Finance Stoch (2010) 14: 375–395
DOI 10.1007/s00780-009-0107-y

On measuring nonlinear risk with scarce observations

**Alexander Cherny · Raphael Douady ·
Stanislav Molchanov**

Received: 25 March 2008 / Accepted: 15 January 2009 / Published online: 7 November 2009

Abstract We propose a methodology for estimating the risk of portfolios that exhibit nonlinear dependence on the risk driving factors and have scarce observations, which is typical for portfolios of investments in hedge funds. The methodology consists of two steps: first, regressing the portfolio return on nonlinear functions of each single risk driving factor and second, merging together the obtained estimates taking into account the dependence between different factors. Performing the second step leads us to a certain probabilistic problem, for which we propose an analytic and computationally feasible solution for the case where the joint law of the factors is a Gaussian copula. A typical practical application can be to estimate the risk of a hedge fund or a portfolio of hedge funds. As a theoretical consequence of our results, we propose a new definition of the factor risk, i.e., the risk of a portfolio brought by a given factor.

Keywords Factor risk · Gaussian copula · Hedge fund replication · Hedge fund risk · Nonlinear risk

Mathematics Subject Classification (2000) 60E99 · 91B28 · 91B30

JEL Classification C35 · G29

A. Cherny (✉)
Department of Probability Theory, Faculty of Mechanics and Mathematics, Moscow State University, 119992 Moscow, Russia
e-mail: alexander.cherny@gmail.com

R. Douady
Head of Research Department, RiskData, 10 East 40th Street, 33rd Floor, New York, NY 10016, USA
e-mail: raphael.douady@riskdata.com

S. Molchanov
Department of Mathematics and Statistics, University of North Carolina at Charlotte, 376 Fretwell Bldg., 9201 University City Blvd., Charlotte, NC 28223-0001, USA
e-mail: smolchan@uncc.edu

1 Introduction

1.1 Motivation

In constructing portfolios of hedge funds, which is the fund of funds problem, one of the central tasks is to estimate the risk of a hedge fund or basket of hedge funds. In order to estimate the risk of a hedge fund, one could simply estimate some risk measure (for example, standard deviation or Value at Risk) using its return history. A deeper approach, however, would be to relate the hedge fund return to risk driving factors like equity, bond, currency, commodity, and volatility indices, credit spreads, etc. This provides the advantages of:

– Taking into account predictions about the future moves of factors (for example, a prediction that oil would exhibit high volatility in the next month).
– Using the long and frequent time series available for factors, as opposed to short-time and infrequent (monthly) time series available for hedge funds.
– Estimating the joint risk of portfolios of hedge funds.

The most classical approach to relating hedge fund returns to risk factors is the linear regression on a collection of factors. This is closely connected with the linear hedge funds replication (see Fung and Hsieh [3], Hasanhodzic and Lo [7]). However, the findings of Hasanhodzic and Lo [7] show that such a method explains only 15–20% of the variance (i.e., risk) in hedge fund returns. One of the reasons is the well-recognized nonlinearity in the dependence of hedge fund returns on the factors (see Agarwal and Naik [1], Fung and Hsieh [4, 5], Lo [8]).

In order to capture this nonlinearity, one might take call and put options on factors as additional factors, as is done in Agarwal and Naik [1]. However, the following problem then arises. Hedge fund returns typically have quite a short observation history and are observed only on a monthly basis; thus, if a fund has a 2-year history, then we have only two dozens of observations of its return. On the other hand, the number of nonlinear factors might well exceed the number of observations, in which case it is impossible to do the regression of the hedge fund return on all the nonlinear factors due to overfitting. One way to overcome this problem is to extract from factors only the most important ones and thus avoid overfitting. For example, Agarwal and Naik [1] choose as the factors one index and four options on the index.

We propose another way to regress the return on a collection of nonlinear factors, which allows us to leave all the linear factors and a reasonable number of options on each of them. Let R denote the hedge fund return over a fixed time period (typically, the monthly return) and $F_1, \dots, F_N$ be the returns/increments of factors over the same time period. The methodology we propose consists of two steps:

1. Regress the return on the nonlinear functions of each *single* factor, i.e., for each $n = 1, \dots, N$, find the best approximation of R by a sum of the form

$$\varphi_n(F_n) = a_n + b_n F_n + \sum_{i=1}^{I_n} c_{ni} f_{ni}(F_n),$$

where f_{ni} is some set of nonlinear functions (for example, $f_{ni}(x) = (x - K_{ni})^+$).

2. Join together the functions $\varphi_1(F_1), \dots, \varphi_N(F_N)$, taking into account the dependence between the factors, to produce a nonlinear approximation of R by a function of all factors.

The methodology for Step 1 is clear: we perform simply a linear regression of R on a factor and a family of nonlinear functions of the factor. The methodology for Step 2 is not obvious since different factors are dependent. Performing Step 2 is the main objective of this paper.

1.2 Formulation

We formalize the above problem in the following way. We assume that we know the conditional expectations

$$\varphi_n(x) = \mathsf{E}[R \mid F_n = x], \quad n = 1, \dots, N$$

and want to recover the conditional expectation

$$\varphi(x_1, \dots, x_N) = \mathsf{E}[R \mid F_1 = x_1, \dots, F_N = x_N].$$

To be more precise, the inputs we have are the joint law P of $X_1, \dots, X_N$[1] and functions $\varphi_1, \dots, \varphi_N$ from $\mathbb{R}$ to $\mathbb{R}$. We want to find a function $\varphi : \mathbb{R}^N \to \mathbb{R}$ such that

$$\mathsf{E}\big[\varphi(X_1, \dots, X_N) \mid X_n\big] = \varphi_n(X_n), \quad n = 1, \dots, N, \tag{1.1}$$

where X_n denotes the nth coordinate projection of $\mathbb{R}^N$ on $\mathbb{R}$ and expectations are taken with respect to P. This is the main problem we consider in the paper.

A reformulation of (1.1) is to recover a function $\varphi : \mathbb{R}^N \to \mathbb{R}$ if one knows its integrals over all the N hyperplanes orthogonal to the coordinate axes. This problem is somehow similar to recovering a function from its Radon transform (which means the knowledge of the integrals over all the one-dimensional lines in $\mathbb{R}^N$). However, there exists a crucial difference between the two problems: knowing the Radon transform is sufficient to recover the function, while knowing the integrals over hyperplanes orthogonal to coordinate axes is far from being sufficient. As an example, consider the situation where P is concentrated on a lattice $\{a_1, \dots, a_M\}^N$. Then we have M^N unknown parameters, while (1.1) provides us only with NM equations. In another example, if X_1, X_2 are independent standard Gaussian and $\varphi(x_1, x_2) = x_1 x_2$, then $\mathsf{E}[\varphi(X_1, X_2) \mid X_n] = 0$, so that adding φ to any solution of (1.1) provides another solution.

Thus, there is no hope to get a unique solution of (1.1) in any reasonable model. On the other hand, for practical applications, one definitely needs to choose a unique solution. So one has to impose additional conditions on the unknown function φ that guarantee the uniqueness of a solution (at least, in reasonable models) and also allow

[1] One might wonder why we assume the knowledge of the joint law of $(X_1, \dots, X_N)$ and do not assume the knowledge of the joint law of $(R, X_1, \dots, X_N)$ (if the latter is known, the whole problem disappears). The reason is that factors have a long history of frequent observations, while hedge fund returns have a short history of monthly observations.

for an efficient procedure of calculating the solution (at least, in some models). In this paper, we propose to look for the solution which is "the most moderate one" as measured by its variance. In other words, we consider the problem

$$\begin{cases} \text{Minimize} & \mathsf{Var}\,\varphi(X_1,\dots,X_N) \\ \text{subject to} & \mathsf{E}[\varphi(X_1,\dots,X_N)\mid X_n]=\varphi_n(X_n), \quad n=1,\dots,N, \end{cases} \tag{1.2}$$

where Var denotes the variance.

