Optimal Portfolio Diversification via Independent Component Analysis

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

Victor DeMiguel
London Business School (UK), Management Science and Operations
avmiguel@london.edu

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

Abstract

A popular approach for enhancing diversification in portfolio selection is to rely on the *factor-risk-parity portfolio*, which is often defined as the portfolio whose return variance is spread equally among the principal components (PCs) of asset returns. Although PCs are useful for dimensionality reduction, they are arbitrary because any rotation of the PC basis yields an equally uncorrelated basis. This is problematic because we theoretically demonstrate that any portfolio is the factor-risk-parity portfolio corresponding to a specific uncorrelated basis. To overcome this problem, we rely on the factor-risk-parity portfolio based on the independent components (ICs), which are the rotation of the PCs that are maximally independent, and thus, account for higher-order moments. We propose a shrinkage portfolio that is obtained by combining the minimum-variance portfolio and the IC-risk-parity portfolio. We also show how to exploit the near independence of the ICs to parsimoniously estimate the factor-risk-parity portfolio with respect to Value-at-Risk. Finally, we empirically demonstrate that shrinkage portfolios based on the IC basis outperform those based on the PC basis, as well as the minimum-variance portfolio.

Keywords

Portfolio selection, risk parity, factor analysis, principal component analysis, higher-order moments.

1. Introduction

The mean-variance efficient portfolio of Markowitz (1952) is diversified in the sense that it has the lowest return variance among all portfolios with the same mean return. However, mean-variance portfolios estimated from historical data are usually not well diversified in terms of portfolio weights because of estimation error (Green and Hollifield 1992). Recently, researchers have thus proposed incorporating portfolio-weight diversification as an explicit objective of investment strategies. For instance, one can shrink the mean-variance portfolio toward the equally weighted portfolio, which is the most diversified portfolio in terms of weights (DeMiguel et al. 2009b, Tu and Zhou 2011). However, the equally weighted portfolio is not optimally diversified *in terms of risk* when some of the assets are riskier than others.

For this reason, investment managers often prefer to diversify their portfolios in terms of *asset risk contributions* rather than in terms of asset weights. This leads to the so-called *asset-risk-parity portfolio*, which is the portfolio whose risk is spread equally among the different assets; see, for instance, Maillard et al. (2010), Ji and Lejeune (2015), Bai et al. (2016) and Haugh et al. (2017). However, the asset-risk-parity portfolio is not optimally diversified in terms of risk unless the asset returns are uncorrelated.

To address this weakness of the asset-risk-parity portfolio, researchers have turned to the *factor-risk-parity portfolio*, which is the portfolio whose risk is equally spread among a set of *uncorrelated factors*; see, for instance, Meucci (2009), Deguest et al. (2013), Meucci et al. (2015) and Roncalli and Weisang (2016). To implement this portfolio, one needs to choose a specific basis of uncorrelated factors. The standard choice is to rely on the principal components of asset returns (the PC basis), provided by *principal component analysis* (PCA).

Although the PC basis is useful for dimensionality reduction (Campbell et al. 1997, Xu et al. 2016), it is problematic for factor-risk parity for three reasons. First, relying on the PC basis is arbitrary because once dimensionality has been reduced, any further rotation of the PC basis yields an equally uncorrelated basis, which, therefore, is equally good in terms of diversification. This is worrying because as we demonstrate, any portfolio is a factor-risk-parity portfolio corresponding to a specific uncorrelated basis. Second, the PC basis is not optimal because it is merely uncorrelated, rather than independent. Correlation may not be sufficient, in particular, when asset returns are not Gaussian. For instance, Poon et al. (2004, p.583) argue that "[correlation] assumes a linear relationship and a multivariate Gaussian distribution, which might lead to a significant underestimation of the risk from joint extreme events." Thus, just as relying on assets is suboptimal because they are correlated, relying on the PC factors is suboptimal because they may feature higher-order dependencies. Third, it is well-known that PC factors are not robust because they are defined as the orthogonal combinations of assets that maximize *variance*, which is sensitive to outliers. We empirically demonstrate that this sensitivity leads to poor out-of-sample performance and large turnover of the PC-risk-parity portfolio.

To overcome these three challenges, we propose using a factor-risk-parity portfolio based on *maximally independent* factors instead of merely *uncorrelated* ones. This is achieved via *independent component analysis* (ICA), an extension of PCA that identifies the rotation of the PC basis that results in factors that are as independent as possible (Hyvärinen et al. 2001). Independence is particularly relevant in portfolio selection because asset-return distributions often display fat tails and non-linear dependencies. For instance, Massacci (2017) shows that global equity markets feature tail connectedness, which increases during periods of turmoil. The IC basis overcomes the three aforementioned challenges related to PCA: (i) it discriminates among all the existing uncorrelated-factor bases, (ii) it enhances diversification by accounting for higher-order dependencies and (iii) it is more robust to outliers as shown by Martin-Clemente and Zarzoso (2017). In addition, we show that the IC basis allows one to parsimoniously deal with higher-order risk measures.

Our contribution is fourfold. First, we demonstrate that the arbitrariness of the PC basis is particularly worrying in the context of factor-risk parity. To do this, we first provide a closed-form expression for the factor-risk-parity portfolio corresponding to a generic uncorrelated-factor basis. We then use this closed-form expression to show that any portfolio is the factor-risk-parity portfolio corresponding to a specific factor basis. Thus, the decorrelation criterion is not sufficient to discriminate among all the possible portfolios, and the standard choice based on the PC basis does not seem justified because any rotation of the PC basis remains uncorrelated.

Second, to overcome the difficulties associated with the PC-risk-parity portfolio, we propose relying instead on the maximally independent factors provided by ICA. We theoretically demonstrate that provided at most one of the principal components is Gaussian, there is a unique IC-risk-parity portfolio. To go beyond strict risk parity, we propose a portfolio that combines variance minimization and factor-risk parity, using linear shrinkage as in Kan and Zhou (2007) and Tu and Zhou (2011). Specifically, we consider the portfolio that is obtained by shrinking the sample minimum-variance portfolio toward the IC-risk-parity portfolio, with the shrinkage intensity estimated via cross-validation; see Hastie et al. (2001). The variance of this shrinkage portfolio is spread more evenly among the independent components than that of the minimum-variance portfolio, and thus, it has lower factor-risk concentration and tail risk.

Third, we show how the IC basis can be used to estimate the portfolio that achieves factor-risk parity with respect to Value-at-Risk (VaR), which accounts for the higher-order moments of asset returns. The IC basis provides a parsimonious way to estimate higher-order moments of portfolio returns because although decorrelation suffices for decomposing the portfolio-return variance into a sum of factor variance exposures, independence is required to obtain a similar decomposition for higher-order moments. For the case where the IC factors are sufficiently independent, the higher-order co-moments of the IC factors are negligible, and this greatly reduces the number of parameters to be estimated. We exploit this fact to propose a shrinkage portfolio that combines the minimum-risk and IC-risk-parity portfolios based on the VaR risk measure.

Fourth, we evaluate the out-of-sample performance of shrinkage portfolios combining minimum-risk and factor-risk-parity portfolios. We find that portfolios based on the IC basis substantially outperform those based on the PC basis in terms of portfolio mean return, Sharpe ratio, kurtosis and VaR. In addition, IC-based portfolios are more stable than PC-based ones as demonstrated by their lower turnover. The IC-based shrinkage portfolios also outperform the corresponding minimum-risk portfolios obtained without shrinkage.

This work is connected to three streams of literature. First, this work contributes to the literature on risk-parity portfolios. Maillard et al. (2010) showed that the *long-only* asset-risk-parity portfolio is unique and its volatility is in between that of the minimum-variance and equally weighted portfolios. Bai et al. (2016) showed that *long-short* asset-risk-parity portfolios are no longer unique. Two recent remarkable contributions are those in Haugh et al. (2017) and Ji and Lejeune (2015). Haugh et al. (2017) propose optimizing the investor's mean-risk utility subject to a risk-budgeting constraint, in which pre-specified budgets of risk are allocated to *sub-portfolios* of assets, rather than to individual assets. As we do, Haugh et al. (2017) consider both variance and VaR risk measures, but for tractability the authors rely on the Gaussian VaR approximation, while we exploit the properties of independent components to incorporate skewness and kurtosis as well. Ji and Lejeune (2015) assign downside-risk budgets not only to sub-portfolios of assets but also to individual assets, and the authors account for estimation risk via stochastic programming.

The main difference between our work and these papers is that we perform risk parity with respect to *factors*, not assets. Factor-risk parity was introduced by Meucci (2009), and its theoretical properties were explored by Deguest et al. (2013); both studies relied on the PC basis. Our work extends these two studies by allowing for dimensionality reduction, emphasizing the importance of the selection of the factor basis, proposing an alternative to the PC basis based on ICA and considering higher-order risk measures. An alternative to the PC basis was also proposed by Meucci et al. (2015) who derive the uncorrelated factors that track another (possibly correlated) set of factors, such as the original assets. This factor basis benefits from improved economic intuition but similar to PC basis, may still be far from independent. Finally, Roncalli and Weisang (2016) propose a risk-budgeting approach based on factors instead of assets. Our paper differs from these papers mainly because we use maximally independent factors, which account for higher-order moments.

Second, our work is related to the literature on portfolio selection under parameter uncertainty. Several approaches have been proposed to reduce estimation error in the moments of asset returns: Bayesian approaches (Jorion 1986), shrinkage covariance matrix estimators (Ledoit and Wolf 2003), robust optimization (Goldfarb and Iyengar 2003), bias adjustment of moment estimators (Siegel and Woodgate 2007), robust estimation (DeMiguel and Nogales 2009) and machine learning (Ban et al. 2018). Other approaches aim at reducing estimation error in the portfolio weights directly, including shortselling constraints (Jagannathan and Ma 2003), portfolio-norm constraints (DeMiguel et al. 2009a), cardinality constraints (Bertsimas and Shioda 2009, Gao and Li 2013), portfolio-turnover constraints (Olivares-Nadal and DeMiguel 2018) and most closely related to our work, shrinkage estimators of portfolio weights (Kan and Zhou 2007, DeMiguel et al. 2009b, Tu and Zhou 2011 and DeMiguel et al. 2013). We introduce shrinkage portfolios that shrink the sample minimum-risk portfolio toward the IC-risk-parity portfolio, using variance and VaR as risk measures. In contrast to DeMiguel et al. (2009b) and Tu and Zhou (2011) who shrink the sample mean-variance portfolio toward the naively diversified equally weighted portfolio, we shrink the sample minimum-variance portfolio toward a portfolio with equally weighted IC-factor exposures.

Third, our work is related to the recent literature that applies independent components to factor analysis of financial time series; see, for instance, Chen et al. (2010), García-Ferrer et al. (2012) and Fabozzi et al. (2016). To the best of our knowledge, the only other paper in which ICA is applied in a portfolio selection context is Hitaj et al. (2015), which proposes maximizing the expected CARA utility of an investor, estimated parsimoniously by assuming the IC factors are independent. Our work is fundamentally different because we focus on the use of IC factors to define *factor-risk-parity portfolios* that are optimally diversified.

