Optimal Portfolio Diversification via Independent Component Analysis

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

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

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

This version: January 2020

Abstract

A popular approach to enhance portfolio diversification is to use the *factor-risk-parity portfolio*, which is the portfolio whose return variance is spread equally among the principal components (PCs) of asset returns. Although PCs are unique and useful for dimension reduction, they are an arbitrary choice because any rotation of the PCs remains uncorrelated. This is problematic because we demonstrate that any portfolio is the factor-variance-parity portfolio corresponding to a specific rotation of the PCs. 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 moments in asset returns. We demonstrate that using the IC-variance-parity portfolio helps to reduce the return kurtosis. We also show how to exploit the near independence of the ICs to parsimoniously estimate the factor-risk-parity portfolio based on Value-at-Risk. Finally, we empirically demonstrate that portfolios based on ICs outperform those based on PCs, and the minimum-variance portfolio.

Keywords

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

1. Introduction

The mean-variance portfolio of Markowitz (1952) is optimally 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 (Green and Hollifield 1992). Recently, to improve the out-of-sample performance of mean-variance portfolios, researchers have thus proposed incorporating portfolio-weight diversification as an explicit objective. 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. For example, Roncalli (Section 2.5.1.1, 2013) shows that it is not duplication invariant: when adding an asset identical to one in the current investment universe, the weights on the other assets in the asset-risk-parity portfolio are no longer the same.

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 spread equally among a set of *uncorrelated factors*; see, for instance, Meucci (2009), Deguest et al. (2013), Meucci et al. (2015) and Roncalli and Weisang (2015). To implement this portfolio, one needs to rely on a specific set of uncorrelated factors. The standard choice is to rely on the principal components (PCs) of asset returns, provided by *principal component analysis* (PCA).

Although the PCs are useful for dimension reduction (Campbell et al. 1997, Xu et al. 2016), they are problematic for factor-risk parity for two main reasons. First, relying on the PCs is arbitrary because once dimension has been reduced, any further rotation of the PCs remains uncorrelated, and thus, is equally valid for factor-risk-parity purposes. This is worrying because as we demonstrate, any portfolio is a factor-variance-parity portfolio corresponding to a specific rotation of the PCs. Second, the PCs are not optimal because they are 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 for risk parity because they are correlated, relying on PCs is suboptimal because they may feature higher-moment dependencies.

To overcome these two challenges, we propose using the 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 PCs 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 present tail connectedness, which increases during periods of turmoil. The independent components (ICs) overcome the two challenges related to PCA: they discriminate among all the uncorrelated factors and they enhance diversification by accounting for higher-moment dependence. In addition, we show that the ICs allow one to parsimoniously deal with higher-moment risk measures. Finally, we show that the factor-variance-parity portfolio based on ICs can be seen as a shrinkage estimator of the tangent mean-variance portfolio that imposes that the return of every independent component has the same Sharpe ratio.

Our contribution is fourfold. First, we demonstrate that the arbitrariness of the PCs is worrying in the context of factor-risk parity. To do this, we first provide a closed-form expression for the factor-variance-parity portfolio corresponding to a generic rotation of the PCs. We then use this closed-form expression to show that any portfolio is a factor-variance-parity portfolio corresponding to a specific rotation of the PCs. Thus, the factor-decorrelation criterion is not sufficient to discriminate among all the possible portfolios and, although they are unique, the PCs are an arbitrary choice because any rotation of them remains uncorrelated.

Second, to overcome the difficulties associated with the PC-risk-parity portfolio, we propose relying instead on the independent components. We theoretically demonstrate that, under some assumptions, using the IC-variance-parity portfolio helps to reduce the return kurtosis. To do this, we first characterize the minimum and maximum excess kurtosis achievable by any portfolio. Then, we show that the excess kurtosis of the IC-variance-parity portfolio return is $1/K$ times the *arithmetic* mean of the excess kurtosis of the ICs, where K is the number of factors, and that it is often close to the minimum achievable excess kurtosis. Thus, diversifying the portfolio-return variance among ICs provides a practical approach to reduce the portfolio-return *tail* risk, which is

an important consideration for many investors; see Dittmar (2002) and Ang et al. (2006).

Third, we show that the ICs provide a parsimonious way to estimate higher moments of portfolio returns because although decorrelation suffices to decompose the portfolio-return variance as a sum of factor variance exposures, independence is required to obtain a similar decomposition for higher moments. For the case where the ICs are sufficiently independent, the higher comoments of ICs are negligible, and this greatly reduces the number of parameters to be estimated. We exploit this fact to show how to parsimoniously estimate the IC-risk-parity portfolio based on a four-moment approximation of Value-at-Risk (VaR).

Fourth, we evaluate the out-of-sample performance of the shrinkage portfolios obtained by combining minimum-risk portfolios and factor-risk-parity portfolios. We use the linear shrinkage of Kan and Zhou (2007) and Tu and Zhou (2011) with the shrinkage intensity estimated via cross-validation (Hastie et al. 2001). We find that the portfolios based on ICs substantially outperform those based on PCs in terms of Sharpe ratio and tail risk. The IC-based shrinkage portfolios also outperform the traditional minimum-risk portfolios, while keeping a similar level of turnover.

Our work is connected to four streams of literature. First, risk-parity portfolios. Maillard et al. (2010) show that the *long-only* asset-variance-parity portfolio is unique and its variance is in between that of the minimum-variance and equally weighted portfolios. Bai et al. (2016) show that *long-short* asset-variance-parity portfolios are no longer unique. Zhu et al. (2010), Cui et al. (2016) and Lejeune (2017) do not impose strict risk parity but, instead, consider mean-variance portfolios under marginal risk control. Boudt et al. (2013) and Baitinger et al. (2017) consider higher moments in asset-risk parity. 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 individual assets. Like us, Haugh et al. (2017) consider both variance and VaR risk measures, but for tractability they use the Gaussian VaR approximation, while we exploit the properties of ICs to rely on an approximation of VaR that incorporates skewness and kurtosis. Ji and Lejeune (2015) assign downside-risk budgets not only to sub-portfolios of assets but also to individual assets, and they 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 PCs as factors. Our work extends these two studies by allowing for dimension reduction, emphasizing the importance of

the selection of factors, proposing an alternative to the PCs based on ICA and considering higher-moment risk measures. An alternative to the PCs was also proposed by Meucci et al. (2015) who derive the uncorrelated factors that track another correlated set of factors, such as the original assets. These factors benefit from improved economic intuition but similar to the PCs, may still be far from independent. Finally, Roncalli and Weisang (2015) 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 moments.

Second, our work contributes to the literature on higher-moment portfolio selection by showing theoretically and empirically that IC-risk-parity portfolios help improve tail risk. Most existing higher-moment portfolios are obtained via a *direct* approach that finds a portfolio on a higher-moment efficient surface, and thus, requires estimating high-dimensional higher-comoment matrices. For example, Jondeau and Rockinger (2006), Guidolin and Timmermann (2008) and Martellini and Ziemann (2010) use a Taylor-series expansion of the investor's utility function to include higher moments, Athayde and Flores (2004) maximize the portfolio-return skewness for a fixed mean and variance and Briec et al. (2007) look for the highest improvements in mean, variance and skewness starting from a benchmark portfolio. In contrast, the IC-risk-parity portfolio is an *indirect* approach because it does not explicitly optimize the portfolio-return higher moments: it improves kurtosis simply by diversifying risk across ICs. Consequently, it is much more parsimonious because, in addition to the covariance matrix, one just needs to estimate the rotation matrix determining the ICs and, for the VaR risk measure, the marginal skewness and kurtosis of the ICs.

Third, 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). 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 exposures on the ICs.

Fourth, our work is related to the recent literature that applies ICA to factor analysis of financial time series; see, for instance, Chen et al. (2010), García-Ferrer et al. (2012) and Fabozzi et al. (2016). In portfolio selection, other papers use ICA to ease the modeling and estimation of the dependence between asset returns. Chen et al. (2007) focus on the computation of the VaR of portfolio returns. Hitaj et al. (2015) consider the maximization of the expected CARA utility. Finally, Lassance and Vrins (2019) apply the approach that we propose in Section 5.1 to parsimoniously estimate the higher moments of portfolio returns. Their purpose is fundamentally different from ours as they use this approach to enhance the robustness of minimum-risk portfolios, whereas we focus on the use of ICs to define *factor-risk-parity portfolios* that are optimally diversified. Masson (2019) in his master's thesis implements the approach proposed by Lassance and Vrins (2019) and compares its performance to that of various benchmarks.

The remainder of the article is organized as follows. Section 2 reviews PCA and ICA. Section 3 shows that factor decorrelation is not a discriminant criterion for factor-variance parity. Section 4 shows that the IC-variance-parity portfolio helps reduce the portfolio-return kurtosis, and derives the conditions under which this portfolio is mean-variance efficient. Section 5 considers the IC-risk-parity approach with VaR as risk measure. Section 6 evaluates the out-of-sample empirical performance of the proposed shrinkage portfolios. Section 7 concludes. The e-companion at the end of the paper contains proofs for all results. Finally, we denote deterministic vectors in bold lower case ($\boldsymbol{a}$), univariate random variables in upper case (A), multivariate random variables and matrices in bold upper case ($\boldsymbol{A}$), and all vectors are column vectors.

2. PCA versus ICA: from decorrelation to independence

In this section, we first review *principal component analysis* (PCA), a standard dimension-reduction technique that generates a specific basis of uncorrelated factors. We argue that, although the principal components (PCs) are unique, any rotation of them remains uncorrelated, and thus, the PCs are an arbitrary choice for factor-risk parity. We then introduce *independent component analysis* (ICA), which extends PCA by rotating the PCs to make them as independent as possible. This allows one to discriminate among all bases of uncorrelated factors.

2.1. Principal component analysis

Let $\boldsymbol{X} = (X_1, \ldots, X_N)'$ be a random vector of asset returns with $N \times N$ covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$. The covariance matrix can be diagonalized as $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{V}_N \boldsymbol{\Lambda}_N \boldsymbol{V}_N'$, where the N columns of $\boldsymbol{V}_N$ are the eigenvectors, and $\boldsymbol{\Lambda}_N := \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$ is the diagonal matrix made of the eigenvalues sorted in decreasing order. This factorization is at the heart of *principal component analysis* (PCA); see, for instance, Section 6.4 of Campbell et al. (1997). To simplify notation, we denote $\boldsymbol{\Lambda}$ the $K \times K$ diagonal matrix containing the K largest eigenvalues, assumed strictly positive, and $\boldsymbol{V}$ the $N \times K$ matrix whose columns contain the corresponding eigenvectors.

Definition 1 (PCA)**.** The *principal components* (PCs) are the projection of the asset returns $\boldsymbol{X}$ onto the first K standardized eigenvectors:

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

PCA is useful for dimension reduction because the first few PCs explain most of the variance in asset returns (Kozak et al. 2019). However, once dimension has been reduced, it is the fact that PCs are uncorrelated that makes them interesting for factor-risk parity, which spreads the portfolio-return risk across a set of *uncorrelated* factors. Indeed, most existing papers on factor-risk parity rely on the PCs as uncorrelated factors; see Meucci (2009), Deguest et al. (2013) and Roncalli and Weisang (2015). However, there is no sound motivation behind this particular choice because any rotation of the PCs remains uncorrelated, and therefore is as suitable for factor-risk parity as the PCs. To see this, let $\mathcal{SO}(K)$ be the *special orthonormal group* of order K; that is, the set of $K \times K$ rotation matrices $\boldsymbol{R}$ such that $\boldsymbol{R}\boldsymbol{R}' = \boldsymbol{I}_K$ and $\det(\boldsymbol{R}) = 1$. Then, for any $\boldsymbol{R} \in \mathcal{SO}(K)$, the factors $\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{Y}^\star$ remain uncorrelated: $\boldsymbol{\Sigma}_{\boldsymbol{Y}} = \boldsymbol{R}\boldsymbol{\Sigma}_{\boldsymbol{Y}^\star}\boldsymbol{R}' = \boldsymbol{R}\boldsymbol{I}_K\boldsymbol{R}' = \boldsymbol{I}_K$. Using the PCs amounts to choose $\boldsymbol{R} = \boldsymbol{I}_K$, a decision that looks arbitrary.

Therefore, there is an indeterminacy in the standard formulation of factor-risk parity because it does not motivate which uncorrelated factors one must consider. This would not be a problem if performing factor-risk parity over any choice of uncorrelated factors led to the same portfolio. However, as we show in Section 3.2, any portfolio is the factor-variance-parity portfolio associated with a particular rotation of the PCs. This suggests that one should think carefully about which set of uncorrelated factors, among the infinitely many rotations of the PCs, one should consider when performing factor-risk parity.

2.2. Independent component analysis

To resolve the arbitrariness inherent in the decorrelation criterion and discriminate among all the uncorrelated factors, one can look for the factors that are as *independent* as possible. Contrary to the PCs, the resulting independent components enhance tail risk (kurtosis) when used for factor-risk parity; see Section 4.2.

Independence is a much stronger requirement than decorrelation; it requires decorrelation for every (non-linear) transformation. In particular, independence not only requires the covariance matrix to be diagonal, but all the higher comoments must vanish as well. A standard criterion to measure the "distance to independence" of a random vector is mutual information.

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

$$I(\boldsymbol{Y}) := \left\langle f_{\boldsymbol{Y}}, \prod_{i=1}^{K} f_{Y_i} \right\rangle = \int_{\boldsymbol{y}} f_{\boldsymbol{Y}}(\boldsymbol{y}) \ln \frac{f_{\boldsymbol{Y}}(\boldsymbol{y})}{\prod_{i=1}^{K} f_{Y_i}(y_i)} \mathrm{d}\boldsymbol{y}. \tag{2}$$

Mutual information is always nonnegative, and it is zero if and only if the components of $\boldsymbol{Y}$ are mutually independent; see Cover and Thomas (2006, p.253). Thus, one can extend the notion of *uncorrelated factors* to that of *independent components* by using mutual information instead of linear correlation as dependence measure.

