ON MEASURING RISK WITH SCARCE OBSERVATIONS

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

**Department of Probability Theory*
Faculty of Mathematics and Mechanics
Moscow State University
119992 Moscow Russia
E-mail: alexander.cherny@gmail.com

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

****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

March 2008

Abstract. We consider the problem of measuring the risk of a portfolio with scarce observations by linking it to several risk factors. A typical example is measuring the risk of a hedge fund. It is assumed that from the available data one can estimate the joint law of all the factors as well as all the 2-dimensional joint laws of the portfolio's return and increments of each factor. The problem is to recover the conditional mean of the portfolio's return given the values for all factors. We present an analytic computationally feasible solution of this problem for the case when the joint law of factors is a Gaussian copula.

Key words: Cross-term risk, factor risk, Gaussian copula, hedge fund replication, hedge fund risk, idiosyncratic risk, monomial risk, non-linear regression, risk measurement.

1 Introduction

1. Financial problem. Suppose that we want to measure the risk of a hedge fund. The returns of hedge funds are reported monthly, so we are measuring the risk of a monthly return. The simplest way of doing that would be to take just the empirical distribution. However, a more advanced procedure would consist in: first, relating the return of the hedge fund to the values of certain financial factors, like the price of oil or the S&P 500 index; second, estimating the joint distribution of factors by a certain time series model. This would yield the distribution of the future return.

In this paper, we will deal with the first problem. Fix a time period $[0, T]$. Let R denote the return of the hedge fund over this period and let $X_1, \dots, X_N$ be the increments of the factors over this period (nothing will change if instead we consider their values at time T). Alternatively, R might be the return or the Profit&Loss of an arbitrary portfolio over $[0, T]$. The problem is then to estimate the conditional distribution

$$\mathrm{Law}(R \,|\, X_1 = x_1, \dots, X_N = x_N). \tag{1.1}$$

If the joint law of $X_1, \dots, X_N$ is known (which is reasonable since otherwise the whole methodology makes no sense), then the above problem is equivalent to estimating the joint law of $R, X_1, \dots, X_N$. This is the point when the major problem arises.

The variable R has scarce observations. If, for example, the hedge fund has a two-year history, then we would have two dozens observations of R. This is sufficient to estimate the distribution of R and might also be sufficient to estimate the joint distributions $\mathrm{Law}(R, X_n)$. However, this data is insufficient to estimate the triple distributions $\mathrm{Law}(R, X_n, X_m)$ and is totally insufficient to estimate $\mathrm{Law}(R, X_1, \dots, X_N)$ as in modern risks measurement schemes N is of order of several hundreds. So, the problem we are faced with is to recover $\mathrm{Law}(R, X_1, \dots, X_N)$ from the knowledge of $\mathrm{Law}(X_1, \dots, X_N)$ and $\mathrm{Law}(R, X_n)$, $n = 1, \dots, N$.

2. Mathematical formulations. In this paper, instead of trying to recover the whole conditional distribution (1.1), we will recover only the conditional expectation $\mathsf{E}[R \mid X_1 = x_1, \dots, X_N = x_n]$. For this, we will need only the conditional expectations $\mathsf{E}[R \,|\, X_n = x]$ instead of the distributions $\mathrm{Law}(R, X_n)$.

The strict mathematical formulation is as follows. We are given a measure P on $\mathbb{R}^N$ (this is the law of $X_1, \dots, X_N$) and the functions $\varphi_n : \mathbb{R} \to \mathbb{R}$ ($\varphi_n(x)$ means $\mathsf{E}[R \mid X_n = x]$). We want to find a function $\varphi : \mathbb{R}^N \to \mathbb{R}$ ($\varphi(x_1, \dots, x_N)$ has the meaning $\mathsf{E}[R \,|\, X_1 = x_1, \dots, X_N = x_n]$) such that

$$\mathsf{E}[\varphi(X_1, \dots, X_N) \,|\, X_n] = \varphi_n(X_n), \quad n = 1, \dots, N, \tag{1.2}$$

where X_n denotes the n-th coordinate projection of $\mathbb{R}^N$ on $\mathbb{R}$.