The remainder of the article is organized as follows. Section 2 reviews the theory behind PCA and ICA. Section 3 shows that factor decorrelation is not a discriminant criterion for factor-risk parity. Section 4 introduces our approach, which consists of combining risk minimization and IC-risk parity. Section 5 extends the IC-risk parity approach to the case with VaR as risk measure. Section 6 evaluates the out-of-sample empirical performance of the proposed shrinkage portfolios. Section 7 concludes. Proofs for all propositions are available in the Appendix. Table 1 lists the main mathematical symbols used in the manuscript.

2. PCA versus ICA: from decorrelation to independence

2.1. Principal component analysis

Let $X = (X_1, \ldots, X_N)' \in \mathbb{R}^N$ be a random asset-return vector with stationary distribution and nonsingular covariance matrix $\Sigma_X \in \mathbb{R}^{N\times N}$. The covariance matrix can be diagonalized as $\Sigma_X = V\Lambda V'$, where the columns of V are the eigenvectors, and $\Lambda := \text{diag}(\lambda_1, \ldots, \lambda_N)$ is the diagonal matrix that contains the (strictly positive) eigenvalues. This factorization is at the heart of *principal component analysis* (PCA); see Section 6.4 of Campbell et al. (1997).

Definition 1 (PC basis)**.** The *PC basis* $Y^\star$ is given by the orthogonal linear transformation of X,

$$Y^\star := \Lambda^{-1/2} V' X. \tag{1}$$

Note that the PC basis $Y^\star$ is the basis formed by the (standardized) *principal components* of asset returns. To alleviate the impact of estimation error, we keep only the first $K \leq N$ principal components. For simplicity, we use the same notation as in (1), where $Y^\star \in \mathbb{R}^K$, $V \in \mathbb{R}^{N\times K}$ and $\Lambda := \text{diag}(\lambda_1, \ldots, \lambda_K) \in \mathbb{R}^{K\times K}$.

It is clear from Definition 1 that the PC basis $Y^\star$ is a *white basis*; that is, its covariance matrix is equal to the identity matrix, $\Sigma_{Y^\star} = I_K$. However, whiteness being invariant under rotation, the PC basis is merely one white basis out of a continuum of many others. Indeed, let $SO(K)$ be the *special orthonormal group* of order K; that is, the set of rotation matrices $R \in \mathbb{R}^{K\times K}$ such that $RR' = I_K$ and $\det(R) = 1$. Then, for any $R \in SO(K)$ we have that $Y = RY^\star$ is a white basis.

Thus, when it comes to selecting an uncorrelated basis of K factors, there is no reason to choose the PC factors over any of their rotations. PC factors are useful for dimensionality reduction (Campbell et al. 1997, Xu et al. 2016) but once dimensionality has been reduced, the basis formed by the first K principal components is merely one orthogonal basis out of a continuum of others. This is problematic in our context because as we show in Section 3.2, any portfolio is the factor-risk-parity portfolio corresponding to a specific rotation of the PC basis.

2.2. Independent component analysis

To resolve the arbitrariness inherent in the decorrelation criterion and discriminate among all the white bases, one can look for the white basis that is as *independent* as possible.

2.2.1. Definition of the IC basis

Independence is a much stronger requirement than decorrelation; it corresponds to asking decorrelation for every (non-linear) transformation. In particular, independence requires no dependence with respect to higher-order moments. It is possible to measure the distance to independence of a particular random vector by relying on mutual information.

Definition 2 (Mutual information)**.** The *mutual information* of a random vector Y is the Kullback-Leibler divergence between the joint density and the corresponding product density of Y:

$$I(Y) := \left\langle f_Y, \prod_{i=1}^{K} f_{Y_i} \right\rangle = \mathbb{E}\left(\ln \frac{f_Y(Y)}{\prod_{i=1}^{K} f_{Y_i}(Y_i)} \right). \tag{2}$$

Mutual information is always nonnegative, and it is zero if and only if the components of Y are mutually independent; see Cover and Thomas (2006, p.253). This criterion can be standardized to take values in $[0, 1]$ by defining the relative mutual information

$$\bar{I}(Y) = \frac{I(Y)}{\frac{1}{K}\sum_{i=1}^{K} h(Y_i)}, \tag{3}$$

where

$$h(Y_i) := h[f_{Y_i}] := -\,\mathbb{E}(\ln f_{Y_i}(Y_i)) \tag{4}$$

is the entropy of Y_i. Thus, one can extend the notion of *white basis* to that of *IC basis* by using mutual information instead of linear correlation as dependence measure.

Definition 3 (IC basis)**.** Let X be a random asset-return vector with PC basis $Y^\star$. The *IC basis* $Y^\dagger \in \mathbb{R}^K$ is the rotation of the PC basis that minimizes mutual information:

$$Y^\dagger := R^\dagger Y^\star \ , \ R^\dagger \in \left\{ \underset{R \in SO(K)}{\operatorname{argmin}} \ I(RY^\star) \right\}. \tag{5}$$

A few comments are in order. First, $I(R^\dagger Y^*)$ may be strictly positive in general, and thus, the IC factors $Y^\dagger$ may not be perfectly independent. Second, the rotation matrix $R^\dagger$ is not unique because mutual information is invariant to a change of sign and permutation of the factors. As we show in Section 3, this is a mild indeterminacy that has no impact for factor-risk parity. Third, it follows from Comon (1994) that the independent components are unique (up to a change of sign and permutation) provided at most one of the principal components is Gaussian.[1] The reason for this is that independence and decorrelation are equivalent for Gaussian random variables.

2.2.2. Computing the IC basis

The IC basis can be computed using *independent component analysis* (ICA), a well-known technique in machine learning; see Hyvärinen et al. (2001) and Vrins (2007). Although the PC basis is known in closed form via the PCA decomposition in (1), there is no similar expression for the matrix $R^\dagger$. Thus, ICA uses adaptive learning techniques, such as gradient descent, to recover the factors.

The computation of the IC basis from (5) may look like a daunting problem because $I(Y)$ contains the *joint* K-dimensional density of Y. However, the problem simplifies greatly in our context because one can show that the objective function collapses to the sum of individual entropies, which features only K one-dimensional densities, but not the joint density; see, for instance, Vrins and Verleysen (2005). This is a critical result in the context of ICA; thus we reiterate it below.

Proposition 1. *The IC basis in* (5) *can be equivalently defined as*

$$Y^\dagger := R^\dagger Y^\star \ , \ R^\dagger \in \left\{ \underset{R \in SO(K)}{\operatorname{argmin}} \sum_{i=1}^{K} h(R_i Y^\star) \right\}, \tag{6}$$

where R_i is the ith row of R, and $h(\cdot)$ is the entropy defined in (4)*.*

To implement the formulation in (6), we follow Hyvärinen's (1999) *FastICA* algorithm. In this algorithm, the individual entropies $h(R_i Y^\star)$ are estimated using the Taylor expansion

$$h(Y) \approx h\big[\mathcal{N}(0,1)\big] - \frac{1}{2}\mathbb{E}(G(Y))^2, \tag{7}$$

where $h\big[\mathcal{N}(0,1)\big]$ is the entropy of a standard Gaussian random variable, and G is any smooth non-quadratic even function. *FastICA* then computes the matrix $R^\dagger$ using a gradient-descent scheme that shows at least quadratic local convergence. The function G controls the robustness of the estimation—an important concern for out-of-sample purposes—given that the influence function of each row $R_i^\dagger$ grows as a function of $xG'(x)$; see Hyvärinen (1999). Thus, we follow Hyvärinen's recommendation and set $G(x) = \ln\cosh(x)$, a choice that has proved robust and well-performing

[1] Rigorously speaking, there may be more than one IC basis in the unlikely case that there are two global minimizers to the minimum mutual-information problem that are not equal up to a change of sign and permutation.

for most practical purposes. The resulting influence function increases only with $x \tanh x \approx |x|$. This close relationship with the L1 norm implies that ICA retrieves factors that are more robust to outliers than those found by PCA, which finds each orthogonal direction by maximizing an L2-norm; see Martin-Clemente and Zarzoso (2017).

2.3. Numerical illustration

Examples 1 and 2 below illustrate the arbitrariness of the PC basis, and in general of the decorrelation criterion, to select the basis of factors. They also illustrate the advantages of the IC basis over any other white basis. We focus on the case $K = N = 2$ case for graphical illustration purposes. We parameterize the rotation matrix $R \in SO(2)$ as a function of the rotation angle θ:

$$R = R(\theta) := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in SO(2). \tag{8}$$

Example 1 (Independent non-Gaussian asset returns)**.** Consider two assets with returns independently distributed as Student's t distributions $X_1 \sim T(\nu_1 = 3)$ and $X_2 \sim T(\nu_2 = 20)$. Because the two asset returns are uncorrelated and X_1 has a higher variance than X_2, we have that the PC basis coincides with the assets; that is, $Y^\star = X$. Figure 1a illustrates how the relative mutual information $\bar{I}(Y)$ and correlation $\mathrm{Corr}(Y)$ of the factors $Y = R(\theta)Y^\star$ obtained by rotating the principal components depend on the rotation angle θ. Note that while the correlation is zero for all θ, the mutual information is zero only when $\theta = k\pi/2$, $k \in \mathbb{Z}$; that is, the IC basis $Y^\dagger$ coincides with the asset returns up to a change of sign and permutation. Thus, while the decorrelation criterion is not discriminating, the independence criterion results in a unique factor basis.

Example 2 (Dependent non-Gaussian asset returns)**.** Now consider the case where the joint distribution of two asset returns is obtained by mixing, via a Gaussian copula with correlation $\rho = 0.2$, two random variables independently distributed as Student's t distributions $X_1 \sim T(\nu_1 = 3)$ and $X_2 \sim T(\nu_2 = 20)$. The PC basis $Y^\star$ is obtained by rotating the two asset returns by an angle of -0.17 radians. Figure 1b depicts the relative mutual information $\bar{I}(Y)$ and correlation $\mathrm{Corr}(Y)$ for rotations of the principal components $Y = R(\theta)Y^\star$. Independence is never achieved because the asset returns are nonlinear functions of random variables with non-Gaussian distribution. Nevertheless, the mutual information varies with the rotation angle θ and this allows one to discriminate between the continuum of white bases of factors. In particular, we find that the IC basis is retrieved for $\theta^\dagger = 0.35 + k\pi/2 \neq 0$, and thus, the IC basis differs from the PC basis.

Finally, to gauge the practical relevance of the independent components, we use three empirical datasets to evaluate the gain in independence that can be achieved by moving from the PC basis to the IC basis. In particular, we consider the three equity datasets from Ken French's library that we use in Section 6: six size and book-to-market portfolios (*6BTM*), six size and operating-profitability portfolios (*6Prof*) and 30 U.S. industry portfolios (*30Ind*). Figure 2 shows that the gain in independence is particularly substantial around the 2007-2008 financial crisis, a period characterized by pronounced tail risk and extreme-value dependence of asset returns; that is, by pronounced non-Gaussianity.

