criteria, using the annualized volatility $\sqrt{252m_2(P)}$ and the MVaR as risk measures.

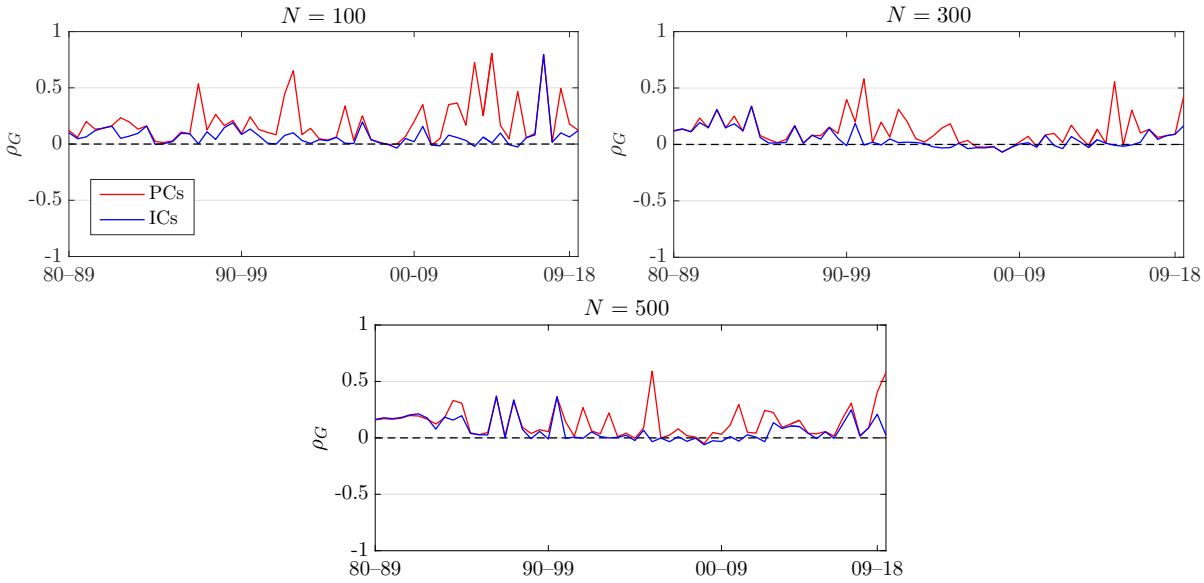
In addition, we report the daily turnover, and the returns are computed net of transaction costs. We assume a proportional transaction cost parameter of $c = 50$ basis points, as in DeMiguel et al. (2009b, p.1929–1930). Transaction costs can have a severe impact on performance when accounting for higher moments, but most previous studies on higher-moment strategies under estimation risk do not take them into account.[3]

Finally, we compute two-sided p-values for the difference in the modified Sharpe ratio adjusted to transaction costs of two portfolios via circular block bootstrap following Ardia and Boudt (2015).[4] We compare PC versus MV, IC versus MV, and IC versus PC.

[3]However, we note that Boudt et al. (2020a, 2020b) report the breakeven transaction cost compared with the EW portfolio, but our empirical results suggest that the MV portfolio might be a better benchmark than the EW portfolio.

[4]We set the block size b via the rule-of-thumb $b = [\tau^{1/3}] = 20$, where $\tau = 30\times 252$ is the size of the out-of-sample period. As in DeMiguel et al. (2009a), the number of bootstrap resamples is set equal to $B = 200\times b$.

Figure 5: Nonlinear correlation of principal versus independent components



Notes. The figure depicts the nonlinear correlation over time of the $K = 2$ principal (in red) and independent (in blue) components for the three datasets of individual stocks and a sample size of $T = 10$ years. The nonlinear 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 (32). The ten-year estimation window is rolled by six months over time and covers the period from January 1980 to April 2019. The corresponding years are reported on the x-axis.

5.2. Independence of PCs versus ICs

The parsimonious representation of the asset-return comoment tensors via ICA is based on the assumption that the ICs are independent. The error resulting from this approximation is expected to be smaller when considering the ICs, instead of the PCs, as factors. To illustrate this finding, we report in Figure 5 the evolution over time of the nonlinear correlation between the $K = 2$ PCs and ICs, which is defined as

$$\rho_G(Y_1, Y_2) := \frac{\mathrm{Cov}\left(G(Y_1), G(Y_2)\right)}{\sqrt{m_2(G(Y_1))m_2(G(Y_2))}}, \tag{32}$$

where $G(x) = \ln \cosh x$ is the nonlinear mapping used by default in *FastICA*. This definition is motivated by the fact that independence between two variables (X, Y) amounts to decorrelation between any of their nonlinear transforms; that is, $\mathrm{Cov}(f(X), g(Y)) = 0$ for all functions f, g. We consider the case in which the sample size is $T = 10$ years. Figure 5 shows that ρ_G is much closer to zero for the ICs than the PCs, illustrating the validity of our approach.

5.3. Comparison of out-of-sample performance

Table 1 compares the out-of-sample performance of the considered portfolios. In terms of risk-adjusted performance measured by the modified Sharpe ratio, the main observation is that, consistent with the Monte Carlo simulations in Section 4, the out-of-sample performance of the IC portfolio improves with the sample size T and, regardless of the number of assets N, the best performance is achieved by the IC portfolio in all considered cases. These results illustrate the superiority of our approach. We subsequently describe the results in greater detail.

First, we compare the IC and PC portfolios. The numerous higher comoments of factors that must be estimated to compute the PC portfolio appear to deteriorate its performance compared with that of the IC portfolio. Indeed, IC systematically achieves a larger modified Sharpe ratio than PC, and the difference is often statistically significant.[5] In

[5]Surprisingly, when $T = 10$ years, the outperformance is statistically significant for $N = 100$ and 300 but not $N = 500$. This can be explained

Table 1: Turnover and out-of-sample performance net of transaction costs on empirical data

	$N = 100$	$N = 300$	$N = 500$	$N = 100$	$N = 300$	$N = 500$
	Modified Sharpe ratio			Modified Value-at-Risk		
EW	0.176	0.186	0.188	6.17%	6.45%	6.27%
$T = 3$ years						
MV	0.184	0.267	0.273	4.49%	3.31%	3.28%
PC	0.209	$0.030_{\star\star\star}$	$0.041_{\star\star\star}$	5.04%	46.39%	32.49%
IC	0.229	$0.271^{\star\star\star}$	$0.302^{\star\star\star}$	5.14%	4.59%	3.82%
$T = 5$ years						
MV	0.158	0.256	0.265	5.30%	3.36%	3.25%
PC	0.214	0.260	0.079	5.39%	4.79%	14.62%
IC	0.235	0.290	$0.305^{\star}_{\star\star}$	5.21%	4.67%	4.01%
$T = 10$ years						
MV	0.127	0.238	0.254	6.41%	3.73%	3.59%
PC	$0.209_{\star}$	$0.330_{\star\star\star}$	$0.345_{\star\star}$	5.73%	4.18%	3.98%
IC	$0.242^{\star\star}_{\star\star}$	$0.365^{\star\star\star}_{\star\star\star}$	$0.378_{\star\star\star}$	5.27%	4.07%	3.67%
	Volatility			Turnover		
EW	16.38%	15.70%	15.48%	1.81%	1.78%	1.78%
$T = 3$ years						
MV	10.96%	9.22%	8.46%	2.32%	2.47%	2.52%
PC	13.94%	35.80%	45.53%	2.17%	2.44%	2.49%
IC	14.61%	11.50%	10.22%	2.26%	2.46%	2.55%
$T = 5$ years						
MV	11.45%	9.40%	8.65%	2.36%	2.48%	2.54%
PC	14.87%	12.05%	12.79%	2.25%	2.50%	2.63%
IC	15.29%	12.08%	10.69%	2.37%	2.55%	2.65%
$T = 10$ years						
MV	12.34%	10.02%	9.23%	2.37%	2.47%	2.51%
PC	15.90%	12.52%	11.02%	2.23%	2.51%	2.61%
IC	16.35%	12.76%	11.18%	2.35%	2.56%	2.65%

