UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA

ANA FLÁVIA PAIVA RODRIGUES

DEFORMED EXPONENTIALS AND FINANCIAL MARKETS: APPLICATIONS TO PORTFOLIO SELECTION AND ASSET PRICING

FORTALEZA
2018

ANA FLÁVIA PAIVA RODRIGUES

DEFORMED EXPONENTIALS AND FINANCIAL MARKETS: APPLICATIONS TO PORTFOLIO SELECTION AND ASSET PRICING

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Teleinformática do Departamento de Engenharia de Teleinformática da Universidade Federal do Ceará, como parte dos requisitos necessários para a obtenção do título de Doutor em Engenharia de Teleinformática. Área de concentração: Sinais e Sistemas.

Orientador: Prof. Dr. Charles Casimiro Cavalcante

FORTALEZA

2018

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

R611d Rodrigues, Ana Flávia Paiva.
Deformed exponentials and financial markets : Applications to portfolio selection and asset pricing / Ana Flávia Paiva Rodrigues. – 2018.
116 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2018.
Orientação: Prof. Dr. Charles Casimiro Cavalcante.

1. Seleção de carteiras. 2. Modelos de fator único. 3. Exponenciais deformadas. 4. Divergências estatísticas. 5. Curvas principais. I. Título.

CDD 621.38

ANA FLÁVIA PAIVA RODRIGUES

DEFORMED EXPONENTIALS AND FINANCIAL MARKETS:
APPLICATIONS TO PORTFOLIO SELECTION AND ASSET PRICING

Thesis presented to the Graduate Program in Teleinformatics Engineering of the Department of Teleinformatics Engineering of the Federal University of Ceará in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Teleinformatics Engineering. Research area: Signals and Systems.

Approved in: 28/03/2018.

THESIS COMMITTEE

Prof. Dr. Charles Casimiro Cavalcante (Advisor)
Federal University of Ceará (UFC)

Prof. Dr. Guilherme de Alencar Barreto
Federal University of Ceará (UFC)

Prof. Dr. Paulo Rogério Faustino Matos
Federal University of Ceará (UFC)

Prof. Dr. Aline de Oliveira Neves Panazio
Federal University of ABC (UFABC)

Prof. Dr. Renato da Rocha Lopes
Campinas State University (UNICAMP)

To my husband, Jorge, and son, Tomás, with love and affection.

ACKNOWLEDGMENTS

First and foremost, I would like to thank God and St. Francis of Assisi, to whom I am devoted, for the enlightenment and blessings in this long and winding road.

To my husband, Jorge Lira, for his support and encouragement. To my son, Tomás Paiva de Lira, a wonderful gift in my life! To my mother, Francisca Paiva, for her unconditional love everyday in my life. To my father, *in memoriam*, Manoel Ribeiro Rodrigues, for his admirable efforts in a life history dedicated to his family.

To my advisor, Charles Casimiro Cavalcante, for his guidance and confidence in this project. To my professors and colleagues in the Graduate Programs of Economics and Teleinformatics Engineering, for the uncountable learning opportunities. To professor Davi Máximo, advisor in the split site doctoral program in Stanford University, for the opening of new opportunities.

To the members of the thesis committee, professors Guilherme Barreto, Paulo Matos, Aline Panazio and Renato Lopes, for the valuable suggestions and questions that considerably improved this work.

To the Federal University of Ceará and, in particular, to the staff of the Graduate Program in Teleinformatics Engineering for the institutional support. To CAPES, for the financial support, specially in the form of the scholarship during the split term doctoral program. To Lemann Foundation, for the partial support during the visit to Stanford University

RESUMO

Propomos, neste trabalho, modelo de seleção de carteiras de ativos financeiros via um critério de média-divergência, adaptado a retornos com distribuições dadas por exponenciais deformadas. Fixado o retorno esperado desejado, trata-se de minimizar o prêmio de risco definido em termos de uma divergência estatística. No caso de retornos gaussianos, a abordagem proposta reduz-se ao clássico modelo de média-variância concebido por H. Markowitz. Na sequência, reformulamos o método de apreçamento por projeções ortogonais desenvolvido por Luenberger para o contexto de mínima divergência, o que nos permite propor modelos de fator único, dentre os quais uma variante do CAPM com betas dependendo de uma matriz de covariância generalizada. Os valores principais dessa matriz nos permitem, por fim, definir e aplicar uma noção estendida de curvas principais, o que adapta os conceitos desenvolvidos por Hastie e Stuetzle ao caso de exponenciais deformadas e divergências de Bregman.

Palavras-chave: Seleção de carteiras. Modelos de fator único. Exponenciais deformadas. Divergências estatísticas. Curvas principais.

ABSTRACT

In this work, we propose a portfolio selection model based on a mean-divergence criteria, adapted to financial returns distributed according deformed exponential probability densities. Fixed a desired expected return, the method reduces to the minimization of a risk premium defined in terms of a statistical divergence, In the particular case of Gaussian returns, we recover the classical mean-divergence model by H. Markowitz. Next, we reformulate the projection pricing theory by Luenberger in the context of divergences as risk measures. This allowed us to define single factor models, including a variant of the CAPM whose beta coefficients depend on a Fisher metric that plays the role of a generalized covariance matrix. The eigenvalues of this matrix are used to define an extended notion of principal curves that adapts the work by Hastie and Stuetzle to the case of deformed exponentials and their correspondent Bregman divergences.

Keywords: Portfolio selection. Single factor models. Deformed exponentials. Statistical divergences. Principal curves.

LIST OF FIGURES

CONTENTS

1 INTRODUCTION

The formulation of a non-extensive Statistical Physics by C. Tsallis (35), (36) and collaborators has been developed along the last two decades in a wide range of applications to complex systems, particularly in Finance (37), (38), (7), (40). In this work, we propose a model of portfolio selection of financial assets that explores the non-additivity and non-normality aspects of Tsallis' Thermostatistics.

As highlighted by J. Naudts, deformed exponentials play a central role in the foundations of that Generalized Thermostatistics. Indeed, Naudts' work established deep and fruitful connections between Statistical Physics and Information Geometry, (24), (25), (26), (27). 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 A. Plastino and C. Vignat (32), (33), (27), (28), (29). The analytic and geometric features of deformed exponentials suggest that they are well suited to model non-normally distributed returns of financial assets. 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 (7), (8), (9), (40), (22). Other related developments are summarized in (38), (39).

Up to our knowledge, however, a systematic theory of portfolio optimization in the context of deformed exponentials has not yet been fully formalized. 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 functions, see Section 2.1.

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 the earlier developments of the Quantitative Finance under different methods. In (30), (31), R. 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 measures 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 (1), (3).

For the reader's benefit, we have collected basic fundamental facts on Finance Theory and Information Theory in Chapters 2 and 3, respectively. The notation in these chapters is fixed in order to be consistent with the presentation of the original material in the subsequent chapters.The third chapter presents a brief but essentially self-contained presentation of the Classical Portfolio Selection Theory, needed to guarantee a precise statement of our results. In Chapter 4 we extend the mean-divergence model in (30), (31) to deformed exponentials families. This part of the thesis is structured as follows. Earlier in Section 3.2 we recall basic notions and facts on the statistical manifold of deformed exponentials. We refer the reader to (1) for a comprehensive mathematical description of these manifolds. We also briefly describe in Section 4.1 a generalized family of hyperbolic risk aversion (HARA) functions as the natural choice of utility functions associated to returns with deformed exponential distributions. The mean-divergence model is presented in Section 4.2 as an extension of both Markowitz and R. Nock el al. models. Theorem 4.1 in Section 4.3 states that the optimal portfolio for the generalized mean-divergence model is given by a closed expression involving the information metric, that is, the Hessian of the cumulant function of the deformed exponential family. In the particular case of q-Gaussian distributions, the optimal portfolio is explicitly given by

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

where Σ_q stands for the q-Gaussian variance, which corresponds to the variance-covariance matrix in the Gaussian case. This theorem motivates the steepest descent algorithm by the natural (Riemannian) gradient of the risk premium in Section 4.4. Our method differs from the one in (30), (31) in some relevant aspects. For instance, the iterations in the search algorithm are indeed defined in terms of the projection of the natural gradient on the simplex of admissible portfolios. Moreover, we stress the fact that this machine learning procedure stems naturally from the generalized Markowitz's formula in Theorem 4.1 what provides an analytical justification that was not entirely evident in (30), (31). Some empirical support to the proposed method is discussed in Section 4.5. There we compare the cumulated returns and the evolution of the divergence for optimal portfolios according to the mean-divergence model and the classical one by Markowitz. Chapter 4 finishes with a brief discussion of prospective empirical tests and theoretical developments.

Chapter 5 is a natural sequel of Chapter 4 in the sense that we propose a generalization of beta pricing models adapted to a mean-divergence portfolio selection (34), (17), (23). In particular, we present an extension of Capital Asset Pricing Model (CAPM) flexible enough to be applied for financial returns with deformed exponential distributions. Our method relies on a geometric approach to the classical mean-variance analysis developed by S. LeRoy and J. Werner (16) and D. Luenberger (19), see also (13). We refer the reader to (6) to a precise and comprehensive exposition of the classical CAPM model.

This chapter begins with the definition in sections 5.1 and 5.2 of the geometric setting of the space of contingent claims $\mathcal{M}$ and the subspace of traded financial assets $\mathcal{M}'$ in terms of statistical manifolds of probability distributions as the manifold of ϕ-

deformed exponentials. Some geometric features of that manifold are summarized in Section 5.4. Following (16), we define expectation and price kernels in terms of a Bregman statistical divergence in $\mathcal{M}$. These two distinguished assets span the mean-divergence efficient frontier in $\mathcal{M}'$. As in (30), (31) and (10), the underlying idea is that the divergence defined a novel risk measure that replaces the variance in the case of normal distributions.

Taking this into account, we deduce in Section 5.5 an expression of a minimum divergence portfolio in the efficient frontier. Roughly speaking, the efficient frontier is the set of tradable assets with the minimum risk among those with the same expected return. As in the classical beta pricing models, the proportions of market portfolio and risk-free assets in this optimal portfolio are dictated by a linear regression coefficient

$$\beta = -\frac{g(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{e}})}{g(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{q}} - R_{\mathsf{e}})}, \tag{1}$$

where R_{q} and R_{e} correspond to the market return and risk-free return in the standard beta pricing equation. Here, g stands for the Riemannian metric in $\mathcal{M}'$ given by the Hessian of the cumulant function K of the deformed exponential probability density. In the classical case of Gaussian distributions, we have a flat metric given by the variance. In our general approach, the Riemann curvature of $\mathcal{M}'$ encodes third and fourth order moments of the distribution as follows from equation (300). Using this machinery, one can obtain further developments for applications of the theoretical model deduced in this work. Some prospective directions are related to estimation techniques of the generalized beta factors, specially useful for valuation models in Corporate Finance.

Now, we briefly discuss the contents of Chapter 6. In their seminal paper (14), T. Hastie and W. 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 pass 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 $\mathbf{f}$ is critical for the variance of the *Euclidean* distance between the data and any locally defined perturbation of $\mathbf{f}$. In particular, a straight line is a principal curve if and only if its direction is an eigenvector of the covariance matrix of z.

The original idea by Hastie and Stuetzle has been developed into relevant improvements, applications and extensions. We point out however that most of the times the criticality of a principal curve is defined in terms of the Euclidean distance. Hence, although $\mathbf{f}$ 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-dimensional 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 whose probabilities are given by exponencial and deformed exponential distributions. In the context of exponential and ϕ-exponential statistical families, straight lines are replaced by affine geodesics and the Hessian of the cumulant function plays the role of a generalized covariance.

Chapter 6 is organized as follows. In Section 6.1, 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 6.2. In Section 6.4, 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.

2 MODERN FINANCE THEORY: A SHORT REVIEW

Financial assets, more precisely their payoffs at a fixed time, say $t = 1$, are represented by random variables of the form

$$z = z(s),$$

where s are the states of the world with probability distribution specified by some density $p(s; \boldsymbol{\vartheta})$. Here, $\boldsymbol{\vartheta}$ is the distribution parameter of a family of probability distributions whose densities define a n-dimensional statistical manifold

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

with $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_n)$ taking values in some open subset U of the n-dimensional Euclidean space $\mathbb{R}^n$. Given a time series of payoffs $\{z^{(i)}(s)\}_{i=0}^{T}$ one defines the returns of the asset as the percentual ratio

$$\mathbf{r}^{(i+1)}(s) = \frac{z^{(i+1)}(s) - z^{(i)}(s)}{z^{(i)}(s)} \tag{2}$$

for $i = 0, \ldots, T-1$.

Example 2.1 *It is an usual however possibly unrealistic assumption that the returns of a financial asset are (log)-normally distributed. This means that the random variable $\mathbf{r}(s)$ has a Gaussian distribution whose density is of the form*

$$p(s, \mu, \sigma) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\left(-\frac{1}{2}(\mathbf{r} - \mu)^2 \frac{1}{\sigma^2} \right),$$

where $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_+$ are, respectively, the mean and variance of the distribution. If one considers N financial assets, then it is commonly supposed that their returns are distributed according to a multivariate Gaussian distribution with density given by

$$p(s, \boldsymbol{\vartheta}) = p(s, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{r} - \mu)^T \Sigma^{-1} (\mathbf{r} - \mu) \right).$$

Note that in those examples the parameters $\boldsymbol{\vartheta} = (\vartheta, \Theta) \in \mathbb{R}^N \times M(N, \mathbb{R})$ are explicitly given by

$$\vartheta = \Sigma^{-1}\mu \in \mathbb{R}^N$$

and

$$\Theta = \frac{1}{2}\Sigma^{-1} \in M(N, \mathbb{R}),$$

where $\mu \in \mathbb{R}^N$ and $\Sigma \in M(N, \mathbb{R})$ are, respectively, the mean and the covariance matrix of the distribution.

In sum, multivariate Gaussian distributions are a family of probability distributions parameterised by their mean and variance. We will see later that those distributions determine a subset of a larger family of probability distributions, namely an exponential family.

We refer the reader to (6) and (12) as comprehensive and formal presentations of the main aspects of Finance Theory.

One of the fundamental principles in Modern Finance Theory is the No Arbitrage Theorem whose far reaching consequences encompass the whole theory of asset pricing. As a simpler version of the No Arbitrage Theorem, we restrict ourselves to consider a risky asset with payoff $z(s)$, an option $c(s)$ whose underlying asset payoff is z and a risk-free asset **1**. This last notation indicates that this asset yields the same risk-free return, say r, in every state of the world, that is, under any circumstances. On the contrary, both $z(s)$ and $c(s)$ are by definition sensitive to the effect of distinct states of the world. Supposing by the sake of simplicity that there are two different states of the world, we can arrange all the possible payoffs in a matrix as

$$D = \begin{bmatrix} 1+r & 1+r \\ z^{(1)}(\text{down}) & z^{(1)}(\text{up}) \\ c^{(1)}(\text{down}) & c^{(1)}(\text{up}) \end{bmatrix}$$

Let Q be the vector of current market values (at time $t = 0$) of these securities, that is,

$$Q = \begin{bmatrix} 1 \\ z^{(0)} \\ c^{(0)} \end{bmatrix}$$

The possible returns of z are given by

$$r(\text{down}) = \frac{z^{(1)}(\text{down}) - z^{(0)}}{z^{(0)}} \quad \text{and} \quad r(\text{up}) = \frac{z^{(1)}(\text{up}) - z^{(0)}}{z^{(0)}}.$$

We distinguish between these two scenarios declaring that

$$r(\text{down}) < r(\text{up}).$$

Suppose that

$$r < r(\text{down}) < r(\text{up}). \tag{3}$$

In this case, an investor could get a *long* position in the asset z buying shares of it with money borrowed at a risk-free rate r, a sort of financial leverage. Even in the worst scenario, (3) implies that the investor would obtain positive returns at time $t = 1$. Now, under the assumption that

$$r(\text{down}) < r(\text{up}) < r, \tag{4}$$

an investor could have positive returns in both scenarios getting a *short* position in z, that is, selling shares of z and investing in the risk-free security. We conclude that if either (3) or (4) are valid, then there are *arbitrage* possibilities in the market: in both cases, it is characterized the existence of a portfolio composed by the risky asset and the risk-free security that costs nothing to the investor and, in spite of that, yields positive returns.

Those arbitrage opportunities are ruled out by imposing that

$$r(\text{down}) < r < r(\text{up}). \tag{5}$$

This condition has important implications on the pricing of the risky assets as we will see in the sequel. We would like to prove the existence of a state-price vector $\pi = (\pi(\text{down}), \pi(\text{up}))$, that is, a vector with positive components for which it holds that

$$Q = D\,\pi,$$

that is,

$$\begin{bmatrix} 1 \\ z^{(0)} \\ c^{(0)} \end{bmatrix} = \begin{bmatrix} 1+r & 1+r \\ z^{(1)}(\text{down}) & z^{(1)}(\text{up}) \\ c^{(1)}(\text{down}) & c^{(1)}(\text{up}) \end{bmatrix} \begin{bmatrix} \pi(\text{down}) \\ \pi(\text{up}) \end{bmatrix}.$$

Then we would like to guarantee the existence of positive solutions to the system of linear equations

$$\begin{cases} (1+r)\pi(\text{down}) + (1+r)\pi(\text{up}) = 1 \\ z^{(1)}(\text{down})\pi(\text{down}) + z^{(1)}(\text{up})\pi(\text{up}) = z^{(0)}, \end{cases} \tag{6}$$

or, equivalently,

$$\begin{cases} (1+r)\pi(\text{down}) + (1+r)\pi(\text{up}) = 1 \\ (1+r(\text{down}))\pi(\text{down}) + (1+r(\text{up}))\pi(\text{up}) = 1, \end{cases} \tag{7}$$

with $\pi(\text{down}) > 0, \pi(\text{up}) > 0$. Subtracting the first equation from the second, we have

$$(r(\text{down}) - r)\pi(\text{down}) + (r(\text{up}) - r)\pi(\text{up}) = 0$$

Hence, in order to get *positive* solutions, a sufficient and necessary condition is

$$r(\text{down}) < r < r(\text{up}),$$

a no-arbitrage condition with strict inequalities. Hence, it follows that in the absence of arbitrage opportunities, there is a well-defined state-price vector π.

Note that the first equation in 6 implies that the components

$$\widehat{\pi}(\text{down}) = (1+r)\pi(\text{down}), \quad \widehat{\pi}(\text{up}) = (1+r)\pi(\text{up}) \tag{8}$$

can be interpreted as set of probabilities, referred to as the *risk-neutral* probabilities. The second equation in 6 may be written in terms of this notation as

$$z^{(0)} = \frac{1}{1+r}\big(z^{(1)}(\text{down})\widehat{\pi}(\text{down}) + z^{(1)}(\text{up})\widehat{\pi}(\text{up})\big),$$

that is,

$$z^{(0)} = \mathbb{E}_{\widehat{\pi}}\left[\frac{1}{1+r}z^{(1)}\right]. \tag{9}$$

This last equation asserts that the current market value of a risky asset is given, in the absence of arbitrage opportunities, as the expected value of the payoffs of this asset, discounted at a risk-free return data, where the expectation is calculated with respect to *risk-neutral* probabilities $\widehat{\pi}$.

Note that, intuitively, $z(s)$ does not behave as in (9) if one considers *actual* subjective probabilities since the expected return of a risky asset must exceed the risk-free return, that is,

$$z^{(0)} < \mathbb{E}\left[\frac{1}{1+r}z^{(1)}\right]$$

if the expectation is computed with respect to some set of subjective probabilities. This last inequality means that the discounted (with respect to a risk-free rate r) expected *subjetive* return of a risky asset must be positive in order to motivate an investment decision.

We have ignored so far the consequence of the existence of a state-price vector π on the option pricing. It follows from our earlier considerations that $c(s)$ also satisfies the condition

$$c^{(0)} = \mathbb{E}_{\widehat{\pi}}\left[\frac{1}{1+r}c^{(1)}\right]. \tag{10}$$

This is the starting point of the binomial method of option pricing proposed by Cox, Ross and Rubinstein. In the limit, that binomial process restores the celebrated Black-Scholes-Merton formula for option pricing in the context of time continuous stochastic processes.

Now we state the No Arbitrage Theorem in a more general setting. We partially follow here the terms and notations in (12). In the general case of N securities $z_1(s), \ldots, z_N(s)$, a *portfolio* is defined by an allocation vector

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N) \in \mathbb{R}^N \tag{11}$$

This means that an investor allocates α_i shares of her initial endowment (at time $t = 0$, say) on the asset z_i. Supposing that we have only a finite set of states of the world $\Omega = \{s_1, \ldots, s_K\}$, we denote the payoff $N \times K$ matrix by

$$D = \begin{bmatrix} z_1(s_1) & \ldots & z_1(s_K) \\ \vdots & \ddots & \vdots \\ z_N(s_1) & \ldots & z_N(s_K) \end{bmatrix}$$

The payoff portfolio is given by the random variable

$$\boldsymbol{\alpha} D = \begin{bmatrix} \alpha_1 z_1(s_1) + \ldots + \alpha_N z_N(s_1) & \ldots & \alpha_1 z_1(s_K) + \ldots + \alpha_N z_N(s_K), \end{bmatrix}$$

where the possible payoffs correspond to the time 1. The set of payoffs available via trades in security markets is the linear span of a basis of traded assets $z_1, \ldots, z_N$, that is,

$$\mathcal{M} = \{z \in \mathbb{R}^K : z = \boldsymbol{\alpha} D\}. \tag{12}$$

A market is said to be complete if $\mathcal{M} = \mathbb{R}^K$, that is, every possible payoff $z \in \mathbb{R}^K$ can be replicated by a traded portfolio. This is equivalent to the condition that the rank of D is exactly K. In particular, $N \geq K$.

If the rank of D is N (in particular $N \leq K$), there are no *redundant* portfolios: if there exist two portfolios in $\mathcal{M}$ with the same payoff, they are equal. Indeed, given $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ such that

$$\boldsymbol{\alpha} D = \boldsymbol{\alpha}' D$$

then

$$(\boldsymbol{\alpha} - \boldsymbol{\alpha}')D = 0.$$

Since the rank of D is N, it defines an injective linear map and then

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}'.$$

Given the current value market, that is, the price vector of all the traded assets

$$Q = \begin{bmatrix} z_1^{(0)} \\ \vdots \\ z_N^{(0)} \end{bmatrix}$$

the price of a portfolio $\boldsymbol{\alpha}$ is

$$\langle Q, \boldsymbol{\alpha} \rangle = \sum_{i=1}^{N} z_i^{(0)} \alpha_i.$$

Definition 2.1 *An* arbitrage *portfolio* $\boldsymbol{\alpha} \in \mathbb{R}^N$ *satisfies by definition one of the following conditions:*

i. $\langle Q, \boldsymbol{\alpha} \rangle \leq 0$ *and* $\boldsymbol{\alpha} D(s) > 0$ *for every* $s \in \Omega$*.*
ii. $\langle Q, \boldsymbol{\alpha} \rangle < 0$ *and* $\boldsymbol{\alpha} D(s) \geq 0$ *for every* $s \in \Omega$*.*

According to this definition, the arbitrage portfolio $\boldsymbol{\alpha}$ can be purchased without (respectively, negative) costs and guarantees some positive (respectively, nonnegative) return in all states of the world.

Theorem 2.1 (Fundamental Theorem of Finance Theory) *There is no arbitrage if and only if there is a state-price vector* $\pi(s) > 0$ *such that*

$$Q = D\,\pi. \tag{13}$$

For a proof of this theorem, we refer the reader to (12), p. 10.

As above we define risk-neutral probabilities from the state-price vector: denote

$$\pi_0 = \pi_1 + \ldots . + \pi_N$$

and set

$$\widehat{\pi} = \frac{1}{\pi_0}\pi$$

in such a way that $\widehat{\pi}_i > 0$ and

$$\widehat{\pi}_1 + \ldots + \widehat{\pi}_N = 1.$$

Hence

$$Q_i = \sum_{j=1}^{N} \pi_j D_{ij} = \frac{1}{\pi_0} \sum_{j=1}^{N} \widehat{\pi}_j D_{ij} = \mathbb{E}_{\widehat{\pi}}\left[\frac{1}{\pi_0} D_i\right],$$

recovering the fact that, using risk-neutral probabilities, the current price of an asset z_i is given by the expected payoff, discounted at a risk-free rate π_0. Indeed, if there exists a risk-less asset **1** with same payoff in every state of the world $s \in \Omega$, then

$$1 \cdot \pi_1 + \ldots + 1 \cdot \pi_N = \text{expected future value of } \mathbf{1} = 1 + r,$$

where r is the rate of return of the risk-less asset. Hence,

$$\pi_0 = 1 + r.$$

2.1 Utility and equilibrium

An agent is characterized by an endowment $\mathsf{e} \in \mathbb{R}^K_+$ and a differentiable strictly increasing *utility* function

$$u : \mathbb{R}^K_+ \to \mathbb{R}$$

where

$$\mathbb{R}^K_+ = \{c = (c_1, \ldots, c_K) \in \mathbb{R}^K : c_i \geq 0\}.$$

Note that c is a random variable depending on the K states of the world and representing the investor's consumption choices. More generally, one may consider a utility function depending on the investor's choices (consumption plans) at two distinct times, say $t = 0$ and $t = 1$. In other terms, we consider an utility function defined in $\mathbb{R}_+ \times \mathbb{R}_+^K$ of the form

$$u(c^{(0)}, c^{(1)}).$$

Here $c^{(0)} \geq 0$ and $c^{(1)} \geq 0$ are, respectively, the consumption plans of the investor at times $t = 0$ and $t = 1$. Both e and c are modelled as random variables $\mathsf{e} = \mathsf{e}(s)$ and $c = c(s)$ depending on the states of the world $s \in \Omega$. We refer the reader to (6), (16), (18) for further details on the notions and fundamental results in this section.