3. Factor-risk parity via uncorrelated factors

Let $P := w'X$ be the portfolio return, where $w = (w_1, \ldots, w_N)'$ is the vector of portfolio weights, which we assume belongs to the set

$$\mathcal{W} := \{w \in \mathbb{R}^N \mid \mathbf{1}_N' w = 1\}, \tag{9}$$

where $\mathbf{1}_N$ is the N-dimensional vector of ones. The factor-risk-parity portfolio is the portfolio for which each uncorrelated factor contributes equally to its return risk. The standard approach in risk parity is to measure risk via the variance, which is considered in this section and the next one. In Section 5, the case of higher-order risk measures is tackled. All portfolios below are computed from sample moment estimators obtained from a sample of X with T observations. In particular, the PC basis is obtained via the eigendecomposition of the sample covariance matrix, $\widehat{\Sigma}_X$.

3.1. General formulation of the factor-risk-parity portfolio

Meucci (2009) relied on the PC basis to derive the factor-risk-parity portfolio. However, as shown in Section 2.1, this is merely one possible choice out of a continuum of white bases. Thus, we consider below any rotation of the PC basis, $Y = RY^\star$. Meucci (2009) also restricted to the full-dimensional case, $K = N$. However, limiting the number of factors helps reduce problem dimensionality and alleviate the impact of estimation error. Moreover, Kozak et al. (2018) show that a small number of principal components explain the cross section of stock returns out of sample. Thus, we consider the case with $K \leq N$ factors, as in Roncalli and Weisang (2016).

The portfolio return $P = w'X$ can be projected on the factor basis Y as $\tilde{w}(R)'Y$, where

$$\tilde{w}(R) := R\Lambda^{1/2}V'w \tag{10}$$

are the factor exposures. The portfolio weights can be retrieved from the factor exposures $\tilde{w}(R)$ as

$$w = (R\Lambda^{1/2}V')^{+}\tilde{w}(R) = V\Lambda^{-1/2}R'\tilde{w}(R), \tag{11}$$

where A^{+} is the pseudo-inverse of A. The projected portfolio-return variance is given by

$$\tilde{w}(R)'\Sigma_Y\tilde{w}(R) = \sum_{i=1}^{K}\tilde{w}_i(R)^2 = \|\tilde{w}(R)\|^2, \tag{12}$$

so that the relative variance contribution of each component Y_i of Y is computed as

$$p_i(R) := \frac{\tilde{w}_i(R)^2}{\|\tilde{w}(R)\|^2}. \tag{13}$$

Denoting $p(R) := (p_1(R)\ldots, p_K(R))'$, we can define the factor-risk-parity portfolio as follows.

Definition 4 (Factor-risk-parity portfolio)**.** The *factor-risk-parity portfolio* associated with the factor basis $Y = RY^{\star}$ is the solution to the optimization problem

$$w^{\star}(R) := \left\{\underset{w\in\mathcal{W}}{\operatorname{argmax}}\,\mathcal{H}[p(R)]\right\}, \tag{14}$$

where $\mathcal{H}[p]$ is the *inverse Herfindahl (diversification) index* of a probability mass function p:

$$\mathcal{H}[p] := \|p\|^{-2} \in [1, K]. \tag{15}$$

The diversification index $\mathcal{H}[p]$ attains a maximum of K when p is uniform; that is, $p_i = 1/K$ for all i, and a minimum of one when $p_i = 1$ for some i. Thus, maximizing $\mathcal{H}[p(R)]$ yields the portfolio for which the factors $Y = RY^{\star}$ contribute the most homogeneously to the portfolio-return variance.

The factor-risk-parity portfolio associated with the first K PC factors is obtained for $R = I_K$.

Definition 5 (PC-risk-parity portfolio)**.** The *PC-risk-parity portfolio* is $w^{\star}(I_K)$.

The analytical expression for the factor-risk-parity portfolio is given in the next proposition.

Proposition 2. *There are 2^{K-1} maximizers to* (14) *for which $\mathcal{H}[p(R)] = K$, given by*

$$w^{\star}(R) = \frac{V\Lambda^{-1/2}R'\mathbf{1}_K^{\pm}}{\mathbf{1}_N'V\Lambda^{-1/2}R'\mathbf{1}_K^{\pm}}, \tag{16}$$

where $\mathbf{1}_K^{\pm}$ is any K-dimensional column vector with ± 1 on each of its entries. Moreover, the set of maximizers $w^{\star}(R)$ is invariant to a change of sign and permutation of the factors $Y = RY^{\star}$.

To resolve this multiplicity of factor-risk-parity portfolios, we show in Section 4.2 that there is a unique factor-risk-parity portfolio that minimizes the portfolio-return variance.

3.2. Arbitrariness of the decorrelation criterion for factor-risk parity

As discussed in Section 2, the PC basis is merely one arbitrarily chosen factor basis: any further rotation of the PC basis yields another white basis, which therefore is equally good in terms of diversification. This arbitrariness is worrying because, as clearly seen from (16), the set of factor-risk-parity portfolios $w^\star(R)$ varies with the rotation R chosen. Even further, we show in the next proposition that *any* portfolio $w \in \mathcal{W}$ is the factor-risk-parity portfolio corresponding to a specific uncorrelated-factor basis.

Proposition 3. *Given any portfolio $w \in \mathcal{W}$, there exists a basis $Y = RY^\star$ for which $\mathcal{H}[p(R)] = K$.*

Note that this result is not due to the multiplicity of solutions in (16): it holds for any choice of vector $\mathbf{1}_K^{\pm}$. In other words, the set of all possible factor-risk-parity portfolios associated with all matrices $R \in SO(K)$ spans the entire space of portfolios $\mathcal{W} \subset \mathbb{R}^N$. Thus, although sufficient to decompose the portfolio-return variance as a sum of individual factor contributions in (12), decorrelation is not sufficient to discriminate among all the possible portfolios. To circumvent this issue, one could choose the PC basis, corresponding to $R = I_K$. Although useful for dimensionality reduction, the PC basis is not better than any of its rotations in terms of diversification.

The example below illustrates the result in Proposition 3 for the bivariate case $K = N = 2$.

Example 3 (Arbitrariness of the factor-risk-parity portfolio)**.** We generate $T = 1000$ observations from the asset-return distribution considered in Example 2. The PC basis $Y^\star$ is obtained by rotating the two asset returns by an angle of -0.14 radians. Proposition 2 shows that there are two different factor-risk-parity portfolios for each white basis in the case with $K = 2$ factors. Figure 3a depicts the weight on the first asset of the two factor-risk-parity portfolios as a function of the rotation angle θ that generates the white basis $Y = R(\theta)Y^\star$. Note that in line with Proposition 3, the weight on the first asset of the two factor-risk-parity portfolios can take any real value. However, if for each θ we choose among the two portfolios the one associated with the lowest portfolio-return variance, then the optimal weight on the first asset is no longer unbounded as a function of θ. This is illustrated in Figure 3b. Nonetheless, the solution remains arbitrary with respect to θ: *any* portfolio allocating a weight to the first asset between -0.25 and 0.7 is a factor-risk-parity portfolio with minimum return variance.

4. Factor-risk parity via independent component analysis

The previous section showed that decorrelation is not a discriminant criterion for factor-risk parity. In contrast, looking for the maximally independent white basis, given by the IC basis, resolves this indeterminacy and improves diversification. Indeed, at the extreme where the IC basis is perfectly independent, allocating equal exposures to each IC factor yields a policy that not only diversifies the portfolio-return variance but that is also diversified with respect to higher-order factor co-moments.

4.1. IC-risk-parity portfolio

The IC basis is a particular white basis corresponding to the rotation matrix $R^{\dagger}$ in (5) such that the K components of $Y^{\dagger} = R^{\dagger} Y^{\star}$ are as independent as possible. Thus, the IC-risk-parity portfolio is a particular factor-risk-parity portfolio.

Definition 6 (IC-risk-parity portfolio)**.** The *IC-risk-parity portfolio* is given by $w^{\star}(R^{\dagger})$ in (16).

From Proposition 2, the set of IC-risk-parity portfolios is unaffected by the change-of-sign-and-permutation indeterminacy of the IC basis. Thus, if at most one of the principal components is Gaussian, then this set of portfolios is unique. The arbitrariness inherent in the decorrelation criterion is fully addressed.

Example 3 (continued)**.** The IC basis is retrieved for the rotation angle $\theta^{\dagger} = 0.18 + k\pi/2$, $k \in \mathbb{Z}$, which gives two IC-risk-parity portfolios with weights on the first asset $w_1^{\star}(R(\theta^{\dagger})) = \{0.38, -0.40\}$. Figure 3b depicts the portfolio with $w_1 = 0.38$, which is preferred because it has a lower return variance. Moreover, because the IC basis does not depend on the rotation angle θ, this weight is constant and there is no arbitrariness left in the IC-risk-parity portfolio.

4.2. Combining risk minimization and factor-risk parity

From Proposition 2, there are 2^{K-1} solutions to the IC-risk-parity portfolio problem. As stressed in Example 3, notice that these solutions may lead to portfolios with different variances. However, investors are concerned not only with risk diversification but also with controlling the overall portfolio risk. Indeed, the minimum-variance portfolio has been widely studied among researchers and shown to exhibit favorable out-of-sample performance; see, for instance, Jagannathan and Ma (2003), DeMiguel et al. (2009a) and DeMiguel and Nogales (2009).

Therefore, we propose using a shrinkage portfolio obtained by combining the minimum-variance portfolio and the IC-risk-parity portfolio. Haugh et al. (2017) propose optimizing the investor's mean-risk utility subject to a strict asset-risk-budgeting constraint. Our approach differs from theirs in that we focus on factors, and we do not impose strict risk parity.

When looking only at minimizing risk, we obtain the sample minimum-variance portfolio

$$w^{\star}_{MV} := \frac{\widehat{\Sigma}_X^{-1}\mathbf{1}_N}{\mathbf{1}'_N\widehat{\Sigma}_X^{-1}\mathbf{1}_N}, \tag{17}$$

where $\widehat{\Sigma}_X$ is the sample covariance matrix of asset returns. When factor-risk parity is imposed, the next proposition gives an analytical expression for the portfolio with least return variance out of the 2^{K-1} solutions to the IC-risk-parity portfolio problem, resolving the issue of multiple solutions.[2]

Proposition 4. *The portfolio with minimum return variance out of the 2^{K-1} solutions to the IC-risk-parity portfolio problem is unique and is given by*

$$w^{\star}_{IC} := \frac{V\Lambda^{-1/2}(R^{\dagger})'\mathbf{1}_K^{MV}}{\mathbf{1}'_N V\Lambda^{-1/2}(R^{\dagger})'\mathbf{1}_K^{MV}}, \tag{18}$$

where $\mathbf{1}_K^{MV} := \mathrm{sign}(R^{\dagger}V'\mathbf{1}_N)$.

One possible alternative to combining the two portfolios in (17) and (18) is to follow Deguest et al. (2013), who, restricting to the PC basis and the full-dimensional case $K = N$, propose minimizing the portfolio-return variance subject to a factor-risk diversification constraint. However, one drawback of this approach is that this portfolio cannot be characterized in closed form. Therefore, we consider the shrinkage portfolio that combines the sample minimum-variance portfolio $w^{\star}_{MV}$ and the IC-risk-parity portfolio $w^{\star}_{IC}$, which can be characterized in closed form.