Notes. The table reports the turnover and out-of-sample modified Sharpe ratio, modified Value-at-Risk, and volatility of the equally weighted portfolio (EW), minimum-variance portfolio (MV), PC estimate of the minimum-MVaR portfolio (PC), and IC estimate of the minimum-MVaR portfolio (IC). We consider three datasets of individual stocks from CRSP following the methodology in Section 5.1. We compute the portfolios under short-selling constraints. The modified Sharpe ratio in (31) and the volatility are annualized. The modified Value-at-Risk in (28) and the turnover are in daily terms. All performance criteria and computed net of transaction costs, assuming a proportional transaction cost of 50 basis points. The number K of factors retained is selected using the method of Bai and Ng (2002) in Equation (30). The out-of-sample period spans January 1990 to December 2019, and we estimate the portfolios on 3-, 5-, and 10-year windows rolled over time by six months. The stars $\star$, $\star\star$, $\star\star\star$ indicate that the two-sided p-value of the difference in the modified Sharpe ratio of two portfolios, computed using the method of Ardia and Boudt (2015), is smaller than 15%, 10%, and 5%. We compare PC vs. MV (subscript next to PC), IC vs. MV (subscript next to IC), and IC vs. PC (superscript next to IC).

terms of risk, IC has a similar volatility to PC and, importantly, a smaller MVaR in all cases except one. Moreover, IC has a slightly larger turnover than PC, which indicates that the outperformance of IC is even more prominent before transaction costs. Interestingly, for all three criteria, the performance of the IC portfolio improves with the number of assets N, meaning that it benefits from the further diversification potential that arises when considering a larger investment universe. In contrast, the performance of PC severely worsens with the estimation risk (higher N and smaller T). These results confirm that the curse of dimensionality has a severe impact on higher-moment strategies, which our ICA parsimonious strategy successfully resolves.

Second, we evaluate whether higher-moment strategies manage to outperform the EW and MV benchmark portfolios. Despite its lower turnover, EW is systematically outperformed by IC—and even by PC when $T = 10$ years—suggesting that naive diversification does not work well. Similarly, MV generally outperforms EW and has the smallest volatility in all cases, as expected. The small volatility of MV combined with its reduced estimation risk explain its good performance in terms of MVaR. However, for the case with a sample size of 10 years, the IC portfolio has a much smaller MVaR for 100 stocks and only a slightly worse one for 300 and 500 stocks. Moreover, in terms of the modified Sharpe ratio, IC outperforms MV in all cases, with large statistical significance when $T = 10$ years. Thus, IC achieves a larger average return than MV, which combined with its low risk makes it a more attractive strategy. This result is illustrated in Figure 6, which depicts the cumulative wealth of the two portfolios over the 1990-2019 out-of-sample period for the case with a sample size of 10 years. The figure clearly shows that the IC portfolio reaches a larger cumulative wealth than the MV portfolio, which confirms the practical impact of our method.

These results clearly show that addressing the curse of dimensionality substantially improves the performance of higher-moment strategies. This finding corroborates Martellini and Ziemann (2010), who also find that, for daily returns, improved higher-comoment estimates help outperform the MV portfolio. However, our setting is much more high-dimensional and prone to estimation error because we consider up to $N = 500$ stocks, whereas Martellini and Ziemann (2010) consider only up to $N = 30$. Our results also mitigate the conclusions in Simaan (2014) and Khashanah et al. (2020), for example, who find for several utility functions that the portfolio on the mean-variance efficient frontier optimizing the chosen utility function outperforms the fully optimized portfolio that takes all moments into account. Finally, note that we use sample averages to estimate higher moments, which means that more sophisticated estimation, such as the nearest-comoment method of Boudt et al. (2020b), is not necessary once we use ICA to estimate higher-comoment tensors, although it could provide additional gains.

5.4. Robustness tests

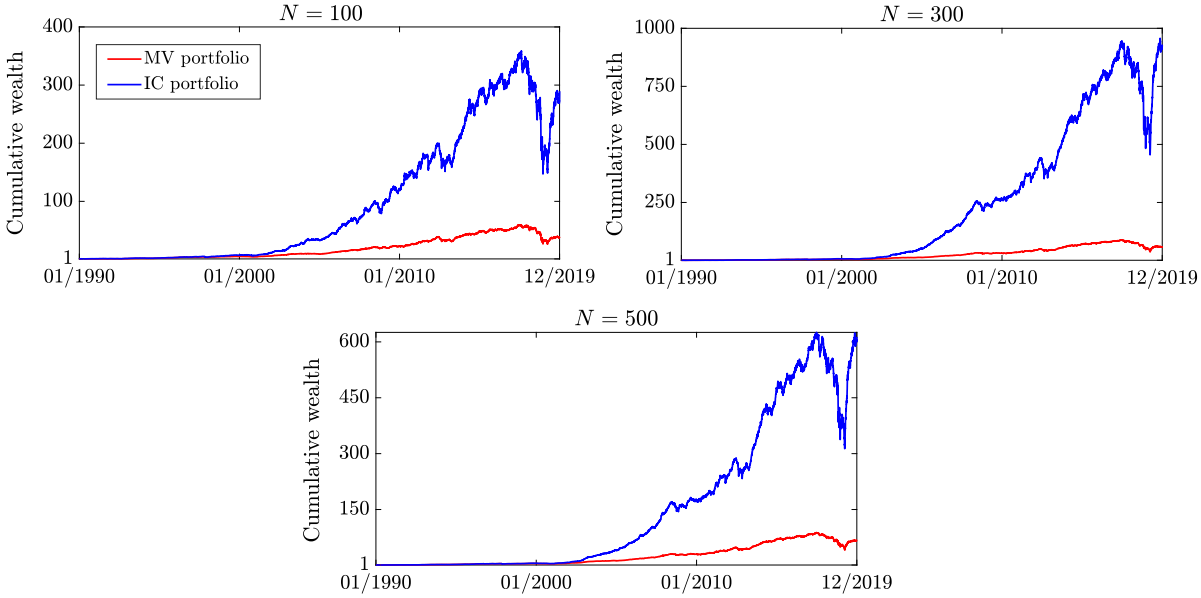
We test the robustness of our results in three ways, focusing on the modified Sharpe ratio as the criterion for brevity. First, following Ghalanos et al. (2015) who consider an independent factor model without dimension reduction, we test the effect of setting $K = N$ on the out-of-sample performance of the IC portfolio. We consider the dataset of size $N = 100$ for brevity. Table 2a) shows that reducing the dimension with $K < N$ leads to substantially better performance than that with $K = N$. Thus, both dimension reduction and an independent factor model are needed to make higher-moment strategies appealing in large investment universes.

