

Article

# Principal Curves for Statistical Divergences and an Application to Finance

Ana Flávia P. Rodrigues and Charles Casimiro Cavalcante \* 

Department of Teleinformatics Engineering, Federal University of Ceará, Fortaleza-CE 60440-900, Brazil; anaflaviapaivar@gmail.com

\* Correspondence: charles@ufc.br; Tel.: +55-85-3366-9470

Received: 22 February 2018; Accepted: 4 April 2018; Published: 2 May 2018



**Abstract:** This paper proposes a method for the beta pricing model under the consideration of non-Gaussian returns by means of a generalization of the mean-variance model and the use of principal curves to define a divergence model for the optimization of the pricing model. We rely on the  $q$ -exponential model so consider the properties of the divergences which are used to describe the statistical model and fully characterize the behavior of the assets. We derive the minimum divergence portfolio, which generalizes the Markowitz's (mean-divergence) approach and relying on the information geometrical aspects of the distributions the Capital Asset Pricing Model (CAPM) is then derived under the geometrical characterization of the distributions which model the data, all by the consideration of principal curves approach. We discuss the possibility of integration of our model into an adaptive procedure that can be used for the search of optimum points on finance applications.

**Keywords:** principal curves; information geometry; deformed exponential; finance application

## 1. Introduction

In their seminal paper [1], Hastie and Stuetzle proposed a notion of principal curves as an elegant and geometric non-linear generalization of factor models as the principal component analysis. A principal curve has the property of self-consistence in the sense that it passes through the middle of the data set representing a sample of some random variable. More precisely, any point of the curve coincides with the expected value of the data projected on it. This is a direct consequence of the fact that a principal curve  $f$  is critical for the variance of the Euclidean distance between the data and any locally defined perturbation of  $f$ . In particular, a straight line is a principal curve if and only if its direction is an eigenvector of the covariance matrix of  $z$ , where  $z$  stands for the a vector containing the observed data.

The original idea by Hastie and Stuetzle has been developed into relevant improvements, applications and extensions. We point out however that the criticality of a principal curve is usually defined in terms of the Euclidean distance. Hence, although  $f$  itself could represent non-Euclidean features of the model, some underlying least-squares approach is still in force. Our main contribution here is to rephrase the notion of principal curves (and, more generally, of principal  $p$ -submanifolds) in terms of a general statistical divergence which replaces the Euclidean divergence, that is, the variance used in the original definition.

Considering statistical divergences as Kullback–Leibler or Bregman divergence allows us to deal with random variables with probabilities given by exponential and deformed exponential distributions. In the context of exponential and  $\phi$ -exponential statistical families, straight lines are replaced by affine geodesics and the Hessian of the cumulant function plays the role of a generalized covariance.

As highlighted by Naudts, deformed exponentials play a central role in the foundations of the Generalized Thermostatistics formulated by Tsallis [2,3] and collaborators. This new approach to

Thermodynamics has been evolved along the last two decades in a wide range of applications to complex systems, particularly in Finance [4–7]. Indeed, Naudts' work established deep and fruitful connections between Statistical Physics and Information Geometry [8–11]. For instance, both Rényi's and Tsallis' entropies are described by Naudts in terms of statistical divergences in the family of  $q$ -exponential distributions that includes  $q$ -Gaussian distributions, defined in details by Plastino and Vignat [11–15]. The analytic and geometric features of deformed exponentials suggest that they are well suited to model non-normally distributed returns of contingent claims. In this direction, for instance, a non-Gaussian option pricing theory has been successfully proposed in terms of diffusion processes associated to  $q$ -Gaussian distributions [4,7,16–18]. Other related developments are summarized in [6,19].

In [20,21], the authors elaborated some preliminary results towards a theory of portfolio optimization in the context of deformed exponentials. One of the cornerstones of the modern Finance Theory, the classical Markowitz's mean-variance model of portfolio selection, relies on the assumptions that the returns of assets are normally distributed and that the investor preferences are described by constant risk-aversion utility function. Some works have dealt with other variations of the portfolio optimization problem. For instance, Zagrodny [22] proposed a convex programming solution to portfolio selection by considering the Hilbert space and a reinsurance approach. In the multi-objective model, Sawik [23] used the expected return as the performance metric and expected worst-case return to measure the risk and providing good interpretation about the consideration of the variance of the risks.

The traditional criticism to the normality assumption in Markowitz's theory raises the need of alternative models for dealing with non-Gaussian distributions. This question has been addressed since then under different methods. In [24,25], Nock et al. extended the Markowitz's model to the wider family of exponential distributions, replacing the mean-variance by a mean-divergence model. Bregman divergences replace the variance as risk measures for non-Gaussian distributions, eventually encompassing information from higher order momenta. On the other hand, since statistical divergences define geometric notions on the statistical manifold of exponential distributions, their method has a geometric interpretation in terms of a steepest descent by the natural gradient of the risk premium [26–28].

In [20], the authors proposed a model of portfolio selection of financial assets that explores the non-additivity and non-normality aspects of Tsallis' Thermostatistics. More precisely, they have extended the mean-divergence model in [24,25] to deformed exponentials families.

In the sequel, the authors formulated [21] a generalization of beta pricing models adapted to a mean-divergence portfolio selection [29–31]. In particular, it is presented an extension of Capital Asset Pricing Model (CAPM) flexible enough to be applied for financial returns with deformed exponential distributions. The method relies on a geometric approach to the classical mean-variance analysis developed by LeRoy and Werner [32] and Luenberger [33] (see also [34]). The main results in [20,21] are summarized in Section 3.

This paper is structured as follows. In Section 2, we define the generalized notion of principal curve and principal submanifold in the geometric context of a given statistical divergence. The earlier contributions for portfolio selection and asset pricing in the case of financial returns distributed according deformed exponential probability densities are schematically resumed in Section 3. In Section 4, we apply the generalized notion of principal submanifolds and the correspondent version of the principal component analysis to obtain an explicit expression of optimal principal portfolios is provided in Section 5.