Definition 7 (ICMV portfolio)**.** The *ICMV portfolio* is defined as the shrinkage portfolio

$$w^{\star}_{ICMV} := (1-\delta)w^{\star}_{MV} + \delta w^{\star}_{IC}, \tag{19}$$

where $\delta \in [0, 1]$ is a given shrinkage intensity.

Shrinkage portfolios as in (19) have been considered by several authors in the literature. For instance, Kan and Zhou (2007) shrink the sample mean-variance portfolio toward the sample minimum-variance portfolio. Perhaps more closely related to our work are DeMiguel et al. (2009b)

[2] We recover the result of Deguest et al. (2013) for the PC basis and $K = N$ as a special case.

and Tu and Zhou (2011), who shrink the sample minimum- and mean-variance portfolio, respectively, toward the *equally weighted portfolio*. A distinguishing feature of our approach is that we shrink toward a target portfolio for which the *IC-factor exposures are equally weighted*.

If one is willing to assume the Gaussianity of asset returns, one can follow DeMiguel et al. (2013) and compute an analytical estimate of the optimal shrinkage intensity δ according to various criteria, such as variance or quadratic utility. However, in the context of ICA, the Gaussianity assumption is not palatable. Therefore, as detailed in Section 6.1, we propose instead calibrating δ non-parametrically via cross-validation, avoiding assumptions on the distribution of asset returns.

Finally, although in this section we focused on the IC basis, the above shrinkage approach works for any white basis $Y = RY^\star$. In particular, in Section 6, we compare the ICMV portfolio to the PCMV portfolio $w^\star_{PCMV} := (1-\delta)w^\star_{MV} + \delta w^\star_{PC}$, where $w^\star_{PC}$ is found by setting $R^\dagger = I_K$ in (18).

5. Factor-risk parity under higher-order risk measures

In Sections 3 and 4, we showed how to apply independent components to the traditional approach to risk parity, which diversifies the *variance* of the portfolio return across a set of underlying factors. In this section, we show how to apply independent components to diversify the *Value-at-Risk* (VaR) of the portfolio return, which accounts for higher-order moments.

Haugh et al. (2017) consider an asset-risk-parity portfolio under the VaR risk measure, but for tractability, they rely on the Gaussian VaR approximation. In contrast, we propose diversifying the portfolio-return VaR with respect to factors, and we exploit the properties of independent components to efficiently incorporate skewness and kurtosis.

5.1. Estimating higher-order moments with independent components

We now show that the IC basis provides a parsimonious approach for estimating portfolio-return higher-order moments. To do this, we introduce some notation. Given a random variable Z, we denote the following sample estimators: $\mu(Z)$ the mean, $m_k(Z) := \mathbb{E}((Z-\mu(Z))^k)$ the kth central moment, $s(Z) := m_3(Z)/m_2(Z)^{3/2}$ the skewness and $\kappa(Z) := m_4(Z)/m_2(Z)^2-3$ the excess kurtosis.

Given a white basis of K factors, $Y = RY^\star$, the third and fourth central moments of the projected portfolio return, $P = \tilde{w}(R)'Y$, are

$$m_3(P) = \tilde{w}(R)'\Phi_Y(\tilde{w}(R)\otimes\tilde{w}(R)), \tag{20}$$

$$m_4(P) = \tilde{w}(R)'\Psi_Y(\tilde{w}(R)\otimes\tilde{w}(R)\otimes\tilde{w}(R)), \tag{21}$$

where $\otimes$ is the Kronecker product, and $\Phi_Y \in \mathbb{R}^{K \times K^2}$ and $\Psi_Y \in \mathbb{R}^{K \times K^3}$ are the sample co-skewness and co-kurtosis matrices of Y, respectively. It is easy to show (Boudt et al. 2015) that estimating these two matrices requires estimating a large number of parameters:

$$\frac{K(K+1)(K+2)(K+7)}{24}.$$

To develop a more parsimonious estimation approach, we rely on the IC basis $Y^\dagger = R^\dagger Y^\star$ and assume that the IC factors are perfectly independent. In this case, all IC co-moments vanish, and thus, estimating the matrices $\Phi_{Y^\dagger}$ and $\Psi_{Y^\dagger}$ requires estimating only $2K$ parameters, which correspond to the third and fourth central moments of the K independent components.

Proposition 5. *Let the independent components be perfectly independent and have finite fourth central moments, $Y^\dagger \in \mathcal{L}^4$. Then, the third and fourth central moments of the projected portfolio return, $P = \tilde{w}(R^\dagger)'Y^\dagger$, are given by*

$$m_3(P) = \sum_{i=1}^{K} \tilde{w}_i(R^\dagger)^3 m_3(Y_i^\dagger), \tag{22}$$

$$m_4(P) = \sum_{i=1}^{K} \tilde{w}_i(R^\dagger)^4 m_4(Y_i^\dagger) + 3\sum_{i=1}^{K}\sum_{j\neq i}(\tilde{w}_i(R^\dagger)\tilde{w}_j(R^\dagger))^2, \tag{23}$$

where $m_3(Y_i^\dagger)$, $m_4(Y_i^\dagger)$ are the third and fourth central moments of the ith independent component.

The proposition above shows that the third and fourth central moments are affine with respect to the independent components' corresponding moments, implying that the IC basis provides a parsimonious approximation of the portfolio-return higher-order moments.

5.2. Factor-risk parity with Value-at-Risk

Proposition 5 allows one to develop parsimonious estimators of factor-risk-parity portfolios with respect to higher-order risk measures. We show how to do this for the popular VaR risk measure, which is a difficult problem because contrary to the central moments in (22) and (23), the quantile of a linear combination of independent random variables is not an affine function of the individual quantiles. To circumvent this difficulty, we employ the Favre and Galeano's (2002) modified VaR (MVaR), which can be conveniently decomposed into factor risk contributions. Boudt et al. (2008) provide closed-form expressions for the risk contributions of *assets* to the portfolio-return MVaR. We extend their results to the case of risk contributions of *factors* and in particular, of IC factors.

Definition 8 (Modified VaR)**.** The *modified VaR* (MVaR) of the random variable $Z \in \mathcal{L}^4$ is the second-order Cornish-Fisher expansion of the quantile function of Z:

$$\mathrm{MVaR}_Z(\alpha) := -\mu(Z) - \sqrt{m_2(Z)}\, q, \tag{24}$$

where

$$q := q_\alpha + \frac{1}{6}(q_\alpha^2 - 1)s(Z) + \frac{1}{24}(q_\alpha^3 - 3q_\alpha)\kappa(Z) - \frac{1}{36}(2q_\alpha^3 - 5q_\alpha)s(Z)^2 \tag{25}$$

and q_α is the α-quantile of a standard Gaussian random variable.

MVaR is an approximation of VaR because the Cornish-Fisher expansion is truncated at the second-order term, but there is evidence that incorporating additional terms barely improves the approximation (Baillie and Bollerslev 1992), while it increases the number of moments to estimate.

Because MVaR is a positive homogeneous risk measure, one can decompose the portfolio-return MVaR into IC-factor risk contributions. As shown in Proposition 5, we can do this parsimoniously by assuming the IC factors are perfectly independent.

Proposition 6. *Let $Y^\dagger \in \mathcal{L}^4$ be the basis of independent components and assume they are perfectly independent. Then, the risk contribution of $Y_i^\dagger$ to $\mathrm{MVaR}_P(\alpha)$, $P = \tilde{w}(R^\dagger)'Y^\dagger$, is given by*

$$\begin{aligned}\mathrm{MVaR}_i(\alpha) := \tilde{w}_i(R^\dagger) \times \Bigg[-\mu(Y_i^\dagger) - \frac{\partial_i m_2(P)}{2\sqrt{m_2(P)}} q + \sqrt{m_2(P)} \Bigg(-\frac{1}{6}(q_\alpha^2 - 1)\partial_i s(P) \\ - \frac{1}{24}(q_\alpha^3 - 3q_\alpha)\partial_i \kappa(P) + \frac{1}{18}(2q_\alpha^3 - 5q_\alpha)s(P)\partial_i s(P) \Bigg) \Bigg],\end{aligned} \tag{26}$$

where

$$\partial_i s(P) = \frac{2m_2(P)^{3/2}\partial_i m_3(P) - 3m_3(P)\sqrt{m_2(P)}\partial_i m_2(P)}{2m_2(P)^3}, \tag{27}$$

$$\partial_i \kappa(P) = \frac{m_2(P)\partial_i m_4(P) - 2m_4(P)\partial_i m_2(P)}{m_2(P)^3}, \tag{28}$$

and $m_2(P)$, $m_3(P)$, and $m_4(P)$ are given by (12), (22), *and* (23), *respectively.*[3]

Although, in general, the independent components are not *perfectly* independent, their higher-order co-moments are negligible if they are sufficiently independent, and thus, Proposition 6 provides a convenient approximation of the MVaR contributions. Equipped with this approximation, we define the IC-risk-parity portfolios with respect to MVaR.

[3] We assume the confidence level α is low enough that the MVaR contributions, $\mathrm{MVaR}_i(\alpha)$, are all positive. This assumption typically holds for $\alpha = 5\%$ or below.

Definition 9 (IC-risk-parity portfolios with respect to MVaR)**.** Let $Y^\dagger \in \mathcal{L}^4$ be the IC basis. Then, the *IC-risk-parity portfolios with respect to MVaR* are the portfolios with $\mathcal{H}[p(\alpha)] = K$, where $p(\alpha)$ is the vector made of the K relative MVaR contributions of IC factors:

$$p_i(\alpha) := \frac{\mathrm{MVaR}_i(\alpha)}{\sum_{j=1}^{K} \mathrm{MVaR}_j(\alpha)}. \tag{29}$$

Note that although there may be multiple IC-risk-parity portfolios with respect to MVaR, similar to the variance case, we can resolve this indeterminacy by selecting the IC-risk-parity portfolio that minimizes MVaR, which we denote as $w^\star_{IC}(\alpha)$. This portfolio can be estimated numerically by solving the following optimization problem:

$$\hat{w}^\star_{IC}(\alpha) := \operatorname*{argmin}_{w \in \mathcal{W}} \mathrm{MVaR}_P(\alpha) \quad \text{subject to} \quad \mathcal{H}[p(\alpha)] \geq K - \epsilon, \tag{30}$$

where ϵ is a tolerance parameter set to $\epsilon = 5 \times 10^{-4}$. We now define the ICMVaR portfolio as the shrinkage portfolio that combines the sample minimum-MVaR portfolio, $w^\star_{MMVaR}(\alpha)$, and $\hat{w}^\star_{IC}(\alpha)$.[4]

Definition 10 (ICMVaR portfolio)**.** The *ICMVaR portfolio* is the shrinkage portfolio

$$w^\star_{ICMVaR}(\alpha) := (1-\delta) w^\star_{MMVaR}(\alpha) + \delta \hat{w}^\star_{IC}(\alpha), \tag{31}$$

where $\delta \in [0, 1]$ is a given shrinkage intensity, $w^\star_{MMVaR}(\alpha)$ is the sample minimum-MVaR portfolio and $\hat{w}^\star_{IC}(\alpha)$ is the estimator of the IC-risk-parity portfolio that minimizes MVaR, given in (30).

