Optimal Portfolio Diversification via Independent Component Analysis

Article *in* Operations Research · April 2021

3 authors:

Nathan Lassance
Université Catholique de Louvain - UCLouvain

Victor Demiguel
London Business School

Frederic Vrins
Université Catholique de Louvain - UCLouvain

Optimal Portfolio Diversification via Independent Component Analysis

Forthcoming in *Operations Research* (2021)

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

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

A natural approach to enhance portfolio diversification is to rely on *factor-risk parity*, which yields the portfolio whose risk is equally spread among a set of uncorrelated factors. The standard choice is to take the variance as risk measure, and the principal components (PCs) of asset returns as factors. Although PCs are unique and useful for dimension reduction, they are an arbitrary choice: any rotation of the PCs results in uncorrelated factors. This is problematic because we demonstrate that any portfolio is a factor-variance-parity portfolio for some rotation of the PCs. More importantly, choosing the PCs does not account for the higher moments of asset returns. To overcome these issues, we propose to use the independent components (ICs) as factors, which are the rotation of the PCs that are maximally independent, and care about higher moments of 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 several state-of-the-art benchmarks.

Key words: 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 (2013, Section 2.5.1.1) 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 and most importantly, 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. Second, relying on the PCs is arbitrary because once dimension has been reduced, any further rotation of the PCs remains uncorrelated, and thus, is an equally valid set of factors for factor-risk-parity purposes. This is worrying because we demonstrate that *any* portfolio spanned by the PCs is actually a factor-variance-parity portfolio, with associated factors corresponding to a specific rotation of the PCs.

To overcome these two drawbacks, 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), a statistical technique that can be used to extract a set of K independent components (ICs) from the N observed asset returns (Hyvärinen et al. 2001). Following a popular approach, we define the ICs as the rotation of the PCs that results in factors as independent as possible. Independence is particularly relevant in portfolio selection because asset returns often display fat tails and higher-moment dependencies. For instance, Massacci (2017) shows that global equity markets present tail connectedness, which increases during periods of turmoil. Performing risk parity over the ICs overcomes the two drawbacks that arise when opting for the PCs: the ICs enhance diversification by accounting for higher-moment dependence, and provide a meaningful way to discriminate among all the uncorrelated factors.

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 set of factor-variance-parity portfolios corresponding to a generic rotation of the PCs. We then use this closed-form expression to show that any portfolio spanned by the PCs is a factor-variance-parity portfolio corresponding to a specific rotation of the PCs. Thus, the factor-decorrelation criterion is not sufficient to find a single set of factor-variance-parity portfolios.

Second, to overcome the difficulties associated with the PC-risk-parity portfolio, we propose relying instead on the ICs. We theoretically demonstrate that, under some assumptions, using the IC-variance-parity portfolio helps to reduce the return kurtosis. To do this, 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), Ang et al. (2006), and Gao et al. (2019).

Third, we show how to obtain a parsimonious estimator of the higher moments of portfolio returns by ignoring the higher comoments of the ICs. This is reasonable because, although the ICs are not fully independent in general, their higher comoments are small compared to those of other factors, such as the PCs. More importantly, higher-comoment estimators are noisy and thus ignoring them results in robust estimators. 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 that combine minimum-risk and IC-risk-parity portfolios, using variance and VaR risk measures, on three datasets containing the returns of six portfolios sorted on size and book-to-market, six portfolios sorted on size and profitability, and 30 industry 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. 2009). We find that the portfolios based on ICs substantially outperform those based on PCs in terms of Sharpe ratio and tail risk, perform well compared to several state-of-the-art benchmarks, and have low turnover when relying on the variance risk measure. We also show in Section EC.6.1 of the e-companion that our findings are robust to considering a larger dataset with 100 stocks. A battery of robustness tests confirm the relevance of the base-case results.

Our work is connected to four streams of literature. First, the literature on risk parity. 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. 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. 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 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, Guidolin and Timmermann (2008) and Martellini and Ziemann (2010) use a Taylor-series expansion of expected utility 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 indirectly, 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. Empirically, we find that our approach outperforms two mean-variance-kurtosis efficient portfolios based on the framework of Briec et al. (2007).

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). A recent modeling paradigm that tackles parameter uncertainty is distributionally robust optimization, which bridges robust optimization and stochastic programming. This approach finds the optimal portfolio for the worst-case distribution, where the true distribution lies within a specified set of distributions; see Calafiore (2007), Popescu (2007) and Delage and Ye (2010) for applications to portfolio selection. 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 the analysis of financial data; see, for instance, Chen et al. (2010) for exchange rates, Kumiega et al. (2011) and García-Ferrer et al. (2012) for equity returns, and Fabozzi et al. (2016) for CDS spreads. 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 (2021) 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 (2021) and compares its performance to that of various benchmarks.

The remainder of the paper 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. Section EC.1 of the e-companion contains proofs for all results.

2. PCA versus ICA: from decorrelation to independence

We first review *principal component analysis* (PCA), a dimension-reduction technique that generates a basis of uncorrelated factors ordered by their ability to explain the variance of the projected asset returns. Although the principal components (PCs) are unique, any rotation of the PCs remains uncorrelated, and thus, the PCs are an arbitrary choice for factor-risk parity. We then introduce a popular approach to *independent component analysis* (ICA) that finds the rotation of the PCs that are as independent as possible, which discriminates among all bases of uncorrelated factors.

2.1. Principal component analysis

Let $\boldsymbol{X} = (X_1, \ldots, X_N)'$ be a random vector of N asset returns with covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$, which 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 $\boldsymbol{\Lambda}_N :=$ diag$(\lambda_1, \ldots, \lambda_N)$ contains the eigenvalues sorted in decreasing order. This factorization is at the heart of *principal component analysis* (PCA); see 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 are the corresponding eigenvectors.

Definition 1 (PCA). The *principal components* (PCs) are the standardized projection of the asset returns $\boldsymbol{X}$ onto the first K 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. 2020). However, once dimension has been reduced, it is decorrelation that makes the PCs appealing for factor-risk parity, which spreads the portfolio-return risk among a set of *uncorrelated* factors. Indeed, most existing papers on factor-risk parity rely on the PCs as 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 thus is as suitable for factor-risk parity as the PCs. To see this, let $\mathcal{SO}(K)$ be the *special orthogonal 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: $\boldsymbol{\Sigma}_{\boldsymbol{Y}} = \boldsymbol{R}\boldsymbol{\Sigma}_{\boldsymbol{Y}^\star}\boldsymbol{R}' = \boldsymbol{R}\boldsymbol{I}_K\boldsymbol{R}' = \boldsymbol{I}_K$. Note that this is also true for orthogonal matrices with $\det(\boldsymbol{R}) = -1$. However, we focus on rotations because the sign of the factors is irrelevant for factor-risk parity:

factor-risk contributions are obtained as the product of the factor exposures and the derivative of the risk measure with respect to the factor exposures, so that any change of sign cancels out.

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 is problematic because we show in Section 3.2 that any portfolio spanned by the K PCs is a factor-variance-parity portfolio associated with a particular rotation of the PCs.

2.2. Independent component analysis

To discriminate among all possible uncorrelated factors, one can look for factors that are as *independent* as possible. Independence is a much stronger requirement than decorrelation: it not only requires the covariance matrix to be diagonal, but also that higher-moment dependence vanishes. A standard criterion to measure the independence of a random vector is the 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 $\boldsymbol{Y}$:

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

Mutual information is nonnegative, and it is zero if and only if the components of $\boldsymbol{Y}$ are mutually independent (Cover and Thomas 2006, p.253). We now introduce the *independent components* that use mutual information instead of linear correlation as dependence measure.

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

$$\boldsymbol{Y}^{\dagger} := \boldsymbol{R}^{\dagger}\boldsymbol{Y}^{\star}, \quad \boldsymbol{R}^{\dagger} \in \left\{ \underset{\boldsymbol{R} \in \mathcal{SO}(K)}{\arg\min}\, I(\boldsymbol{R}\boldsymbol{Y}^{\star}) \right\}. \tag{3}$$

Observe that mutual information is invariant with respect to a permutation or change of sign of the factors $\boldsymbol{Y}$. Therefore, the ICs are determined only up to a change of sign and permutation. However, as we show in Section 3.1, whatever the choice of factors (e.g., the PCs or ICs), there are always 2^{K-1} factor-variance-parity portfolios. Thus, ICs do not introduce any additional indeterminacy. Note also that the ICs may not be perfectly independent; they are the *maximally independent* factors found by a linear transformation of the asset returns.

An implication of Definition 3 is that we rely on PCA to reduce the problem dimension and choose the number of factors K. This PCA preprocessing step is common in the ICA literature because while PCA is designed to perform dimension reduction by selecting the PCs that maximize the explained variance, ICA is not. Indeed, ICA provides maximally independent factors (the ICs) from asset returns, but it does not provide a criterion to choose *how many* ICs should be retained.

Hereafter we assume that at most one of the K principal components is Gaussian. We require this assumption because, if more than one PC is Gaussian, then the factors obtained by rotating the Gaussian PCs remain jointly Gaussian and uncorrelated, and thus mutually independent. Thus, in this case the ICs are undetermined beyond a change of sign and permutation. Nonetheless, a modified version of our approach can be applied even if several PCs are Gaussian, provided there are at least two non-Gaussian PCs, because one can still identify the ICs *in the subset of the non-Gaussian PCs*. In particular, one can rely on the set of factors consisting of the Gaussian PCs plus the factors obtained by rotating the non-Gaussian PCs to minimize mutual information. Also, we find that the non-Gaussianity assumption holds for the three empirical datasets we use in Section 6 because their PCs have significantly positive excess kurtosis. In addition, Section EC.2 of the e-companion lists the type of asset-return datasets for which, according to the existing literature, one can expect the non-Gaussianity assumption to hold.

The ICs are computed using *independent component analysis* (ICA), a well-known technique in machine learning; see Hyvärinen et al. (2001), Vrins (2007), and Vrins et al. (2007). To compute the rotation matrix $\boldsymbol{R}^{\dagger}$ associated with the ICs, we use the *FastICA* algorithm of Hyvärinen (1999), described in Section EC.3 of the e-companion. The advantages of *FastICA* are that it is simple, robust to outliers, computationally efficient, and has been shown to perform well for a wide range of distribution types and applications. Nonetheless, as explained in Section EC.3 of the e-companion, *FastICA* is derived under approximations, and thus other algorithms may be preferred for certain problems. For instance, two popular alternative approaches to ICA are JADE (Cardoso and Souloumiac 1993) and SOBI (Belouchrani et al. 1997); see García-Ferrer et al. (2012) for a discussion in a financial context. Expectation-maximization algorithms have also proved successful; see Ablin et al. (2019) for such a recent ICA algorithm that does not require any hyperparameter tuning, whose stochastic loss function is guaranteed to decrease at each iteration, and which outperforms state-of-the-art algorithms for high-dimensional datasets.

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

$$\boldsymbol{R} = \boldsymbol{R}(\theta) := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in \mathcal{SO}(2). \tag{4}$$

The examples are illustrated with polar plots, which use polar coordinates (θ, r) with r 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: $\boldsymbol{Y}^\star = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{X}$. The left plot of Figure 1 depicts how the relative mutual information $\bar{I}$ (a standardized version of the mutual information taking values in $[0,1]$) and correlation of the factors $\boldsymbol{Y} = \boldsymbol{R}(\theta)\boldsymbol{Y}^\star$, obtained by rotating the PCs, depend on θ. The correlation is zero for all θ, but 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 two assets with returns distributed as a Gaussian copula of correlation $\rho = 0.2$ and X_1, X_2 above as marginals. The right plot of Figure 1 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 θ. Independence is never achieved but the mutual information varies with θ and this allows one to discriminate between all uncorrelated factors. In particular, the ICs $\boldsymbol{Y}^\dagger$ are obtained for $\theta = 0.35 + k\pi/2 \neq 0$, $k \in \mathbb{Z}$, and thus, differ from the PCs.

Finally, to gauge the practical relevance of the ICs, we use the three equity datasets from Ken French's library that we use in Section 6 to evaluate the gain in independence that can be achieved by moving from the PCs to the ICs. Figure 2 shows that the gain in independence, measured by $I(\boldsymbol{Y}^\dagger) - I(\boldsymbol{Y}^\star)$, is particularly substantial in the period preceding and during the 2007-2009 financial crisis, a period characterized by pronounced tail risk and higher-moment dependence of asset returns. We refer to Section EC.5 of the e-companion for a more detailed discussion.

3. Factor-variance parity via uncorrelated factors

We now introduce 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

$$\mathcal{W} := \{\boldsymbol{w} \in \mathbb{R}^N \mid \boldsymbol{1}_N'\boldsymbol{w} = 1\}. \tag{5}$$

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 consider higher-moment risk measures. The original proposal of Meucci (2009) is to diversify the portfolio-return variance among the PCs. As discussed in Section 2.1, this choice is arbitrary because 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 in the subspace spanned by the PCs is an FVP portfolio for a set of factors obtained by rotating the PCs.

3.1. Derivation of the factor-variance-parity portfolio

Consider K factors obtained by rotating the PCs, $\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{Y}^\star$. Unlike Meucci (2009) and Deguest et al. (2013), we allow K to be smaller than the number of assets N, because this reduces the dimension and thus alleviates the impact of estimation error (Hastie et al. 2009, Kozak et al. 2020).

Because we reduce dimension by considering K PCs, we start by considering the asset returns reconstructed from the subspace spanned by the K factors $\boldsymbol{Y}$:

$$\tilde{\boldsymbol{X}} := \boldsymbol{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{R}'\boldsymbol{Y} = \boldsymbol{V}\boldsymbol{V}'\boldsymbol{X}. \tag{6}$$

We then define the *reduced portfolio return* $\tilde{P}$ as the projection of the return of the portfolio $\boldsymbol{w}$ onto the subspace spanned by the K factors:

$$\tilde{P} := \boldsymbol{w}'\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{w}}(\boldsymbol{R})'\boldsymbol{Y}, \tag{7}$$

where the factor exposures $\tilde{\boldsymbol{w}}(\boldsymbol{R})$ are defined as

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

Although the matrix $\boldsymbol{R}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}'$ is not invertible for $K < N$, the reduced portfolio return associated with $\boldsymbol{w}$, $\tilde{P} = \boldsymbol{w}'\tilde{\boldsymbol{X}}$, is recovered from $\tilde{\boldsymbol{w}}(\boldsymbol{R})$ by taking the pseudo inverse:

$$\boldsymbol{w} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\tilde{\boldsymbol{w}}(\boldsymbol{R}). \tag{9}$$

Indeed, $\tilde{P} = \boldsymbol{w}'\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{w}}(\boldsymbol{R})'\boldsymbol{Y} = \tilde{\boldsymbol{w}}(\boldsymbol{R})'\boldsymbol{R}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}'\boldsymbol{X} = \tilde{\boldsymbol{w}}(\boldsymbol{R})'\boldsymbol{R}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}'\tilde{\boldsymbol{X}}$ as $\boldsymbol{V}'\boldsymbol{X} = \boldsymbol{V}'\tilde{\boldsymbol{X}}$, and comparing the first and last equalities leads to (9). Consistent with the requirement $\boldsymbol{w} \in \mathcal{W}$, the factor exposures $\tilde{\boldsymbol{w}}(\boldsymbol{R})$ must belong to

$$\widetilde{\mathcal{W}}(\boldsymbol{R}) := \left\{\tilde{\boldsymbol{w}}(\boldsymbol{R}) \in \mathbb{R}^K \;\middle|\; \boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\tilde{\boldsymbol{w}}(\boldsymbol{R}) \in \mathcal{W}\right\}. \tag{10}$$

Observe now that because $\boldsymbol{\Sigma}_{\boldsymbol{Y}} = \boldsymbol{I}_K$, we have from (7) that the variance of $\tilde{P}$ is simply given by

$$m_2(\tilde{P}) = \sum_{i=1}^{K} \tilde{w}_i(\boldsymbol{R})^2 = \|\tilde{\boldsymbol{w}}(\boldsymbol{R})\|^2, \tag{11}$$

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 $\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{Y}^\star$ satisfies $\tilde{w}_i(\boldsymbol{R})^2 = \tilde{w}_j(\boldsymbol{R})^2$ for all $i, j = 1, \ldots, K$.