Definition 3 (ICA)**.** The *independent components* (ICs) are the rotation of the PCs that minimize the mutual information:

$$\boldsymbol{Y}^{\dagger} := \boldsymbol{R}^{\dagger} \boldsymbol{Y}^{\star}, \quad \boldsymbol{R}^{\dagger} \in \mathcal{R}_I, \tag{3}$$

where $\mathcal{R}_I$ is the set

$$\mathcal{R}_I := \left\{ \operatorname*{argmin}_{\boldsymbol{R} \in \mathcal{SO}(K)} I(\boldsymbol{R}\boldsymbol{Y}^{\star}) \right\}. \tag{4}$$

Observe that mutual information is invariant with respect to a permutation or change of sign of the factors. Therefore, just like the set of PCs, the set of ICs are determined only up to a change of sign and permutation. As we show in Section 3, this indeterminacy is irrelevant for factor-risk-parity purposes. Moreover, in general, the ICs may not be perfectly independent; they are the *maximally independent* factors found by a linear transformation of the asset returns.

In what follows, we assume that at most one of the K principal components is Gaussian. This is a mild assumption because it holds in the data. In particular, for the equity return data used in

Section 6, the PCs have significantly positive excess kurtosis, and thus, are not Gaussian. The reason we require this assumption is that, if more than one PC is Gaussian, any rotation of the Gaussian PCs remains jointly Gaussian and uncorrelated—hence, mutually independent—and are thus ICs. Therefore, in this case, the ICs are undetermined, even up to a change of sign and permutation. This is because higher-moment information is needed to discriminate the maximally independent factors, and thus ICA is not useful in the purely Gaussian case. If the non-Gaussianity assumption is violated, the ICs are arbitrary like PCs, *but only in the subset of purely Gaussian PCs*.

The ICs are computed using *independent component analysis* (ICA), a well-known technique in machine learning; see Hyvärinen et al. (2001) and Vrins (2007). In general, the rotation matrix $\boldsymbol{R}^\dagger$ associated with the ICs is found via adaptive techniques such as gradient descent. To that aim, we rely on the *FastICA* algorithm of Hyvärinen (1999), a robust algorithm whose optimization criterion is derived from the mutual information and that provides $\boldsymbol{R}^\dagger$ in a fraction of a second even in high dimensions. Details about this algorithm are provided in the e-companion.

2.3. Numerical illustration

In Examples 1 and 2 below, we illustrate the arbitrariness of the decorrelation criterion to select a set of factors, contrary to the independence (mutual information) criterion. We focus on the case $K = N = 2$ for graphical illustration purposes. We parametrize the rotation matrix $\boldsymbol{R} \in \mathcal{SO}(2)$ as a function of the rotation angle θ:

$$\boldsymbol{R} = \boldsymbol{R}(\theta) := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in \mathcal{SO}(2). \tag{5}$$

The examples are illustrated with polar plots, in which the standard cartesian coordinates (x, y) are replaced by polar coordinates (θ, r), where θ is the angle in (5) and r is the value taken by the variable of interest, measured by the radial distance from the origin $(\theta, 0)$.

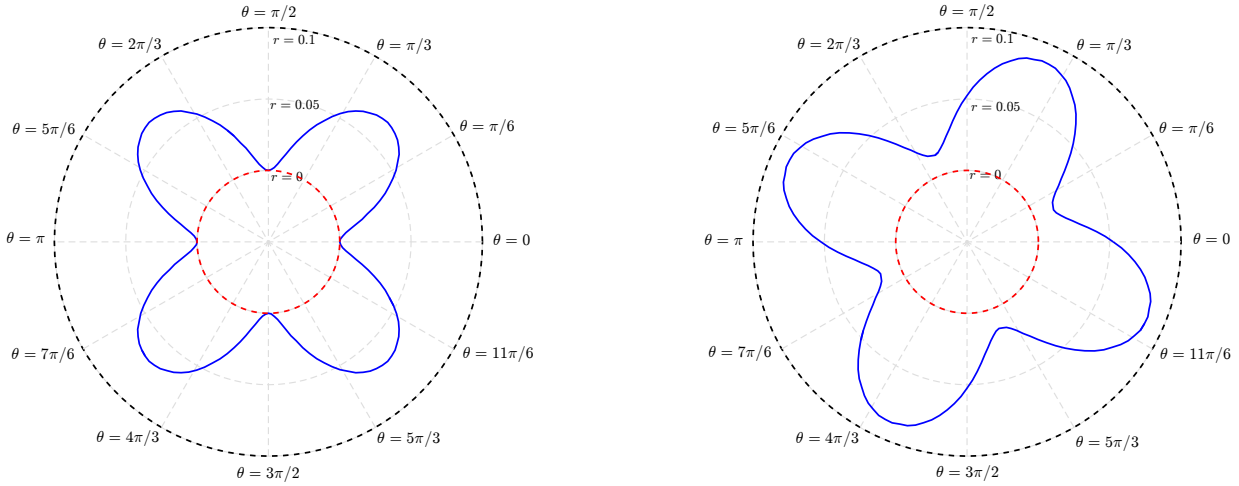
Example 1 (Independent non-Gaussian asset returns)**.** Consider two assets with returns independently distributed as Student t distributions: $X_1 \sim T(\nu_1 = 3)$, $X_2 \sim T(\nu_2 = 20)$. Because the two asset returns are uncorrelated and X_1 has a higher variance than X_2, the PCs coincide with the standardized asset returns; that is, $\boldsymbol{Y}^\star = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{X}$ up to sign and permutation. Figure 1a depicts how the relative mutual information[1] and correlation of the factors $\boldsymbol{Y} = \boldsymbol{R}(\theta)\boldsymbol{Y}^\star$, obtained by

[1]The relative mutual information is a standardized version of the mutual information taking values in $[0, 1]$ defined as $\bar{I}(\boldsymbol{Y}) := KI(\boldsymbol{Y})/\sum_{i=1}^K h(Y_i)$, where $h(Y_i) := -\mathbb{E}[\ln f_{Y_i}(Y_i)]$ is the Shannon entropy of Y_i.

Figure 1 Mutual information versus correlation

(a) Independent non-Gaussian asset returns

(b) Dependent non-Gaussian asset returns

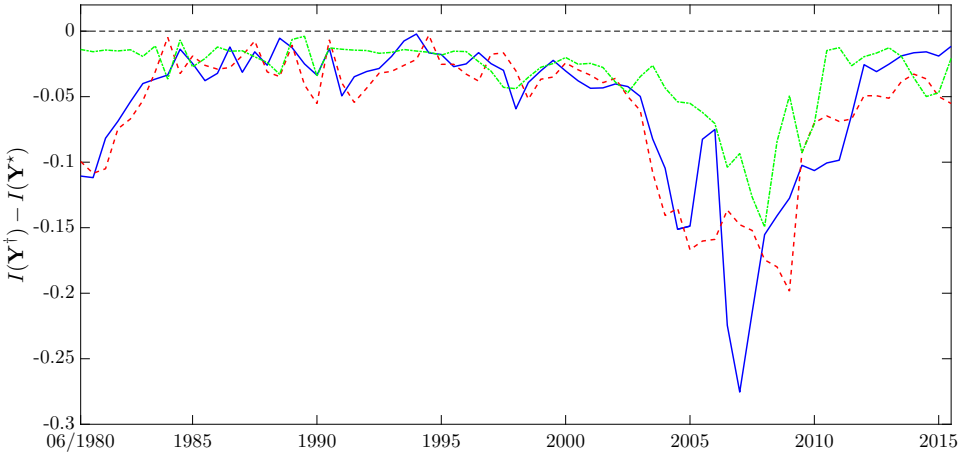


Notes. The figure depicts the relative mutual information (in solid blue) and correlation (in dashed red) of the $K = N = 2$ factors $\boldsymbol{Y}$ obtained as a rotation of the principal components, $\boldsymbol{Y} = \boldsymbol{R}(\theta)\boldsymbol{Y}^\star$. Figure 1a considers the case with independent 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 (θ, r), where θ is the angle determining the rotation matrix $\boldsymbol{R}(\theta)$ defined in (5) and r is the radial distance from the origin $(\theta, 0)$. The principal components correspond to $\theta = 0$. Note that, for clarity of the figures, the center of the polar plots corresponds to $r = -0.05$ rather than $r = 0$.

rotating the PCs, depend on the rotation angle θ. Observe that the correlation is zero for all θ, but that the relative mutual information is zero only when $\theta = k\pi/2$, $k \in \mathbb{Z}$; that is, the ICs $\boldsymbol{Y}^\dagger$ coincide with the asset returns up to sign and permutation. Thus, while the decorrelation criterion is not discriminating, the independence criterion results in a unique set of factors.

Example 2 (Dependent non-Gaussian asset returns)**.** Consider now a joint distribution of two asset returns obtained by mixing, via a Gaussian copula of correlation $\rho = 0.2$, two independent Student t random variables with degrees of freedom $(\nu_1, \nu_2) = (3, 20)$. The PCs $\boldsymbol{Y}^\star$ are obtained by rotating the two asset returns by an angle of -0.17 radians; that is, $\boldsymbol{V} = \boldsymbol{R}(-0.17)$. Figure 1b depicts how the relative mutual information and correlation of the factors $\boldsymbol{Y} = \boldsymbol{R}(\theta)\boldsymbol{Y}^\star$, obtained by rotating the PCs, depend on the rotation angle θ. Independence is never achieved because the asset returns are nonlinear functions of random variables with non-Gaussian distributions. Nevertheless, the mutual information varies with θ and this allows one to discriminate between the continuum of uncorrelated factors. In particular, we find that the ICs are retrieved for $\theta^\dagger = 0.35 + k\pi/2 \neq 0$, $k \in \mathbb{Z}$, and thus, the ICs $\boldsymbol{Y}^\dagger = \boldsymbol{R}(\theta^\dagger)\boldsymbol{Y}^\star$ differ from the PCs $\boldsymbol{Y}^\star$.

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, $I(\boldsymbol{Y}^\dagger) - I(\boldsymbol{Y}^\star)$, 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.

Finally, to gauge the practical relevance of the ICs, we use three empirical datasets to evaluate the gain in independence that can be achieved by moving from the PCs to the ICs. 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, measured by $I(\boldsymbol{Y}^\dagger) - I(\boldsymbol{Y}^\star)$, 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-variance parity via uncorrelated factors

Let us introduce some additional notation. Let $P := \boldsymbol{w}'\boldsymbol{X}$ be the portfolio return, where $\boldsymbol{w}$ is the vector of portfolio weights—simply called *portfolio* hereafter—that we assume belongs to the set

$$\mathcal{W} := \{\boldsymbol{w} \in \mathbb{R}^N \mid \mathbf{1}_N'\boldsymbol{w} = 1\}, \tag{6}$$

where $\mathbf{1}_N$ is the N-dimensional vector of ones. Given a random variable X, we denote $\mu(X)$ the mean, $m_k(X)$ the kth central moment, $\zeta(X) := m_3(X)/m_2(X)^{3/2}$ the skewness and $\kappa(X) := m_4(X)/m_2(X)^2 - 3$ the excess kurtosis.

The factor-variance-parity (FVP) portfolio is the portfolio for which each uncorrelated factor contributes equally to its return variance. In Section 5, we tackle the case of higher-moment risk

measures. The original proposal of Meucci (2009) is to diversify the portfolio-return variance on the PCs. As discussed in Section 2.1, this choice appears arbitrary because the only motivation for using the PCs is that they are uncorrelated, but any rotation of the PCs remains uncorrelated, and is thus equally valid for FVP purposes. We divide this section in two parts. First, we derive the FVP portfolio for an arbitrary rotation of the PCs. Second, we formally establish the arbitrariness of the decorrelation criterion by showing that any portfolio is a FVP portfolio on a set of factors obtained by some rotation of the PCs.

3.1. Derivation of the factor-variance-parity portfolio

We consider below factors given by an arbitrary rotation of the K principal components, $\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{Y}^\star$. Contrary to Meucci (2009) and Deguest et al. (2013), we do not restrict to the full-dimensional case $K = N$. Allowing for $K \leq N$ reduces dimension and alleviates the impact of estimation error (Hastie et al. 2001, Kozak et al. 2019).

To proceed, we rewrite the portfolio return as a linear combination of the factors $\boldsymbol{Y}$. Defining $\tilde{P} := \tilde{\boldsymbol{w}}(\boldsymbol{R})'\boldsymbol{Y}$ with $\tilde{\boldsymbol{w}}(\boldsymbol{R})$ the factor exposures, we have that $\tilde{P} = \tilde{\boldsymbol{w}}(\boldsymbol{R})'\boldsymbol{R}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}'\boldsymbol{X}$, and thus, $\boldsymbol{w}$ is retrieved from the factor exposures $\tilde{\boldsymbol{w}}(\boldsymbol{R})$ as

$$\boldsymbol{w} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\tilde{\boldsymbol{w}}(\boldsymbol{R}). \tag{7}$$

Taking the pseudoinverse leads to the following definition of $\tilde{\boldsymbol{w}}(\boldsymbol{R})$:

$$\tilde{\boldsymbol{w}}(\boldsymbol{R}) := \boldsymbol{R}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}'\boldsymbol{w}. \tag{8}$$

Consistent with the requirement $\boldsymbol{w} \in \mathcal{W}$, the factor exposures $\tilde{\boldsymbol{w}}(\boldsymbol{R})$ belong to

$$\widetilde{\mathcal{W}}(\boldsymbol{R}) := \big\{\tilde{\boldsymbol{w}}(\boldsymbol{R}) \in \mathbb{R}^K \ | \boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\tilde{\boldsymbol{w}}(\boldsymbol{R}) \in \mathcal{W}\big\}. \tag{9}$$

Because we reduce dimension, we naturally have that

$$\tilde{P} = \tilde{\boldsymbol{w}}(\boldsymbol{R})'\boldsymbol{Y} = \boldsymbol{w}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{X} \neq P,$$

except when $K = N$. We call $\tilde{P}$ the *reduced portfolio return*.

Observe now that because $\boldsymbol{\Sigma}_{\boldsymbol{Y}} = \boldsymbol{I}_K$, the reduced-portfolio-return variance is simply given by

$$m_2(\tilde{P}) = \sum_{i=1}^{K} \tilde{w}_i(\boldsymbol{R})^2 = \|\tilde{\boldsymbol{w}}(\boldsymbol{R})\|^2, \tag{10}$$

which allows us to define the FVP portfolio as follows.

Definition 4 (Factor-variance-parity portfolio)**.** A *factor-variance-parity (FVP) portfolio* associated with the factors $\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{Y}^\star$ is a portfolio for which $\tilde{w}_i(\boldsymbol{R})^2 = \tilde{w}_j(\boldsymbol{R})^2$ for all $i, j = 1, \dots, K$.