6. Out-of-sample performance

We now evaluate the out-of-sample performance of the proposed shrinkage portfolios. In Section 6.1, we discuss the empirical data and methodology employed, and in Section 6.2, we discuss the results.

6.1. Data and methodology

We consider three equity datasets from Ken French's library: six size and book-to-market portfolios (*6BTM*), six size and operating-profitability portfolios (*6Prof*) and 30 U.S. industry portfolios (*30Ind*). We download value-weighted daily portfolio returns from 1978 to 2017. We use daily return data because they produce better estimates of risk than weekly or monthly return data; see Jagannathan and Ma (2003). Daily data also improve the estimation accuracy for PCA and ICA.

[4] See El Ghaoui et al. (2003) for a study of the minimum-VaR portfolio.

We consider five portfolio policies. First, the sample minimum-variance portfolio (MV), given by (17). Second and third, the shrinkage portfolio obtained by combining the MV portfolio and the factor-risk-parity portfolio with respect to variance based on the *independent components* (ICMV), given by (19), and *principal components* (PCMV). Fourth, the sample minimum-modified-VaR portfolio (MMVaR). Fifth, the shrinkage portfolio obtained by combining the MMVaR portfolio and the factor-risk-parity portfolio with respect to modified VaR based on the independent components (ICMVaR), given by (31). We fix the VaR confidence level at $\alpha = 1\%$.[5]

To alleviate the impact of estimation error, we reduce problem dimensionality by keeping only the first two principal components, $K = 2$. This is consistent with results in the literature that show that a few principal components are sufficient to explain the cross section of stock returns; see, for instance, Kozak et al. (2018). Figure 4 reports the cumulative percentage variance explained by the first K principal components and shows that $K = 2$ principal components explain the majority of the variance for the *6BTM* and *6Prof* datasets, and offer a good trade-off between explanatory power and parsimony for the *30Ind* dataset.

To evaluate the out-of-sample performance of the different portfolios, we employ a rolling-horizon methodology. At a given time t, we estimate the different portfolio policies using the past five years of daily data (sample size $T = 1260$). We keep these portfolio weights constant for the next six months, thus rebalancing the portfolio every day to account for the impact of asset returns on portfolio weights, and compute the corresponding out-of-sample daily portfolio returns. After six months, we roll forward the estimation window by six months and repeat this procedure. At the end of this process, we obtain 35 years of out-of-sample daily portfolio returns.

We compare the out-of-sample performance of the different portfolios in terms of mean-variance trade-off, tail risk and portfolio-weight stability. For mean-variance trade-off, we report the annualized sample mean, volatility and Sharpe ratio. For tail risk, we report the daily sample skewness, excess kurtosis, 1% MVaR and 1% modified Sharpe ratio, which is the ratio between the mean and the MVaR of the portfolio return as in Gregoriou and Gueyie (2003).[6] For portfolio-weight stability, we report the daily portfolio turnover, computed as the average of the absolute value of daily rebalancing trades across the N assets, over the 35 years of daily trading dates.

To test the statistical significance of the difference between the Sharpe ratio or the modified

[5] We use the Matlab function *fmincon* with the *GlobalSearch* option to compute the MMVaR and ICMVaR portfolios, and we compute the IC basis with the *FastICA* Matlab package available on Aapo Hyvärinen's website: `www.cs.helsinki.fi/u/ahyvarin/papers/fastica.shtml`.

[6] We rely on daily, not annualized, measures for tail risk because their orders of magnitude are more intuitive.

Sharpe ratio of two given portfolios, we use circular-bootstrapping methods that are well-suited for returns that feature fat tails and non-stationarity. Specifically, we compute the two-sided p-value associated with the null hypothesis that the Sharpe ratio or the modified Sharpe ratio of the shrinkage portfolio coincides with that of the MV portfolio (for ICMV and PCMV) or the MMVaR portfolio (for ICMVaR). The bootstrapped p-values are computed as proposed by Ledoit and Wolf (2008) for the Sharpe ratio and Ardia and Boudt (2015) for the modified Sharpe ratio. As in DeMiguel et al. (2009a), we use $B = 1000$ bootstrap resamples and block size $b = 5$.[7]

We calibrate the shrinkage intensity δ via 10-fold cross-validation; see Chapter 7 in Hastie et al. (2001). The calibration criteria considered are the Sharpe ratio and the modified Sharpe ratio.[8] We find that the qualitative insights are robust to using the Sharpe ratio or the modified Sharpe ratio as calibration criterion. Thus, for brevity, we report only the results for the modified Sharpe ratio criterion. Figure 5 depicts these two cross-validated criteria over the entire sample as a function of the shrinkage intensity δ. The figure provides six different panels with the results for the three datasets and two calibration criteria. The figure shows that the ICMV portfolio is likely to dominate the PCMV portfolio out of sample. To see this, note that shrinking the MV portfolio toward the PC-risk-parity portfolio yields systematically worse cross-validated criteria than shrinking toward the IC-risk-parity portfolio *for all* δ. Moreover, although for the PCMV portfolio cross-validation indicates that it is better to stick to the MV portfolio by setting the shrinkage intensity $\delta = 0$, for the ICMV portfolio the optimal shrinkage intensity δ is always substantially larger than zero.

6.2. Results

Table 2 reports the out-of-sample performance of the five portfolio policies across the three datasets. We first compare the performance of the ICMV and PCMV portfolios to test whether the shrinkage portfolios based on independent components outperform those based on principal components. The results confirm that portfolios based on the IC basis have superior out-of-sample performance. In terms of Sharpe ratio, the ICMV portfolio systematically outperforms the PCMV portfolio. The gains come mostly from a higher portfolio mean return. Regarding tail risk, the ICMV portfolio also outperforms the PCMV portfolio in terms of kurtosis, modified VaR and modified Sharpe

[7] We employ the R package `PeerPerformance` to compute the bootstrapped p-values.

[8] The 10-fold cross-validation method works as follows. First, set $\delta = 0$. Then, divide the estimation window in 10 distinct intervals of equal length. Select one interval, remove it from the estimation window, and use the remaining data to compute the portfolio policy. Use these portfolio weights to compute out-of-sample returns on the interval that was removed. After repeating this procedure for each of the 10 intervals, compute the Sharpe ratio or the modified Sharpe ratio on all the out-of-sample returns. Increase δ by 0.01 and repeat the process above. Finally, when $\delta = 1$ is reached, select the value of δ associated with the maximum Sharpe ratio or modified Sharpe ratio.

ratio. This result highlights that using maximally independent factors effectively reduces tail risk. Moreover, the ICMV portfolio has a smaller turnover, which makes its practical implementation less costly and demonstrates that the IC basis is more robust out of sample than the PC basis.

Comparing the ICMV and MV portfolios, we observe that there are clear benefits from shrinking the MV portfolio toward the IC-risk-parity portfolio. In terms of mean-variance trade-off, the out-of-sample Sharpe ratio of the ICMV portfolio is higher than that of the MV portfolio. The gains can be traced back to a substantially higher mean accompanied by an only moderately higher volatility. In terms of tail risk, the ICMV portfolio also outperforms the MV portfolio in terms of kurtosis, modified VaR and modified Sharpe ratio. The improvement in the Sharpe ratio and the modified Sharpe ratio is statistically significant for *6BTM* and *6Prof*, but not for *30Ind*. This is to be expected because the optimal shrinkage intensity is, on average, lower for *30Ind*. Last, the superior performance of the ICMV portfolio is achieved with a turnover that is only slightly higher than that of the MV portfolio, and thus, remains appealing in practice.

We then turn to the two portfolios based on the VaR risk measure: the sample minimum-modified-VaR portfolio (MMVaR) and the portfolio that shrinks the MMVaR portfolio towards the IC-risk-parity portfolio with respect to modified VaR (ICMVaR). The ICMVaR portfolio outperforms the MMVaR portfolio in terms of Sharpe ratio and modified Sharpe ratio, while generating a similar turnover. The ICMVaR portfolio is also an appealing alternative to the ICMV portfolio. In terms of Sharpe ratio, the ICMVaR portfolio performs similarly for the *6BTM* and *30Ind* datasets, but it substantially outperforms the ICMV portfolio for the *6Prof* dataset. More importantly, the ICMVaR portfolio systematically outperforms all other portfolios in terms of modified Sharpe ratio, which demonstrates that the returns of factor-risk-parity portfolios based on the *VaR risk measure* have better higher-order-moment properties. However, the ICMVaR portfolio has a higher turnover than the ICMV portfolio, which is expected because higher-order risk measures like VaR are more sensitive to estimation error than the variance.

7. Conclusion

Factor-risk parity is an effective approach to portfolio diversification. Compared to asset-risk parity, it has the advantage of diversifying risk on an uncorrelated basis, which improves diversification. However, we show that, for any portfolio, there is an uncorrelated-factor basis for which the portfolio is a factor-risk-parity portfolio. Thus, this framework is arbitrary when decorrelation is the only criterion governing the choice of factor basis. The basis formed by the first principal components of asset returns is not more meaningful because, once dimensionality has been reduced, any further rotation of the principal components yields an equally uncorrelated basis. To circumvent these concerns, we propose using the independent-components basis, which is the uncorrelated basis that features the least (higher-order) dependence. This basis offers three benefits: it generally leads to a unique factor-risk-parity portfolio, it enhances diversification because it accounts for higher-order moments and it provides a parsimonious way to deal with higher-order risk measures. Empirically, shrinking the minimum-risk portfolio toward the IC-risk-parity portfolio helps improve the out-of-sample performance in terms of mean-variance trade-off and tail risk.