## 2. Statistical Divergences and Principal Curves

Let  $\mathcal{M}$  be a space of random variables  $z = z(s)$  whose probability distributions are given by densities lying in a  $n$ -dimensional statistical manifold

$$\mathcal{S} = \{p(s, \boldsymbol{\vartheta}) : \boldsymbol{\vartheta} \in U \subset \mathbb{R}^n\},$$

where  $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_n)$  are statistical parameters ranging in some open subset  $U$  of the  $n$ -dimensional Euclidean space  $\mathbb{R}^n$ . Let  $D$  be a given statistical divergence in  $\mathcal{S}$ . Given a curve  $\mathbf{f} : \Lambda \subset \mathbb{R} \rightarrow \mathcal{M}$ , a projection of  $z$  on the trace of  $\mathbf{f}$  is a point  $\mathbf{f}(\lambda_*)$ , for some  $\lambda_* \in \Lambda$ , such that

$$D(z|\mathbf{f}(\lambda_*)) = \inf_{\lambda \in \Lambda} D(z|\mathbf{f}(\lambda)). \quad (1)$$

In the following, we suppose that such a projection exists and it is unique for any curve  $\mathbf{f} : \Lambda \rightarrow \mathcal{M}$  we are going to consider. Under this assumption, we denote

$$\pi_{\mathbf{f}}(z) = \mathbf{f}(\lambda_*). \quad (2)$$

In this notation, we propose the following variational notion of principal curve relative to  $D$ :

**Definition 1.** A curve  $\mathbf{f} : \Lambda \rightarrow \mathcal{M}$  is a principal curve in  $(\mathcal{M}, D)$  if

$$D(z|\pi_{\mathbf{f}}(z)) = \inf_{s \in (-\epsilon, \epsilon)} D(z|\pi_{\mathbf{f}_s}(z)) \quad (3)$$

for all one-parameter family of curves  $\mathbf{f}_s : \Lambda \rightarrow \mathcal{M}$ ,  $s \in (-\epsilon, \epsilon)$ , such that  $\mathbf{f}_0 = \mathbf{f}$ .

Recall that a statistical divergence  $D$  determines a dually flat structure in  $\mathcal{M}$  for which affine geodesics are parameterized as straight lines of the form

$$\mathbf{f}(\lambda) = \mathbf{a} + \lambda \mathbf{u}, \quad \lambda \in \Lambda,$$

where  $\mathbf{a}$  and  $\mathbf{u}$  are constant vectors. By definition, the projection of the random variable  $z$  on a principal curve  $\mathbf{f}$  minimizes the divergence among the projections on curves close to  $\mathbf{f}$ . Projections satisfy a Pythagorean theorem, one of the fundamental results in Information Geometry that can be stated as follows.

**Theorem 1** (Theorem 1.2 and Theorem 1.3, [26]). Given  $o, z, w \in \mathcal{M}$  such that the dual affine geodesic connecting  $z$  and  $w$  is orthogonal to the affine geodesic connecting  $w$  and  $o$ , the following generalized Pythagorean relation holds

$$D(z|o) = D(w|o) + D(z|w). \quad (4)$$

Similarly, if the affine geodesic connecting  $z$  and  $w$  is orthogonal to the dual affine geodesic connecting  $w$  and  $o$ , we have the dual relation

$$D^*(z|o) = D^*(w|o) + D^*(z|w), \quad (5)$$

where  $D^*$  is the dual divergence. The dual divergence is defined with respect to the dual connection as defined in [26].

In view of this proposition, it is natural to draw our attention to one-parameter families of affine geodesics in  $\mathcal{M}$ .

**Theorem 2.** An affine geodesic in  $(\mathcal{M}, D)$  is a principal curve with respect to one-parameter families of affine geodesics if and only if its direction is an eigenvector of the Fisher metric  $\mathcal{G} = \nabla^2 D$  associated to  $D$ .

**Proof.** Denote by  $\nabla D$  and  $\nabla^2 D$ , respectively, the differential and Hessian of  $D$  with respect to the second variable. Hence, we have for a fixed  $s \in (-\epsilon, \epsilon)$  that

$$0 = \frac{d}{d\lambda} D(z|\mathbf{f}(s, \lambda)) = \frac{d}{d\lambda} D(z|\mathbf{c}(s)) = \left\langle \nabla D(z|\mathbf{c}(s)), \frac{\partial \mathbf{f}}{\partial \lambda} \right\rangle$$

where the derivative is computed at the critical value  $\lambda = \lambda_*(s)$  and for a fixed value of  $s$ ,  $\mathbf{f}_s$  is a geodesic parameterized by  $\lambda$  which we can write  $\mathbf{f}_s(\lambda) = \mathbf{f}(s, \lambda)$ .

Denote

$$\mathbf{c}(s) = \mathbf{f}(s, \lambda_*(s)), \quad s \in (-\epsilon, \epsilon).$$

Note that

$$\mathbf{c}'(s) = \frac{\partial \mathbf{f}}{\partial s}(s, \lambda_*(s)) + \frac{\partial \mathbf{f}}{\partial \lambda}(s, \lambda_*(s)) \frac{d\lambda_*}{ds}.$$

Hence, we obtain

$$\begin{aligned} \frac{d}{ds} \left\langle \nabla D(z|\mathbf{c}(s)), \frac{\partial \mathbf{f}}{\partial \lambda}(s, \lambda_*(s)) \right\rangle &= \left\langle \nabla D(z|\mathbf{c}(s)), \frac{d}{ds} \frac{\partial \mathbf{f}}{\partial \lambda}(s, \lambda_*(s)) \right\rangle \\ &+ \mathbf{c}'(s)^\top \nabla^2 D(z|\mathbf{c}(s)) \frac{\partial \mathbf{f}}{\partial \lambda}(s, \lambda_*(s)) = \left\langle \nabla D(z|\mathbf{c}(s)), \frac{d}{ds} \frac{\partial \mathbf{f}}{\partial \lambda}(s, \lambda_*(s)) \right\rangle \\ &+ \left( \frac{\partial \mathbf{f}}{\partial s}(s, \lambda_*(s)) + \frac{\partial \mathbf{f}}{\partial \lambda}(s, \lambda_*(s)) \frac{d\lambda_*}{ds} \right)^\top \nabla^2 D(z|\mathbf{c}(s)) \frac{\partial \mathbf{f}}{\partial \lambda}(s, \lambda_*(s)). \end{aligned}$$