Second, we test the effect of allowing short-selling on the out-of-sample performance of the MV, PC, and IC portfolios. We consider the datasets of size $N = 100$ for brevity. Table 2b) shows that our base-case results with no short-selling are always better than those with short-selling allowed, which is consistent with Jagannathan and Ma (2003) and Behr et al. (2013). Moreover, the conclusion that the IC portfolio outperforms the MV and PC portfolios is robust to allowing short-selling.

Third, we assess whether the good performance of the IC portfolio still applies when considering other risk measures. As an alternative to MVaR, we rely on the four-moment Gram-Charlier expansion of the lower partial moment (LPM); see Leòn and Moreno (2017). The methodology and results are reported in the appendix. We find that the results are robust to using the LPM as an objective function. Most importantly, regardless of the number of assets N,

by the fact that the outperformance in terms of the modified Sharpe ratio results both from a larger average return and a smaller MVaR for $N = 100$ and 300, whereas it results nearly exclusively from a smaller MVaR for $N = 500$. Indeed, the annualized average returns of the PC and IC portfolios are, respectively, $(0.191, 0.203)$ for $N = 100$, $(0.219, 0.236)$ for $N = 300$, and $(0.219, 0.220)$ for $N = 500$. That the outperformance is statistically significant for $N = 100$ and 300 but not $N = 500$ is thus likely a result of this difference. This can be explained by the fact that the MVaR depends on higher moments that have large sampling errors; hence, it is more difficult to achieve statistical significance from an improvement in MVaR than an improvement in average return.

Figure 6: Cumulative wealth of the MV and IC portfolios for a sample size of 10 years



Notes. The figure depicts the cumulative wealth attained by the minimum-variance portfolio (MV) and the IC estimate of the minimum-MVaR portfolio (IC) for the case with a sample size of 10 years of daily returns. We consider an initial wealth of one dollar, and then compute the cumulative wealth using the out-of-sample returns adjusted to transaction costs of each portfolio strategy, assuming a proportional transaction cost of 50 basis points. The datasets consist of $N = 100, 300$ and 500 individual stocks from CRSP. We compute the portfolios under short-selling constraints.

the best out-of-sample performance is achieved by the IC portfolio estimated with a sample size of $T = 10$ years, which is similar to the case with MVaR as a risk measure. Thus, the applicability of our method is rather general.

6. Conclusion

Although investors care about higher-moment features in asset returns, such as asymmetry and fat tails, these features are often dismissed in portfolio strategies. The reason is found in the dimensionality of the portfolio problem: the number of comoments to estimate quickly increases with the order of the moments, as well as the number of assets at hand. This curse of dimensionality severely impacts the sensitivity to estimation error of optimal portfolio strategies and leads to disappointing out-of-sample performance. Factor models were previously considered to alleviate this problem but are limited in their applicability because numerous higher comoments of factors must still be estimated, no dimension reduction is performed, or a high computational burden is imposed. In this paper, we propose projecting asset returns onto a small set of maximally independent factors found via independent component analysis. This procedure is shown to resolve the curse of dimensionality: by exploiting the near independence of these factors, the number of free parameters in the comoment tensors of asset returns no longer grows according to a power law but is a linear function of the number of factors. In contrast to previous studies that consider small investment universes, our parsimonious approach is scalable to high-dimensional settings. In particular, we consider up to $N = 500$ assets in our Monte Carlo simulations and empirical study on individual stocks. We find, for the value-at-risk and lower partial moment as risk measures, that our method substantially improves out-of-sample risk-adjusted performance net of transaction costs compared with a shrinkage estimator of the minimum-variance portfolio and a factor model based on principal components.

Table 2: Out-of-sample modified Sharpe ratio ratio net of transaction costs for two robustness tests

a) Effect of dimension reduction on IC portfolio

	$K < N$	$K = N$
$T = 3$ years	0.229	0.192
$T = 5$ years	0.235	0.195
$T = 10$ years	0.242	0.144

b) Effect of short-selling on MV, PC and IC portfolios

	MV	PC	IC
$T = 3$ years			
Short-selling	0.182	0.186	0.194
No short-selling	0.184	0.209	0.229
$T = 5$ years			
Short-selling	0.158	0.196	0.213
No short-selling	0.158	0.214	0.235
$T = 10$ years			
Short-selling	0.124	0.164	0.215
No short-selling	0.127	0.209	0.242

Notes. The table reports the out-of-sample modified Sharpe ratio for the robustness tests of Section 5.4. Table a) considers the IC estimate of the minimum-MVaR portfolio (IC) for the case in which $K < N$ is set via the method of Bai and Ng (2002) in Equation (30), and for the case in which $K = N$. Table b) considers the minimum-variance portfolio (MV), PC estimate of minimum-MVaR portfolio (PC), and IC portfolio for the cases in which short selling is allowed and not allowed. We consider the dataset of individual stocks from CRSP of size $N = 100$, and the empirical methodology is available in Section 5.1. The modified Sharpe ratio in (31) is annualized and computed net of transaction costs, assuming a proportional transaction cost of 50 basis points. The out-of-sample period spans January 1990 to December 2019, and we estimate the portfolios on 3-, 5-, and 10-year windows rolled over time by six months.

Appendix A. Out-of-sample performance for the lower partial moment as an objective function

In this appendix, we replicate the empirical results for a different objective function—the lower partial moment (LPM)—which is a well-studied downside risk measure in finance; see, for instance, Price et al. (1982), Jarrow and Zao (2006), and Anthonisz (2012). We consider the LPM of order one, which is the risk measure used in the popular omega ratio of Keating and Shadwick (2002). Given a threshold u, the LPM of order one of the portfolio return P with density f_P is

$$\mathrm{LPM}_u(P) := \int_{-\infty}^{u} (u - x) f_P(x) \mathrm{d}x. \tag{A.1}$$

We fix $u = \mu(P)$ and, similar to the definition of the MVaR in (28), we approximate the LPM via the four-moment Gram-Charlier expansion of Leòn and Moreno (2017). As we formally prove in Online Appendix A.6, this corresponds to the objective function

$$\psi(m_2(P), m_4(P)) = \sqrt{\frac{m_2(P)}{2\pi}} \left(1 + \frac{23}{24}\kappa(P)\right). \tag{A.2}$$

Consistent with the modified Sharpe ratio used in (31), we consider, as a portfolio-performance criterion, the annualized ratio between the portfolio mean return and the LPM: $\sqrt{252} \times \mu(P)/\mathrm{LPM}_{\mu(P)}(P)$. We replicate the results of Table 1 for the LPM in (A.2), and they are reported in Table A.3.

References

Alexander, G., Baptista, A. 2004. A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model. *Management Science* 50(9), 1261–1273.

Ang, A., Chen, J., Yu, H. 2006. Downside risk. *Review of Financial Studies* 19(4), 1191–1239.

Anthonisz, S. 2012. Asset pricing with partial-moments. *Journal of Banking and Finance* 36(7), 2122–2135.

Ardia, D., Boudt, K. 2015. Testing equality of modified Sharpe ratios. *Finance Research Letters* 13, 97–104.

Bai, J., Ng, S. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70(1), 191–221.