In the next proposition, we give an analytical expression for the FVP portfolios.

Proposition 1. *Given a rotation matrix $\boldsymbol{R} \in \mathcal{SO}(K)$, the following holds:*

(i) There are 2^{K-1} factor-variance-parity portfolios with respect to the factors $\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{Y}^\star$, given by the set

$$\mathcal{W}_{FVP}(\boldsymbol{R}) := \left\{ \boldsymbol{w} \in \mathcal{W} \;\middle|\; \boldsymbol{w} = \frac{\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\mathbf{1}_K^{\pm}}{\mathbf{1}_N'\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\mathbf{1}_K^{\pm}} \right\}, \tag{11}$$

where $\mathbf{1}_K^{\pm}$ is any K-dimensional vector of ± 1's.

(ii) The set of portfolios $\mathcal{W}_{FVP}(\boldsymbol{R})$ is invariant to a change of sign and permutation of rows of $\boldsymbol{R}$: $\mathcal{W}_{FVP}(\boldsymbol{R}) = \mathcal{W}_{FVP}(\boldsymbol{B}\boldsymbol{R})$ for all rotation matrices $\boldsymbol{B}$ of the form $\boldsymbol{B} = \boldsymbol{I}_K^{\pm}\boldsymbol{\Pi}$ with $\boldsymbol{I}_K^{\pm}$ a diagonal matrix of ± 1's and $\boldsymbol{\Pi}$ a permutation matrix;

(iii) The factor-variance-parity portfolio with minimum return variance is unique and obtained by considering the particular case $\mathbf{1}_K^{\pm} \leftarrow \mathbf{1}_K^{MV}$ in (11), where

$$\mathbf{1}_K^{MV} := \operatorname{sign}(\boldsymbol{R}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}'\mathbf{1}_N). \tag{12}$$

Point *(i)* shows that there are 2^{K-1} FVP portfolios, which is a natural consequence of the fact that a portfolio is a FVP portfolio if all the *squared* factor exposures are equal; their signs are irrelevant. This holds for any choice of factors, including the PCs and ICs. Point *(ii)* shows that the sign and order of the considered factors is irrelevant for factor-variance parity. Point *(iii)* shows that there is a single FVP portfolio minimizing the portfolio-return variance, which allows us to choose one of the 2^{K-1} FVP portfolios.

3.2. Arbitrariness of the factor-decorrelation criterion

As we now show, the factor-decorrelation criterion is arbitrary for FVP purposes because *any* portfolio is a FVP portfolio for factors corresponding to some rotation of the PCs.

Proposition 2. *Let $K = N$. Then, given any portfolio $\boldsymbol{w} \in \mathcal{W}$, there exists a rotation matrix $\boldsymbol{R} \in \mathcal{SO}(K)$ for which $\boldsymbol{w}$ is a FVP portfolio; that is, for which $\boldsymbol{w} \in \mathcal{W}_{FVP}(\boldsymbol{R})$.*

Note that this proposition does not depend on the multiplicity of solutions in (11); it holds for any fixed choice of vector $\mathbf{1}_K^{\pm}$. Thus, although decorrelation is sufficient to decompose the portfolio-return variance as a sum of individual factor contributions in (10), decorrelation alone is not sufficient to discriminate among all the possible portfolios. When $K < N$, Proposition 2 no longer holds: the rotation matrix $\boldsymbol{R}$ does not give enough degrees of freedom to generate any portfolio. Still, the FVP portfolio continues to depend on the rotation $\boldsymbol{R}$ of PCs chosen, and thus, the choice of uncorrelated factors remains arbitrary.

Choosing $\boldsymbol{R} = \boldsymbol{I}_K$ leads to a unique set of PC-variance-parity (PCVP) portfolios, which consists of taking the PCs as factors, but this constitutes an arbitrary choice. As explained in Section 2.1, the only motivation to use PCs for factor-risk parity is that they are uncorrelated, but there are infinitely many rotations of the PCs sharing this feature, which, as shown in Proposition 2, lead to all possible portfolios. It is therefore natural to ask ourselves whether the PCs provide the best factors for FVP purposes. We answer this question in Section 4.2 in the context of higher moments. Specifically, we show that diversifying the portfolio-return variance on ICs helps to reduce kurtosis, but that diversifying it on merely uncorrelated factors, such as the PCs, does not.

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

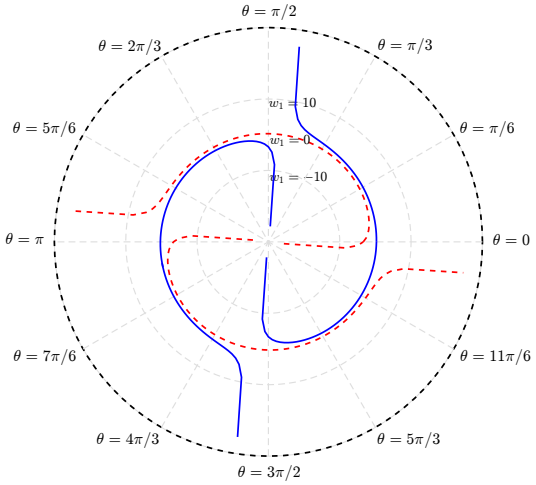
Example 3 (Arbitrariness of decorrelation for FVP)**.** We generate 1000 observations from the asset-return distribution considered in Example 2. Proposition 1 shows that there are two different FVP portfolios for each rotation of the PCs $\boldsymbol{Y} = \boldsymbol{R}(\theta)\boldsymbol{Y}^{\star}$ when $K = 2$. Figure 3a depicts the weight on the first asset, w_1, of the two FVP portfolios as a function of the rotation angle θ. In line with Proposition 2, w_1 can take any real value for each of the two FVP portfolios. In other words, any portfolio $\boldsymbol{w} = (w_1, 1 - w_1)'$ is a FVP portfolio for some θ. However, if for each θ we choose among the two FVP portfolios the one 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 FVP portfolio with minimum return variance.

4. Higher-moment diversification via independent component analysis

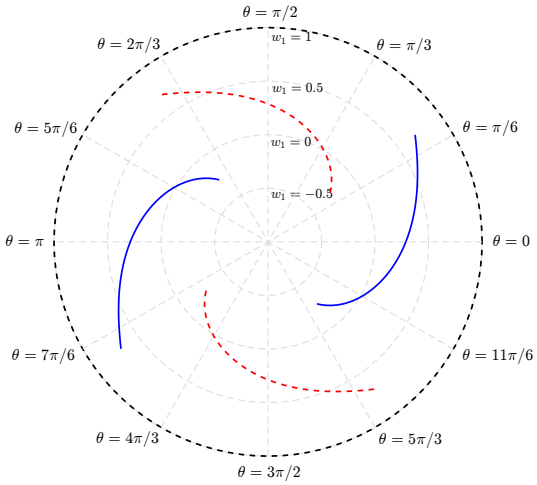
In Section 3, we showed that decorrelation is not a discriminant criterion for factor-variance parity; any portfolio is a FVP portfolio for some rotation of the PCs. Going from decorrelation to independence by relying on the ICs resolves this indeterminacy. In this section, we first show that

Figure 3 Arbitrariness of the factor-variance-parity portfolio

(a) Weight on first asset of the two FVP portfolios



(b) Weight on first asset of FVP portfolio with minimum return variance



Notes. The figures are generated for the $N = 2$ asset returns in Example 3. We consider $K = 2$ factors $\boldsymbol{Y}$ given by some rotation of the principal components (PCs), $\boldsymbol{Y} = \boldsymbol{R}(\theta)\boldsymbol{Y}^\star$. From Proposition 2, there exist two factor-variance-parity (FVP) portfolios for any rotation of the PCs. Figure 3a depicts the weight on the first asset return, w_1, of the two FVP portfolios (in solid blue and dashed red, respectively) as a function of the rotation angle θ. Figure 3b depicts the weight on the first asset return, w_1, of the FVP portfolio with minimum return variance as a function of the rotation angle θ. The plots are in polar coordinates (θ, r), where θ is the angle determining the rotation matrix $\boldsymbol{R}(\theta)$ defined in (5) and $r = w_1$ is the radial distance from the origin $(\theta, 0)$. The PCs correspond to $\theta = 0$. Note that r can be negative as it represents the portfolio weight w_1.

diversifying variance across ICs provides a natural way of reducing the kurtosis of portfolio returns. In contrast, the PCs are only based on the asset-return covariance matrix, and thus, ignore higher moments (such as kurtosis) in asset returns. Second, we derive the conditions under which the tangent mean-variance portfolio is an IC-variance-parity portfolio. Finally, we introduce a shrinkage portfolio that combines the minimum-variance portfolio and the IC-variance-parity portfolio.

4.1. IC-variance-parity portfolios

The ICs are obtained by finding the rotation matrix $\boldsymbol{R}^\dagger$ in (3) such that the factors $\boldsymbol{Y}^\dagger = \boldsymbol{R}^\dagger\boldsymbol{Y}^\star$ are as independent as possible, according to the mutual-information criterion. Thus, the IC-variance-parity portfolios are particular FVP portfolios.

Definition 5 (IC-variance-parity portfolios)**.** The set of 2^{K-1} *IC-variance-parity (ICVP) portfolios* is given by $\mathcal{W}_{FVP}(\boldsymbol{R}^\dagger)$ in (11).

From Proposition 1, the set of ICVP portfolios is unaffected by the sign-and-permutation indeterminacy of the ICs. Thus, under the identifiability assumption explained in Section 2.2 that

at most one of the PCs is Gaussian, relying on the ICVP portfolios addresses the arbitrariness inherent in the factor-decorrelation criterion.

4.2. Kurtosis reduction

We now show that using ICs for variance parity is better than using merely uncorrelated factors, such as the PCs, because it helps to reduce the kurtosis of portfolio returns.

Proposition 3. *Let the N asset returns be given by a linear combination of K independent factors with strictly positive excess kurtosis. Let the arithmetic and harmonic means be $A(x_1,\dots,x_K) := \frac{1}{K}\sum_{i=1}^{K} x_i$ and $H(x_1,\dots,x_K) := K\big(\sum_{i=1}^{K} 1/x_i\big)^{-1}$, respectively. Then, the following holds:*

(i) The minimum and maximum return excess kurtosis achievable by any portfolio are

$$\kappa_{\min} = \frac{1}{K} H\big(\kappa\big(Y_1^\dagger\big),\dots,\kappa\big(Y_K^\dagger\big)\big) \text{ and} \tag{13}$$

$$\kappa_{\max} = \max\big(\kappa\big(Y_1^\dagger\big),\dots,\kappa\big(Y_K^\dagger\big)\big), \tag{14}$$

where $\kappa\big(Y_i^\dagger\big)$ is the excess kurtosis of the ith independent component.

(ii) The 2^{K-1} IC-variance-parity portfolios have an identical return excess kurtosis, equal to the arithmetic mean of the excess kurtosis of the ICs divided by K,

$$\kappa_{IC} = \frac{1}{K} A\big(\kappa\big(Y_1^\dagger\big),\dots,\kappa\big(Y_K^\dagger\big)\big). \tag{15}$$

(iii) The excess kurtosis of any FVP portfolio return, other than an ICVP one, can take any possible value in $[\kappa_{\min},\kappa_{\max}]$ as a function of $\boldsymbol{R}^\dagger$.

The intuition behind Proposition 3 comes from the central limit theorem, which tells us that the asymptotic distribution of a sum of standardized *independent* random variables with the same variance is Gaussian, and therefore has zero excess kurtosis. Though the excess kurtosis becomes zero only asymptotically, the excess kurtosis of the ICVP portfolio return, which is an equally weighted sum of K independent components with unit variance, moves toward zero as a function of K. This argument cannot be applied to the PC-variance-parity portfolios because the PCs are not independent, and thus, the central limit theorem does not apply.

There are two important implications from Proposition 3. First, concomitantly with diversifying the portfolio-return variance over independent factors, one obtains a kurtosis that is close to the minimum one. In particular, the excess kurtosis of the IC-variance-parity portfolios is equal to

the minimum one, $\kappa_{IC} = \kappa_{\min}$, if the excess kurtosis of all ICs are equal, and it remains close to the minimum excess kurtosis as long as the excess kurtosis of the ICs are not very different from one another.[2] Note that the ICVP portfolios share the same kurtosis because the kurtosis of a weighted sum of independent factors only depends on the *squared* weights. This is no longer true for non-independent factors.

Second, by diversifying the portfolio-return variance over non-independent factors, we do not have any control over kurtosis. Specifically, depending on the rotation matrix $\boldsymbol{R}^\dagger$ that is data-driven and over which we know nothing a priori, the FVP portfolio return can have *any* kurtosis in the range $[\kappa_{\min}, \kappa_{\max}]$: the FVP portfolio has no specific benefit in terms of kurtosis. This is true in particular for the PCVP portfolios and is a fundamental difference with the kurtosis of the ICVP portfolios which does not depend on $\boldsymbol{R}^\dagger$.

Thus, the ICVP portfolio strategy offers a natural way of reducing the tail risk of portfolio returns, which is an important concern for many investors; see Dittmar (2002) and Ang et al. (2006). Moreover, the ICVP portfolio accounts for kurtosis in a parsimonious way: in addition to the asset-return covariance matrix, one does not need to estimate the large-dimensional cokurtosis matrix, but merely the ICA rotation matrix $\boldsymbol{R}^\dagger$ made of $K(K-1)/2$ distinct parameters only.

The following example illustrates Proposition 3 for the case $K = 2$.