We may write

$$\mathbf{f}(s, \lambda) = \mathbf{f}(\lambda) + s\mathbf{v}(\lambda) + O(s^2),$$

where

$$\mathbf{v}(\lambda) = \frac{\partial \mathbf{f}}{\partial s}(0, \lambda)$$

is the variational field that corresponds to  $\mathbf{f}$ . Thus, we have

$$\frac{\partial \mathbf{f}}{\partial \lambda}(s, \lambda_*(s)) = \mathbf{f}'(\lambda_*(s)) + s\mathbf{v}'(\lambda_*(s)) + O(s^2)$$

and

$$\frac{d}{ds} \Big|_{s=0} \frac{\partial \mathbf{f}}{\partial \lambda}(s, \lambda_*(s)) = \mathbf{f}''(\lambda_*) \frac{d\lambda_*}{ds} \Big|_{s=0} + \mathbf{v}'(\lambda_*(s)).$$

If  $\mathbf{f} = \mathbf{f}(0, \cdot)$  is a critical curve, we have

$$\begin{aligned} 0 &= \frac{d}{ds} \Big|_{s=0} \left\langle \nabla D(z|\mathbf{c}(s)), \frac{\partial \mathbf{f}}{\partial \lambda}(s, \lambda_*(s)) \right\rangle = \left\langle \nabla D(z|\mathbf{f}(\lambda_*)), \frac{d}{ds} \frac{\partial \mathbf{f}}{\partial \lambda}(0, \lambda_*) \right\rangle \\ &+ \left( \mathbf{v}(\lambda_*) + \frac{d\lambda_*}{ds} \Big|_{s=0} \mathbf{f}'(\lambda_*) \right)^\top \nabla^2 D(z|\mathbf{f}(\lambda_*)) \mathbf{f}'(\lambda_*). \end{aligned}$$

We conclude that

$$\begin{aligned} &\left( \mathbf{v}(\lambda_*) + \frac{d\lambda_*}{ds} \Big|_{s=0} \mathbf{f}'(\lambda_*) \right)^\top \nabla^2 D(z|\mathbf{f}(\lambda_*)) \mathbf{f}'(\lambda_*) \\ &+ \left\langle \nabla D(z|\mathbf{f}(\lambda_*)), \mathbf{f}''(\lambda_*) \frac{d\lambda_*}{ds} \Big|_{s=0} + \mathbf{v}'(\lambda_*(s)) \right\rangle = 0. \end{aligned}$$

Setting  $\mathbf{f}''(\lambda) = 0$  and  $\mathbf{v}'(\lambda) = 0$ , one gets

$$\left( \mathbf{v}(\lambda_*) + \frac{d\lambda_*}{ds} \Big|_{s=0} \mathbf{f}'(\lambda_*) \right)^\top \nabla^2 D(z|\mathbf{f}(\lambda_*)) \mathbf{f}'(\lambda_*) = 0.$$

Since  $\mathbf{v}(\lambda_*)$  can be arbitrarily chosen in such a way that  $\mathbf{v}(\lambda_*)$  and  $\mathbf{f}'(\lambda_*)$  are linearly independent, we conclude there exists  $\mu \in \mathbb{R}$  such that

$$\nabla^2 D(z|\mathbf{f}(\lambda_*)) \mathbf{f}'(\lambda_*) = \mu \mathbf{f}'(\lambda_*). \tag{6}$$

This means that  $\mathbf{f}'(\lambda_*)$  is an eigenvector of the Fisher information metric at the point  $\mathbf{f}(\lambda_*)$

$$\mathcal{G}_{\mathbf{f}(\lambda_*)} = \nabla^2 D(z|\mathbf{f}(\lambda_*)) \tag{7}$$

associated to the divergence  $D$ . This finishes the proof.  $\square$

A result concerning principal submanifolds similar to Theorem 2 follows easily as a scholia of its proof: we may consider the projection of the random variable  $z$  onto a  $p$ -dimensional affine submanifold in  $\mathcal{M}$  parameterized by a smooth map  $\mathbf{f} : \Lambda \subset \mathbb{R}^p \rightarrow \mathcal{M}$  whose differential has rank  $p$ . The submanifold  $\mathbf{f}(\Lambda)$  is principal with respect to families of affine submanifolds if and only if it is spanned by  $p$  geodesics whose velocities are linearly independent and are eigenvectors of the Hessian matrix  $\mathcal{G} = \nabla^2 D$  at the projection point.

A fundamental example of divergence is the Euclidean  $L^2$ -norm

$$D_{\text{euc}}(z|w) = \frac{1}{2} |z - w|^2, \quad z, w \in \mathcal{M}$$

on which is based both the least-squares method and the principal component analysis. In their seminal work [1], Hastie and Stuetzle proved that a Euclidean straight line is a principal curve with respect to their definition if and only if its direction is an eigenvector of the covariance matrix of the random variable  $z$ .

Now, we obtain an extension of this result by Hastie and Stuetzle valid in the context of non-Euclidean statistical divergences. In our setting, the role of the covariance matrix is played by its non-Euclidean and non-Gaussian counterpart, namely the Hessian matrix  $\nabla^2 K$ , where  $K$  is the cumulant generating function.