A reformulation of (1.2) is: to recover a function $\varphi : \mathbb{R}^N \to \mathbb{R}$ by knowing its integrals over all the hyperplanes orthogonal to each of 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 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 (1.2) provides us only with NM equations. One more example, if X_1, X_2 are independent standard Gaussian and $\varphi(x_1, x_2) = x_1 x_2$, then $\mathsf{E}[\varphi(X_1, X_2) \,|\, X_n] = 0$, so that adding φ to any solution of (1.2) provides another solution.

Thus, there is no hope to get a unique solution of (1.2) in any reasonable model. On the other hand, for the practical applications, one definitely needs to choose a unique solution. So, one has to impose additional conditions on the unknown function φ that would guarantee the uniqueness of a solution (at least, in reasonable models) and would also allow for an efficient procedure of calculating the solution. 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 will consider the problem

$$\begin{cases} \text{Minimize } \mathsf{Var}\,\varphi(X_1, \dots, X_N) \text{ subject to} \\ \mathsf{E}[\varphi(X_1, \dots, X_N) \,|\, X_n] = \varphi_n(X_n), \quad n = 1, \dots, N, \end{cases} \tag{1.3}$$

where Var denotes the variance.

3. Solution. We first consider (1.3) 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 will 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 (1.3) 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 the Gaussian P. Then the solution of (1.3) is provided by

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

where H_m are Hermite polynomials and α_{nm} are found through solving certain N-dimensional linear systems (if in a practical calculation one cuts off the above summation in m at some $M \in \mathbb{N}$, then one 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. Practically, this is very useful as Gaussian copulas are probably the most popular class of models for the joint distribution of risk factors.

4. Structure of the paper. In Section 2, we describe some other interpretations of the problem under consideration. In Section 3, problem (1.3) is studied for a general P. Three particular situations described above are considered in Section 4. Section 5 deals with the extension for multidimensional factors. In Section 6, we provide an application of our results to decomposing the risk of a portfolio into three components. Section 7 concludes.

2 Problem Interpretations

As described above, the problem under consideration comes from measuring the risk of portfolios with scarce observations. It admits also other statistical and financial interpretations, which are described in this section. The reader interested only in the solution of the above problem can skip this section.

2.1 Linear and Non-Linear Regression

Let $R, X_1, \dots, X_N$ be square integrable random variables with mean zero. For the convenience of presentation, we also assume that $\mathsf{E}X_n^2 = 1$. The problem of linear regression of R on X_n is

$$\mathsf{E}(R - aX)^2 \longrightarrow \min, \quad a \in \mathbb{R}.$$

The solution is given by $a_n = \mathsf{E}RX_n$. The problem of linear regression of R on $X_1, \dots, X_N$ is

$$\mathsf{E}\Big(R - \sum_{n=1}^{N} \alpha_n X_n\Big)^2 \longrightarrow \min, \quad \alpha_1, \dots, \alpha_N \in \mathbb{R}.$$

The solution is given by

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} = C^{-1} \begin{pmatrix} \mathsf{E}RX_1 \\ \vdots \\ \mathsf{E}RX_N \end{pmatrix} = C^{-1} \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix},$$

where C is the covariance matrix of $X_1, \dots, X_N$. Thus, the solution of the multidimensional regression problem can be recovered from the solutions of the one-dimensional problems if $\mathrm{Law}(X_1, \dots, X_N)$ is known.

The non-linear regression of R on X_n is the problem

$$\mathsf{E}(R - \varphi(X_n))^2 \longrightarrow \min, \quad \varphi : \mathbb{R} \to \mathbb{R},$$

whose solution is $\varphi_n(x) = \mathsf{E}[R \,|\, X_n = x]$. The non-linear regression of R on $X_1, \dots, X_N$ is the problem

$$\mathsf{E}(R - \varphi(X_1, \dots, X_N))^2 \longrightarrow \min, \quad \varphi : \mathbb{R}^N \to \mathbb{R},$$

whose solution is $\varphi(x_1, \dots, x_N) = \mathsf{E}[R \,|\, X_1 = x_1, \dots, X_N = x_N]$. Thus, the problem we are considering in this paper might be seen as the problem of recovering the multidimensional non-linear regression function from one-dimensional regression functions.

2.2 Factor Risks

Along with the problem of measuring the risk of a portfolio, an important problem is to measure separately the risks brought by each of the economic factors affecting the portfolio's return. This is important to understand which factors affect the risk most and is also useful if large moves of some factors are expected.