1.3 Solution

We first consider (1.2) in complete generality, i.e., for an arbitrary measure P. A necessary condition for the existence of a solution is that all $\varphi_n(X_n)$ have the same expectation. We impose this condition and assume without loss of generality that this expectation is zero. The result we prove shows that in typical cases the solution exists, is unique, and has the form

$$\varphi(x_1,\dots,x_N)=\sum_{n=1}^{N}\psi_n(x_n)$$

with some functions $\psi_n:\mathbb{R}\to\mathbb{R}$. In order to get explicit expressions for ψ_n, we are then considering three particular cases (each corresponding to a quite wide class of measures P).

Our first example is the case when $X_1,\dots,X_N$ are independent. In this case the solution of (1.2) has the form

$$\varphi(x_1,\dots,x_N)=\sum_{n=1}^{N}\varphi_n(x_n).$$

However, this example is of theoretical interest only as in practice the factors are always dependent.

The second particular case corresponds to a Gaussian P. Then the solution of (1.2) is provided by

$$\varphi(x_1,\dots,x_N)=\sum_{n,m=1}^{N,\infty}\alpha_{nm}H_m(x_n),$$

where the H_m are Hermite polynomials and the α_{nm} are found through solving certain N-dimensional linear systems. In practical calculations, one cuts off the above summation in m at some $M\in\mathbb{N}$ and then has to solve M such linear systems.

Finally, we consider the case when P is a Gaussian copula. Mathematically, this is trivial because the problem is immediately reduced to the Gaussian case. But practically, this is very useful as Gaussian copulas are probably the most popular class of models for the joint distribution of risk factors.

1.4 Application

An important problem of modern risk measurement is to determine which part of the portfolio risk is brought by each of the risk driving factors.[2] Recent papers on the subject include the following. Cherny and Madan [2] consider the conditional expectation of the portfolio return/P&L with respect to the factor and call the risk of this new random variable the factor risk brought by that factor. Martin and Tasche [10] also consider the same conditional expectation, but then apply the Euler principle taking the derivative of the portfolio risk in the direction of this conditional expectation and call the result the risk impact. Rosen and Saunders [12] apply the Hoeffding decomposition of the portfolio return/P&L with respect to a set of factors; the first several terms of this decomposition coincide with the conditional expectations mentioned above. However, each of the above approaches has certain problems coming from the correlation between different factors. The factor risks or risk impacts brought by different factors do not necessarily sum up to the overall risk; in particular, this sum might be considerably smaller meaning that factor risks or risk impacts do not control the overall risk (see Example 5.3 below). As for the Hoeffding decomposition, it has 2^N terms, where N is the number of factors (N might be of order 100); one might try to avoid this huge number by considering the first N terms and aggregating together all the rest, but then the same problem as described above arises.

The methodology of this paper allows us to transform the conditional expectations of the return/P&L with respect to the factors to other nonlinear functions of the factors in a way that takes into account the correlation between the factors and results in the best L^2-approximation of the return/P&L by a sum of nonlinear functions of single factors. This approach has some similarities with the work of Rosen and Saunders [11], but the difference is that the authors of that paper consider the best linear approximation of the return/P&L by the factors, while we are considering a nonlinear one. As a result, we obtain a new definition of factor risks and provide a new decomposition of portfolio risk into the sum of $N+2$ parts (N is the number of factors): N factor risks brought by each of the factors, the risk coming from nonlinear cross-correlation between the factors, and the idiosyncratic risk.

1.5 Structure of the paper

In Sect. 2, problem (1.2) is studied for a general P. The three particular situations described above are considered in Sect. 3. Section 4 deals with the extension to multidimensional factors. In Sect. 5, we propose a new definition of factor risks and provide a new decomposition of the portfolio risk. Section 6 concludes.

[2]This is not to be confused with the problem of determining which part of the portfolio risk is brought by each of the subportfolios.

2 General setup

In this section, we study the problem (1.2) for an arbitrary measure P on $\mathbb{R}^N$. Let us set

$$\Phi = \big\{(\varphi_1, \dots, \varphi_N) : \mathsf{E}\varphi_n^2(X_n) < \infty \text{ and } \mathsf{E}\varphi_n(X_n) = 0\ \forall n = 1, \dots, N\big\}.$$

We denote by $\|\cdot\|$ the L^2-norm and by Pr_E the orthogonal projection in the $L^2(P)$-sense on a space E.

We begin with a useful lemma, which sheds light on the structure of solutions.

Lemma 2.1 *Let* $(\varphi_n) \in \Phi$. *Suppose that* $\psi_n : \mathbb{R} \to \mathbb{R}$ *are measurable functions with* $\mathsf{E}\psi_n^2(X_n) < \infty$ *such that the function*

$$\varphi(x_1, \dots, x_N) = \sum_{n=1}^{N} \psi_n(x_n) \tag{2.1}$$

satisfies (1.1). *Then* φ *is the unique solution of* (1.2).

Proof Let $\widetilde{\varphi}$ be a function satisfying (1.1). Denote

$$E_n = \big\{\xi \in L^2 : \xi \text{ is } X_n\text{-measurable}, \mathsf{E}\xi = 0\big\}, \tag{2.2}$$

where $L^2 = L^2(\mathbb{R}^N, P)$ is the space of square-integrable random variables. Then $\mathrm{Pr}_{E_n}\widetilde{\varphi} = \varphi_n(X_n) = \mathrm{Pr}_{E_n}\varphi$, where φ and $\widetilde{\varphi}$ are considered as random variables on $(\mathbb{R}^N, P)$. As $\varphi \in E_1 + \dots + E_N$ (this is the space of sums $\xi_1 + \dots + \xi_N$, where $\xi_n \in E_n$), we get that $\widetilde{\varphi} - \varphi$ is orthogonal to φ. This implies that $\|\widetilde{\varphi}\| \ge \|\varphi\|$, and the equality is possible only if $\widetilde{\varphi} = \varphi$ P-a.s. □

The next theorem provides a condition on P ensuring the existence of a solution for any $(\varphi_n) \in \Phi$. It also shows that in this case the solution enjoys a number of nice properties.

Theorem 2.2 *The following conditions are equivalent*:

(a) *For any* $(\varphi_n) \in \Phi$, *there exists a solution of* (1.1).
(b) (Lower ellipticity) *There exists a constant* $c > 0$ *such that, for any* $(\varphi_n) \in \Phi$,

$$\bigg\|\sum_{n=1}^{N} \varphi_n(X_n)\bigg\| \ge c \sum_{n=1}^{N} \big\|\varphi_n(X_n)\big\|.$$

If the above conditions are satisfied, then

(c) (Existence, uniqueness, and form of the solution) *For any* $(\varphi_n) \in \Phi$, *there exists a unique solution of* (1.2), *and it has the form* (2.1) *with some* $(\psi_n) \in \Phi$.
(d) (Linearity) *If* φ (*resp.* φ') *is a solution of* (1.2) *corresponding to* $(\varphi_n) \in \Phi$ (*resp.* $(\varphi'_n) \in \Phi$), *then the solution corresponding to* $(\alpha\varphi_n + \alpha'\varphi'_n)$ *is* $\alpha\varphi + \alpha'\varphi'$.

(e) (Continuity) *There exists a constant* $C > 0$ *such that, for any* $(\varphi_n) \in \Phi$,

$$\big\|\varphi(X_1, \dots, X_N)\big\| \le C \sum_{n=1}^{N} \big\|\varphi_n(X_n)\big\|,$$

where φ *is the solution of* (1.2) *corresponding to* (φ_n).