Bali, T., Brown, S., Demirtas, K. 2013. Do hedge funds outperform stocks and bonds? *Management Science* 59(8), 1887–1903.

Table A.3: Turnover and out-of-sample mean-LPM ratio net of transaction costs on empirical data

	Mean-LPM ratio			Turnover		
	$N = 100$	$N = 300$	$N = 500$	$N = 100$	$N = 300$	$N = 500$
EW	0.175	0.175	0.178	1.81%	1.78%	1.78%
$T = 3$ years						
MV	0.171	0.273	0.269	2.32%	2.47%	2.52%
PC	0.247	0.211	$0.313_{\star\star}$	2.20%	2.51%	2.67%
IC	$0.255_{\star}$	$0.303^{\star\star}_{\star}$	0.212	3.21%	3.69%	3.78%
$T = 5$ years						
MV	0.137	0.257	0.261	2.36%	2.48%	2.54%
PC	0.218	0.206	$0.328_{\star\star\star}$	2.28%	2.58%	2.75%
IC	0.249	$0.295^{\star}_{\star\star}$	$0.196^{\star\star\star}$	3.18%	3.46%	3.56%
$T = 10$ years						
MV	0.107	0.232	0.238	2.37%	2.47%	2.51%
PC	$0.232_{\star}$	$0.354_{\star\star\star}$	0.337	2.26%	2.56%	2.70%
IC	$0.334^{\star}_{\star\star}$	$0.425_{\star\star\star}$	$0.426_{\star\star\star}$	2.69%	2.84%	2.95%

Notes. The table reports the out-of-sample turnover and mean-lower-partial-moment (LPM) ratio of the equally weighted portfolio (EW), minimum-variance portfolio (MV), PC estimate of the minimum-LPM portfolio (PC), and IC estimate of the minimum-LPM portfolio (IC). We consider three datasets of individual stocks from CRSP following the methodology in Section 5.1 and the appendix. We compute the portfolios under short-selling constraints. The mean-LPM ratio is annualized and computed net of transaction costs, assuming a proportional transaction cost of 50 basis points. The number K of factors retained is selected via the method of Bai and Ng (2002) in Equation (30). The out-of-sample period spans January 1990 to December 2019, and we estimate the portfolios on 3-, 5-, and 10-year windows rolled over time by six months. The stars $\star$, $\star\star$, $\star\star\star$ indicate that the two-sided p-value of the difference in the mean-LPM ratio of two portfolios, computed by adapting the methodology of Ardia and Boudt (2015) to the LPM in (A.2) as risk measure, is smaller than 15%, 10%, and 5%. We compare PC vs. MV (subscript next to PC), IC vs. MV (subscript next to IC), and IC vs. PC (superscript next to IC).

Behr, P., Guettler, A., Miebs, F. 2013. On portfolio optimization: Imposing the right constraints. *Journal of Banking and Finance* 37(4), 1232–1242.

Boudt, K., Cornilly, D., Verdonck, T. 2020a. A coskewness shrinkage approach for estimating the skewness of linear combinations of random variables. *Journal of Financial Econometrics* 18(1), 1–23.

Boudt, K., Cornilly, D., Verdonck, T. 2020b. Nearest comoment estimation with unobserved factors. *Journal of Econometrics* 217(2), 381–397.

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

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

Brandt, M., Santa-Clara, P., Valkanov, R. 2009. Parametric portfolio policies: Exploiting characteristics in the cross section of equity returns. *Review of Financial Studies* 22(9), 3411–3447.

Briec, W., Kerstens, K., Jokung, O. 2007. Mean-variance-skewness portfolio performance gauging: A general shortage function and dual approach. *Management Science* 53(1), 135–149.

Briec, W., Kerstens, K., Van de Woestyne, I. 2013. Portfolio selection with skewness: A comparison of methods and a generalized one fund result. *European Journal of Operational Research* 230(2), 412–421.

Campbell, R., Huisman, R., Koedijk, K. 2001. Optimal portfolio selection in a Value-at-Risk framework. *Journal of Banking and Finance* 25(9), 1789–1804.

Campbell, J., Lo, A., MacKinlay, A. 1997. *The Econometrics of Financial Markets*. New Jersey, Princeton University Press.

Chen, Y., Härdle, W., Spokoiny, V. 2007. Portfolio value at risk based on independent component analysis. *Journal of Computational and Applied Mathematics* 205(1), 594–607.

Chen, J., Yuan, M. 2016. Efficient portfolio selection in a large market. *Journal of Financial Econometrics* 14(3), 496–524.

Cover, T., Thomas, J. 2006. *Elements of Information Theory* (2nd ed.). New Jersey, Wiley.

De Nard, G., Ledoit, O., Wolf, M. 2019. Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics* 33.

DeMiguel, V., Garlappi, L., Nogales, F., Uppal, R. 2009a. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science* 55(5), 798–812.

DeMiguel, V., Garlappi, L., Uppal, R. 2009b. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies* 22(5), 1915–1953.

Eling, M., Schuhmacher, F. 2007. Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking and Finance* 31(9), 2632–2647.

Favre, L., Galeano, J. 2002. Mean-modified value-at-risk optimization with hedge funds. *Journal of Alternative Investments* 5(2), 21–25.

Ghalanos, A., Rossi, E., Urga, G. 2015. Independent factor autoregressive conditional density model. *Econometric Reviews* 34(5), 594–616.

Green, J., Hand, J., Zhang, X. 2013. The supraview of return predictive signals. *Review of Accounting Studies* 18, 692–730.

Gregoriou, G., Gueyie, J. 2003. Risk-adjusted performance of funds of hedge funds using a modified Sharpe ratio. *Journal of Wealth Management* 6(3), 77–83.

Guidolin, M., Timmermann, A. 2008. International asset allocation under regime switching, skew, and kurtosis preferences. *Review of Financial Studies* 21(2), 889–935.

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

Harvey, C., Siddique, A. 2000. Conditional skewness in asset pricing tests. *Journal of Finance* 55(3), 1263–1295.

Hitaj, A., Mercuri, L., Rroji, E. 2015. Portfolio selection with independent component analysis. *Finance Research Letters* 15, 146–159.

Hyvärinen, A. 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks* 10(3), 626–634.

Hyvärinen, A., Karhunen, J., Oja, E. 2001. *Independent Component Analysis*. New York, Wiley.

Jagannathan, R., and Ma, T. 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance* 58(4), 1651–1684.

Jarrow, R., Zhao, F. 2006. Downside loss aversion and portfolio management. *Management Science* 52(4), 558–566.

Jondeau, E., Jurczenko, E., Rockinger, M. 2018. Moment component analysis: An illustration with international stock markets. *Journal of Business & Economic Statistics* 36(4), 576–598.

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

Keating, C., Shadwick, W. 2002. A universal performance measure. *Journal of Performance Measurement* 6(3), 59–84.

Khashanah, K., Simaan, M., Simaan, Y. 2020. Do we need higher-order comoments to enhance mean-variance portfolios? SSRN 3523379.

Lassance, N., DeMiguel, V., Vrins, F. 2020. Optimal portfolio diversification via independent component analysis. Working paper, London Business School.

Lassance, N., Vrins, F. 2019. Minimum Rényi entropy portfolios. *Annals of Operations Research*, `https://doi.org/10.1007/s10479-019-03364-2`.