The following notion of *factor risk* was introduced independently in industry by the RiskData company and in the literature by Cherny and Madan [1]. The aim of this notion, as opposed to the sensitivity coefficients, is to take into account the non-linear dependence of the portfolio's return on the values of the factors. Let ρ be a risk measure (for example, the standard deviation, Value at Risk, or a coherent risk measure), R be the return or the Profit&Loss of a portfolio over a unit time period, and X be the increment of a factor over this period. The factor risk of R brought by X is

$$\rho^f(R;X) := \rho(\mathsf{E}[R\,|\,X]) = \rho(\varphi(X)),$$

where $\varphi(x) = \mathsf{E}[R\,|\,X = x]$. The intuitive meaning of this definition is as follows: R is represented as $R^1 + R^2$, where $R^1 = \mathsf{E}[R\,|\,X]$ means the uncertainty in R brought by the factor X (recall that this is the projection of R on the space of X-measurable random variables), while $R^2 = R - R^1$ means the uncertainty in R brought by other factors.

In particular, if ρ is a coherent risk measure, i.e.

$$\rho(R) = -\inf_{Z\in\mathcal{D}} \mathsf{E}[ZR],$$

where $\mathcal{D}$ is some set of random variables with $Z \geq 0$, $\mathsf{E}Z = 1$, then

$$\rho^f(R;X) = -\inf_{Z\in\mathcal{D}} \mathsf{E}[Z\mathsf{E}[R\,|\,X]] = -\inf_{Z\in\mathcal{D}} \mathsf{E}[\mathsf{E}[Z\,|\,X]\mathsf{E}[R\,|\,X]] = -\inf_{Z\in\mathsf{E}[\mathcal{D}|X]} \mathsf{E}[ZR],$$

where $\mathsf{E}[\mathcal{D}\,|\,X_n] = \{\mathsf{E}[Z\,|\,X] : Z \in \mathcal{D}]\}$. Thus, the map $R \mapsto \rho^f(R;X)$ is again a coherent risk measure.

It might seem that if we take all the relevant factors into account, i.e. R is measurable with respect to $X_1, \dots, X_N$, then the overall risk is dominated (or at least "controlled") by the sum of the factor risks. However, this is not true if the factors are dependent, which is the case in practice. To illustrate this effect, consider a simple example. Let X_1, X_2 be Gaussian random variables with $\mathsf{E}X_n = 0$, $\mathsf{E}X_n^2 = 1$, and $\mathsf{corr}(X_1, X_2) = -1 + \varepsilon$, where ε is a small number. Consider $R = \varepsilon^{-1}(X_1 + X_2)$. As R is centered Gaussian, then, for any reasonable risk measure, $\rho(R)$ should be proportional to the standard deviation of R, which is of order $\varepsilon^{-1/2}$. On the other hand, $\rho^f(R;X_n)$ is proportional to the standard deviation of $\mathsf{E}[R \mid X_n]$, which is of order 1. In other words, the factor risks do not necessarily "control" the overall risk.

The problem we are considering in this paper might be seen as the problem of joining factor risks together in such a way that this procedure would provide a right assessment of the overall risk of the portfolio. □

2.3 Hedge Fund Replication

The problem of *hedge fund replication* consists in approximating the return R of a hedge fund by a function of factor increments $X_1, \dots, X_N$. A popular way to do that is to consider a linear function $\sum_n \alpha_n X_n$; see [2], [6]. However, as stressed in [3], [4], often the dependence of R on X_n is clearly non-linear.

The problem of this paper might be seen as finding the best non-linear approximation of R by $X_1, \dots, X_N$, through knowing the best non-linear approximations of R by each X_n.

3 General Setup

In this section, we will study problem (1.3) for an arbitrary measure P. Let us set

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

We will denote by $\|\cdot\|$ the L^2-norm and by Pr_E the orthogonal projection on a space E.

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

Lemma 3.1. *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*

$$\varphi(x_1, \dots, x_N) = \sum_{n=1}^{N} \psi_n(x_n) \tag{3.1}$$

satisfies (1.2). *Then it is the unique solution of* (1.3).

Proof. Let $\widetilde{\varphi}$ be a function satisfying (1.2). Denote

$$E_n = \{\xi \in L^2 : \xi \text{ is } X_n\text{-measurable}, \mathsf{E}\xi = 0\}, \tag{3.2}$$

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 $\|\widetilde{\varphi}\| \geq \|\varphi\|$, 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 3.2. *The following conditions are equivalent:*

(a) *For any $(\varphi_n) \in \Phi$, there exists a solution of* (1.2).