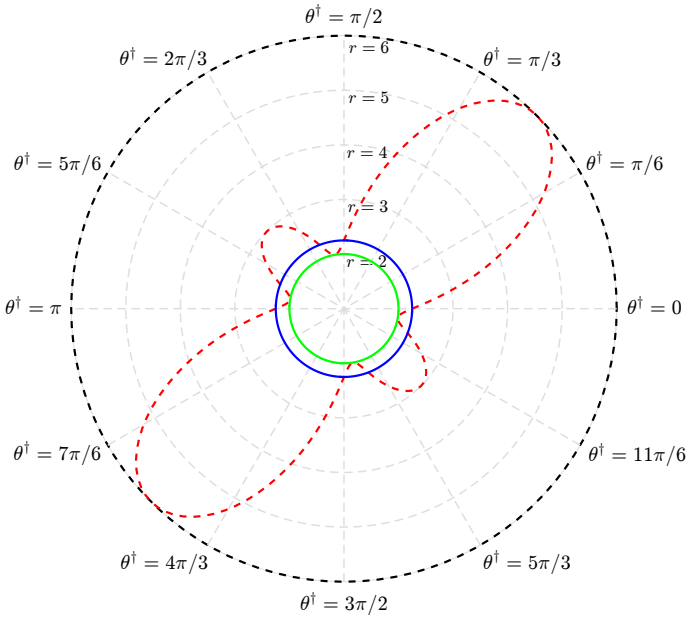
Example 4 (Kurtosis of ICVP versus PCVP portfolios)**.** When $K = 2$, the excess kurtosis of the two PCVP portfolios is given by

$$\kappa_{PC} \in \frac{1}{8}\left(\pm 4 \sin 2\theta^\dagger \big(\kappa\big(Y_1^\dagger\big) - \kappa\big(Y_2^\dagger\big)\big) + (3 - \cos 4\theta^\dagger)\big(\kappa\big(Y_1^\dagger\big) + \kappa\big(Y_2^\dagger\big)\big)\right), \tag{16}$$

where $\boldsymbol{R}^\dagger = \boldsymbol{R}(\theta^\dagger)$. In line with point *(iii)* of Proposition 3, it is easy to check that κ_{PC} spans the range $[\kappa_{\min}, \kappa_{\max}]$ as a function of $\theta^\dagger$. This is shown in Figure 4, which depicts $\kappa_{\min}$, κ_{IC} and κ_{PC} (with the plus sign in (16)) as a function of $\theta^\dagger$ for $\big(\kappa\big(Y_1^\dagger\big), \kappa\big(Y_2^\dagger\big)\big) = (3, 6)$. By chance, the kurtosis of the PCVP portfolio could be lower than the kurtosis of the ICVP portfolio, but then only slightly so because $\kappa_{IC} = 9/4$ is very close to $\kappa_{\min} = 2$. In contrast, it can also be much higher because it takes values in the range $[\kappa_{\min}, \kappa_{\max}] = [2, 6]$.

[2] For example, taking $K = 3$ and $\big(\kappa\big(Y_1^\dagger\big), \kappa\big(Y_2^\dagger\big), \kappa\big(Y_3^\dagger\big)\big) = (3, 5, 7)$, the minimum excess kurtosis is given by $\kappa_{\min} = 105/71 = 1.48$, versus $\kappa_{IC} = 5/3 = 1.67$ for the ICVP portfolios.

Figure 4 Excess kurtosis of PC- and IC-variance-parity portfolios



Notes. The figure is an illustration of Proposition 3 with $(\kappa(Y_1^\dagger), \kappa(Y_2^\dagger)) = (3, 6)$. It depicts, as a function of the rotation angle $\theta^\dagger$ determining the ICA rotation matrix $\boldsymbol{R}^\dagger = \boldsymbol{R}(\theta^\dagger)$, the excess kurtosis of the minimum-kurtosis portfolio ($\kappa_{\min} = 2$ in solid green), of the two ICVP portfolios ($\kappa_{IC} = 9/4$ in solid blue) and of one of the two PCVP portfolios (κ_{PC} with the plus sign in (16) in dashed red). The plot is in polar coordinates $(\theta^\dagger, r)$, where r is the radial distance from the origin $(\theta^\dagger, 0)$. Note that the center of the polar plot corresponds to $r = 1$.

4.3. Mean-variance efficiency

In this section, we characterize the relation between the ICVP portfolio and the portfolios based on mean-variance utility. In particular, we show that the ICVP portfolio can be seen as a shrinkage estimator of the mean-variance portfolio. To do this, we build on the result by Maillard et al. (2010), who show that the tangent mean-variance portfolio of Markowitz (1952) corresponds to the asset-variance-parity portfolio when all assets have the same Sharpe ratio and pairwise correlations. In our context, this condition reduces to the independent components having the same (absolute) Sharpe ratio.

Proposition 4. *Let $K = N$ and let the return of every independent component have the same Sharpe ratio in absolute value. Then, the mean-variance tangent portfolio is an IC-variance-parity portfolio.*

Note that the above condition for mean-variance efficiency is reasonable because, as explained by Kozak et al. (2019), many asset-pricing models assume that if a factor earns high returns, it

must also be a major source of variance. For example, the Bayesian prior used by Pástor (2000) and Pástor and Stambaugh (2000) is precisely that all factors have the same Sharpe ratio.

4.4. Data-driven shrinkage portfolio

In this section, we introduce the portfolio-selection strategy that we propose to implement when using the variance as risk measure: a shrinkage of the minimum-variance portfolio toward the minimum-variance ICVP portfolio.

Proposition 3 supports the use of the *minimum-variance* ICVP (MV-ICVP) portfolio,

$$\boldsymbol{w}_{IC}^{\star} := \frac{\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}^{\dagger\prime}\mathbf{1}_K^{MV}}{\mathbf{1}_N'\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}^{\dagger\prime}\mathbf{1}_K^{MV}}, \quad \mathbf{1}_K^{MV} := \text{sign}(\boldsymbol{R}^{\dagger}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}'\mathbf{1}_N) \tag{17}$$

out of the 2^{K-1} ICVP portfolios because, under the assumptions of Proposition 3, all the ICVP portfolios have the same return kurtosis but not the same return variance. Thus, the MV-ICVP portfolio helps reduce variance without increasing kurtosis. This resembles Haugh et al. (2017) who propose optimizing the investor's mean-risk utility subject to an asset-risk-budgeting constraint.

The minimum-variance (MV) portfolio,

$$\boldsymbol{w}_{MV}^{\star} := \frac{\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{1}_N}{\mathbf{1}_N'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{1}_N}, \tag{18}$$

has been widely shown to exhibit favorable out-of-sample performance in terms of Sharpe ratio; see, for instance, Jagannathan and Ma (2003), DeMiguel et al. (2009a) and DeMiguel and Nogales (2009). The reason is that it is the only portfolio on the efficient frontier that does not require any estimation of asset mean returns, and thus, it is much less impacted by estimation risk.

Next, we introduce a shrinkage portfolio that combines the two portfolios $\boldsymbol{w}_{MV}^{\star}$ and $\boldsymbol{w}_{IC}^{\star}$.

Definition 6 (ICMV portfolio)**.** The *ICMV portfolio* is defined as the shrinkage portfolio

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

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

We calibrate the shrinkage intensity δ in a data-driven way using the bootstrap methodology of 10-fold cross-validation that avoids assumptions on the asset-return distribution; see Section 6.1.

The empirical results show that, out of sample, the ICMV portfolio achieves an appealing trade-off between the large Sharpe ratio of the MV portfolio and the low kurtosis of the MV-ICVP

portfolio. In contrast, shrinking the MV portfolio toward the minimum-variance PCVP portfolio, which relies on the principal components, does not reduce the return kurtosis as much.

Shrinkage portfolios as in (19) have been considered by several authors in the literature. For instance, Kan and Zhou (2007) shrink the mean-variance portfolio toward the MV portfolio. Perhaps more closely related to our work are DeMiguel et al. (2009b) and Tu and Zhou (2011), who shrink the MV and mean-variance portfolios, respectively, toward the *equally weighted portfolio*. A distinguishing feature of our approach is that we shrink toward a target portfolio for which the *IC exposures are equally weighted*.

5. Factor-risk parity with higher-moment risk measures

In Sections 3 and 4, we showed how to apply ICs to the traditional factor-risk-parity approach based on *variance*. In this section, we show how to apply ICs to diversify the modified Value-at-Risk (MVaR), which is an approximation of the VaR that accounts for the first four 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, the modified VaR accounts for skewness and kurtosis, which we parsimoniously incorporate by exploiting the maximally independent property of the ICs.

5.1. Parsimonious estimation of higher moments with ICs

We now show that relying on the ICs provides a natural way to parsimoniously estimate the portfolio-return higher moments.

Consider a risk measure that depends on the portfolio-return third and fourth moments. Then, diversifying the portfolio-return risk with respect to K uncorrelated factors $\boldsymbol{Y}$ given by some rotation of the PCs, $\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{Y}^\star$, requires estimating the third and fourth moments of the reduced portfolio return $\tilde{P} = \tilde{\boldsymbol{w}}(\boldsymbol{R})'\boldsymbol{Y}$. Briec et al. (2007) show that they can be written concisely as

$$m_3(\tilde{P}) = \tilde{\boldsymbol{w}}(\boldsymbol{R})'\boldsymbol{M}_3(\tilde{\boldsymbol{w}}(\boldsymbol{R}) \otimes \tilde{\boldsymbol{w}}(\boldsymbol{R})), \tag{20}$$

$$m_4(\tilde{P}) = \tilde{\boldsymbol{w}}(\boldsymbol{R})'\boldsymbol{M}_4(\tilde{\boldsymbol{w}}(\boldsymbol{R}) \otimes \tilde{\boldsymbol{w}}(\boldsymbol{R}) \otimes \tilde{\boldsymbol{w}}(\boldsymbol{R})), \tag{21}$$

where $\boldsymbol{M}_3 := \mathbb{E}\big[(\boldsymbol{Y} - \mu(\boldsymbol{Y}))(\boldsymbol{Y} - \mu(\boldsymbol{Y}))' \otimes (\boldsymbol{Y} - \mu(\boldsymbol{Y}))'\big]$ and $\boldsymbol{M}_4 := \mathbb{E}\big[(\boldsymbol{Y} - \mu(\boldsymbol{Y}))(\boldsymbol{Y} - \mu(\boldsymbol{Y}))' \otimes (\boldsymbol{Y} - \mu(\boldsymbol{Y}))' \otimes (\boldsymbol{Y} - \mu(\boldsymbol{Y}))'\big]$ are the coskewness and cokurtosis matrices of $\boldsymbol{Y}$ and the kronecker

product $\otimes$ between two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is

$$\boldsymbol{A} \otimes \boldsymbol{B} := \begin{bmatrix} A_{11}\boldsymbol{B} & \dots & A_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ A_{m1}\boldsymbol{B} & \dots & A_{mn}\boldsymbol{B} \end{bmatrix}. \tag{22}$$

Boudt et al. (2015) show that estimating $\boldsymbol{M}_3$ and $\boldsymbol{M}_4$ requires estimating a large number of parameters:

$$\frac{K(K+1)(K+2)(K+7)}{24} = \mathcal{O}(K^4).$$

To develop a more parsimonious estimation approach, we rely on the ICs $\boldsymbol{Y}^\dagger = \boldsymbol{R}^\dagger \boldsymbol{Y}^\star$ and assume that they are truly independent. In this case, all comoments of ICs vanish, and estimating the matrices $\boldsymbol{M}_3$ and $\boldsymbol{M}_4$ requires estimating only $2K = \mathcal{O}(K)$ parameters, which correspond to the third and fourth central moments of the K independent components.

Proposition 5. *Let the K ICs be independent and have finite fourth central moments, $\boldsymbol{Y}^\dagger \in \mathcal{L}^4$. Then, the third and fourth central moments of the reduced portfolio return are given by*

$$m_3(\tilde{P}) = \sum_{i=1}^{K} \tilde{w}_i (\boldsymbol{R}^\dagger)^3 m_3(Y_i^\dagger), \tag{23}$$

$$m_4(\tilde{P}) = \sum_{i=1}^{K} \tilde{w}_i (\boldsymbol{R}^\dagger)^4 m_4(Y_i^\dagger) + 3 \sum_{i=1}^{K} \sum_{j \neq i}^{K} (\tilde{w}_i (\boldsymbol{R}^\dagger) \tilde{w}_j (\boldsymbol{R}^\dagger))^2. \tag{24}$$

In case the ICs are not truly independent, the approximations in (23) and (24) correspond to shrinking the IC higher comoments to zero, resulting in a bias that is expected to be largely offset by the benefits in terms of estimation risk.[3]

5.2. IC-risk parity with modified Value-at-Risk

Proposition 5 allows one to develop parsimonious estimators of factor-risk-parity portfolios with higher-moment risk measures. We show how to do this for an approximation of the popular VaR risk measure. The difficulty in using the standard VaR is that contrary to the central moments in (23) and (24), the quantile of a linear combination of independent random variables is not an affine function of the individual quantiles.

[3]For example, Martellini and Ziemann (2010) introduce estimators of the coskewness and cokurtosis matrices that shrink the sample estimator toward a target estimator obtained under the assumption of a constant correlation or a single-factor model. Empirically, they find that the shrinkage intensity minimizing the mean-squared estimation error is around 80-90% on average for the cokurtosis matrix, and nearly 100% for the coskewness matrix. This shows that higher-comoment matrices are highly tainted by estimation error, supporting our parsimonious estimation approach.

To circumvent this difficulty, it is useful to employ the Cornish-Fisher expansion of the reduced-portfolio-return quantile at confidence level α (Cornish and Fisher 1938),

$$Q_\alpha(\tilde{P}, r) := \mu(\tilde{P}) + \sqrt{m_2(\tilde{P})} \left(z_\alpha + \sum_{i=1}^{r} p_i(z_\alpha) \right), \tag{25}$$

where z_α is the standard Gaussian quantile and the p_i's are polynomials whose coefficients depend on the standardized moments of $\tilde{P}$. The exact value of the quantile is given by $Q_\alpha(\tilde{P}, \infty)$, which requires estimating all the reduced-portfolio-return moments. To avoid doing this, it is common to truncate the infinite sum and choose $r < \infty$. Choosing $r = 0$ corresponds to the Gaussian VaR, studied by Alexander and Baptista (2002) and Haugh et al. (2017). We choose $r = 2$, which allows us to capture the skewness and kurtosis, and yields the modified VaR of Favre and Galeano (2002).[4]

Definition 7 (Modified VaR)**.** The *modified VaR* (MVaR) of the reduced portfolio return $\tilde{P} \in \mathcal{L}^4$ is the negative second-order Cornish-Fisher expansion of the quantile function in (25):

$$\mathrm{MVaR}_\alpha(\tilde{P}) := -Q_\alpha(\tilde{P}, 2) = -\mu(\tilde{P}) - \sqrt{m_2(\tilde{P})} \left(z_\alpha + p_1(z_\alpha) + p_2(z_\alpha) \right), \tag{26}$$

where $p_1(x) = \frac{1}{6}(x^2 - 1)\zeta(\tilde{P})$ and $p_2(x) = \frac{1}{24}(x^3 - 3x)\kappa(\tilde{P}) - \frac{1}{36}(2x^3 - 5x)\zeta(\tilde{P})^2$.[5]

The Cornish-Fisher approximation is very useful in our context because it relates the reduced-portfolio-return quantile to its moments, which can easily be differentiated with respect to the factor exposures, allowing an easy computation of the MVaR contribution of each IC. Specifically, because MVaR is positive homogeneous, it can be decomposed as

$$\mathrm{MVaR}_\alpha(\tilde{P}) = \sum_{i=1}^{K} \mathrm{MVaR}_\alpha(Y_i^\dagger | \tilde{P}) := \sum_{i=1}^{K} \tilde{w}_i(\boldsymbol{R}^\dagger) \frac{\partial \mathrm{MVaR}_\alpha(\tilde{P})}{\partial \tilde{w}_i(\boldsymbol{R}^\dagger)}. \tag{27}$$

Boudt et al. (2008) provide closed-form expressions for the MVaR contributions in the case of *assets*. We extend their results to the case of MVaR contributions of *factors* and in particular, of ICs. As shown in Proposition 5, we can do this parsimoniously by assuming the ICs are independent.