**Corollary 1.** *Let  $K$  be a convex function in  $\mathcal{S}$  and  $D$  be the Bregman divergence in  $\mathcal{M}$  determined by  $K$ . Then an affine geodesic is a principal curve with respect to one-parameter families of affine geodesics in  $\mathcal{M}$  if and only if its direction is an eigenvector of the Hessian of  $K$ .*

**Proof.** This follows directly from Theorem 2 once we have observed that the Fisher metric in this case coincides with the Hessian of  $K$ . This is however a well-known fact that may be deduced easily from the definition of the Bregman divergence itself as

$$D(z|w) = K(z) - K(w) - \langle \nabla K(w), z - w \rangle. \tag{8}$$

For details, we refer the reader to [26].  $\square$

In the sequel, we are going to consider more general examples, not necessarily quadratic. For instance, we may fix the Kullback–Leibler divergence

$$D_{KL}(z(\cdot, \boldsymbol{\theta})|w(\cdot, \boldsymbol{\theta}')) = \int p(s, \boldsymbol{\theta}) \log \left( \frac{p(s, \boldsymbol{\theta})}{p(s, \boldsymbol{\theta}')} \right) ds$$

or, more generally, the relative  $\phi$ -entropy associated to  $\phi$ -exponential distributions.

### 3. The Space of Financial Assets

From now on,  $\mathcal{M}$  stands for the linear span of financial assets traded in a securities market. More precisely, every point in  $\mathcal{M}$  corresponds to the payoff  $z$  of a contingent claim at a fixed time, say  $t = 1$ , a random variable

$$z = z(s),$$

where  $s$  are the states of the world with probability distribution specified by some density  $p(s; \boldsymbol{\theta})$ . Recall that  $\boldsymbol{\theta}$  is the distribution parameter of a family of probability distributions in a  $n$ -dimensional statistical manifold  $\mathcal{S}$ .

In the following, we will consider the statistical manifold of  $\phi$ -exponential probability densities

$$p(s, \boldsymbol{\theta}) = \exp_{\phi}(\langle T(s), \boldsymbol{\theta} \rangle - K(s, \boldsymbol{\theta})) p_0(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \mathbb{R}^n,$$

where  $T$  is a sufficient statistics of the random variable  $z(s)$  and  $K$  is the moment generating function. Here,  $p_0$  is a fixed reference density and  $\exp_{\phi}$  is the  $\phi$ -exponential defined as the inverse function of the  $\phi$ -logarithm [8,9]

$$\log_{\phi}(t) = \int_1^t \frac{1}{\phi(s)} ds,$$

where  $\phi : (0, +\infty) \rightarrow (0, +\infty)$  is a strictly positive, nondecreasing and continuous real function. A particular case of this deformed exponential is given by the  $q$ -exponential function

$$\exp_q(t) = (1 + (1 - q)t)^{\frac{1}{1-q}}$$

with  $q > 0$ , which corresponds to set  $\phi(t) = t^q$ . Hence, the  $q$ -logarithm is defined by

$$\log_q(t) = \int_1^t \frac{1}{s} ds = \frac{1}{1-q}(t^{1-q} - 1).$$

The moment generating function  $K$  defines a Bregman divergence given by

$$D(z|w) = K(z) - K(w) - \langle \nabla K(z), z - w \rangle.$$

where the probability distributions of  $z(s)$  and  $w(s)$  are, respectively, given by the densities  $p(s, \boldsymbol{\theta})$  and  $p(s, \boldsymbol{\theta}')$ .

#### 3.1. Deformed Exponentials and Portfolio Selection

Setting  $\phi(t) = t$ , one gets the family of exponential distributions, in particular multivariate Gaussian distributions. For this family, Nock et al. [24,25] represented the key concepts of Portfolio Selection theory in terms of the moment generating function and the associated Bregman divergence. More precisely, they proved that, for constant absolute risk aversion (CARA) utility functions, the certainty equivalent and risk premium of risky assets are, respectively, given by

$$C = \frac{1}{a}(K(z(\cdot, \boldsymbol{\theta})) - K(w(\cdot, \boldsymbol{\theta}')))$$

and

$$\Pi = \frac{1}{a}D[z(\cdot, \boldsymbol{\theta})|w(\cdot, \boldsymbol{\theta}')],$$

where  $a > 0$  is a risk-aversion parameter. Hence, they extended the classical mean-variance portfolio selection to a general mean-divergence model for which an optimal allocation  $\boldsymbol{\alpha}$  is a solution of the minimization problem

$$\min_{\boldsymbol{\alpha}} \left( \langle \nabla K(\boldsymbol{\theta} - a\boldsymbol{\alpha}), \boldsymbol{\alpha} \rangle + \frac{1}{a}D_{\phi}(\boldsymbol{\theta} | (\boldsymbol{\theta} - a\boldsymbol{\alpha})) \right).$$

In the particular case of Gaussian distributed returns, they easily recover the classical Markowitz's optimal portfolio allocation vector

$$\alpha = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^\top \Sigma^{-1}\mathbf{1}},$$

where  $\Sigma$  is the variance-covariance matrix of the returns on the assets.

In [20], the authors extended this approach to  $\phi$ -exponential distributions, in particular to  $q$ -exponential distributions. They proved that the optimal portfolio for their extended mean-divergence model is given in terms of the cumulant function by

$$\alpha = \frac{\nabla^2 K(\boldsymbol{\theta})^{-1}\mathbf{1}}{\mathbf{1}^\top \nabla^2 K(\boldsymbol{\theta})^{-1}\mathbf{1}}. \quad (9)$$