The financial meaning of the linearity is as follows. If R denotes the return of a portfolio, then the solution corresponding to a weighted average of several portfolios is the weighted average of solutions. If R denotes the Profit&Loss of a portfolio, then the solution corresponding to a sum of portfolios is the sum of solutions. This property is very convenient in constructing an optimal portfolio of hedge funds, i.e., for the fund of funds problem.

The financial meaning of the continuity is that the solution is stable under small misspecifications of the distributions $\mathrm{Law}(R, X_n)$ (for a fixed $\mathrm{Law}(X_1, \dots, X_N)$).

First, we prove an auxiliary lemma.

Lemma 2.3 *Let* $H_1, \dots, H_N$ *be closed linear subspaces of a Hilbert space* H. *Suppose that there exists a constant* $c > 0$ *such that, for any* $x_n \in H_n$,

$$\left\|\sum_{n=1}^{N} x_n\right\| \ge c \sum_{n=1}^{N} \|x_n\|.$$

Then, for any $x_1 \in H_1, \dots, x_N \in H_N$, *there exist* $y_1 \in H_1, \dots, y_N \in H_N$ *such that* $\mathrm{Pr}_{H_n} \sum_m y_m = x_n$ *for any* n.

Proof We prove this statement by induction in N. Let $N = 2$. Consider the sequence $(z_k) \in H$ defined by $z_1 = x_1$,

$$z_{k+1} = \begin{cases} z_k + x_2 - \mathrm{Pr}_{H_2} z_k & \text{if } k \text{ is odd,} \\ z_k + x_1 - \mathrm{Pr}_{H_1} z_k & \text{if } k \text{ is even.} \end{cases}$$

Denote $\delta_k = z_k - z_{k-1}$. Then

$$\delta_{k+1} = x_2 - \mathrm{Pr}_{H_2} z_k = \mathrm{Pr}_{H_2} z_{k-1} - \mathrm{Pr}_{H_2} z_k = -\mathrm{Pr}_{H_2} \delta_k \quad \text{if } k \text{ is odd,}$$
$$\delta_{k+1} = x_1 - \mathrm{Pr}_{H_1} z_k = \mathrm{Pr}_{H_1} z_{k-1} - \mathrm{Pr}_{H_1} z_k = -\mathrm{Pr}_{H_1} \delta_k \quad \text{if } k \text{ is even.}$$

It is easy to see that there exists $\gamma < 1$ such that $\|\mathrm{Pr}_{H_1} z\| \le \gamma\|z\|$ for any $z \in H_2$ (indeed, otherwise we can find $(u_n) \in H_2$ with $\|\mathrm{Pr}_{H_1} u_n\|/\|u_n\| \to 1$; then we obtain $\|u_n - \mathrm{Pr}_{H_1} u_n\|/(\|u_n\| + \|\mathrm{Pr}_{H_1} u_n\|) \to 0$, which is a contradiction). Clearly, we can choose $\gamma < 1$ such that we also have $\|\mathrm{Pr}_{H_2} z\| \le \gamma\|z\|$ for any $z \in H_1$. Then $\|\delta_k\| \le \gamma\|\delta_{k-1}\|$. This means that (z_k) has a limit (z_∞). As $\delta_k \in H_1$ for odd k and $\delta_k \in H_2$ for even k, we see that z_∞ is represented as $y_1 + y_2$ with $y_n \in H_n$. It is clear that y_1, y_2 satisfy the desired condition.

Suppose now that the statement is true for $N - 1$ and let us prove it for N. Denote $\widetilde{H}_1 = H_1 + \cdots + H_{N-1}$, $\widetilde{H}_2 = H_N$. Then the pair $(\widetilde{H}_1, \widetilde{H}_2)$ satisfies the conditions of the lemma. So, for the pair $\widetilde{x}_1 = x_1 + \cdots + x_{N-1}$, $\widetilde{x}_2 = x_N$, there exist $\widetilde{y}_1 \in \widetilde{H}_1$ and $\widetilde{y}_2 \in \widetilde{H}_2$ such that $\mathrm{Pr}_{\widetilde{H}_1}(\widetilde{y}_1 + \widetilde{y}_2) = \widetilde{x}_1$ and $\mathrm{Pr}_{\widetilde{H}_2}(\widetilde{y}_1 + \widetilde{y}_2) = \widetilde{x}_2$. We have $\widetilde{x}_1 = y_1 + \cdots + y_{N-1}$ with $y_n \in H_n$. Then the collection $y_1, \dots, y_{N-1}, \widetilde{y}_2$ satisfies the desired conditions. □

Proof of Theorem 2.2 (b) ⇒ (a) This implication follows from the above lemma.

(a) ⇒ (c) Denote by E the L^2-closure of $E_1 + \cdots + E_N$, where E_n are given by (2.2). Let φ be a solution of (1.1) corresponding to (φ_n). It is easy to see that the set of all solutions of (1.1) consists of the functions $\widetilde{\varphi}$ such that $\widetilde{\varphi} - \varphi$ is orthogonal to each E_n. In other words, the set of all solutions is $\varphi + E^\perp$, where $E^\perp$ is the orthogonal complement to E. Now, it is clear that the solution of (1.2) exists, is unique, and is given by $\mathrm{Pr}_E\, \varphi$.

We shall prove the representation of the solution later.

(a) ⇒ (d) This property easily follows from the description of the solution provided above.

(a) ⇒ (e) Consider the space $F = \prod_n E_n$ (i.e., F consists of collections $(\xi_1, \dots, \xi_N)$), where E_n are given by (2.2), endowed with the norm $\|(\xi_1, \dots, \xi_N)\| = (\sum_n \|\xi_n\|^2)^{1/2}$. Let E be the same as above. Then the map

$$E \ni \xi \quad \longmapsto \quad (\mathrm{Pr}_{E_1}\, \xi, \dots, \mathrm{Pr}_{E_n}\, \xi) \in F$$

is continuous (E is equipped with the L^2-norm), one-to-one (if $\xi, \xi' \in E$ have the same projections on each E_n, then $\xi - \xi'$ must be orthogonal to E, which is possible only if $\xi = \xi'$), and onto (due to (a)). Both E and F are Banach spaces. By the Banach theorem, the inverse map is continuous. This is just what we need.

(a) ⇒ (b) Fix $(\varphi_n) \in \Phi$ with $\|\varphi_n(X_n)\| = 1$. Denote $G_n = \{x\varphi_n(X_n) : x \in \mathbb{R}\}$, $G = \{\sum_n x_n\varphi_n(X_n) : x_n \in \mathbb{R}\}$. Then, for any $(x_1, \dots, x_N) \in \mathbb{R}^N$, there exists $\xi \in L^2$ such that $\mathrm{Pr}_{G_n}\, \xi = x_n\varphi_n(X_n)$ for any n. The same will be true for $\mathrm{Pr}_G\, \xi$ instead of ξ. So, for any $(x_1, \dots, x_N) \in \mathbb{R}^N$, we can find $\xi \in G$ such that $\mathrm{Pr}_{G_n}\, \xi = x_n\varphi_n(X_n)$ for any n. The random variables ξ corresponding to different collections $(x_1, \dots, x_N)$ must be different because $\|\varphi_n(X_n)\| = 1$. This implies that G has dimension N, i.e., $\varphi_1(X_1), \dots, \varphi_N(X_N)$ are linearly independent.