Given consumption plans c, c_- and c_+ one has by assumption that

$$u(c_-^{(0)}, c^{(1)}) < u(c_+^{(0)}, c^{(1)})$$

if $c_-^{(0)} < c_+^{(0)}$ and

$$u(c^{(0)}, c_-^{(1)}) < u(c^{(0)}, c_+^{(1)})$$

if $c_-^{(1)} < c_+^{(1)}$. The partial order for vectors (consumption bundles) in $\mathbb{R}_+^K$ is defined here by

$$c_- < c_+ \quad \text{if and only if} \quad c_{-i} < c_{+i},$$

for every $i = 1, \dots, K$. We then consider the following constrained optimization problem

$$\max_{c,\,\boldsymbol{\alpha}} u(c^{(0)}, c^{(1)}) \tag{14}$$

subject to the constraints

$$c^{(0)} \leq \mathsf{e}^{(0)} - \langle Q, \boldsymbol{\alpha} \rangle \tag{15}$$

$$c^{(1)} \leq \mathsf{e}^{(1)} + \boldsymbol{\alpha} D \tag{16}$$

Theorem 2.2 *If there is no arbitrage, then the problem (14), (15)-(16) has a solution, that is, there exists an optimal consumption plan and optimal portfolio.*

Proof. Without loss of generality, we may assume that there are no redundant portfolios. Indeed redundant portfolios $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ have the same payoffs, that is,

$$\boldsymbol{\alpha} D = \boldsymbol{\alpha}' D.$$

Since there are no arbitrage portfolios, we conclude that $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ also have the same prices:

$$\langle Q, \boldsymbol{\alpha} \rangle = \langle Q, \boldsymbol{\alpha}' \rangle.$$

Then the maximization problem does not distinguish between redundant portfolios and then we may take only a representative of possible redundant portfolios. In technical terms, this means that we may consider the quotient space of the feasible set by $\ker D$.

Since u is a continuous function, the optimization problem can be solved once we prove that the constraints define a closed and bounded subset in $\mathbb{R}^K_+$. Given admissible sequences $\{c_n\}$ and $\{\boldsymbol{\alpha}_n\}$ of consumption plans and portfolio allocations, respectively, we have

$$c_n^{(0)} \leq \mathsf{e}^{(0)} - \langle Q, \boldsymbol{\alpha}_n \rangle \tag{17}$$

$$c_n^{(1)} \leq \mathsf{e}^{(1)} + \boldsymbol{\alpha}_n D \tag{18}$$

If $c_n \to c$ and $\boldsymbol{\alpha}_n \to \boldsymbol{\alpha}$ then $\{c, \boldsymbol{\alpha}\}$ lies in the feasible set, that is,

$$c^{(0)} \leq \mathsf{e}^{(0)} - \langle Q, \boldsymbol{\alpha} \rangle \tag{19}$$

$$c^{(1)} \leq \mathsf{e}^{(1)} + \boldsymbol{\alpha} D \tag{20}$$

Therefore, the feasible set is closed. In order to prove that it is also bounded, suppose by contradiction that there exists an *unbounded* sequence $\{c_n, \boldsymbol{\alpha}_n\}$ in the feasible set. This implies that the sequence of portfolios $\{\boldsymbol{\alpha}_n\}$ is unbounded. Otherwise, the constraint inequalities

$$c_n^{(0)} \leq \mathsf{e}^{(0)} - \langle Q, \boldsymbol{\alpha}_n \rangle \tag{21}$$

$$c_n^{(1)} \leq \mathsf{e}^{(1)} + \boldsymbol{\alpha}_n D \tag{22}$$

would imply that $\{c_n\}$ would be bounded too. Hence, $\{\boldsymbol{\alpha}_n\}$ is unbounded in the sense that

$$|\boldsymbol{\alpha}_n| = (\alpha_1^2 + \ldots + \alpha_n^2)^{1/2} \to +\infty$$

as $n \to +\infty$. On the other hand, since the consumption plans are non-negative, that is, $c_n^{(0)} \geq 0$ and $c_n^{(1)} \geq 0$, we have

$$\langle Q, \boldsymbol{\alpha}_n \rangle \leq \mathsf{e}^{(0)} \tag{23}$$

$$- \boldsymbol{\alpha}_n D \leq \mathsf{e}^{(1)} \tag{24}$$

Therefore, dividing both sides of the constraint inequalities by $|\boldsymbol{\alpha}_n|$ yields

$$\left\langle Q, \frac{\boldsymbol{\alpha}_n}{|\boldsymbol{\alpha}_n|} \right\rangle \leq \frac{1}{|\boldsymbol{\alpha}_n|} \mathsf{e}^{(0)} \tag{25}$$

$$- \frac{\boldsymbol{\alpha}_n}{|\boldsymbol{\alpha}_n|} D \leq \frac{1}{|\boldsymbol{\alpha}_n|} \mathsf{e}^{(1)} \tag{26}$$

However a subsequence of

$$\frac{\boldsymbol{\alpha}_n}{|\boldsymbol{\alpha}_n|}$$

converges to some non-zero portfolio $\boldsymbol{\alpha} \in \mathbb{R}^N$. Taking limits on both sides of the inequalities 25-26 above, one concludes that

$$\langle Q, \boldsymbol{\alpha} \rangle \leq 0 \tag{27}$$

$$-\boldsymbol{\alpha} D \leq 0 \tag{28}$$

Since $\boldsymbol{\alpha} \neq 0$ and due to the fact that only the trivial portfolio has zero payoff, we conclude that

$$\boldsymbol{\alpha} D > 0$$

with $\langle Q, \boldsymbol{\alpha} \rangle \leq 0$ what means that $\boldsymbol{\alpha}$ is an arbitrage portfolio. This contradiction proves that the feasible set is closed and bounded. This is enough to ensure the existence of an optimal solution, finishing the proof. □

Theorem 2.3 *If there exist an optimal consumption plan and optimal portfolio for the problem (14), (15)-(16), then there is no arbitrage portfolio.*

Proof. Suppose by contradiction that there exists an arbitrage portfolio $\boldsymbol{\alpha}_0$. Then, given any feasible consumption plan and portfolio $\{c, \boldsymbol{\alpha}\}$ one has

$$\langle Q, \boldsymbol{\alpha}_0 \rangle \leq 0$$

(respectively, $\langle Q, \boldsymbol{\alpha}_0 \rangle < 0$) and

$$\boldsymbol{\alpha}_0 D > 0$$

(respectively, $\boldsymbol{\alpha}_0 D \geq 0$.) In both arbitrage cases, one gets

$$c^{(0)} \leq \mathsf{e}^{(0)} - \langle Q, \boldsymbol{\alpha} + \boldsymbol{\alpha}_0 \rangle \tag{29}$$

$$c^{(1)} \leq \mathsf{e}^{(1)} + (\boldsymbol{\alpha} + \boldsymbol{\alpha}_0) D, \tag{30}$$

that is,

$$c^{(0)} + \langle Q, \boldsymbol{\alpha}_0 \rangle \leq \mathsf{e}^{(0)} - \langle Q, \boldsymbol{\alpha} \rangle \tag{31}$$

$$c^{(1)} - \boldsymbol{\alpha}_0 D \leq \mathsf{e}^{(1)} + \boldsymbol{\alpha} D, \tag{32}$$

Since u is strictly increasing and at least one of the inequalities above is strict, we conclude that the consumption plan $(c^{(0)} - \langle Q, \boldsymbol{\alpha}_0 \rangle, c^{(1)} - \boldsymbol{\alpha}_0 D)$ is strictly preferred to the optimal consumption plan $(c^{(0)}, c^{(1)})$. This contradiction proves that there is no arbitrage portfolio, finishing the proof. □

In sum, the no-arbitrage condition is necessary and sufficient to the existence of an optimal consumption plan and portfolio for investors whose utility functions are strictly increasing with respect to both present and future consumption plans.

The constrained maximization problem 14, 15-16 may be formulated in terms of the Lagrangian

$$L = u(c^{(0)}, c^{(1)}) + \lambda(c^{(0)} - \mathsf{e}^{(0)} + \langle Q, \boldsymbol{\alpha} \rangle) + \mu(c^{(1)} - \mathsf{e}^{(1)} - \boldsymbol{\alpha} D) \tag{33}$$

It follows that the first-order conditions (at a regular maximum point) are given by

$$\partial_{c^{(0)}} u + \lambda = 0, \tag{34}$$

$$\partial_{c^{(1)}} u + \mu = 0, \tag{35}$$

$$\lambda Q - \mu D = 0. \tag{36}$$

Combining these equations gives

$$Q = D \frac{\partial_{c^{(1)}} u}{\partial_{c^{(0)}} u} \tag{37}$$

We conclude that the state-price vector π is given by the marginal rate of substitution, that is,

$$\pi = \frac{\partial_{c^{(1)}} u}{\partial_{c^{(0)}} u} \cdot \tag{38}$$

Equations 34-36 are also sufficient conditions to the optimality of some feasible $\{c, \boldsymbol{\alpha}\}$ in the case when the utility function is differentiable and concave, see (12), p. 13. The concavity assumption is also useful to guarantee the existence of a general equilibrium in securities market, see Theorem 1.8.1 in (16).

Theorem 2.4 *If each agent's admissible consumption plans are restricted to be non-negative, her utility function is strictly increasing and concave, her initial endowment is strictly positive, and there exists a portfolio with strictly positive payoff, then there exists an equilibrium in security markets.*

An equilibrium in a market composed by M investors $(u_{(\ell)}, \mathsf{e}_\ell)$, $\ell = 1, \ldots, M$, and N assets with payoffs D is by definition a state-price vector π and individual optimal allocations $\boldsymbol{\alpha}_\ell$, $\ell = 1, \ldots, M$, for (14), (15)-(16) such that

$$\sum_{\ell=1}^{M} \boldsymbol{\alpha}_\ell = 0 \tag{39}$$

This expression means that a short position of an agent trading in the market corresponds to a long position of other agent. The total sum of market operations, encoded in the sum of individual portfolio allocations, is equal to zero.

Expected utility

In this section, we restrict ourselves to the case of expected utility functions, a central concept in von Neumann and Morgenstern theory of choice under uncertainty. A detailed account of this theory and criticism may be found in (21) and (6). The theoretical apparatus of von Neumann and Morgenstern relies on a list of axioms, the most controversial being the independence axiom that states that consumption plans, represented by random variables of the form

$$c = (c_1, \ldots, c_K) \in \mathbb{R}^K,$$

satisfy the preferences ordering relations

$$(c_1, \ldots, c_{i-1}, x, c_{i+1}, \ldots, c_K) \succsim (c'_1, \ldots, \hat{c}'_{i-1}, x, \hat{c}'_{i+1}, \ldots, c'_K)$$

if

$$(c_1, \ldots, c_{i-1}, c_i, c_{i+1}, \ldots, c_K) \succsim (c'_1, \ldots, \hat{c}'_{i-1}, c'_i, \hat{c}'_{i+1}, \ldots, c'_K)$$

for all $c, c' \in \mathbb{R}^K$ and $x \in \mathbb{R}$. Here, $A \succsim B$ is a preference ordering that means that "A is preferred to B". Besides that, one of the main hypothesis in this theory is that expectations are calculated with respect to given *objective* probabilities. Hence, a portfolio is described in terms of a "gamble"

$$\big((c_1, p_1), \ldots, (c_K, p_K)\big),$$

that is, a *given* list of possible payoffs of the consumption plan and their respective probabilities.

It follows from that axiomatic approach that rational optimization under uncertainty could be modeled in terms of a particular form of the utility function that we are going to use in the sequel. Indeed, we consider in what follows the case of an expected utility function defined in terms of a set of probabilities $\{p_1, \ldots, p_K\}$ by

$$u(c^{(0)}, c^{(1)}) = u_0(c^{(0)}) + \mathbb{E}[u_1(c^{(1)})] = u_0(c^{(0)}) + \sum_{j=1}^{K} p_j u_1(c^{(1)}(s_j)). \tag{40}$$

This means that u is separated into two components, u_0 and u_1 that depend respectively on the consumption plans at times $t = 0$ and $t = 1$. Note that the function u_1 does not depend on the parameter j, that is, does not depend on the states of the world $s_j \in \Omega$. In this case, condition 37 reads as

$$Q = \frac{1}{\partial_{c^{(0)}} u_0} \sum_{j=1}^{K} p_j D u'_1(c^{(1)}(s_j)),$$

what implies that

$$Q = \frac{1}{\partial_{c^{(0)}} u_0} \mathbb{E}[D u_1'(c^{(1)})]. \tag{41}$$

For the sake of simplicity we will assume for a while that $u_0 \equiv 1$. An agent is said to be *risk-averse* if she prefers the expectation of any consumption plan to the consumption plan itself, that is,

$$\mathbb{E}[u_1(c^{(1)})] \leq u_1(\mathbb{E}[c^{(1)}]), \tag{42}$$

for every consumption plan c. Intuitively, the right-hand side $u_1(\mathbb{E}[c^{(1)}])$ is the utility of the expected value of the lottery represented by the random variable $c^{(1)}$. This is the risk-less position to the investor. The left-hand side $\mathbb{E}[u_1(c^{(1)})]$ is the expected utility of the lottery. The inequality means that, in order to move from the risk-less position to a riskier one, the risk-averse investor demands a positive risk premium that we are going to define shortly afterwards. For risk neutral agents instead, we have

$$\mathbb{E}[u_1(c^{(1)})] = u_1(\mathbb{E}[c^{(1)}]), \tag{43}$$

for every consumption plan c. It turns out that an agent is risk averse (respectively, risk neutral) if and only if her von Neumann-Morgenstern utility function u_1 is concave (respectively, linear). This follows from Jensen's inequality for concave functions. Indeed, if u_1 is concave we have

$$p_1 u_1(c^{(1)}(s_1)) + \ldots + p_K u_1(c^{(1)}(s_K)) \leq u_1(p_1 c^{(1)}(s_1) + \ldots + p_K u^{(1)}(s_K)),$$

that is,

$$\mathbb{E}[u_1(c^{(1)})] \leq u_1(\mathbb{E}[c^{(1)}]).$$

The Jensen's gap

$$J = u_1(\mathbb{E}[c^{(1)}]) - \mathbb{E}[u_1(c^{(1)})]$$

increases with the concavity of the graph of u_1. This indicates that the second derivative of u_1 encodes some information about the investor's aversion to risk. Indeed, one of the commonly used risk measures is the Arrow-Pratt *absolute* risk aversion coefficient defined by

$$a = -\frac{u_1''(c^{(1)})}{u_1'(c^{(1)})}. \tag{44}$$

We now define the risk premium, a notion that is closely related to Jensen's gap. Let $\bar{c}^{(1)}$ be the mean of the random variable $c^{(1)}$, that is,

$$\bar{c}^{(1)} = \mathbb{E}[c^{(1)}]$$

Setting

$$c^{(1)} = \bar{c}^{(1)} + \varepsilon,$$

where ε is a random variable with zero mean, we define the risk premium Π by

$$\mathbb{E}[u_1(\bar{c}^{(1)} + \varepsilon)] = u_1(\bar{c}^{(1)} - \Pi) \tag{45}$$

The certainty equivalent is by definition

$$C = \bar{c}^{(1)} - \Pi. \tag{46}$$

The risk premium Π can be regarded as the maximum amount the investor renounce in order to avoid the "gamble" represented by the random factor ε. We have for an arbitrary von Neumann-Morgenstern utility function the following approximation

$$u_1(\bar{c}^{(1)} + \varepsilon) = u_1(\bar{c}^{(1)}) + u_1'(c^{(1)})\varepsilon + \frac{1}{2}u_1''(c^{(1)})\varepsilon^2 + O(\varepsilon^3) \tag{47}$$

and since $\mathbb{E}[\varepsilon] = 0$ one gets

$$\mathbb{E}[u_1(\bar{c}^{(1)} + \varepsilon)] \sim u_1(\bar{c}^{(1)}) + \frac{1}{2}u''(c^{(1)})\mathbb{E}[\varepsilon^2].$$

up to $O(\varepsilon^3)$ remainder. Recall that the variance of ε is given by

$$\mathrm{var}[\varepsilon] = \mathbb{E}[\varepsilon^2] - (\mathbb{E}[\varepsilon])^2 = \mathbb{E}[\varepsilon^2].$$

Therefore

$$\mathbb{E}[u_1(\bar{c}^{(1)} + \varepsilon)] \sim u_1(\bar{c}^{(1)}) + \frac{1}{2}u''(c^{(1)})\mathrm{var}[\varepsilon].$$

On the other hand

$$u_1(\bar{c}^{(1)} - \Pi) \sim u_1(\bar{c}^{(1)}) - u'(\bar{c}^{(1)})\Pi$$

up to second order terms in Π. Then, we obtain the approximation

$$\Pi \sim -\frac{1}{2}\frac{u''(\bar{c}^{(1)})}{u'(\bar{c}^{(1)})}\mathrm{var}[\varepsilon] \tag{48}$$

or, in terms of the risk aversion coefficient

$$\Pi \sim -\frac{1}{2}a\,\mathrm{var}[\varepsilon] \tag{49}$$

One deduces from (49) that

i. If $u_1''(c^{(1)}) > 0$ then $C > u(c^{(1)})$ and the investor prefers a greater certainty equivalent on an uncertain investiment (that is, a "gamble") than its expected value.

ii. If $u_1''(c^{(1)}) < 0$ then $C < u(c^{(1)})$ and the investor prefers a lower certainty equivalent on an uncertain investiment than its expected value.

iii. If $u_1''(c^{(1)}) = 0$ then $C = u(c^{(1)})$ and the investor has a certainty equivalent equal to the expected return.

In the special case of a CARA (constant absolute risk aversion) utility function one has

$$\frac{u_1''(\bar{c}^{(1)})}{u_1'(\bar{c}^{(1)})} = -a$$

for some constant a. Hence u can be taken as

$$u_1(\bar{c}^{(1)}) = -\exp(-a\bar{c}^{(1)}). \tag{50}$$

In this case we have

$$u_1(\bar{c}^{(1)} - \Pi) = -\exp(-a\bar{c}^{(1)} + a\Pi) = -\exp(-a\bar{c}^{(1)})\exp(a\Pi).$$

We obtain in a similar way that

$$u_1(\bar{c}^{(1)} + \varepsilon) = -\exp(-a\bar{c}^{(1)})\exp(-a\varepsilon)$$

from what follows that

$$\mathbb{E}[u_1(\bar{c}^{(1)} + \varepsilon)] = -\exp(-a\bar{c}^{(1)})\mathbb{E}[\exp(-a\varepsilon)]$$

Comparing both expressions, one obtains

$$\exp(a\Pi) = \mathbb{E}[\exp(-a\varepsilon)].$$

At this point, suppose that the random deviations ε are normally distributed with mean μ and covariance matrix Σ. Using that $\mathbb{E}[\varepsilon] = 0$ and denoting $\sigma^2 = \mathrm{var}[\varepsilon]$ one has

$$\begin{aligned}
\mathbb{E}[\exp(-a\varepsilon)] &= \frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\int \exp\Big(-a\varepsilon - \frac{1}{2}\frac{1}{\sigma^2}\varepsilon^2\Big)\,\mathrm{d}\varepsilon \\
&= \frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\int \exp\Big(-\frac{1}{2}\Big(\frac{1}{\sigma}\varepsilon + a\sigma\Big)^2\Big)\exp\Big(\frac{1}{2}a^2\sigma^2\Big)\mathrm{d}\varepsilon \\
&= \frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\int \exp\Big(-\frac{1}{2}t^2\Big)\exp\Big(\frac{1}{2}a^2\sigma^2\Big)\sigma\mathrm{d}t \\
&= \exp\Big(\frac{1}{2}a^2\sigma^2\Big)\frac{1}{(2\pi)^{\frac{1}{2}}}\int \exp\Big(-\frac{1}{2}t^2\Big)\mathrm{d}t = \exp\Big(\frac{1}{2}a^2\sigma^2\Big)
\end{aligned}$$

We conclude that in the case of CARA utility functions and normally distributed increments ε, we have exactly

$$\Pi = \frac{1}{2} a \operatorname{var}[\varepsilon]. \tag{51}$$

The certainty equivalent in this case is given by

$$C = \mu - \frac{1}{2} a \sigma^2, \tag{52}$$

where

$$\mu = \mathbb{E}[c^{(1)}].$$

2.2 Optimal portfolios

We refer the reader to (6), (16), (18) for further details on the notions and fundamental results in this section.

Suppose that $c^{(0)} = 0$. We recast the optimization problem 14, 15-16 in the context of expected utility function as

$$\max_{\boldsymbol{\alpha}'} \mathbb{E}[u_1(c^{(1)})] \tag{53}$$

subject to

$$\mathsf{e}^{(0)} = \langle Q, \boldsymbol{\alpha}' \rangle, \tag{54}$$

$$c^{(1)} = \mathsf{e}^{(1)} + \boldsymbol{\alpha}' D. \tag{55}$$

Suppose that the endowment at time $t = 1$ lies in the asset span $\mathcal{M}$. Then there exists a portfolio $\boldsymbol{\alpha}_0$ such that

$$\mathsf{e}^{(1)} = \boldsymbol{\alpha}_0 D.$$

and the price of this portfolio generating $\mathsf{e}^{(1)}$ is given by

$$\langle Q, \boldsymbol{\alpha}_0 \rangle.$$

Denoting

$$\boldsymbol{\alpha}' = \boldsymbol{\alpha} - \boldsymbol{\alpha}_0$$

we write 54 and 55 respectively as

$$\mathsf{e}^{(0)} + \langle Q, \boldsymbol{\alpha}_0 \rangle = \langle Q, \boldsymbol{\alpha} \rangle$$

and

$$c^{(1)} = \boldsymbol{\alpha}_0 D + \boldsymbol{\alpha}' D = (\boldsymbol{\alpha}_0 + \boldsymbol{\alpha}') D = \boldsymbol{\alpha} D.$$

In this case, we define the *wealth* of the investor by

$$w^{(0)} = \mathsf{e}^{(0)} + \langle Q, \boldsymbol{\alpha}_0 \rangle \tag{56}$$

and recast the problem 53, 54-55 as

$$\max_{\boldsymbol{\alpha}} \mathbb{E}[u_1(\boldsymbol{\alpha} D)] \tag{57}$$

subject to

$$w^{(0)} = \langle Q, \boldsymbol{\alpha} \rangle. \tag{58}$$

According to this formulation of the problem, we can regard the portfolio payoff $\boldsymbol{\alpha} D$ as the investor's wealth at time $t = 1$, that is,

$$w^{(1)} = \boldsymbol{\alpha} D. \tag{59}$$

The Lagrangian associated to 57-58 is of the form

$$L = \mathbb{E}[u_1(\boldsymbol{\alpha} D)] - \lambda(\langle Q, \boldsymbol{\alpha} \rangle - w^{(0)}).$$

The first-order condition may be formally written as

$$\mathbb{E}[u_1'(w^{(1)})D_i] = \lambda Q_i \tag{60}$$

for all i. Since

$$\frac{D_i - Q_i}{Q_i} = \mathbf{r}_i =: R_i - 1$$

after dividing both sides by λQ_i one gets

$$\mathbb{E}\left[\frac{1}{\lambda} u_1'(w^{(1)})R_i\right] = 1 \tag{61}$$

for all i. Therefore, given two assets with returns R_i and R_j we have

$$0 = \mathbb{E}\left[\frac{1}{\lambda} u_1'(w^{(1)})R_i\right] - \mathbb{E}\left[\frac{1}{\lambda} u_1'(w^{(1)})R_j\right] = \mathbb{E}\left[\frac{1}{\lambda} u_1'(w^{(1)})(R_i - R_j)\right].$$

It follows that

$$\mathbb{E}\left[\frac{1}{\lambda} u_1'(w^{(1)})(R_i - R_j)\right] = 0.$$

Choosing one of the assets to be risk-free, say setting $R_j = R_{\mathsf{f}}$, the risk-free return rate, and obtain

$$\mathbb{E}\left[\frac{1}{\lambda} u_1'(w^{(1)})(R_i - R_{\mathsf{f}})\right] = 0 \tag{62}$$

The random variable $R_i - R_{\mathsf{f}}$ is the *excess return* of the asset z_i compared with the risk-free return. Applying 60 to a risk-free asset with return R_{f} (that is, setting $R_i = R_{\mathsf{f}}$ in 60) one has

$$R_{\mathsf{f}}\,\mathbb{E}\big[u_1'(w^{(1)})\big] = \mathbb{E}\big[u_1'(w^{(1)})R_{\mathsf{f}}\big] = \lambda$$

This determines λ in 60. Hence, considering again the expression 60 for an arbitrary R_i gives

$$\mathbb{E}\left[\frac{1}{\lambda}u_1'(w^{(1)})D\right] = Q \tag{63}$$

is a stochastic version of the price law

$$D\,\pi = Q.$$

Then we refer to

$$m := \frac{1}{\lambda}u_1'(w^{(1)}) = \frac{1}{R_{\mathsf{f}}}\frac{u_1'(w^{(1)})}{\mathbb{E}\big[u_1'(w^{(1)})\big]}$$

as a *stochastic discount factor* playing a role similar to the state-price vector π. It follows from 61 that

$$\mathbb{E}[mR_i] = 1. \tag{64}$$

On the other hand

$$\mathbb{E}[mR_i] = \mathbb{E}[m]\mathbb{E}[R_i] + \operatorname{cov}[m, R_i]. \tag{65}$$

Applying 64 to $R_i = R_{\mathsf{f}}$ one has

$$\mathbb{E}[m]R_{\mathsf{f}} = 1,$$

that is,

$$\mathbb{E}[m] = \frac{1}{R_{\mathsf{f}}}\cdot$$

This determines $\mathbb{E}[m]$ in terms of R_{f}. Replacing this expression in 65 yields

$$\frac{1}{R_{\mathsf{f}}}\mathbb{E}[R_i] + \operatorname{cov}[m, R_i] = 1.$$

Rearranging terms, one obtains

$$\mathbb{E}[R_i] - R_{\mathsf{f}} = -R_{\mathsf{f}}\operatorname{cov}[m, R_i] \tag{66}$$