Note that the Hessian of the (convex) function  $K$  is positive-definite and plays the role of the variance-covariance matrix in the Gaussian case. In the particular case of  $q$ -Gaussian distributions [14], the optimal allocation portfolio is given by

$$\alpha = \frac{\Sigma_q^{-1}\mathbf{1}}{\mathbf{1}^\top \Sigma_q^{-1}\mathbf{1}} \quad (10)$$

where

$$\Sigma_q = \gamma_q C_{q,n}^{1-q} |\Sigma|^{\frac{1-q}{2}} \Sigma \quad (11)$$

with

$$\gamma_q = \frac{1}{2}((n+4) - (n+2)q) \quad (12)$$

and

$$C_{q,n} = \begin{cases} \frac{\Gamma(\frac{1}{q-1} - \frac{n}{2})\sqrt{\pi}}{\Gamma(\frac{1}{q-1})} \left(\frac{1}{q-1}\right)^{\frac{n}{2}} ((n+4) - (n+2)q)^{\frac{n}{2}}, & \text{for } 1 < q < \frac{n+4}{n+2}, \\ \frac{\Gamma(\frac{2-q}{1-q})\sqrt{\pi}}{\Gamma(\frac{2-q}{q-1} + \frac{n}{2})} \left(\frac{1}{1-q}\right)^{\frac{n}{2}} ((n+4) - (n+2)q)^{\frac{n}{2}}, & \text{for } q < 1. \end{cases} \quad (13)$$

Here,  $|\Sigma|$  is the determinant of  $\Sigma$ . We refer the reader to [14] for further details in  $q$ -multivariate Gaussian distributions. It is evident that one re-obtains the Markowitz's portfolio for  $q = 1$  in Equation (10).

In view of Equation (9), the authors have elaborated in [20] a steepest descent algorithm by the natural (Riemannian) gradient of the risk premium. Some empirical support to the proposed method is provided by comparing the cumulated returns and the evolution of the divergence for optimal portfolios according to the mean-divergence model and the classical one by Markowitz. The numerical evaluations in [20] show the proposal is able to yield better tracking of deep changes in the stock market, such as the ones present in crisis scenarios, and yet produce a higher return than the classical mean-variance strategy.

### 3.2. Mean-Divergence Efficient Frontier

In Markowitz's model, the optimal portfolio allocation lies in the mean-variance efficient frontier that bounds the feasible set of allowed returns and risks of traded risky portfolios. In [32,33], LeRoy, Werner and Luenberger have developed a geometric approach to the mean-variance analysis in terms of the geometry of orthogonal projections onto a mean-variance efficient frontier. From this approach, they easily deduce an elegant geometric interpretation of the celebrated Capital Asset Pricing Method (CAPM) as well as other factor pricing models.

In [21], the authors have extended the geometric pricing method to general divergence geometries in  $\mathcal{M}$  instead of the Hilbert space  $L^2$ -norm.

Since  $K$  is a strictly convex function, its Hessian is positive-definite and then defines a Riemannian metric in  $\mathcal{M}$ , that is, for each  $z \in \mathcal{M}$ , we define an inner product in the tangent space  $T_z\mathcal{M}$  by

$$\mathcal{G}_z = \nabla^2 K(z). \tag{14}$$

This metric can be expanded in local coordinates around a fixed reference point  $o \in \mathcal{M}$  as

$$\mathcal{G}_z \sim \nabla^2 K(o) + o(|z|^2), \tag{15}$$

where quadratic terms are determined in terms of the Riemann curvature of the Riemannian manifold  $(\mathcal{M}, g)$ , see [35].

Denote by  $k_e$  the expectation kernel, that is, an asset in  $\mathcal{M}$  that yields the expected payoffs of the assets in  $\mathcal{M}$ . More precisely,

$$\mathcal{G}_z(k_e, z) = \mathbb{E}[z]$$

for any  $z \in \mathcal{M}$ . We define the pricing kernel  $k_q$  as an asset in  $\mathcal{M}$  that gives the price of any contingent claim  $z \in \mathcal{M}$  as the expected discounted payoff

$$g(k_q, z) = \mathbb{E}[mz] = Q(z),$$

where  $m$  is a stochastic discount factor. Here,  $Q : \mathcal{M} \rightarrow \mathbb{R}$  is the price functional, that is, the present value of the expected returns of the asset, discounted at rate  $m$ . The existence of this functional is one of the consequences of the Fundamental Theorem of Finance Theory whose key assumption is that there are no arbitrage portfolios in  $\mathcal{M}$ . For a comprehensive treatment of those fundamentals on Finance, we refer the reader to [32,36].

Denote by  $\mathcal{E}$  the subspace in  $\mathcal{M}$  spanned by  $k_e$  and  $k_q$ . The projection  $z^\mathcal{E}$  of  $z \in \mathcal{M}$  onto  $\mathcal{E}$  is defined by

$$D(z^\mathcal{E}|z) = \min_{w \in \mathcal{E}} D(w|z).$$

It follows from the generalized Pythagorean Theorem for divergences (Theorem 1) that, fixing a reference point  $o \in \mathcal{M}$ , one has

$$D(z|o) = D(z^\mathcal{E}|o) + D(z|z^\mathcal{E}), \quad z \in \mathcal{M}'. \tag{16}$$

If the case of the divergence given by the Euclidean  $L^2$ -norm in  $\mathcal{M}$

$$D_{\text{euc}}(z|w) = \frac{1}{2}|z - w|^2$$

Equation (16) reduces to the Euclidean decomposition

$$\mathbb{E}[|z|^2] = \mathbb{E}[z]^2 + \text{var}[z], \tag{17}$$

where

$$\text{var}[z] = \mathbb{E}[(z - \mathbb{E}[z])^2]$$

is the variance, the classical risk measure in Portfolio Theory [36,37].

Motivated by the analogy between Equations (16) and (17), the authors proposed in [21] the projection

$$\Pi(z) = D(z|z^\mathcal{E})$$

as a novel risk measure for assets  $z \in \mathcal{M}$ . Since it depends on the whole information about the probability densities  $p(s, \theta)$ , this measure encodes higher moments of  $z$  instead of only the variance. Moreover, one easily verifies that  $\Pi$  is the variance in the case of normally distributed returns



and Euclidean divergence. Hence, we have defined a risk measure that embodies non-normality and non-Euclidean features of the returns of financial assets and the estimation of their statistical parameters, respectively.

The main result in [21] is that the two reference assets  $k_e$  and  $k_q$  determine the efficient frontier for portfolios of assets in  $\mathcal{M}$  with respect to the risk measure  $\Pi$ . Indeed, we have the following theorem.