(b) (Lower ellipticity) *There exists a constant $c > 0$ such that, for any $(\varphi_n) \in \Phi$,*

$$\left\| \sum_{n=1}^{N} \varphi_n(X_n) \right\| \geq c \sum_{n=1}^{N} \|\varphi_n(X_n)\|.$$

If the above conditions are satisfied, then

(c) (Existence, uniqueness, and form of solution) *For any $(\varphi_n) \in \Phi$, there exists a unique solution of* (1.3)*, and it has the form* (3.1) *with some $(\psi_n) \in \Phi$.*

(d) (Linearity) *If φ (resp., φ') is a solution of* (1.3) *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* $(\varphi_n) \in \Phi$,

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

where φ *is the solution of* (1.3) *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. the fund of funds problem.

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

First, we will prove an auxiliary lemma.

Lemma 3.3. *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* $x_n \in H_n$,

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

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 will prove this statement by the 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

$$\delta_k = \begin{cases} -\mathrm{Pr}_{H_1} \delta_{k-1} & \text{if } k \text{ is odd,} \\ -\mathrm{Pr}_{H_2} \delta_{k-1} & \text{if } k \text{ is even.} \end{cases}$$

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 $\|u_n - \mathrm{Pr}_{H_1} u_n\|/(\|u_n\| + \|\mathrm{Pr}_{H_1} u_n\|) \to 0$, which is a contradiction). Clearly, we can choose $\gamma < 1$ such that we will 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 $\widetilde{H}_1 = H_1 + \dots + H_{N-1}$, $\widetilde{H}_2 = H_N$. Then the pair $(\widetilde{H}_1, \widetilde{H}_2)$ satisfies the conditions of the lemma. So, for the pair $\widetilde{x}_1 = x_1 + \dots + x_{N-1}$, $\widetilde{x}_2 = x_N$, 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 $\widetilde{x}_1 = y_1 + \dots + y_{N-1}$ 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 3.2. (b)$\Rightarrow$(a) This implication follows from the above lemma.

(a)$\Rightarrow$(c) Denote by E the L^2-closure of $E_1 + \dots + E_N$, where E_n are given by (3.2). Let φ be a solution of (1.2) corresponding to (φ_n). It is easy to see that the set of all solutions of (1.2) 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 (1.3) exists, is unique, and is given by $\mathrm{Pr}_E\, \varphi$.

We will prove the representation of the solution later.

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

(a)$\Rightarrow$(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 (3.2), endowed with the norm $\|(\xi_1, \dots, \xi_N)\| = \bigl(\sum_n \|\xi_n\|^2\bigr)^{1/2}$. Let E be the same as above. Then the map

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

is continuous (E is denoted with the L^2-norm), one-to-one (if $\xi, \xi' \in E$ have the same projections on each E_n, then $\xi - \xi'$ should 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)$\Rightarrow$(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 $(x_1, \dots, x_N)$ should 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

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

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 $A_{nm} = \langle \varphi_n(X_n), \varphi_m(X_m) \rangle$. Using condition (c), which follows from (a), we can write

$$\begin{aligned}\langle A^{-1}x, x\rangle^{1/2} = \langle y, Ay\rangle^{1/2} &= \Bigl\| \sum_{n=1}^N y_n \varphi_n(X_n) \Bigr\| \le \|\xi\| \\ &\le C \sum_{n=1}^N \|x_n \varphi_n(X_n)\| = C \sum_{n=1}^N |x_n| \le NC \langle x, x\rangle^{1/2},\end{aligned}$$

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 $(x_1, \dots, x_N)$

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

(a)⇒(c) It remains to prove that the solution has the form (3.1). For this, we note that (a) implies (b), which, in turn, implies that the sum $E_1 + \dots + E_N$ is L^2-closed (we are using the same notation as above). Hence, $E = E_1 + \dots + E_N$. As shown above, any solution of (1.3) belongs to E. The proof is completed. □

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

Example 3.4. **(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 (1.3).

(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 delta-mass concentrated at zero. Let $\varphi_1(x) = I(x = 0)$, $\varphi_2(x) = -I(x = 0)$. If there exists a solution φ of (1.2), 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

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

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 3.2 is not satisfied, which means that, for some $(\varphi_1, \varphi_2) \in \Phi$, there is no solution of (1.3) (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. □

4 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.