Appendix

A. Proofs of propositions

A.1. Proposition 1

Proof. Observe first that for any random vector Y, we have $I(Y) = \sum_{i=1}^{K} h(Y_i) - h(Y)$, where $h(Y)$ is the joint entropy of Y; see Cover and Thomas (2006, p.251). Moreover, it is easy to show that $h(MY) = h(Y) + \ln|\det(M)|$ for every square matrix M. Thus, because we optimize over $SO(K)$, $\det(R) = 1$ and $h(RY) = h(Y)$ is a constant. Therefore, minimizing $I(Y)$ over $SO(K)$ amounts to minimizing the sum of the individual entropies $\sum_{i=1}^{K} h(Y_i)$. □

A.2. Proposition 2

Proof. To achieve factor-risk parity, we want $\tilde{w}_i(R)^2 = c$ for all i and some constant $c \in \mathbb{R}_0$. This can be rewritten as $\tilde{w}(R) = c\mathbf{1}_K^{\pm}$, which, from (11), translates in portfolio weights given by $w = cV\Lambda^{-1/2}R'\mathbf{1}_K^{\pm}$. Rescaling this expression to get $\mathbf{1}_N' w = 1$, c vanishes, and the final expression for the portfolio $w^\star(R)$ in (2) is obtained. Each choice of $\mathbf{1}_K^{\pm}$ gives equal squared factor exposures $\tilde{w}_i(R)^2$, and thus, satisfies $\mathcal{H}[p(R)] = K$. There are 2^K different vectors $\mathbf{1}_K^{\pm}$, which, combined with the normalization constraint, means that there are 2^{K-1} factor-risk-parity portfolios. Finally, the set of maximizers $w^\star(R)$ is invariant to a change of sign and permutation because, from (10), changing the sign or permuting the components of Y changes the corresponding factor exposures $\tilde{w}(R)$ in the same way, so that the projected portfolio return $\tilde{w}(R)'Y$ remains identical. □

A.3. Proposition 3

Proof. The factor-risk-parity portfolio

$$w^\star(R) = \frac{V\Lambda^{-1/2}R'\mathbf{1}_K^{\pm}}{\mathbf{1}_N' V\Lambda^{-1/2}R'\mathbf{1}_K^{\pm}} \tag{32}$$

for any $\mathbf{1}_K^{\pm}$ is the portfolio belonging to $\mathcal{W}$ for which all squared components of $\tilde{w}(R)$ are equal. Observe that the denominator of (32) is just a scaling factor projecting the numerator onto $\mathcal{W}$; thus, we need not worry about it. To prove the proposition, we merely need to show that, for any non-zero vector $w \in \mathbb{R}^n$, we can find a rotation matrix $R \in SO(K)$ such that each squared entry of $\tilde{w}(R)$ is equal; that is, such that $\tilde{w}(R) = c\mathbf{1}_K^{\pm}$ for some $c \in \mathbb{R}_0$. From (11), this amounts to showing that given any non-zero vector $w \in \mathbb{R}^N$, there exists $R \in SO(K)$ such that

$$R\Lambda^{1/2}V'w = c\mathbf{1}_K^{\pm} \tag{33}$$

holds for some $c \in \mathbb{R}_0$. We have that $||\mathbf{1}_K^{\pm}|| = \sqrt{K}$ and the scaling coefficient c given by

$$c = \frac{||\Lambda^{1/2}V'w||}{\sqrt{K}}$$

means that (33) can be written as

$$Rx = y, \tag{34}$$

where $x := \frac{\Lambda^{1/2}V'w}{||\Lambda^{1/2}V'w||}$ and $y := \mathbf{1}_K^{\pm}/\sqrt{K}$ are unit-norm vectors. The claim then comes from the fact that the space of unit-norm K-dimensional vectors is spanned by $SO(K)$. More explicitly, given two arbitrary unit-norm vectors $x, y \in \mathbb{R}^K$, there exists a rotation matrix $R \in SO(K)$ such that $Rx = y$. To see this, notice that a rotation around the origin is merely a specific *direct isometry*. But every isometry in $\mathbb{R}^K$ can be written as a composite of at most $K+1$ *reflections*; see Theorem A.1.5 of Pressley (2010, p.387). Actually, R can be built as a sequence of two reflections: the first reflection S is done with respect to the hyperplane orthogonal to $x + y$, and maps x to $-y$. The second reflection Q maps $-y$ to y. Eventually, because a reflection is an indirect isometry, these two reflections form a direct isometry, and $R = QS$ is the rotation matrix mapping x to $y = Rx$. $\square$

A.4. Proposition 4

Proof. The sample return variance of the portfolios $w^\star(R^\dagger)$ develops as

$$w^\star(R^\dagger)'\widehat{\Sigma}_X w^\star(R^\dagger) = w^\star(R^\dagger)'V_N\Lambda_N V_N' w^\star(R^\dagger) = \frac{(\mathbf{1}_K^\pm)'R^\dagger\Lambda^{-1/2}V'V_N\Lambda_N V_N' V\Lambda^{-1/2}(R^\dagger)'\mathbf{1}_K^\pm}{(\mathbf{1}_N' V\Lambda^{-1/2}(R^\dagger)'\mathbf{1}_K^\pm)^2}, \quad (35)$$

where we explicitly write the subscript N to mean that V_N and Λ_N are the full $N\times N$ eigenvectors and eigenvalues matrices. $V'V_N$ is given by the first K rows of I_N, and thus $V'V_N\Lambda_N$ is made of the first K rows of Λ_N. Multiplying $V_K'V_N\Lambda_N$ by $V_N'V$, the first K columns of I_N, thus gives

$$V'V_N\Lambda_N V_N'V = \Lambda.$$

As a result, (35) simplifies to

$$w^\star(R^\dagger)'\widehat{\Sigma}_X w^\star(R^\dagger) = \frac{(\mathbf{1}_K^\pm)'\mathbf{1}_K^\pm}{(\mathbf{1}_N' V\Lambda^{-1/2}(R^\dagger)'\mathbf{1}_K^\pm)^2} = \frac{K}{(\mathbf{1}_N' V\Lambda^{-1/2}(R^\dagger)'\mathbf{1}_K^\pm)^2}. \quad (36)$$

Clearly, the vector $\mathbf{1}_K^\pm$ that minimizes (36) is unique and corresponds to

$$\mathbf{1}_K^\pm = \operatorname{sign}((\mathbf{1}_N' V\Lambda^{-1/2}(R^\dagger)')') = \operatorname{sign}(R^\dagger\Lambda^{-1/2}V'\mathbf{1}_N) = \operatorname{sign}(R^\dagger V'\mathbf{1}_N) =: \mathbf{1}_K^{MV}. \qquad \square$$

A.5. Proposition 5

Proof. To prove the proposition, we use the well-known fact that cumulants are additive for independent random variables. The third and fourth cumulants of the projected portfolio return $P = \tilde{w}(R^\dagger)'Y^\dagger$ correspond to $m_3(P)$ and $m_4(P) - 3m_2(P)^2$, respectively. Therefore, under the independence of the IC factors, we directly find that the third central moment of P is given by

$$m_3(P) = \sum_{i=1}^K m_3(\tilde{w}_i(R^\dagger)Y_i^\dagger) = \sum_{i=1}^K \tilde{w}_i(R^\dagger)^3 m_3(Y_i^\dagger).$$

Similarly, the fourth central moment of P is found as

$$\begin{aligned}
m_4(P) - 3m_2(P)^2 &= \sum_{i=1}^K m_4(\tilde{w}_i(R^\dagger)Y_i^\dagger) - 3m_2(\tilde{w}_i(R^\dagger)Y_i^\dagger)^2 \\
\Leftrightarrow\ m_4(P) &= \sum_{i=1}^K \tilde{w}_i(R^\dagger)^4 m_4(Y_i^\dagger) - 3\sum_{i=1}^K \tilde{w}_i(R^\dagger)^4 + 3\Big(\sum_{i=1}^K \tilde{w}_i(R^\dagger)^2\Big)^2 \\
\Leftrightarrow\ m_4(P) &= \sum_{i=1}^K \tilde{w}_i(R^\dagger)^4 m_4(Y_i^\dagger) - 3\sum_{i=1}^K \tilde{w}_i(R^\dagger)^4 + 3\sum_{i=1}^K\sum_{j=1}^K (\tilde{w}_i(R^\dagger)\tilde{w}_j(R^\dagger)^2
\end{aligned}$$

$$\Leftrightarrow m_4(P) = \sum_{i=1}^{K} \tilde{w}_i(R^\dagger)^4 m_4(Y_i^\dagger) + 3\sum_{i=1}^{K}\sum_{j\neq i}(\tilde{w}_i(R^\dagger)\tilde{w}_j(R^\dagger)^2. \qquad \square$$

A.6. Proposition 6

Proof. Boudt et al. (2008) show that the risk contribution of the ith asset X_i to the modified VaR of the portfolio return $P = w'X$ can be computed via the Euler risk contribution

$$w_i \frac{\partial \mathrm{MVaR}_P(\alpha)}{\partial w_i} = w_i \times \Bigg[-\mu(X_i) - \frac{\partial_i m_2(P)}{2\sqrt{m_2(P)}} q + \sqrt{m_2(P)} \Bigg(-\frac{1}{6}(q_\alpha^2 - 1)\partial_i s(P) \\ - \frac{1}{24}(q_\alpha^3 - 3q_\alpha)\partial_i \kappa(P) + \frac{1}{18}(2q_\alpha^3 - 5q_\alpha)s(P)\partial_i s(P) \Bigg) \Bigg], \tag{37}$$

where the portfolio-return central moments are given by

$$m_2(P) = w'\Sigma_X w, \quad m_3(P) = w'\Phi_X(w \otimes w) \;\text{ and }\; m_4(P) = w'\Psi_X(w \otimes w \otimes w), \tag{38}$$

with Φ_X and Ψ_X the co-skewness and co-kurtosis matrices of X, respectively. The MVaR contribution of the ith IC factor $Y_i^\dagger$ is equivalent to (37)-(38) with respect to $P = \tilde{w}(R^\dagger)'Y^\dagger$, and as we assume $Y^\dagger$ is perfectly independent, $m_2(P)$, $m_3(P)$ and $m_4(P)$ simplify to (12), (22) and (23). $\square$

Acknowledgements

Nathan Lassance's research is funded by the Fonds de la Recherche Scientifique (F.R.S.-FNRS) under grant ASP (grant number FC 17775). We thank participants of the 2018 INFORMS Annual Meeting Phoenix for comments and discussions.

References

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

Bai X, Scheinberg K, Tutuncu R (2016) Least-squares approach to risk parity in portfolio selection. *Quant. Finance* 16(3):357–376.

Baillie R, Bollerslev T (1992) Prediction in dynamic models with time-dependent conditional variances. *J. Econometrics* 52(1–2):91–113.

Ban G, El Karoui N, Lim A (2018) Machine learning and portfolio optimization. *Management Sci.* 64(3):1136–1154.

Bertsimas D, Shioda R (2009) Algorithm for cardinality-constrained quadratic optimization. *Comp. Optim. Appl.* 43(1):1–22.

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

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

Campbell J, Lo A, MacKinlay A (1997) *The Econometrics of Financial Markets* (Princeton University Press, Princeton, NJ).

Chen Y, Härdle W, Spokoiny V (2010) GHICA—Risk analysis with GH distributions and independent components. *J. Empirical Finance* 17(2):255-269.

Comon P (1994) Independent component analysis, a new concept? *Signal Processing* 36(3):287–314.

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

Deguest R, Martellini L, Meucci A (2013) Risk parity and beyond—From asset allocation to risk allocation decisions. EDHEC Working paper.

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

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

DeMiguel V, Martin-Utrera A, Nogales F (2013) Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *J. Banking Finance* 37(8):3018–3034.

DeMiguel V, Nogales F (2009) Portfolio selection with robust estimation. *Oper. Res.* 55(5):798–812.

El Ghaoui L, Oks M, Oustry F (2003) Worst-case Value-At-Risk and robust portfolio optimization: A conic programming approach. *Oper. Res.* 51(4):543–556.

Fabozzi F, Giacometti R, Tsuchida N (2016) Factor decomposition of the Eurozone sovereign CDS spreads. *J. Inter. Money Finance* 65:1–23.

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

Gao J, Li D (2013) Optimal cardinality constrained portfolio selection. *Oper. Res.* 61(3):745–761.