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

$$\mathcal{W}_{FVP}(\boldsymbol{R}) := \left\{ \boldsymbol{w} \in \mathcal{W} \;\middle|\; \boldsymbol{w} = \frac{\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\boldsymbol{1}_K^{\pm}}{\boldsymbol{1}_N'\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\boldsymbol{1}_K^{\pm}} \right\}, \tag{12}$$

where $\boldsymbol{1}_K^{\pm}$ 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 orthogonal 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 obtained by considering the particular case $\boldsymbol{1}_K^{\pm} \leftarrow \boldsymbol{1}_K^{MV}$ in (12), where

$$\boldsymbol{1}_K^{MV} := \operatorname{sign}(\boldsymbol{R}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}'\boldsymbol{1}_N). \tag{13}$$

Part (i) shows that there are 2^{K-1} FVP portfolios. This is because a portfolio is an FVP portfolio if all the *squared* factor exposures are equal, and thus the factor signs are irrelevant. This holds for any choice of factors, including the PCs and ICs. Part (ii) shows that the sign and order of the factors is irrelevant for factor-variance parity. Part (iii) identifies the FVP portfolio that minimizes portfolio-return variance, which we call MV-FVP portfolio, and this allows us to choose one of the 2^{K-1} FVP portfolios. We rely on the sign vector $\boldsymbol{1}_K^{MV}$ in the empirical analysis.

3.2. Arbitrariness of the factor-decorrelation criterion

The following proposition shows that the factor-decorrelation criterion is arbitrary because *any* portfolio in the subspace spanned by the first K PCs is an FVP portfolio with respect to a set of uncorrelated factors obtained by rotating the PCs. To set the stage for the proposition, note that there is a one-to-one mapping between the portfolios spanned by the PCs, $\boldsymbol{w}$, and their PC exposures, $\tilde{\boldsymbol{w}}(\boldsymbol{I}_K)$, defined in (8)–(9). Therefore, the set of portfolios spanned by the K PCs is

$$\mathcal{W}_K := \left\{ \boldsymbol{w} \in \mathcal{W} \;\middle|\; \boldsymbol{w} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\tilde{\boldsymbol{w}}(\boldsymbol{I}_K),\; \tilde{\boldsymbol{w}}(\boldsymbol{I}_K) \in \widetilde{\mathcal{W}}(\boldsymbol{I}_K) \right\}.$$

PROPOSITION 2. *Any portfolio in the subspace spanned by the K principal compoments, $\boldsymbol{w} \in \mathcal{W}_K$, is an FVP portfolio for a set of uncorrelated factors obtained by rotating the K principal components.*

Proposition 2 demonstrates that factor decorrelation is not sufficient to discriminate among all portfolios in $\mathcal{W}_K$. Choosing $\boldsymbol{R} = \boldsymbol{I}_K$ leads to a unique set of PC-variance-parity (PCVP) portfolios, 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 different FVP portfolios. 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 $K = 2$. The left plot of Figure 3 depicts the weight on the first asset, w_1, of the two FVP portfolios as a function of the rotation angle θ. When $K = N$, $\mathcal{W}_K = \mathcal{W}$ and 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 an FVP portfolio for some uncorrelated factors $\boldsymbol{R}(\theta)\boldsymbol{Y}^\star$. However, if for each θ we choose among the two FVP portfolios the one with the lowest return variance according to (13), then the optimal weight on the first asset is no longer unbounded as a function of θ. This is illustrated in the right plot of Figure 3. 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 an FVP portfolio with minimum return variance.

Example 3 suggests that although not every portfolio is a *minimum-variance* FVP (MV-FVP) for a rotation of the PCs, a wide range of portfolios are MV-FVP for some rotation of the PCs. Identifying theoretically the set of MV-FVP portfolios corresponding to all rotations of the PCs is complex because the vector of exposure signs that leads to minimum variance, $\mathbf{1}_K^{MV}$, depends on the rotation matrix $\boldsymbol{R}$. However, Section EC.1.3 of the e-companion shows that for the case where $K = 2$ and the covariance matrix is diagonal, any portfolio of the two assets with largest variance whose weights satisfy certain bounds is an MV-FVP portfolio for some rotation of the PCs.

4. Kurtosis diversification via independent component analysis

In Section 3, we showed that decorrelation is not a discriminant criterion for factor-variance parity; Proposition 2 shows that any portfolio spanned by the K PCs is an FVP portfolio, whereas Example 3 suggests that focusing on MV-FVP portfolios does not alleviate this problem. Going from decorrelation to independence by relying on the ICs resolves this indeterminacy. In this section, we first show that diversifying variance among ICs provides a natural way of reducing the kurtosis of portfolio returns, contrary to PCs. Second, we characterize the relation between the tangent mean-variance portfolio and the 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 $\boldsymbol{Y}^\dagger = \boldsymbol{R}^\dagger\boldsymbol{Y}^\star$ have least mutual information. 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 (12).

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 standardized independent factors with strictly positive excess kurtosis, which amounts to the factor model $\boldsymbol{X} = \boldsymbol{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{R}^{\dagger\prime}\boldsymbol{Y}^{\dagger}$ for some rotation matrix $\boldsymbol{R}^{\dagger}$. Let the arithmetic and harmonic means be $A(x_1,\ldots,x_K) := \frac{1}{K}\sum_{i=1}^{K} x_i$ and $H(x_1,\ldots,x_K) := K\left(\sum_{i=1}^{K} 1/x_i\right)^{-1}$, respectively.*

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

$$\kappa_{\min} = \frac{1}{K} H\big(\kappa(Y_1^{\dagger}),\ldots,\kappa(Y_K^{\dagger})\big) \text{ and} \tag{14}$$

$$\kappa_{\max} = \max\big(\kappa(Y_1^{\dagger}),\ldots,\kappa(Y_K^{\dagger})\big), \tag{15}$$

where $\kappa(Y_i^{\dagger})$ 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,

$$\kappa_{IC} = \frac{1}{K} A\big(\kappa(Y_1^{\dagger}),\ldots,\kappa(Y_K^{\dagger})\big). \tag{16}$$

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

The intuition behind Proposition 3 is that the central limit theorem implies that the asymptotic distribution of a sum of standardized *independent* random variables is Gaussian, and thus has zero excess kurtosis. Although there is only a finite number K of ICs, Equation (16) shows that the return excess kurtosis of ICVP portfolios decreases toward zero with K. Although ICs are not fully independent in practice, they are maximally independent, and thus we may expect that the return excess kurtosis of the ICVP portfolios is closer to zero than those of FVP portfolios based on other less independent factors such as the PCs.

There are two important implications from Proposition 3. First, concomitantly with diversifying the portfolio-return variance among independent factors, one obtains a kurtosis that is close to the minimum one. In particular, for the case where the ICs are fully independent, the excess kurtosis of the ICVP 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 $\kappa_{\min}$ as long as the excess kurtosis of the ICs are not too different from one another. For example, taking $K=3$ and $\left(\kappa\left(Y_1^\dagger\right),\kappa\left(Y_2^\dagger\right),\kappa\left(Y_3^\dagger\right)\right)=(3,5,7)$, $\kappa_{\min}=105/71=1.48$, versus $\kappa_{IC}=5/3=1.67$. Note that in this case the ICVP portfolios have equal kurtosis because the kurtosis of a weighted sum of independent factors only depends on the *squared* weights. This argument does not hold for non-independent factors, such as the PCs.

Second, by diversifying the portfolio-return variance among non-independent factors, we do not have any control over kurtosis. Specifically, depending on the rotation matrix $\boldsymbol{R}^\dagger$ underlying the asset returns, the FVP portfolio return can have *any* kurtosis in the range $[\kappa_{\min},\kappa_{\max}]$. This is true in particular for the PCVP portfolios and is a fundamental difference with the kurtosis of the ICVP portfolios that does not depend on $\boldsymbol{R}^\dagger$.

Thus, the ICVP portfolio strategy naturally reduces the tail risk of portfolio returns, which is an important concern for many investors; see Dittmar (2002), Ang et al. (2006), and Gao et al. (2019). 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 rotation matrix can be parametrized as $\boldsymbol{R}^\dagger=\boldsymbol{R}(\theta^\dagger)$. Then, the PCs $\boldsymbol{Y}^\star=\boldsymbol{R}(\theta^\dagger)'\boldsymbol{Y}^\dagger$ depend on $\theta^\dagger$ and the excess kurtosis of the two PCVP portfolios is given by

$$\kappa_{PC}\in\frac{1}{8}\left(\pm4\sin2\theta^\dagger\left(\kappa\left(Y_1^\dagger\right)-\kappa\left(Y_2^\dagger\right)\right)+(3-\cos4\theta^\dagger)\left(\kappa\left(Y_1^\dagger\right)+\kappa\left(Y_2^\dagger\right)\right)\right); \tag{17}$$

see Section EC.1.5 of the e-companion. In line with Proposition 3(iii), κ_{PC} spans $[\kappa_{\min},\kappa_{\max}]$ as a function of $\theta^\dagger$. For $\left(\kappa\left(Y_1^\dagger\right),\kappa\left(Y_2^\dagger\right)\right)=(3,6)$ for example, $\kappa_{IC}=9/4$ is very close to $\kappa_{\min}=2$, whereas κ_{PC} takes values in $[\kappa_{\min},\kappa_{\max}]=[2,6]$. This shows that the kurtosis of the PCVP portfolio strongly depends on the matrix $\boldsymbol{R}^\dagger$ underlying the asset returns, on which we have no control.

4.3. Relation to mean-variance portfolio

We now characterize the relation between the ICVP portfolio and the mean-variance portfolio of Markowitz (1952). In particular, we show that the ICVP portfolio can be interpreted as a shrinkage estimator of the tangent mean-variance portfolio. To do this, we derive a similar result to Maillard et al. (2010), who show that the tangent mean-variance portfolio is equal 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. *If the Sharpe ratio of the returns of the K independent components are equal in absolute value, then the tangent mean-variance portfolio of assets spanned by the K principal components is an IC-variance-parity portfolio.*

Proposition 4 shows that the ICVP portfolio can be seen as a shrinkage estimator of the tangent mean-variance portfolio, where the estimator of the absolute value of the Sharpe ratio of each IC is fully shrunk towards a common value. Although empirically we do not expect the ICs to have equal absolute Sharpe ratios, this condition provides a sensible shrinkage target because, as explained by Kozak et al. (2020), many asset-pricing models assume that if a factor earns high returns, it must also be a major source of volatility. Indeed, 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,

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

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,

$$\boldsymbol{w}_{MV}^{\star} := \frac{\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{1}_N}{\mathbf{1}_N'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{1}_N}, \tag{19}$$

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 has reduced sensitivity to 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

$$\boldsymbol{w}_{ICMV}^{\star} := (1-\delta)\boldsymbol{w}_{MV}^{\star} + \delta\boldsymbol{w}_{IC}^{\star}, \tag{20}$$

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

We calibrate the shrinkage intensity δ using 10-fold cross-validation, which is a nonparametric approach that does not require any 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 (20) have been considered by several authors in the literature. In particular, DeMiguel et al. (2009b) and Tu and Zhou (2011) 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, which consists of using the *variance* as risk measure. In this section, we show how to extend our ICVP approach to other risk measures. In particular, we seek 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 MVaR accounts for skewness and kurtosis, which we parsimoniously estimate via the ICs.

5.1. Parsimonious estimation of higher moments with ICs

We now show that relying on the ICs allows 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 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 $\tilde{P} = \tilde{\boldsymbol{w}}(\boldsymbol{R})'\boldsymbol{Y}$. From Briec et al. (2007), they can be written concisely as

$$m_3(\tilde{P}) = \tilde{\boldsymbol{w}}(\boldsymbol{R})'\boldsymbol{M}_3(\tilde{\boldsymbol{w}}(\boldsymbol{R}) \otimes \tilde{\boldsymbol{w}}(\boldsymbol{R})), \tag{21}$$

$$m_4(\tilde{P}) = \tilde{\boldsymbol{w}}(\boldsymbol{R})'\boldsymbol{M}_4(\tilde{\boldsymbol{w}}(\boldsymbol{R}) \otimes \tilde{\boldsymbol{w}}(\boldsymbol{R}) \otimes \tilde{\boldsymbol{w}}(\boldsymbol{R})), \tag{22}$$

where $\boldsymbol{M}_3 := \mathbb{E}\big[\overline{\boldsymbol{Y}}\,\overline{\boldsymbol{Y}}' \otimes \overline{\boldsymbol{Y}}'\big]$ and $\boldsymbol{M}_4 := \mathbb{E}\big[\overline{\boldsymbol{Y}}\,\overline{\boldsymbol{Y}}' \otimes \overline{\boldsymbol{Y}}' \otimes \overline{\boldsymbol{Y}}'\big]$, $\overline{\boldsymbol{Y}} := \boldsymbol{Y} - \mu(\boldsymbol{Y})$, are the coskewness and cokurtosis matrices of $\boldsymbol{Y}$ and $\otimes$ is the kronecker product. Boudt et al. (2015) show that estimating $\boldsymbol{M}_3$ and $\boldsymbol{M}_4$ requires estimating many 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, estimating the matrices $\boldsymbol{M}_3$ and $\boldsymbol{M}_4$ requires estimating only $2K = \mathcal{O}(K)$ parameters, which are the third and fourth moments of the K ICs.

Proposition 5. *Let the K ICs be independent and have finite fourth moments. Then, the third and fourth central moments of the reduced portfolio return are given by*

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

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

Although in practice the ICs are not fully independent, the estimators of the third and fourth central moments provided by Proposition 5 can be interpreted as *shrinkage estimators* that use as shrinkage target a K-dimensional *independent* factor model and set the shrinkage intensity to 100%. The independent-factor-model assumption is popular in financial econometrics—see Ghalanos et al. (2015) and Boudt et al. (2020) for applications to portfolio selection—and it is a reasonable shrinkage target in our context because the ICs are maximally independent by construction.

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

We now use Proposition 5 to develop a parsimonious estimator of the factor-risk-parity portfolio 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.

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

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

where z_α is the standard Gaussian quantile and the p_i's are polynomials whose coefficients depend on the standardized moments of $\tilde{P}$. To avoid estimating all moments, 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 (2004) and Haugh et al. (2017). We choose $r = 2$, the modified VaR of Favre and Galeano (2002), which allows us to capture the skewness and kurtosis.