In order to fix ideas, we are going to perform some explicit calculation in the case of CARA utility function and normally distributed returns. We have in this particular setting that

$$u_1(c^{(1)}) = u_1(w^{(1)}) = -\exp(-a\,\boldsymbol{\alpha}D).$$

where we have used 50 and 59. We denote the expected payoff of the portfolio and its variance respectively by μ and σ^2. Then

$$\mu = \mathbb{E}[\boldsymbol{\alpha} D] = \boldsymbol{\alpha}\bar{D} \quad \text{and} \quad \sigma^2 = \boldsymbol{\alpha}\,\Sigma\,\boldsymbol{\alpha}^\top,$$

where $\bar{D}$ and Σ are, respectively, the mean vector and the covariance matrix of the payoffs D. Denoting $\boldsymbol{\alpha} D = z$ one has as above

$$\begin{aligned}
&\mathbb{E}[u_1(w^{(1)})] = \mathbb{E}[-\exp(-a\,\boldsymbol{\alpha} D)] \\
&= -\frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\int \exp(-az)\exp\Big(-\frac{1}{2}\frac{1}{\sigma^2}(z-\mu)^2\Big)\,\mathrm{d}z \\
&= -\frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\int \exp\Big(-\frac{1}{2}\Big(\frac{z-\mu}{\sigma}+a\sigma\Big)^2\Big)\exp\Big(\frac{1}{2}a^2\sigma^2-a\mu\Big)\mathrm{d}z \\
&= -\frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\int \exp\Big(-\frac{1}{2}t^2\Big)\exp\Big(\frac{1}{2}a^2\sigma^2-a\mu\Big)\sigma\mathrm{d}t \\
&= -\exp\Big(\frac{1}{2}a^2\sigma^2-a\mu\Big)\frac{1}{(2\pi)^{\frac{1}{2}}}\int \exp\Big(-\frac{1}{2}t^2\Big)\mathrm{d}t = -\exp\Big(\frac{1}{2}a^2\sigma^2-a\mu\Big) \\
&= -\exp\Big(-a\Big(\mu-\frac{1}{2}a\sigma^2\Big)\Big) = u_1\Big(\mu-\frac{1}{2}a\sigma^2\Big).
\end{aligned}$$

We conclude that the certainty equivalent of the random wealth $w^{(1)}$ is given by

$$C = \mu - \frac{1}{2}a\sigma^2, \tag{67}$$

that is,

$$C = \boldsymbol{\alpha}\bar{D} - \frac{1}{2}a\,\boldsymbol{\alpha}\Sigma\,\boldsymbol{\alpha}^\top, \tag{68}$$

where Σ is the covariance matrix of the payoffs of the basic assets $z_1, \ldots, z_N$. Hence the risk premium is completely determined by the variance of the portfolio and the absolute risk aversion coefficient:

$$\Pi = \frac{1}{2}a\,\boldsymbol{\alpha}\Sigma\,\boldsymbol{\alpha}^\top. \tag{69}$$

More importantly, we deduce that maximizing the CARA expected utility is equivalent to maximizing the utility of the certainty equivalent, which amounts to be equivalent to maximizing C itself. The corresponding first order condition is

$$\bar{D} - a\,\boldsymbol{\alpha}\Sigma = 0.$$

Therefore, the optimal portfolio is given by

$$\boldsymbol{\alpha} = \frac{1}{a}\Sigma^{-1}\bar{D}. \tag{70}$$

2.3 Mean-variance analysis

The earlier computations motivate the problem of maximizing the certainty equivalent. As we have seen, under the assumption of CARA utility function and normally distributed returns this problem reduces to the mean-variance criterion for the choice of optimal portfolios:

$$\max_{\boldsymbol{\alpha}} \left(\boldsymbol{\alpha} \bar{D} - \frac{1}{2} a \, \boldsymbol{\alpha} \, \Sigma \, \alpha^{\top} \right) \tag{71}$$

Then we consider the mean-variance problem formulated by Harry Markowitz (20):

$$\min_{\boldsymbol{\alpha}} \boldsymbol{\alpha} \, \Sigma \, \boldsymbol{\alpha}^{\top} \tag{72}$$

subject to

$$\langle \boldsymbol{\alpha}, D \rangle = \mu_* \tag{73}$$

$$\langle \boldsymbol{\alpha}, \mathbf{1} \rangle = 1, \quad \boldsymbol{\alpha} \geq 0, \tag{74}$$

where μ_* is the desired expected return for the portfolio. Note that here we are representing D as a N-dimensional random vector (depending on K states of the world instead of a $N \times K$ matrix. The second constraint implies that

$$\boldsymbol{\alpha} \in \mathcal{S}_{N-1} \subset \mathbb{R}^N$$

where $\mathcal{S}_{N-1}$ is the $(N-1)$-dimensional simplex defined by

$$\mathcal{S}_{N-1} = \left\{ \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N) \in \mathbb{R}^N : 0 \leq \alpha_i \leq 1 \text{ and } \sum_{i=1}^{N} \alpha_i = 1 \right\}$$

First, we consider the Lagrangian

$$L_\mu = \frac{1}{2} \boldsymbol{\alpha} \, \Sigma \, \boldsymbol{\alpha}^{\top} - \lambda(\langle \boldsymbol{\alpha}, D \rangle - \mu_*) \tag{75}$$

subject to the restriction

$$\langle \boldsymbol{\alpha}, \mathbf{1} \rangle = \sum_{i=1}^{N} \alpha_i = 1. \tag{76}$$

The first order condition is

$$\alpha_i = \lambda (D \Sigma^{-1})_i =: \lambda \psi_i.$$

Then, considering the restriction 76 one has

$$\lambda \sum_{i=1}^{N} \psi_i = \sum_{i=1}^{N} \alpha_i = 1.$$

Therefore
$$\lambda = \frac{1}{\sum_{i=1}^{N} \psi_i}.$$
We conclude that the optimal portfolio is given by
$$\boldsymbol{\alpha}_\mu = \frac{D\Sigma^{-1}}{\langle D\Sigma^{-1}, \mathbf{1}\rangle}. \tag{77}$$

Next, we define the Lagrangian
$$L_{\mathbf{1}} = \frac{1}{2}\boldsymbol{\alpha}\,\Sigma\,\boldsymbol{\alpha}^\top - \nu(\langle \boldsymbol{\alpha}, \mathbf{1}\rangle - 1) \tag{78}$$

subject to the constraint of prescribed expected portfolio return:
$$\langle \boldsymbol{\alpha}, D\rangle = \mu_*. \tag{79}$$

Now, the first order condition is
$$\alpha_i = \nu(\mathbf{1}\Sigma^{-1})_i =: \nu\phi_i.$$

Note that as before
$$\nu\sum_{i=1}^{N} \phi_i = \sum_{i=1} \alpha_i = 1$$
from what follows that the optimal portfolio for $L_{\mathbf{1}}$ is
$$\boldsymbol{\alpha}_{\mathbf{1}} = \frac{\mathbf{1}\Sigma^{-1}}{\langle \mathbf{1}\Sigma^{-1}, \mathbf{1}\rangle}. \tag{80}$$

In view of the restriction 79 one obtains
$$\nu\langle \boldsymbol{\phi}, D\rangle = \langle \boldsymbol{\alpha}, D\rangle = \mu_*$$

what implies that
$$\langle \mathbf{1}\Sigma^{-1}, D\rangle\,\nu = \mu_* \tag{81}$$
Similarly one gets
$$\langle D\Sigma^{-1}, D\rangle\,\lambda = \mu_*. \tag{82}$$

It follows that an optimal portfolio for the full Lagrangian
$$L = \frac{1}{2}\boldsymbol{\alpha}\,\Sigma\,\boldsymbol{\alpha}^\top - \lambda(\boldsymbol{\alpha}D - \mu_*) - \nu(\langle \boldsymbol{\alpha}, \mathbf{1}\rangle - 1)$$

can be expressed as
$$\boldsymbol{\alpha} = \widetilde{\lambda}\boldsymbol{\alpha}_\mu + \widetilde{\nu}\boldsymbol{\alpha}_{\mathbf{1}}, \tag{83}$$

that is,

$$\boldsymbol{\alpha} = \widetilde{\lambda}\frac{D\Sigma^{-1}}{\langle D\Sigma^{-1}, \mathbf{1}\rangle} + \widetilde{\nu}\frac{\mathbf{1}\Sigma^{-1}}{\langle \mathbf{1}\Sigma^{-1}, \mathbf{1}\rangle} \tag{84}$$

where the coefficients $\widetilde{\lambda}$ and $\widetilde{\nu}$ are solutions of the system

$$\widetilde{\lambda} + \widetilde{\nu} = 1, \tag{85}$$

$$\widetilde{\lambda}\frac{\langle D\Sigma^{-1}, D\rangle}{\langle D\Sigma^{-1}, \mathbf{1}\rangle} + \widetilde{\nu}\frac{\langle \mathbf{1}\Sigma^{-1}, D\rangle}{\langle \mathbf{1}\Sigma^{-1}, \mathbf{1}\rangle} = \mu_* \tag{86}$$

Arbitrary choices of $\widetilde{\lambda}$ and $\widetilde{\nu}$ determine the mean-variance frontier. For instance, both portfolios $\boldsymbol{\alpha}_\mu$ and $\boldsymbol{\alpha}_{\mathbf{1}}$ are in the mean-variance frontier since they correspond to set $\widetilde{\lambda} = 1$, $\widetilde{\nu} = 0$ and $\widetilde{\lambda} = 0$, $\widetilde{\nu} = 1$, respectively.

It follows from (83) and (85)-(86) that

$$\widetilde{\lambda}(\langle \boldsymbol{\alpha}_\mu, D\rangle - \langle \boldsymbol{\alpha}_{\mathbf{1}}, D\rangle) + \langle \boldsymbol{\alpha}_{\mathbf{1}}, D\rangle = \mu_*.$$

Hence

$$\widetilde{\lambda} = \frac{\mu_* - \langle \boldsymbol{\alpha}_{\mathbf{1}}, D\rangle}{\langle \boldsymbol{\alpha}_\mu, D\rangle - \langle \boldsymbol{\alpha}_{\mathbf{1}}, D\rangle}. \tag{87}$$

Note that the variance of $\boldsymbol{\alpha}$ is given by

$$\begin{aligned}\sigma^2 &= \boldsymbol{\alpha}\Sigma\,\boldsymbol{\alpha}^\top = (\widetilde{\lambda}\boldsymbol{\alpha}_\mu + \widetilde{\nu}\boldsymbol{\alpha}_{\mathbf{1}})\Sigma(\widetilde{\lambda}\boldsymbol{\alpha}_\mu + \widetilde{\nu}\boldsymbol{\alpha}_{\mathbf{1}})^\top \\ &= \widetilde{\lambda}^2\frac{\langle D\Sigma^{-1}, D\rangle}{\langle D\Sigma^{-1}, \mathbf{1}\rangle^2} + 2\widetilde{\lambda}(1-\widetilde{\lambda})\frac{\langle D, \mathbf{1}\Sigma^{-1}\rangle}{\langle D\Sigma^{-1}, \mathbf{1}\rangle\langle \mathbf{1}\Sigma^{-1}, \mathbf{1}\rangle} + (1-\widetilde{\lambda}^2)\frac{1}{\langle \mathbf{1}\Sigma^{-1}, \mathbf{1}\rangle}\end{aligned}$$

Denoting

$$A = \langle D\Sigma^{-1}, D\rangle, \quad B = \langle \mathbf{1}\Sigma^{-1}, D\rangle, \quad C = \langle \mathbf{1}\Sigma^{-1}, \mathbf{1}\rangle \tag{88}$$

one has

$$\sigma^2 = \frac{A - 2B\mu_* + C\mu_*^2}{AC - B^2}.$$

Therefore, the mean-variance frontier is parameterized in terms of $\widetilde{\lambda}$ as the geometric locus

$$\left(\mu_*, \sigma^2 = \frac{A - 2B\mu_* + C\mu_*^2}{AC - B^2}\right).$$

Fixed a desired expected return μ_*, the minimum variance efficient portfolio is determined differentiating the expression above for σ^2 with respect to $\widetilde{\lambda}$. We obtain

$$\mu_* = \frac{B}{C} = \frac{\langle \mathbf{1}\Sigma^{-1}, D\rangle}{\langle \mathbf{1}\Sigma^{-1}, \mathbf{1}\rangle}$$

what means that $\widetilde{\lambda} = 0$ and that the minimum variance efficient portfolio is indeed

$$\boldsymbol{\alpha}_{\mathbf{1}} = \frac{\mathbf{1}\Sigma^{-1}}{\langle \mathbf{1}\Sigma^{-1}, \mathbf{1} \rangle} \tag{89}$$

with return μ_*, the optimal return rate in Markowitz's portfolio selection theory.

2.4 Projection pricing

We refer the reader to (6), (12), (16), (18), (19) for further details on the notions and fundamental results in this section. The formalism below relies on orthogonal projections in a vector space and provides a geometric setting to the pricing theory of financial assets.

Let $Z = Z(s)$ be an arbitrary contingent claim, depending on $s \in \Omega$ but not necessarily contained in the asset span $\mathcal{M}$. We may define the orthogonal projection $z = z(s)$ of Z onto $\mathcal{M}$ using the L^2-inner product

$$(Z, Z') = \mathbb{E}[ZZ'],$$

which is well-defined for any random variables Z, Z' with finite variance, that is, whenever

$$\mathbb{E}[Z^2] < +\infty, \quad \mathbb{E}[Z'^2] < +\infty.$$

Since $z \in \mathcal{M}$ there exists $\boldsymbol{\alpha} \in \mathbb{R}^N$ such that

$$z = \boldsymbol{\alpha} D = \sum_{i=1}^{N} \alpha_i z_i.$$

Hence,

$$(Z - \boldsymbol{\alpha} D, z_j) = 0$$

for every $j = 1, \ldots, N$, what implies that

$$\alpha_i = \sum_{j=1}^{N} g_{ij}^{-1}(Z, Z_j),$$

where the matrix g is given by

$$g_{ij} = (z_i, z_j) = \mathbb{E}[z_i z_j]$$

and it is suppose to be invertible. Hence,

$$z = \sum_{j=1}^{N} g_{ij}^{-1}(Z, z_i) z_j \tag{90}$$

and

$$Z = z + \varepsilon,$$

where

$$(\varepsilon, z_j) = \mathbb{E}[\varepsilon z_j] = 0,$$

for every $j = 1, \ldots, N$. We may verify that z is the solution of the least squares problem

$$\min_{z \in \mathcal{M}} (Z - z, Z - z) = \min_{z \in \mathcal{M}} \mathbb{E}[(Z - z)^2].$$

We now consider the particular case when $Z = m$, a stochastic discount factor. Recall that this means that the price of an arbitrary asset Z is given by

$$Q(z) = \mathbb{E}[mZ] = (m, Z).$$

In this case,

$$(m, z_i) = Q_i$$

is the price of the basic asset z_i and the projection of m onto the asset span is given by

$$k_{\mathsf{q}} = \sum_{j=1}^{N} g_{ij}^{-1} Q_i z_j \tag{91}$$

We obviously have

$$Q(z) = \mathbb{E}[k_{\mathsf{q}} z] = (k_{\mathsf{q}}, z), \tag{92}$$

for every $z \in \mathcal{M}$. We refer to k_{q} as a *pricing kernel*. In mathematical terms, this means that k_{q} is the vector in $\mathcal{M}$ correspondent to the price functional $Q(z)$ through the inner product g.

Now, we determine the projection k_{e} onto $\mathcal{M}$ of a risk-free asset $\mathbf{1}$. We have for any contingent claim Z that

$$\mathbb{E}[Z] = \mathbb{E}[Z\mathbf{1}] = (Z, \mathbf{1})$$

and

$$\mathrm{var}[Z] = \mathbb{E}[(Z - \mathbb{E}[Z])^2] = (Z - (Z, \mathbf{1})\mathbf{1}, Z - (Z, \mathbf{1})\mathbf{1}).$$

We conclude that the variance of Z is the squared norm of the projection of Z onto the subspace orthogonal to $\mathbf{1}$ whereas the component of Z in the direction of $\mathbf{1}$ is the expected value of Z. Since

$$(\mathbf{1}, z_i) = \mathbb{E}[z_i]$$

we have

$$k_{\mathsf{e}} = \sum_{j=1}^{N} g_{ij}^{-1} \mathbb{E}[z_i] z_j$$

and

$$\mathbb{E}[z] = \mathbb{E}[k_{\mathsf{e}} z], \tag{93}$$

for every $z \in \mathcal{M}$. In what follows we refer to k_{e} as the *expectation kernel.*

The theorem below highlights the importance of both pricing and expectation returns.

Definition 2.2 *A traded asset $z \in \mathcal{M}$ is in the mean-variance frontier if there exists no $z' \in \mathcal{M}$ such that*

$$\mathbb{E}[z] = \mathbb{E}[z'], \quad Q(z) = Q(z') \tag{94}$$

and

$$\mathrm{var}[z] > \mathrm{var}[z']. \tag{95}$$

Theorem 2.5 *The mean-variance frontier payoff is spanned by the expectation and pricing kernels.*

Proof. Let $\mathcal{E}$ be the (one or two-dimensional) linear space spanned by the expectation kernel k_{e} and pricing kernel k_{q}. Given $z \in \mathcal{M}$, we have the orthogonal decomposition

$$z = z^{\mathcal{E}} + \varepsilon,$$

where $z^{\mathcal{E}}$ is a linear combination of the kernels $\{k_{\mathsf{e}}, k_{\mathsf{q}}\}$ and

$$0 = (k_{\mathsf{e}}, \varepsilon) = \mathbb{E}[k_{\mathsf{e}} \varepsilon] = \mathbb{E}[\varepsilon]$$
$$0 = (k_{\mathsf{q}}, \varepsilon) = \mathbb{E}[k_{\mathsf{q}} \varepsilon] = Q(\varepsilon)$$

This implies that

$$\mathbb{E}[z] = \mathbb{E}[z^{\mathcal{E}}],$$
$$Q(z) = Q(z^{\mathcal{E}}).$$

Moreover,

$$\mathbb{E}[z^2] = (z, z) = (z^{\mathcal{E}} + \varepsilon, z^{\mathcal{E}} + \varepsilon) = (z^{\mathcal{E}}, z^{\mathcal{E}}) + (\varepsilon, \varepsilon) = \mathbb{E}[(z^{\mathcal{E}})^2] + (\varepsilon, \varepsilon)$$

and

$$(\mathbb{E}[z])^2 = (\mathbb{E}[z^{\mathcal{E}}])^2.$$

Therefore

$$\mathrm{var}[z] = \mathbb{E}[z^2] - (\mathbb{E}[z])^2 = \mathbb{E}[(z^{\mathcal{E}})^2) - (\mathbb{E}[z^{\mathcal{E}}])^2 + (\varepsilon, \varepsilon)$$

what implies that

$$\mathrm{var}[z] = \mathrm{var}[(z^{\mathcal{E}})^2] + (\varepsilon, \varepsilon), \tag{96}$$

with equality if and only if $\varepsilon = 0$, that is, if and only if $z \in \mathcal{E}$. □

Note that the price and return of the pricing kernel are respectively given by

$$Q(k_{\mathsf{q}}) = \mathbb{E}[k_{\mathsf{q}} k_{\mathsf{q}}] = \mathbb{E}[k_{\mathsf{q}}^2] \tag{97}$$

and

$$R_{\mathsf{q}} = \frac{k_{\mathsf{q}}}{Q(k_{\mathsf{q}})} = \frac{k_{\mathsf{q}}}{\mathbb{E}[k_{\mathsf{q}}^2]}, \tag{98}$$

whereas the price and returns of the expectation kernel are given by

$$Q(k_{\mathsf{e}}) = \mathbb{E}[k_{\mathsf{q}} k_{\mathsf{e}}] = \mathbb{E}[k_{\mathsf{e}} k_{\mathsf{q}}] = \mathbb{E}[k_{\mathsf{q}}]$$

and

$$R_{\mathsf{e}} = \frac{k_{\mathsf{e}}}{\mathbb{E}[k_{\mathsf{q}}]},$$

respectively.

2.5 Beta models and Capital Asset Pricing Model

In what follows, we propose a version of the Capital Asset Pricing Model adapted to non-Gaussian distributions. The model described below relies essentially on a geometric definition of generalized beta coefficients. These betas are related to the Fisher metric defined by the Hessian of the cumulant function that plays the role of the variance in the case of non-Gaussian returns.

We refer the reader to (6), (12), (16), (18), (19) for introductory expositions of the classical Capital Asset Pricing Model (CAPM).

We have proved thet $\mathcal{E} = \mathrm{span}\{k_{\mathsf{q}}, k_{\mathsf{e}}\}$ is the mean-variance frontier in $\mathcal{M}$. Suppose that $\mathcal{E}$ is indeed a two-dimensional linear space. Now we address the problem of minimizing the variance among points in $z \in \mathcal{E}$, that is,

$$\min_{z \in \mathcal{E}} \mathrm{var}[z], \tag{99}$$

Any point $z \in \mathcal{E}$ is of the form

$$z = a k_{\mathsf{q}} + b k_{\mathsf{e}},$$

for some $a, b \in \mathbb{R}$. The price of this portfolio is

$$Q(z) = \mathbb{E}[mz] = a\mathbb{E}[m k_{\mathsf{q}}] + b\mathbb{E}[m k_{\mathsf{e}}] = aQ(k_{\mathsf{q}}) + bQ(k_{\mathsf{e}}) \tag{100}$$

Fixing the constraint that the price of the portfolio is $Q(z) = 1$, we denote

$$\beta = aQ(k_{\mathsf{q}})$$

and therefore

$$1 - \beta = bQ(k_{\mathsf{e}}).$$

Therefore the portfolios with unit price are parameterized by

$$z = \beta \frac{k_{\mathsf{q}}}{Q(k_{\mathsf{q}})} + (1-\beta)\frac{k_{\mathsf{e}}}{Q(k_{\mathsf{e}})} = \beta R_{\mathsf{q}} + (1-\beta)R_{\mathsf{e}} = R_{\mathsf{e}} + \beta(R_{\mathsf{q}} - R_{\mathsf{e}}) \tag{101}$$

with $\beta \in \mathbb{R}$. Here R_{q} and R_{e} are the returns of k_{q} and k_{e}, respectively. We conclude from (99) that the optimal portfolio with unit price is determined by

$$\beta_0 := -\frac{(R_{\mathsf{q}} - R_{\mathsf{e}})^\top \Sigma R_{\mathsf{e}}}{(R_{\mathsf{q}} - R_{\mathsf{e}})^\top \Sigma (R_{\mathsf{q}} - R_{\mathsf{e}})} \tag{102}$$

Note that the expected return of this portfolio is

$$\mathbb{E}[z_0] = \mathbb{E}[R_{\mathsf{e}}] + \beta_0\, \mathbb{E}[R_{\mathsf{q}} - R_{\mathsf{e}}] \tag{103}$$

We have in the case when the risk-free asset $\mathbf{1}$ with riskless return R_{f} is an element in $\mathcal{M}$ that

$$k_{\mathsf{e}} = \mathbf{1}$$

and $R_{\mathsf{e}} = R_{\mathsf{f}}$. We have in this case

$$\mathbb{E}[z_0] = R_{\mathsf{f}} + \beta_0\,(\mathbb{E}[R_{\mathsf{q}}] - R_{\mathsf{f}}). \tag{104}$$

In that way, we recover the classical beta pricing equation.

Generalized beta pricing

Recall that we are assuming that $\mathcal{E}$ has dimension two. It is then convenient to restate the results above using two linearly independent assets other than k_{e} and k_{q}. We fix such assets, say k_λ and k_ν, with respective returns

$$r_\lambda = R_{\mathsf{e}} + \lambda(R_{\mathsf{q}} - R_{\mathsf{e}})$$

and

$$r_\nu = R_{\mathsf{e}} + \mu(R_{\mathsf{q}} - R_{\mathsf{e}})$$

in such a way that

$$\operatorname{cov}(r_\lambda, r_\nu) = 0. \tag{105}$$

Hence, ν is given by

$$\nu = -\frac{\operatorname{cov}(R_{\mathsf{e}}, R_{\mathsf{e}}) + \lambda \operatorname{cov}(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{e}})}{\operatorname{cov}(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{e}}) + \lambda \operatorname{cov}(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{q}} - R_{\mathsf{e}})} \tag{106}$$

Note that ν is well-defined if and only if $\lambda \neq \beta_0$ in (102), that is, if k_λ is not the minimum variance portfolio in $\mathcal{E}$.

Given an asset $z \in \mathcal{M}$ with unit price we have the decomposition

$$z = z^{\mathcal{E}} + \varepsilon$$

where

$$z^{\mathcal{E}} = a k_\lambda + b k_\nu$$

with $\varepsilon \perp \mathcal{E}$ and $k_{\mathsf{e}}(\varepsilon) = \mathbb{E}[\varepsilon] = 0$. It follows that

$$\begin{aligned} \mathbb{E}[z] &= a\mathbb{E}[k_\lambda] + b\mathbb{E}[k_\mu] = a\mathsf{q}(k_\lambda)\mathbb{E}[r_\lambda] + b\mathsf{q}(k_\nu)\mathbb{E}[r_\nu] \\ &=: \mathbb{E}[r_\nu] + \beta(\mathbb{E}[r_\lambda] - \mathbb{E}[r_\nu]) \end{aligned}$$

with $\beta = a\mathsf{q}(k_\lambda)$. Denoting by r the return of z one obtains

$$r = z = a\mathsf{q}(k_\lambda) r_\lambda + b\mathsf{q}(k_\nu) r_\nu + \varepsilon = r_\nu + \beta(r_\lambda - r_\nu) + \varepsilon$$

from what follows that

$$\begin{aligned} \operatorname{cov}(r, r_\lambda) &= \operatorname{cov}(r_\mu, r_\lambda) + \beta \operatorname{cov}(r_\lambda - r_\nu, r_\lambda) + \operatorname{cov}(\varepsilon, r_\lambda) \\ &= \operatorname{cov}(r_\mu, r_\lambda) + \beta \operatorname{cov}(r_\lambda - r_\nu, r_\lambda) \\ &= \beta \operatorname{cov}(r_\lambda - r_\nu, r_\lambda). \end{aligned}$$

We conclude that

$$\beta = \frac{\operatorname{cov}(r, r_\lambda)}{\operatorname{cov}(r_\lambda - r_\nu, r_\lambda)} = \frac{\operatorname{cov}(r, r_\lambda)}{\operatorname{cov}(r_\lambda, r_\lambda)}.$$

In sum, we have obtained a generalized beta pricing equation

$$\mathbb{E}[z] = \mathbb{E}[r_\nu] + \beta(\mathbb{E}[r_\lambda] - \mathbb{E}[r_\nu]) \tag{107}$$

for assets in $z \in \mathcal{M}$. If the risk-free asset $\mathbf{1}$ with return R_{f} lies in the asset span $\mathcal{M}$, we fix $r_\nu = R_{\mathsf{f}}$. With this choice, (107) reduces to

$$\mathbb{E}[z] = R_{\mathsf{f}} + \beta(\mathbb{E}[r_\lambda] - R_{\mathsf{f}}), \tag{108}$$

a generalized beta pricing equation written in terms of an asset k_λ instead of the pricing kernel k_{q} as in (104).