4.1 Independent Components

Suppose that $X_1, \dots, X_N$ are independent under P. It is clear from Lemma 3.1 that the solution of (1.3) is given by

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

4.2 Gaussian Distribution

Suppose that P is a Gaussian non-degenerate distribution. We will 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 into such a distribution.

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

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

As an example,

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= x, \\ H_2(x) &= (x^2 - 1)/\sqrt{2}, \\ H_3(x) &= (x^3 - 3x)/\sqrt{6}. \end{aligned}$$

Denote

$$a_{nm} = \mathsf{E}[\varphi_n(X_n) H_m(X_n)] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \varphi_n(x) H_m(x) e^{-x^2/2} dx, \quad n = 1, \dots, N, \; m \in \mathbb{N}. \tag{4.1}$$

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

$$\begin{pmatrix} \alpha_{1m} \\ \vdots \\ \alpha_{Nm} \end{pmatrix} = (C^m)^{-1} \begin{pmatrix} a_{1m} \\ \vdots \\ a_{Nm} \end{pmatrix} \tag{4.2}$$

is well defined.

Theorem 4.1. *The solution of* (1.3) *is given by the* L^2*-convergent series*

$$\varphi(x_1, \dots, x_N) = \sum_{n,m=1}^{N,\infty} \alpha_{nm} H_m(x_n), \quad x_n \in \mathbb{R}. \tag{4.3}$$

Proof. Let us prove the L^2-convergence of series (4.3). The matrices C^m, $m \in \mathbb{N}$ are non-degenerate by Lemma A.2. As they converge to the unit matrix, there exists $\mu > 0$ such that

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

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

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

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 (1.2). Consider the spaces E_n given by (3.2). Then $(H_m(X_n))_{m=1}^{\infty}$ forms an orthonormal basis in E_n. Employing Lemma A.1, we can write

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

An application of Lemma 3.1 completes the proof. □

As is well known, the Hermite polynomials are the result of orthogonalization of 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 4.2. (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* (1.3) *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* (1.3) *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* (1.3) *is given by* (4.3) *with the summation in* m *going up to* M.

Let us remark that the solution provided by Corollary 4.2 (i) has the same form as the function described in Subsection 2.1. In other words, if P is Gaussian and φ_n are linear, our methodology provides the linear regression estimate of R given $X_1, \dots, X_N$.

4.3 Gaussian Copula

Suppose that P is a non-degenerate Gaussian copula, i.e. there exists a non-degenerate Gaussian vector $(\widetilde{X}_1, \dots, \widetilde{X}_N)$ and increasing functions $f_n : \mathbb{R} \to \mathbb{R}$ such that $P = \mathrm{Law}(f_1(\widetilde{X}_1), \dots, f_N(\widetilde{X}_N))$. 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 $\widetilde{P} = \mathrm{Law}(\widetilde{X}_1, \dots, \widetilde{X}_N)$ 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

$$\varphi(x_1, \dots, x_N) = \widetilde{\varphi}(f_1^{-1}(x_1), \dots, f_N^{-1}(x_N)), \tag{4.4}$$

where f_n^{-1} is the right-continuous inverse of f_n, satisfies (1.2). As $\widetilde{\varphi}$ is the sum of functions of one variable, the same is true for φ. By Lemma 3.1, φ is the solution of (1.3).

To sum up, the procedure for solving (1.3) consists of the following steps:

1. Estimate from the data the distribution $P = \mathrm{Law}(X_1, \dots, X_N)$ (assumed here to be a Gaussian copula) and the conditional expectations $\varphi_n(x) = \mathsf{E}[R \,|\, X_n = x]$.
2. Find the functions f_n such that $P = \mathrm{Law}(f_1(\widetilde{X}_1), \dots, f_N(\widetilde{X}_N))$, 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 find the coefficients $(a_{nm}; n = 1, \dots, N,\ m = 1, \dots, M)$ given by (4.1) 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 linear systems (4.2) 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 (4.3) and find the function φ given by (4.4). This is the desired solution.