Definition 7 (Modified VaR). The *modified VaR* (MVaR) of the reduced portfolio return is

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

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

Note that, for $\alpha = 1\%$ used in the empirical analysis, the MVaR is a sensible higher-moment criterion because is has the right preferences with respect to skewness and kurtosis: $\frac{\partial \mathrm{MVaR}_\alpha(\tilde{P})}{\partial \zeta(\tilde{P})} \leq 0$ if $\zeta(\tilde{P}) \geq \frac{3(z_\alpha^2-1)}{2z_\alpha^3-5z_\alpha} = -0.9769$, a mild condition, and $\frac{\partial \mathrm{MVaR}_\alpha(\tilde{P})}{\partial \kappa(\tilde{P})} = \frac{1}{24}(3z_\alpha - z_\alpha^3)\sqrt{m_2(\tilde{P})} \geq 0$.

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 K ICs be independent and have finite fourth moments. Then, the MVaR contribution* $\mathrm{MVaR}_\alpha(Y_i^\dagger|\tilde{P})$ *in* (27) *is given by*

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

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 $m_2(\tilde{P})$, $m_3(\tilde{P})$, $m_4(\tilde{P})$ *given by* (11), (23), (24), *respectively.*

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.

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:

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

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 $\epsilon = 5\times 10^{-4}$. This is 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

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

where $\delta \in [0,1]$ is a given shrinkage intensity, $\boldsymbol{w}^{\star}_{MMVaR}(\alpha)$ is the minimum-MVaR portfolio (see El Ghaoui et al. 2003 for a study of the minimum-VaR portfolio) and $\hat{\boldsymbol{w}}^{\star}_{IC}(\alpha)$ is defined in (31).

Section EC.6.5 of the e-companion shows empirically that ignoring the higher comoments of ICs as proposed above leads to superior out-of-sample performance for the ICMVaR portfolio.

6. Out-of-sample performance

We now evaluate the out-of-sample performance of the proposed shrinkage portfolios. Section 6.1 discusses the empirical data and methodology employed, Section 6.2 discusses parameter calibration, and Section 6.3 discusses the results. We discuss several robustness checks in Section 6.4.

6.1. Methodology

6.1.1. Data. 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 use value-weighted daily returns from 1978 to end of 2017. We use daily return data because they produce better estimates of risk than weekly or monthly returns; see Jagannathan and Ma (2003). Daily data also improve the estimation accuracy for PCA and ICA. In the robustness tests of Section 6.4, we also consider a dataset of 100 individual stocks.

6.1.2. Performance criteria. We evaluate out-of-sample performance in terms of mean-variance trade-off, tail risk and transaction costs. 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 mean and MVaR (Gregoriou and Gueyie 2003). For transaction costs, we report the daily turnover, which is the average of the absolute value of daily rebalancing trades across the N assets over the whole period. Standard errors computed via non-parametric bootstrap are available upon request.

6.1.3. Portfolio strategies. We consider several portfolio strategies. First, the minimum-variance asset-variance-parity portfolio (AVP). Bai et al. (2016) show that there are 2^{N-1} AVP 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; we rely instead on the heuristic algorithm of Bai et al. (2016, p.7) that provides a solution close to the minimum-variance AVP portfolio. Second, the minimum-variance (MV) portfolio. Third, the minimum-MVaR (MMVaR) portfolio. Although two common choices of the confidence level are $\alpha = 1\%$ and 5%, we use $\alpha = 1\%$ because Cavenaile and Lejeune (2012) show that setting $\alpha > 4.16\%$ is inconsistent with investors' negative preferences for kurtosis. Fourth and fifth, the shrinkage portfolio obtained by combining the MV and minimum-variance FVP portfolios using the ICs (ICMV) and PCs (PCMV), computed analogously to (31). Sixth, the shrinkage portfolio obtained by combining the MMVaR portfolio and the minimum-MVaR factor-MVaR-parity portfolio based on the ICs (ICMVaR). Finally, the seventh and eighth portfolios are mean-variance-kurtosis efficient portfolios based on the framework of Briec et al. (2007); see Section EC.4 of the e-companion for details on the implementation. This portfolio requires specifying a benchmark portfolio as a starting point. We rely on the equally weighted portfolio as in Boudt et al. (2020a), and also the minimum-variance IC-variance-parity portfolio introduced in this paper. We call these portfolios MVKEW and MVKIC.

We calibrate the shrinkage intensity δ for the ICMV, PCMV and ICMVaR portfolios using 10-fold cross-validation with modified Sharpe ratio as calibration criterion; see Section 6.1.4 for details. To make a fair comparison, the three benchmark portfolios (AVP, MV, MMVaR) are also shrunk toward the equally weighted portfolio using the same calibration approach. We call the resulting benchmark portfolios EWAVP, EWMV and EWMVaR.

All portfolios are computed using sample moment averages. Portfolios that require optimization are found using the Matlab function *fmincon* with the *GlobalSearch* option based on Ugray et al. (2007). We find the ICs with the *FastICA* Matlab package available on Aapo Hyvärinen's website.

6.1.4. Calibration of K and δ. We select the number of PCs retained, K, using the minimum-average-partial-correlation method of Velicer (1976). Let $\boldsymbol{R}(K) := \boldsymbol{R_X} - \boldsymbol{V}\boldsymbol{V}'$ 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 $\sum_{i=1}^{N} \sum_{j\neq i}^{N} R_{ij}(K)^2$.

We calibrate the shrinkage intensity δ via 10-fold cross-validation; see Chapter 7 in Hastie et al. (2009). We use as calibration criterion the modified Sharpe ratio, which accounts for higher moments. Specifically, 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.

6.1.5. Rolling-horizon methodology. To evaluate out-of-sample performance, we use the rolling-horizon methodology commonly employed in the existing literature on static portfolio selection; see, for instance, Jagannathan and Ma (2003), DeMiguel et al. (2009b), Tu and Zhou (2011) and the references therein. 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. This rolling-window methodology helps to address the existence of potential nonstationarities because stock return data is closer to stationary within each estimation window than over the whole sample.

This methodology also means that the number of factors K and the shrinkage intensities δ are calibrated on the training data. However, the cross-validation method to calibrate δ attempts to recreate an out-of-sample calibration and is consistent with other papers in the literature such as DeMiguel et al. (2009a) and Ban et al. (2018).

6.1.6. Statistical significance. 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 $b = 5$. We employ the R package `PeerPerformance` to compute the bootstrapped p-values. The stars $\star$, $\star\star$ and $\star\star\star$ in Table 1 mean that the p-value is less than 10%, 5% and 1%. We compare the following pairs of portfolios: PCMV vs EWMV (stars in superscript next to PCMV quantity), ICMV vs EWMV (stars in superscript next to ICMV quantity), ICMV vs PCMV (stars in subscript next to ICMV quantity) and ICMVaR vs EWMVaR (stars in superscript next to ICMVaR quantity). Other p-values are available upon request.

6.2. Discussion on calibrated K and δ

We now discuss the results from the calibration of the number of factors K and the shrinkage intensities δ. Table 1 shows that 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 firm characteristics: size and book-to-market or size and operating profitability. For the *30Ind* dataset, the average K over the 70 rolling windows is 2.73 and it attains a maximum of $K = 5$. A small K is consistent with the literature that shows that a few PCs can explain the cross-section of stock returns (Kozak et al. 2020). In Section EC.5 of the e-companion, we show that the first K PCs capture a large proportion of the total variance over time. We also show, for the *30Ind* dataset, that K is larger in more volatile periods. Note also that the 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) rejects the Gaussianity of all K PCs for all estimation windows and datasets, with very low associated p-values.

Regarding the shrinkage intensity δ, we observe in Table 1 that it is higher on average for the ICMV 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. In Section EC.5 of the e-companion, we discuss how the shrinkage intensities vary over time.

6.3. Results

Table 1 reports the out-of-sample performance of the eight portfolios for the three datasets. First, we observe that the EWAVP portfolio performs similar to the EWMV portfolio, but with a systematically higher turnover. It is also outperformed by the ICMV portfolio, which shows that diversifying the portfolio-return variance among assets is less useful than among ICs.

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 among ICs effectively reduces tail risk, in line with 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*. Moreover, the ICMV portfolio has lower turnover, which makes its implementation less costly.

Comparing the ICMV and EWMV portfolios, there are clear benefits from shrinking the MV portfolio toward the minimum-variance ICVP portfolio rather than the naively diversified equally weighted portfolio. The ICMV portfolio has a substantially higher mean return for a similar volatility, which results in a Sharpe ratio that is systematically larger. In terms of tail risk, the ICMV portfolio also outperforms the EWMV portfolio in terms of skewness, kurtosis, modified VaR and modified Sharpe ratio. The improvement in the Sharpe ratio and the modified Sharpe ratio is statistically significant. Lastly, the superior performance of the ICMV portfolio is achieved with a turnover that is not much higher than that of EWMV, and thus, remains appealing in practice.

We then turn to the two portfolios based on the modified VaR. Like for the variance risk measure, we find that there are substantial benefits in shrinking the minimum-modified-VaR portfolio toward a portfolio that is diversified among ICs rather than the equally weighted portfolio. Indeed, the ICMVaR shrinkage portfolio has a superior Sharpe ratio and modified Sharpe ratio than EWMVaR on all datasets, the only exception being for the Sharpe ratio for *30Ind*. A drawback of MVaR-based portfolios is however that they have larger turnover than variance-based strategies.

Finally, we consider the two mean-variance-kurtosis efficient portfolios (MVKEW, MVKIC). Interestingly, relying on the MV-ICVP portfolio as benchmark (MVKIC) yields a better performance than relying on the equally weighted portfolio (MVKEW), which indicates its usefulness in terms of higher moments. However, compared to MVKIC, ICMV has a better Sharpe ratio, kurtosis, MVaR and modified Sharpe ratio, with a lower turnover. This is because, as the literature extensively shows, in-sample portfolio efficiency does not necessarily translate in satisfying out-of-sample performance. In contrast, the ICMV portfolio is generally not efficient in sample but is parsimonious as it only requires an estimate of the covariance matrix and the ICA rotation matrix.

6.4. Robustness tests

We now test the robustness of our results. To conserve space, we report here only the main conclusions; detailed tables and discussion can be consulted in Section EC.6 of the e-companion.

In Section EC.6.1, we assess the performance of the portfolios for a high-dimensional dataset of $N = 100$ individual stocks from the Center for Research in Security Prices constructed as in Jagannathan and Ma (2003), DeMiguel et al. (2009a) and Olivares-Nadal and DeMiguel (2018). In this high-dimensional setting, the best performance is achieved for the variance-based rather than MVaR-based portfolios and, among those, the ICMV portfolio has superior Sharpe ratio and tail risk. It also performs similarly to MVKEW and MVKIC.

In Section EC.6.2, we consider the ICMV portfolio based on the worst-case-robust MV portfolio of Goldfarb and Iyengar (2003), which accounts for parameter uncertainty in the asset-return covariance matrix via a minmax formulation. We find that this additional layer of robustness does not help: the worst-case-robust ICMV portfolio is generally outperformed by the non-robust one. Nonetheless, is has less excess kurtosis during the financial crisis, as expected.

In Section EC.6.3, we consider the EWMV, 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).

In Section EC.6.4, we evaluate the performance of the EWMV, ICMV and PCMV portfolios in two subsamples, 1983-1999 and 2000-2012, which from Figure 2 correspond, respectively, to periods of low and high mutual-information reduction by using ICs instead of PCs. We observe as expected that the outperformance of ICMV is particularly significant during the 2000 to 2012 period.

In Section EC.6.5, we compare the performance of the ICMVaR portfolios constructing ignoring and considering the higher co-moments of the ICs. We find that ignoring the higher co-moments of ICs as proposed in Section 5 improves out-of-sample performance and reduces turnover.

In Section EC.6.6, we consider the plain Value-at-Risk (VaR) with $\alpha = 1\%$ as performance criterion, estimated via the empirical quantile. We find that our proposed portfolios are the best in terms of modified Sharpe ratio regardless of whether MVaR or VaR is used to evaluate the modified Sharpe ratio: ICMVaR is the best for *6BTM* and *6Prof* and ICMV for *30Ind*.

In Section EC.6.7, we consider the MVaR and modified-Sharpe-ratio criteria for a more challenging α of 0.5% and 0.1% and find that the results are consistent with those with $\alpha = 1\%$.

7. Conclusion

Factor-risk-parity portfolios spread their risk equally across a set of uncorrelated factors, and thus are typically better diversified in terms of risk than the traditional asset-risk-parity portfolios. A common approach is to use the first K principal components as the uncorrelated factors, but we show that any portfolio in the subspace spanned by the first K principal components is a factor-variance-parity portfolio with respect to a particular rotation of the principal components. This shows that the factor-variance-parity framework is arbitrary when decorrelation is the only criterion governing the choice of factors. Moreover, the standard approach based on principal components is not comprehensive because it ignores the higher-moment dependence in asset returns. Instead, we propose using as factors the independent components, which are the rotation of the principal components 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 generates portfolios with favorable out-of-sample performance in terms of mean-variance trade-off, tail risk and turnover.

Acknowledgments

Part of this research was conducted while N. Lassance was supported by the Fonds de la Recherche Scientifique (F.R.S.-FNRS) under the ASP grant FC 17775. The work of F. Vrins is supported by the research grant ARC 18-23/089 of the Belgian Federal Science Policy Office. The authors are grateful to the area editor (Steven Kou), the associate editor, and three anonymous referees for their comments, which greatly improved the paper. We also 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), 2020 virtual seminar at Universidad Miguel Hernández, Wolfe Research 3rd Annual Virtual Quantitative and Macro Investment Conference, and SYZ Asset Management (London, UK) for comments and discussions.

References

Ablin P, Gramfort A, Cardoso J, Bach F (2019) Stochastic algorithms with descent guarantees for ICA. *Proc. Mach. Learn. Res.* 89:1564–1573.

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

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.

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.

Belouchrani A, Abed-Meraim K, Cardoso J, Moulines E (1997) A blind source separation technique based on second order statistics. *IEEE T. Signal Proces.* 45(2):434–444.

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, Cornilly D, Verdonck T (2020a) A coskewness shrinkage approach for estimating the skewness of linear combinations of random variables. *J. Financ. Econom.* 18(1):1–23.

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

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.

Calafiore G (2007) Ambiguous risk measures and optimal robust portfolios. *SIAM J. Optim.* 18(3):853–877.

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

Cardoso J, Souloumiac A (1993) Blind beamforming for non-Gaussian signals. *IEE Proc-F.* 140(6):362–370.

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.

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

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

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* 58(3):595–612.

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.* 57(3):560–577.

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.

Gao G, Lu X, Song Z (2019) Tail risk concerns everywhere. *Management Sci.* 65(7):3111–3130.

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.

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

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 (2009) *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.

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 (2020) Shrinking the cross-section. *J. Financial Econom.* 135(2):271–292.

Kumiega A, Neururer T, Van Vliet B (2011). Independent component analysis for realized volatility: Analysis of the stock market crash of 2008. *Q. Rev. Econ. Finance* 51(3):292–302.

Lassance N, Vrins F (2021) Portfolio selection with parsimonious higher comoments estimation. *J. Bank. Finance* 126(9):106–115.

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

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

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

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

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.

Popescu I (2007) Robust mean-covariance solutions for stochastic optimization. *Oper. Res.* 55(1):98–112.

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(3):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).

Vrins F, Pham D-T, Verleysen M (2007) Mixing and non-mixing local minima of the entropy contrast for blind source separation. *IEEE T. Inform. Theory* 53(3):1030–1042.

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

Table and figures

Table 1 Out-of-sample performance of portfolio policies.