Fix $(x_1, \dots, x_N) \in \mathbb{R}^N$ and find $\xi \in L^2$ such that $\mathsf{E}[\xi \mid X_n] = x_n\varphi_n(X_n)$ for any n. The projection $\mathrm{Pr}_G\, \xi$ can be represented as $\sum_n y_n\varphi_n(X_n)$, and $y_1, \dots, y_N$ are determined uniquely due to the linear independence of $\varphi_1(X_1), \dots, \varphi_N(X_N)$. We then have

$$\mathrm{Pr}_{G_n} \sum_{m=1}^{N} y_m\varphi_m(X_m) = x_n\varphi_n(X_n), \quad n = 1, \dots, N,$$

which means that

$$\left\langle \sum_{m=1}^{N} y_m\varphi_m(X_m), \varphi_n(X_n) \right\rangle = x_n, \quad n = 1, \dots, N,$$

where $\langle \cdot, \cdot \rangle$ denotes the L^2-scalar product. This means that the vectors $x = (x_1, \dots, x_N)$ and $y = (y_1, \dots, y_N)$ are related by the equality $Ay = x$, where the matrix A is given by $A_{nm} = \langle \varphi_n(X_n), \varphi_m(X_m)\rangle$. Using condition (c), which follows from (a), we can write

$$\left\langle A^{-1}x, x\right\rangle^{1/2} = \langle y, Ay\rangle^{1/2} = \left\| \sum_{n=1}^{N} y_n \varphi_n(X_n) \right\| \le \|\xi\|$$
$$\le C \sum_{n=1}^{N} \left\| x_n \varphi_n(X_n) \right\| = C \sum_{n=1}^{N} |x_n| \le NC \langle x, x\rangle^{1/2},$$

where the scalar product $\langle x, y\rangle$ for $x, y \in \mathbb{R}^N$ is defined as $\sum_n x_n y_n$. As the above inequality is true for any $x \in \mathbb{R}^N$, we see that all the eigenvalues of the matrix A^{-1} do not exceed N^2C^2. Hence, all the eigenvalues of A are greater than or equal to $N^{-2}C^{-2}$. Finally, we get for any $(x_1, \dots, x_N)$

$$\left\| \sum_{n=1}^{N} x_n \varphi_n(X_n) \right\| = \langle x, Ax\rangle^{1/2} \ge N^{-1}C^{-1} \langle x, x\rangle^{1/2}$$
$$\ge N^{-2}C^{-1} \sum_{n=1}^{N} |x_n| = N^{-2}C^{-1} \sum_{n=1}^{N} \left\| x_n \varphi_n(X_n) \right\|.$$

(a) ⇒ (c) It remains to prove that the solution has the form (2.1). For this, we note that (a) implies (b), which, in turn, implies that the sum $E_1 + \cdots + E_N$ is L^2-closed (we are using the same notation as above). Hence, $E = E_1 + \cdots + E_N$. As shown above, any solution of (1.2) belongs to E. The proof is completed. □

Remark If $X_1, \dots, X_N$ are independent, then condition (b) of Theorem 2.2 is clearly satisfied. More generally, if there exists a measure Q on $\mathbb{R}^N$ with independent marginals and a constant c such that $c^{-1} \le dP/dQ \le c$, then the above condition is also satisfied since, for any random variable Z, we have $c^{-1/2}\|Z\|_Q \le \|Z\|_P \le c^{1/2}\|Z\|_Q$, where $\|Z\|_P$ (resp. $\|Z\|_Q$) denotes the $L^2(P)$-norm (resp. $L^2(Q)$-norm).

We next provide examples illustrating situations when a solution does not exist.

Example 2.4

(i) Let $N = 2$, P be concentrated on the line $\{y = x\}$, and $\varphi_1(x) = x$, $\varphi_2(x) = -x$. Then clearly there exists no solution of (1.2).

(ii) This is a non-degenerate example. Let $N = 2$ and $P = \frac{1}{2}(P_1 + P_2)$, where P_1 is the standard Gaussian distribution and P_2 is the point mass concentrated at zero. Let $\varphi_1(x) = I(x = 0)$, $\varphi_2(x) = -I(x = 0)$. If there exists a solution φ of (1.1), then $\mathrm{Pr}_{G_n} \varphi = \varphi_n(X_n)$, $n = 1, 2$, where $G_n = \{x\varphi_n(X_n) : x \in \mathbb{R}\}$. But this is impossible because $\varphi_1(X_1) = -\varphi_2(X_2)$ P-a.s.

(iii) This is an example of P having a density. Let ξ_1, ξ_2 be independent random variables, ξ_1 having the density $\frac{1}{2}\exp\{-|x|\}$ and ξ_2 being standard Gaussian.

Let $\eta_1 = \xi_1 + \xi_2$, $\eta_2 = \xi_1 - \xi_2$, and $P = \mathrm{Law}(\eta_1, \eta_2)$. The density of P is given by

$$p(x_1, x_2) = \frac{1}{4\sqrt{2\pi}} \exp\left\{-\frac{|x_1 + x_2|}{2} - \frac{(x_1 - x_2)^2}{8}\right\}.$$

It is easy to see from this expression that, for any $\varepsilon > 0$,

$$P\big(|x_1| > k, |x_2| > k, |x_2/x_1 - 1| < \varepsilon \,\big|\, |x_1| > k, |x_2| > k\big) \xrightarrow[k\to\infty]{1} .$$

Consider the functions

$$\varphi_1^k(x) = c_k x I(|x| > k), \qquad \varphi_2^k(x) = -c_k x I(|x| > k),$$

where c_k are chosen so that $\|\varphi_n^k(X_n)\| = 1$. Then $\|\varphi_1^k(X_1) + \varphi_2^k(X_2)\| \to 0$. So, condition (b) of Theorem 2.2 is not satisfied, which means that, for some $(\varphi_1, \varphi_2) \in \Phi$, there is no solution of (1.2) (actually, one can see that there is no solution for $(\varphi_1^k, \varphi_2^k)$ with k being large enough).

This example shows that if P is "built" from measures with different heaviness of the tail, then the solution might not exist.

3 Three particular cases

Throughout this section, we assume that $\mathsf{E}\varphi_n(X_n) = 0$ and $\mathsf{E}\varphi_n^2(X_n) < \infty$ for any n.

3.1 Independent components

Suppose that $X_1, \dots, X_N$ are independent under P. It is clear from Lemma 2.1 that the solution of (1.2) is given by

$$\varphi(x_1, \dots, x_N) = \sum_{n=1}^{N} \varphi_n(x_n).$$

3.2 Gaussian distribution

Suppose that P is a Gaussian non-degenerate distribution. We assume that $\mathsf{E}X_n = 0$ and $\mathsf{E}X_n^2 = 1$. This will slightly simplify the formulas and is sufficient for our applications. Indeed, the main application we have in mind is the Gaussian copula (next subsection), and a non-degenerate Gaussian copula can always be transformed to such a distribution.

Consider the Hermite polynomials $H_m(x)$, $m = 0, 1, 2, \dots$. Recall that one way to define them is as

$$H_m(x) = \frac{1}{\sqrt{m!}} \frac{\partial^m}{\partial a^m}\bigg|_{a=0} \exp\{ax - a^2/2\}, \quad x \in \mathbb{R}.$$

For example,

$$H_0(x) = 1,$$
$$H_1(x) = x,$$
$$H_2(x) = \left(x^2 - 1\right)/\sqrt{2},$$
$$H_3(x) = \left(x^3 - 3x\right)/\sqrt{6}.$$

Denote, for $n = 1, \dots, N, m \in \mathbb{N}$,

$$a_{nm} = \mathsf{E}\left[\varphi_n(X_n)H_m(X_n)\right] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \varphi_n(x)H_m(x)e^{-x^2/2}\,dx. \tag{3.1}$$

Denote by C the covariance matrix of $(X_1, \dots, X_N)$ and let C^m denote its mth componentwise power, i.e., C^m_{kn} is the mth power of C_{kn}. As follows from Lemma A.2, each C^m is symmetric, positive definite, and non-degenerate. Hence, for each $m \in \mathbb{N}$, the vector

$$\begin{pmatrix} \alpha_{1m} \\ \vdots \\ \alpha_{Nm} \end{pmatrix} = \left(C^m\right)^{-1} \begin{pmatrix} a_{1m} \\ \vdots \\ a_{Nm} \end{pmatrix} \tag{3.2}$$

is well defined.