5 Multidimensional Factors

In this section, we consider problem (1.3) 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 some others, like different parts of the yield curve, are inherently multidimensional. Considerations of Section 3 as well as Subsections 4.1 and 4.3 admit a straightforward extension to multidimensional X_n. This is not the case for Subsection 4.2, 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, while each $\varphi_n : \mathbb{R}^{d_n} \to \mathbb{R}$ satisfies $\mathsf{E}\varphi_n(X_n) = 0$ and $\mathsf{E}\varphi_n^2(X_n) < \infty$.

5.1 Two Factors

We first consider the case $N = 2$. Without loss of generality, $d_1 \geq d_2$. 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 [5]), and procedures for finding U_1, U_2 are implemented in most of 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 nonzero multiindices of length d_n, i.e. M_n consists of collections $(m(1), \dots, m(d_N))$, where $m(i)$ take the values $0, 1, 2, \dots$ and at least one of $m(i)$ is greater than zero. For $m \in M_n$, we consider the corresponding multidimensional Hermite polynomial

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

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

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

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

$$\rho^m = (\rho^1)^{m(1)} \cdot \dots \cdot (\rho^{d_2})^{m(d_2)},$$

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 (1.3). 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 $|\rho^i| \leq \lambda$ for any $i = 1, \dots, d_2$. Then $|\rho^m| \leq \lambda$ for any $m \in M_2$, which implies that $|\alpha_{nm}| \leq (1 - \lambda^2)^{-1}(|a_{1m}| + |a_{2m}|)$ 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

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

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

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

For $m \in M_2$, we have

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

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 [7; Ch. II, § 2]), we see that $\mathrm{Pr}_{E_1} \widetilde{\varphi}(Z_1, Z_2) = \widetilde{\varphi}_1(Z_1)$. In other words, $\mathsf{E}[\widetilde{\varphi}(Z_1, Z_2) \,|\, Z_1] = \widetilde{\varphi}_1(Z_1)$, which means that $\mathsf{E}[\varphi(X_1, X_2) \,|\, X_1] = \varphi_1(X_1)$. In the same way, we prove that $\mathsf{E}[\varphi(X_1, X_2) \,|\, X_2] = \varphi_2(X_2)$. Now, the multidimensional analog of Lemma 3.1 guarantees that φ is the desired solution.

5.2 Multiple Factors

Consider now an arbitrary N. The solution of (1.3) will be constructed in a sequence of steps. Without loss of generality, $d_1 \ge \dots \ge d_N$.

Step 1. Applying the above described 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

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

Step 2. Applying the above described 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

$$\begin{aligned} &\mathsf{E}[\varphi_{123}(X_1, X_2, X_3) \,|\, X_1, X_2] = \varphi_{12}(X_1, X_2), \\ &\mathsf{E}[\varphi_{123}(X_1, X_2, X_3) \,|\, X_3] = \varphi_3(X_3). \end{aligned}$$

The first equality implies that

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

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

6 Three Types of Risk

The risk of a portfolio can be decomposed into a sum of systematic and idiosyncratic risk. The results of this paper allow us to decompose further the systematic risk into two components. This is the topic of the present section.

Let $R, X_1, \dots, X_N$ be random variables with $\mathsf{E}R^2 < \infty$. Denote

$$P = \mathrm{Law}(X_1, \dots, X_N),$$
$$\varphi_n(x) = \mathsf{E}[R \,|\, X_n = x],$$
$$\psi(x_1, \dots, x_N) = \mathsf{E}[R \,|\, X_1 = x_1, \dots, X_N = x_N].$$

Suppose that there exists a solution φ of (1.3) (clearly, it is unique up to P-indistinguishability). Consider the decomposition

$$R = R^1 + R^2 + R^3, \tag{6.1}$$

where

$$R^1 = \varphi(X_1, \dots, X_N),$$
$$R^2 = \psi(X_1, \dots, X_N) - \varphi(X_1, \dots, X_N),$$
$$R^3 = R - \psi(X_1, \dots, X_N).$$

Proposition 6.1. *The random variables R^1, R^2, R^3 are uncorrelated.*

Proof. Without loss of generality, $\mathsf{E}R = 0$. Let E_n be given by (3.2) and E denote the L^2-closure of $E_1 + \dots + E_N$. The random variable $\mathrm{Pr}_E\, R$ is $(X_1, \dots, X_N)$-measurable and hence, it can be written as $\widetilde{\varphi}(X_1, \dots, X_N)$. We have

$$\mathrm{Pr}_{E_n}\, \widetilde{\varphi}(X_1, \dots, X_N) = \mathrm{Pr}_{E_n}\, R = \varphi_n(X_n).$$