6BTM dataset (average $K = 2$)

	EWAVP	EWMV	ICMV	PCMV	EWMVaR	ICMVaR	MVKEW	MVKIC
Mean	19.76%	19.63%	21.39%	20.84%	23.27%	**23.93%**	19.93%	20.41%
Volatility	13.75%	13.50%	**13.41%**	13.60%	15.40%	15.14%	13.65%	13.64%
Sharpe ratio	1.44	1.45	**1.59****	1.53	1.51	1.58**	1.46	1.50
Skewness	-0.60	-0.63	-0.33	-0.23	-0.18	**-0.16**	-0.47	-0.44
Excess kurtosis	16.92	17.54	11.53	18.53	8.74	**7.22**	19.97	19.55
1% modified Value-at-Risk	5.63%	5.65%	4.33%	5.75%	4.27%	**3.84%**	6.16%	6.06%
Modified Sharpe ratio ($\times 10^2$)	1.39	1.38	1.96$^{\star\star\star}_{\star}$	1.44	2.16	**2.47***	1.28	1.34
Turnover	2.91%	**2.38%**	2.79%	3.29%	5.67%	5.75%	2.62%	2.84%
Average shrinkage intensity δ	0.05	0.05	0.63	0.59	0.03	0.24	/	/

6Prof dataset (average $K = 2$)

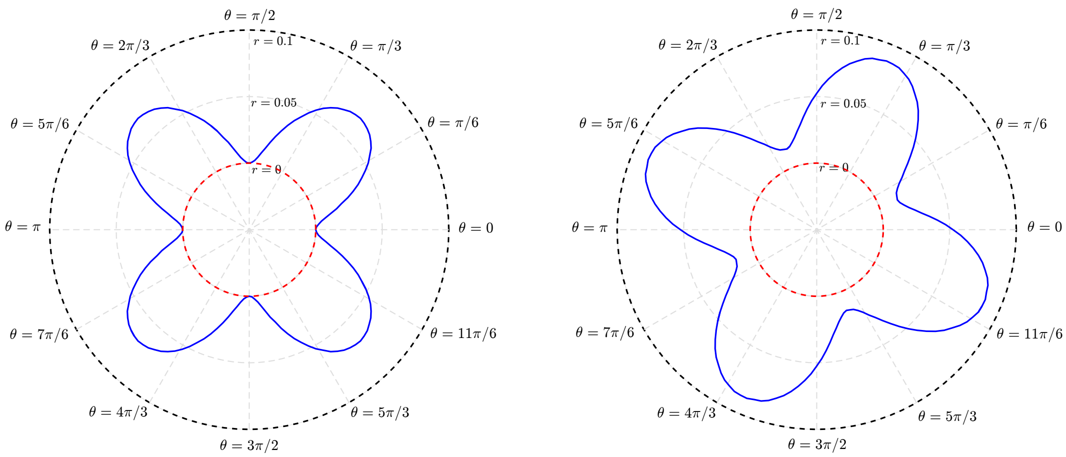
	EWAVP	EWMV	ICMV	PCMV	EWMVaR	ICMVaR	MVKEW	MVKIC
Mean	17.23%	16.93%	20.41%	17.83%	21.81%	**22.55%**	17.22%	19.25%
Volatility	**14.21%**	14.23%	14.83%	14.41%	15.94%	15.60%	14.51%	14.72%
Sharpe ratio	1.21	1.19	1.38$^{\star\star\star}_{\star\star}$	1.24	1.37	**1.45***	1.19	1.31
Skewness	-0.26	-0.37	-0.35	**-0.04**	-0.05	-0.25	-0.07	-0.16
Excess kurtosis	17.00	15.38	14.86	20.06	13.55	**10.99**	19.46	17.59
1% modified Value-at-Risk	5.72%	5.44%	5.53%	6.32%	5.47%	**4.88%**	6.26%	5.99%
Modified Sharpe ratio ($\times 10^2$)	1.20	1.24	1.46**	1.12	1.58	**1.83***	1.09	1.27
Turnover	3.12%	**2.37%**	2.93%	3.09%	4.62%	5.00%	2.70%	3.13%
Average shrinkage intensity δ	0.06	0.08	0.71	0.39	0.11	0.23	/	/

30Ind dataset (average $K = 2.73$)

	EWAVP	EWMV	ICMV	PCMV	EWMVaR	ICMVaR	MVKEW	MVKIC
Mean	12.79%	12.57%	**15.14%**	10.58%	13.16%	11.70%	11.98%	11.97%
Volatility	14.70%	14.39%	14.60%	12.92%	15.00%	14.14%	**12.13%**	12.30%
Sharpe ratio	0.87	0.87	**1.04**$_{\star}$	0.82	0.88	0.83	0.99	0.97
Skewness	-0.72	-0.85	**-0.49**	-0.56	-0.82	-0.78	-0.66	-0.70
Excess kurtosis	21.93	23.07	**12.95**	16.96	22.65	17.46	22.64	19.91
1% modified Value-at-Risk	7.16%	7.27%	**5.11%**	5.32%	7.48%	5.97%	6.02%	5.62%
Modified Sharpe ratio ($\times 10^2$)	0.71	0.69	**1.18****	0.79	0.70	0.78	0.79	0.85
Turnover	4.11%	**2.01%**	2.52%	2.53%	2.78%	3.99%	2.64%	2.99%
Average shrinkage intensity δ	0.34	0.38	0.42	0.27	0.50	0.22	/	/

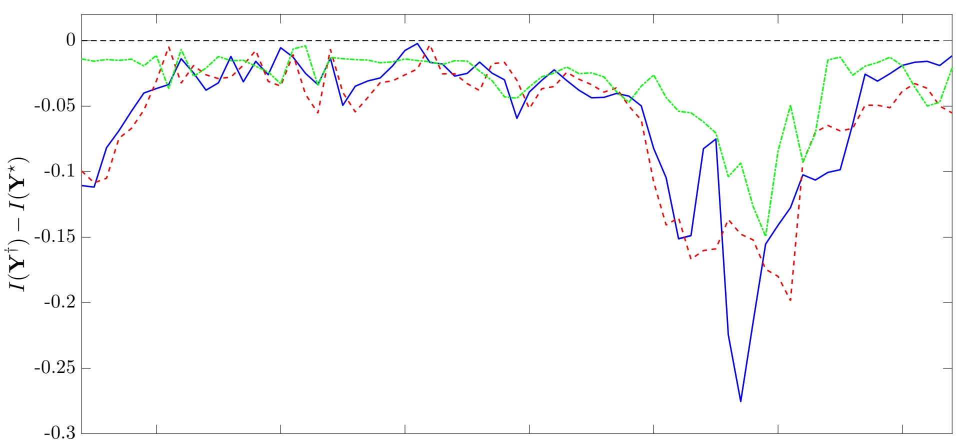
This table reports the out-of-sample performance of the eight portfolio policies on the three datasets following the methodology in Section 6.1. The portfolio mean return, volatility and Sharpe ratio are annualized, while all other performance criteria are in daily terms. The shrinkage intensity δ is 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 and dataset. 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 10%, 5% and 1%. We compare the following pairs of portfolios: PCMV vs EWMV (stars in superscript next to PCMV quantity), ICMV vs EWMV (stars in superscript next to ICMV quantity), ICMV vs PCMV (stars in subscript next to ICMV quantity) and ICMVaR vs EWMVaR (stars in superscript next to ICMVaR quantity).

Figure 1 Mutual information versus correlation.



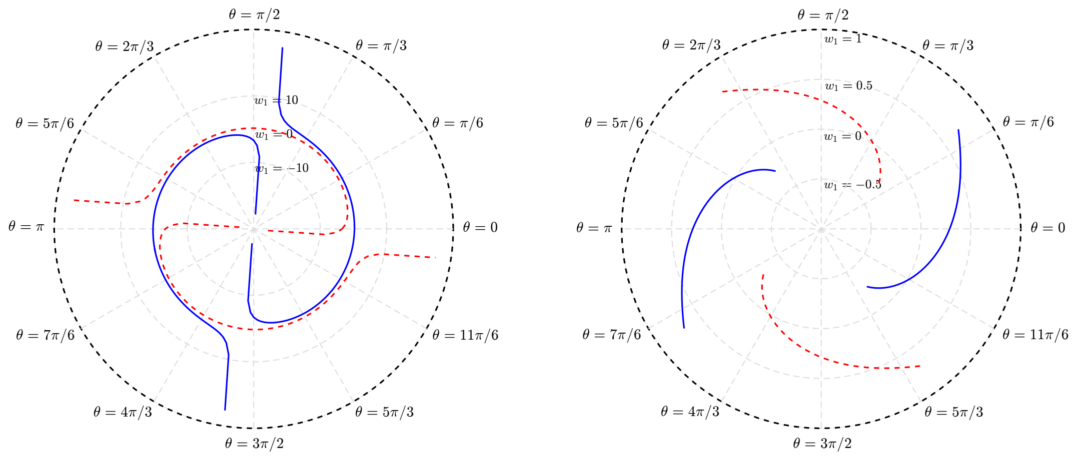
Note. The figure depicts the relative mutual information (in solid blue) and correlation (in dashed red) of the $K = N = 2$ factors $\boldsymbol{Y}$ obtained as a rotation of the principal components, $\boldsymbol{Y} = \boldsymbol{R}(\theta)\boldsymbol{Y}^\star$. The left plot considers the case with independent non-Gaussian asset returns and the right plot 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 (4) and r is the radial distance from the origin $(\theta, 0)$. The principal components correspond to $\theta = 0$. The independent components $\boldsymbol{Y}^\dagger$ are obtained for $\theta = 2k\pi$ in the left plot, and $\theta = 0.35 + 2k\pi$ in the right plot, $k \in \mathbb{Z}$. Note that, for clarity of the figures, the center of the polar plots corresponds to $r = -0.05$ rather than $r = 0$.

Figure 2 Mutual information of independent versus principal components.



Note. 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: six size and book-to-market portfolios (solid blue), six size and operating-profitability portfolios (dashed red) and 30 U.S. industry portfolios (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 last year of the window; for instance, the value at the date 2005 is computed from data from 2001 to 2005.

Figure 3 Arbitrariness of the factor-variance-parity portfolio.



Note. We consider the $N = 2$ asset returns in Example 3 and $K = 2$ 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. The left plot depicts the weight on the first asset, w_1, of the two FVP portfolios (in solid blue and dashed red, respectively) as a function of θ. w_1 continues up to $\pm\infty$ at the asymptotes. The right plot depicts the weight on the first asset, w_1, of the FVP portfolio with minimum return variance as a function of θ. The plots are in polar coordinates (θ, r), where θ is the angle determining the rotation matrix $\boldsymbol{R}(\theta)$ in (4) and $r = w_1$ is the radial distance from the origin $(\theta, 0)$, which can be negative. The PCs correspond to $\theta = 0$.

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

This e-companion contains six sections. Section EC.1 gives proofs of all results in the paper. Section EC.2 reviews the literature on when the non-Gaussianity assumption is likely to hold. Section EC.3 describes the *FastICA* algorithm. Section EC.4 details the implementation of the MVKEW and MVKIC portfolios. Section EC.5 displays several figures related to the empirical analysis. Section EC.6 provides detailed results on the robustness tests of our empirical results.

EC.1. Proofs of all results

EC.1.1. 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, where $c \in \mathbb{R}_0$ is a constant ensuring 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 (9), 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 (12) 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 sign vector $\mathbf{1}_K^{\pm}$, the variance of the FVP portfolio return is given by

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

where $\boldsymbol{V}_N$ and $\boldsymbol{\Lambda}_N$ are the full $N \times N$ eigenvectors and eigenvalues matrices. The matrix $\boldsymbol{V}'\boldsymbol{V}_N\boldsymbol{\Lambda}_N\boldsymbol{V}_N'\boldsymbol{V}$ is the covariance matrix of the PCs $\boldsymbol{Y}^{\star} = \boldsymbol{V}'\boldsymbol{X}$, and thus,

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

As a result, (EC.1) simplifies to

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

Clearly, the sign vector $\mathbf{1}_K^{\pm}$ that minimizes (EC.2) corresponds to

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

□

EC.1.2. Proposition 2

Proof. To prove the proposition, we need to show that given any portfolio $\boldsymbol{w} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\tilde{\boldsymbol{w}}(\boldsymbol{I}_K) \in \mathcal{W}_K$, there exists an orthogonal matrix $\boldsymbol{R}$ and a scalar $c \in \mathbb{R}_0$ such that $\boldsymbol{w}$ is an FVP portfolio on the factors $\boldsymbol{R}\boldsymbol{Y}^\star$; that is, such that

$$\boldsymbol{w} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\tilde{\boldsymbol{w}}(\boldsymbol{I}_K) = c\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\mathbf{1}_K^{\pm}. \tag{EC.3}$$

It is sufficient to find an orthogonal matrix $\boldsymbol{R}$ because if $\det(\boldsymbol{R}) = -1$, one can just change the sign of one of its rows; the resulting matrix would be a rotation matrix belonging to $\mathcal{SO}(K)$ and, following Proposition 1(ii), would lead to the same FVP portfolios. Equation (EC.3) amounts to showing that there exists an orthogonal matrix $\boldsymbol{R}$ and $c \in \mathbb{R}_0$ such that

$$\tilde{\boldsymbol{w}}(\boldsymbol{I}_K) = c\boldsymbol{R}\mathbf{1}_K^{\pm}. \tag{EC.4}$$

We have that $||\mathbf{1}_K^{\pm}|| = \sqrt{K}$ and the scaling coefficient c given by

$$c = \frac{||\tilde{\boldsymbol{w}}(\boldsymbol{I}_K)||}{\sqrt{K}}$$

means that (EC.4) can be written as

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

where $\boldsymbol{x} := \mathbf{1}_K^{\pm}/\sqrt{K}$ and $\boldsymbol{y} := \tilde{\boldsymbol{w}}(\boldsymbol{I}_K)/||\tilde{\boldsymbol{w}}(\boldsymbol{I}_K)||$ are unit-norm vectors. The claim follows because given two arbitrary unit-norm vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^K$, there always exists an orthogonal matrix $\boldsymbol{R}$ such that $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{y}$. To see this, define $\tilde{\boldsymbol{y}} := (\boldsymbol{y} - (\boldsymbol{x}\boldsymbol{y}')\boldsymbol{x})/||\boldsymbol{y} - (\boldsymbol{x}\boldsymbol{y}')\boldsymbol{x}||$, so that $(\boldsymbol{x}, \tilde{\boldsymbol{y}})$ forms an orthonormal basis in the plane spanned by $(\boldsymbol{x}, \boldsymbol{y})$. The matrix $\boldsymbol{R}$ takes the form $\boldsymbol{P} + \boldsymbol{Q}$ where

$$\boldsymbol{P} := \boldsymbol{I}_K - \boldsymbol{x}\boldsymbol{x}' - \tilde{\boldsymbol{y}}\tilde{\boldsymbol{y}}' \quad \text{and} \quad \boldsymbol{Q} := [\boldsymbol{x}\ \tilde{\boldsymbol{y}}]\boldsymbol{R}(\theta)[\boldsymbol{x}\ \tilde{\boldsymbol{y}}]' \quad \text{with} \quad \theta := \arccos(\boldsymbol{x}'\boldsymbol{y}).$$

The orthogonality of $\boldsymbol{R}$ can be checked by observing that $\boldsymbol{P}$ is a projection matrix (hence, $\boldsymbol{P}\boldsymbol{P}' = \boldsymbol{P}$) and that $\boldsymbol{Q}\boldsymbol{Q}' = \boldsymbol{x}\boldsymbol{x}' + \tilde{\boldsymbol{y}}\tilde{\boldsymbol{y}}' = \boldsymbol{I}_K - \boldsymbol{P}$ whereas $\boldsymbol{P}\boldsymbol{Q}' = \boldsymbol{Q}\boldsymbol{P}' = \mathbf{0}$. □

EC.1.3. Arbitrariness of FVP portfolios with minimum return variance

In this section, we formally state and prove the result mentioned in Section 3.2 in the paper that in the case where $K = 2$ and the covariance matrix is diagonal, any portfolio of the two assets with largest variance, whose weights satisfy certain bounds, is an FVP portfolio with minimum return variance (MV-FVP portfolio) for some rotation of the PCs.