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

Leòn, A., Moreno, M. 2017. One-sided performance measures under Gram-Charlier distributions. *Journal of Banking and Finance* 74, 38–50.

Lwin, K., Qu, R., MacCarthy, B. 2017. Mean-VaR portfolio optimization: A nonparametric approach. *European Journal of Operational Research* 260(2), 751–766.

MacKinlay, A., Pastor, L. 2000. Asset pricing models: Implications for expected returns and portfolio selection. *Review of Financial Studies* 13(4), 883–916.

Markowitz, H. 1952. Portfolio selection. *Journal of Finance* 7(1), 77–91.

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

Merton, R. 1980. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8(4), 323–361.

Price, K., Price, B., Nantell, T. 1982. Variance and lower partial moment measures of systematic risk: Some analytical and empirical results. *Journal of Finance* 37(3), 843–855.

Scott, R., Horvath, P. 1980. On the direction of preference for moments of higher order than the variance. *Journal of Finance* 35(4), 915–919.

Simaan, Y. 2014. The opportunity cost of mean-variance choice under estimation risk. *European Journal of Operational Research* 234(2), 382–391.

Zakamouline, V., Koekebakker, S. 2009. Portfolio performance evaluation with generalized Sharpe ratios: Beyond the mean and variance. *Journal of Banking and Finance* 33(7), 1242–1254.

Zoia, M., Biffi, P., Nicolussi, F. 2018. Value at risk and expected shortfall based on Gram-Charlier-like expansions. *Journal of Banking and Finance* 93, 92–104.

Online Appendix to "Portfolio selection with parsimonious higher comoments estimation"

A. Proofs of results

A.1. Proposition 1

Proof. The result hinges on the following identities for the moments of order 2, 3, and 4 of a sum of N independent random variables $Z_1 \perp Z_2 \perp \cdots \perp Z_N$:

$$\begin{aligned} m_2\left(\sum_{i=1}^{N} Z_i\right) &= \sum_{i=1}^{N} m_2(Z_i), \\ m_3\left(\sum_{i=1}^{N} Z_i\right) &= \sum_{i=1}^{N} m_3(Z_i), \\ m_4\left(\sum_{i=1}^{N} Z_i\right) &= \sum_{i=1}^{N} m_4(Z_i) + 3\sum_{i=1}^{N}\sum_{j\neq i}^{N} m_2(Z_i)m_2(Z_j). \end{aligned} \tag{A.1}$$

$\hat{P} = \boldsymbol{w}'\boldsymbol{A}_K\boldsymbol{Y}_K + \boldsymbol{w}'\boldsymbol{\varepsilon}$ is the sum of two independent random variables because we assume that $\boldsymbol{Y}_K$ and $\boldsymbol{\varepsilon}$ are independent. As a result, the moments of orders 2, 3, and 4 of $\hat{P}$ become

$$\begin{aligned} m_2(\hat{P}) &= m_2(\boldsymbol{w}'\boldsymbol{A}_K\boldsymbol{Y}_K) + m_2(\boldsymbol{w}'\boldsymbol{\varepsilon}), \\ m_3(\hat{P}) &= m_3(\boldsymbol{w}'\boldsymbol{A}_K\boldsymbol{Y}_K) + m_3(\boldsymbol{w}'\boldsymbol{\varepsilon}), \\ m_4(\hat{P}) &= m_4(\boldsymbol{w}'\boldsymbol{A}_K\boldsymbol{Y}_K) + m_4(\boldsymbol{w}'\boldsymbol{\varepsilon}) + 6m_2(\boldsymbol{w}'\boldsymbol{A}_K\boldsymbol{Y}_K)m_2(\boldsymbol{w}'\boldsymbol{\varepsilon}). \end{aligned} \tag{A.2}$$

Because $\boldsymbol{A}_K\boldsymbol{A}_K' = \boldsymbol{V}_K\boldsymbol{\Lambda}_K\boldsymbol{V}_K'$ and the errors $\boldsymbol{\varepsilon}$ are assumed to be mutually independent, we have a simplification of the following terms by application of (A.1): $m_2(\boldsymbol{w}'\boldsymbol{A}_K\boldsymbol{Y}_K) = \boldsymbol{w}'\boldsymbol{V}_K\boldsymbol{\Lambda}_K\boldsymbol{V}_K'\boldsymbol{w}$, $m_2(\boldsymbol{w}'\boldsymbol{\varepsilon}) = \sum_{i=1}^{N} w_i^2 m_2(\varepsilon_i)$, $m_3(\boldsymbol{w}'\boldsymbol{\varepsilon}) = \sum_{i=1}^{N} w_i^3 m_3(\varepsilon_i)$, $m_4(\boldsymbol{w}'\boldsymbol{\varepsilon}) = \sum_{i=1}^{N} w_i^4 m_4(\varepsilon_i) + 3\sum_{i=1}^{N}\sum_{j\neq i}^{N} w_i^2 w_j^2 m_2(\varepsilon_i)m_2(\varepsilon_j)$. Plugging these values into (A.2) proves the proposition. □

A.2. Proposition 2

Proof. Because we assume that the errors $\boldsymbol{\varepsilon}$ are independent from the principal components (PCs), we can decompose the cardinality of $\boldsymbol{M}_k(\hat{\boldsymbol{X}})$, as in (11).

The cardinality of $\boldsymbol{A}_K = \boldsymbol{V}_K \boldsymbol{\Lambda}_K^{1/2}$ is $\sharp(\boldsymbol{A}_K) = \sharp(\boldsymbol{V}_K) + \sharp(\boldsymbol{\Lambda}_K)$. Clearly, $\sharp(\boldsymbol{\Lambda}_K) = K$. The matrix $\boldsymbol{V}_K$ consists 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 $\sharp(\boldsymbol{A}_K)$ given by (12).

The cardinality of $\boldsymbol{M}_k(\boldsymbol{Y}_K)$ in (13) is obtained by plugging $N \leftarrow K$ into (3), except $\sharp(\boldsymbol{M}_2(\boldsymbol{Y}_K)) = 0$ because the PCs are standardized to have unit variance.

We turn finally to the errors $\boldsymbol{\varepsilon}$. To count the number of distinct parameters in each comoment tensor $\boldsymbol{M}_k(\boldsymbol{\varepsilon})$ for $k \geq 2$, knowing that the errors $\boldsymbol{\varepsilon}$ are mutually independent, let us start with some specific values of k. Note that, because of independence, the expectation of a product of several ε_i's simplifies to the product of the corresponding moments. For $k = 2$, the tensor corresponds to the covariance matrix and is diagonal, corresponding to N entries $m_2(\varepsilon_i)$. For $k = 3$, the entries (i, j, k) and (i, i, j) (and their permutations) are all zero because the mean of each error is zero. Thus, only the N entries (i, i, i), equal to $m_3(\varepsilon_i)$, must be counted. However, for $k = 4$, one must count not only the N moments $m_4(\varepsilon_i)$ but also the moments corresponding to the entries (i, i, j, j) (and their permutations) that equal $m_2(\varepsilon_i)m_2(\varepsilon_j)$. Thus, one must add the N moments $m_2(\varepsilon_i)$, yielding $2N$ moments in total for $k = 4$. Note that the third moments $m_3(\varepsilon_i)$ are not needed because they appear in the entries (i, i, i, j) (and their permutations) that equal zero because the mean of each error is zero. By generalizing this result, we have that $\sharp(\boldsymbol{M}_2(\boldsymbol{\varepsilon})) = N$ and, for $k \geq 3$, $\boldsymbol{M}_k(\boldsymbol{\varepsilon})$ requires all marginal moments of order 2 to k, except those of order $k-1$. Thus, $\sharp(\boldsymbol{M}_k(\boldsymbol{\varepsilon})) = K(k-2)$ for $k \geq 3$, proving Equation (14). □