In a market of M investors, the *market payoff* is by definition the projection of the aggregate endowment at time $t = 1$

$$\mathsf{e} := \sum_{\ell=1}^{M} \mathsf{e}_\ell^{(1)}$$

and the market return r_{m} is the market payoff divided by the equilibrium price of the market portfolio (that is, the portfolio that replicates in $\mathcal{M}$ the projection of e.)

An agent has mean-variance preferences if his utility function $u_1(c^1)$ is strictly increasing and has the representation

$$u_1(c^{(1)}) = v(\mathbb{E}[c^{(1)}], \mathrm{var}[c^{(1)}]) \tag{109}$$

where v is strictly increasing and concave with respect to the the second variable, that is, with respect to variance.

As a consequence of the existence of a market equilibrium (for mean-variance preferences), we have the central theorem of the Capital Asset Pricing Model (CAPM), that defines the *market security line*

Theorem 2.6 *The equilibrium prices for efficient assets z with returns r in a market with agents' preferences described by an utility function of the form (109) are given by*

$$\mathbb{E}[r] - \mathbb{E}[r_\nu] = \beta_{\mathsf{m}}(\mathbb{E}[r_{\mathsf{m}}] - \mathbb{E}[r_\nu]), \tag{110}$$

where r_{m} is the return of the market portfolio and

$$\beta_{\mathsf{m}} = \frac{\mathrm{cov}(r, r_{\mathsf{m}})}{\mathrm{cov}(r_{\mathsf{m}}, r_{\mathsf{m}})}. \tag{111}$$

The coefficient β measures the asset return's sensitivity to fluctuations in the market return: it is a measure of the systematic risk of the asset, that risk component that cannot be suppressed by diversification.

3 FUNDAMENTALS OF INFORMATION THEORY

The *entropy*, more precisely, Shannon's entropy, of a probability distribution P with density $p(s, \boldsymbol{\vartheta})$ is defined by

$$H(P) = -\int_\Omega p(s) \log p(s, \boldsymbol{\vartheta})\, \mathrm{d}s = \mathbb{E}[-\log p] \tag{112}$$

where $\mathrm{d}s$ is some fixed measure in the sample space Ω of states of the world. Here, ϑ stands for the statistical parameters of the distribution p. If $\Omega = \{s_1, \ldots, s_K\}$ and P is a discrete probability function defined by

$$P(s_i) = p_i$$

for every $i = 1, \ldots, K$, then the entropy reduces to

$$H(P) = -\sum_{i=1}^{K} \log(p_i)\, p_i = \mathbb{E}[-\log p]$$

Note that $H(P) \geq 0$ in both cases.

Intuitively, Shannon's entropy is a measure of average uncertainty in a random variable $\mathbf{r}$ with probability distribution P, represented in the discrete case as the number of *bits* needed to describe it.

In order to fix ideas, we will discuss a couple of examples. For instance, given an uniform distribution supported in $C = [a_1, b_1] \times \ldots \times [a_K, b_K]$ we have

$$p(s) = \begin{cases} \frac{1}{|C|}1, & \text{if } s \in C, \\ 0, & \text{otherwise,} \end{cases}$$

where

$$|C| = \operatorname{volume} C = (b_1 - a_1) \cdot \ldots \cdot (b_K - a_K).$$

Then

$$H(P) = -\int_C \frac{1}{|C|} \log\left(\frac{1}{|C|}\right) \mathrm{d}s = -\frac{|C|}{|C|} \log\left(\frac{1}{|C|}\right) = -\log\left(\frac{1}{|C|}\right) = \log |C|.$$

In the discrete case, one easily verifies that the uniform distribution has the greatest entropy among discrete probability distributions in a finite sample space. We also have

$$H(P) \leq \log |\Omega| = \log K$$

with the upper bound attained for the uniform distribution.

Now, given a multivariate normal distribution with density

$$p(\mathbf{r}, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{r} - \mu)\Sigma^{-1}(\mathbf{r} - \mu)^\top \right)$$

one has

$$\log p = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2}(\mathbf{r} - \mu)\Sigma^{-1}(\mathbf{r} - \mu)^\top$$

Setting $\Sigma^{-1} = AA^\top$ and denoting $\mathbf{w} = \mathbf{r} - \mu$ one computes

$$\begin{aligned}
H(P) &= \int -p \log p \, \mathrm{d}\mathbf{r} = \left(\frac{N}{2}\log(2\pi) + \frac{1}{2}\log|\Sigma|\right)\int p \, \mathrm{d}\mathbf{r} \\
&+ \frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}}\int \frac{1}{2}(\mathbf{r}-\mu)\Sigma^{-1}(\mathbf{r}-\mu)^\top \exp\left(\frac{1}{2}(\mathbf{r}-\mu)\Sigma^{-1}(\mathbf{r}-\mu)^\top\right)\mathrm{d}\mathbf{r} \\
&= \frac{N}{2}\log(2\pi) + \frac{1}{2}\log|\Sigma| + \frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}}\int \frac{1}{2}\mathbf{w}AA^\top\mathbf{w}^\top \exp\left(\frac{1}{2}\mathbf{w}AA^\top\mathbf{w}^\top\right)\mathrm{d}\mathbf{w} \\
&= \frac{N}{2}\log(2\pi) + \frac{1}{2}\log|\Sigma| + \frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}}\int \frac{1}{2}\mathbf{w}AA^\top\mathbf{w}^\top \exp\left(\frac{1}{2}\mathbf{w}AA^\top\mathbf{w}^\top\right)\mathrm{d}\mathbf{w}.
\end{aligned}$$

Now, considering the change of variables $\mathbf{z} = \mathbf{w}A$ one gets

$$\begin{aligned}
H(P) &= \frac{N}{2}\log(2\pi) + \frac{1}{2}\log|\Sigma| + \frac{1}{(2\pi)^{\frac{N}{2}}}\int \frac{1}{2}|\mathbf{z}|^2 \exp\left(\frac{1}{2}|\mathbf{z}|^2\right)\mathrm{d}\mathbf{z} \\
&= \frac{N}{2}\log(2\pi) + \frac{1}{2}\log|\Sigma| + \frac{1}{2}\left(\mathrm{var}[\mathbf{z}] + \mathbb{E}[\mathbf{z}]^2\right) \\
&= \frac{N}{2}\log(2\pi) + \frac{1}{2}\log|\Sigma| + \frac{1}{2}\mathrm{var}[\mathbf{z}] = \frac{N}{2}\log(2\pi) + \frac{1}{2}\log|\Sigma| + \frac{N}{2}.
\end{aligned}$$

Hence, we conclude that the entropy of a multivariate Gaussian distribution depends on its covariance matrix:

$$H(P) = \frac{N}{2}(1 + \log(2\pi)) + \frac{1}{2}\log|\Sigma|. \tag{113}$$

Moreover, it holds that the Gaussian distribution has the greatest entropy among the distributions with zero mean and covariance Σ, see (24).

Given a random variable $\mathbf{r}$ with distribution probability P, we refer to $H(P)$ also as the entropy of the random variable itself and denote

$$H(\mathbf{r}) = H(P)$$

is a slight abuse of notation. Hence, given two random variables $(\mathbf{r}, \mathbf{r}')$ whose joint distribution has density p, we define the joint entropy by

$$H(\mathbf{r}, \mathbf{r}') = -\int\int p(s, s') \log p(s, s') \, \mathrm{d}s \, \mathrm{d}s'. \tag{114}$$

In its discrete version, this expression reads as

$$H(\mathbf{r}, \mathbf{r}') = -\sum_{i=1}^{K}\sum_{j=1}^{K} p(s_i, s'_j) \log p(s_i, s'_j).$$

Therefore, using the decomposition of the total probability in conditional probabilities, one has

$$
\begin{aligned}
H(\mathbf{r},\mathbf{r}') &= -\sum_{j=1}^{K}\left(\sum_{i=1}^{K} p(s_i,s_j')\log\big(p(s_i|s_j')p(s_j')\big)\right)\\
&= -\sum_{j=1}^{K}\left(\sum_{i=1}^{K} p(s_i,s_j')\big(\log p(s_i|s_j') + \log p(s_j')\big)\right)\\
&= -\sum_{j=1}^{K}\sum_{i=1}^{K} p(s_i,s_j')\log p(s_i|s_j') - \sum_{j=1}^{K}\log p(s_j')\left(\sum_{i=1}^{K} p(s_i,s_j')\right).
\end{aligned}
$$

However, the last term between parenthesis is the marginal probability of $\mathbf{r}'$. Hence,

$$
\begin{aligned}
H(\mathbf{r},\mathbf{r}') &= -\sum_{j=1}^{K}\sum_{i=1}^{K} p(s_i,s_j')\log p(s_i|s_j') - \sum_{j=1}^{K}\log p(s_j')p(s_j')\\
&= -\sum_{j=1}^{K}\sum_{i=1}^{K} p(s_i,s_j')\log p(s_i|s_j') + H(\mathbf{r}').
\end{aligned}
$$

Defining the *conditional entropy* by

$$
H(\mathbf{r}|\mathbf{r}') = -\sum_{j=1}^{K}\sum_{i=1}^{K} p(s_i,s_j')\log p(s_i|s_j') \tag{115}
$$

which is natural since the expression involves the conditional probabilities, one obtains

$$
H(\mathbf{r},\mathbf{r}') = H(\mathbf{r}|\mathbf{r}') + H(\mathbf{r}'). \tag{116}
$$

Next, we define the *mutual information*

$$
I(\mathbf{r},\mathbf{r}') = H(\mathbf{r}) - H(\mathbf{r}|\mathbf{r}'). \tag{117}
$$

Alternatively, $I(\mathbf{r},\mathbf{r}')$ can be regarded as the *relative entropy* between the joint probability distribution $p(\mathbf{r},\mathbf{r}')$ and the product of the marginal distributions $p(\mathbf{r})p(\mathbf{r}')$, that is,

$$
I(\mathbf{r},\mathbf{r}') = \sum_{i=1}^{K}\sum_{j=1}^{K} p(s_i,s_j')\log\left(\frac{p(s_i,s_j')}{p(s_i)p(s_j')}\right).
$$

In particular, if $\mathbf{r}$ and $\mathbf{r}'$ are independent random variables, then

$$
I(\mathbf{r},\mathbf{r}') = 0.
$$

It turns out that

$$
I(\mathbf{r},\mathbf{r}') \geq 0
$$

with equality only in the case of independent random variables.

In general, if we intuitively take H as a measure of uncertainty of the random variable, $I(\mathbf{r}, \mathbf{r}')$ represent how much information about $\mathbf{r}$ comes out from the knowledge of $\mathbf{r}'$: knowing $\mathbf{r}'$ may reduce the ignorance about $\mathbf{r}$ from the level $H(\mathbf{r})$ to a possibly lower level of uncertainty $H(\mathbf{r}|\mathbf{r}')$. This content of information is quantified by the mutual information.

More generally, one defines the *relative entropy* between random variables $\mathbf{r}, \mathbf{r}'$ (to be precise, between their probability distributions p and p', respectively) by

$$D(\mathbf{r}||\mathbf{r}') = \sum_{i=1}^{K}\sum_{j=1}^{K} p(s_i)\log\left(\frac{p(s_i)}{p(s'_j)}\right) \tag{118}$$

We also refer to this quantity as *Kullback-Leibler statistical divergence*. Since

$$D(\mathbf{r}||\mathbf{r}') \geq 0$$

with equality if and only if $p(s_i) = p(s'_i)$ for every $i = 1, \ldots, K$, it seems suggestive to consider the Kullback-Leibler divergence as a *distance* between probability distributions. We will discuss this point of view in further details in the following section.

3.1 Statistical divergences and Fisher metric

A *statistical manifold* can be thought of as a set $\mathcal{S}$ of probability distributions whose densities $p(s, \boldsymbol{\vartheta})$, $s \in \Omega\, \mathrm{d}s$, depend on a list of parameters $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_n)$ varying along an open subset $U \subset \mathbb{R}^n$. Formally, we have a parametrization

$$\boldsymbol{\vartheta} \in U \mapsto p(s, \boldsymbol{\vartheta})\, \mathrm{d}s \in \mathcal{S}.$$

Precise definitions may be found for instance in (1), (15).

The geometry in a statistical manifold can be introduced through the notion of a distance between probability distributions: the relevant notion here is that of a statistical *divergence*

$$D(P||P') = D(\boldsymbol{\vartheta}||\boldsymbol{\vartheta}') \tag{119}$$

between probability distributions P and P' in $\mathcal{S}$ with respective densities given by $p(s, \boldsymbol{\vartheta})$ and $p(s, \boldsymbol{\vartheta}')$ for some $\boldsymbol{\vartheta}, \boldsymbol{\vartheta}'$ in U.

A trivial yet fundamental example is the Euclidean divergence that is, up to a constant, merely the Euclidean norm in U, that is,

$$D_{\mathrm{euc}}(P||P') = \frac{1}{2}|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'|^2 = \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}'_i)^2. \tag{120}$$

Another well-known example is the Kullback-Leibler divergence or relative entropy we have defined above

$$D_{KL}(P||P') = \mathbb{E}_P\left[\log\frac{p(\cdot,\boldsymbol{\vartheta})}{p(\cdot,\boldsymbol{\vartheta}')}\right] = \int p(s,\boldsymbol{\vartheta})\log\frac{p(s,\boldsymbol{\vartheta})}{p(s,\boldsymbol{\vartheta}')}\,\mathrm{d}s \tag{121}$$

Some important statistical manifolds are defined in terms of a convex real function $K : U \subset \mathbb{R}^n \to \mathbb{R}$. This is the case of the exponential family whose densities are of the form

$$p(\mathbf{r};\boldsymbol{\vartheta}) = \exp(\langle \mathbf{r},\boldsymbol{\vartheta}\rangle - K(\boldsymbol{\vartheta}))\,p_0(\mathbf{r}), \quad \boldsymbol{\vartheta}\in U, \tag{122}$$

where p_0 is a fixed reference density. Note that K is the moment-generating function of the distribution. Indeed, for each $i = 1,\dots,n$ we have

$$\begin{aligned}
\mathbb{E}[\mathbf{r}_i] &= \int \mathbf{r}_i p(\mathbf{r};\boldsymbol{\vartheta})\,\mathrm{d}\mathbf{r} \\
&= \int \mathbf{r}_i \exp(\langle \mathbf{r},\boldsymbol{\vartheta}\rangle - K(\boldsymbol{\vartheta}))p_o(\mathbf{r})\,\mathrm{d}\mathbf{r} \\
&= \int \left(\frac{\partial}{\partial\vartheta_i}p(\mathbf{r},\boldsymbol{\vartheta}) + \frac{\partial K}{\partial\vartheta_i}p(\mathbf{r},\boldsymbol{\vartheta})\right)\mathrm{d}\mathbf{r} \\
&= \frac{\partial}{\partial\vartheta_i}\int p(\mathbf{r},\boldsymbol{\vartheta})\,\mathrm{d}\mathbf{r} + \frac{\partial K}{\partial\vartheta_i}\int p(\mathbf{r},\boldsymbol{\vartheta})\,\mathrm{d}\mathbf{r} = \frac{\partial K}{\partial\vartheta_i},
\end{aligned}$$

where we used the fact that

$$\int p(\mathbf{r};\boldsymbol{\vartheta})\,\mathrm{d}\mathbf{r} = 1.$$

We conclude that

$$\mathbb{E}[\mathbf{r}] = \nabla K(\boldsymbol{\vartheta}). \tag{123}$$

Similarly, one computes

$$\mathrm{cov}(\mathbf{r}_i,\mathbf{r}_j) = \frac{\partial^2 K}{\partial\vartheta_i\partial\vartheta_j}$$

and so on. The natural divergence in the case of such a family is the *Bregman* divergence defined in terms of K by

$$D(P||P') = D(\boldsymbol{\vartheta}||\boldsymbol{\vartheta}') = K(\boldsymbol{\vartheta}) - K(\boldsymbol{\vartheta}') - \langle\nabla K(\boldsymbol{\vartheta}'),\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'\rangle \tag{124}$$

The Taylor expansion of an arbitrary divergence has the form

$$D(\boldsymbol{\vartheta}||\boldsymbol{\vartheta}') = \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n g_{ij}(\boldsymbol{\vartheta})(\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}'_i)(\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}'_j) + O(|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'|^3). \tag{125}$$

The positive-definite matrix $(g_{ij})_{i,j=1}^n$ is the *Fisher information* metric defined by K and then associated to the Bregman divergence (124) as the second order term in the Taylor expansion of D.

The parameters $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_n) \in U$ define coordinates in the statistical manifold $\mathcal{S}$. An associated dual system of coordinates is provided by

$$\boldsymbol{\eta} = \nabla K(\boldsymbol{\vartheta}) = \left(\frac{\partial K}{\partial \vartheta_1}, \dots, \frac{\partial K}{\partial \vartheta_n}\right). \tag{126}$$

At this point, it is worth to define the dual function $K^*(\boldsymbol{\eta})$ given by the Legendre transform of K:

$$K^*(\boldsymbol{\eta}) = \max_{\boldsymbol{\vartheta}} \big(\langle \boldsymbol{\vartheta}, \boldsymbol{\eta} \rangle - K(\boldsymbol{\vartheta})\big) \tag{127}$$

Differentiating the right-hand side at a maximum point we recover (126). The dual Bregman divergence is given by

$$D^*(\boldsymbol{\eta}||\boldsymbol{\eta}') = K^*(\boldsymbol{\eta}) - K^*(\boldsymbol{\eta}') - \langle \nabla K^*(\boldsymbol{\eta}'), \boldsymbol{\eta} - \boldsymbol{\eta}' \rangle.$$

It is easy to verify that

$$D(\boldsymbol{\vartheta}||\boldsymbol{\vartheta}') = D^*(\boldsymbol{\eta}'||\boldsymbol{\eta}) \tag{128}$$

if

$$\boldsymbol{\eta} = \nabla K(\boldsymbol{\vartheta}), \quad \boldsymbol{\eta}' = \nabla K(\boldsymbol{\vartheta}').$$

It also follows that the primal Bregman divergence may be written as

$$D(\boldsymbol{\vartheta}||\boldsymbol{\vartheta}') = K(\boldsymbol{\vartheta}) + K^*(\boldsymbol{\eta}) - \langle \boldsymbol{\vartheta}, \boldsymbol{\eta} \rangle. \tag{129}$$

Straight lines of the form

$$\boldsymbol{\gamma}(t) = \boldsymbol{\vartheta}' + t(\boldsymbol{\theta} - \boldsymbol{\vartheta}') \tag{130}$$

and

$$\boldsymbol{\gamma}^*(t) = \boldsymbol{\eta}' + t(\boldsymbol{\eta} - \boldsymbol{\eta}') \tag{131}$$

define locally shortest paths in $\mathcal{S}$ by means of the parameterizations given by $\boldsymbol{\vartheta}$ and $\boldsymbol{\eta}$, respectively. Those shortest curves are called, respectively, *geodesics* and *dual geodesics*. A geodesic and a (dual) geodesic intersect orthogonally at a point $p = \gamma(0) = \eta(0)$ if their velocity tangent vectors satisfy

$$\langle \gamma'(0), \eta'(0) \rangle = 0,$$

that is,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}(p) \gamma_i'(0) \eta_j'(0) = 0.$$

Now, we may state the fundamental Pytaghorean theorem in the context of the Information Geometry of statistical divergence.

Theorem 3.1 *Given three points P, P' and Q in $\mathcal{S}$ such that the dual geodesic connecting*

P and P' is orthogonal to the geodesic connecting P' and Q, we have

$$D(P||Q) = D(P||P') + D(P'||Q). \tag{132}$$

The dual theorem is stated as follows

Theorem 3.2 *Given three points P, P' and Q in $\mathcal{S}$ such that the geodesic connecting P and P' is orthogonal to the dual geodesic connecting P' and Q, we have*

$$D^*(P||Q) = D^*(P||P') + D^*(P'||Q). \tag{133}$$

3.2 Deformed exponentials

In this section we fix some notation and recall basic notions and facts about deformed exponentials and logarithms. We consider a strictly positive, nondecreasing and continuous real function $\phi : (0, +\infty) \to (0, +\infty)$ and define the ϕ-logarithm as

$$\log_\phi t = \int_1^t \frac{ds}{\phi(s)}, \quad t > 0. \tag{134}$$

Therefore $\log_\phi$ is invertible and its inverse is denoted by $\exp_\phi$. We refer to the function $\exp_\phi$ as the *deformed exponential* defined by ϕ. Following (26) we also define

$$\psi(t) = \phi\big(\exp_\phi(t)\big). \tag{135}$$

In the case of $\phi(t) = t^q$ with $q > 0$ we obtain the q-exponential (25) as

$$\exp_q(t) = \big(1 + (1-q)t\big)^{\frac{1}{1-q}}. \tag{136}$$

This q-exponential function is suitable to describe non-additive processes since

$$\exp_q(t)\exp_q(t') = \exp_q(t + t' + (1-q)tt').$$

This motivates the definition of the q-sum

$$t \oplus t' = t + t' + (1-q)tt'. \tag{137}$$

The usual exponential function may be recovered as the limit of $\exp_q$ as $q \to 1$. Other examples of deformed exponentials in the literature are the Kaniadakis exponential given as

$$\exp_{\mathsf{k}}(t) = \exp \int_0^t \frac{\mathrm{d}\varsigma}{\sqrt{1 + q^2\varsigma^2}}, \quad \text{for } 0 \le q \le 1, \tag{138}$$

and the Newton exponential

$$\exp_{\mathsf{n}}(t) = \left(\frac{1+qt}{1-qt}\right)^{1/2q}, \quad \text{for } -\frac{1}{q} < x < \frac{1}{q} \quad \text{and } 0 \le q \le 1. \tag{139}$$

Another model was proposed by Vigelis and Cavalcante where the Naudts' model for deformed exponential was extended to a generic reference density. We refer the reader to (1), (24), (40), (7) and references therein for a detailed account on the analytical properties and applications of ϕ-exponentials and logarithms.

Example 3.1 *Our definitions and conventions for q-Gaussian multidimensional distribution follow closely the ones presented in* (25), (33). *Given a vector $\mu \in \mathbb{R}^N$ and a positive definite $N \times N$ matrix Σ, a q-Gaussian multivariate distribution with parameters (μ, Σ) is defined by*

$$p(\mathbf{r};\mu,\Sigma) = \frac{1}{\Gamma_q}\exp_q\left(-\frac{1}{2\gamma_q}(\mathbf{r}-\mu)\Sigma^{-1}(\mathbf{r}-\mu)^\top\right), \tag{140}$$

with

$$\gamma_q = \frac{1}{2}\big((N+4)-(N+2)q\big),$$

and

$$\Gamma_q = K_{q,N}\sqrt{|\Sigma|},$$

where

$$K_{q,N} = \begin{cases} \frac{\Gamma(\frac{1}{q-1}-\frac{N}{2})\sqrt{\pi}}{\Gamma(\frac{1}{q-1})}\left(\frac{1}{q-1}\right)^{\frac{N}{2}}\big((N+4)-(N+2)q\big)^{\frac{N}{2}}, & \text{for} \quad 1 < q < \frac{N+4}{N+2}, \\ \frac{\Gamma(\frac{2-q}{1-q})\sqrt{\pi}}{\Gamma(\frac{2-q}{q-1}+\frac{N}{2})}\left(\frac{1}{1-q}\right)^{\frac{N}{2}}\big((N+4)-(N+2)q\big)^{\frac{N}{2}}, & \text{for} \quad q < 1, \end{cases} \tag{141}$$

and $|\Sigma|$ denotes the determinant of Σ.

Geometry of deformed exponentials

Now we briefly digress on the geometry of statistical manifolds of deformed exponentials. An overview of the key concepts on Information Geometry of deformed exponentials is useful to set up the notation and to provide some geometric intuition. Indeed, in what follows we will describe our strategy of portfolio diversification in terms of Information Geometry invariants. More precisely, we will propose in this work a search algorithm for portfolios which maximizes the certainty equivalent under the assumption of q-Gaussian probability densities for returns on assets. In particular, this will require the notion of natural gradient associated to a statistical manifold (1).

Probability densities modeled upon ϕ-exponential functions determine a statistical manifold

$$\mathcal{S} = \{p(\cdot\,;\boldsymbol{\vartheta}) : \boldsymbol{\vartheta} \in \mathbb{R}^n\}, \tag{142}$$

where $n = N + N^2$, whose points are probability density functions locally parameterized in terms of n statistical parameters $\boldsymbol{\vartheta} \in \mathbb{R}^n$ by the map

$$p(\mathbf{r};\boldsymbol{\vartheta}) = \exp_\phi(\langle T(\mathbf{r}), \boldsymbol{\vartheta}\rangle - K(\boldsymbol{\vartheta}))\, p_0(\mathbf{r}), \quad \boldsymbol{\vartheta} \in \mathbb{R}^n, \tag{143}$$

where p_0 is a fixed reference probability density and K is the moment-generating function. In what follows, the random variable $\mathbf{r} : \mathcal{S} \to \mathbb{R}^N$ represents the returns of a portfolio with N assets which randomly depend on a set of states of the world. We define the sufficient statistics $T(\boldsymbol{r})$ by

$$T(\boldsymbol{r}) = (\boldsymbol{r}, \boldsymbol{r} \otimes \boldsymbol{r}) \in \mathbb{R}^N \times M(N, \mathbb{R})$$

where $\boldsymbol{r} \otimes \boldsymbol{r}$ is the tensor with components $\boldsymbol{r}_i \boldsymbol{r}_j$.