We expect the range of portfolios spanned by MV-FVP portfolios to be smaller than when choosing a fixed sign vector $\mathbf{1}_K^{\pm}$ as in Proposition 2 because, out of the 2^{K-1} FVP portfolios given a rotation $\boldsymbol{R}$ of the PCs, we choose the one that has the lowest variance, which dismisses many portfolios whose weights are too extreme. Identifying the range of portfolios spanned by MV-FVP portfolios is more complex because the sign vector $\mathbf{1}_K^{MV}$ in (13) also depends on $\boldsymbol{R}$. The proposition below states a specific result for $K=2$ and a diagonal covariance matrix.

Proposition EC.1. *Let $K=2$, $\boldsymbol{\Sigma_X}$ be diagonal and suppose, without loss of generality, that the assets are sorted in decreasing order of variance: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_N$, $N \geq 2$. Define*

$$\widetilde{\mathcal{W}}_2 := \left\{ \boldsymbol{w} \in \mathcal{W} \;\middle|\; \boldsymbol{w} = (w_1, w_2, 0, \ldots, 0), \quad w_1 \in \left[\frac{\sigma_2(\sigma_2 - \sigma_1)}{\sigma_1^2 + \sigma_2^2}, \frac{\sigma_2(\sigma_1 + \sigma_2)}{\sigma_1^2 + \sigma_2^2} \right] \right\}, \tag{EC.6}$$

a subset of portfolios investing exclusively in the first two assets according to some constraints. Then, any portfolio $\boldsymbol{w} \in \widetilde{\mathcal{W}}_2$ is an MV-FVP portfolio for factors given by some rotation of the K principal components.

Proof. We begin with an intermediary result. Under the assumption that $\boldsymbol{\Sigma_X}$ is diagonal and sorted in decreasing order of variance, the $N \times K$ eigenvectors matrix $\boldsymbol{V}$ corresponds to the first K columns of $\boldsymbol{I}_N$. It follows that every FVP portfolio constructed from factors that are a rotation of the first K PCs takes the form $\boldsymbol{w} = c\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\mathbf{1}_K^{\pm} = (\boldsymbol{w}_1, \boldsymbol{w}_2)'$, where $\boldsymbol{w}_1 = c\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\mathbf{1}_K^{\pm} \in \mathbb{R}^K$, $\boldsymbol{w}_2 = \mathbf{0} \in \mathbb{R}^{N-K}$, c is a normalization constant and $\boldsymbol{R} \in \mathcal{SO}(K)$. From Proposition 2 for the case $K=N$, this set of portfolios corresponds to the set of portfolios investing in the first K assets only.

We now turn to the main proposition. Because $K=2$ and $\boldsymbol{\Sigma_X}$ is diagonal and sorted in decreasing order of variance, we know from the intermediary result above that any portfolio $\boldsymbol{w} \in \mathcal{W}$ investing exclusively in the first two assets is an FVP portfolio for some $\boldsymbol{R} \in \mathcal{SO}(K)$. Moreover, the PCs $\boldsymbol{Y}^\star$ correspond to the first two standardized asset returns. Let us now focus on the first portfolio weight, w_1. For $K=2$, $\boldsymbol{R} = \boldsymbol{R}(\theta)$ in (4), and it is easy to check that

$$\mathbf{1}_K^{MV} = \operatorname{sign}\left(\begin{pmatrix} \frac{\cos\theta}{\sigma_1} - \frac{\sin\theta}{\sigma_2} \\ \frac{\sin\theta}{\sigma_1} + \frac{\cos\theta}{\sigma_2} \end{pmatrix} \right).$$

Therefore, w_1 is given by the first entry of the vector

$$c\boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}'\mathbf{1}_K^{MV} = c \begin{pmatrix} \frac{\cos\theta}{\sigma_1} & \frac{\sin\theta}{\sigma_1} \\ -\frac{\sin\theta}{\sigma_2} & \frac{\cos\theta}{\sigma_2} \end{pmatrix} \operatorname{sign}\left(\begin{pmatrix} \frac{\cos\theta}{\sigma_1} - \frac{\sin\theta}{\sigma_2} \\ \frac{\sin\theta}{\sigma_1} + \frac{\cos\theta}{\sigma_2} \end{pmatrix} \right).$$

After some developments, we find that

$$w_1 = \frac{\cos\theta \times \operatorname{sign}\left(\frac{\cos\theta}{\sigma_1} - \frac{\sin\theta}{\sigma_2}\right) + \sin\theta \times \operatorname{sign}\left(\frac{\sin\theta}{\sigma_1} + \frac{\cos\theta}{\sigma_2}\right)}{\left|\cos\theta - \frac{\sigma_1}{\sigma_2}\sin\theta\right| + \left|\sin\theta + \frac{\sigma_1}{\sigma_2}\cos\theta\right|}. \tag{EC.7}$$

To find the bounds of w_1 as a function of θ, we need to look the value $\theta_{\lim}$ of θ for which one of the sign quantities in the numerator of (EC.7) switches sign. Indeed, a switch of sign indicates that the MV-FVP portfolio switches from one of the two FVP portfolios to the other, and thus, that w_1 has attained its minimum or maximum value. Formally, it is easy to verify that, within the bounds derived below, w_1 is always an increasing function of θ. Thus, w_1 achieves its maximum (minimum) by taking the left (right) limit as $\theta \to \theta_{\text{lim}}$. Focusing on the first sign quantity in the numerator of (EC.7), we find that $\frac{\cos\theta}{\sigma_1} - \frac{\sin\theta}{\sigma_2} = 0$ for $\theta_{\text{lim}} = \tan^{-1}(\sigma_2/\sigma_1)$. The limit of w_1 as $\theta \to \theta_{\text{lim}}$ takes two different values for the left and right limits, corresponding to $\text{sign}\big(\frac{\cos\theta}{\sigma_1} - \frac{\sin\theta}{\sigma_2}\big) = \pm 1$. Thus, we obtain the following two bounds for w_1 by taking the limit as $\theta \to \theta_{\text{lim}}$ in (EC.7):

$$w_1^{\pm} = \frac{\pm\cos\theta_{\text{lim}} + \sin\theta_{\text{lim}} \times \text{sign}\big(\frac{\sin\theta_{\text{lim}}}{\sigma_1} + \frac{\cos\theta_{\text{lim}}}{\sigma_2}\big)}{\big|\cos\theta_{\text{lim}} - \frac{\sigma_1}{\sigma_2}\sin\theta_{\text{lim}}\big| + \big|\sin\theta_{\text{lim}} + \frac{\sigma_1}{\sigma_2}\cos\theta_{\text{lim}}\big|}. \tag{EC.8}$$

Finally, using the identities $\cos(\tan^{-1}(x)) = (x^2+1)^{-1/2}$ and $\sin(\tan^{-1}(x)) = x(x^2+1)^{-1/2}$, we obtain after simplifications that $w_1^{\pm}$ in (EC.8) correspond to the bounds in $\widetilde{\mathcal{W}}_2$ in (EC.6). □

EC.1.4. Proposition 3

Proof. First, we explain why assuming that the N asset returns are given by a linear combination of K standardized independent factors amounts to the factor model $\boldsymbol{X} = \boldsymbol{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{R}^{\dagger\prime}\boldsymbol{Y}^{\dagger}$. The general independent factor model for the asset returns $\boldsymbol{X}$ assumed in the proposition is given by

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}, \tag{EC.9}$$

where $\boldsymbol{A}$ is a full-rank $N \times K$ matrix, $\boldsymbol{\Sigma_S} = \boldsymbol{I}_K$, $I(\boldsymbol{S}) = 0$ 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 indeed find, as claimed in the statement of the proposition, that the factor model becomes $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{R}^{\dagger\prime}$.

From (8), 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:

$$\tilde{P} = \tilde{\boldsymbol{w}}(\boldsymbol{R}^{\dagger})'\boldsymbol{Y}^{\dagger} = \boldsymbol{w}'\boldsymbol{A}\boldsymbol{Y}^{\dagger} = \boldsymbol{w}'\boldsymbol{A}\boldsymbol{S} = \boldsymbol{w}'\boldsymbol{X} = P. \tag{EC.10}$$

This is a natural result as the asset returns follow an exact K-dimensional factor model without errors. It 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:

$$\kappa(P) = \frac{\sum_{i=1}^{K}\kappa_i\tilde{w}_i(\boldsymbol{R}^{\dagger})^4}{\Big(\sum_{i=1}^{K}\tilde{w}_i(\boldsymbol{R}^{\dagger})^2\Big)^2} = \frac{\sum_{i=1}^{K}\kappa_i z_i^2}{\Big(\sum_{i=1}^{K} z_i\Big)^2}, \tag{EC.11}$$

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 (EC.11), 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

$$\kappa(P) = \sum_{i=1}^{K} \kappa_i z_i^2 = \boldsymbol{z}'\mathcal{K}\boldsymbol{z}, \tag{EC.12}$$

where $\boldsymbol{z} = (z_1, \ldots, z_K)'$ and $\mathcal{K} := \text{diag}(\kappa_1, \ldots, \kappa_K)$. 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:

$$\min_{\boldsymbol{z}} \boldsymbol{z}'\mathcal{K}\boldsymbol{z} \quad \text{subject to} \quad \mathbf{1}_K'\boldsymbol{z} = 1,\ \boldsymbol{z} \geq 0. \tag{EC.13}$$

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 (14):

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

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

$$\max_{\boldsymbol{z}} \boldsymbol{z}'\mathcal{K}\boldsymbol{z} \quad \text{subject to} \quad \mathbf{1}_K'\boldsymbol{z} = 1,\ \boldsymbol{z} \geq 0. \tag{EC.14}$$

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 (15).

(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 (EC.11) 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, depending on the rotation matrix $\boldsymbol{R}^\dagger$ hidden behind the mixing matrix $\boldsymbol{A}$ associated with the linear-mixture model $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}$, the FVP portfolio associated with $\boldsymbol{Y}$ could be any portfolio, and thus, could have any attainable kurtosis.

To show this, we first need to show that any portfolio return P is an 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 holds by the same reasoning than in the proof of Proposition 2.

Suppose now without loss of generality that a singular-value decomposition of the mixing matrix is of the form $\boldsymbol{A} = \boldsymbol{R}_1 \boldsymbol{D} \boldsymbol{R}' \tilde{\boldsymbol{R}}$ where $\boldsymbol{R}_1$ is a $N \times K$ orthogonal matrix, $\boldsymbol{D}$ is a $K \times K$ diagonal matrix, and $\boldsymbol{R}, \tilde{\boldsymbol{R}}$ are defined above. It results that the ICA rotation matrix is $\boldsymbol{R}^\dagger = \tilde{\boldsymbol{R}}' \boldsymbol{R}$. Then, it is easy to see that the factors $\boldsymbol{Y}$ correspond to $\boldsymbol{Y} = \tilde{\boldsymbol{R}} \boldsymbol{Y}^\dagger$. Therefore, 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 the FVP portfolio return associated with the factors $\boldsymbol{Y}$. In other words, depending on the ICA rotation matrix $\boldsymbol{R}^\dagger$ hidden behind the mixing matrix $\boldsymbol{A}$, 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 (16). □

EC.1.5. 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 3, 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 a normalization constant. Then, from (EC.11), 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 (17). □

EC.1.6. Proposition 4

Proof. The tangent mean-variance portfolio computed from the asset returns spanned by the K principal components, $\tilde{\boldsymbol{X}}$ in (6), is given by

$$\boldsymbol{w} = \frac{\boldsymbol{\Sigma}_{\tilde{\boldsymbol{X}}}^{-1} \boldsymbol{\mu}_{\tilde{\boldsymbol{X}}}}{\mathbf{1}_N' \boldsymbol{\Sigma}_{\tilde{\boldsymbol{X}}}^{-1} \boldsymbol{\mu}_{\tilde{\boldsymbol{X}}}}. \tag{EC.15}$$

Because $\tilde{\boldsymbol{X}} = \boldsymbol{V}\boldsymbol{V}'\boldsymbol{X}$, we have $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{X}}}^{-1} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}'$ and $\boldsymbol{\mu}_{\tilde{\boldsymbol{X}}} = \boldsymbol{V}\boldsymbol{V}'\boldsymbol{\mu}_{\boldsymbol{X}}$. Thus, the tangent mean-variance portfolio in (EC.15) simplifies to

$$\boldsymbol{w} = \frac{\boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}'\boldsymbol{\mu}_{\boldsymbol{X}}}{\mathbf{1}_N' \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}'\boldsymbol{\mu}_{\boldsymbol{X}}}, \tag{EC.16}$$

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

$$\tilde{\boldsymbol{w}}(\boldsymbol{R}^{\dagger}) = \boldsymbol{R}^{\dagger}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}'\boldsymbol{w} = \frac{\boldsymbol{R}^{\dagger}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}'\boldsymbol{\mu}_{\boldsymbol{X}}}{\boldsymbol{1}_N'\boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}'\boldsymbol{\mu}_{\boldsymbol{X}}}. \tag{EC.17}$$

Given that the ICs are defined as $\boldsymbol{Y}^{\dagger} = \boldsymbol{R}^{\dagger}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}'\boldsymbol{X}$, the numerator of (EC.17) 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 mean-variance portfolio is an IC-variance-parity portfolio if each IC has the same mean return in absolute value:

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

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

EC.1.7. 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:

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

□

EC.1.8. Proposition 6

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

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

where the portfolio-return central moments are given by

$$m_2(P) = \boldsymbol{w}'\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{w}, \quad m_3(P) = \boldsymbol{w}'\boldsymbol{M}_3(\boldsymbol{w}\otimes\boldsymbol{w}) \quad \text{and} \quad m_4(P) = \boldsymbol{w}'\boldsymbol{M}_4(\boldsymbol{w}\otimes\boldsymbol{w}\otimes\boldsymbol{w}), \tag{EC.19}$$

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 (EC.18)-(EC.19) 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). □

EC.2. When is the non-Gaussianity assumption likely to hold?

In this section, we discuss the evidence in the literature on when we can expect asset returns to deviate from normality, and thus, that the non-Gaussianity assumption stated in Section 2.2 in the paper is likely to hold. We have surveyed the literature and we have identified four general themes:

1. *Return frequency.* High-frequency returns deviate more from the Gaussian distribution because there is less time aggregation. Martellini and Ziemann (2010, p.1497) for example note that "[...] daily returns are notoriously exhibiting stronger deviations from normality compared to weekly and monthly returns." Cont (2001, p.224) lists this as a stylized fact called aggregational Gaussianity: "[...] as one increases the time scale Δt over which returns are calculated, their distribution looks more and more like a normal distribution."