Proposition 6. *Let the ICs* $\boldsymbol{Y}^\dagger \in \mathcal{L}^4$ *be independent. Then, the MVaR contribution* $\mathrm{MVaR}_\alpha(Y_i^\dagger | \tilde{P})$

[4] Baillie and Bollerslev (1992) provide evidence that incorporating additional terms barely improves the approximation, while it increases the number of moments to estimate.

[5] For $\alpha = 1\%$ used in the empirical analysis, the MVaR is a sensible higher-moment criterion in the sense that $\frac{\partial \mathrm{MVaR}_\alpha(\tilde{P})}{\partial \zeta(\tilde{P})} \leq 0$ if $\zeta(\tilde{P}) \geq \frac{3(z_\alpha^2 - 1)}{2z_\alpha^3 - 5z_\alpha} = -0.9769$, a mild condition, and $\frac{\partial \mathrm{MVaR}_\alpha(\tilde{P})}{\partial \kappa(\tilde{P})} = \frac{1}{24}(3z_\alpha - z_\alpha^3)\sqrt{m_2(\tilde{P})} \geq 0$.

in (27) *is given by*

$$\begin{aligned}\mathrm{MVaR}_\alpha(Y_i^\dagger|\tilde{P}) = \tilde{w}_i(\boldsymbol{R}^\dagger) \times \Bigg[&- \mu(Y_i^\dagger) - \frac{\partial_i m_2(\tilde{P})}{2\sqrt{m_2(\tilde{P})}}\left(z_\alpha + p_1(z_\alpha) + p_2(z_\alpha)\right) + \\ &\sqrt{m_2(\tilde{P})}\Bigg(- p_1(z_\alpha)\partial_i\zeta(\tilde{P}) - \frac{1}{24}(z_\alpha^3 - 3z_\alpha)\partial_i\kappa(\tilde{P}) + \frac{1}{18}(2z_\alpha^3 - 5z_\alpha)\zeta(\tilde{P})\partial_i\zeta(\tilde{P})\Bigg)\Bigg],\end{aligned} \tag{28}$$

where

$$\partial_i\zeta(\tilde{P}) = \frac{2m_2(\tilde{P})^{3/2}\partial_i m_3(\tilde{P}) - 3m_3(\tilde{P})\sqrt{m_2(\tilde{P})}\partial_i m_2(\tilde{P})}{2m_2(\tilde{P})^3}, \tag{29}$$

$$\partial_i\kappa(\tilde{P}) = \frac{m_2(\tilde{P})\partial_i m_4(\tilde{P}) - 2m_4(\tilde{P})\partial_i m_2(\tilde{P})}{m_2(\tilde{P})^3}, \tag{30}$$

with $m_2(\tilde{P})$, $m_3(\tilde{P})$ *and* $m_4(\tilde{P})$ *given by* (10), (23) *and* (24), *respectively.*[6]

Equipped with Proposition 6, we define the IC-MVaR-parity portfolios.

Definition 8 (IC-MVaR-parity portfolios)**.** An *IC-MVaR-parity portfolio* is a portfolio for which $\mathrm{MVaR}_\alpha(Y_i^\dagger|\tilde{P}) = \mathrm{MVaR}_\alpha(Y_j^\dagger|\tilde{P})$ for all $i, j = 1, \ldots, K$.

Because there may be multiple IC-MVaR-parity portfolios, similar to the variance case, we consider the IC-MVaR-parity portfolio that minimizes MVaR, which we estimate numerically by solving the following optimization problem:

$$\hat{\boldsymbol{w}}_{IC}^\star(\alpha) := \underset{w\in\mathcal{W}}{\operatorname{argmin}}\ \mathrm{MVaR}_\alpha(P) \quad \text{subject to} \quad \left(\sum_{i=1}^K \%\mathrm{MVaR}_\alpha(Y_i^\dagger|\tilde{P})^2\right)^{-1} \geq K - \epsilon, \tag{31}$$

where $\%\mathrm{MVaR}_\alpha(Y_i^\dagger|\tilde{P}) := \frac{\mathrm{MVaR}_\alpha(Y_i^\dagger|\tilde{P})}{\sum_{j=1}^K \mathrm{MVaR}_\alpha(Y_j^\dagger|\tilde{P})}$ and ϵ is a tolerance parameter set to $\epsilon = 5\times10^{-4}$. This is clearly a non-convex optimization problem that we solve using Matlab *GlobalSearch* algorithm (Ugray et al. 2007). Finally, we propose to implement the following shrinkage portfolio.

Definition 9 (ICMVaR portfolio)**.** The *ICMVaR portfolio* is the shrinkage portfolio

$$\boldsymbol{w}_{ICMVaR}^\star(\alpha) := (1-\delta)\boldsymbol{w}_{MMVaR}^\star(\alpha) + \delta\hat{\boldsymbol{w}}_{IC}^\star(\alpha), \tag{32}$$

where $\delta \in [0,1]$ is a given shrinkage intensity, $\boldsymbol{w}_{MMVaR}^\star(\alpha)$ is the minimum-MVaR portfolio[7] and $\hat{\boldsymbol{w}}_{IC}^\star(\alpha)$ is the IC-MVaR-parity portfolio that minimizes MVaR, computed in (31).

[6] We assume that the confidence level α is low enough for the MVaR contributions to be all positive. This assumption typically holds for $\alpha = 5\%$ or below.

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

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, in Section 6.2 we discuss the calibration of parameters, and in Section 6.3 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.

We consider six portfolio policies.[8] First, the minimum-variance asset-variance-parity portfolio (AVP).[9] Second, the minimum-variance portfolio (MV). Third and fourth, the shrinkage portfolio obtained by combining the MV portfolio and the minimum-variance FVP portfolio based on the ICs (ICMV) and PCs (PCMV). Fifth, the minimum-MVaR portfolio (MMVaR). Sixth, the shrinkage portfolio obtained by combining the MMVaR portfolio and the minimum-MVaR factor-MVaR-parity portfolio based on the ICs (ICMVaR).[10] We fix the MVaR confidence level at $\alpha = 1\%$.[11] All portfolios are computed using sample moment averages.

To alleviate the impact of estimation error, we reduce dimension and keep only the first $K \leq N$ PCs. We select K using the minimum-average-partial-correlation method of Velicer (1976). Let $\boldsymbol{R}(K) := \boldsymbol{R}_{\boldsymbol{X}} - \boldsymbol{V}\boldsymbol{V}'$ be the correlation matrix of $\boldsymbol{X}$ after the effect of the first K principal components has been partialed out. Then, we select the $K \in [2, N-1]$ that minimizes $\sum_{i=1}^{N} \sum_{j \neq i}^{N} R_{ij}(K)^2$.

[8] We have also considered the equally weighted portfolio but we do not report the results here because, with daily returns, it is largely outperformed by the other portfolio strategies we consider. In the e-companion, we also compute the MV, ICMV and PCMV portfolios under the no-short-selling constraint and find the conclusions are qualitatively similar, although the performance is worse than that of the unconstrained portfolios because, with daily data, preventing short-selling hurts the out-of-sample performance (Jagannathan and Ma 2003).

[9] Bai et al. (2016) show that there are 2^{N-1} asset-variance-parity portfolios and provide an optimization problem to compute them. For *6BTM* and *6Prof*, we compute all 2^5 solutions and choose the one with minimum return variance. For *30Ind*, it is not feasible to compute all 2^{29} solutions, and thus, we rely instead on the heuristic algorithm of Bai et al. (2016, p.7) that provides a solution close to the minimum-variance asset-variance-parity portfolio with much less computational work.

[10] We use the Matlab function *fmincon* with the *GlobalSearch* option based on Ugray et al. (2007) to compute the MMVaR and ICMVaR portfolios, and we find the ICs with the *FastICA* Matlab package available on Aapo Hyvärinen's website: `www.cs.helsinki.fi/u/ahyvarin/papers/fastica.shtml`.

[11] Two common choices in the literature are $\alpha = 1\%$ and 5%. We choose $\alpha = 1\%$ because Cavenaile and Lejeune (2012) show that setting $\alpha > 4.16\%$ is inconsistent with investors' negative preferences for kurtosis.

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 (a sample size of $T = 1260$). 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). 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 Sharpe ratio of two given portfolios, we compute two-sided p-values with the circular-bootstrapping methods proposed by Ledoit and Wolf (2008) for the Sharpe ratio and Ardia and Boudt (2015) for the modified Sharpe ratio, which are well-suited for returns that have fat tails and non-stationarity. As in DeMiguel et al. (2009a), we use $B = 1000$ bootstrap resamples and block size $b = 5$.[12] The stars $\star$, $\star\star$ and $\star\star\star$ in Table 1 mean that the p-value is less than 20%, 10% and 5%. We compare the following pairs of portfolios: PCMV vs MV (stars in superscript next to PCMV quantity), ICMV vs MV (stars in superscript next to ICMV quantity), ICMV vs PCMV (stars in subscript next to ICMV quantity) and ICMVaR vs MMVaR (stars in superscript next to ICMVaR quantity).

We calibrate the shrinkage intensity δ via 10-fold cross-validation; see Chapter 7 in Hastie et al. (2001). As discussed in Section 4.2, the FVP portfolio based on the ICs is expected to improve upon the MV portfolio in terms of tail risk. Thus, we use as calibration criterion the modified Sharpe ratio, which accounts for higher moments.[13]

[12] The idea behind circular bootstrapping is to wrap time-series data in a circle; that is, to assume $\boldsymbol{X}_{T+1} = \boldsymbol{X}_1$, $\boldsymbol{X}_{T+2} = \boldsymbol{X}_2$ and so on. Then, one randomly (uniformly) chooses a sequence of b consecutive observations on this circle, B amounts of times. We employ the R package `PeerPerformance` to compute the bootstrapped p-values.

[13] 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 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 modified Sharpe ratio.

6.2. Calibration of K and δ

Let us discuss the results from the calibration of the parameters K and δ. As one can observe in Table 1, the calibration method of Velicer (1976) systematically chooses $K = 2$ for the *6BTM* and *6Prof* datasets, which is sensible because they are driven by two dimensions: size and book-to-market or size and operating profitability. For the *30Ind* dataset, which is a higher-dimensional dataset, the average K over the 70 rolling windows is 2.73, with a maximum of $K = 5$ in three windows. These low values of K are 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. (2019). Note also that these PCs are far from being Gaussian, supporting the identifiability assumption in Section 2.2 that at most one of the PCs is Gaussian. Specifically, the Gaussianity test of Jarque and Bera (1980) systematically rejects the Gaussianity of all K principal components for all estimation windows and datasets, with very low associated p-values.

Regarding the shrinkage intensity δ, calibrated by maximizing the 10-fold cross-validation modified Sharpe ratio, we observe in Table 1 that it is higher on average for the ICMV portfolio than the PCMV portfolio. This indicates that the ICs provide a superior target portfolio than the PCs with regards to higher moments. For the ICMVaR portfolio, the optimal δ is quite low because the MMVaR portfolio already minimizes the denominator of the modified Sharpe ratio.

6.3. Results

Table 1 reports the out-of-sample performance of the six portfolio policies for the three datasets. First, we observe that the AVP portfolio does not compare favorably to the MV portfolio. For *6BTM* it is outperformed in terms of Sharpe ratio and modified Sharpe ratio, for *6Prof* it performs very similarly, and for *30Ind* it improves tail risk but has a much worse Sharpe ratio and turnover. These results show that diversifying the portfolio-return variance across assets is not helpful.

We then compare the performance of the ICMV and PCMV portfolios to test whether the shrinkage portfolios based on ICs outperform those based on PCs. The results confirm that portfolios based on the ICs 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 ratio. This result highlights that diversifying the portfolio-return variance with respect to ICs effectively reduces tail risk, in line with

Table 1 Out-of-sample performance of portfolio policies (continued on next page)

6BTM dataset (average $K = 2$)

	AVP	MV	ICMV	PCMV	MMVaR	ICMVaR
Mean	19.92% (2.31%)	19.75% (2.19%)	21.39% (2.28%)	20.84% (2.19%)	23.50% (2.56%)	**23.93%** (2.52%)
Volatility	13.81% (0.35%)	**13.36%** (0.34%)	13.41% (0.27%)	13.60% (0.32%)	15.26% (0.26%)	15.14% (0.25%)
Sharpe ratio	1.44 (0.18)	1.48 (0.18)	**1.59***** (0.18)	1.53 (0.17)	1.54 (0.17)	1.58* (0.18)
Skewness	-0.42 (0.51)	-0.32 (0.53)	-0.33 (0.31)	-0.23 (0.50)	**-0.04** (0.27)	-0.16 (0.20)
Excess kurtosis	21.39 (4.65)	20.51 (4.95)	11.53 (2.40)	18.53 (4.60)	8.72 (2.36)	**7.22** (1.39)
1% modified Value-at-Risk	6.51% (1.05%)	6.08% (0.99%)	4.33% (0.50%)	5.75% (0.95%)	4.13% (0.48%)	**3.84%** (0.33%)
Modified Sharpe ratio ($\times 10^2$)	1.21 (0.29)	1.29 (0.29)	1.96$^{\star\star}_{\star\star}$ (0.34)	1.44** (0.33)	2.26 (0.39)	**2.47*** (0.38)
Portfolio turnover	3.05%	**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 (average $K = 2$)

	AVP	MV	ICMV	PCMV	MMVaR	ICMVaR
Mean	17.78% (2.41%)	17.53% (2.42%)	20.41% (2.57%)	17.83% (2.46%)	22.38% (2.69%)	**22.55%** (2.54%)
Volatility	14.13% (0.37%)	**13.98%** (0.36%)	14.83% (0.32%)	14.41% (0.36%)	16.14% (0.34%)	15.60% (0.31%)
Sharpe ratio	1.26 (0.17)	1.25 (0.17)	1.38$^{\star\star\star}_{\star\star\star}$ (0.17)	1.24 (0.17)	1.39 (0.17)	**1.45*** (0.18)
Skewness	-0.04 (0.64)	**-0.03** (0.62)	-0.35 (0.42)	-0.04 (0.59)	-0.03 (0.46)	-0.25 (0.34)
Excess kurtosis	22.09 (7.52)	22.55 (7.98)	14.86 (4.11)	20.06 (6.90)	14.66 (5.02)	**10.99** (3.12)
1% modified Value-at-Risk	6.63% (1.31%)	6.64% (1.37%)	5.53% (0.97%)	6.32% (1.25%)	5.78% (1.04%)	**4.88%** (0.74%)
Modified Sharpe ratio ($\times 10^2$)	1.06 (0.31)	1.05 (0.30)	1.46$^{\star}_{\star}$ (0.37)	1.12* (0.28)	1.54 (0.37)	**1.83** (0.38)
Portfolio turnover	3.23%	**2.44%**	2.93%	3.09%	5.16%	5.00%
Average shrinkage intensity δ	\	0.00	0.71	0.39	0.00	0.23

the theoretical predictions of Proposition 3. The improvement in Sharpe ratio is statistically significant for *6Prof* and *30Ind*, and the improvement in modified Sharpe ratio for *6BTM* and *6Prof*. Moreover, the ICMV portfolio has lower turnover, which makes its implementation less costly.