Theorem 3.1 *The solution of* (1.2) *is given by the* L^2*-convergent series*

$$\varphi(x_1, \dots, x_N) = \sum_{n,m=1}^{N,\infty} \alpha_{nm} H_m(x_n), \quad x_n \in \mathbb{R}. \tag{3.3}$$

Proof Let us prove the L^2-convergence of the series (3.3). The matrices $C^m, m \in \mathbb{N}$, are non-degenerate by Lemma A.2. As they converge to the identity matrix, there exists $\mu > 0$ such that

$$\left\|C^m x\right\| \geq \mu\|x\|, \quad x \in \mathbb{R}^N,\ m \in \mathbb{N},$$

where by $\|\cdot\|$ we denote the Euclidean norm. Then

$$\sum_{n,m=1}^{N,\infty} \alpha_{nm}^2 \leq \mu^{-2} \sum_{n,m=1}^{N,\infty} a_{nm}^2 \leq \sum_{n=1}^{N} \left\|\varphi_n(X_n)\right\|^2 < \infty.$$

The second inequality follows from the orthonormality of $(H_m(X_n))_{m=1}^{\infty}$ (which is well known and follows, in particular, from Lemma A.1). As $\|H_m(X_n)\| = 1$, we get the claim.

Let us prove that φ satisfies (1.1). Consider the spaces E_n given by (2.2). Then $(H_m(X_n))_{m=1}^{\infty}$ forms an orthonormal basis in E_n. Employing Lemma A.1, we can write

$$\begin{aligned}\mathsf{E}[\varphi \mid X_n] = \mathrm{Pr}_{E_n}\varphi &= \sum_{m=1}^{\infty}\langle \varphi, H_m(X_n)\rangle H_m(X_n)\\&= \sum_{k,m=1}^{N,\infty} \alpha_{km}\langle H_m(X_k), H_m(X_n)\rangle H_m(X_n)\\&= \sum_{k,m=1}^{N,\infty} \alpha_{km} C_{kn}^m H_m(X_n)\\&= \sum_{m=1}^{\infty} a_{nm} H_m(X_n)\\&= \varphi_n(X_n), \quad n = 1, \dots, N.\end{aligned}$$

An application of Lemma 2.1 completes the proof. □

As is well known, the Hermite polynomials result from orthogonalizing the system of polynomials with respect to the Gaussian measure. Therefore, $\mathsf{E}[X_n^M H_m(X_n)] = 0$ for $m > M$. Consequently, if each φ_n is a polynomial of degree M, then $a_{nm} = 0$ for $m > M$. Thus, we get

Corollary 3.2

(i) *Suppose that*

$$\varphi_n(x) = a_n x, \quad n = 1, \dots, N.$$

Set $\alpha = C^{-1}a$, *where* $a = (a_1, \dots, a_N)$ (*we are considering all vectors as column vectors*). *Then the solution of* (1.2) *is given by*

$$\varphi(x_1, \dots, x_N) = \sum_{k=1}^{N} \alpha_n x_n.$$

(ii) *Suppose that*

$$\varphi_n(x) = a_n(x^2 - 1), \quad n = 1, \dots, N.$$

Set $\alpha = (C^2)^{-1}a$, *where* C^2 *is the componentwise square of* C. *Then the solution of* (1.2) *is given by*

$$\varphi(x_1, \dots, x_N) = \sum_{n=1}^{N} \alpha_n(x_n^2 - 1).$$

(iii) *Suppose that each* φ_n *is a polynomial of degree* M. *Then the solution of* (1.2) *is given by* (3.3) *with the summation in* m *going up to* M.

Remark Let $(R, X_1, \dots, X_N)$ be a Gaussian random vector with $\mathsf{E}R = 0$, $\mathsf{E}X_n = 0$, $\mathsf{E}X_n^2 = 1$. In this case, $\mathsf{E}[R \mid X_n] = a_n x$, where $a_n = \mathsf{E}RX_n$, while

$$\mathsf{E}[R \mid X_1, \dots, X_N] = \sum_{n=1}^{N} \alpha_n x_n,$$

where $\alpha = C^{-1}a$ and C is the correlation matrix of $X_1, \dots, X_N$. This coincides with the function provided by Corollary 3.2(i). In other words, in the case of jointly Gaussian factors and hedge fund returns, our technique allows us to completely recover the conditional expectation of the hedge fund return given the factors.

3.3 Gaussian copula

Suppose that P is a non-degenerate Gaussian copula, i.e., there exist a non-degenerate Gaussian vector $(\widetilde{X}_1, \dots, \widetilde{X}_N)$ and increasing functions $f_n : \mathbb{R} \to \mathbb{R}$ such that $P = \text{Law}(f_1(\widetilde{X}_1), \dots, f_N(\widetilde{X}_N))$. We can arrange f_n and $\widetilde{X}_n$ in such a way that $\mathsf{E}\widetilde{X}_n = 0$ and $\mathsf{E}\widetilde{X}_n^2 = 1$ for any n.

Denote $\widetilde{P} = \text{Law}(\widetilde{X}_1, \dots, \widetilde{X}_N)$ and $\widetilde{\varphi}_n = \varphi_n \circ f_n$. Let $\widetilde{\varphi}$ be the solution of the problem based on $\widetilde{P}$ and $(\widetilde{\varphi}_n)$, which was provided in the previous subsection. Then clearly, the function

$$\varphi(x_1, \dots, x_N) = \widetilde{\varphi}\big(f_1^{-1}(x_1), \dots, f_N^{-1}(x_N)\big), \tag{3.4}$$

where f_n^{-1} is the right-continuous inverse of f_n, satisfies (1.1). As $\widetilde{\varphi}$ is the sum of functions of one variable, the same is true for φ. By Lemma 2.1, φ is the solution of (1.2).

To sum up, the procedure for solving (1.2) consists of the following steps:

1. Estimate from data the distribution $P = \text{Law}(X_1, \dots, X_N)$ (assumed here to be a Gaussian copula) and the conditional expectations $\varphi_n(x) = \mathsf{E}[R \mid X_n = x]$.
2. Find functions f_n such that $P = \text{Law}(f_1(\widetilde{X}_1), \dots, f_N(\widetilde{X}_N))$, where $(\widetilde{X}_1, \dots, \widetilde{X}_N)$ is Gaussian with $\mathsf{E}\widetilde{X}_n = 0$ and $\mathsf{E}\widetilde{X}_n^2 = 1$.
3. Fix a number $M \in \mathbb{N}$ (for example, 30) and determine the coefficients $(a_{nm}; n = 1, \dots, N, m = 1, \dots, M)$ given by (3.1) with X_n replaced by $\widetilde{X}_n$ and φ_n replaced by $\widetilde{\varphi}_n = \varphi_n \circ f_n$.
4. Find the values $(\alpha_{nm}; n = 1, \dots, N, m = 1, \dots, M)$ by solving the linear systems (3.2) for $m = 1, \dots, M$, where C is the covariance matrix of $(\widetilde{X}_1, \dots, \widetilde{X}_N)$.
5. Define the function $\widetilde{\varphi}$ by (3.3) and find the function φ given by (3.4). This is the desired solution.

4 Multidimensional factors

In this section, we consider the problem (1.2) for multidimensional X_n, i.e., we suppose that $X_n = (X_n^1, \dots, X_N^{d_N})$ is a random vector (the dimensions d_n are different for different n). The practical motivation comes from the fact that some factors, like the price of oil, are inherently one-dimensional, while others, like different parts of the yield curve, are inherently multidimensional. The considerations of Sect. 2 as well as Sects. 3.1 and 3.3 admit a straightforward extension to multidimensional X_n. This is not the case for Sect. 3.2, and the present section will deal with the corresponding extension.