A.3. Proposition 3

Proof. To count the number of distinct parameters in each higher-comoment tensor $\boldsymbol{M}_k(\boldsymbol{S}_K)$, $k \geq 3$, knowing that the factors $\boldsymbol{S}_K$ are mutually independent, we can follow the same line of reasoning as for the errors in Proposition 2. The only difference is that the factors $\boldsymbol{S}_K$ have unit variance and, thus, second moments must not be counted. Therefore, for $k \geq 3$, $\boldsymbol{M}_k(\boldsymbol{S}_K)$ requires all marginal moments of order 3 to k, except for those of order $k-1$. This results in $\sharp(\boldsymbol{M}_3(\boldsymbol{S}_K)) = K$ and $\sharp(\boldsymbol{M}_k(\boldsymbol{S}_K)) = K(k-3)$ for $k \geq 4$, which corresponds to Equation (18). □

A.4. Proposition 4

Proof. The result is direct given that the independent components (ICs) $\boldsymbol{Y}_K^{\perp}$ are assumed independent and, thus, $\hat{m}_3\left(\boldsymbol{w}'\boldsymbol{A}_K^{\perp}\boldsymbol{Y}_K^{\perp}\right)$ and $\hat{m}_4\left(\boldsymbol{w}'\boldsymbol{A}_K^{\perp}\boldsymbol{Y}_K^{\perp}\right)$ in (24)–(25) are obtained by applying (A.1). □

A.5. Proposition 5

Proof. We have that $m_k(\hat{P}) = \boldsymbol{w}'\boldsymbol{M}_k(\hat{\boldsymbol{X}})\bigotimes_{i=1}^{k-1}\boldsymbol{w}$ and $\hat{m}_k(\hat{P}) = \boldsymbol{w}'\widehat{\boldsymbol{M}}_k(\hat{\boldsymbol{X}})\bigotimes_{i=1}^{k-1}\boldsymbol{w}$. Thus, the difference in cardinality is $\sharp\left(\boldsymbol{M}_k(\hat{\boldsymbol{X}})\right) - \sharp\left(\widehat{\boldsymbol{M}}_k(\hat{\boldsymbol{X}})\right)$. From Proposition 2, $\sharp\left(\boldsymbol{M}_k(\hat{\boldsymbol{X}})\right) = \sharp(\boldsymbol{A}_K) + \sharp(\boldsymbol{M}_k(\boldsymbol{Y}_K)) + \sharp(\boldsymbol{M}_k(\boldsymbol{\varepsilon})) = \sharp(\boldsymbol{V}_K) + \sharp(\boldsymbol{\Lambda}_K) + \sharp(\boldsymbol{M}_k(\boldsymbol{Y}_K)) + \sharp(\boldsymbol{M}_k(\boldsymbol{\varepsilon}))$. From Proposition 3, $\sharp\left(\widehat{\boldsymbol{M}}_k(\hat{\boldsymbol{X}})\right) = \sharp(\boldsymbol{A}_K^{\perp}) + \sharp\left(\widehat{\boldsymbol{M}}_k(\boldsymbol{Y}_K^{\perp})\right) + \sharp(\boldsymbol{M}_k(\boldsymbol{\varepsilon})) = \sharp(\boldsymbol{V}_K) + \sharp(\boldsymbol{\Lambda}_K) + \sharp(\boldsymbol{R}_K^{\perp}) + \sharp\left(\widehat{\boldsymbol{M}}_k(\boldsymbol{Y}_K^{\perp})\right) + \sharp(\boldsymbol{M}_k(\boldsymbol{\varepsilon}))$. Thus, the difference becomes $\sharp(\boldsymbol{M}_k(\boldsymbol{Y}_K)) - \left(\sharp\left(\boldsymbol{R}_K^{\perp}\right) + \sharp\left(\widehat{\boldsymbol{M}}_k(\boldsymbol{Y}_K^{\perp})\right)\right)$ and simplifies to (26) from $\sharp(\boldsymbol{M}_k(\boldsymbol{Y}_K))$ in (13), $\sharp\left(\widehat{\boldsymbol{M}}_k(\boldsymbol{Y}_K^{\perp})\right)$ in (18), and $\sharp(\boldsymbol{R}_K^{\perp}) = K(K-1)/2$. □

A.6. Gram-Charlier expansion of the lower partial moment in (35)

Proof. For a general threshold u, Leòn and Moreno (2017) show that the four-moment Gram-Charlier expansion of $\mathrm{LPM}_u(P)$ in (34) is given by

$$\mathrm{LPM}_u(P) \approx \mathrm{LPM}_{u,\phi}(P) + \frac{\zeta(P)}{\sqrt{6}}\theta_{2,1} + \frac{\kappa(P)}{\sqrt{24}}\theta_{3,1}, \tag{A.3}$$

where $\zeta(P), \kappa(P)$ are the standardized skewness and excess kurtosis, and

$$\begin{aligned}
\mathrm{LPM}_{u,\phi}(P) &= (u-\mu(P))\Phi(u^\star) + \sqrt{m_2(P)}\phi(u^\star), \\
\theta_{j,1} &= (u-\mu(P))A_{0,j} - \sqrt{m_2(P)}A_{1,j}, \\
A_{k2} &= \frac{1}{\sqrt{6}}(B_{k+3} - 3B_{k+1}), \\
A_{k3} &= \frac{1}{\sqrt{24}}(B_{k+4} - 6B_{k+2} + 3B_k), \\
B_k &= \int_{-\infty}^{u^\star} x^k \phi(x)\mathrm{d}x,
\end{aligned} \tag{A.4}$$

with ϕ and Φ the pdf and cdf of the standard Gaussian, and $u^\star := (u-\mu(P))/\sqrt{m_2(P)}$. When $u = \mu(P)$, $u^\star = 0$ and (A.3)–(A.4) simplify to

$$\mathrm{LPM}_u(P) \approx \sqrt{m_2(P)}\left(\frac{1}{\sqrt{2\pi}} - \frac{\zeta(P)}{6}(B_4 - 3B_2) - \frac{\kappa(P)}{24}(B_5 - 6B_3 + 3B_1)\right), \quad B_k = \int_{-\infty}^{0} x^k \phi(x)\mathrm{d}x. \tag{A.5}$$

Finally, for $u^\star = 0$, $B_1 = -1/\sqrt{2\pi}$, $B_2 = 1/2$, $B_3 = \sqrt{2/\pi}$, $B_4 = 3/2$ and $B_5 = -4\sqrt{2/\pi}$; hence, $B_4 - 3B_2 = 0$ and $B_5 - 6B_3 + 3B_1 = -23/\sqrt{2\pi}$, resulting in (35). □

B. Data generation process for the Monte Carlo simulations

We describe in detail how we generate the components of the factor model in (27), which is used for the Monte Carlo simulations in Section 4.