Comparing the ICMV and MV portfolios, we observe that there are clear benefits from shrinking the MV portfolio toward the minimum-variance IC-variance-parity portfolio. Naturally, the MV portfolio achieves a lower volatility, but the ICMV portfolio has a substantially higher mean return because the modified-Sharpe-ratio calibration criterion favors a higher mean return in the numerator. This results in a Sharpe ratio that is systematically larger for the ICMV portfolio. In terms of tail risk, the ICMV portfolio also outperforms the MV portfolio in terms of kurtosis, modified VaR and modified Sharpe ratio, as expected. The improvement in the Sharpe ratio and

Table 1 Out-of-sample performance of portfolio policies (continued)

30Ind dataset (average $K = 2.73$)

	AVP	MV	ICMV	PCMV	MMVaR	ICMVaR
Mean	9.66%	11.18%	**15.14%**	10.58%	10.22%	11.70%
	(2.31%)	(1.95%)	(2.49%)	(2.19%)	(2.27%)	(2.28%)
Volatility	13.34%	**11.71%**	14.60%	12.92%	13.29%	14.14%
	(0.27%)	(0.31%)	(0.29%)	(0.30%)	(0.34%)	(0.33%)
Sharpe ratio	0.72	0.95	**1.04**$_{\star\star}$	0.82*	0.77	0.83
	(0.17)	(0.17)	(0.17)	(0.16)	(0.17)	(0.17)
Skewness	**-0.47**	-0.74	-0.49	-0.56	-0.97	-0.78
	(0.45)	(0.76)	(0.42)	(0.58)	(0.75)	(0.67)
Excess kurtosis	13.45	23.99	**12.95**	16.96	21.59	17.46
	(5.67)	(12.56)	(5.75)	(8.76)	(14.13)	(11.20)
1% modified Value-at-Risk	**4.78%**	6.06%	5.11%	5.32%	6.43%	5.97%
	(1.27%)	(2.30%)	(1.36%)	(1.85%)	(2.80%)	(2.54%)
Modified Sharpe ratio ($\times 10^2$)	0.80	0.73	**1.18**	0.79	0.63	0.78
	(0.36)	(0.46)	(0.43)	(0.43)	(0.55)	(0.56)
Turnover	6.75%	**2.45%**	2.52%	2.53%	4.35%	3.99%
Average shrinkage intensity δ	\	0.00	0.42	0.27	0.00	0.22

Notes. This table reports the out-of-sample performance of the six 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. The number of PCs, K, is selected via the minimum-average-partial-correlation method of Velicer (1976). Bold figures indicate the best strategy for each criterion. Figures in parentheses indicate the standard errors, computed via non-parametric bootstrap with Matlab *bootstat* function and $B = 1000$ bootstrap resamples. We evaluate the statistical significance of the difference in Sharpe ratio and the modified Sharpe ratio by computing the two-sided p-values via the circular 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%. We compare the following pairs of portfolios: PCMV vs MV (stars in superscript next to PCMV quantity), ICMV vs MV (stars in superscript next to ICMV quantity), ICMV vs PCMV (stars in subscript next to ICMV quantity) and ICMVaR vs MMVaR (stars in superscript next to ICMVaR quantity).

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 δ 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 modified VaR as risk measure. Just like for the variance as risk measure, the results show that there are substantial benefits in shrinking the sample minimum-modified-VaR portfolio (MMVaR) toward a portfolio that is diversified with respect to ICs. Indeed, the resulting shrinkage portfolio (ICMVaR) has a superior Sharpe ratio and modified Sharpe ratio on all datasets. Notably, even though the MMVaR portfolio aims at minimizing the modified VaR, the ICMVaR portfolio achieves a lower MVaR than the MMVaR portfolio out of sample, which indicates that it is less sensitive to estimation risk and that the parsimonious estimation procedure introduced in Section 5.1 works well. This is confirmed by the lower turnover of the ICMVaR portfolio compared to the MMVaR portfolio.

7. Conclusion

Factor-risk parity is an effective approach to portfolio diversification. Compared to asset-risk parity, it has the advantage of diversifying risk across uncorrelated factors, which improves diversification. However, we show that for any portfolio there is a basis of uncorrelated factors for which the portfolio is a factor-variance-parity portfolio. Thus, this framework is arbitrary when decorrelation is the only criterion governing the choice of factors. Choosing as factors the first principal components of asset returns is not more meaningful because, once dimension has been reduced, any further rotation of the principal components remains uncorrelated. To circumvent these concerns, we propose using the independent components, which are the factors with least dependence. They offer three benefits: they discriminate among all the factor-risk-parity portfolios, they reduce the portfolio-return kurtosis, and they provide a parsimonious way to deal with higher-moment risk measures. Empirically, we find that shrinking the minimum-risk portfolio toward the IC-risk-parity portfolio improves the out-of-sample performance in terms of mean-variance trade-off and tail risk.

Acknowledgements

Nathan Lassance's research is funded by the Fonds de la Recherche Scientifique (F.R.S.-FNRS) under grant ASP [grant number FC 17775]. We thank Kris Boudt, Guofu Zhou and participants of the 2018 INFORMS Annual Meeting (Phoenix, USA), 2019 Actuarial and Financial Mathematics Conference (Brussels, BE), 2019 Financial Econometrics Conference (Toulouse, FR) and SYZ Asset Management (London, UK) for comments and discussions.

References

Alexander G, Baptista, A (2002) Economic implications of using a mean-VaR model for portfolio selection: A comparison with mean-variance analysis. *J. Econ. Dyn. Control* 26(7-8):1159–1193.

Ang A, Chen J, Yu H (2006) Downside risk. *Rev. Financial Stud.* 19(4):1191–1239.

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

Athayde G, Flores R (2004) Finding a maximum skewness portfolio: A general solution to three-moments portfolio choice. *J. Econ. Dyn. Control* 28(7):1335–1352.

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.

Baitinger E, Dragosch A, Topalova A (2017) Extending the risk parity approach to higher moments: Is there any value added? *J. Portfolio Management* 43(2):24–36.

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, Carl P, Peterson B (2013) Asset allocation with conditional value-at-risk budgets. *J. Risk* 15(3):39–68.

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.

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

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

Cavenaile L, Lejeune T (2012). A note on the use of modified value-at-risk. *J. Alternative Investments* 14(4):79–83.

Chen Y, Härdle W, Spokoiny V (2007) Portfolio value at risk based on independent component analysis. *J. Comput. Appl. Math* 205(1):594–607.

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

Cornish E, Fisher R (1938) Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Stat.* 5(4):307–320.

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

Cui X, Zhu S, Li D, Sun X (2016) Mean-variance portfolio optimization with parameter sensitivity control. *Optim. Methods Softw.* 31(4):755–774.

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

Dittmar R (2002) Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *J. Finance* 57(1):369–403.

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.

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

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.

Jarque C, Bera A (1980) Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econ. Lett.* 6(3):255–259.

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

Jondeau E, Rockinger M (2006) Optimal portfolio allocation under higher moments. *Eur. Financ. Manag.* 12(1):29–55.

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 (2019) Shrinking the cross section. *J. Financial Econom.*, forthcoming.

Lassance N, Vrins F (2019) Robust portfolio selection using sparse estimation of comoment tensors. Louvain Finance working paper 2019/7.

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.

Lejeune M (2017) Stochastic optimization investment models with portfolio and marginal risk constraints. In Terlaky T, Anjos M, Ahmed S *Advances and Trends in Optimization with Engineering Applications* MOS-SIAM Series on Optimization 427–436.

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.

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

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

Masson E (2019) Portfolio selection via independent component analysis. Louvain School of Management, master's thesis.

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.

Pástor L (2000) Portfolio selection and asset pricing models. *J. Finance* 55(1):179–223.

Pástor L, Stambaugh R (2000) Comparing asset pricing models: an investment perspective. *J. Financial Econom.* 56(3):335-381.

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.

Roncalli T (2013) *Introduction to Risk Parity and Budgeting* (Chapman & Hall, Boca Raton).

Roncalli T, Weisang G (2015) 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.

Ugray Z, Lasdon L, Plummer J, Glover F, Kelly J, Martí R (2007) Scatter search and local NLP solvers: A multistart framework for global optimization. *INFORMS J. Comput.* 19(3):328–340.

Velicer W (1976) Determining the number of components from the matrix of partial correlations. *Psychometrika* 41:321–327.

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).

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

Zhu S, Li D, Sun X (2010) Portfolio selection with marginal risk control. *J. Comput. Finance* 14(1):3–28.

Electronic companion to "Optimal Portfolio Diversification via Independent Component Analysis"

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

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

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

This e-companion contains three sections. Section A gives proofs of all results in the paper. Section B describes the *FastICA* algorithm. Section C discusses the performance of the long-only factor-variance-parity portfolios.

A. Proofs of all results

A.1. Proposition 1

Proof. Let us prove the results one by one:

(i) From Definition 4, the FVP portfolios are obtained for $\tilde{w}_i(\boldsymbol{R})^2 = c$ for all i and some constant $c \in \mathbb{R}_0$ which ensures that $\tilde{\boldsymbol{w}}(\boldsymbol{R}) \in \widetilde{\mathcal{W}}(\boldsymbol{R})$. This can be rewritten as $\tilde{\boldsymbol{w}}(\boldsymbol{R}) = c\mathbf{1}_K^{\pm}$, which, from (7), translates in portfolio weights given by $\boldsymbol{w} = c\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\mathbf{1}_K^{\pm}$. After normalization, c vanishes and the final expression for the FVP portfolios in (11) is obtained. There are 2^K different vectors $\mathbf{1}_K^{\pm}$, which, combined with the normalization constraint, means that there are 2^{K-1} FVP portfolios.

(ii) The set of FVP portfolios $\mathcal{W}_{FVP}(\boldsymbol{R})$ is invariant to a change of sign and permutation because, from (8), changing the sign or permuting the components of $\boldsymbol{Y}$ changes the corresponding factor exposures $\tilde{\boldsymbol{w}}(\boldsymbol{R})$ in the same way, so that the reduced portfolio return remains identical.

(iii) Given some vector $\mathbf{1}_K^{\pm}$, the variance of the FVP portfolio return is given by

$$m_2(P) = \frac{(\mathbf{1}_K^{\pm})'\boldsymbol{R}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}'\boldsymbol{V}_N\boldsymbol{\Lambda}_N\boldsymbol{V}_N'\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\mathbf{1}_K^{\pm}}{(\mathbf{1}_N'\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\mathbf{1}_K^{\pm})^2}, \tag{A.1}$$

where $\boldsymbol{V}_N$ and $\boldsymbol{\Lambda}_N$ are the full $N \times N$ eigenvectors and eigenvalues matrices. $\boldsymbol{V}'\boldsymbol{V}_N$ is given by the first K rows of $\boldsymbol{I}_N$, and thus $\boldsymbol{V}'\boldsymbol{V}_N\boldsymbol{\Lambda}_N$ is made of the first K rows of $\boldsymbol{\Lambda}_N$. Multiplying $\boldsymbol{V}'\boldsymbol{V}_N\boldsymbol{\Lambda}_N$ by $\boldsymbol{V}_N'\boldsymbol{V}$, the first K columns of $\boldsymbol{I}_N$, thus gives

$$\boldsymbol{V}'\boldsymbol{V}_N\boldsymbol{\Lambda}_N\boldsymbol{V}_N'\boldsymbol{V} = \boldsymbol{\Lambda}.$$

As a result, (A.1) simplifies to

$$m_2(P) = \frac{(\mathbf{1}_K^{\pm})'\mathbf{1}_K^{\pm}}{(\mathbf{1}_N'\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\mathbf{1}_K^{\pm})^2} = \frac{K}{(\mathbf{1}_N'\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\mathbf{1}_K^{\pm})^2}. \tag{A.2}$$

Clearly, the vector $\mathbf{1}_K^{\pm}$ that minimizes (A.2) is unique and corresponds to

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

A.2. Proposition 2

Proof. To prove the proposition, we need to show that given any portfolio $\boldsymbol{w} \in \mathcal{W}$, there exists $\boldsymbol{R} \in \mathcal{SO}(N)$ for which $\boldsymbol{w}$ can be written in the form

$$\boldsymbol{w} = \frac{\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\mathbf{1}_N^{\pm}}{\mathbf{1}_N'\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\mathbf{1}_N^{\pm}}. \tag{A.3}$$