In order to define a Fisher information metric tensor and a statistical divergence in $\mathcal{S}$ we consider the escort distribution (25) given by

$$\hat{p}(\mathbf{r};\boldsymbol{\vartheta}) \;=\; \frac{1}{h_\phi(\boldsymbol{\vartheta})}\phi\big(p(\mathbf{r};\boldsymbol{\vartheta})/p_0(\mathbf{r})\big)p_0(\mathbf{r}) = \frac{1}{h_\phi(\boldsymbol{\vartheta})}\psi\big(\langle T(\mathbf{r}), \boldsymbol{\vartheta}\rangle - K(\boldsymbol{\vartheta})\big)p_0(\mathbf{r}),$$

with

$$h_\phi(\boldsymbol{\vartheta}) = \int \phi(p(\mathbf{r};\boldsymbol{\vartheta})/p_0(\mathbf{r}))p_0(\mathbf{r})d\mathbf{r}. \tag{144}$$

This permits to generalize the notion of statistical divergence for deformed exponential families as follows

$$D_\phi[\boldsymbol{\vartheta}'||\boldsymbol{\vartheta}] = \mathbb{E}_{\hat{p}(\boldsymbol{r};\boldsymbol{\vartheta})}[\log_\phi p(\mathbf{r};\boldsymbol{\vartheta})/p_0(\mathbf{r}) - \log_\phi p(\mathbf{r};\boldsymbol{\vartheta}')/p_0(\mathbf{r})], \tag{145}$$

where the ϕ-expectation operator is given by

$$\mathbb{E}_\phi[u] := \mathbb{E}_{\hat{p}(\mathbf{r};\boldsymbol{\vartheta})}[u] = \frac{1}{h_\phi(\boldsymbol{\vartheta})}\int u(\mathbf{r})\hat{p}(\mathbf{r};\boldsymbol{\vartheta})\,\mathrm{d}\mathbf{r}. \tag{146}$$

Using the fact that $\mathbb{E}_{\hat{p}(\mathbf{r};\boldsymbol{\vartheta})}[\mathbf{r}] = \nabla K(\boldsymbol{\vartheta})$ we write the Amari-Ohara divergence (145) as a Bregman divergence which can be written as

$$D_\phi[\boldsymbol{\vartheta}||\boldsymbol{\vartheta}'] = K(\boldsymbol{\vartheta}) - K(\boldsymbol{\vartheta}') - \langle \nabla K(\boldsymbol{\vartheta}'), \boldsymbol{\vartheta} - \boldsymbol{\vartheta}'\rangle. \tag{147}$$

In geometric terms, (147) yields two sets of dual affine coordinates for $\mathcal{S}$. The dual relations are $\boldsymbol{\eta} = \nabla K(\boldsymbol{\vartheta})$ and the duality between K and K^* is given by

$$K^*(\boldsymbol{\eta}) = \max_{\boldsymbol{\vartheta}}(\langle \boldsymbol{\vartheta}, \boldsymbol{\eta} \rangle - K(\boldsymbol{\vartheta})) \tag{148}$$

and then $\boldsymbol{\vartheta} = \nabla K^*(\boldsymbol{\eta})$. We have

$$D_\phi[\boldsymbol{\vartheta}||\boldsymbol{\vartheta}'] = K(\boldsymbol{\vartheta}) + K^*(\boldsymbol{\eta}) - \langle \boldsymbol{\eta}, \boldsymbol{\vartheta} \rangle, \tag{149}$$

what implies that D_ϕ is the canonical divergence associated to the dual system $\{\boldsymbol{\vartheta}, \boldsymbol{\eta}\}$ and to the convex function K. We observe that the dual potential function is given by

$$K^*(\boldsymbol{\eta}) = \mathbb{E}_\phi[\log_\phi p(\mathbf{r}; \boldsymbol{\vartheta})/p_0(\mathbf{r})], \tag{150}$$

a relative negative entropy.

We refer to (1) and (15) as comprehensive presentations of the mathematical background on Information Geometry.

3.3 Some technical computations

In this section we denote, for notational convenience, the dimension of the random variable $\mathbf{r}$ by n instead of N.

Proposition 3.1 *The q-Gaussian multivariate distributions belong to the statistical manifold of q-exponential distributions. Indeed, the density*

$$p(\mathbf{r}; \mu, \Sigma) = \frac{1}{\Gamma_q} \exp_q \left(-\frac{1}{2\gamma_q} (\mathbf{r} - \mu)^\top \Sigma^{-1} (\mathbf{r} - \mu) \right) \tag{151}$$

may be written as

$$p(\mathbf{r}, \boldsymbol{\vartheta}) = \exp_q \left(\langle T(\mathbf{r}), \boldsymbol{\vartheta} \rangle - K(\boldsymbol{\vartheta}) \right) \tag{152}$$

where

$$T(\mathbf{r}) = (\mathbf{r}, \mathbf{r} \otimes \mathbf{r}) \in \mathbb{R}^n \times M(n, \mathbb{R}) \tag{153}$$

and the moment-generating function is given by

$$K(\boldsymbol{\vartheta}) = \frac{1}{4} \vartheta^T \Theta^{-1} \vartheta - \log_q (J_q |\Theta|^{\frac{1}{2+(1-q)n}}) \tag{154}$$

The statistical parameters $\boldsymbol{\vartheta} = (\vartheta, \Theta) \in \mathbb{R}^N \times M(N, \mathbb{R})$ are defined by

$$\vartheta = \Sigma_q^{-1} \mu, \quad \textit{and} \quad \Theta = \frac{1}{2} \Sigma_q^{-1}. \tag{155}$$

Here, J_q is the constant

$$J_q = (2\gamma_q)^{\frac{n}{2+(1-q)n}} K_{q,n}^{\frac{-2}{2+(1-q)n}}. \tag{156}$$

Proof. A ϕ-Gaussian multivariate distribution ((25), (33) and references therein) can be defined by

$$p(\mathbf{r};\mu,\Sigma)=\frac{1}{\Gamma_\phi}\exp_\phi\Big(-\frac{1}{2\gamma_\phi}(\mathbf{r}-\mu)^\top\Sigma^{-1}(\mathbf{r}-\mu)\Big) \tag{157}$$

where Γ_ϕ, γ_ϕ are normalizing constants to be fixed in the sequel and Σ is a positive-definite matrix. More precisely, multivariate q-Gaussian densities can be defined as follows. For $1<q<\frac{n+4}{n+2}$ we set

$$p(\mathbf{r};\mu,\Sigma)=\frac{1}{\Gamma_q}\exp_q\Big(-\frac{1}{(n+4)-(n+2)q}(\mathbf{r}-\mu)^\top\Sigma^{-1}(\mathbf{r}-\mu)\Big), \tag{158}$$

that is

$$p(\mathbf{r};\mu,\Sigma)=\frac{1}{\Gamma_q}\Big(1-(1-q)\frac{1}{(n+4)-(n+2)q}(\mathbf{r}-\mu)^\top\Sigma^{-1}(\mathbf{r}-\mu)\Big)^{\frac{1}{1-q}} \tag{159}$$

where Σ is the usual covariance matrix of $\mathbf{r}$,

$$\Gamma_q=\frac{\Gamma(\frac{1}{q-1}-\frac{n}{2})\sqrt{\pi}}{\Gamma(\frac{1}{q-1})}\Big(\frac{1}{q-1}\big((n+4)-(n+2)q\big)\Big)^{\frac{n}{2}}\sqrt{|\Sigma|} \tag{160}$$

and

$$\gamma_q=\frac{1}{2}\big((n+4)-(n+2)q\big). \tag{161}$$

In the case $q<1$ we set

$$p(\mathbf{r};\mu,\Sigma)=\frac{1}{\Gamma_q}\exp_q\Big(-\frac{1}{(n+4)-(n+2)q}(\mathbf{r}-\mu)^\top\Sigma^{-1}(\mathbf{r}-\mu)\Big)_+, \tag{162}$$

(where $+$ denotes the positive part of the function) that is

$$p(\mathbf{r};\mu,\Sigma)=\frac{1}{\Gamma_q}\Big(1-(1-q)\frac{1}{(n+4)-(n+2)q}(\mathbf{r}-\mu)^\top\Sigma^{-1}(\mathbf{r}-\mu)\Big)_+^{\frac{1}{1-q}} \tag{163}$$

where

$$\Gamma_q=\frac{\Gamma(\frac{2-q}{1-q})\sqrt{\pi}}{\Gamma(\frac{2-q}{q-1}+\frac{n}{2})}\Big(\frac{1}{1-q}\big((n+4)-(n+2)q\big)\Big)^{\frac{n}{2}}\sqrt{|\Sigma|} \tag{164}$$

and

$$\gamma_q=\frac{1}{2}\big((n+4)-(n+2)q\big). \tag{165}$$

In order to express these densities in terms of ϕ-exponential parameters we consider the linear change of coordinates

$$\mathbf{z}=A(\mathbf{r}-\mu),$$

where the matrix A is given by

$$A^\top A = \Sigma^{-1}.$$

We have

$$-\frac{1}{2\gamma_\phi}(\mathbf{r}-\mu)^\top \Sigma^{-1}(\mathbf{r}-\mu) = -\frac{1}{2\gamma_\phi}(\mathbf{r}-\mu)^\top A^\top A(\mathbf{r}-\mu) = -\frac{1}{2\gamma_\phi}\mathbf{z}^\top \mathbf{z}$$

with

$$\mathrm{d}\mathbf{r} = |A^{-1}|\mathrm{d}\mathbf{z} = \sqrt{|\Sigma|}\,\mathrm{d}\mathbf{z}.$$

Hence

$$\int_{\mathbb{R}^n} \exp_\phi\Big(-\frac{1}{2\gamma_\phi}(\mathbf{r}-\mu)^T\Sigma^{-1}(\mathbf{r}-\mu)\Big)\,\mathrm{d}\mathbf{r} = \sqrt{|\Sigma|}\int_{\mathbb{R}^n} \exp_\phi\Big(-\frac{1}{2\gamma_\phi}\mathbf{z}^T\mathbf{z}\Big)\,\mathrm{d}\mathbf{z}$$
$$= \sqrt{|\Sigma|}\int_{\mathbb{R}^n} \exp_\phi\Big(-\frac{1}{2\gamma_\phi}|\mathbf{z}|^2\Big)\,\mathrm{d}\mathbf{z} = \sqrt{|\Sigma|}|\mathbb{S}^{n-1}|\int_0^\infty \exp_\phi\Big(-\frac{1}{2\gamma_\phi}r^2\Big)\,r^{n-1}\mathrm{d}r.$$

In the particular case of q-exponentials when $\phi(t) = t^q$, we have

$$\int_{\mathbb{R}^n} \exp_q\Big(-\frac{1}{2\gamma_q}(\mathbf{r}-\mu)^T\Sigma^{-1}(\mathbf{r}-\mu)\Big)\,\mathrm{d}\mathbf{r} = \sqrt{|\Sigma|}\int_{\mathbb{R}^n}\Big(1-(1-q)\frac{1}{2\gamma_q}|\mathbf{z}|^2\Big)^{\frac{1}{1-q}}\,\mathrm{d}\mathbf{z}$$
$$= \sqrt{|\Sigma|}|\mathbb{S}^{n-1}|\int_0^\infty \Big(1-\frac{1}{2\gamma_q}(1-q)r^2\Big)^{\frac{1}{1-q}}\,r^{n-1}\mathrm{d}r = \Gamma_q.$$

Therefore

$$\Gamma_q = I_q\sqrt{|\Sigma|}$$

where

$$I_q = \int_{\mathbb{R}^n}\Big(1-(1-q)\frac{1}{2\gamma_q}|\mathbf{z}|^2\Big)^{\frac{1}{1-q}}\,\mathrm{d}\mathbf{z} = |\mathbb{S}^{n-1}|\int_0^\infty\Big(1-\frac{1}{2\gamma_q}(1-q)r^2\Big)^{\frac{1}{1-q}}\,r^{n-1}\mathrm{d}r$$
$$= \gamma_q^{\frac{n}{2}}|\mathbb{S}^{n-1}|\int_0^\infty\Big(1-\frac{1}{2}(1-q)\tau^2\Big)^{\frac{1}{1-q}}\,\tau^{n-1}\,\mathrm{d}\tau$$

The normalizing constant fixed above is

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

and with this choice

$$\Gamma_q = I_q\sqrt{|\Sigma|} = K_{q,n}\sqrt{|\Sigma|} \tag{166}$$

where

$$K_{q,n} = \begin{cases} \frac{\Gamma(\frac{1}{q-1}-\frac{n}{2})\sqrt{\pi}}{\Gamma(\frac{1}{q-1})}\big(\frac{1}{q-1}\big)^{\frac{n}{2}}\big((n+4)-(n+2)q)\big)^{\frac{n}{2}}, & \text{for} \quad 1<q<\frac{n+4}{n+2}, \\ \frac{\Gamma(\frac{2-q}{1-q})\sqrt{\pi}}{\Gamma(\frac{2-q}{q-1}+\frac{n}{2})}\big(\frac{1}{1-q}\big)^{\frac{n}{2}}\big((n+4)-(n+2)q)\big)^{\frac{n}{2}}, & \text{for} \quad q<1. \end{cases} \tag{167}$$

We conclude that

$$|\mathbb{S}^{n-1}|\int_0^\infty \Big(1-\frac{1}{2}(1-q)\tau^2\Big)^{\frac{1}{1-q}}\tau^{n-1}\,\mathrm{d}\tau = \begin{cases} \frac{\Gamma(\frac{1}{q-1}-\frac{n}{2})\sqrt{\pi}}{\Gamma(\frac{1}{q-1})}\big(\frac{2}{q-1}\big)^{\frac{n}{2}}, & \text{for} \quad 1<q<\frac{n+4}{n+2}, \\ \frac{\Gamma(\frac{2-q}{1-q})\sqrt{\pi}}{\Gamma(\frac{2-q}{q-1}+\frac{n}{2})}\big(\frac{2}{1-q}\big)^{\frac{n}{2}}, & \text{for} \quad q<1. \end{cases}$$

Now we rewrite (151) for $\phi(t)=t^q$ as a q-exponential distribution as follows. We have

$$\begin{aligned} p(\mathbf{r};\mu,\Sigma) &= \exp_q\Big(\log_q\frac{1}{\Gamma_q}\Big)\exp_q\Big(-\frac{1}{2\gamma_q}(\mathbf{r}-\mu)^\top\Sigma^{-1}(\mathbf{r}-\mu)\Big) \\ &= \exp_q\Big(-\frac{1}{2\gamma_q}(\mathbf{r}-\mu)^\top\Sigma^{-1}(\mathbf{r}-\mu)\oplus\log_q\frac{1}{\Gamma_q}\Big). \end{aligned}$$

However

$$\begin{aligned} &-\frac{1}{2\gamma_q}(\mathbf{r}-\mu)^\top\Sigma^{-1}(\mathbf{r}-\mu)\oplus\log_q\frac{1}{\Gamma_q} \\ &= -\frac{1}{2\gamma_q}(\mathbf{r}-\mu)^\top\Sigma^{-1}(\mathbf{r}-\mu)\Big(1+(1-q)\log_q\frac{1}{\Gamma_q}\Big)+\log_q\frac{1}{\Gamma_q} \\ &= -\frac{1}{2\gamma_q}(\mathbf{r}-\mu)^\top\Sigma^{-1}(\mathbf{r}-\mu)\Big(\exp_q\Big(\log_q\frac{1}{\Gamma_q}\Big)\Big)^{1-q}+\log_q\frac{1}{\Gamma_q} \\ &= -\frac{1}{2\gamma_q}(\mathbf{r}-\mu)^\top\Sigma^{-1}(\mathbf{r}-\mu)\,\frac{1}{\Gamma_q^{1-q}}+\log_q\frac{1}{\Gamma_q}. \end{aligned}$$

Since

$$\Gamma_q = I_q\sqrt{|\Sigma|},$$

we write

$$-\frac{1}{2\gamma_q}(\mathbf{r}-\mu)^\top\Sigma^{-1}(\mathbf{r}-\mu)\oplus\log_q\frac{1}{\Gamma_q} = -\frac{1}{2}(\mathbf{r}-\mu)^\top\Sigma_q^{-1}(\mathbf{r}-\mu)+\log_q\frac{1}{\Gamma_q},$$

where

$$\Sigma_q = \gamma_q I_q^{1-q}|\Sigma|^{\frac{1-q}{2}}\,\Sigma. \tag{168}$$

Hence we denote

$$T(\mathbf{r}) = \begin{bmatrix} \mathbf{r}_i & -\mathbf{r}_i\mathbf{r}_j \end{bmatrix}, \quad \boldsymbol{\vartheta} = \begin{bmatrix} \underbrace{(\Sigma_q^{-1}\mu)_i}_{=\vartheta_i} & \underbrace{\frac{1}{2}(\Sigma_q^{-1})_{ij}}_{=\Theta_{ij}} \end{bmatrix}.$$

We obtain

$$p(\mathbf{r};\mu,\Sigma) = p(\mathbf{r};\boldsymbol{\vartheta}) = \exp_q\left(\langle T(\mathbf{r}),\boldsymbol{\vartheta}\rangle - \frac{1}{4}\vartheta^\top\Theta^{-1}\vartheta + \log_q \frac{1}{\sqrt{|\Sigma|}\,I_q}\right).$$

Indeed,

$$\begin{aligned}-\frac{1}{2}(\mathbf{r}-\mu)^\top\Sigma_q^{-1}(\mathbf{r}-\mu) &= \mathbf{r}_i(\Sigma_q^{-1})^i_j\mu^j - \frac{1}{2}\,\mathbf{r}_i(\Sigma_q^{-1})^i_j\mathbf{r}^j - \frac{1}{2}\,\mu_i(\Sigma_q^{-1})^i_j\mu^j \\ &= \left[\begin{array}{cc}\mathbf{r}_i & -\mathbf{r}_i\mathbf{r}_j\end{array}\right]\left[\begin{array}{c}(\Sigma_q^{-1}\mu)_i \\ \frac{1}{2}(\Sigma_q^{-1})_{ij}\end{array}\right] - \frac{1}{2}\mu^\top\Sigma_q^{-1}\mu = \langle T(\mathbf{r}),\boldsymbol{\vartheta}\rangle - \frac{1}{2}\mu^\top\Sigma_q^{-1}\mu,\end{aligned}$$

However

$$I_q\sqrt{|\Sigma|} = \gamma_q^{\frac{-n}{2+(1-q)n}}\,I_q^{\frac{2}{2+(1-q)n}}\,|\Sigma_q|^{\frac{1}{2+(1-q)n}} = (2\gamma_q)^{\frac{-n}{2+(1-q)n}}\,I_q^{\frac{2}{2+(1-q)n}}\,|\Theta|^{\frac{-1}{2+(1-q)n}}.$$

Therefore

$$K(\boldsymbol{\vartheta}) = \frac{1}{4}\vartheta^\top\Theta^{-1}\vartheta - \log_q\frac{1}{\sqrt{|\Sigma|}I_q} = \frac{1}{4}\vartheta^T\Theta^{-1}\theta - \log_q((2\gamma_q)^{\frac{n}{2+(1-q)n}}\,I_q^{\frac{-2}{2+(1-q)n}}\,|\Theta|^{\frac{1}{2+(1-q)n}}).$$

Denoting

$$J_q = (2\gamma_q)^{\frac{n}{2+(1-q)n}}\,K_{q,n}^{\frac{-2}{2+(1-q)n}} \tag{169}$$

one has

$$K(\boldsymbol{\vartheta}) = \frac{1}{4}\vartheta^T\Theta^{-1}\vartheta - \log_q(J_q|\Theta|^{\frac{1}{2+(1-q)n}}). \tag{170}$$

This finishes the proof. □

Proposition 3.2 *For a q-Gaussian multivariate distribution, the gradient of the cumulant function is given by*

$$\nabla K(\boldsymbol{\vartheta}) = \Big(\frac{1}{2}\Theta^{-1}\vartheta, -\frac{1}{2+(1-q)n}J_q^{1-q}|\Theta|^{\frac{1-q}{2+(1-q)n}}\,\Theta^{-1} - \frac{1}{4}(\Theta^{-1}\vartheta)^\top(\Theta^{-1}\vartheta)\Big) \tag{171}$$

Proof. It follows from Proposition 3.1 that for $q=1$ it holds that

$$K(\boldsymbol{\vartheta}) = \frac{1}{4}(\Theta^{-1})^r_s\vartheta_r\vartheta^s - \frac{1}{2}\log\det\Theta + \frac{n}{2}\log\pi. \tag{172}$$

Given a curve of matrices of the form

$$\Theta(s) = \Theta + sA,$$

we compute

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0}\det\Theta(s) = \det\Theta\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0}\det(I+s\Theta^{-1}A) = \det\Theta\,\mathrm{tr}\,\Theta^{-1}A.$$

Setting A as the matrix whose only non-zero element is 1 at the i and column j and using Einstein summation convention on repeated indices in the sequel, one gets

$$\operatorname{tr}\Theta^{-1}A = \sum_{k,\ell}(\Theta^{-1})^k_\ell A^\ell_k = \sum_\ell (\Theta^{-1})^j_\ell A^\ell_j = (\Theta^{-1})^j_i$$

On the other hand, this choice of A yields

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} \det\Theta(s) = \frac{\partial}{\partial\Theta^i_j}\det\Theta.$$

We conclude that

$$\frac{\partial}{\partial\Theta^i_j}\det\Theta = \det\Theta\,(\Theta^{-1})^j_i$$

and

$$\frac{\partial}{\partial\Theta^i_j}\log\det\Theta = (\Theta^{-1})^j_i. \tag{173}$$

Now, differentiating both sides in

$$\Theta^{-1}(s)\Theta(s) = I$$

with respect to s, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0}(\Theta^{-1})^r_k\Theta^k_\ell + (\Theta^{-1})^r_k\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0}\Theta^k_\ell = 0$$

what implies

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0}(\Theta^{-1})^r_s = -(\Theta^{-1})^r_k\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0}\Theta^k_\ell(\Theta^{-1})^\ell_s.$$

In particular, choosing A as above gives

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0}\Theta^k_\ell = \frac{\partial\Theta^k_\ell}{\partial\Theta^i_j} = \delta^k_i\delta^j_\ell$$

and

$$\frac{\partial}{\partial\Theta^i_j}(\Theta^{-1})^r_s = -(\Theta^{-1})^r_k\delta^k_i\delta^j_\ell(\Theta^{-1})^\ell_s = -(\Theta^{-1})^r_i(\Theta^{-1})^j_s$$

In sum,

$$\frac{\partial}{\partial\Theta^i_j}(\Theta^{-1})^r_s = -(\Theta^{-1})^r_i(\Theta^{-1})^j_s \tag{174}$$

We conclude that

$$\frac{\partial}{\partial\Theta^i_j}(\Theta^{-1})^r_s\theta_r\theta^s = -(\Theta^{-1})^r_i\theta_r\,(\Theta^{-1})^j_s\theta^s.$$

Hence,

$$\frac{\partial K}{\partial \Theta^i_j} = -\frac{1}{4}(\Theta^{-1})^r_i \theta_r\, (\Theta^{-1})^j_s \theta^s - \frac{1}{2}(\Theta^{-1})^j_i \tag{175}$$

Moreover, we have

$$\frac{\partial}{\partial \theta^i}(\Theta^{-1})^r_s \vartheta_r \vartheta^s = (\Theta^{-1})^i_s \vartheta^s + (\Theta^{-1})^r_i \vartheta_r.$$

Therefore

$$\frac{\partial K}{\partial \vartheta^i} = \frac{1}{2}(\Theta^{-1})^r_i \vartheta_r. \tag{176}$$

We conclude that

$$\nabla K(\vartheta, \Theta) = \Big(\frac{1}{2}\Theta^{-1}\vartheta, -\frac{1}{2}\Theta^{-1} - \frac{1}{4}(\Theta^{-1}\vartheta)^\top(\Theta^{-1}\vartheta)\Big) \tag{177}$$

Note that we have

$$\nabla K(\boldsymbol{\vartheta}) = \big(\mu, -\Sigma - \mu^T\mu\big) \tag{178}$$

in terms of the original parameters (μ, Σ),

Now, we consider the case $q \neq 1$ when

$$K(\boldsymbol{\vartheta}) = \frac{1}{4}(\Theta^{-1})^r_s \vartheta_r \vartheta^s - \log_q\big(J_q|\Theta|^{\frac{1}{2+(1-q)n}}\big). \tag{179}$$

Using the expressions deduced above, one obtains

$$\frac{\partial}{\partial \Theta^i_j}\log_q\big(J_q|\Theta|^{\frac{1}{2+(1-q)n}}\big) = \frac{1}{J_q^q|\Theta|^{\frac{q}{2+(1-q)n}}} J_q \frac{1}{2+(1-q)n}|\Theta|^{\frac{1}{2+(1-q)n}}(\Theta^{-1})^j_i$$
$$= \frac{1}{2+(1-q)n} J_q^{1-q}|\Theta|^{\frac{1-q}{2+(1-q)n}}\,(\Theta^{-1})^j_i.$$

Therefore

$$\frac{\partial K}{\partial \Theta^i_j} = -\frac{1}{4}(\Theta^{-1})^r_i \vartheta_r\, (\Theta^{-1})^j_s \vartheta^s - \frac{1}{2+(1-q)n} J_q^{1-q}|\Theta|^{\frac{1-q}{2+(1-q)n}}\,(\Theta^{-1})^j_i. \tag{180}$$

Moreover

$$\frac{\partial K}{\partial \vartheta^i} = \frac{1}{2}(\Theta^{-1})^r_i \vartheta_r. \tag{181}$$