First, similar to the PCs, the K factors $\boldsymbol{Y}$ are uncorrelated and have unit variance but are not independent. Specifically, we generate the factors from a Clayton copula[1] (see Section 5.4 of McNeil et al., 2009) with marginal distributions being unit-variance Student's t distributions that have a number of degrees of freedom ν that is randomly sampled from a uniform distribution between 3 and 10 for daily returns and 5 and 15 for weekly returns. This accounts for the fact that daily

[1]Contrary to popular elliptical copulas such as the multivariate Gaussian or Student's t, the Clayton copula is asymmetric and features more dependence among negative than positive returns, consistent with observed asset returns.

returns exhibit stronger deviations from normality than lower-frequency returns (Martellini and Ziemann, 2010). The parameter θ of the Clayton copula that controls for the dependence is set equal to $\theta = 0.5$, which corresponds to a Kendall tau of 20%. We then make the covariance matrix of the factors unitary by projecting them via PCA.

Second, the errors $\boldsymbol{\varepsilon}$ are drawn from a multivariate Gaussian distribution, which means that the non-Gaussian and nonlinear dependency features of the asset returns come from the factors and not the errors. We set the Gaussian mean-return vector to zero. The standard deviation of each error is randomly sampled from a uniform distribution between $0.05/\sqrt{d}$ and $0.15/\sqrt{d}$, where d is the number of returns per year; that is, a 10% yearly standard deviation on average. Regarding the correlation matrix, we simulate it using the methodology of Lewandowski et al. (2009), which allows us to generate positive-definite correlation matrices while leaving free the marginal distribution of each correlation coefficient ρ_{ij}. We take as a marginal distribution a Beta(20,20) shifted between -1 and 1. This means that ρ_{ij} is zero on average, and the probability that $\rho_{ij} \in [-0.2, 0.2]$ is 80%. Thus, whereas the errors are not assumed to be independent, they are independent on average, and the dependence is not large.

Third, the loading matrix $\boldsymbol{B}$ is set as in (5) as $\boldsymbol{B} = \boldsymbol{V}_K \boldsymbol{\Lambda}_K^{1/2}$, where $(\boldsymbol{V}_K, \boldsymbol{\Lambda}_K)$ comes from the eigendecomposition of a randomly generated $N \times N$ covariance matrix. The standard deviations in this covariance matrix are randomly sampled from a uniform distribution between $0.10/\sqrt{d}$ and $0.30/\sqrt{d}$; that is, twice the standard deviation of the errors on average.[2] We simulate the correlation matrix similarly to that of the errors, except the marginal distribution of each correlation coefficient ρ_{ij} is a Beta(26, 53/3). This choice yields a mode for ρ_{ij} of 20% and makes the probability of a negative correlation $\rho_{ij} < 0$ equal to 10%. This result is consistent with observed stock returns, which often have a small positive correlation.

[2]This is consistent with Goldfarb and Iyengar (2003, p.4): "The eigenvalues of the residual covariance matrix $\mathbf{D}$ are typically much smaller than those of the covariance matrix $\mathbf{V}^\top \mathbf{F} \mathbf{V}$ implied by the factors."

Table C.1 Turnover and out-of-sample modified Sharpe ratio of the fast IC estimate of the minimum-MVaR portfolio

		$T = 3$ years		$T = 5$ years		$T = 10$ years	
		IC	Fast IC	IC	Fast IC	IC	Fast IC
$N = 100$	MSR	0.229	0.134	0.235	0.215	0.242	0.159
	Turnover	2.26%	3.08%	2.37%	2.94%	2.35%	2.88%
$N = 300$	MSR	0.271	0.238	0.290	0.230	0.365	0.257
	Turnover	2.46%	2.77%	2.55%	2.86%	2.56%	2.64%
$N = 500$	MSR	0.302	0.179	0.305	0.233	0.426	0.208
	Turnover	2.55%	2.87%	2.65%	2.69%	2.95%	2.60%

Notes. The table reports the out-of-sample modified Sharpe ratio of the IC estimate (Definition 3) and fast IC estimate (Definition C.1) of the minimum-MVaR portfolio for the three datasets of individual stocks from CRSP, following the methodology in Section 5.1. We compute the portfolios under short-selling constraints. The modified Sharpe ratio in (32) is annualized and computed net of transaction costs, assuming a proportional transaction cost of 50 basis points. The number K of factors retained is selected using the method of Bai and Ng (2002) in Equation (31). The out-of-sample period spans January 1990 to December 2019, and we estimate the portfolios on 3-, 5-, and 10-year windows rolled over time by six months.

C. Fast IC estimate

Following Remark 3, we introduce the *fast IC estimate* of portfolio $\boldsymbol{w}_\psi$. Note that we focus on ICA because it is our contribution, but the approach can be adapted to the PC estimate in a similar fashion. The idea is to discard the errors $\boldsymbol{\varepsilon}$ and approximate the asset returns $\boldsymbol{X}$ as $\hat{\boldsymbol{X}} = \boldsymbol{A}_K^{\perp}\boldsymbol{Y}_K^{\perp}$. The approximation of the portfolio return P then becomes $\hat{P} = \boldsymbol{w}'\hat{\boldsymbol{X}} = \tilde{\boldsymbol{w}}'\boldsymbol{Y}_K^{\perp}$, where $\tilde{\boldsymbol{w}} = (\boldsymbol{A}_K^{\perp})'\boldsymbol{w}$ are the exposures on the K ICs $\boldsymbol{Y}_K^{\perp}$. Thus, the portfolio of N assets can be written equivalently as a portfolio of K ICs. The fast IC estimate is then defined as follows.

Definition C.1. *The fast IC estimate of the portfolio $\boldsymbol{w}_\psi$ in (2) is*

$$\hat{\boldsymbol{w}}_{\psi,K}^{\perp} := \boldsymbol{A}_K^{\perp}\tilde{\boldsymbol{w}}_{\psi,K}^{\perp},$$

where

$$\tilde{\boldsymbol{w}}_{\psi,K}^{\perp} := \operatorname*{argmin}_{\tilde{\boldsymbol{w}}\in\widetilde{\mathscr{W}}^{\perp}} \psi\big(\hat{m}_j(\hat{P}(\tilde{\boldsymbol{w}})),\ldots,\hat{m}_k(\hat{P}(\tilde{\boldsymbol{w}})),\ldots,\hat{m}_l(\hat{P}(\tilde{\boldsymbol{w}}))\big)$$

with

$$\widetilde{\mathscr{W}^{\perp}} := \left\{ \boldsymbol{v} \in \mathbb{R}^K \;\middle|\; \boldsymbol{A}_K^{\perp} \boldsymbol{v} \in \mathscr{W} \right\}.$$

Although it is not realistic to assume that the small-dimensional factor model does not leave space for errors, it has the practical benefit of accelerating the optimization: the fast IC estimate in Definition C.1 optimizes only K IC exposures versus N portfolio weights for the original IC estimate in Definition 3. Moreover, it avoids estimating the marginal moments of the errors $\boldsymbol{\varepsilon}$, which might reduce the estimation error and compensate for the bias induced by ignoring errors. We replicate in Table C.1 the out-of-sample performance of the IC portfolio in Table 1 for the case without errors and compare it to the case with errors included. The table shows that the faster computation achieved by ignoring errors comes at a cost: the fast IC estimate has a larger turnover and worse out-of-sample performance.