Thus, we assume that the vector $(X_n^i : n = 1, \dots, N, i = 1, \dots, d_n)$ is Gaussian and non-degenerate, and each $\varphi_n : \mathbb{R}^{d_n} \to \mathbb{R}$ satisfies $\mathsf{E}\varphi_n(X_n) = 0$ and $\mathsf{E}\varphi_n^2(X_n) < \infty$.

4.1 Two factors

We first consider the case $N = 2$. Without loss of generality, $d_1 \ge d_2$. We can represent each X_n as $A_n Y_n$, where A_n is a non-degenerate $d_n \times d_n$ matrix and Y_n has a standard Gaussian distribution in $\mathbb{R}^{d_n}$, i.e., $\mathsf{cov}(Y_n^i, Y_n^j) = I\ (i = j)$. Let C denote the covariance matrix between Y_1 and Y_2, i.e., $C^{ij} = \mathsf{cov}(Y_1^i, Y_2^j)$. According to the *singular value decomposition* (SVD), there exist a $d_1 \times d_1$ unitary matrix U_1 and a $d_2 \times d_2$ unitary matrix U_2 such that $C = U_1 D U_2^t$ ("t" denotes the transpose) with a $d_1 \times d_2$ diagonal matrix D, i.e., $D_{ij} = 0$ for $i \neq j$. SVD is a standard numerical tool (see [6]), and procedures for finding U_1, U_2 are implemented in most mathematical packages, like MATLAB. Thus, Y_n can be represented as $Y_n = U_n Z_n$ (we are considering all random vectors as column vectors), where $Z_1 = (Z_1^1, \dots, Z_1^{d_1})$, $Z_2 = (Z_2^1, \dots, Z_2^{d_2})$ are jointly Gaussian vectors with $\mathsf{cov}(Z_1^i, Z_2^j) = 0$ for $i \neq j$. As U_n is unitary and Y_n has a standard Gaussian distribution, the same is true for Z_n. As a result, we can represent X_n as $B_n Z_n$ with $B_n = A_n U_n$.

Denote by M_n the space of non-zero multiindices of length d_n, i.e., M_n consists of collections $(m(1), \dots, m(d_N))$, where the $m(i)$ take the values $0, 1, 2, \dots$ and at least one $m(i)$ is greater than zero. For $m \in M_n$, we consider the corresponding multidimensional Hermite polynomial

$$\bar{H}_m\big(z^1, \dots, z^{d_n}\big) = H_{m(1)}\big(z^1\big) \cdots H_{m(d_n)}\big(z^{d_n}\big), \quad \big(z^1, \dots, z^{d_n}\big) \in \mathbb{R}^{d_n}.$$

Consider the functions $\widetilde{\varphi}_n(z) = \varphi_n(B_n z)$ and set

$$a_{nm} = \big\langle \widetilde{\varphi}_n(Z_n), \bar{H}_m(Z_n)\big\rangle, \quad n = 1, 2,\ m \in M_n.$$

There exists a natural inclusion of M_2 in M_1; we informally write $M_2 \subseteq M_1$. For $m \in M_2$, we denote

$$\rho^m = \big(\rho^1\big)^{m(1)} \cdots \big(\rho^{d_2}\big)^{m(d_2)},$$

where $\rho^i = \mathsf{cov}(Z_1^i, Z_2^i)$. For $m \in M_2$, set

$$\alpha_{1m} = \frac{a_{1m} - \rho^m a_{2m}}{1 - (\rho^m)^2}, \qquad \alpha_{2m} = \frac{a_{2m} - \rho^m a_{1m}}{1 - (\rho^m)^2}.$$

For $m \in M_1 \setminus M_2$, we set $\alpha_{1m} = a_{1m}$. Consider the function

$$\widetilde{\varphi}(z_1, z_2) = \sum_{m \in M_1} \alpha_{1m} \bar{H}_m(z_1) + \sum_{m \in M_2} \alpha_{2m} \bar{H}_m(z_2), \quad z_1 \in \mathbb{R}^{d_1},\ z_2 \in \mathbb{R}^{d_2}$$

and let $\varphi(x_1, x_2) = \widetilde{\varphi}(B_1^{-1}x_1, B_2^{-1}x_2)$.

Let us prove that φ is the desired solution of (1.2). To verify the L^2-convergence of the above series, note that the vector (Z_1, Z_2) is non-degenerate, and therefore, there exists $\lambda < 1$ such that $|\rho^i| \le \lambda$ for any $i = 1, \dots, d_2$. Then $|\rho^m| \le \lambda$ for any $m \in M_2$, which implies that $|\alpha_{nm}| \le (1-\lambda^2)^{-1}(|a_{1m}| + |a_{2m}|)$ for $m \in M_2$. As the system $(\bar{H}_m(X_n))_{m\in M_n}$ is orthonormal, we get $\sum_{m\in M_n} a_{nm}^2 < \infty$, and hence, $\sum_{m\in M_n} \alpha_{nm}^2 < \infty$.

Clearly, $\|\bar{H}_m(Z_n)\| = 1$ for any $m \in M_n$ and

$$\big\langle H_{m_1}(Z_1), H_{m_2}(Z_2)\big\rangle = \begin{cases} 0 & \text{if } m_1 \ne m_2, \\ \rho^{m_2} & \text{if } m_1 = m_2, \end{cases}$$

where the equality $m_1 = m_2$ means that $m_1(i) = m_2(i)$ for $i \le d_2$ and $m_1(i) = 0$ for $i > d_2$. For $m \in M_1 \setminus M_2$, we have

$$\big\langle \widetilde{\varphi}(Z_1, Z_2), \bar{H}_m(Z_1)\big\rangle = \alpha_{1m} = a_{1m} = \big\langle \widetilde{\varphi}_1(Z_1), \bar{H}_m(Z_1)\big\rangle.$$

For $m \in M_2$, we have

$$\big\langle \widetilde{\varphi}(Z_1, Z_2), \bar{H}_m(Z_1)\big\rangle = \alpha_{1m} + \alpha_{2m}\rho^m = a_{1m} = \big\langle \widetilde{\varphi}_1(Z_1), \bar{H}_m(Z_1)\big\rangle.$$

As the system $(\bar{H}_m(Z_1))_{m\in M_1}$ forms an orthonormal basis in the space E_1 of Z_1-measurable square-integrable random variables with zero mean (see [9, Chap. II, §2]), we see that $\mathrm{Pr}_{E_1} \widetilde{\varphi}(Z_1, Z_2) = \widetilde{\varphi}_1(Z_1)$. In other words, $\mathsf{E}[\widetilde{\varphi}(Z_1, Z_2) \mid Z_1] = \widetilde{\varphi}_1(Z_1)$, which means that $\mathsf{E}[\varphi(X_1, X_2) \mid X_1] = \varphi_1(X_1)$. In the same way, we prove that $\mathsf{E}[\varphi(X_1, X_2) \mid X_2] = \varphi_2(X_2)$. Now, the multidimensional analogue of Lemma 2.1 guarantees that φ is the desired solution.

4.2 Multiple factors

Consider now an arbitrary N. The solution of (1.2) will be constructed in a sequence of steps. Without loss of generality, $d_1 \ge \dots \ge d_N$.