This means that $\widetilde{\varphi}$ satisfies (1.2). Furthermore, $\|\widetilde{\varphi}(X_1, \dots, X_N)\| \le \|\varphi(X_1, \dots, X_N)\|$, and the equality is possible only if $\widetilde{\varphi} = \varphi$ P-a.s. As a result,

$$\varphi(X_1, \dots, X_N) = \mathrm{Pr}_E\, R.$$

Consider now the space

$$L = \{\xi \in L^2 : \xi \text{ is } (X_1, \dots, X_N)\text{-measurable}, \mathsf{E}\xi = 0\}.$$

Then

$$\psi(X_1, \dots, X_N) = \mathrm{Pr}_L\, R,$$

so that

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

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

Decomposition (6.1) might be interpreted as the decomposition of the risk of R into three components. As seen from Theorem 3.2, the component R^1 is typically a sum of monomials, i.e. functions of single random variables X_n. We can say that it corresponds to the *monomial risk*. The component R^2 captures the non-linearity of ψ coming from non-monomial terms, such as $X_k X_n$. We might say that it corresponds to the *cross-term risk*. Finally, R^3 accounts for the the *idiosyncratic risk*. Thus, the above decomposition of R might be interpreted as:

$$\text{Risk} \;=\; \text{Monomial risk} \;+\; \text{Cross-term risk} \;+\; \text{Idiosyncratic risk}.$$

In these terms, the topic of this paper is recovering the monomial risk.

7 Conclusion

We consider the problem of estimating the risk of a portfolio with scarce observations, for example, the returns of a hedge fund. We assume that from the data one can derive the joint laws of the portfolio's return and each of the factors, as well as the joint law of all the factors. The problem under consideration is recovering the conditional mean of the return given all the factors. This problem also admits a statistical interpretation of recovering the non-linear multidimensional regression function from the one-dimensional regression functions. It also admits an interpretation of assessing the risk of a portfolio from one-dimensional non-linear factor risks.

This problem admits multiple solutions. In order to get a single one, we are looking for the "most moderate one" in the sense that it has the minimal variance. In this formulation, the solution would typically exist, be unique, and possess a number of nice features like linearity (i.e. the solution corresponding to the sum of portfolios is the sum of solutions) and continuity (stability under misspecification of the data). Moreover, it has the form of a sum of functions of single factors. This result admits an interpretation of decomposing the portfolio's risk in three components: the risk coming from sums of non-linear functions of single factors, the risk coming from the cross-terms, and the idiosyncratic risk.

We have provided an explicit form of the solution of the above problem for the practically important case when the joint law of factors is a Gaussian copula. This case is immediately reduced to the Gaussian one. In the latter case, we provide a solution, whose practical calculation consists in finding several integrals of one-dimensional functions and solving several N-dimensional linear systems, where N is the number of factors (the explicit procedure is described in Subsection 4.3). Thus, our solution is easily implementable and practically meaningful because the Gaussian copula is a very popular model for the joint distribution of factors in the modern risk measurement.

Appendix

For the 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

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

On the other hand,

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

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 positively 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] *A.S. Cherny, D.B. Madan.* Coherent measurement of factor risks. SSRN, 2006, www.ssrn.com.

[2] *W. Fung, D.A. Hsieh.* Empirical characteristics of dynamic trading strategies: the case of hedge funds. Review of Financial Studies, **10** (1997), p. 275–302.

[3] *W. Fung, D.A. Hsieh.* Survivorship bias and investment style in the returns of CTAs. Journal of Portfolio Management, **23** (1997), p. 30–41.

[4] *W. Fung, D.A. Hsieh.* The risk in hedge fund strategies: theory and evidence from trend followers. Review of Financial Studies, **14** (2001), p. 313–341.

[5] *G.H. Golub, C.F. Van Loan.* Matrix computations. 3rd Ed., Johns Hopkins University Press, Baltimore, MD, USA, 1996.

[6] *J. Hasanhodzic, A. Lo.* Can hedge-fund returns be replicated?: the linear case. Journal of Investment Management, **5** (2007), p. 5–45.

[7] *V.A. Malyshev, R.A. Minlos.* Gibbs random fields. Kluwer, 1991.