In sum

$$\nabla K(\boldsymbol{\vartheta}) = \Big(\frac{1}{2}\Theta^{-1}\vartheta, -\frac{1}{2+(1-q)n} J_q^{1-q}|\Theta|^{\frac{1-q}{2+(1-q)n}}\,\Theta^{-1} - \frac{1}{4}(\Theta^{-1}\vartheta)^\top(\Theta^{-1}\vartheta)\Big) \tag{182}$$

Recall that

$$J_q = (2\gamma_q)^{\frac{n}{2+(1-q)n}} I_q^{-\frac{2}{2+(1-q)n}}, \tag{183}$$

where

$$2\gamma_q = (n+4) - (n+2)q \tag{184}$$

and

$$I_q = K_{q,n} = \begin{cases} \frac{\Gamma(\frac{1}{q-1} - \frac{n}{2})\sqrt{\pi}}{\Gamma(\frac{1}{q-1})} \left(\frac{1}{q-1}\right)^{\frac{n}{2}} \big((n+4) - (n+2)q\big)\big)^{\frac{n}{2}}, & \text{for} \quad 1 < q < \frac{n+4}{n+2}, \\ \frac{\Gamma(\frac{2-q}{1-q})\sqrt{\pi}}{\Gamma(\frac{2-q}{q-1} + \frac{n}{2})} \left(\frac{1}{1-q}\right)^{\frac{n}{2}} \big((n+4) - (n+2)q\big)\big)^{\frac{n}{2}}, & \text{for} \quad q < 1. \end{cases} \tag{185}$$

Hence, we have in terms of the original parameters (μ, Σ_q) that

$$\nabla K(\boldsymbol{\vartheta}) = \left(\mu, -\frac{2^{\frac{2}{2+(1-q)n}}}{2+(1-q)n} |\Sigma_q|^{\frac{q-1}{2+(1-q)n}} \Sigma_q - \mu^T \mu\right) \tag{186}$$

where

$$\Sigma_q = \gamma_q I_q^{1-q} |\Sigma|^{\frac{1-q}{2}} \Sigma$$

and Σ is the variance-covariance matrix. This finishes the proof. □

Proposition 3.3 *Given $\boldsymbol{\alpha} = (\alpha, \mathcal{A}) \in \mathbb{R}^n \times M(n, \mathbb{R})$, the solution of the system*

$$\nabla K(\boldsymbol{\vartheta}) = \boldsymbol{\alpha} \tag{187}$$

is given by

$$\begin{aligned} \Theta^\top &= (2 + (1-q)n)^{\frac{(q-1)n-2}{2}} J_q^{(1-q)(1+(1-q)\frac{n}{2})} | - \mathcal{A} - \alpha^\top \alpha|^{\frac{q-1}{2}} (-\mathcal{A} - \alpha^\top \alpha)^{-1} \\ \theta &= 2\alpha^\top \Theta \end{aligned}$$

Proof. We first consider the case $q = 1$ when (187) reduces to

$$\frac{1}{2}(\Theta^{-1})_i^r \theta_r = \alpha_i \tag{188}$$

$$-\frac{1}{4}(\Theta^{-1})_i^r \theta_r \, (\Theta^{-1})_s^j \theta^s - \frac{1}{2}(\Theta^{-1})_i^j = \mathcal{A}_j^i. \tag{189}$$

Replacing the first equation in the second, one gets

$$-\frac{1}{2}(\Theta^{-1})_i^j = \mathcal{A}_j^i + \alpha^i \alpha_j.$$

Hence,

$$\Theta_j^i = -\frac{1}{2}\big((\mathcal{A} + \alpha^\top \alpha)^{-1}\big)_i^j, \tag{190}$$

that is,

$$\Theta^\top = -\frac{1}{2}(\mathcal{A} + \alpha^\top \alpha)^{-1}. \tag{191}$$

Moreover

$$\theta_i = 2\Theta_i^j \alpha_j,$$

that is,

$$\theta = 2\alpha^\top \Theta. \tag{192}$$

For the case $q \neq 1$, given

$$\boldsymbol{\alpha} = (\alpha, \mathcal{A}),$$

the system

$$\nabla K(\boldsymbol{\theta}) = \boldsymbol{\alpha}$$

is written as

$$\frac{1}{2}(\Theta^{-1})_i^r \theta_r = \alpha_i \tag{193}$$

$$-\frac{1}{4}(\Theta^{-1})_i^r \theta_r \, (\Theta^{-1})_s^j \theta^s - \frac{1}{2+(1-q)n} J_q^{1-q} |\Theta|^{\frac{1-q}{2+(1-q)n}} (\Theta^{-1})_i^j = \mathcal{A}_j^i. \tag{194}$$

Replacing the first equation in the second, one gets

$$-\frac{1}{2+(1-q)n} J_q^{1-q} |\Theta|^{\frac{1-q}{2+(1-q)n}} (\Theta^{-1})_i^j = \mathcal{A}_j^i + \alpha^i \alpha_j.$$

Therefore

$$|\Theta|^{\frac{1-q}{2+(1-q)n}} (\Theta^{-1})_i^j = (2+(1-q)n) J_q^{q-1} \, (-\mathcal{A}_j^i - \alpha^i \alpha_j).$$

In particular

$$|\Theta|^{-\frac{2}{2+(1-q)n}} = (2+(1-q)n)^n J_q^{(q-1)n} \, |-\mathcal{A} - \alpha^\top \alpha|,$$

from what follows that

$$|\Theta|^{\frac{1-q}{2+(1-q)n}} = (2+(1-q)n)^{(q-1)\frac{n}{2}} J_q^{(q-1)^2 \frac{n}{2}} \, |-\mathcal{A} - \alpha^\top \alpha|^{\frac{q-1}{2}}$$

Hence,

$$(\Theta^{-1})_i^j = (2+(1-q)n)^{1+(1-q)\frac{n}{2}} J_q^{(q-1)(1+(1-q)\frac{n}{2})} |-\mathcal{A} - \alpha^\top \alpha|^{\frac{1-q}{2}} \, (-\mathcal{A}_j^i - \alpha^i \alpha_j)$$

that is,

$$\Theta^\top = (2+(1-q)n)^{\frac{(q-1)n-2}{2}} J_q^{(1-q)(1+(1-q)\frac{n}{2})} |-\mathcal{A} - \alpha^\top \alpha|^{\frac{q-1}{2}} \, (-\mathcal{A} - \alpha^\top \alpha)^{-1} \tag{195}$$

Moreover,

$$\theta_i = 2\Theta_i^j \alpha_j,$$

that is,

$$\theta = 2\alpha^T \Theta. \tag{196}$$

Then, $\boldsymbol{\alpha} = (\alpha, \mathcal{A})$ determines the parameters $\boldsymbol{\vartheta} = (\vartheta, \Theta)$ via

$$\Theta^\top = (2 + (1-q)n)^{\frac{(q-1)n-2}{2}} J_q^{(1-q)(1+(1-q)\frac{n}{2})} | -\mathcal{A} - \alpha^\top \alpha|^{\frac{q-1}{2}} (-\mathcal{A} - \alpha^\top \alpha)^{-1}$$
$$\theta = 2\alpha^\top \Theta.$$

This finishes the proof of the proposition □

Proposition 3.4 *The dual cumulant function K^* is given in the case of q-exponential distributions in terms of the negative relative q-entropy:*

$$K^*(\boldsymbol{\eta}) = \mathbb{E}_q[\log_q(p(\mathbf{r}; \boldsymbol{\vartheta})/p_0(\mathbf{r}))], \tag{197}$$

that is,

$$K^*(\boldsymbol{\eta}) = \frac{1}{1-q}\left(\frac{1}{h_q(\boldsymbol{\vartheta})} - 1\right), \tag{198}$$

where h_q is given by 144.

Proof. We have from the definition of the dual moment-generating function K^* that setting $\boldsymbol{\eta} = \nabla K(\boldsymbol{\vartheta})$ one obtains

$$\begin{aligned}
K^*(\boldsymbol{\eta}) &= \langle \boldsymbol{\vartheta}, \boldsymbol{\eta} \rangle - K(\boldsymbol{\vartheta}) = \langle \boldsymbol{\vartheta}, \nabla K(\boldsymbol{\vartheta}) \rangle - K(\boldsymbol{\vartheta}) = \langle \boldsymbol{\vartheta}, \mathbb{E}_{\hat{p}(\cdot;\boldsymbol{\vartheta})}[T(\mathbf{r})] \rangle - \mathbb{E}_{\hat{p}(\cdot;\boldsymbol{\vartheta})}[K(\boldsymbol{\vartheta})] \\
&= \mathbb{E}_{\hat{p}(\cdot;\boldsymbol{\vartheta})}[\langle \vartheta, T(\mathbf{r}) \rangle - K(\boldsymbol{\vartheta})] = \mathbb{E}_q[\log_q(p(\mathbf{r}; \boldsymbol{\vartheta})/p_0(\mathbf{r}))] \\
&= \frac{1}{h_q(\boldsymbol{\vartheta})} \int \big(\langle \vartheta, T(\mathbf{r}) \rangle - K(\boldsymbol{\vartheta})\big)\big(1 + (1-q)(\langle \vartheta, T(\mathbf{r}) \rangle - K(\boldsymbol{\vartheta}))\big)^{\frac{q}{1-q}} p_0(\mathbf{r})\, \mathrm{d}\mathbf{r} \\
&= \frac{1}{h_q(\boldsymbol{\vartheta})} \int \frac{1}{1-q}\big((p(\mathbf{r}; \boldsymbol{\vartheta})/p_0(\mathbf{r}))^{1-q} - 1\big)\big(p(\mathbf{r}; \boldsymbol{\vartheta})/p_0(\mathbf{r})\big)^q p_0(\mathbf{r})\, \mathrm{d}\mathbf{r} \\
&= \frac{1}{1-q} \frac{1}{h_q(\boldsymbol{\vartheta})} \int \big(p(\mathbf{r}; \boldsymbol{\vartheta})/p_0(\mathbf{r}) - (p(\mathbf{r}; \boldsymbol{\vartheta})/p_0(\mathbf{r}))^q\big) p_0(\mathbf{r})\, \mathrm{d}\mathbf{r} \\
&= \frac{1}{1-q} \frac{1}{h_q(\boldsymbol{\vartheta})} \int p(\mathbf{r}; \boldsymbol{\vartheta})\, \mathrm{d}\mathbf{r} - \frac{1}{1-q} \frac{1}{h_q(\boldsymbol{\vartheta})} \int \big(p(\mathbf{r}; \boldsymbol{\vartheta})/p_0(\mathbf{r})\big)^q p_0(\mathbf{r})\, \mathrm{d}\mathbf{r} \\
&= \frac{1}{1-q} \frac{1}{h_q(\boldsymbol{\vartheta})} - \frac{1}{1-q} \frac{1}{h_q(\boldsymbol{\vartheta})} h_q(\boldsymbol{\vartheta})
\end{aligned}$$

We conclude that

$$H_q(\boldsymbol{\vartheta}) = K^*(\boldsymbol{\eta}) = \frac{1}{1-q}\left(\frac{1}{h_q(\boldsymbol{\vartheta})} - 1\right) \tag{199}$$

where H_q is the relative q-entropy

$$H_q(\boldsymbol{\vartheta}) = \mathbb{E}_q[\log_q(p(\mathbf{r}; \boldsymbol{\vartheta})/p_0(\mathbf{r}))]. \tag{200}$$

Proposition 3.5 *The gradient of the dual cumulant function K^* in terms of affine dual parameters*

$$\boldsymbol{\eta} = (\eta, \mathcal{N})$$

is given in the case of q-exponential distributions by

$$\nabla K^*(\boldsymbol{\eta}) = \left(\widetilde{C}_{q,n}|B|^{-\frac{1-q}{2}}B^{-1}\eta, \frac{1}{2}\widetilde{C}_{q,n}|B|^{-\frac{1-q}{2}}B^{-1}\right) \tag{201}$$

where

$$\widetilde{C}_{q,n} = -(1-q)n^2 C_{q,n}^{-\frac{C_{q,n}}{2}} J_q^{(1-q)\frac{C_{q,n}}{2}} + C_{q,n}^{-(1-q)\frac{n}{2}} J_q^{(1-q)\frac{C_{q,n}}{2}},$$

with $C_{q,n} = 2 + (1-q)n$ *and*

$$B = -\mathcal{N} - \eta^T\eta = -\mathcal{N}(I + \eta^T\mathcal{N}^{-1}\eta).$$

Proof. First, we deduce an explicit expression of K^* in terms of the dual coordinate system $\boldsymbol{\eta}$. We have

$$\begin{aligned}\langle\boldsymbol{\vartheta}, \nabla K(\boldsymbol{\vartheta})\rangle = \boldsymbol{\vartheta}^T\nabla K(\boldsymbol{\vartheta}) &= \frac{1}{2}\theta^i(\Theta^{-1})_i^r\theta_r - \frac{1}{2+(1-q)n}J_q^{1-q}|\Theta|^{\frac{1-q}{2+(1-q)n}}\Theta_j^i(\Theta^{-1})_i^j \\ &-\frac{1}{4}\Theta_j^i(\Theta^{-1})_i^r\theta_r(\Theta^{-1})_s^j\theta^s \\ &= \frac{1}{2}\theta^i(\Theta^{-1})_i^r\theta_r - \frac{n^2}{2+(1-q)n}J_q^{1-q}|\Theta|^{\frac{1-q}{2+(1-q)n}} - \frac{1}{4}\theta_r(\Theta^{-1})_s^r\theta^s\end{aligned}$$

Therefore

$$\begin{aligned}K^*(\boldsymbol{\eta}) = \langle\boldsymbol{\vartheta}, \nabla K(\boldsymbol{\vartheta})\rangle - K(\boldsymbol{\vartheta}) &= \frac{1}{4}\theta^i(\Theta^{-1})_i^r\theta_r - \frac{n^2}{2+(1-q)n}J_q^{1-q}|\Theta|^{\frac{1-q}{2+(1-q)n}} \\ &-\frac{1}{4}(\Theta^{-1})_s^r\theta_r\theta^s + \log_q\left(J_q|\Theta|^{\frac{1}{2+(1-q)n}}\right)\end{aligned}$$

Hence,

$$K^*(\boldsymbol{\eta}) = -\frac{n^2}{2+(1-q)n}J_q^{1-q}|\Theta|^{\frac{1-q}{2+(1-q)n}} + \log_q\left(J_q|\Theta|^{\frac{1}{2+(1-q)n}}\right) \tag{202}$$

Since

$$(\eta, \mathcal{N}) := \boldsymbol{\eta} = \nabla K(\boldsymbol{\vartheta})$$

we have

$$\frac{1}{2}(\Theta^{-1})_i^r\theta_r = \eta_i \tag{203}$$

$$-\frac{1}{4}(\Theta^{-1})_i^r\theta_r(\Theta^{-1})_s^j\theta^s - \frac{1}{2+(1-q)n}J_q^{1-q}|\Theta|^{\frac{1-q}{2+(1-q)n}}(\Theta^{-1})_i^j = \mathcal{N}_i^j \tag{204}$$

Replacing the first equation in the second, one obtains

$$-\frac{1}{2+(1-q)n}J_q^{1-q}|\Theta|^{\frac{1-q}{2+(1-q)n}}(\Theta^{-1})_i^j = \mathcal{N}_i^j + \eta_i\eta^j.$$

Denoting

$$B_i^j := -\mathcal{N}_i^j - \eta_i \eta^j, \tag{205}$$

one concludes that

$$\left(\frac{1}{2+(1-q)n}\right)^n J_q^{(1-q)n} |\Theta|^{\frac{(1-q)n}{2+(1-q)n}} |\Theta|^{-1} = |B|,$$

what implies that

$$|\Theta|^{\frac{-2}{2+(1-q)n}} = \left(\frac{1}{2+(1-q)n}\right)^{-n} J_q^{-(1-q)n} |B|$$

or

$$|\Theta|^{\frac{1-q}{2+(1-q)n}} = \left(\frac{1}{2+(1-q)n}\right)^{(1-q)\frac{n}{2}} J_q^{(1-q)^2 \frac{n}{2}} |B|^{-\frac{1-q}{2}}.$$

Thus,

$$\frac{1}{2+(1-q)n} J_q^{1-q} |\Theta|^{\frac{1-q}{2+(1-q)n}} = \left(\frac{1}{2+(1-q)n}\right)^{\frac{2+(1-q)n}{2}} J_q^{(1-q)+(1-q)^2\frac{n}{2}} |B|^{-\frac{1-q}{2}}$$

We conclude that

$$(\Theta^{-1})_i^j = \left(\frac{1}{2+(1-q)n}\right)^{-\frac{2+(1-q)n}{2}} J_q^{-(1-q)-(1-q)^2\frac{n}{2}} |B|^{\frac{1-q}{2}} B_i^j \tag{206}$$

and

$$\theta_i = 2\sum_{j=1}^{n} \Theta_i^j \eta_j \tag{207}$$

Moreover, we get

$$K^*(\boldsymbol{\eta}) = K^*(\eta, \mathcal{N}) = -n^2 \left(\frac{1}{2+(1-q)n}\right)^{\frac{2+(1-q)n}{2}} J_q^{(1-q)+(1-q)^2\frac{n}{2}} |B|^{-\frac{1-q}{2}}$$
$$+ \log_q \left(\left(\frac{1}{2+(1-q)n}\right)^{\frac{n}{2}} J_q^{\frac{2+(1-q)n}{2}} |B|^{-\frac{1}{2}} \right)$$

where

$$B_i^j = -\mathcal{N}_i^j - \eta_i \eta^j.$$

Note that, when $q = 1$, we have

$$I_1 = (2\pi)^{\frac{n}{2}}$$

and

$$J_1 = 2^{\frac{n}{2}} I_1^{-1} = \pi^{-\frac{n}{2}}.$$

In this case,

$$K^*(\boldsymbol{\eta}) = -\frac{n^2}{2} + \log\left(\pi^{-\frac{n}{2}}|B|^{-\frac{1}{2}}\right).$$

Using dual parameters $\boldsymbol{\eta} = (\eta, \mathcal{N})$ and having in mind that

$$|B| = |-\mathcal{N}||I + \eta^T\mathcal{N}^{-1}\eta|,$$

one obtains

$$K^*(\boldsymbol{\eta}) = -\frac{n^2}{2} - \frac{n}{2}\log(2\pi) - \frac{1}{2}\log|-\mathcal{N}| - \frac{1}{2}\log|I + \eta^T\mathcal{N}^{-1}\eta|.$$

Now, using the definition of B as

$$B^i_j = -\mathcal{N}^i_j - \eta^i\eta_j,$$

one computes

$$\frac{\partial}{\partial\eta^i}|B| = \frac{\partial}{\partial B^k_\ell}|B|\frac{\partial B^k_\ell}{\partial\eta^i} = -|B|(B^{-1})^\ell_k(\delta^k_i\eta_\ell + \eta^k\delta_{\ell i}) = -2|B|(B^{-1})^\ell_i\eta_\ell$$

and

$$\frac{\partial}{\partial\mathcal{N}^i_j}|B| = \frac{\partial}{\partial B^k_\ell}|B|\frac{\partial B^k_\ell}{\partial\mathcal{N}^i_j} = -|B|(B^{-1})^\ell_k\delta^k_i\delta^j_\ell = -|B|(B^{-1})^j_i.$$

Denoting

$$C_{q,n} := 2 + (1-q)n \tag{208}$$

one concludes that

$$\frac{\partial}{\partial\eta^i}K^*(\boldsymbol{\eta}) = -2\frac{1-q}{2}n^2C_{q,n}^{-\frac{C_{q,n}}{2}}J_q^{(1-q)\frac{C_{q,n}}{2}}|B|^{-\frac{1-q}{2}}(B^{-1})^\ell_i\eta_\ell + C_{q,n}^{-(1-q)\frac{n}{2}}J_q^{(1-q)\frac{C_{q,n}}{2}}|B|^{-\frac{1-q}{2}}(B^{-1})^\ell_i\eta_\ell.$$

Therefore

$$\frac{\partial}{\partial\eta^i}K^*(\boldsymbol{\eta}) = \left(-(1-q)n^2C_{q,n}^{-\frac{C_{q,n}}{2}}J_q^{(1-q)\frac{C_{q,n}}{2}} + C_{q,n}^{-(1-q)\frac{n}{2}}J_q^{(1-q)\frac{C_{q,n}}{2}}\right)|B|^{-\frac{1-q}{2}}(B^{-1})^\ell_i\eta_\ell. \tag{209}$$

We also have

$$\frac{\partial}{\partial\mathcal{N}^i_j}K^*(\boldsymbol{\eta}) = -\frac{1-q}{2}n^2C_{q,n}^{-\frac{C_{q,n}}{2}}J_q^{(1-q)\frac{C_{q,n}}{2}}|B|^{-\frac{1-q}{2}}(B^{-1})^j_i + \frac{1}{2}C_{q,n}^{-(1-q)\frac{n}{2}}J_q^{(1-q)\frac{C_{q,n}}{2}}|B|^{-\frac{1-q}{2}}(B^{-1})^j_i.$$

Hence,

$$\frac{\partial}{\partial\mathcal{N}^i_j}K^*(\boldsymbol{\eta}) = \frac{1}{2}\left(-(1-q)n^2C_{q,n}^{-\frac{C_{q,n}}{2}}J_q^{(1-q)\frac{C_{q,n}}{2}} + C_{q,n}^{-(1-q)\frac{n}{2}}J_q^{(1-q)\frac{C_{q,n}}{2}}\right)|B|^{-\frac{1-q}{2}}(B^{-1})^j_i \tag{210}$$

Setting

$$\widetilde{C}_{q,n} = -(1-q)n^2 C_{q,n}^{-\frac{C_{q,n}}{2}} J_q^{(1-q)\frac{C_{q,n}}{2}} + C_{q,n}^{-(1-q)\frac{n}{2}} J_q^{(1-q)\frac{C_{q,n}}{2}}, \tag{211}$$

one obtains

$$\nabla K^*(\boldsymbol{\eta}) = \left(\widetilde{C}_{q,n}|B|^{-\frac{1-q}{2}} B^{-1}\eta, \frac{1}{2}\widetilde{C}_{q,n}|B|^{-\frac{1-q}{2}} B^{-1}\right) \tag{212}$$

where

$$B = -\mathcal{N} - \eta^T\eta = -\mathcal{N}(I + \eta^T\mathcal{N}^{-1}\eta),$$

Replacing

$$\eta = \alpha, \quad \mathcal{N} = \mathcal{A},$$

one obtains the correspondent value of the parameter $\boldsymbol{\vartheta}$, namely

$$\nabla K^*(\boldsymbol{\alpha}) = \boldsymbol{\vartheta}.$$

This finishes the proof. □

Proposition 3.6 *The Hessian of K^* with respect to the vector parameter η is given by the q-Gaussian covariance matrix Σ_q.*

Proof. We have

$$\frac{\partial^2 K}{\partial\theta^i\partial\theta^j} = \frac{1}{2}\frac{\partial}{\partial\theta^i}\big((\Theta^{-1})^r_j\theta_r\big) = \frac{1}{2}(\Theta^{-1})^r_j\delta_{ri} = \frac{1}{2}(\Theta^{-1})^j_i = (\Sigma_q)^j_i. \tag{213}$$

The conclusion follows from the fact that the Hessian matrices of K and K^*, calculated at correspondent values of the parameters, are inverses of each other. □

4 MEAN-DIVERGENCE PORTFOLIO SELECTION

4.1 Generalized HARA utility functions

As we commented earlier, portfolio selection models usually rely on a theory of choice under uncertainty whose most traditional formulation is the expected utility theory by J. von Neumann and O. Morgenstern (21). A suitable class of utility functions in the context of deformed exponentials is given by generalized HARA (hyperbolic risk aversion) functions of the form

$$u(w) = C_0\frac{1}{\psi(aw)}, \tag{214}$$

where C_0 is a normalizing constant and $a > 0$ is a risk aversion parameter. In the particular case of q-exponential we recover, up to some adjustment of constants, the well-known HARA functions

$$u(w) = -\frac{1}{q}\big(1 + (1-q)aw\big)^{-\frac{q}{1-q}}.$$

As expected, the limit case $q \to 1$ yields the constant risk aversion (CARA) utility functions $u(w) = -\exp(-aw)$. Note that the Arrow-Pratt absolute risk aversion for generalized HARA functions is given by

$$-\frac{u''(w)}{u'(w)} = a\left(\dot{\phi}(v) - \frac{\ddot{\phi}(v)}{\dot{\phi}(v)}\phi(v)\right), \tag{215}$$

where $v = \exp_\phi(aw)$. Here $'$ and $\cdot$ denote, respectively, derivatives with respect to w and v. This expression reduces to

$$-\frac{u''(w)}{u'(w)} = a$$

in the case of exponential utility functions, an example of constant absolute risk aversion utility function, whereas

$$-\frac{u''(w)}{u'(w)} = aw^{q-1}$$

in the case of q-exponentials, that is, for $\phi(t) = t^q$. Up to a multiplicative constant, HARA utility function (214) may be written in a form usually found in the literature

$$u(w) = \frac{1-\gamma}{\gamma}\left(B + \frac{A}{1-\gamma}w\right)^\gamma,$$

where $\gamma = -q/(1-q)$. Taking $B \to 0$ we recover a power utility function of the form

$$u(w) = Cw^\gamma$$

for some constant $C > 0$. In this case, the coefficient of absolute risk aversion is given by

$$-\frac{u''(w)}{u'(w)} = \frac{1-\gamma}{w} = \frac{1}{1-q}\frac{1}{w}.$$

4.2 Generalized mean-divergence model

We now summarize the theory developed by R. Nock, B. Magdalou, E. Bryis and F. Nielsen (30), (31) according to which $D_\phi(\cdot||\cdot)$ is a risk measure that is at the same time more precise and general than the usual variance of the portfolio. Recall that in Markowitz mean-variance model the optimal portfolio is determined by (20), (6)

$$\max_{\boldsymbol{\alpha}\in\mathcal{S}_{N-1}}\left(\mathbb{E}[\langle \mathbf{r}, \boldsymbol{\alpha}\rangle] - \frac{a}{2}\boldsymbol{\alpha}\Sigma\,\boldsymbol{\alpha}^\top\right), \tag{216}$$

where Σ is the variance-covariance matrix of the random variable $\mathbf{r}$ and

$$\mathcal{S}_{N-1} = \{\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^N_+ : \alpha_1 + \dots + \alpha_N = 1\}$$

is the set of weights for the assets on the portfolio.