García-Ferrer A, González-Prieto E, Peña D (2012) A conditionally heteroskedastic independent factor model with an application to financial stock returns. *Inter. J. Forecast.* 28(1):70–93.

Goldfarb D, Iyengar G (2003) Robust portfolio selection problems. *Math. Oper. Res.* 28(1):1–38.

Green R, Hollifield B (1992) When will mean-variance efficient portfolios be well diversified? *J.*

Finance 47(5):1785–1809.

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

Hastie T, Tibshirani R, Friedman J (2001) *The Elements of Statistical Learning: Prediction, Inference and Data Mining*, 2nd ed. (Springer, New York).

Haugh M, Iyengar G, Song I (2017) A generalized risk budgeting approach to portfolio construction. *J. Comput. Finance* 21(2):29–60.

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

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

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

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

Ji R, Lejeune M (2015) Risk-budgeting multi-portfolio optimization with portfolio and marginal risk constraints. *Ann. Oper. Res.* 262(2):547–578.

Jorion P (1986) Bayes-Stein estimation for portfolio analysis. *J. Financial Quant. Anal.* 21(3):279–292.

Kan R, Zhou G (2007) Optimal portfolio choice with parameter uncertainty. *J. Financial Quant. Anal.* 42(3):621–656.

Kozak S, Nagel S, Santosh S (2018) Interpreting factor models. *J. Finance* 73(3):1183–1223.

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

Ledoit O, Wolf M (2008) Robust performance hypothesis testing with the sharpe ratio. *J. Empirical Finance* 15(5):850–859.

Maillard S, Roncalli T, Teiletche J (2010) On the properties of equally-weighted risk contributions portfolios. *J. Portfolio Management* 36(4):60–70.

Markowitz H (1952) Portfolio selection. *J. Finance* 7(1):77–91.

Martin-Clemente R, Zarzoso V (2017) On the link between L1-PCA and ICA. *IEEE Trans. Pattern Anal. Mach. Intell.* 39(3):515–528.

Massacci D (2017) Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. *Management Sci.* 63(9):2773–3145.

Meucci A (2009) Managing diversification. *Risk* 22(5):74–79.

Meucci A, Santangelo A, Deguest R (2015) Risk budgeting and diversification based on optimized uncorrelated factors. *Risk* 11(29):70–75.

Olivares-Nadal A, DeMiguel V (2018) Technical note—A robust perspective on transaction costs in portfolio optimization. *Oper. Res.* 66(3):733–739.

Poon S, Rockinger M, Tawn J (2004) Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *Rev. Financial Stud.* 17(2):581–610.

Pressley A (2010) *Elementary Differential Geometry* (Springer, London).

Roncalli T, Weisang G (2016) Risk parity portfolios with risk factors. *Quant. Finance* 16(3):377–388.

Siegel A, Woodgate A (2007) Performance of portfolios optimized with estimation error. *Management Sci.* 53(6):1005–1015.

Tu J, Zhou G (2011) Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *J. Financial Econom.* 99(1):204–215.

Vrins F (2007) *Contrast Properties of Entropic Criteria for Blind Source Separation: A Unifying Framework based on Information-Theoretic Inequalities* (Presses Universitaires de Louvain, Louvain-la-Neuve).

Vrins F, Verleysen M (2005) On the entropy minimization of a linear mixture of variables for source separation. *Signal Processing* 85(5):1029–1044.

Xu H, Caraminis C, Mannor S (2016) Statistical optimization in high dimensions. *Oper. Res.* 64(4):958–979.

Tables

Table 1 List of the main mathematical symbols used in the paper

Symbol	Meaning	Page number
X	Asset returns	6
Σ_X	Covariance matrix of X	6
V	Eigenvectors matrix	6
Λ	Eigenvalues matrix	6
$Y^\star$	Principal-components basis	6
$SO(K)$	Special orthonormal group of order K	6
R	Rotation matrix	6
$I(Y)$	Mutual information of Y	7
$\bar{I}(Y)$	Relative mutual information of Y	7
$h(Y)$	Entropy of Y	7
$R^\dagger$	ICA rotation matrix	7
$Y^\dagger$	Independent-components basis	7
G	Entropy estimation functional	8
$R(\theta)$	Rotation matrix in two dimensions	9
P	Portfolio return	10
w	Portfolio weights	10
$\mathcal{W}$	Space of portfolio weights	10
$\widehat{\Sigma}_X$	Sample covariance matrix of X	10
$\tilde{w}(R)$	Exposures on the factors $Y = RY^\star$	10
$p(R)$	Relative variance contributions	11
$\mathcal{H}[p]$	Diversification index of p	11
$w^\star(R)$	Set of factor-risk-parity portfolios	11
$w^\star(I_K)$	Set of PC-risk-parity portfolios	11
$w^\star(R^\dagger)$	Set of IC-risk-parity portfolios	13
$w^\star_{MV}$	Minimum-variance portfolio	14
$w^\star_{IC}$	IC-risk-parity portfolio with minimum variance	14
δ	Shrinkage intensity	14
$w^\star_{ICMV}$	ICMV portfolio	14
$w^\star_{PCMV}$	PCMV portfolio	15
$\mu(Z)$	Mean of Z	15
$m_k(Z)$	kth central moment of Z	15
$s(Z)$	Skewness of Z	15
$\kappa(Z)$	Excess kurtosis of Z	15
Φ_Y	Co-skewness matrix of Y	15
Ψ_Y	Co-kurtosis matrix of Y	15
$\mathrm{MVaR}_Z(\alpha)$	Modified VaR of Z	17
$\mathrm{MVaR}_i(\alpha)$	MVaR contribution of $Y_i^\dagger$	17
$p(\alpha)$	Relative MVaR contributions	18
$\hat{w}^\star_{IC}(\alpha)$	Estimator of IC-risk-parity portfolio w.r.t. MVaR	18
$w^\star_{MMVaR}(\alpha)$	Minimum-MVaR portfolio	18
$w^\star_{ICMVaR}(\alpha)$	ICMVaR portfolio	18

Notes. The symbols are listed in order of their appearance in the text. The page number refers to the first time that the symbol is introduced.

Table 2 Out-of-sample performance of portfolio policies

6BTM dataset

	MV	ICMV	PCMV	MMVaR	ICMVaR
Mean	19.75%	21.39%	20.84%	23.50%	**23.93%**
Volatility	**13.36%**	13.41%	13.60%	15.26%	15.14%
Sharpe ratio	1.48	**1.59*****	1.53	1.54	1.58*
Skewness	-0.32	-0.33	-0.23	**-0.04**	-0.16
Excess kurtosis	20.51	11.53	18.53	8.72	**7.22**
1% modified Value-at-Risk	6.08%	4.33%	5.75%	4.13%	**3.84%**
Modified Sharpe ratio ($\times 10^2$)	1.29	1.96**	1.44**	2.26	**2.47***
Portfolio turnover	**2.47%**	2.79%	3.29%	5.75%	5.75%
Average shrinkage intensity δ	0.00	0.63	0.59	0.00	0.24

6Prof dataset

	MV	ICMV	PCMV	MMVaR	ICMVaR
Mean	17.53%	20.41%	17.83%	22.38%	**22.55%**
Volatility	**13.98%**	14.83%	14.41%	16.14%	15.60%
Sharpe ratio	1.25	1.38***	1.24	1.39	**1.45***
Skewness	**-0.03**	-0.35	-0.04	-0.03	-0.25
Excess kurtosis	22.55	14.86	20.06	14.66	**10.99**
1% modified Value-at-Risk	6.64%	5.53%	6.32%	5.78%	**4.88%**
Modified Sharpe ratio ($\times 10^2$)	1.05	1.46*	1.12*	1.54	**1.83**
Portfolio turnover	**2.44%**	2.93%	3.09%	5.16%	5.00%
Average shrinkage intensity δ	0.00	0.71	0.39	0.00	0.23

30Ind dataset

	MV	ICMV	PCMV	MMVaR	ICMVaR
Mean	11.18%	12.05%	10.74%	10.22%	**12.16%**
Volatility	**11.71%**	12.14%	11.90%	13.29%	12.22%
Sharpe ratio	0.95	0.99	0.90	0.77	**1.00*****
Skewness	-0.74	-0.75	-0.72	-0.97	**-0.66**
Excess kurtosis	23.99	**20.21**	22.58	21.59	20.35
1% modified Value-at-Risk	6.06%	**5.60%**	5.91%	6.43%	5.65%
Modified Sharpe ratio ($\times 10^2$)	0.73	0.85	0.72	0.63	**0.85**
Portfolio turnover	**2.45%**	2.85%	3.23%	4.35%	4.57%
Average shrinkage intensity δ	0.00	0.42	0.25	0.00	0.60

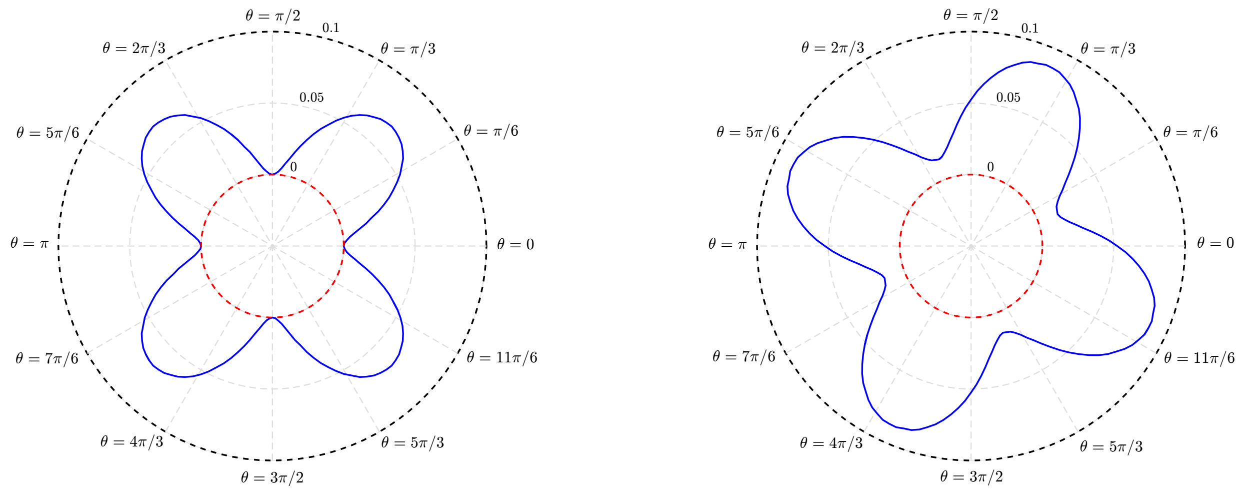
Notes. This table reports the out-of-sample performance of the five portfolio policies on the three datasets following the methodology in Section 6.1. The portfolio mean return, volatility and Sharpe ratio are annualized, while all other performance criteria are reported in daily terms. The shrinkage intensity δ is computed via 10-fold cross-validation, using the modified Sharpe ratio as the calibration criterion. Bold figures indicate the best strategy for each criterion. We evaluate the statistical significance of the Sharpe ratio and the modified Sharpe ratio compared to the MV portfolio (for ICMV and PCMV) or the MMVaR portfolio (for ICMVaR). The two-sided p-values are computed via the bootstrapping methodology of Ledoit and Wolf (2008) for the Sharpe ratio and Ardia and Boudt (2015) for the modified Sharpe ratio. The stars $\star$, $\star\star$ and $\star\star\star$ mean that the p-value is less than 20%, 10% and 5%.