2. *Macroeconomic conditions.* Using monthly returns of the S&P500, Gormsen and Jensen (2020, p.2) show that higher-moment risk is higher in bull versus bear markets: "The higher order moments vary substantially with the state of the financial markets. In bad times, such as during the financial crisis, there is hardly any higher-moment risk. Excess kurtosis is low and skewness is close to zero, meaning that the return distribution is well approximated by a normal distribution. But going into good times, the shape of the return distribution changes. The distribution becomes left-skewed and fat tailed, meaning additional risk builds in the higher order moments." We also note that, while the *marginal* distributions may become closer to a Gaussian in bad periods, there is extensive evidence that left-tail-risk *dependence* between asset returns increases in bad times (Massacci 2017), and such asymmetric dependence dismisses the multivariate Gaussian distribution.

3. *Market geography.* Comparing monthly returns of the U.S. market and emerging markets (e.g., Argentina, Mexico, South Korea, Thailand), Bekaert et al. (1998) find that emerging-market returns deviate more strongly from Gaussianity.

4. *Asset classes.* The literature shows that the returns of several different asset types (equity, bonds, currencies, hedge funds) all have heavy-tailed. This is extensively documented for equities and hedge funds and, in a recent paper called "Tail risk concerns everywhere", Gao et al. (2019) show that tail risk matters more generally not only for equities but also for bonds and currencies.

For a comprehensive list of papers that investigate the importance of higher moments in financial applications, we also refer to Harvey et al. (2010).

EC.3. 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. In 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 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 $\boldsymbol{R}$. Because the second term does not depend on $\boldsymbol{R}$, minimizing $I(\boldsymbol{RY}^\star)$ over the set of rotation matrices $\mathcal{SO}(K)$ 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 $Y_1,\dots,Y_K$.

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, maximizing $\sum_{i=1}^{K}J(\boldsymbol{R}_i\boldsymbol{Y}^\star)$ is equivalent to minimizing $I(\boldsymbol{RY}^\star)$ 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 $\{\boldsymbol{R} \in \mathcal{SO}(K) \mid I(\boldsymbol{R}\boldsymbol{Y}^\star) = 0\}$. 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 kurtosis 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 high sensitivity to outliers 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 et al. 2007). 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 in sample estimates of 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).

EC.4. Implementation of MVKEW and MVKIC portfolios

In this section, we explain how the MVKEW and MVKIC portfolios we consider in the empirical analysis of Section 6 are implemented. These two portfolios are mean-variance-kurtosis efficient-portfolio strategies based on Briec et al. (2007). These authors introduced a method whereby one searches for the largest improvements in mean, variance and skewness relative to a given benchmark portfolio $\boldsymbol{w}_b$. This approach is appealing because it is guaranteed to give a global solution, it does not require specifying a utility function, and it ensures that the portfolio is located on the mean-variance-skewness efficient surface. Because, as we show both theoretically and empirically, our IC-risk-parity approach performs particularly well in terms of kurtosis, we consider a mean-variance-kurtosis version of the portfolio of Briec et al. (2007) that is defined by the following optimization problem:

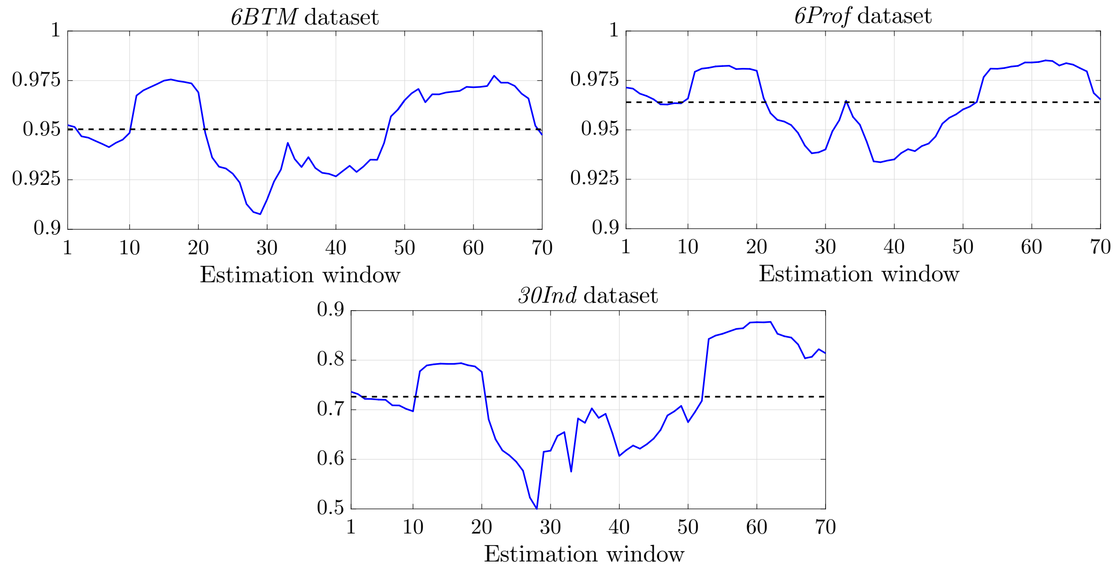
$$\begin{aligned} \max_{\delta \in [0,1], \boldsymbol{w} \in \mathcal{W}} \quad & \delta \\ \text{subject to} \quad & \mu(\boldsymbol{w}'\boldsymbol{X}) \geq \mu(\boldsymbol{w}_b'\boldsymbol{X}), \\ & m_2(\boldsymbol{w}'\boldsymbol{X}) \leq m_2(\boldsymbol{w}_b'\boldsymbol{X})(1-\delta), \\ & m_4(\boldsymbol{w}'\boldsymbol{X}) \leq m_4(\boldsymbol{w}_b'\boldsymbol{X})(1-\delta), \end{aligned}$$

where $m_k(\cdot)$ is the kth central moment. We rely on two different benchmarks $\boldsymbol{w}_b$. First, as in Boudt et al. (2020) for example, the equally weighted portfolio, which represents the uninformed choice (MVKEW). Second, we also consider the minimum-variance IC-variance-parity portfolio introduced in the paper (MVKIC).

EC.5. Further empirical analysis

In this section, we display several figures that provide additional information about the empirical analysis of Section 6. First, Figure EC.1 reports the proportion of variance explained by the first K principal components over time. We observe that, for the *6BTM* and *6Prof* datasets, a very large proportion of the variance is explained by the first K PCs. This makes sense because these two datasets are driven by two firm characteristics (size and book-to-market and size and operating profitability) and the method we use always selects $K = 2$ for these two datasets. For the *30Ind* dataset, the proportion of variance explained is smaller but still remains at 72% on average. Therefore, we can be confident that the method we use to select K captures an important proportion of the variance of the entire set of variables. At the same time, K remains quite small relative to N, which reduces the sensitivity to estimation risk thanks to dimension reduction. Note that

Figure EC.1 Proportion of variance explained by the K principal components over time.



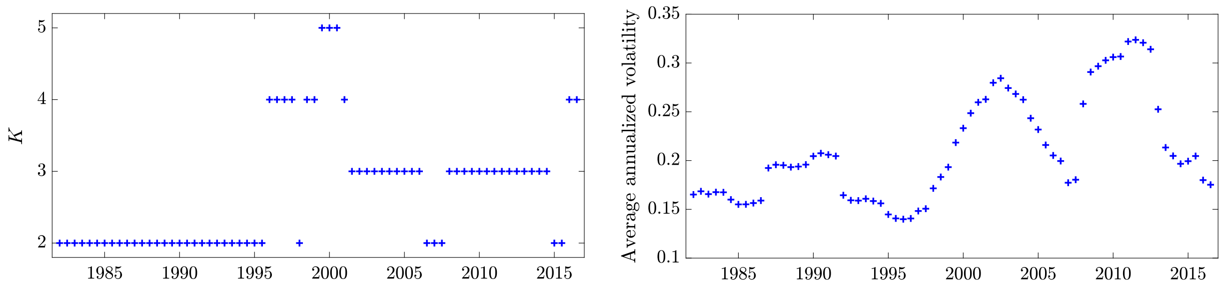
Note. The figure depicts the proportion of variance explained by the first K principal components estimated every six months using an estimation window containing the previous five years of daily asset returns. The dotted black line depicts the average over all windows.

the advantage of the method of Velicer (1976) is that it does not require an arbitrary choice of variance-explained threshold (e.g., above 90%) to fix K.

Second, Figure EC.2 reports the time evolution of K for the *30Ind* dataset (note that K is always equal to two for the *6BTM* and *6Prof* datasets). In the same figure, we report the average annualized volatility of the 30 industry returns over time. The figure clearly shows that volatility and K are positively correlated: on the low-volatility period from 1978 to 1995, K remains equal to two, whereas on the subsequent period 1996 to 2017, which is characterized by a larger volatility, K varies between two and five.

Third, Figure EC.3 reports the time evolution of the shrinkage intensity δ for the three datasets. We focus on the ICMV portfolio and not ICMVaR for brevity. We make two observations. On the one hand, δ varies quite substantially over time, which can be explained by evolving asset-return dynamics and also by the use of the modified Sharpe ratio as cross-validation calibration criterion, which depends on higher moments that also vary substantially over time. However, we note that the time-variation of δ is not problematic because Table 1 in the paper shows that the resulting ICMV shrinkage portfolio performs well out of sample and has a turnover that is not much higher that of the EWMV portfolio. On the other hand, the constraint $\delta \in [0,1]$ is often binding, and it is more often equal to one for the small-dimensional datasets (*6BTM*, *6Prof*) than the higher-dimensional datasets (*30Ind*). The explanation for this phenomenon is that δ is determined as a trade-off between two effects. Firstly, which of the MV and MV-ICVP portfolios has the largest

Figure EC.2 Time evolution of K and average asset-return volatility for the *30Ind* dataset.

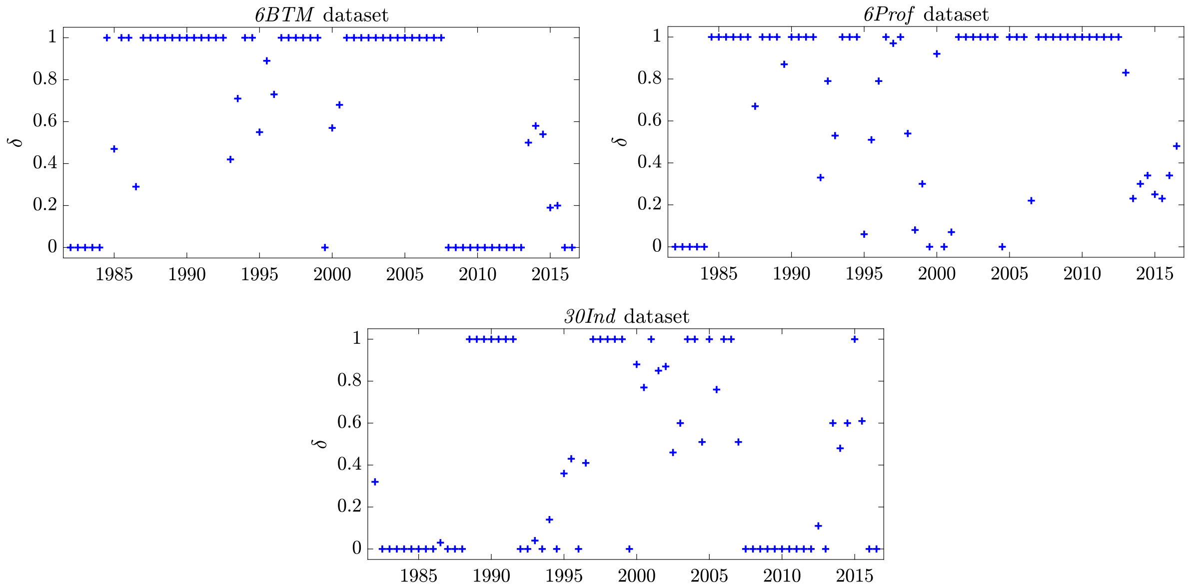


Note. This figure depicts the time evolution of the number of factors K (left graph) and of the average annualized volatility of the 30 industry returns (right graph) estimated every six months using an estimation window containing the previous five years of daily asset returns. The x-axis labels indicate the last year of the window; for instance, the value at the date 2005 is computed from data from 2001 to 2005.

in-sample modified Sharpe ratio. The MV-ICVP is expected to perform best in this dimension because, as shown in Proposition 3 in the paper, it accounts for portfolio-return kurtosis. This effect will drive δ toward one. Secondly, because cross-validation is an out-of-sample calibration method, which of the MV and MV-ICVP portfolios is less sensitive to estimation risk. The MV portfolio is expected to be less sensitive to estimation risk because it does not account for higher moments. This effect will drive δ toward zero. The fact that δ often equals zero or one means that, in many estimation windows, one of the two effects dominates the other. Moreover, the fact that δ is more often equal to one for *6BTM* and *6Prof* makes sense as there is less estimation risk when the dimension N is smaller for a fixed sample size.

Fourth, Figure EC.4 complements Figure 2 in the main body of the manuscript, which reports the mutual information of the PCs and ICs. Although mutual information is a useful dependence criterion in the context of ICA, it is not easy to interpret: in Figure 2 in the main body of the paper, it is not obvious whether a reduction of mutual information of 0.1 by going from the PCs to the ICs is substantial or not, for example. To ease the interpretation, Figure EC.4 depicts the average of the pairwise non-linear correlations between the ICs and the PCs. That is, for each possible pair among the K ICs or K PCs, we compute the pairwise non-linear correlation, and we report the average across all pairs. We define the non-linear correlation as the Pearson linear correlation applied to a transform G of the ICs or PCs, and we take $G(x) = \ln\cosh x$ as in *FastICA*. This yields a number between -1 and 1 that is more easily interpretable. The figure is in line with the observations based on mutual information in Figure 2 in the main body of the paper. That is, the reduction in non-linear correlation allowed by the ICs starts to increase in the period preceding the crisis. This indicates that there is increased non-Gaussianity among asset returns in this period, which ICs leverage upon in contrast to PCs. This observation is consistent with Gormsen and

Figure EC.3 Time evolution of the ICMV shrinkage intensity δ.



Note. The shrinkage intensity δ of the ICMV portfolio is computed every six months using an estimation window containing the previous five years of daily asset returns. The x-axis labels indicate the last year of the window; for instance, the value at the date 2005 is computed from data from 2001 to 2005.

Jensen (2020), who find that higher-moment risk increases in bull markets where variance is low. The period leading up to the 2007-2009 financial crisis is such a period. When the financial crisis enters into the five-year rolling window, non-linear correlation among ICs and PCs increases, but the reduction allowed by the ICs remains important. We refer to Kumiega et al. (2011) for the economic interpretation of ICs during the financial crisis.