Step 1. Applying the above procedure to $\widetilde{X}_1 = X_1$, $\widetilde{X}_2 = X_2$, $\widetilde{\varphi}_1 = \varphi_1$ and $\widetilde{\varphi}_2 = \varphi$, we get a function $\varphi_{12} : \mathbb{R}^{d_1+d_2} \to \mathbb{R}$ such that

$$\mathsf{E}\big[\varphi_{12}(X_1, X_2) \mid X_n\big] = \varphi_n(X_n), \quad n = 1, 2.$$

Step 2. Applying the above procedure to $\widetilde{X}_1 = (X_1, X_2)$, $\widetilde{X}_2 = X_3$, $\widetilde{\varphi}_1 = \varphi_{12}$ and $\widetilde{\varphi}_2 = \varphi_3$, we get a function $\varphi_{123} : \mathbb{R}^{d_1+d_2+d_3} \to \mathbb{R}$ such that

$$\mathsf{E}\big[\varphi_{123}(X_1, X_2, X_3) \mid X_1, X_2\big] = \varphi_{12}(X_1, X_2),$$
$$\mathsf{E}\big[\varphi_{123}(X_1, X_2, X_3) \mid X_3\big] = \varphi_3(X_3).$$

The first equality implies that

$$\mathsf{E}\big[\varphi_{123}(X_1, X_2, X_3) \mid X_n\big] = \mathsf{E}\big[\varphi_{12}(X_1, X_2) \mid X_n\big] = \varphi_n(X_n), \quad n = 1, 2.$$

Proceeding in the same way, we construct the desired solution at the $(N-1)$th step.

5 Factor risks

Let R be the return/P&L of a portfolio over a certain time period and $F_1, \dots, F_N$ the returns/increments of factors over the same period. We assume that $\mathsf{E}R^2 < \infty$. Denote

$$P = \mathrm{Law}(F_1, \dots, F_N),$$
$$\bar{R} = R - \mathsf{E}R,$$
$$\varphi_n(x) = \mathsf{E}[\bar{R} \mid F_n = x],$$
$$\eta(x_1, \dots, x_N) = \mathsf{E}[\bar{R} \mid F_1 = x_1, \dots, F_N = x_N].$$

Suppose that there exists a solution φ of (1.2) (clearly, it is unique up to P-indistinguishability) and that it has the form

$$\varphi(x_1, \dots, x_N) = \sum_{n=1}^{N} \psi_n(x_n), \tag{5.1}$$

which is the typical situation, as discussed in Sect. 2. We assume that $\mathsf{E}\psi_n(F_n) = 0$, which ensures that the ψ_n are determined uniquely in typical situations.

Consider the decomposition

$$R = R^0 + R^1 + R^2 + R^3,$$

where

$$R^0 = \mathsf{E}R,$$
$$R^1 = \varphi(F_1, \dots, F_N),$$
$$R^2 = \eta(F_1, \dots, F_N) - \varphi(F_1, \dots, F_N),$$
$$R^3 = \bar{R} - \eta(F_1, \dots, F_N).$$

Theorem 5.1

(i) *We have*

$$\varphi = \operatorname{argmin} \mathsf{E}\big(\bar{R} - \tilde{\varphi}(F_1, \dots, F_N)\big)^2,$$

where the minimum is taken over all the functions $\tilde{\varphi}$ which are sums of one-dimensional functions of single variables.

(ii) *The random variables R^1, R^2, R^3 are uncorrelated.*

Proof (i) Let

$$E_n = \big\{\xi \in L^2 : \xi \text{ is } F_n\text{-measurable}, \mathsf{E}\xi = 0\big\},$$

and denote by E the L^2-closure of $E_1 + \dots + E_N$. A function $\widetilde{\varphi}$ satisfies (1.1) if and only if $\mathrm{Pr}_{E_n} \widetilde{\varphi}(F_1, \dots, F_N) = \varphi_n(F_n)$ for any n (here Pr is the L^2-projection). This is the same as saying that $\mathrm{Pr}_{E_n} \widetilde{\varphi}(F_1, \dots, F_N) = \mathrm{Pr}_{E_n} \bar{R}$ for any n. The latter property is clearly equivalent to $\mathrm{Pr}_E \widetilde{\varphi}(F_1, \dots, F_N) = \mathrm{Pr}_E \bar{R}$. Now, it is clear that the solution of (1.2) satisfies

$$\varphi(F_1, \dots, F_N) = \mathrm{Pr}_E \bar{R},$$

from which (i) is obvious.

(ii) Consider the space

$$L = \{\xi \in L^2 : \xi \text{ is } (F_1, \dots, F_N)\text{-measurable}, \mathsf{E}\xi = 0\}.$$

Then

$$\eta(F_1, \dots, F_N) = \mathrm{Pr}_L \bar{R},$$

so that

$$R^1 = \mathrm{Pr}_E \bar{R}, \qquad R^2 = \mathrm{Pr}_L \bar{R} - \mathrm{Pr}_E \bar{R}, \qquad R^3 = \bar{R} - \mathrm{Pr}_L \bar{R}.$$

As $E \subseteq L$, we get the desired statement. □

Thus, we obtain for R the L^2-orthogonal decomposition

$$R = \mathsf{E}R + \sum_{n=1}^{N} \psi_n(F_n) + R^2 + R^3.$$

Combining this with the well-known Euler principle, we are led to the following definition. Below ρ is a risk measure (e.g. standard deviation, V@R, TailV@R, coherent risk, etc.). We also assume that the derivatives below exist.

Definition 5.2 The *factor risks* of R brought by the factors $F_1, \dots, F_N$ are defined as

$$\rho_n = \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \rho\big(R + \varepsilon\psi_n(F_n)\big), \quad n = 1, \dots, N.$$

The *cross-term risk* of R is defined as

$$\rho_C = \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \rho\big(R + \varepsilon R^2\big).$$

The *idiosyncratic risk* of R is defined as

$$\rho_I = \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \rho\big(R + \varepsilon R^3\big).$$

To avoid trivial complications, assume now that $\mathsf{E}R = 0$. If ρ is homogeneous (which is the case for all natural risk measures, in particular, those mentioned above), then in natural situations[3]

$$\rho(R) = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \rho(R + \varepsilon R) = \sum_{n=1}^{N} \rho_n + \rho_C + \rho_I. \tag{5.2}$$

Thus, the overall risk of the portfolio is typically the sum of the risks introduced in Definition 5.2, i.e.,

$$\text{Risk} = \text{Sum of factor risks} + \text{Cross-term risk} + \text{Idiosyncratic risk}.$$

To conclude the section, let us compare the notion of the factor risk introduced here with the ones introduced by Cherny and Madan [2] and Martin and Tasche [10]. According to the definition from [2], the *factor risk* of R brought by F_n is defined (in the current notation) as $\rho(\varphi_n(F_n))$. The *risk impact* proposed in [10] is the derivative $d/d\varepsilon|_{\varepsilon=0}\rho(R + \varepsilon\varphi_n(F_n))$ (to be more precise, the authors of [10] divide this quantity by $\rho(R)$, but this is not important for our discussion). The problem with both definitions is that they do not account for the correlation between the factors, so that the sum of factor risks thus defined might be substantially different (smaller or larger) from the overall risk of the portfolio. The essence of the approach proposed here is passing from the functions φ_n to the functions ψ_n, which accounts for the correlation between the factors. The effect of that operation is seen from the example below.

Example 5.3 Suppose that F_1, F_2 are jointly Gaussian with $\mathsf{E}F_n = 0$, $\mathsf{E}F_n^2 = 1$ and $\mathsf{corr}(F_1, F_2) = c$. Let $R = F_1 + F_2$. If ρ is any reasonable risk measure (like standard deviation, V@R, TailV@R, or law invariant coherent risk), then there exists a constant $\gamma > 0$ such that, for any centered Gaussian random variable ξ, we have $\rho(\xi) = \gamma\sigma(\xi)$, where σ denotes the standard deviation.