The minimum variance portfolio for a prescribed expected return is specified by the allocation vector

$$\boldsymbol{\alpha} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}\Sigma^{-1}\mathbf{1}^{\top}}. \tag{217}$$

It turns out that in the particular case of CARA utility functions and Gaussian (more generally, exponential) distributed returns the risk premium is *explicitly* given by

$$\Pi = \frac{a}{2}\boldsymbol{\alpha}\,\Sigma\,\boldsymbol{\alpha}^{\top}, \tag{218}$$

see (20). Hence (216) may be written in terms of the certainly equivalent

$$C = \mathbb{E}[w] - \Pi$$

as

$$\max_{\boldsymbol{\alpha}} \big(\mathbb{E}[w] - \Pi\big), \tag{219}$$

where $w = \langle \mathbf{r}, \boldsymbol{\alpha}\rangle$ is the wealth level corresponding to $\mathbf{r}$. The term between parenthesis in (219) is the certainty equivalent $\mathcal{C}$.

In sum, in (30), (31) the authors proved that for exponential densities and CARA utility functions the certainty equivalent $\mathcal{C}$ and the risk premium Π may be explicitly computed. Indeed, they obtain

$$C = \frac{1}{a}\big(K(\boldsymbol{\vartheta}) - K(\boldsymbol{\vartheta}')\big) \tag{220}$$

and

$$\Pi = \frac{1}{a}D_1[\boldsymbol{\vartheta}'||\boldsymbol{\vartheta}], \tag{221}$$

where $\phi(t) = t$ (which corresponds to the usual exponential function, that is, to $q = 1$) and $\boldsymbol{\vartheta}' = \boldsymbol{\vartheta} - a\boldsymbol{\alpha}$. This implies that (219) may be written as

$$\min_{\boldsymbol{\alpha}} \left(\langle \nabla K(\boldsymbol{\vartheta}'), \boldsymbol{\alpha}\rangle + \frac{1}{a}D_1[\boldsymbol{\vartheta}||\boldsymbol{\vartheta}']\right) \tag{222}$$

where D_1 is the Bregman divergence associated to an exponential probability density of the form

$$p(\mathbf{r};\boldsymbol{\vartheta}) = \exp(\langle T(\mathbf{r}), \boldsymbol{\vartheta}\rangle - K(\boldsymbol{\vartheta}))\, p_0(\mathbf{r}), \quad \boldsymbol{\vartheta} \in \mathbb{R}^N.$$

Motivated by this idea we can define (222) as the correct analog of the mean-variance optimization problem for deformed exponentials and generalized HARA utility functions. Hence we rephrase the mean-variance method in terms of a generalized mean-divergence problem of the form

$$\min_{\boldsymbol{\alpha}} \left(\langle \nabla K(\boldsymbol{\vartheta}'), \boldsymbol{\alpha}\rangle + \frac{1}{a_{\phi}}D_{\phi}[\boldsymbol{\vartheta}||\boldsymbol{\vartheta}']\right), \tag{223}$$

with

$$a_\phi = a\left(\dot{\phi} - \frac{\ddot{\phi}}{\dot{\phi}}\phi\right). \tag{224}$$

We observe that even in the case of q-exponentials their non-additivity prevents us to obtain a closed form for C and Π. However, since the utility function is modeled as a generalized HARA function we may obtain an approximate expression for those quantities. Indeed the expected utility is given by

$$\mathbb{E}_\phi[u] = \frac{1}{h_\phi(\boldsymbol{\vartheta})} \int \frac{1}{\psi(\langle T(\mathbf{r}), a\boldsymbol{\alpha}\rangle)} \psi(\langle T(\mathbf{r}), \boldsymbol{\vartheta}\rangle - K(\boldsymbol{\vartheta})) p_0(\mathbf{r}) \mathrm{d}\mathbf{r}.$$

In the case of q-exponentials, that is, setting $\phi(t) = t^q$ we have

$$\psi(t)\psi(t') = \psi(t \oplus t'),$$

where $\oplus$ stands for the q-sum operation

$$t \oplus t' = t + t + (1-q)tt'.$$

Therefore

$$\mathbb{E}_\phi[u] = \int \psi(\langle T(\mathbf{r}), \boldsymbol{\vartheta}\rangle - K(\boldsymbol{\vartheta}) \oplus \langle \mathbf{r}, -a\boldsymbol{\alpha}\rangle))\, p_0(\mathbf{r})\, \mathrm{d}\mathbf{r}.$$

However, expanding ψ up to second order we obtain

$$\mathbb{E}_\phi[u] \simeq \psi(K(\boldsymbol{\vartheta} - a\boldsymbol{\alpha}) - K(\boldsymbol{\vartheta})) + O(a^2), \tag{225}$$

which allows us to recover the analogs of (221) and (222) in the general case up to quadratic remainder terms. This is important since we can then generalize the Markowitz model for portfolio selection, as described in next section.

4.3 Generalized Markowitz portfolio selection

Motivated by (223) and (216) we consider the divergence D_ϕ as the counterpart of the variance in the case of returns distributed according to a ϕ-exponential distribution. Following the classical mean-variance optimization model we search for an allocation $\boldsymbol{\alpha} \in \mathcal{S}_n$ such that fixed some initial choice of the statistical parameter $\boldsymbol{\vartheta}(t)$ at a time t we have

$$D_\phi(\boldsymbol{\vartheta}(t)||\boldsymbol{\vartheta} - a\boldsymbol{\alpha}) = \min_{\boldsymbol{\vartheta}} D_\phi(\boldsymbol{\vartheta}(t)||\boldsymbol{\vartheta}) \tag{226}$$

subject to

$$\langle \mathbf{1}, \boldsymbol{\alpha}\rangle = 1. \tag{227}$$

Theorem 4.1 *A portfolio allocation $\boldsymbol{\alpha}_1$ solves the optimization problem (226)-(227) if* *and only if*

$$\boldsymbol{\alpha}_{\mathbf{1}}(\tau) = \frac{1}{\mathbf{1}(\nabla^2 K(\boldsymbol{\vartheta}(\tau)))^{-1}\mathbf{1}^\top}(\nabla^2 K(\boldsymbol{\vartheta}(\tau)))^{-1}\mathbf{1}^\top. \tag{228}$$

where $\boldsymbol{\vartheta}(\tau)$ *is the statistical parameter at a time* τ *for which*

$$\min_\tau D_\phi(\boldsymbol{\vartheta}(t)||\boldsymbol{\vartheta}(\tau))$$

is attained.

Proof. We consider the following function from the Lagrangian model:

$$f(\tau) = D_\phi(\boldsymbol{\vartheta}(t)||\boldsymbol{\vartheta}(\tau)) + \lambda(1 - \langle \mathbf{1}, \boldsymbol{\alpha}(\tau)\rangle).$$

Since

$$D_\phi(\boldsymbol{\vartheta}(t)||\boldsymbol{\vartheta}(\tau)) = K(\boldsymbol{\vartheta}(t)) - K(\boldsymbol{\vartheta}(\tau)) - \langle \nabla K(\boldsymbol{\vartheta}(\tau)), \boldsymbol{\vartheta}(t) - \boldsymbol{\vartheta}(\tau)\rangle,$$

the first-order condition reads as

$$\begin{aligned} &-\langle \nabla K(\boldsymbol{\vartheta}(\tau)), \boldsymbol{\vartheta}'(\tau)\rangle + \langle \nabla K(\boldsymbol{\vartheta}(\tau)), \boldsymbol{\vartheta}'(\tau)\rangle - \boldsymbol{\vartheta}'(\tau)\nabla^2 K(\boldsymbol{\vartheta}(\tau))(\boldsymbol{\vartheta}(t) - \boldsymbol{\vartheta}(\tau))^\top \\ &-\lambda\langle \mathbf{1}, \boldsymbol{\alpha}'(\tau)\rangle = 0, \end{aligned}$$

that is,

$$-\boldsymbol{\vartheta}'(\tau)\nabla^2 K(\boldsymbol{\vartheta}(\tau))(\boldsymbol{\vartheta}(t) - \boldsymbol{\vartheta}(\tau))^\top = \lambda\langle \mathbf{1}, \boldsymbol{\alpha}'(\tau)\rangle. \tag{229}$$

Assuming that

$$\boldsymbol{\alpha}'(\tau) = -\boldsymbol{\vartheta}'(\tau),$$

which assumes that updated forecasts on the parameters of the distribution are immediately followed by equal adjustments of the portfolio allocation (similarly to the strategy of self-financing portfolio), one deduces the vectorial equation

$$\nabla^2 K(\boldsymbol{\vartheta}(\tau))(\boldsymbol{\vartheta}(t) - \boldsymbol{\vartheta}(\tau))^\top = \lambda\mathbf{1}.$$

Therefore

$$\boldsymbol{\vartheta}(t) - \boldsymbol{\vartheta}(\tau) = \lambda(\nabla^2 K(\boldsymbol{\vartheta}(\tau)))^{-1}\mathbf{1}^\top.$$

Taking traces, one obtains

$$\langle \boldsymbol{\vartheta}(t) - \boldsymbol{\vartheta}(\tau), \mathbf{1}\rangle = \lambda\mathbf{1}(\nabla^2 K(\boldsymbol{\vartheta}(\tau)))^{-1}\mathbf{1}^\top.$$

For the setting

$$\boldsymbol{\vartheta}(\tau) = \boldsymbol{\vartheta}(t) - \boldsymbol{\alpha}(\tau)$$

(which in particular yields $\boldsymbol{\alpha}'(\tau) = -\boldsymbol{\vartheta}'(\tau)$) we conclude that

$$1 = \lambda \mathbf{1}(\nabla^2 K(\boldsymbol{\vartheta}(t)))^{-1}\mathbf{1}^\top,$$

obtaining an explicit expression for λ, that is, for the shadow price (21). We have

$$\boldsymbol{\alpha}(\tau) = \frac{1}{\mathbf{1}(\nabla^2 K(\boldsymbol{\vartheta}(\tau)))^{-1}\mathbf{1}^\top}(\nabla^2 K(\boldsymbol{\vartheta}(\tau)))^{-1}\mathbf{1}^\top, \tag{230}$$

as a generalized Markowitz formula. Recall that

$$\nabla K(\boldsymbol{\vartheta}(\tau)) = \boldsymbol{\eta}(\tau) = \mathbb{E}_{\hat{p}(\cdot;\boldsymbol{\vartheta}(\tau))}[T(\mathbf{r})].$$

This finishes the proof of the theorem. □

As in the classical case, we may consider the following variant of the optimization problem:

$$D_\phi(\boldsymbol{\vartheta}(t)||\boldsymbol{\vartheta} - a\boldsymbol{\alpha}) = \min_{\boldsymbol{\vartheta}} D_\phi(\boldsymbol{\vartheta}(t)||\boldsymbol{\vartheta}) \tag{231}$$

subject to the constraint

$$\langle \mathbf{r}, \boldsymbol{\alpha}\rangle = \mu_* \tag{232}$$

for a certain fixed expected portfolio return μ_*

Theorem 4.2 *A portfolio allocation $\boldsymbol{\alpha}_\mu$ solves the optimization problem (231)-(232) if and only if*

$$\boldsymbol{\alpha}_\mu(\tau) = \frac{1}{\mathbf{1}(\nabla^2 K(\boldsymbol{\vartheta}(\tau)))^{-1}\mathbf{r}^\top}(\nabla^2 K(\boldsymbol{\vartheta}(\tau)))^{-1}\mathbf{r}^\top. \tag{233}$$

Proof. This time we consider the Lagrangian

$$f(\tau) = D_\phi(\boldsymbol{\vartheta}(t)||\boldsymbol{\vartheta}(\tau)) + \lambda(\mu_* - \langle \mathbf{r}, \boldsymbol{\alpha}(\tau)\rangle).$$

Since

$$D_\phi(\boldsymbol{\vartheta}(t)||\boldsymbol{\vartheta}(\tau)) = K(\boldsymbol{\vartheta}(t)) - K(\boldsymbol{\vartheta}(\tau)) - \langle \nabla K(\boldsymbol{\vartheta}(\tau)), \boldsymbol{\vartheta}(t) - \boldsymbol{\vartheta}(\tau)\rangle,$$

the first order condition is given by

$$-\boldsymbol{\vartheta}'(\tau)\nabla^2 K(\boldsymbol{\vartheta}(\tau))(\boldsymbol{\vartheta}(t) - \boldsymbol{\vartheta}(\tau))^\top = \lambda\langle \mathbf{r}, \boldsymbol{\alpha}'(\tau)\rangle. \tag{234}$$

Assuming again that

$$\boldsymbol{\alpha}'(\tau) = -\boldsymbol{\vartheta}'(\tau),$$

one has

$$\nabla^2 K(\boldsymbol{\vartheta}(\tau))(\boldsymbol{\vartheta}(t) - \boldsymbol{\vartheta}(\tau))^\top = \lambda \mathbf{r}.$$

Therefore

$$\boldsymbol{\vartheta}(t) - \boldsymbol{\vartheta}(\tau) = \lambda(\nabla^2 K(\boldsymbol{\vartheta}(\tau)))^{-1}\mathbf{r}^\top.$$

Taking traces, one obtains

$$\langle \boldsymbol{\vartheta}(t) - \boldsymbol{\vartheta}(\tau), \mathbf{1}\rangle = \lambda \mathbf{1}(\nabla^2 K(\boldsymbol{\vartheta}(\tau)))^{-1}\mathbf{r}^\top.$$

For the setting

$$\boldsymbol{\vartheta}(\tau) = \boldsymbol{\vartheta}(t) - \boldsymbol{\alpha}(\tau)$$

we conclude that

$$1 = \lambda \mathbf{1}(\nabla^2 K(\boldsymbol{\theta}(t)))^{-1}\mathbf{r}^\top,$$

obtaining an explicit expression for λ, that is, for the shadow price (21). We have

$$\boldsymbol{\alpha}(\tau) = \frac{1}{\mathbf{1}(\nabla^2 K(\boldsymbol{\vartheta}(\tau)))^{-1}\mathbf{r}^\top}(\nabla^2 K(\boldsymbol{\vartheta}(\tau)))^{-1}\mathbf{r}^\top, \tag{235}$$

finishing the proof of the theorem. □

Remark 4.1 *These theorems and the mean-divergence problem in (223) motivate a geometric steepest descent of the form*

$$\boldsymbol{\alpha}(\tau_{k+1}) = \boldsymbol{\vartheta}(t) - \boldsymbol{\vartheta}(\tau_{k+1}) = \boldsymbol{\vartheta}(t) + \boldsymbol{\gamma}_{\boldsymbol{\vartheta}(\tau_k)}(-\mu\nabla D_\phi(\boldsymbol{\vartheta}(\tau_k);\, \cdot)) \tag{236}$$

where $\boldsymbol{\gamma}$ is a step-size parameter.

Now we present some explicit expressions for (233) in the case of returns distributed according a q-Gaussian probability density. First, it is worth to recall our definitions and conventions for q-Gaussian multidimensional distribution presented in Section 3.2.

Definition 4.1 *Given a vector $\mu \in \mathbb{R}^N$ and a positive definite $N \times N$ matrix Σ, a q-Gaussian multivariate distribution with parameters (μ, Σ) is defined by*

$$p(\mathbf{r}; \mu, \Sigma) = \frac{1}{\Gamma_q}\exp_q\Big(- \frac{1}{2\gamma_q}(\mathbf{r}-\mu)^\top \Sigma^{-1}(\mathbf{r}-\mu)\Big), \tag{237}$$

with

$$\gamma_q = \frac{1}{2}\big((N+4) - (N+2)q\big),$$

and

$$\Gamma_q = K_{q,N}\sqrt{|\Sigma|},$$

where

$$K_{q,N} = \begin{cases} \frac{\Gamma(\frac{1}{q-1} - \frac{N}{2})\sqrt{\pi}}{\Gamma(\frac{1}{q-1})}\big(\frac{1}{q-1}\big)^{\frac{N}{2}}\big((N+4)-(N+2)q\big)^{\frac{N}{2}}, & \text{for} \quad 1 < q < \frac{N+4}{N+2}, \\ \frac{\Gamma(\frac{2-q}{1-q})\sqrt{\pi}}{\Gamma(\frac{2-q}{q-1} + \frac{N}{2})}\big(\frac{1}{1-q}\big)^{\frac{N}{2}}\big((N+4)-(N+2)q\big)^{\frac{N}{2}}, & \text{for} \quad q < 1, \end{cases} \tag{238}$$

and $|\Sigma|$ *denotes the determinant of* Σ.

In this case, recall that the parameter $\boldsymbol{\vartheta} = (\vartheta, \Theta) \in \mathbb{R}^N \times M(N, \mathbb{R})$ is defined in Proposition 3.1 by

$$\vartheta = \Sigma_q^{-1}\mu, \quad \Theta = \frac{1}{2}\Sigma_q^{-1}, \tag{239}$$

where

$$\Sigma_q = \gamma_q K_{q,N}^{1-q} |\Sigma|^{\frac{1-q}{2}} \Sigma. \tag{240}$$

Hence

$$K(\boldsymbol{\vartheta}) = K(\vartheta, \Theta) = \frac{1}{4}\vartheta^\top \Theta \vartheta - \log_q \left(J_q |\Theta|^{\frac{1}{2+(1-q)N}}\right), \tag{241}$$

where

$$J_q = (2\gamma_q)^{\frac{N}{2+(1-q)N}} K_{q,N}^{\frac{-2}{2+(1-q)N}} \tag{242}$$

we obtain

$$\nabla^2 K(\theta, \cdot) = \Sigma_q.$$

From the result above we may conclude what is described as the following corollary.

Corollary 4.1 *The optimal portfolio with respect to the mean-divergence problem (226)-(227) under the assumption of returns q-Gaussian distributed is given by*

$$\boldsymbol{\alpha} = \frac{\Sigma_q^{-1}\mathbf{1}}{\mathbf{1}\Sigma_q^{-1}\mathbf{1}^\top}, \tag{243}$$

where

$$\Sigma_q = \gamma_q K_{q,N}^{1-q} |\Sigma|^{\frac{1-q}{2}} \Sigma.$$

is the q-Gaussian analog of the covariance matrix.

4.4 A natural gradient search

Recall that inspired by (30), (31) and the considerations above we have defined the risk premium Π associated to a ϕ-exponential family as

$$\Pi(\boldsymbol{\alpha}; \boldsymbol{\vartheta}) = \frac{1}{a_\phi} D_\phi[\boldsymbol{\vartheta} - a\boldsymbol{\alpha} || \boldsymbol{\vartheta}], \tag{244}$$

where

$$a_\phi = a\,\frac{\dot{\phi}^2 - \ddot{\phi}\phi}{\dot{\phi}},$$

with derivatives evaluated at $v = \exp_\phi(a\mathbb{E}_\phi[w])$. See expressions 223 and 224 above. Given a fixed ϕ-expected value of the returns, maximizing the certainty equivalent is equivalent to minimizing the risk premium. It follows directly from (244) that minimizing the risk premium is equivalent to minimizing the divergence. So, the search method is then developed according to this equivalence as described in the sequel.

Let $\boldsymbol{\alpha}(\tau)$ be a curve in $\mathcal{S}_{N-1}$ with $\boldsymbol{\alpha}(0) = \boldsymbol{\alpha}$. It follows from (244) that the gradient of Π with respect to the variable $\boldsymbol{\alpha}$ is given by

$$\nabla\Pi(\boldsymbol{\alpha};\boldsymbol{\vartheta}) = \frac{a}{a_\phi}\big(\nabla K(\boldsymbol{\vartheta}) - \nabla K(\boldsymbol{\vartheta} - a\boldsymbol{\alpha})\big), \tag{245}$$

with

$$\nabla K(\boldsymbol{\vartheta}) = \boldsymbol{\eta} = \mathbb{E}_{\hat{p}(\cdot;\boldsymbol{\vartheta})}[\mathbf{r}].$$

Then we consider a steepest descent algorithm of the form

$$\boldsymbol{\alpha}(t_{k+1}) = \boldsymbol{\alpha}(t_k) - \gamma\nabla^{\mathcal{S}_{N-1}}\Pi(\boldsymbol{\alpha}_{t_k};\boldsymbol{\vartheta}_{t_k}), \tag{246}$$

where γ is the fixed step to each iteration and

$$\boldsymbol{\vartheta}_{t_k} = \boldsymbol{\vartheta}_{t_{k-1}} - a\boldsymbol{\alpha}_{t_k}$$

and the differential operator $\nabla^{\mathcal{S}_{N-1}}$ indicates the natural gradient (with respect to $\boldsymbol{\alpha}$) in the $(N-1)$-dimensional simplex $\mathcal{S}_{N-1}$, whose expression is

$$\nabla^{\mathcal{S}_{N-1}}\Pi(\boldsymbol{\alpha};\boldsymbol{\vartheta}) = \left(\frac{\partial\Pi}{\partial\alpha_1} - \nu, \ldots, \frac{\partial\Pi}{\partial\alpha_{N-1}} - \nu, -\nu\right) \tag{247}$$

with

$$\nu = \frac{1}{N}\left(\frac{\partial\Pi}{\partial\alpha_1} + \ldots + \frac{\partial\Pi}{\partial\alpha_{N-1}}\right).$$

The expression (247) is the projection of the Euclidean gradient that is tangent to the simplex $\mathbb{S}_{N-1}$. In sum, we define the following natural gradient search algorithm: one sets the initial allocation as

$$\boldsymbol{\alpha}_0 = (\bar{\mathbf{r}}, I_{N\times N})$$

where $\bar{\mathbf{r}}$ is the mean of the historical returns on the N assets. Hence, the initial approximation to the parameters $\boldsymbol{\vartheta}_0 = (\vartheta_0, \Theta_0)$ of the distribution is given by

$$\Theta_0^{-1} = (2+(1-q)N)^{\frac{2+(1-q)N}{2}} J_q^{-\frac{1}{2}(1-q)(2+(1-q)N)} |M|^{\frac{1-q}{2}} M,$$
$$\vartheta_0 = 2\Theta_0\bar{\mathbf{r}},$$

where M is the $N\times N$ matrix whose components are given as

$$M_{ij} = -\overline{\mathbf{r}_i\mathbf{r}_j} - \bar{\mathbf{r}}_i\bar{\mathbf{r}}_j.$$

The iterative procedure is then defined by the following steps: at each iteration, set the iterated allocation vector $\boldsymbol{\alpha}_{t_{k+1}} = (\alpha_{t_{k+1}}, A_{t_{k+1}})$ as

$$\alpha_{t_{k+1}} = \alpha_{t_k} - \mu \nabla^{\mathcal{S}_{N-1}} \pi(\boldsymbol{\alpha}_{t_k}; \boldsymbol{\vartheta}_{t_k})$$

and $A_{t_{k+1}} = I_{N\times N}$. Then, define the iterated adjusted parameter of the distribution by

$$\nabla K^*(\boldsymbol{\alpha}_{t_{k+1}}) = \boldsymbol{\vartheta}_{t_{k+1}},$$

where K^* is the Legendre transform of K, that is, the dual cumulating function in (150). In the case of an underlying q-Gaussian multivariate distribution one has

$$\begin{aligned} K^*(\boldsymbol{\eta}) &= \frac{1}{1-q}\left(\frac{1}{h_q(\boldsymbol{\vartheta})} - 1\right), \\ &= -\frac{N^2}{2+(1-q)N} J_q^{1-q} |\Theta|^{\frac{1-q}{2+(1-q)N}} + \log_q\left(J_q |\Theta|^{\frac{1}{2+(1-q)N}}\right). \end{aligned}$$

In this case, one obtains

$$\begin{aligned} &\Theta_{t_{k+1}}^{-1} = (2+(1-q)N)^{\frac{2+(1-q)N}{2}} J_q^{-\frac{1}{2}(1-q)(2+(1-q)N)} |B_{t_{k+1}}|^{\frac{1-q}{2}} B_{t_{k+1}}, \\ &\vartheta_{t_{k+1}} = 2\Theta_{t_{k+1}} \alpha_{t_{k+1}}, \end{aligned}$$

where

$$B_{t_{k+1}} = -A_{t_{k+1}} - \alpha_{t_{k+1}} \otimes \alpha_{t_{k+1}}.$$

The procedure is iterated a number of times determined by the choice of parameters a and μ.

4.5 Numerical examples and analysis

In this section we describe some numerical experiments that support the generalized mean-divergence for q-exponential distributions as a valuable alternative to the classical mean-variance strategy. Here our aim is twofold: (i) to compare performance measures of optimal allocation portfolios according to both methods, and; (ii) to collect some evidence of how the parameter q affects the performance of optimal portfolios, particularly in what concerns their divergence.

The experiments also allowed us to recover the optimal allocation obtained directly from (243) by a learning procedure that adjusts the optimal portfolio at each iteration. In this way, it is not mandatory to previously choose an estimator for the deformed variance-covariance matrix Σ_q. In other terms, the iterative algorithm can be applied even without defining an estimator for the Hessian of the cumulant function as needed to apply directly the formula (233). However, it turns out that the numerical results give some indirect evidence of theoretical estimators as, for instance,

$$\widehat{\Sigma}_q = \gamma_q K_{q,N}^{1-q} |\widehat{\Sigma}|^{\frac{1-q}{2}} \widehat{\Sigma}.$$

Our preliminary dataset is composed by weekly USD total arithmetic returns of four market indices on the years 2000–2016, namely: (i) Russell 2000, benchmark of a stock market index of *small caps* mutual funds; (ii) Euronext 100, a *blue chip* index of the largest and most liquid stocks traded on Euronext; (iii) Hang Seng, stock market index that comprises the largest companies of the Hong Kong stock market; and (iv) Nikkei, the most widely quoted average of Japanese equities. It is obvious that other choices are allowed. It remains the important question about the robustness of our findings with respect to different choices of market indices.