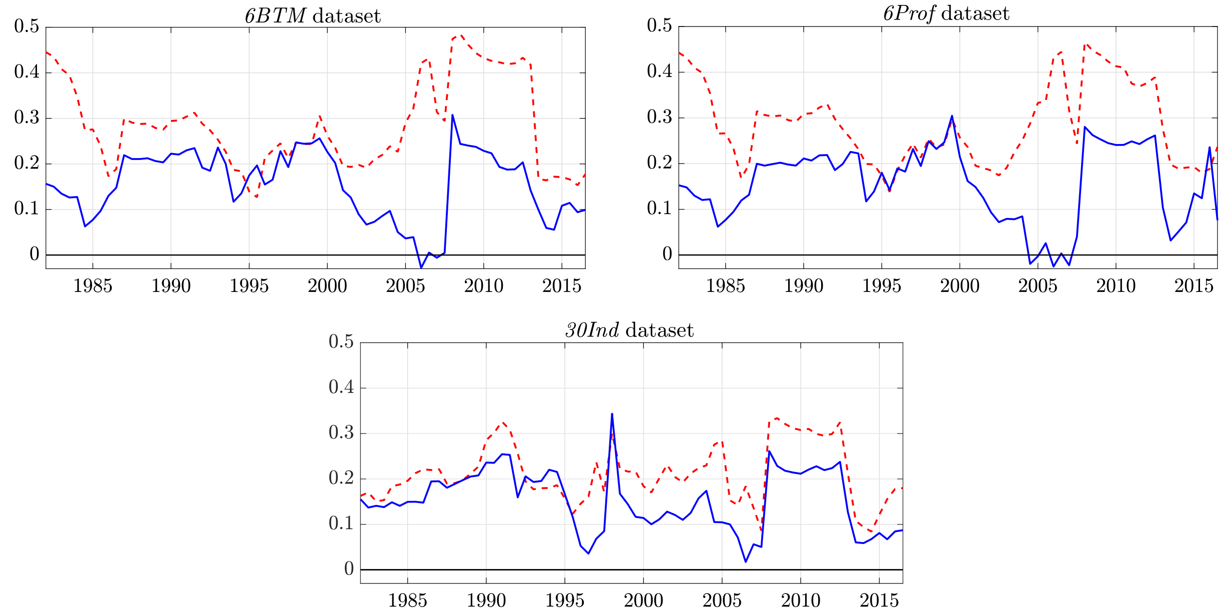
EC.6. Robustness tests

In this section, we report the numerical results and further discussion related to the robustness tests mentioned in Section 6.4 in the paper.

EC.6.1. High-dimensional dataset of individual stocks

In this section, we assess the out-of-sample performance of the portfolios for a high-dimensional dataset of 100 individual stocks. To build the dataset, we follow the methodology in Jagannathan and Ma (2003), DeMiguel et al. (2009) and Olivares-Nadal and DeMiguel (2018), which is designed to avoid survivorship bias. We download individual-stock returns from the Center for Research in Security Prices (CRSP) covering the same period from 1978 to 2017 as for the three other datasets. Starting from January 1983, and then every six months, we identify the subset of stocks that have an history of returns for the five preceding years as well as the next six months. As in the aforementioned papers, we remove microcap and penny stocks, which are the stocks that, in

Figure EC.4 Non-linear correlation of independent versus principal components.



Note. The figure depicts the average pairwise non-linear correlation across the K PCs (dotted red) and ICs (solid blue). The non-linear correlations are computed every six months using an estimation window containing the previous five years of daily asset returns. The x-axis labels indicate the last year of the window; for instance, the value at the date 2005 is computed from data from 2001 to 2005.

the preceding five years, were below the 20th percentile of average market capitalization and had an average price below 5 dollars. Finally, out of all remaining stocks, we randomly select $N = 100$ stocks to form our investment universe for the next six months. We repeat this procedure every six months until the end of the sample is achieved.

Table EC.1 reports the out-of-sample performance of the various portfolios for the individual-stocks dataset *100CRSP*. A few comments are in order. First, the portfolios based on variance tend to outperform those based on MVaR for the high-dimensional *100CRSP* dataset. This is consistent with the results for the three original datasets (*6BTM*, *6Prof*, *30Ind*), for which ICMVaR was the best strategy for $N = 6$ but ICMV did better for $N = 30$. The explanation for this result is that the portfolios based on higher-moment measures of risk are more sensitive to estimation risk, which is more important for high-dimensional asset universes. Second, comparing the four portfolios based on variance as risk measure (EWAVP, EWMV, ICMV, PCMV), we observe that our proposed ICMV portfolio outperforms the others: it has the best Sharpe ratio, kurtosis, MVaR and modified Sharpe ratio, albeit with a larger turnover than EWMV. The improvement in Sharpe ratio and modified Sharpe ratio compared to EWMV is statistically significant. Third, while for the three original datasets ICMV outperformed the two mean-variance-kurtosis portfolios MVKEW and MVKIC, for the *100CRSP* dataset their performance is similar. Overall, we note that the

Table EC.1 Out-of-sample performance for the *100CRSP* dataset.

100CRSP dataset (average $K = 18.37$)

	EWAVP	EWMV	ICMV	PCMV	EWMVaR	ICMVaR	MVKEW	MVKIC
Mean	16.42%	17.37%	16.29%	18.54%	**18.90%**	16.77%	16.55%	16.01%
Volatility	20.56%	13.55%	10.78%	13.55%	15.67%	13.54%	10.82%	**10.48%**
Sharpe ratio	0.80	1.28	$1.51^{\star}$	1.37	1.21	1.24	**1.53**	1.53
Skewness	-0.31	0.04	-0.11	0.04	0.01	**0.47**	-0.03	-0.16
Excess kurtosis	7.78	13.95	**7.73**	8.31	9.77	12.29	8.65	8.37
1% modified Value-at-Risk	5.55%	4.68%	**2.79%**	3.55%	4.47%	4.00%	2.91%	2.84%
Modified Sharpe ratio ($\times 10^2$)	1.17	1.47	$\mathbf{2.31}^{\star\star\star}$	$2.07^{\star\star}$	1.68	1.66	2.25	2.24
Turnover	7.89%	3.50%	4.74%	4.66%	**3.32%**	6.11%	4.64%	4.50%
Average shrinkage intensity δ	0.75	0.26	0.18	0.18	0.57	0.02	/	/

This table reports the out-of-sample performance of the eight portfolio policies for the *100CRSP* dataset following the methodology in Section 6.1 of the paper and Section EC.6.1 of this e-companion. The portfolio mean return, volatility and Sharpe ratio are annualized, while all other performance criteria are 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. 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 10%, 5% and 1%. We compare the following pairs of portfolios: PCMV vs EWMV (stars in superscript next to PCMV quantity), ICMV vs EWMV (stars in superscript next to ICMV quantity), ICMV vs PCMV (stars in subscript next to ICMV quantity) and ICMVaR vs EWMVaR (stars in superscript next to ICMVaR quantity).

empirical finding that our proposed portfolios outperform the benchmark portfolios is robust to considering a large dataset with the returns of 100 individual stocks.

EC.6.2. Worst-case-robust ICMV portfolio

In this section, we test the robustness of our results to parameter uncertainty in the covariance matrix of asset returns. To do this, we evaluate the performance of a worst-case-robust ICMV portfolio that is obtained by combining the minimum-variance IC-variance-parity portfolio with a worst-case-robust minimum-variance portfolio based on the popular approach by Goldfarb and Iyengar (2003), who consider parameter uncertainty in the covariance matrix assuming a factor model for the asset returns.

To do this, we first rewrite the asset returns as a linear combination of the N principal components,

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{\Lambda}_N^{1/2}\boldsymbol{Y}^{\star}, \tag{EC.20}$$

where $\boldsymbol{Y}^{\star}$ is the vector of the N standardized PCs, $\boldsymbol{\Lambda}_N$ is the $N \times N$ covariance matrix of the PCs and $\boldsymbol{A}$ is a $N \times N$ loading matrix. Following Goldfarb and Iyengar (2003), we assume that $\boldsymbol{\Lambda}_N$ is known, but we account for model uncertainty in $\boldsymbol{A}$ via the elliptical uncertainty set

$$\mathcal{A} := \left\{ \boldsymbol{A} \in \mathbb{R}^{N\times N} \;\middle|\; \boldsymbol{A} = \boldsymbol{A}_0 + \boldsymbol{\Delta}, \quad ||\boldsymbol{\Delta}_i||_{\boldsymbol{G}} \leq \rho_i, \quad i = 1, \ldots, N \right\}, \tag{EC.21}$$

where $\boldsymbol{\Delta}_i$ is ith row of $\boldsymbol{\Delta}$ and $||\boldsymbol{x}||_{\boldsymbol{G}} := \sqrt{\boldsymbol{x}'\boldsymbol{G}\boldsymbol{x}}$ is the $\boldsymbol{G}$-norm of $\boldsymbol{x}$ with $\boldsymbol{G}$ a symmetric positive-definite $N \times N$ matrix. To select $\boldsymbol{A}_0$, $\boldsymbol{G}$ and ρ_i, we follow Section 5 of Goldfarb and Iyengar (2003) and set $\boldsymbol{A}_0 = \boldsymbol{V}_N$, the sample eigenvector matrix, and calibrate $\boldsymbol{G}$ and the ρ_i's to the data so as to form a 95% confidence interval around $\boldsymbol{A}_0$.

Given this uncertainty set, the worst-case-robust minimum-variance portfolio is the solution to the maxmin problem

$$\min_{\boldsymbol{w}\in\mathcal{W}} \max_{\boldsymbol{A}\in\mathcal{A}} \boldsymbol{w}'\boldsymbol{A}\boldsymbol{\Lambda}_N\boldsymbol{A}'\boldsymbol{w}. \tag{EC.22}$$

As in Goldfarb and Iyengar (2003), we recast this maxmin problem as a second-order-cone program (SOCP)

$$\begin{aligned}
\min_{\boldsymbol{w}\in\mathcal{W},\,\boldsymbol{\psi}\in\mathbb{R}^N,\,(\nu,\tau,\sigma)\geq 0,\,\boldsymbol{t}\in\mathbb{R}^N_+} \quad & v \\
\text{subject to} \quad & \psi_i \geq w_i,\ \forall i, \\
& \psi_i \geq -w_i,\ \forall i, \\
& v \geq \tau + \mathbf{1}_N'\boldsymbol{t}, \\
& \sigma \leq 1/\lambda_{\max}(\boldsymbol{H}), \\
& 4(\boldsymbol{\rho}'\boldsymbol{\psi})^2 + (\sigma-\tau)^2 \leq (\sigma+\tau)^2, \\
& 4\phi_i^2 + (1-\sigma\lambda_i(\boldsymbol{H}) - t_i)^2 \leq (1-\sigma\lambda_i(\boldsymbol{H}) + t_i)^2, \quad \forall i,
\end{aligned} \tag{EC.23}$$

where $\boldsymbol{\rho}$ is the vector of ρ_i's defined in (EC.21), $\boldsymbol{H} := \boldsymbol{G}^{-1/2}\boldsymbol{\Lambda}_N\boldsymbol{G}^{-1/2}$, $\lambda_i(\boldsymbol{H})$ is the ith eigenvalue of $\boldsymbol{H}$ and $\boldsymbol{\phi} := \boldsymbol{Q}'\boldsymbol{H}^{1/2}\boldsymbol{G}^{1/2}\boldsymbol{A}_0\boldsymbol{w}$ with $\boldsymbol{Q}$ the eigenvector matrix of $\boldsymbol{H}$. The last two constraints in (EC.23) are cone constraints and, because problem (EC.23) is a SOCP, we solve it via Matlab *coneprog* algorithm. Once the worst-case-robust minimum-variance portfolio is found, we then shrink it toward our proposed minimum-variance IC-variance-parity portfolio with the shrinkage intensity calibrated via 10-fold cross-validation as in the manuscript. The final portfolio is the worst-case-robust ICMV portfolio.

In Table EC.2, we report the out-of-sample performance of the ICMV portfolio using (i) sample non-robust estimation as in Table 1 in the manuscript and (ii) worst-case-robust estimation as described above. We report the same performance criteria as in Table 1 in the manuscript, and in addition we also report the excess kurtosis in the period from December 2007 to June 2009, which is the period corresponding to the global financial crisis according to the "NBER based Recession Indicators for the United States from the Peak through the Trough". We look at this indicator because, by construction, worst-case-robust portfolios are expected to have less tail risk in crisis periods.

The table shows that the standard (non-robust) ICMV portfolio generally outperforms the worst-case-robust ICMV portfolio. For instance, the non-robust ICMV portfolio is best in terms of Sharpe ratio for all three datasets and also in terms of modified Sharpe ratio for the *6BTM* and *30Ind* datasets. Only for the *6Prof* dataset does the worst-case-robust ICMV portfolio outperform

Table EC.2 Out-of-sample performance of worst-case-robust versus non-robust portfolios.

	Worst-case-robust			Non-robust		
	6BTM	*6Prof*	*30Ind*	*6BTM*	*6Prof*	*30Ind*
Mean	21.73%	19.99%	11.98%	21.39%	20.41%	15.14%
Volatility	13.91%	15.39%	16.07%	13.41%	14.83%	14.60%
Sharpe ratio	1.56	1.30	0.75	1.59	1.38	1.04
Skewness	-0.46	-0.39	-0.72	-0.33	-0.35	-0.49
Excess kurtosis	11.58	11.57	14.45	11.53	14.86	12.95
1% modified VaR	4.55%	5.02%	6.06%	4.33%	5.53%	5.11%
Modified Sharpe ratio ($\times 10^2$)	1.89	1.58	0.78	1.96	1.46	1.18
Turnover	2.75%	4.11%	1.80%	2.79%	2.93%	2.52%
Average shrinkage intensity δ	0.92	0.91	0.55	0.63	0.71	0.42
Excess kurtosis 07-09	3.39	2.60	1.99	3.67	4.78	4.59

This table reports the out-of-sample performance of the ICMV portfolio estimated via worst-case-robust and sample non-robust estimation following the methodology in Section EC.6.2 of this e-companion. The 07-09 excess kurtosis is computed from December 2007 to June 2009, which corresponds to the global financial crisis according to the "NBER based Recession Indicators for the United States from the Peak through the Trough". The portfolio mean return, volatility and Sharpe ratio are annualized, while all other performance criteria are 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).

the non-robust portfolio and only in terms of modified Sharpe ratio and tail risk. We also find that, as expected, the worst-case-robust ICMV portfolio has less excess kurtosis during the financial crisis. Overall, the out-of-sample performance of the non-robust ICMV portfolio is at least as good as that of the worst-case-robust ICMV portfolio. The explanation for the good performance of the non-robust ICMV portfolio is twofold. First, the ICMV portfolio is already robust because it is based on risk parity and it ignores estimates of mean returns. Thus, adding one more layer of robustness might not be beneficial. Second, we use daily data, which produce more accurate estimators than weekly or monthly data. This implies that the non-robust ICMV portfolio is estimated quite accurately.

EC.6.3. Long-only factor-variance-parity portfolios

In this section, we report the out-of-sample performance of factor-variance-parity (FVP) portfolios, based on PCs and ICs, 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 reduces sensitivity to 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 first step 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 minimum-variance 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 EC.3, 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 portfolio that shrinks the long-only minimum-variance (MV) portfolio toward the equally weighted portfolio (EWMV). 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 for two datasets and modified Sharpe ratio for all datasets. Moreover, compared to the EWMV portfolio, it systematically outperforms both in terms of Sharpe ratio and modified Sharpe ratio. 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) show 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 much less turnover.

EC.6.4. Subsample analysis: low versus high mutual-information reduction

Figure 2 in the main body of the paper and Figure EC.4 above show that the reduction in mutual information and non-linear correlations from using ICs instead of PCs varies over time. To study

Table EC.3 Out-of-sample performance of long-only MV, ICMV and PCMV portfolios.