D. Independent components and Fama-French factors

Although we focus on latent factors (PCs and ICs) in the paper, a large portion of the literature that explains stock returns via factor models focuses on observable economic factors and financial characteristics. Arguably, the most famous model for that purpose is the five-factor model of Fama and French (2015) based on market, size, value, profitability and investment.

In this appendix, we examine the relationship between the FF factors and the ICs computed from them. We collect daily returns on the five FF factors from Kenneth French's library for the period from November 1997 to October 2017. In Figure D.1, we depict several descriptive statistics of the FF factors and ICs, which are standardized to have unit variance. Subfigures (a), (b), and (c) depict the historical returns, skewness and excess kurtosis of the factors. The distribution of the FF factors is largely non-Gaussian, depicting peaks in returns around 2008. Therefore, the ICs, which are obtained as linear combinations of the FF factors, are also non-Gaussian. Regarding

subfigures (b) and (c), most of the non-Gaussianity around the 2008 crisis appears to be captured by the first IC. This can be explained because the ICs are iteratively computed by maximizing non-Gaussianity; see Hyvärinen et al. (2001) for details. Finally, in our context, the most relevant difference between the FF factors and the ICs concerns their dependence structure. Indeed, as subfigure (d) shows, the nonlinear pairwise correlations (computed as in Section 5.2) between the FF factors are larger than those between the ICs: the average over the ten possible pairs is 35% for the FF factors versus 15% for the ICs. This smaller dependence makes the ICs a particularly appealing choice of factors to estimate higher-comoment tensors.

References

Bai, J., Ng, S. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70(1), 191–221.

Fama, E., French, K. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116(1), 1–22.

Goldfarb, D., Iyengar, G. 2003. Robust portfolio selection problems. *Mathematics of Operations Research* 28(1), 1–38.

Hyvärinen, A., Karhunen, J., Oja, E. 2001. *Independent Component Analysis*. New York, Wiley.

Leòn, A., Moreno, M. 2017. One-sided performance measures under Gram-Charlier distributions. *Journal of Banking and Finance* 74, 38–50.

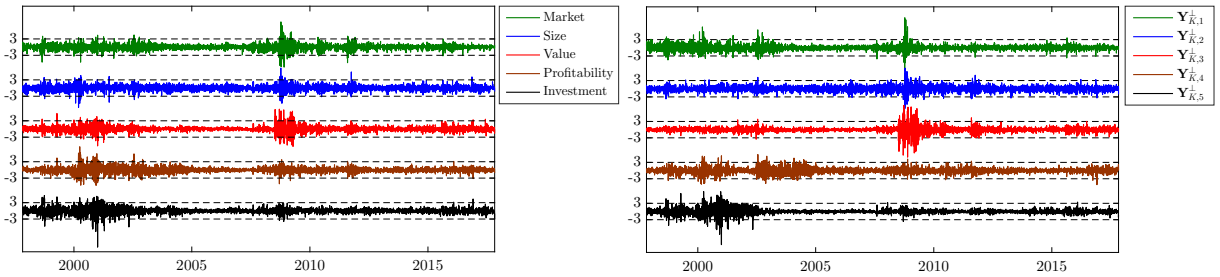
Lewandowski, D., Kurowicka, D., Joe, H. 2009. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* 100(9), 1989–2001.

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

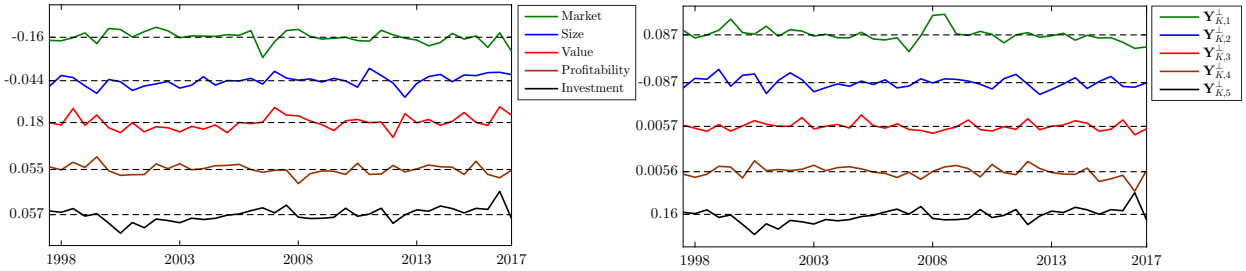
McNeil, A., Frey, R., Embrechts, P. 2005. *Quantitative Risk Management: Concepts, Techniques and Tools*. New Jersey, Princeton University Press.

Figure D.1 Characteristics of independent components and Fama-French factors

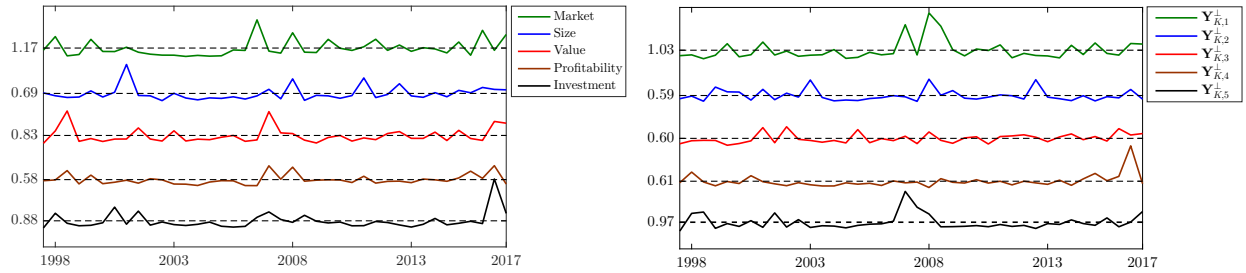
(a) Historical returns of factors



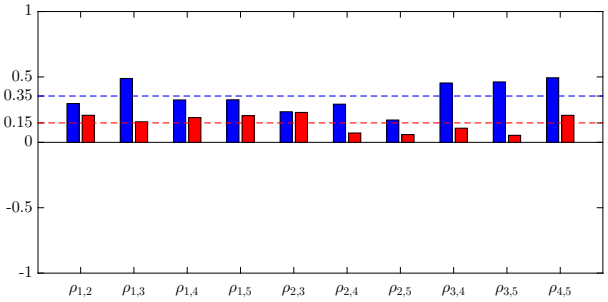
(b) Rolling value of skewness of factors



(c) Rolling value of excess kurtosis of factors



(d) Non-linear correlation of factors



Notes. The figure depicts characteristics of the five factors of Fama and French (2015), and of the independent components computed from them, using daily returns spanning November 1997 to October 2017. The FF factors are standardized to have unit variance. Subfigure (a) depicts the historical returns of the factors. The year on the x-axis indicates that the return is that of the first day of that year. Subfigures (b) and (c) depict the skewness and excess kurtosis of the factors, respectively, using 40 rolling windows of six months. The year on the x-axis indicates that the rolling window spans May to October of that year. The dotted black lines show the average skewness and excess kurtosis over all rolling windows, and are separated by 3 for skewness and 8 for excess kurtosis. Subfigure (d) depicts the nonlinear correlation between all possible pairs of FF factors (in blue) and ICs (in red), computed as in Section 5.2. The dotted lines show the average correlation over all pairs.