**Theorem 3** (Theorem 2 in [21]). *Let  $\mathcal{E} = \text{span}\{k_e, k_q\}$  the subspace in  $\mathcal{M}$  spanned by the expectation and pricing kernels. Given  $z \in \mathcal{M}$ , we have*

$$\mathbb{E}[z] = \mathbb{E}[z^{\mathcal{E}}]$$

and

$$\Pi(z^{\mathcal{E}}) \leq \Pi(z)$$

where  $z^{\mathcal{E}}$  is the projection of  $z$  onto  $\mathcal{E}$ .

Since the efficient frontier is spanned by two assets, this last result can be regarded as a non-Gaussian and non-Euclidean version of the two-fund spanning theorem in Finance. Generalizing the mean-variance case, we can prove in the case of  $\phi$ -exponentials that the efficient mean-divergence frontier for portfolio selection is spanned by two portfolios

$$\alpha_{\mathbf{1}} = \frac{(\nabla^2 K)^{-1} \mathbf{1}}{\mathbf{1}^\top (\nabla^2 K)^{-1} \mathbf{1}} \quad \text{and} \quad \alpha_{\mu} = \frac{(\nabla^2 K)^{-1} \mu}{\mathbf{1}^\top (\nabla^2 K)^{-1} \mu},$$

where  $\mu$  is the desired expected return of the portfolio.

#### 4. Generalized Beta Pricing Models and CAPM

Denote by  $R_e$  and  $R_q$  the returns of  $k_e$  and  $k_q$ , respectively. In [21], the authors have proved that the minimum divergence portfolio in  $\mathcal{M}$  is given by

$$z = R_e + (1 - \beta)(R_q - R_e)$$

where

$$\beta = -\frac{g(R_q - R_e, R_e)}{g(R_q - R_e, R_q - R_e)}.$$

A similar expression holds replacing the basic assets  $k_e$  and  $k_q$  by two efficient assets  $k_\lambda$  and  $k_\nu$  in  $\mathcal{E}$  such that

$$\mathcal{G}(r_\lambda, r_\nu) = 0. \tag{18}$$

These zero-covariance pair of assets is given by

$$r_\lambda = R_e + \lambda(R_q - R_e)$$

and

$$r_\nu = R_e + \nu(R_q - R_e)$$

where  $\nu$  is given by

$$\nu = -\frac{\mathcal{G}(R_e, R_e) + \lambda \mathcal{G}(R_q - R_e, R_e)}{\mathcal{G}(R_q - R_e, R_e) + \lambda \mathcal{G}(R_q - R_e, R_q - R_e)}. \tag{19}$$

Note that  $\nu$  is well-defined if and only if  $k_\lambda$  is not the minimum divergence portfolio in  $\mathcal{E}$ .

We have obtained in [21] a generalized beta pricing equation involving  $k_\lambda$  and  $k_\nu$

$$\mathbb{E}[z] = \mathbb{E}[r_\nu] + \beta(\mathbb{E}[r_\lambda] - \mathbb{E}[r_\nu]) \tag{20}$$

for assets in  $z \in \mathcal{M}$ , where the generalized beta coefficient is given by

$$\beta = \frac{\mathcal{G}(r, r_\lambda)}{\mathcal{G}(r_\lambda, r_\lambda)}. \tag{21}$$

If there exists a risk-free asset **1** with return  $R_f$  in  $\mathcal{M}$ , we fix  $r_\nu = \mathbf{1}$  reducing Equation (20) to

$$\mathbb{E}[z] = R_f + \beta(\mathbb{E}[r_\lambda] - R_f). \tag{22}$$

As in the classical CAPM, we can take  $r_\lambda$  as the market return  $r_m$  since it is possible to prove under some assumptions that  $r_m$  is in the mean-divergence efficient frontier. More precisely, this is the case when every agent in the market has consumption preferences given by a time-separable utility function of the form

$$u(c_0, c_1) = u_0(c_0) + u_1(\mathbb{E}[c_1], \mathcal{G}_{|c_1}(c_1, c_1)) \tag{23}$$

where  $u_1$  is strictly decreasing with respect to the second variable. Here,  $c_0$  is the agent’s consumption plan at time  $t = 0$  and  $c_1 = c_1(s)$  is a random variable in  $\mathcal{M}$  that describes the consumption plan of the agent at time  $t = 1$ .

Under this assumption, we obtained in [21] a generalized CAPM equation

$$\mathbb{E}[r] - R_f = \tilde{\beta}(\mathbb{E}[r_m] - R_f), \tag{24}$$

where  $r_m$  is the return of the market portfolio and

$$\tilde{\beta} = \frac{\mathcal{G}(r, r_m)}{\mathcal{G}(r_m, r_m)} \tag{25}$$

is the generalized beta market. This coefficient measures the generalized covariance between the risk of the asset or portfolio and the market risk. Note that both Equations (20) and (24) define a generalized security market line [32,38].

The Fisher information metric  $\mathcal{G}$  plays the role here of the covariance matrix. In the particular case when the returns of traded assets are distributed according to a  $q$ -Gaussian distribution, it holds that

$$\mathcal{G}_z = \nabla^2 K(z) = \Sigma_q$$

for every  $z \in \mathcal{M}$ , where the  $q$ -variance matrix  $\Sigma_q$  is defined in Section 3.1.

### 5. Generalized Principal Components Analysis (PCA) and Applications to Finance

The results we have quoted in Sections 3.2 and 4 indicate that the Hessian information matrix

$$\mathcal{G} = \nabla^2 D \tag{26}$$

plays a central role in the extension of portfolio selection and asset pricing models in the case of non-Gaussian returns. Even under the assumption of normality of the asset returns,  $\mathcal{G}$  can provide a more accurate risk measure since it is sensitive to higher moments of the underlying probability distributions.