	6BTM dataset			*6Prof* dataset			*30Ind* dataset		
	EWMV	ICMV	PCMV	EWMV	ICMV	PCMV	EWMV	ICMV	PCMV
Mean	14.37%	**14.99%**	14.78%	13.19%	**14.49%**	13.90%	11.13%	**15.18%**	10.77%
Volatility	15.95%	**15.62%**	16.07%	15.51%	16.91%	**15.24%**	14.21%	16.91%	**13.66%**
Sharpe ratio	0.90	**0.96**	0.92	0.85	0.86	**0.91**	0.78	**0.90**	0.79
Skewness	-0.58	-0.72	**-0.42**	-0.59	-0.54	**-0.45**	-0.87	**-0.20**	-0.61
Excess kurtosis	13.63	**13.00**	13.66	12.61	**10.41**	15.06	24.59	**16.75**	18.81
1% modified Value-at-Risk	5.78%	**5.55%**	5.77%	5.40%	**5.32%**	5.80%	7.50%	6.73%	**6.01%**
Modified Sharpe ratio ($\times 10^2$)	0.99	**1.07**	1.02	0.97	**1.08**	0.95	0.59	**0.90**	0.71
Turnover	**0.36%**	0.44%	0.44%	0.30%	0.50%	**0.22%**	**0.81%**	0.95%	0.82%
Average shrinkage intensity δ	0.21	0.36	0.14	0.18	0.52	0.10	0.26	0.37	0.40

This table reports the out-of-sample performance of the variance-based portfolios under the no-short-selling constraint, following the methodology in Section EC.6.3 of this e-companion. The portfolio mean return, volatility and Sharpe ratio are annualized, while all other performance criteria are 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 and dataset.

the impact of this variation on the relative performance of IC-based versus PC-based portfolios, we now evaluate the out-of-sample performance of the different portfolios on low versus high mutual-information subsamples.

Note that the out-of-sample period in our analysis covers the period from January 1983 to December 2017. Based on Figure 2 in the main body of the paper and Figure EC.4 above, we evaluate the out-of-sample performance in the following two subsamples: (i) January 1983 to December 1999 (low reduction in mutual information/non-linear correlation) and (ii) January 2000 to December 2012 (high reduction). We also report the performance of the EWMV portfolio for comparison. The hypothesis underlying the subsample analysis is that we expect the ICMV portfolio to be particularly useful compared to PCMV when the reduction in mutual information is high.

The results are reported in Table EC.4. We make two main observations. First, in terms of modified Sharpe ratio, which accounts for the first four moments of portfolio returns, using ICMV rather than PCMV is particularly appealing during periods of high mutual-information reduction. Indeed, in the 1983-1999 period (low reduction), ICMV is outperformed by PCMV for *6BTM*, performs similarly for *6Prof*, and largely outperforms for *30Ind*. The two strategies are therefore overall on par. In contrast, during the 2000-2012 period (high reduction), ICMV performs much better than PCMV for all three datasets. This finding is intuitive: ICA-based portfolios are likely to outperform PCA-based portfolios when the gain in independence from using ICA is substantial, which in the data corresponds to the period before and around the financial crisis. Second, ICMV nearly systematically outperforms the benchmark EWMV in both subsamples in terms of Sharpe ratio and modified Sharpe ratio. Overall, looking at the two subsamples separately, we conclude that relying on ICMV is always a good strategy because it always performs at least on par with PCMV, and outperforms PCMV for periods of high mutual-information reduction.

Table EC.4 Subsample analysis: low versus high mutual-information reduction.

6BTM dataset	1983-1999			2000-2012		
	EWMV	ICMV	PCMV	EWMV	ICMV	PCMV
Mean	21.96%	**23.66%**	23.36%	17.65%	20.05%	**20.44%**
Volatility	**8.54%**	9.10%	8.73%	18.44%	17.92%	**14.06%**
Sharpe ratio	2.57	2.60	**2.68**	0.96	1.12	**1.45**
Skewness	-1.95	-1.46	**-1.39**	-0.38	**-0.09**	-0.23
Excess kurtosis	30.18	24.89	**20.15**	10.50	**6.48**	19.57
1% modified Value-at-Risk	4.96%	4.73%	**3.94%**	5.74%	**4.33%**	6.16%
Modified Sharpe ratio ($\times 10^2$)	1.76	1.98	**2.35**	1.22	**1.84**	1.32
Turnover	**1.86%**	2.47%	2.50%	**2.78%**	3.19%	3.99%
6Prof dataset	1983-1999			2000-2012		
	EWMV	ICMV	PCMV	EWMV	ICMV	PCMV
Mean	18.20%	**22.12%**	17.63%	14.91%	**19.66%**	17.22%
Volatility	**10.19%**	11.45%	10.96%	18.86%	19.07%	**15.07%**
Sharpe ratio	1.79	**1.93**	1.61	0.79	1.03	**1.14**
Skewness	-1.25	-1.06	**-0.97**	-0.14	-0.11	**-0.02**
Excess kurtosis	18.65	20.19	**14.76**	**10.30**	10.36	20.44
1% modified Value-at-Risk	4.43%	5.25%	**4.16%**	**5.68%**	5.72%	6.69%
Modified Sharpe ratio ($\times 10^2$)	1.63	1.67	**1.68**	1.04	**1.36**	1.02
Turnover	**2.28%**	2.37%	2.96%	**2.44%**	3.77%	3.30%
30Ind dataset	1983-1999			2000-2012		
	EWMV	ICMV	PCMV	EWMV	ICMV	PCMV
Mean	16.53%	**17.13%**	9.97%	7.22%	**14.89%**	9.56%
Volatility	11.92%	14.04%	**11.81%**	18.07%	16.05%	**13.22%**
Sharpe ratio	**1.39**	1.22	0.84	**0.40**	0.93	0.72
Skewness	-2.28	**-1.05**	-1.68	-0.25	**-0.03**	-0.62
Excess kurtosis	46.31	**23.62**	36.82	12.81	**5.23**	18.74
1% modified Value-at-Risk	9.60%	**7.19%**	8.22%	6.21%	**3.55%**	5.81%
Modified Sharpe ratio ($\times 10^2$)	0.68	**0.95**	0.48	0.46	**1.67**	0.65
Turnover	**1.58%**	2.14%	2.29%	**2.37%**	3.06%	2.84%

This table reports the out-of-sample performance of the variance-based portfolios in the low (1983-1999) and high (2000-2012) mutual-information reduction subsamples following the methodology in Section EC.6.3 of this e-companion. The portfolio mean return, volatility and Sharpe ratio are annualized, while all other performance criteria are 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, dataset and subsample.

EC.6.5. ICMVaR portfolio without assuming the independence of ICs

In this section, we compare the performance of the ICMVaR portfolio computed with and without the assumption of independence of the ICs. For the sake of comparison, we keep the same shrinkage intensities in the case without the independence assumption as those originally computed in the paper for the case with the independence assumption. The computation of the IC-MVaR-parity portfolio without the independence assumption is similar to what is described in Section 5 in the paper except that, in Proposition 6 where we derive the MVaR contribution the ICs, $m_3(\tilde{P})$ and

Table EC.5 Out-of-sample performance of ICMVaR portfolio with and without the assumption that ICs are independent.

	With independence			Without independence		
	6BTM	*6Prof*	*30Ind*	*6BTM*	*6Prof*	*30Ind*
Mean	23.93%	22.55%	11.70%	23.81%	22.19%	10.65%
Volatility	15.14%	15.60%	14.14%	15.76%	16.34%	14.41%
Sharpe ratio	1.58	1.45	0.83	1.51	1.36	0.74
Skewness	-0.16	-0.25	-0.78	-0.14	-0.30	-0.77
Excess kurtosis	7.22	10.99	17.46	7.26	11.39	17.60
1% modified Value-at-Risk	3.84%	4.88%	5.97%	4.00%	5.24%	6.11%
Modified Sharpe ratio ($\times 10^2$)	2.47	1.83	0.78	2.36	1.68	0.69
Turnover	5.75%	5.00%	3.99%	6.23%	5.33%	5.15%

This table reports the out-of-sample performance of the ICMVaR portfolio with and without the assumption that ICs are independent following the methodology in Section EC.6.5 of this e-companion. The portfolio mean return, volatility and Sharpe ratio are annualized, while all other performance criteria are 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).

$m_4(\tilde{P})$ now admit the more general formulas in equations (21) and (22), which depend on the full coskewness and cokurtosis matrices of the ICs because we do not assume independence.

The out-of-sample performance is reported in Table EC.5. The table shows that ignoring the higher-moment dependence of ICs helps to improve the out-of-sample performance of the ICMVaR portfolio. In addition, the turnover of the portfolios estimated considering the IC comoments are higher, probably due to the estimation error associated with IC comoments, and thus the relative performance of these portfolios will be worse in practice.

EC.6.6. Value-at-Risk versus modified Value-at-Risk

In the main body of the paper, we compare the performance of the different portfolios in terms of various criteria including modified VaR (MVaR) and modified Sharpe ratio. In this section, we study the robustness of our findings to evaluating the portfolio performance using VaR instead of MVaR, which we estimate via the empirical 1% quantile of portfolio returns.

Table EC.6 reports the out-of-sample MVaR and VaR of the different portfolios. Our proposed portfolios are the best in terms of MVaR for all three datasets (ICMVaR for *6BTM* and *6Prof* and ICMV for *30Ind*), but the best portfolio in terms of VaR varies across the three datasets (PCMV for *6BTM*, EWAVP for *6Prof* and MVKEW for *30Ind*). More importantly, investors are ultimately concerned about risk-adjusted returns, and thus we also report in Table EC.6 the modified Sharpe ratio based on MVaR and VaR, defined as the ratio between the portfolio-return mean and either MVaR or VaR. Our main finding is that our proposed portfolios are the best in terms of modified Sharpe ratio across the three datasets regardless of whether MVaR or VaR is used to evaluate the modified Sharpe ratio: ICMVaR is the best for *6BTM* and *6Prof* and ICMV for *30Ind*. These results

Table EC.6 Comparison of VaR and modified VaR performance criteria.

	EWAVP	EWMV	ICMV	PCMV	EWMVaR	ICMVaR	MVKEW	MVKIC
6BTM dataset								
VaR	2.38%	2.35%	2.36%	**2.27%**	2.65%	2.59%	2.29%	2.29%
MVaR	5.63%	5.65%	4.33%	5.75%	4.27%	**3.84%**	6.16%	6.06%
MSR VaR	3.29	3.31	3.60	3.64	3.48	**3.67**	3.45	3.54
MSR MVaR	1.39	1.38	1.96	1.44	2.16	**2.47**	1.28	1.34
6Prof dataset								
	EWAVP	EWMV	ICMV	PCMV	EWMVaR	ICMVaR	MVKEW	MVKIC
VaR	**2.44%**	2.49%	2.49%	2.48%	2.65%	2.56%	2.52%	2.49%
MVaR	5.72%	5.44%	5.53%	6.32%	5.47%	**4.88%**	6.26%	5.99%
MSR VaR	2.73	2.75	3.27	2.85	3.27	**3.54**	2.73	3.04
MSR MVaR	1.20	1.24	1.46	1.12	1.58	**1.83**	1.09	1.27
30Ind dataset								
	EWAVP	EWMV	ICMV	PCMV	EWMVaR	ICMVaR	MVKEW	MVKIC
VaR	2.37%	2.51%	2.54%	2.28%	2.47%	2.43%	**2.04%**	2.11%
MVaR	7.16%	7.27%	**5.11%**	5.32%	7.48%	5.97%	6.02%	5.62%
MSR VaR	2.14	1.99	**2.37**	1.84	2.11	1.91	2.33	2.25
MSR MVaR	0.71	0.69	**1.18**	0.79	0.70	0.78	0.79	0.85

This table reports, for all portfolio policies, the out-of-sample VaR, modified VaR (MVaR) and modified Sharpe ratio (MSR) based on either VaR and MVaR for a confidence level $\alpha = 1\%$ following the methodology in Section EC.6.6 of this e-companion. The criteria are in daily terms, and the MSR is multiplied by 100. 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 and dataset.

confirm that using MVaR for portfolio selection helps to achieve good out-of-sample risk-adjusted returns whether these are measured in terms of MVaR or VaR. Table EC.6 also shows that MVaR tends to overestimate VaR, but this is not a concern because the table also shows that using MVaR leads to estimated portfolios that have good out-of-sample risk-adjusted returns.

EC.6.7. Modified VaR and modified Sharpe ratio for more challenging α's

In this section, we test whether the ranking of the different portfolios in terms of MVaR and modified Sharpe ratio for $\alpha = 1\%$ in Table 1 in the paper is robust to using a more challenging α of 0.5% and 0.1%.

The results are reported in Table EC.7. The main conclusion is that considering more challenging α's does not change the relative ranking of the portfolio strategies compared to the case $\alpha = 1\%$. In particular, in terms of MVaR and modified Sharpe ratio, ICMVaR remains the best strategy for the *6BTM* and *6Prof* datasets, and ICMV for the *30Ind* dataset.

References

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

Table EC.7 Out-of-sample modified VaR and modified Sharpe ratio for lower confidence levels α.

	6BTM dataset				*6Prof* dataset				*30Ind* dataset			
	$\alpha = 0.5\%$		$\alpha = 0.1\%$		$\alpha = 0.5\%$		$\alpha = 0.1\%$		$\alpha = 0.5\%$		$\alpha = 0.1\%$	
	MVaR	MSR	MVaR	MSR	MVaR	MSR	MVaR	MSR	MVaR	MSR	MVaR	MSR
EWAVP	8.17%	0.96	15.32%	0.51	8.36%	0.82	15.79%	0.43	10.60%	0.48	20.30%	0.25
EWMV	8.23%	0.95	15.48%	0.50	7.86%	0.85	14.65%	0.46	10.78%	0.46	20.69%	0.24
ICMV	6.10%	1.39	11.03%	0.77	7.98%	1.02	14.84%	0.55	**7.25%**	**0.83**	**13.20%**	**0.46**
PCMV	8.47%	0.98	16.18%	0.51	9.40%	0.75	18.13%	0.39	7.72%	0.54	14.46%	0.29
EWMVaR	5.86%	1.57	10.27%	0.90	7.86%	1.10	14.57%	0.59	11.09%	0.47	21.26%	0.25
ICMVaR	**5.18%**	**1.83**	**8.86%**	**1.07**	**6.86%**	**1.30**	**12.34%**	**0.72**	8.65%	0.54	16.15%	0.29
MVKEW	9.10%	0.87	17.41%	0.45	9.28%	0.74	17.84%	0.38	8.95%	0.53	17.22%	0.28
MVKIC	8.94%	0.91	17.07%	0.47	8.80%	0.87	16.72%	0.46	8.25%	0.58	15.67%	0.30

This table reports the out-of-sample modified VaR (MVaR) and modified Sharpe ratio (MSR) of the eight portfolio policies for a confidence level α of 0.5% and 0.1%; see Section EC.6.7. The MVaR and MSR are in daily terms, and the MSR is multiplied by 100. 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, dataset and choice of α.

Bekaert G, Erb C, Harvey C, Viskanta T (1998) Distributional characteristics of emerging market returns and asset allocation. *J. Portfolio Management* 24(2):102–116.

Boudt K, Cornilly D, Verdonck T (2020) A coskewness shrinkage approach for estimating the skewness of linear combinations of random variables. *J. Financ. Econom.* 18(1):1–23.

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.

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

Cont R (2001) Empirical properties of asset returns: Stylized facts and statistical issues. *Quant. Finance* 1(2):223–236.

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.

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

Gao G, Lu X, Song Z (2019) Tail risk concerns everywhere. *Management Sci.* 65(7):3111–3130.

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

Gormsen N, Jensen C (2020) Higher-moment risk. Available at SSRN: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3069617`.

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

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.

Kumiega A, Neururer T, Van Vliet B (2011). Independent component analysis for realized volatility: Analysis of the stock market crash of 2008. *Q. Rev. Econ. Finance* 51(3):292–302.

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

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.

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

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(3):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).

Vrins F, Pham D-T, Verleysen M (2007) Mixing and non-mixing local minima of the entropy contrast for blind source separation. *IEEE T. Inform. Theory* 53(3):1030–1042.