This amount to show that given any non-zero vector $\boldsymbol{v} \in \mathbb{R}^N$, there exists $\boldsymbol{R} \in \mathcal{SO}(N)$ for which

$$\boldsymbol{R}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}'\boldsymbol{v} = c\mathbf{1}_N^{\pm} \tag{A.4}$$

holds for some $c \in \mathbb{R}_0$. The corresponding portfolio $\boldsymbol{w}$ is then obtained by normalizing $\boldsymbol{v}$. We have that $||\mathbf{1}_N^{\pm}|| = \sqrt{N}$ and the scaling coefficient c given by

$$c = \frac{||\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}'\boldsymbol{v}||}{\sqrt{N}}$$

means that (A.4) can be written as

$$\boldsymbol{R}\boldsymbol{x} = \boldsymbol{y}, \tag{A.5}$$

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

A.3. Proposition 3

Proof. This proposition assumes that the asset returns are given by the linear-mixture model

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}, \tag{A.6}$$

where $\boldsymbol{A}$ is a full-rank $N \times K$ matrix, $I(\boldsymbol{S}) = 0$ and $\kappa(S_i) > 0$ for all i. In this setting, we have from Comon (1994) that the ICs $\boldsymbol{Y}^\dagger$ are independent and correspond to $\boldsymbol{S}$, ignoring the irrelevant sign-and-permutation indeterminacy. As a result, we have that $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{R}^{\dagger\prime}$ and the IC exposures are $\tilde{\boldsymbol{w}}(\boldsymbol{R}^\dagger) = \boldsymbol{A}'\boldsymbol{w}$. Thus, the reduced portfolio return $\tilde{P}$ corresponds to the portfolio return P in this setup:

$$\tilde{P} = \tilde{\boldsymbol{w}}(\boldsymbol{R}^\dagger)'\boldsymbol{Y}^\dagger = \boldsymbol{w}'\boldsymbol{A}\boldsymbol{Y}^\dagger = \boldsymbol{w}'\boldsymbol{A}\boldsymbol{S} = \boldsymbol{w}'\boldsymbol{X} = P. \tag{A.7}$$

This result is important for the proofs below. For ease of exposition, we denote $\kappa_i := \kappa\big(Y_i^\dagger\big)$ the excess kurtosis of the ith IC. Let us prove the results one by one:

(i) The excess kurtosis of the portfolio return $P = \tilde{P} = \tilde{\boldsymbol{w}}(\boldsymbol{R}^\dagger)'\boldsymbol{Y}^\dagger$ is found from the excess kurtosis of a weighted sum of independent random variables:

$$\kappa(P) = \frac{\sum_{i=1}^K \kappa_i \tilde{w}_i(\boldsymbol{R}^\dagger)^4}{\Big(\sum_{i=1}^K \tilde{w}_i(\boldsymbol{R}^\dagger)^2\Big)^2} = \frac{\sum_{i=1}^K \kappa_i z_i^2}{\Big(\sum_{i=1}^K z_i\Big)^2}, \tag{A.8}$$

where $z_i := \tilde{w}_i(\boldsymbol{R}^\dagger)^2 \geq 0$. We begin by noting two things. First, the sum $\sum_{i=1}^K z_i \neq 0$,

otherwise the portfolio-return variance $m_2(P)$ vanishes. Second, the vector of z_i's can only be identified up to a multiplicative constant because, from (A.8), it results in the same excess kurtosis $\kappa(P)$. Therefore, without loss of generality, we normalize the z_i's so that $\sum_{i=1}^{K} z_i = 1$. In turn, the portfolio-return excess kurtosis simplifies to

$$\kappa(P) = \sum_{i=1}^{K} \kappa_i z_i^2 = \boldsymbol{z}' \mathcal{K} \boldsymbol{z}, \tag{A.9}$$

where $\boldsymbol{z} = (z_1, \ldots, z_K)'$ and $\mathcal{K} := \operatorname{diag}(\kappa_1, \ldots, \kappa_K)$. Let us now find the global minimum and maximum of the excess kurtosis $\kappa(P)$.

- To find the global minimum of the excess kurtosis $\kappa(P)$, we need to solve the following quadratic optimization program:

$$\min_{\boldsymbol{z}} \boldsymbol{z}' \mathcal{K} \boldsymbol{z} \quad \text{subject to} \quad \mathbf{1}_K' \boldsymbol{z} = 1, \ \boldsymbol{z} \geq 0. \tag{A.10}$$

  This is a convex problem with unique solution similar to that of the minimum-variance portfolio: $\boldsymbol{z} = \mathcal{K}^{-1}\mathbf{1}_K / \mathbf{1}_K' \mathcal{K}^{-1} \mathbf{1}_K$. As desired, $\boldsymbol{z}$ is positive as we assume that the ICs have strictly positive excess kurtosis. The obtained minimum excess kurtosis $\kappa(P)$ corresponds indeed to $\kappa_{\min}$ in (13):

$$\min \kappa(P) = \frac{1}{\mathbf{1}_K' \mathcal{K}^{-1} \mathbf{1}_K} = \frac{1}{\sum_{i=1}^{K} 1/\kappa_i} = \frac{1}{K} H(\kappa_1, \ldots, \kappa_K) =: \kappa_{\min}.$$

- To find the maxima of the excess kurtosis $\kappa(P)$, we need to solve the following quadratic optimization program:

$$\max_{\boldsymbol{z}} \boldsymbol{z}' \mathcal{K} \boldsymbol{z} \quad \text{subject to} \quad \mathbf{1}_K' \boldsymbol{z} = 1, \ \boldsymbol{z} \geq 0. \tag{A.11}$$

  Given that $\boldsymbol{z}' \mathcal{K} \boldsymbol{z}$ is a convex function in $\boldsymbol{z}$, all maxima correspond to corner solutions of the type $z_i = 1$ and $z_j = 0$ for all $j \neq i$, with excess kurtosis $\kappa(P) = \kappa_i$. Therefore, the obtained global maximum corresponds indeed to $\kappa_{\max} := \max(\kappa_1, \ldots, \kappa_K)$ in (14).

(ii) The ICVP portfolio return is given by $P = \tilde{P} = c\mathbf{1}_K^{\pm\prime} \boldsymbol{Y}^{\dagger}$ and the resulting excess kurtosis in

(A.8) is independent of c and the particular vector $\mathbf{1}_K^{\pm}$ chosen, and is given by

$$\kappa(P) = \frac{\sum_{i=1}^{K} \kappa_i}{K^2} = \frac{1}{K} A(\kappa_1, \ldots, \kappa_K) =: \kappa_{IC}.$$

(iii) Consider the factors $\boldsymbol{Y}$ given by an arbitrary rotation $\boldsymbol{R}$ of the PCs, $\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{Y}^{\star}$. Our purpose is to show that the FVP portfolio associated with $\boldsymbol{Y}$ depends on the mixing matrix $\boldsymbol{A}$ associated with the linear-mixture model $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}$ *except in the very specific case corresponding to the rotation associated with the ICs*; that is, except when $\boldsymbol{R} \in \mathcal{R}_I$ in (4).

To show this, we first need to show that any portfolio return P is a FVP portfolio return for some rotation of the ICs; that is, that there exists $\tilde{\boldsymbol{R}} \in \mathcal{SO}(K)$ such that

$$P = \tilde{P} = \tilde{\boldsymbol{w}}(\boldsymbol{R}^{\dagger})'\boldsymbol{Y}^{\dagger} = c\mathbf{1}_K^{\pm\prime}\tilde{\boldsymbol{R}}\boldsymbol{Y}^{\dagger}$$

holds for some $c \in \mathbb{R}_0$. This is equivalent to showing that $\tilde{\boldsymbol{R}}\boldsymbol{A}'\boldsymbol{w} = c\mathbf{1}_K^{\pm}$, which holds by the same argument than in Proposition 2 by setting $c = ||\tilde{\boldsymbol{R}}\boldsymbol{A}'\boldsymbol{w}||/\sqrt{K}$.

Suppose now that the mixture model in the problem at hand is of the form $\boldsymbol{A} = \boldsymbol{R}_1\boldsymbol{D}\boldsymbol{R}'\tilde{\boldsymbol{R}}$ where $\boldsymbol{R}_1$ and $\boldsymbol{D}$ are respectively rotation and full-rank diagonal matrices; that is, $\boldsymbol{A}$ is such that $\boldsymbol{R}^{\dagger} = \tilde{\boldsymbol{R}}'\boldsymbol{R}$. Then, it is easy to see that the factors $\boldsymbol{Y}$ correspond to $\boldsymbol{Y} = \tilde{\boldsymbol{R}}\boldsymbol{Y}^{\dagger}$. Therefore, in this specific case, one obtains

$$P = c\mathbf{1}_K^{\pm\prime}\tilde{\boldsymbol{R}}\boldsymbol{Y}^{\dagger} = c\mathbf{1}_K^{\pm\prime}\tilde{\boldsymbol{R}}\tilde{\boldsymbol{R}}'\boldsymbol{Y} = c\mathbf{1}_K^{\pm\prime}\boldsymbol{Y};$$

that is, P is a FVP portfolio return associated with the factors $\boldsymbol{Y}$. In other words, depending on the ICA rotation matrix $\boldsymbol{R}^{\dagger}$ hidden behind the asset-return generating process $\boldsymbol{A}$, the FVP portfolio return associated to the factors $\boldsymbol{Y}$ could be *any* portfolio return and, consequently, could have any attainable kurtosis value. This is a fundamental difference with the ICVP portfolio returns that are independent from $\boldsymbol{A}$. Indeed, only the P corresponding to the special case $\tilde{\boldsymbol{R}} = \boldsymbol{I}$ (up to a change of sign and permutation of the rows) are ICVP portfolio returns, and all of those 2^{K-1} portfolio returns have the same excess kurtosis in (15). □

A.4. Excess kurtosis of PCVP portfolios in Example 4

Proof. We give the proof for the signs of PC exposures equal to $\mathbf{1}_2^{\pm} = \mathbf{1}_2$. The proof is similar for the other choice of signs, $\mathbf{1}_2^{\pm} = (1, -1)'$. Under the assumptions of Proposition 4, the PCVP portfolio return is given by

$$P = c\mathbf{1}_2'\boldsymbol{R}(\theta^\dagger)'\boldsymbol{Y}^\dagger = c\left(\cos\theta^\dagger - \sin\theta^\dagger\right)Y_1^\dagger + c\left(\cos\theta^\dagger + \sin\theta^\dagger\right)Y_2^\dagger,$$

where $c \in \mathbb{R}_0$ is an irrelevant constant. Then, from (A.8), the return excess kurtosis of the PCVP portfolio is given by

$$\kappa(P) = \frac{\left(\cos\theta^\dagger - \sin\theta^\dagger\right)^4 \kappa\left(Y_1^\dagger\right) + \left(\cos\theta^\dagger + \sin\theta^\dagger\right)^4 \kappa\left(Y_2^\dagger\right)}{\left(\left(\cos\theta^\dagger - \sin\theta^\dagger\right)^2 + \left(\cos\theta^\dagger + \sin\theta^\dagger\right)^2\right)^2},$$

which after algebraic manipulations yields the final expression with minus sign in (16). □

A.5. Proposition 4

Proof. The mean-variance tangent portfolio is given by

$$\boldsymbol{w} = \frac{\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\mu}_{\boldsymbol{X}}}{\mathbf{1}_N'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\mu}_{\boldsymbol{X}}}, \tag{A.12}$$

which, from (8), corresponds to the following exposures on the ICs:

$$\tilde{\boldsymbol{w}}(\boldsymbol{R}^\dagger) = \boldsymbol{R}^\dagger\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}'\boldsymbol{w} = \frac{\boldsymbol{R}^\dagger\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\mu}_{\boldsymbol{X}}}{\mathbf{1}_N'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\mu}_{\boldsymbol{X}}}. \tag{A.13}$$

As $K = N$, we have that $\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}'$, and thus, $\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1} = \boldsymbol{\Lambda}^{1/2}\boldsymbol{V}'\boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}' = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}'$. The IC exposures then become

$$\tilde{\boldsymbol{w}}(\boldsymbol{R}^\dagger) = \frac{\boldsymbol{R}^\dagger\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}'\boldsymbol{\mu}_{\boldsymbol{X}}}{\mathbf{1}_N'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\mu}_{\boldsymbol{X}}}. \tag{A.14}$$

Given that the ICs are defined as $\boldsymbol{Y}^\dagger = \boldsymbol{R}^\dagger\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}'\boldsymbol{X}$, the numerator of (A.14) is the mean-return vector of the ICs, $\boldsymbol{\mu}_{\boldsymbol{Y}^\dagger}$. Because an IC-variance-parity portfolio must have equal squared IC exposures, $\tilde{w}_i(\boldsymbol{R}^\dagger)^2 = \tilde{w}_j(\boldsymbol{R}^\dagger)^2$ for all i, j, the tangent portfolio is an IC-variance-parity portfolio if

each IC has the same mean return in absolute value:

$$\tilde{w}_i(\boldsymbol{R}^\dagger)^2 = \frac{\mu\big(Y_i^\dagger\big)^2}{\big(\mathbf{1}_N'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\mu}_{\boldsymbol{X}}\big)^2} \text{ is constant for all } i \text{ if } \big|\mu\big(Y_i^\dagger\big)\big| = \big|\mu\big(Y_j^\dagger\big)\big| \text{ for all } i, j.$$

Given that the ICs all have unit variance, this condition is equivalent to the ICs having the same Sharpe ratio in absolute value. □

A.6. 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 reduced portfolio return $\tilde{P} = \tilde{\boldsymbol{w}}(\boldsymbol{R}^\dagger)'\boldsymbol{Y}^\dagger$ correspond to $m_3(\tilde{P})$ and $m_4(\tilde{P}) - 3m_2(\tilde{P})^2$, respectively. Therefore, under the independence of the ICs, we directly find that the third central moment of $\tilde{P}$ is given by

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

Similarly, the fourth central moment of $\tilde{P}$ is found as follows:

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

□

A.7. Proposition 6

Proof. Boudt et al. (2008) show that the MVaR contribution of the ith asset X_i is given by

$$\begin{aligned}
\mathrm{MVaR}_\alpha(X_i|P) := w_i \times \Bigg[& -\mu(X_i) - \frac{\partial_i m_2(P)}{2\sqrt{m_2(P)}}\left(z_\alpha + p_1(z_\alpha) + p_2(z_\alpha)\right) + \\
& \sqrt{m_2(P)}\Bigg(- p_1(z_\alpha)\partial_i\zeta(P) - \frac{1}{24}(z_\alpha^3 - 3z_\alpha)\partial_i\kappa(P) + \frac{1}{18}(2z_\alpha^3 - 5z_\alpha)\zeta(P)\partial_i\zeta(P)\Bigg)\Bigg],
\end{aligned} \tag{A.15}$$

where the portfolio-return central moments are given by

$$m_2(P) = \boldsymbol{w}'\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{w}, \quad m_3(P) = \boldsymbol{w}'\boldsymbol{M}_3(\boldsymbol{w}\otimes\boldsymbol{w}) \quad \text{and} \quad m_4(P) = \boldsymbol{w}'\boldsymbol{M}_4(\boldsymbol{w}\otimes\boldsymbol{w}\otimes\boldsymbol{w}), \tag{A.16}$$

with $\boldsymbol{M}_3$ and $\boldsymbol{M}_4$ the coskewness and cokurtosis matrices of $\boldsymbol{X}$, respectively. The MVaR contribution of the ith IC $Y_i^\dagger$ is equivalent to (A.15)-(A.16) with respect to $\tilde{P} = \tilde{\boldsymbol{w}}(\boldsymbol{R}^\dagger)'\boldsymbol{Y}^\dagger$, and because we assume the ICs are independent, $m_3(\tilde{P})$ and $m_4(\tilde{P})$ simplify to (23)-(24). □

B. The *FastICA* algorithm

The theory supporting the foundations of *FastICA* is built upon the independence assumption; that is, the ICs are truly independent ($I(\boldsymbol{Y}^\dagger) = 0$). Thus, we present the *FastICA* algorithm under this assumption, keeping in mind that its practical application does not require this assumption to hold. In the non-independent case, the algorithm still delivers a special set of factors, associated with the linear combination of asset returns that minimizes mutual information. Most results below can be consulted in more details in the textbook of Hyvärinen et al. (2001).