A portfolio composed by  $N$  risky assets  $z_1, \dots, z_N$  in  $\mathbb{M}$  is determined by an allocation vector  $\alpha = (\alpha_1, \dots, \alpha_N) \in \mathbb{R}^N$

$$\alpha D, \tag{27}$$

where  $D$  is the vector of payoffs  $(z_1, \dots, z_N)^\top$ . We assume that the payoffs have probability distributions given by densities  $p(s, \vartheta_i) \in \mathcal{S}, i = 1, \dots, N$ . The expected return of this portfolio is

$$\mu = \alpha \mathbb{E}[D] = \sum_{i=1}^N \alpha_i \mathbb{E}[z_i]$$

whereas its generalized covariance is given by

$$\pi = \mathcal{G}(\alpha D, \alpha D) = \sum_{i,j=1}^N \alpha_i \mathcal{G}(z_i, z_j) \alpha_j.$$

The matrix

$$\mathcal{G}_{ij} := \mathcal{G}(z_i, z_j) \tag{28}$$

is referred to as the generalized covariance matrix of the assets  $z_1, \dots, z_n$ . Thus, we consider the optimization problem

$$\min_{\alpha} \pi \tag{29}$$

subject to the constraint

$$\alpha \alpha^T = 1. \tag{30}$$

Setting the Lagrangian

$$L = \pi - \lambda_1 (\alpha \alpha^T - 1),$$

one easily verifies that the first order necessary condition for the optimal portfolio  $F^{(1)}$  is

$$\sum_{j=1}^N \mathcal{G}_{ij} F_j^{(1)} = \lambda_1 F_i^{(1)}, \tag{31}$$

that is,  $F^{(1)}$  is an eigenvector of the generalized covariance matrix  $\mathcal{G}$  relative to the eigenvalue  $\lambda_1$ . Supposing the  $\mathcal{G}$  has  $N$  distinct eigenvalues and iterating this same optimization procedure in subspaces orthogonal to the span of the already given eigenvectors, one obtains the principal directions  $F^{(1)}, \dots, F^{(N)}$  correspondent to the eigenvalues

$$\lambda_1 \geq \dots \geq \lambda_N \geq 0.$$

We then define a matrix  $R$  by

$$z_j = \sum_{i=1}^N R_{ij} F^{(i)}$$

in such a way that an arbitrary portfolio's payoffs may be rewritten as

$$\alpha D = \sum_{j=1}^N \left( \sum_{i=1}^N \alpha_j R_{ij} \right) F^{(i)} =: \sum_{i=1}^N \beta_i F^{(i)}.$$

Next, we restrict ourselves to the projections of portfolios onto the (totally geodesic) affine subspace spanned by the first  $p < N$  principal directions  $F^{(1)}, \dots, F^{(p)}$ , taken as the most significant ones because they represent the largest  $p$  diagonal elements in the generalized covariance matrix in diagonal form, that is,

$$\mathcal{G}_{\text{diag}} = R^{-1} \mathcal{G} R.$$

Hence, we obtain a multi-factor linear model of the form

$$\alpha D = \sum_{i=1}^p \beta_i F^{(i)} + \varepsilon, \tag{32}$$

where

$$\varepsilon = \sum_{i=p+1}^N \beta_i F^{(i)}$$

satisfies

$$\mathcal{G}\left(\varepsilon, \sum_{i=1}^p \beta_i F^{(i)}\right) = 0.$$

The expected return of the  $p$ -principal portfolio

$$\sum_{i=1}^p \beta_i F^{(i)}$$

is

$$\sum_{i=1}^p \beta_i \mathbb{E}[F^{(1)}]$$

and its generalized variance is given by

$$\sum_{i=1}^p \lambda_i \beta_i^2.$$

We claim that the  $p$ -principal portfolio with expected return  $\mu_*$  and minimum generalized variance is determined by the weights

$$\beta_i = \left( \sum_{j=1}^p \frac{\mathbb{E}[F^{(j)}]}{\sqrt{\lambda_j}} \right)^{-1} \frac{\mathbb{E}[F^{(i)}]}{\lambda_i}. \tag{33}$$

To prove this claim, we denote

$$\tilde{\beta}_i = \sqrt{\lambda_i} \beta_i, \quad R_i = \frac{1}{\sqrt{\lambda_i}} \mathbb{E}[F^{(i)}]$$

and then we set the Lagrangian

$$\frac{1}{2} \sum_{i=1}^p \tilde{\beta}_i^2 - \nu \left( \sum_{i=1}^p \tilde{\beta}_i R_i - \mu_* \right)$$

with a constraint given by

$$\sum_{i=1}^p \frac{1}{\sqrt{\lambda_i}} \tilde{\beta}_i = 1.$$

The first order condition is

$$\tilde{\beta}_i = \nu R_i,$$

for all  $i = 1, \dots, p$ . Taking traces and using the constraint condition, one gets

$$\nu \sum_{i=1}^p \frac{R_i}{\sqrt{\lambda_i}} = 1.$$

We conclude that

$$\beta_i = \left( \sum_{j=1}^p \frac{R_j}{\sqrt{\lambda_j}} \right)^{-1} \frac{R_i}{\sqrt{\lambda_i}} \tag{34}$$

as claimed. In sum, we have proven the following theorem.

**Theorem 4.** The  $p$ -principal portfolio with minimum generalized variance is given by

$$\beta_i = \left( \sum_{j=1}^p \frac{\mathbb{E}[F^{(j)}]}{\sqrt{\lambda_j}} \right)^{-1} \frac{\mathbb{E}[F^{(i)}]}{\lambda_i}, \quad (35)$$

where  $\mathbb{E}[F^{(i)}]$  and  $\lambda_i$ ,  $i = 1, \dots, p$ , are, respectively, the expected return and the generalized variance of the first  $p$  eigenvectors  $F^{(1)}, \dots, F^{(p)}$  of the generalized covariance matrix  $\mathcal{G} = \nabla^2 D$ . This portfolio coincides with the projection of the random variable  $z = \alpha D$  over the principal  $p$ -dimensional submanifold spanned by the eigenvectors.