Figures

Figure 1   Mutual information versus correlation

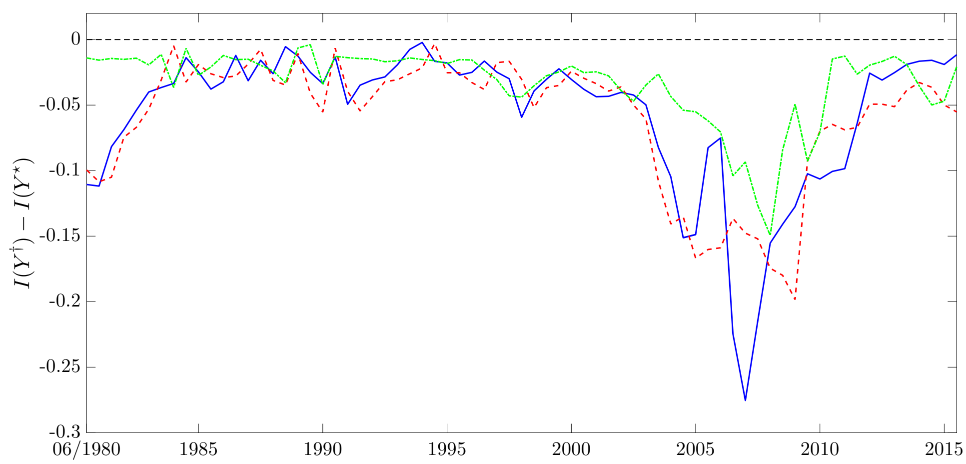
(a) Independent non-Gaussian asset returns

(b) Dependent non-Gaussian asset returns



Notes. The figure depicts the relative mutual information $\bar{I}(Y)$ in solid blue and correlation $\mathrm{Corr}(Y)$ in dashed red for the rotation of the principal components for two asset returns. Figure 1a considers the case with independent non-Gaussian asset returns and Figure 1b the case with dependent non-Gaussian asset returns, corresponding to Examples 1 and 2, respectively. The plots are in polar coordinates. The angle θ determines the rotation of the principal components, $Y := R(\theta)Y^\star$. The values of $\bar{I}(Y)$ and $\mathrm{Corr}(Y)$ are given by the distance from the origin.

Figure 2   Mutual information of independent components versus principal components

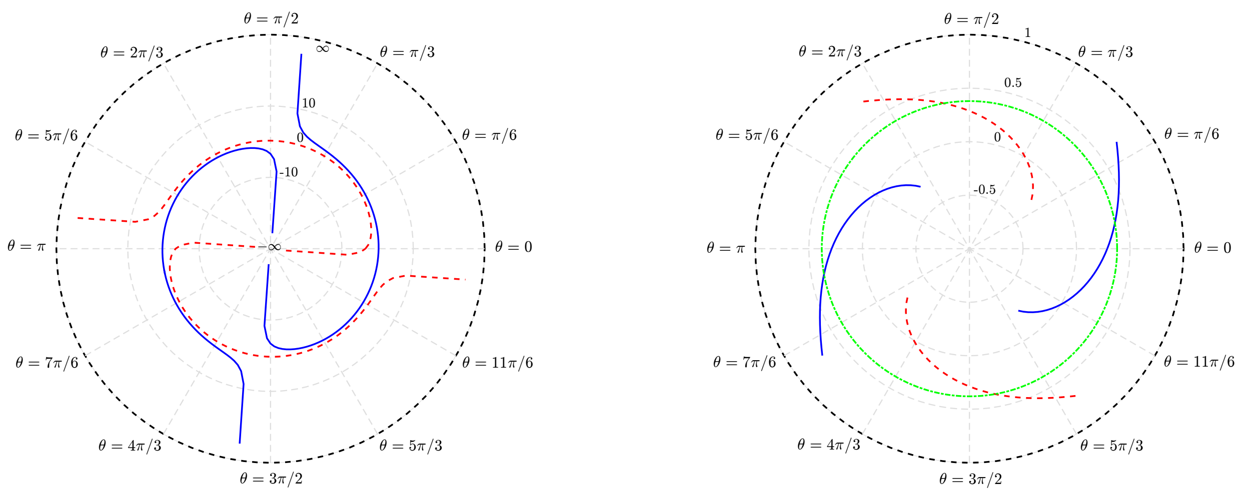


Notes. The figure depicts the time evolution of the difference in mutual information between the $K = 2$ independent and principal components for three datasets: *6BTM* in solid blue, *6Prof* in dashed red and *30Ind* in dotted green. The mutual-information difference is computed every six months using an estimation window containing the previous five years of daily asset returns. The x-axis labels indicate the window midyear.

Figure 3 Arbitrariness of the factor-risk-parity portfolio

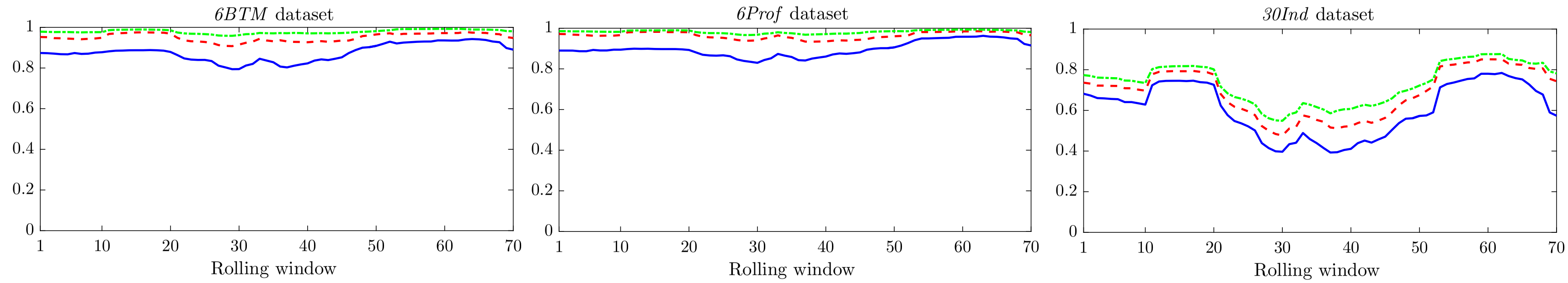
(a) Weight on first asset of the two factor-risk-parity portfolios

(b) Weight on first asset of factor-risk-parity portfolio with minimum return variance



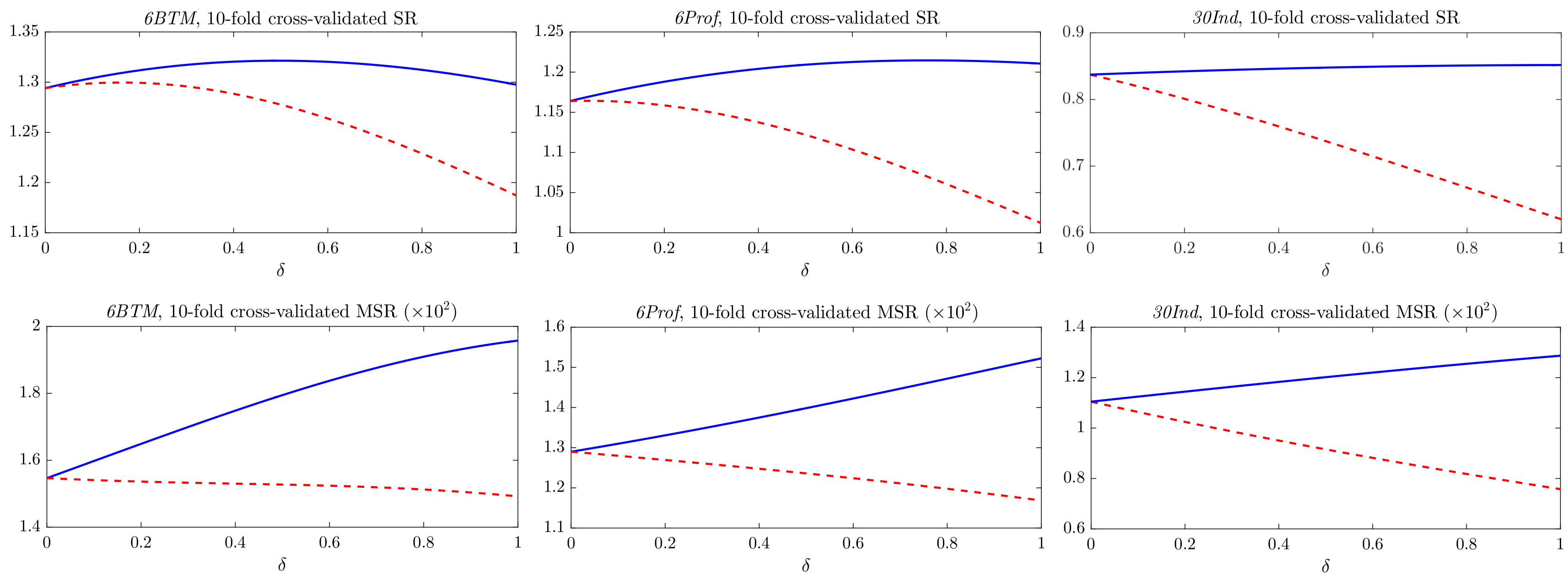
Notes. The figure depicts the weight on the first asset of the two factor-risk-parity portfolios for the dependent non-Gaussian asset returns in Example 3. The plots are in polar coordinates. The angle θ determines the rotation of the principal components, $Y := R(\theta)Y^{\star}$. The value of the portfolio weight is given by the distance from the origin. Figure 3a depicts the weight of the two factor-risk-parity-portfolios (solid blue and dashed red, respectively). Figure 3b depicts, for each angle θ, the factor-risk-parity-portfolio weight on the first asset associated with the factor-risk-parity portfolio with minimum return variance, and the IC-risk-parity-portfolio weight associated with the lowest portfolio-return variance in dotted green.

Figure 4 Cumulative percentage variance explained by the principal components



Notes. The figure depicts the time evolution of the cumulative percentage variance explained by the first K principal components for the three datasets, where $K = 1$ is in solid blue, $K = 2$ is in dashed red and $K = 3$ is in dotted green. The cumulative percentage variances are computed every six months using an estimation window containing the previous five years of daily asset returns.

Figure 5 Cross-validation of the shrinkage intensity δ



Notes. The figure depicts the 10-fold cross-validated Sharpe ratio (SR) and modified Sharpe ratio (MSR) of the ICMV and PCMV portfolios as a function of the shrinkage intensity δ, computed on the whole sample of the three datasets with $K = 2$. The results for the ICMV portfolio are in solid blue and for the PCMV portfolio in dashed red.