Computing the ICs requires to find the rotation matrix such that the outputs have zero mutual information. By contrast with PCA, ICA needs to be achieved in an adaptive way, using gradient-descent techniques for instance. This triggers computational issues since mutual information requires the joint distribution of the factors $\boldsymbol{Y}$, which would need to be estimated at every step of the algorithm. Fortunately, it is easy to show that (Cover and Thomas 2006, p.251)

$$I(\boldsymbol{Y}) = \sum_{i=1}^{K} h(Y_i) - h(\boldsymbol{Y})$$

where $h(X) := -\mathbb{E}[\ln f_X(X)]$ is the Shannon (differential) entropy of X. Because $h(\boldsymbol{AX}) = h(\boldsymbol{X}) + \ln|\det \boldsymbol{A}|$, we have that $I(\boldsymbol{RY}^\star) = \sum_{i=1}^{K} h(\boldsymbol{R}_i\boldsymbol{Y}^\star) - h(\boldsymbol{Y}^\star)$, where $\boldsymbol{R}_i$ is the ith row of $\boldsymbol{R}$. Clearly, the second term does not depend on $\boldsymbol{R}$ and, over the set of rotation matrices $\mathcal{SO}(K)$, minimizing $I(\boldsymbol{RY}^\star)$ is the same as minimizing the sum of the entropies $h(\boldsymbol{R}_i\boldsymbol{Y}^\star)$, a criterion that only features the marginal densities of $Y_1, \ldots, Y_K$.

Another standard formulation of mutual information consists of using negentropy, defined as $J(X) := h(Z) - h(X)$ where $Z \sim \mathcal{N}(0,1)$ is a standard Gaussian random variable. Because this criterion is positive and vanishes if and only if $X \sim \mathcal{N}(0,1)$ whenever X is a standard random variable with support on the real line, it is trivial to see that maximizing $\sum_{i=1}^{K} J(\boldsymbol{R}_i\boldsymbol{Y}^\star)$ is equivalent

to minimizing $I(\boldsymbol{R}\boldsymbol{Y}^\star)$ over $\mathcal{SO}(K)$. This provides another intuitive interpretation of ICA in terms of non-Gaussianity. From the central limit theorem, it is known that the asymptotic distribution of a mixture of independent random variables is Gaussian. The above formulation suggests indeed that the independent (assumed non-Gaussian) factors can be recovered by maximizing a non-Gaussianity measure—here, negentropy—applied to every output $Y_i = \boldsymbol{R}_i\boldsymbol{Y}^\star$.

Interestingly, it has been proven by relying on a truncated development of the density around the standard Gaussian density using a set of orthogonal basis functions (a generalization of Gram-Charlier expansion, but using more general functions than simple monomials) that many "non-Gaussianity measures" can serve as a criterion to perform ICA. The fact that they correspond to a crude negentropy estimator is not an issue since it is possible to show that replacing $J(X)$ by

$$\hat{J}_m(X) := \sum_{i=1}^{m} c_i \left(\mathbb{E}^2[G_i(Z)] - \mathbb{E}^2[G_i(X)]\right),$$

for a set of functions G_i's satisfying some technical conditions, leads to a criterion $\sum_{i=1}^{K} \hat{J}(\boldsymbol{R}_i\boldsymbol{Y}^\star)$ with stationary points on the set $\{\boldsymbol{R} \in \mathcal{SO}(K) | I(\boldsymbol{R}\boldsymbol{Y}^\star) = 0\}$. As an illustration, plugging a fourth-order Gram-Charlier expansion of the density f_X in the negentropy definition leads to the above expression with $m = 2$, $c_1 = 1/12$, $c_2 = 1/48$, $G_1(x) = x^3$ and $G_2(x) = x^4$. In fact, it can even be shown that the restriction to $m = 1$, leading to

$$\hat{J}(X) := \hat{J}_1(X) = \mathbb{E}^2[G(Z)] - \mathbb{E}^2[G(X)]$$

is enough to perform ICA, in the sense that the stationary points of the above criterion include the ICs provided that G satisfies some conditions. For instance, taking $m = c_1 = 1$ and $G(x) = x^4$ corresponds to using the kurtosis as non-Gaussianity measure. Even though it is arguably a crude estimator of negentropy, it is theoretically proven that maximizing the sum of the square of the output kurtoses does also lead to recover the ICs; see Delfosse and Loubaton (1995). In practice however, neither the criterion corresponding to the actual negentropy nor that associated with the kurtosis is an appealing choice: the first one requires the estimation of the density of each of the factor Y_i at every iteration, and the second suffers from obvious robustness issues when estimated from finite samples. When the ICs underlying the data all have a positive excess kurtosis, it has been suggested to use $\hat{J}$ with $G(x) := \log\cosh x$. This function satisfies the technical conditions ensuring that the gradient of $\sum_{i=1}^{K} \hat{J}(\boldsymbol{R}_i\boldsymbol{Y}^\star)$ vanishes when $\boldsymbol{R}\boldsymbol{Y}^\star = \boldsymbol{Y}^\dagger$ (up to sign and permutation, as

usual), and leads to a very robust algorithm. This choice corresponds to the default setup of the *FastICA* algorithm developed by Hyvärinen (1999), the most popular algorithm to perform ICA. This approach has been successfully used in many applications, from the analysis of hyperspectral data to biomedical engineering, and thus, is used in the empirical part of the paper.

Finally, one may wonder whether the choice of the criterion used to perform ICA affects the existence of potential spurious local optima. The choice of the criterion does indeed matter. For instance, it has been shown that the actual negentropy criterion exhibits spurious (local) minima when the independent factors are strongly multimodal (Vrins and Verleysen 2005). In contrast, the kurtosis criterion is, in theory, free from spurious solutions: all stationary points of the criterion agree with the recovery of the ICs (Delfosse and Loubaton 1995). This result is however purely theoretical, and is of little practical interest given the large estimation errors resulting from the fourth power defining the kurtosis. The criterion optimized in *FastICA* is very appealing in this respect: choosing a smooth non-quadratic function G that does not grow too fast, like $G(x) = \ln \cosh x$, gives a robust algorithm, a feature that has been extensively studied with the help of simulated data. For further details about spurious solutions of ICA algorithms, we refer to Vrins (2007).

C. Long-only factor-variance-parity portfolios

In this section, we report the out-of-sample performance of factor-variance-parity (FVP) portfolios, based on PCA and ICA, under the no-short-selling constraint. Preventing short-selling is an important consideration in practice because many investors are not allowed to short assets and that it accounts for estimation risk (Jagannathan and Ma 2003).

Let us explain how we compute the long-only FVP portfolios. As shown by Roncalli and Weisang (2015), when short-selling is not allowed there may not exist a portfolio for which all the factors contribute equally to the variance of the portfolio return; that is, for which factor-variance parity is achieved. Therefore, we compute the long-only FVP portfolio in two steps. We consider here factors $\boldsymbol{Y}$ given by some rotation of the PCs: $\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{Y}^\star$. First, we compute the maximum factor-variance diversification that can be achieved by computing the maximum inverse Herfindahl index under no short-selling:

$$\max_{\boldsymbol{w}\in\mathcal{W}} \ \mathcal{H}[\boldsymbol{p}(\boldsymbol{R})] \quad \text{subject to} \quad w_i \geq 0 \ \forall i,$$

where

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

is the relative variance contribution of the ith factor Y_i and $\mathcal{H}[\boldsymbol{p}] := ||\boldsymbol{p}||^{-2} \in [1, K]$ is the inverse Herfindahl index, which equals K when $p_i(\boldsymbol{R}) = 1/K$ for all i. This gives the maximum inverse Herfindahl index $\mathcal{H}_{\max}$, which can be lower than K. Second, to be consistent with our proposal of relying on the FVP portfolio with minimum variance, we compute the long-only FVP portfolio as follows:

$$\min_{\boldsymbol{w}\in\mathcal{W}} \ \boldsymbol{w}'\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{w} \quad \text{subject to} \quad \mathcal{H}[\boldsymbol{p}(\boldsymbol{R})] \geq \mathcal{H}_{\max} - \epsilon, \ w_i \geq 0 \ \forall i,$$

where ϵ is a tolerance parameter set to $\epsilon = 5 \times 10^{-4}$.

Then, in Table I, we report the out-of-sample performance of three portfolio strategies, using the same methodology than for the unconstrained portfolios in the paper. First, the long-only minimum-variance (MV) portfolio. Second, the portfolio that is a shrinkage between the long-only MV portfolio and the long-only IC-variance-parity portfolio (ICMV). Third, the portfolio that is a shrinkage between the long-only MV portfolio and the long-only PC-variance-parity portfolio (PCMV). The shrinkage intensity is calibrated as for the unconstrained portfolios; that is, via 10-fold cross-validation using modified Sharpe ratio as calibration criterion.

The table shows that the results obtained for the long-only portfolios remain qualitatively similar to the unconstrained portfolios in the paper. Specifically, the ICMV portfolio outperforms the PCMV portfolio in terms of Sharpe ratio and modified Sharpe ratio for all three datasets and, compared to the MV portfolio, performs slightly worse in terms of Sharpe ratio but substantially better in terms of tail risk. Finally, the out-of-sample performance of long-only portfolios is systematically worse than that of unconstrained portfolios. The reason may be that we use daily data, and Jagannathan and Ma (2003) have shown that for daily data the no-short-selling constraint is likely to hurt performance. One benefit of long-only portfolios however is that they require less turnover.

References

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.

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

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

Delfosse N, Loubaton P (1995) Adaptive blind separation of independent sources: A deflation approach. *Signal Process.* 45(1):59–83.

Hyvärinen A (1999) Fast and robust fixed-point algorithms for independent component analysis.

Table I Out-of-sample performance of long-only MV, ICMV and PCMV portfolios

	6BTM dataset			*6Prof* dataset			*30Ind* dataset		
	MV	ICMV	PCMV	MV	ICMV	PCMV	MV	ICMV	PCMV
Mean	14.54%	**14.99%**	14.78%	13.82%	**14.49%**	13.90%	12.36%	**15.18%**	10.77%
	(2.57%)	(2.66%)	(2.80%)	(2.57%)	(2.67%)	(2.56%)	(2.22%)	(2.87%)	(2.32%)
Volatility	**14.82%**	15.62%	16.07%	**15.18%**	16.91%	15.24%	**13.12%**	16.91%	13.66%
	(2.57%)	(2.66%)	(2.80%)	(2.57%)	(2.67%)	(2.56%)	(2.22%)	(2.87%)	(2.32%)
Sharpe ratio	**0.98**	0.96	0.92	**0.91**	0.86	0.91	**0.94**	0.90	0.79
	(0.17)	(0.18)	(0.18)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)
Skewness	-0.55	-0.72	**-0.42**	-0.46	-0.54	**-0.45**	-0.78	**-0.20**	-0.61
	(0.39)	(0.29)	(0.32)	(0.42)	(0.35)	(0.39)	(0.71)	(0.45)	(0.61)
Excess kurtosis	15.51	**13.00**	13.66	15.32	**10.41**	15.06	24.89	**16.75**	18.81
	(3.25)	(2.36)	(2.37)	(3.71)	(2.42)	(3.62)	(10.81)	(4.50)	(8.62)
1% modified Value-at-Risk	5.77%	**5.55%**	5.77%	5.84%	**5.32%**	5.80%	6.97%	6.73%	**6.01%**
	(0.81%)	(0.67%)	(0.64%)	(0.88%)	(0.65%)	(0.90%)	(2.25%)	(1.27%)	(1.94%)
Modified Sharpe ratio ($\times 10^2$)	1.00	**1.07**	1.02	0.94	**1.08**	0.95	0.70	**0.90**	0.71
	(0.24)	(0.26)	(0.24)	(0.25)	(0.24)	(0.24)	(0.35)	(0.27)	(0.35)
Portfolio turnover	**0.24%**	0.44%	0.44%	**0.19%**	0.50%	0.22%	**0.61%**	0.95%	0.82%
Average shrinkage intensity δ	/	0.36	0.14	/	0.52	0.10	/	0.37	0.40

Notes. This table reports the out-of-sample performance of the variance-based portfolios under the no-short-selling constraint, following the methodology in Section C of this e-companion and Section 6.1 of the paper. 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. The number of PCs kept, K, is fixed via the minimum-average-partial-correlation method of Velicer (1976). Bold figures indicate the best strategy for each criterion and dataset. Figures in parentheses indicate the standard errors, computed via non-parametric bootstrap with Matlab *bootstat* function and $B = 1000$ bootstrap resamples.

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.

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

Roncalli T, Weisang G (2015) Risk parity portfolios with risk factors. *Quant. Finance* 16(3):377–388.

Velicer W (1976) Determining the number of components from the matrix of partial correlations. *Psychometrika* 41:321–327.

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

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).