**Acknowledgments:** The authors would like to thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Procs. 309055/2014-8 and 408609/2016-8) for partial funding of this research.

**Author Contributions:** The two authors contributed equally to this work.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Hastie, T.; Stuetzle, W. Principal Curves. *J. Am. Stat. Assoc.* **1989**, *84*, 502–516.
- Tsallis, C. Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487.
- Tsallis, C.; Bukman, D.J. Anomalous diffusion in the presence of external forces: Exact time-dependent solutions and their thermostistical basis. *Phys. Rev. E* **1996**, *54*, R2197–R2200.
- Borland, L. A theory of non-Gaussian option pricing. *Quant. Finan.* **2002**, *2*, 415–431.
- Tsallis, C. *Introduction to Nonextensive Statistical Mechanics*; Springer: New York, NY, USA, 2009.
- Tsallis, C.; Anteneodo, C.; Borland, L.; Osorio, R. Nonextensive statistical mechanics and economics. *Physica A* **2003**, *324*, 89–100.
- Trivellato, B. Deformed exponentials and applications to finance. *Entropy* **2013**, *15*, 3471–3489.
- Naudts, J. *Generalised Thermostatistics*; Springer: London, UK, 2011.
- Naudts, J. Estimators, escort probabilities, and  $\phi$ -exponential families in statistical physics. *J. Inequal. Pure Appl. Math.* **2004**, *5*, 102.
- Naudts, J. On the maximum entropy principle in non-extensive thermostatics. *arXiv* **2004**, arXiv:cond-mat/0405508v1.
- Naudts, J. Generalised exponential families and associated entropy functions. *Entropy* **2008**, *10*, 131–149.
- Naudts, J. Deformed exponentials and logarithms in generalized thermostatics. *Physica A* **2002**, *316*, 323–334.
- Naudts, J.; Anthonis, B. The exponential family in abstract information theory. In *Geometric Science of Information*; Lecture Notes Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8085, pp. 265–272.
- Plastino, A.R.; Plastino, A. Non-extensive statistical mechanics and generalized Fokker-Planck equation. *Physica A* **1995**, *222*, 347–354.
- Vignat, C.; Plastino, A. Central limit theorem and deformed exponentials. *J. Phys. A Math. Theor.* **2007**, *40*, F969–F978.
- Borland, L.; Bouchaud, J.-P. A non-Gaussian option pricing model with skew. *Quant. Finan.* **2004**, *4*, 499–514.
- Borland, L. Option Pricing Formulas based on a non-Gaussian Stock Price Model. *Phys. Rev. Lett.* **2002**, *89*, 098701.
- Moretto, E.; Pasquali, S.; Trivellato, B. Option pricing under deformed Gaussian distributions. *Physica A* **2016**, *446*, 246–263.
- Tsallis, C. Economics and Finance:  $q$ -Statistical stylized features galore. *Entropy* **2017**, *19*, 457.
- Rodrigues, A.F.P.; Guerreiro, I.M.; Cavalcante, C.C. Deformed exponentials and portfolio selection. *Int. J. Mod. Phys. C* **2018**, doi:10.1142/S0129183118500298.
- Rodrigues, A.F.P.; Guerreiro, I.M.; Cavalcante, C.C. A projection pricing model for non-Gaussian financial asset returns. *Inf. Geom.* **2018**, submitted.

22. Zagrodny, D. An optimality of change loss type strategy. *Optimization* **2003**, *52*, 757–772.
23. Sawik, B. Downside Risk Approach for Multi-Objective Portfolio Optimization. In Proceedings of the 2011 (Selected Papers) Operations Research, Zurich, Switzerland, 30 August–2 September 2011; Klatte, D., Lüthi, H.-J., Schmedders, K., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 191–196.
24. Nock, R.; Magdalou, B.; Briys, E.; Nielsen, F. On tracking portfolios with certainty equivalents on a generalization of Markowitz model: The fool, the wise and the adaptive. In Proceedings of the 28th International Conference on Machine Learning, Bellevue, WA, USA, 28 June–2 July 2011; Omnipress: Madison, WI, USA, 2011; pp. 73–80.
25. Nock, R.; Magdalou, B.; Briys, E.; Nielsen, F. *Mining Matrix Data with Bregman Matrix Divergences for Portfolio Selection*; Nielsen, F., Bhatia, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 373–402.
26. Amari, S.-I. *Information Geometry and Its Applications*; Springer: Tokyo, Japan, 2016.
27. Amari, S.-I. Natural gradient works efficiently in learning. *Neural Comput.* **1998**, *10*, 251–276.
28. Amari, S.-I.; Cichocki, A. Information geometry of divergence functions. *Tech. Sci.* **2010**, *58*, 183–195.
29. Lintner, J. The aggregation of investor's diverse judgments and preferences in purely competitive security markets. *J. Financ. Quant. Anal.* **1969**, *4*, 347–400.
30. Mossin, J. Equilibrium in a capital asset market. *Econometrica* **1966**, *34*, 768–783.
31. Sharpe, W. Capital asset prices: A theory of market equilibrium under conditions of risk. *J. Finan.* **1964**, *19*, 425–442.
32. LeRoy, S.; Werner, J. *Principles of Financial Economics*; Cambridge University Press: Cambridge, UK, 2014.
33. Luenberger, D. Projection pricing. *J. Optim. Theory Appl.* **2001**, *109*, 1–25.
34. Fontana, C.; Schweizer, M. Simplified mean-variance portfolio optimisation. *Math. Finan. Econ.* **2012**, *6*, 125–152.
35. Jost, J. *Riemannian Geometry and Geometric Analysis*; Springer: Berlin/Heidelberg, Germany, 2017.
36. Back, K. *Asset Pricing and Portfolio Choice Theory*; Oxford University Press: Oxford, UK, 2010.
37. Markowitz, H. Portfolio selection. *J. Finan.* **1952**, *7*, 77–91.
38. Luenberger, D. *Investment Science*; Oxford University Press: New York, NY, USA, 1998.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).