In this case, $\rho(R) = \gamma\sigma(R)$. Furthermore, $\varphi_n(x) = (1 + c)x, n = 1, 2$. So, for the factor risk from [2], we have

$$\rho\big(\varphi_n(F_n)\big) = \gamma(1 + c), \quad n = 1, 2.$$

For the risk impact from [10], we have (skipping some trivial calculations)

$$\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \rho\big(R + \varepsilon(1 + c)F_n\big) = \gamma\sigma(R)\frac{1 + c}{2}, \quad n = 1, 2.$$

If c is close to -1, then the risk of R is of order $(1 + c)^{1/2}$, while the sum of factor risks (resp. risk impacts) is of order $1 + c$ (resp. $(1 + c)^{3/2}$); so in this case the sum of factor risks or risk impacts is considerably smaller than the risk of R. If c is close to 1, then the risk of R is close to 2γ, while the sum of factor risks or risk impacts is close to 4γ; so in this case the sum of factor risks or risk impacts is larger than the risk of R, and this effect is strengthened further if we have multiple rather than two

[3] Of course, it is easy to provide more or less degenerate examples, in which this relation is violated, so we are making no strict statement here.

correlated factors. Thus, neither of the discussed definitions provides reliable results in case of high (positive or negative) correlation between the factors.

On the other hand, the functions ψ_n given by (5.1) have the form $\psi_n(x) = x$, $n = 1, 2$, as follows from Corollary 3.2(i). Then the factor risks of Definition 5.2 are

$$\rho_n = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \rho(R + \varepsilon F_n) = \frac{\gamma\sigma(R)}{2}, \quad n = 1, 2.$$

As a result, $\rho(R) = \gamma\sigma(R) = \rho_1 + \rho_2$, which is also clear from (5.2) since in this example $\rho_C = \rho_I = 0$.

In this example, irrespectively of the correlation between the factors, the sum of factor risks equals the risk of the portfolio. In the general situation, the equality might not be true due to the existence of cross-term risk and idiosyncratic risk, but still the sum of factor risks would provide a reasonable approximation to the overall risk, while each factor risk would provide a reasonable estimate of the contribution of that factor to the overall portfolio risk.

6 Conclusion

We propose a new methodology for estimating the risk of portfolios. It can be particularly useful in cases when portfolios exhibit nonlinear dependence on the risk driving factors and have scarce observations. This is typical for the case of hedge funds, so that our results can be useful for estimating portfolios of hedge funds, which is the fund of funds problem. Our methodology consists of two steps: first, performing a nonlinear regression of the portfolio return on each of the risk driving factors and second, joining the obtained nonlinear estimates together in a way that would capture the dependence between the factors.

In this paper, we propose an approach to performing the second step, which consists in solving a certain optimization problem. In the general setup, we provide a necessary and sufficient condition for the existence and uniqueness of a solution. Furthermore, we provide an explicit and numerically computable solution in the case when the joint law of the factors is a Gaussian copula, which is a very popular model in modern risk measurement.

The obtained results lead us to a new definition of factor risks (i.e., risks of a portfolio brought by each of the factors), which takes into account both the nonlinear dependence of the portfolio return on the factors and the correlation between the factors. We also propose a new decomposition of the portfolio risk into an orthogonal sum of factor risks, cross-term risk, and idiosyncratic risk.

Acknowledgements We express our thanks to three anonymous referees and the Associate Editor for the very careful reading of the paper and a number of valuable comments and suggestions that led to improving its quality.

Appendix

For completeness of exposition, we give here the proofs of some known facts from probability and linear algebra.

Lemma A.1 *Let X, Y be a jointly Gaussian vector with* $\mathsf{E}X = \mathsf{E}Y = 0$, $\mathsf{E}X^2 = \mathsf{E}Y^2 = 1$ *and* $\mathsf{E}[XY] = \rho$. *Then*

$$\langle H_m(X), H_k(Y)\rangle = \begin{cases} 0 & \text{if } m \neq k, \\ \rho^m & \text{if } m = k. \end{cases}$$

Proof Write down the Taylor expansion

$$\exp\{ax - a^2/2\} = \sum_{m=0}^{\infty} H_m(x)\frac{a^m}{m!}, \quad a, x \in \mathbb{R}.$$

Then

$$\mathsf{E}\big[\exp\{aX - a^2/2\}\exp\{bY - b^2/2\}\big] = \sum_{m,k=0}^{\infty} \mathsf{E}\big[H_m(X)H_k(Y)\big]\frac{a^m b^k}{m!k!}, \quad a, b \in \mathbb{R}.$$

On the other hand,

$$\begin{aligned} &\mathsf{E}\big[\exp\{aX - a^2/2\}\exp\{bY - b^2/2\}\big] \\ &\quad = \mathsf{E}\exp\big\{aX + bY - \big(a^2 + 2\rho ab + b^2\big)/2 + \rho ab\big\} \\ &\quad = \exp\{\rho ab\} = \sum_{m=0}^{\infty} \rho^m \frac{a^m b^m}{m!}, \quad a, b \in \mathbb{R}. \end{aligned}$$

Equating the coefficients in the two series, we get the result. □

The next result goes back to Jacobi.

Lemma A.2 *Let A, B be two symmetric positive definite non-degenerate N-dimensional matrices. Then their componentwise product $(A_{nk}B_{nk})_{n,k=1}^N$ has the same properties.*

Proof Consider independent N-dimensional Gaussian vectors X, Y with mean zero and covariance matrices A, B, respectively. Then the covariance matrix of the vector $(X_1Y_1, \dots, X_NY_N)$ is exactly the componentwise product of A and B. Thus, the matrix is positive definite. To show its non-degeneracy, assume that there exists a vector $(a_1, \dots, a_N)$ such that $\sum_n a_n X_n Y_n = 0$. We can find an equivalent measure Q under which X, Y remain independent and have independent components. Then the above equality holds Q-a.s., which implies that each $a_n X_n Y_n$ is degenerate under Q and hence, under the original measure. As $\mathsf{E}[X_n Y_n] = 0$, we get $a_n X_n Y_n = 0$, i.e., $a_n = 0$. □

References

1. Agarwal, V., Naik, N.: Risks and portfolio decisions involving hedge funds. Rev. Financ. Stud. **17**, 63–98 (2004)

2. Cherny, A.S., Madan, D.B.: Coherent measurement of factor risks. Preprint (2006). www.ssrn.com
3. Fung, W., Hsieh, D.A.: Empirical characteristics of dynamic trading strategies: the case of hedge funds. Rev. Financ. Stud. **10**, 275–302 (1997)
4. Fung, W., Hsieh, D.A.: Survivorship bias and investment style in the returns of CTAs. J. Portf. Manag. **23**, 30–41 (1997)
5. Fung, W., Hsieh, D.A.: The risk in hedge fund strategies: theory and evidence from trend followers. Rev. Financ. Stud. **14**, 313–341 (2001)
6. Golub, G.H., Van Loan, C.F.: Matrix Computations, 3rd edn. Johns Hopkins University Press, Baltimore (1996)
7. Hasanhodzic, J., Lo, A.: Can hedge-fund returns be replicated?: The linear case. J. Invest. Manag. **5**, 5–45 (2007)
8. Lo, A.: Risk management for hedge funds: Introduction and overview. Financ. Anal. J. **57**, 16–33 (2001)
9. Malyshev, V.A., Minlos, R.A.: Gibbs Random Fields. Kluwer, Dordrecht (1991)
10. Martin, R., Tasche, D.: Shortfall: A tail of two parts. Risk **20**, 84–89 (2007)
11. Rosen, D., Saunders, D.: Analytical methods for hedging systematic credit risk with linear factor portfolios. J. Econ. Dyn. Control **33**, 37–52 (2009)
12. Rosen, D., Saunders, D.: Risk factor contributions in portfolio credit risk models. Preprint (2006). www.defaultrisk.com