In order to determine the initial optimal portfolio weights $\boldsymbol{\alpha}$ we run the natural gradient search algorithm described in Section 4.4 considering the estimation window with the data from the first five years. The optimal portfolios, chosen relatively to a choice of the parameter q, are evaluated in the next twelve months. The estimation window is then shifted twelve months and then the same procedure is applied. In this way, we have registered the performance evolution of risk measures as the evolution of the Bregman divergences in a total of 12 out-of-sample annual rolling windows.

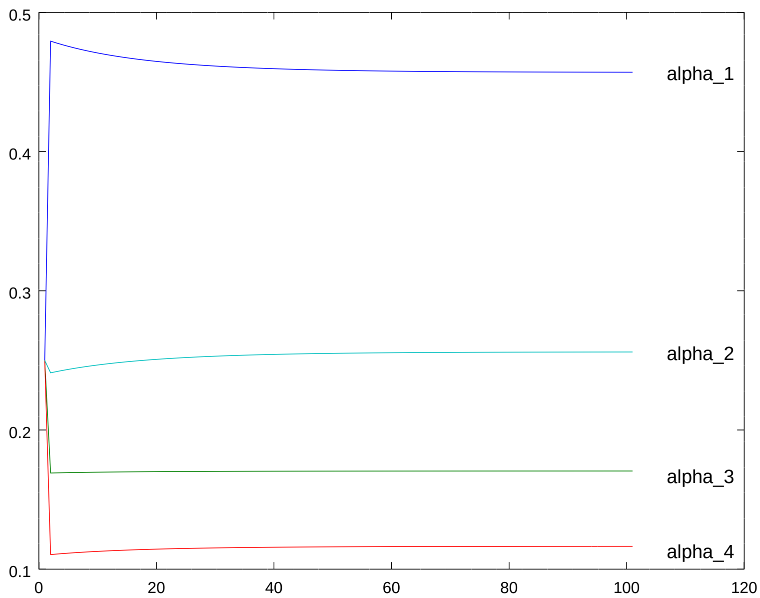

Figura 1: Temporal evolution of weights $\alpha_1, \ldots, \alpha_4$ for the q-portfolio method.

The evolution of the weights for the assets in the portfolio (values of each element of $\boldsymbol{\alpha}$) across the iterations of the gradient method is shown in Figure 1. The components of $\boldsymbol{\alpha}$ illustrated in this figure correspond to the *initial* allocation portfolio obtained by the iterative gradient descent starting with the data of the first five years.

As one can easily see, the method converges fast enough to track the stationary points within a short time interval. In this figure, $alpha_i$ is the weight for the i-th asset. Figure 1 above shows the outcomes of the numerical algorithm starting from a naive allocation of 1/4 to each one of the four indices and then converging after a relatively small number of iterations (numbered at the horizontal axis) to the optimal allocation (shown at the vertical axis) found by the steepest descent. The coloured lines describe the proportion of each one of the four indices at each iteration.

Figure 2 shows the cumulated returns from two different portfolios choices: the blue line stands for the cumulated returns of the mean-divergence portfolio modeled by the assumption of q-Gaussian distribution for returns whereas the cumulated returns of the Markowitz's mean-variance portfolio are plotted in red. The horizontal axis displays the time measured in weeks whereas the vertical axis corresponds to the cumulated returns.

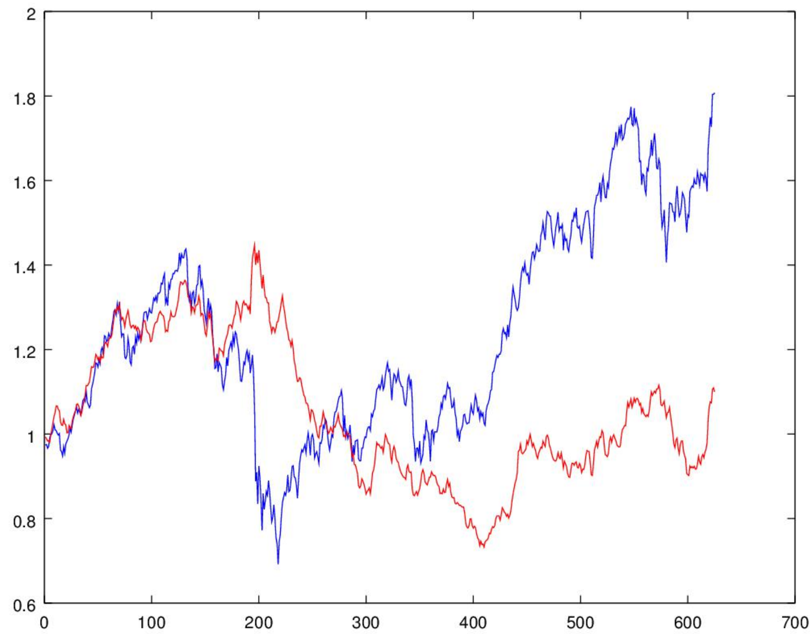

Figura 2: Cumulated returns for proposed method (q-portfolio) and Markowitz's portfolio.

One can easily see that both methods have a similar performance in the initial period (up to ∼150 weeks) when the proposed method (q-portfolio) is outperformed by the Markowitz's method. However, the q-portfolio is more suitable to learn from the system presenting a much faster and consistent recovering of the returns, presenting then a much greater result in terms of the cumulative returns. This indicates the ability of the proposal to deal with heavy tail distributions, usually observed in crisis periods.

Table 1 summarizes the performance of the q-portfolio in every twelve-months windows compared with the Markowitz's classical portfolio. The first (respectively, second) row in the table gives the annual cumulative returns of a q-portfolio determined once and for all from the first five years sample (respectively, rebalanced with respect to the moving five years windows) whereas the third row lists the returns of the Markowitz's portfolio with annual recompositions. In sum, the cumulative returns corresponding to the first (respectively, second) q-portfolio obtained by the steepest descent algorithm for the mean-divergence method outperform that resulting from the mean-variance portfolio in 9 of the 12 (respectively, in 10 of the 12) annual periods under analysis.

It is worth remarking in particular that the out-of-sample performance of the q-portfolios is better in comparison with the Markowitz's classical portfolio in the years of economic recovery after the financial crisis of 2007-2008.

Tabela 1: Annual performance comparison for portfolio selection methods.

Annual Return (%)	2005	2006	2007	2008	2009	2010	2011	2012
q-portfolio (fixed)	19.91	12.12	-5.22	-22.02	-8.39	1.29	-9.36	9.43
q-portfolio (rebalanced)	19.91	16.03	-3.60	-4.35	16.50	9.59	-11.04	9.67
Markowitz	18.56	11.82	-3.59	-5.66	-21.6	-9.71	-8.62	-15.03

Annual Return (%)	2013	2014	2015	2016
q-portfolio (fixed)	40.98	7.55	3.92	8.72
q-portfolio (rebalanced)	40.34	8.72	3.46	17.63
Markowitz	27.04	1.16	1.12	1.03

In what concerns risk measures, we refer to Table 2 that displays the Bregman divergences of the first q-portfolio and the Markowitz's portfolio in two different periods.

Tabela 2: Risk measures: Bregman divergences.

Divergence	2007-2008	2008-2009	2010-2011	2011-2012
q-portfolio	0.2373	0.1920	0.0072	0.0435
Markowitz	0.0459	0.0053	0.0364	0.1263

The statistical divergences measure in what extent the portfolios were rebalanced from one year to another: by definition, higher divergences correspond to higher differences between successive portfolios. The table indicates that the Markowitz's portfolio would have been only slighted rebalanced along the critical years from 2007 to 2009 and considerably modified on the recovery period from 2010 to 2012. In terms of returns, this reinforces the fact that the Markowitz's strategy would allow to earn smaller returns in the recovery years than the q-portfolio. The higher q-divergences of the q-portfolio are an evidence that its allocation strategy is more responsive to riskier scenarios even taking into account initial losses.

.

4.6 A technical appendix

Proposition 4.1 *The natural gradient of the risk premium* Π *as a function defined in the* $(N-1)$*-dimensional simplex* $\mathcal{S}_{N-1}$ *is given by*

$$\nabla^{\mathcal{S}_{N-1}}\Pi(\boldsymbol{\alpha};\boldsymbol{\vartheta}) = \left(\frac{\partial \Pi}{\partial \alpha_1} - \frac{1}{N}\left(\frac{\partial \Pi}{\partial \alpha_1} + \ldots + \frac{\partial \Pi}{\partial \alpha_{N-1}}\right), \ldots, \right.$$
$$\left. \frac{\partial \Pi}{\partial \alpha_{N-1}} - \frac{1}{N}\left(\frac{\partial \Pi}{\partial \alpha_1} + \ldots + \frac{\partial \Pi}{\partial \alpha_{N-1}}\right), -\frac{1}{N}\left(\frac{\partial \Pi}{\partial \alpha_1} + \ldots + \frac{\partial \Pi}{\partial \alpha_{n-1}}\right)\right). \tag{248}$$

Proof. Recall that inspired by (30) and the considerations above we have defined the risk premium associated to a ϕ-exponential family as the quantity

$$\Pi(\boldsymbol{\alpha};\boldsymbol{\vartheta}) = \frac{1}{a_\phi} D_\phi[\boldsymbol{\vartheta} - a\boldsymbol{\alpha}||\boldsymbol{\vartheta}],$$

where

$$a_\phi = a\,\frac{\dot{\phi}^2 - \ddot{\phi}\phi}{\dot{\phi}}$$

with derivatives evaluated at $v = \exp_\phi(a\mathbb{E}_\phi[w])$. Given a fixed ϕ-expected value of the returns, maximizing the certainty equivalent is equivalent to minimizing the risk premium. It turns out that minimizing the risk premium is equivalent to minimizing the divergence.

Let $\boldsymbol{\alpha}(\tau)$ be a curve in $\mathcal{S}_{N-1}$ with $\boldsymbol{\alpha}(0) = \boldsymbol{\alpha}$. Then the gradient of Π with respect to the variables $\boldsymbol{\alpha}$ is given by

$$\langle \nabla\Pi(\boldsymbol{\alpha};\boldsymbol{\vartheta}), \alpha'(\tau)\rangle = \frac{1}{a_\phi}\frac{\mathrm{d}}{\mathrm{d}\tau}\Big|_{\tau=0} D_\phi[\boldsymbol{\vartheta} - a\boldsymbol{\alpha}(\tau)||\boldsymbol{\vartheta}]$$
$$= \frac{1}{a_\phi}\big(\langle \nabla K(\boldsymbol{\vartheta} - a\boldsymbol{\alpha}(\tau)), -a\boldsymbol{\alpha}'(\tau)\rangle - \langle \nabla K(\boldsymbol{\vartheta}), -a\boldsymbol{\alpha}'(\tau)\rangle\big)$$

from what we conclude that

$$\nabla\Pi(\boldsymbol{\alpha};\boldsymbol{\vartheta}) = \frac{a}{a_\phi}\big(\nabla K(\boldsymbol{\vartheta}) - \nabla K(\boldsymbol{\vartheta} - a\boldsymbol{\alpha})\big). \tag{249}$$

Note that

$$\nabla K(\boldsymbol{\vartheta}) = \boldsymbol{\eta} = \mathbb{E}_{\hat{p}(\cdot;\boldsymbol{\vartheta})}[\mathbf{r}].$$

Then we consider a steepest descent algorithm of the form

$$\boldsymbol{\alpha}(t_{k+1}) = \boldsymbol{\alpha}(t_k) - \mu\nabla^{\mathcal{S}_{N-1}}\Pi(\boldsymbol{\alpha}_{t_k};\boldsymbol{\vartheta}_{t_k}), \tag{250}$$

where

$$\boldsymbol{\vartheta}_{t_k} = \boldsymbol{\vartheta}_{t_{k-1}} - a\boldsymbol{\alpha}_{t_k}$$

and the differential operator $\nabla^{\mathcal{S}_{N-1}}$ standing for the natural gradient (with respect to $\boldsymbol{\alpha}$)

in the $(N-1)$-dimensional simplex

$$\mathcal{S}_{N-1} = \{\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{N-1}, \alpha_N) \in \mathbb{R}^N : \alpha_1 + \ldots + \alpha_N = 1,\, 0 \le \alpha_i \le 1\}. \tag{251}$$

whose expression is

$$\nabla^{\mathcal{S}_{N-1}} \Pi(\boldsymbol{\alpha}; \boldsymbol{\vartheta}) = \nabla \Pi - \frac{1}{N} \langle \nabla \Pi, (1, \ldots, 1, 0) \rangle (1, \ldots, 1, 1)$$

where $\nabla\Pi$ is the Euclidean gradient of Π as a function defined in the $(N-2)$-dimensional simplex

$$\mathcal{F}_N = \{(\alpha_1, \ldots, \alpha_{N-1}, 0) \in \mathbb{R}^N : \alpha_1 + \ldots + \alpha_{N-1} \le 1,\ \alpha_i \ge 0\}.$$

Note that

$$\alpha_N = 1 - \alpha_1 - \ldots - \alpha_{N-1}$$

if and only if $(\alpha_1, \ldots, \alpha_{N-1}, \alpha_N) \in \mathcal{S}_{N-1}$ with $(\alpha_1, \ldots, \alpha_{N-1}, 0) \in \mathcal{F}_N$. This means that we are describing $\mathcal{S}_{N-1}$ as the graph of the function

$$\alpha_N = f(\alpha_1, \ldots, \alpha_{N-1}) = 1 - \alpha^1 - \ldots - \alpha_{N-1}$$

defined em $\mathcal{F}_N$. Hence, for a fixed $\boldsymbol{\vartheta}$,

$$\nabla \Pi = \left(\frac{\partial \mathcal{P}}{\partial \alpha_1}, \ldots, \frac{\partial \mathcal{P}}{\partial \alpha_{N-1}}, 0 \right),$$

and

$$\nabla^{\mathcal{S}_{N-1}} \Pi(\boldsymbol{\alpha}; \boldsymbol{\vartheta}) = \left(\frac{\partial \Pi}{\partial \alpha_1}, \ldots, \frac{\partial \Pi}{\partial \alpha_{N-1}}, 0 \right) - \frac{1}{N} \left(\frac{\partial \mathcal{P}}{\partial \alpha_1} + \ldots + \frac{\partial \Pi}{\partial \alpha_{N-1}} \right) (1, \ldots, 1, 1). \tag{252}$$

In other terms

$$\begin{aligned} \nabla^{\mathcal{S}_{N-1}} \Pi(\boldsymbol{\alpha}; \boldsymbol{\vartheta}) = \Bigg(& \frac{\partial \Pi}{\partial \alpha_1} - \frac{1}{N} \left(\frac{\partial \Pi}{\partial \alpha_1} + \ldots + \frac{\partial \Pi}{\partial \alpha_{N-1}} \right), \ldots, \\ & \frac{\partial \Pi}{\partial \alpha_{N-1}} - \frac{1}{N} \left(\frac{\partial \Pi}{\partial \alpha_1} + \ldots + \frac{\partial \Pi}{\partial \alpha_{N-1}} \right), -\frac{1}{N} \left(\frac{\partial \Pi}{\partial \alpha_1} + \ldots + \frac{\partial \Pi}{\partial \alpha_{n-1}} \right) \Bigg) \end{aligned} \tag{253}$$

This finishes the proof of the proposition. □

Note that in order to have each new iteration in (250) yielding a new point in $\mathcal{S}_{N-1}$ we impose for a fixed $\boldsymbol{\vartheta}$ that

$$0 \le \alpha^i - \mu \left(\frac{\partial \Pi}{\partial \alpha^i} - \frac{1}{N} \left(\frac{\partial \Pi}{\partial \alpha_1} + \ldots + \frac{\partial \Pi}{\partial \alpha_{N-1}} \right) \right) \le 1. \tag{254}$$

It is enough to assume that

$$\mu \leq \frac{\alpha^i}{\left(\frac{\partial \Pi}{\partial \alpha_i} - \frac{1}{N}\left(\frac{\partial \Pi}{\partial \alpha_1} + \ldots + \frac{\partial \Pi}{\partial \alpha_{N-1}}\right)\right)}, \tag{255}$$

the derivatives taken at $\boldsymbol{\alpha}$, and

$$\mu \geq \frac{\alpha_i - 1}{\left(\frac{\partial \Pi}{\partial \alpha_i} - \frac{1}{N}\left(\frac{\partial \Pi}{\partial \alpha_1} + \ldots + \frac{\partial \Pi}{\partial \alpha_{N-1}}\right)\right)}. \tag{256}$$

5 SOME GENERALIZATIONS OF CAPM

5.1 The space of financial assets

Following (16) and (19), one models the set $\mathcal{M}$ spanned by financial assets as a subspace of a linear space $\mathcal{H}$ of contingent claims. More precisely, every point in $\mathcal{M}$ corresponds to the payoff z of a contingent claim at a fixed time, say $t = 1$, that is, a random variable

$$z = z(s),$$

where s are the states of the world with probability distribution specified by some density $p(s; \boldsymbol{\vartheta})$. Here, $\boldsymbol{\vartheta}$ is the distribution parameter of a family of probability distributions whose densities define a n-dimensional statistical manifold

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

where $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_n)$ takes values in some open subset U of the n-dimensional Euclidean space $\mathbb{R}^n$.

Example 5.1 *Suppose that $\mathcal{M}$ is spanned by finitely many assets and that the space of states of the world is also finite-dimensional. Then each $z \in \mathcal{M}$ is determined by their possible payoffs in distinct and also finitely many states of the world, say, $\{s_1, \ldots, s_L\}$. In other terms, each $z \in \mathcal{M}$ is described by a L-dimensional vector*

$$z = (z(s_1), \ldots, z(s_L)).$$

The discrete *probability distribution of the states of the world is denoted by $p_i = p(s_i; \boldsymbol{\vartheta})$, $i = 1, \ldots, L$, where $\boldsymbol{\vartheta}$ only indicates the parameters of this distribution.*

In sum, a point $z \in \mathcal{M}$ corresponds to the possible payoffs of a given asset under the distinct states of the world. The probability distribution of these states is given by a probability density in a statistical manifold.

For the sake of simplicity, in what follows we will restrict ourselves to a finite-dimensional asset span $\mathcal{M}$. The case of infinite-dimensional linear (Hilbert) spaces can be handled with some notational and technical adjustments.

5.2 Deformed exponentials and optimal ϕ-portfolios

In order to fix ideas, we are going to consider a statistical manifold of ϕ-exponential probability densities of the form

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

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 ϕ-exponential defined as the inverse function of the ϕ-logarithm (24), (25)

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

where $\phi : (0,+\infty) \to (0,+\infty)$ 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 $q > 0$, what corresponds to set $\phi(t) = t^q$, Hence, the q-logarithm is defined by

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

The cumulant function K defines a Bregman divergence given by

$$D(z||w) = K(\boldsymbol{\vartheta}) - K(\boldsymbol{\vartheta}') - \langle \nabla K(\boldsymbol{\vartheta}'), \boldsymbol{\vartheta} - \boldsymbol{\vartheta}'\rangle.$$

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

Setting $\phi(t) = t$ one gets the family of exponential distributions, in particular multivariate Gaussian distributions. For this family, R. Nock, B. Magdalou, E. Bryis and F. Nielsen (30), (31) represented the key concepts of Portfolio Selection theory in terms of the cumulant function and the associated Bregman divergence. More precisely, they proved that for CARA utility functions the certainty equivalent and risk premium of risky assets are respectively given by

$$C = \frac{1}{a}\big(K(\boldsymbol{\vartheta}) - K(\boldsymbol{\vartheta}')\big)$$

and

$$\Pi = \frac{1}{a}D[\boldsymbol{\vartheta}|\boldsymbol{\vartheta}'],$$

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

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

In the particular case of Gaussian distributed returns, one recovers the classical Markowitz's optimal portfolio allocation vector

$$\boldsymbol{\alpha} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}, \tag{257}$$

where Σ is the covariance matrix of the returns on the assets.

As summarized in Chapter 4, C. Cavalcante, I. Guerreiro and A. F. Rodrigues extended in (10) this approach to ϕ-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

$$\boldsymbol{\alpha} = \frac{\nabla^2 K(\boldsymbol{\vartheta})^{-1}\mathbf{1}}{\mathbf{1}^T \nabla^2 K(\boldsymbol{\vartheta})^{-1} \mathbf{1}}. \tag{258}$$

Note that the Hessian of the moment-generating (convex) function is positive-definite and plays the role of the variance-covariance matrix in the Gaussian case. In the particular case of q-Gaussian distributions (32), the optimal allocation portfolio is given by

$$\boldsymbol{\alpha} = \frac{\Sigma_q^{-1}\mathbf{1}}{\mathbf{1}^T \Sigma_q^{-1} \mathbf{1}} \tag{259}$$

where

$$\Sigma_q = \gamma_q K_{q,N}^{1-q} |\Sigma|^{\frac{1-q}{2}} \Sigma \tag{260}$$

with

$$\gamma_q = \frac{1}{2}\big((N+4) - (N+2)q\big) \tag{261}$$

and

$$K_{q,N} = \begin{cases} \frac{\Gamma(\frac{1}{q-1} - \frac{N}{2})\sqrt{\pi}}{\Gamma(\frac{1}{q-1})} \left(\frac{1}{q-1}\right)^{\frac{N}{2}} \big((N+4) - (N+2)q\big)^{\frac{N}{2}}, & \text{for} \quad 1 < q < \frac{N+4}{N+2}, \\ \frac{\Gamma(\frac{2-q}{1-q})\sqrt{\pi}}{\Gamma(\frac{2-q}{q-1} + \frac{N}{2})} \left(\frac{1}{1-q}\right)^{\frac{N}{2}} \big((N+4) - (N+2)q\big)^{\frac{N}{2}}, & \text{for} \quad q < 1. \end{cases} \tag{262}$$

Here $|\Sigma|$ is the determinant of Σ. We refer the reader to (32) for further details in q-multivariate Gaussian distributions. It is evident that one reobtains the Markowitz's portfolio for $q = 1$ in (259).

In view of (257), we have elaborated in Chapter 4 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 (10) 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.

5.3 Notation and main results

In (16) and (19), S. LeRoy, J. Werner and D. Luenberger have developed a geometric approach to the mean-variance analysis in terms of a vector space geometry of projections onto a mean-variance efficient frontier. From this approach, they easily deduce an elegant geometric interpretation of CAPM and factor pricing models.

In what follows, we extend their geometric methods to divergence geometries in $\mathcal{M}$ instead of the Hilbert space norm.

Before stating our main results, we fix some notation and basic definitions. Let K^* be the Legendre transform of K, that is, the dual cumulant function

$$K^*(\boldsymbol{\eta}) = \max_{\boldsymbol{\vartheta}} \big(\langle \boldsymbol{\vartheta}, \boldsymbol{\eta} \rangle - K(\boldsymbol{\vartheta})\big)$$

given by the relative negative ϕ-entropy (4)

$$K^*(\boldsymbol{\eta}) = \mathbb{E}_\phi[\log_\phi p(s, \boldsymbol{\vartheta})/p_0(s)]$$

where $\boldsymbol{\eta}$ is the dual affine coordinate defined by

$$\boldsymbol{\eta} = \nabla K(\boldsymbol{\vartheta}).$$

In the case of q-exponential distributions one has

$$K^*(\boldsymbol{\eta}) = \frac{1}{1-q}\left(\frac{1}{h(\boldsymbol{\vartheta})-1}\right)$$

with

$$h(\boldsymbol{\vartheta}) = \int \big(p(s, \boldsymbol{\vartheta})/p_0(s)\big)^q \, p_0(s)\, \mathrm{d}s.$$

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

$$g|_{T_z\mathcal{M}} = \nabla^2 K(\boldsymbol{\vartheta}), \tag{263}$$

where $\boldsymbol{\vartheta}$ is the statistical parameter of the distribution $p(s, \boldsymbol{\vartheta})$ of the payoff $z = z(s)$. This metric can be expanded in local coordinates around a fixed reference point $o \in \mathcal{M}$

as

$$g \sim \nabla^2 K(o) + o(|z|^2), \tag{264}$$

where quadratic terms are determined in terms of the Riemann curvature of the Riemannian manifold $(\mathcal{M}, g)$, see (15).

Let $\mathcal{M}'$ be a finite dimensional subspace of $\mathcal{M}$ spanned by N traded assets. 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

$$g(k_{\mathsf{e}}, z) = \mathbb{E}[z]$$

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

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

where m is a stochastic discount factor. Here $Q : \mathcal{M} \to \mathbb{R}$ is the price functional, that is, the present value of the expected returns of the asset, discounted at rate m. Since the expectation is not necessarily taken with respect to risk-neutral probabilities, m is a risk-adjusted discount rate, possibly distinct from the risk-free return rate (6), (16).

The Riemannian metric g in (264) is defined in such a way that it is possible to obtain precise expressions for both kernels in terms of the Bregman divergence associated to K, at least for assets in the finite-dimensional subspace $\mathcal{M}'$.

Theorem 5.1 *The expectation and pricing kernels in a N-dimensional subspace $\mathcal{M}'$ of contingent claims in $\mathcal{M}$ are respectively given by*

$$k_{\mathsf{e}} = g^{-1}\big(\sqrt{D(z_1|o)}, \ldots, \sqrt{D(z_N|o)}\big) \tag{265}$$

and

$$k_{\mathsf{q}} = g^{-1}\big(Q(z_1), \ldots, Q(z_N)\big) \tag{266}$$

where $\mathcal{M}' = \operatorname{span}\{z_1, \ldots, z_N\}$.

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 Pytaghorean Theorem for divergences (Theorem 5.4 below) that fixed a reference point $o \in \mathcal{M}'$ one has

$$D(z|o) = D(z^{\mathcal{E}}|o) + D(z|z^{\mathcal{E}}), \quad z \in \mathcal{M}'. \tag{267}$$

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

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

expression (267) reduces to the Euclidean decomposition

$$|z|^2 = \mathbb{E}[z]^2 + \text{var}[z], \tag{268}$$

where

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

is the variance, the classical risk measure in Portfolio Theory (20), (6).

Motivated by the analogy between (267) and (268), we propose in the sequel the projection

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

as a novel risk measure for assets $z \in \mathcal{M}'$ that encodes higher moments of the probability distributions.

The next result states that the two reference assets k_{e} and k_{q} determine the efficient frontier for portfolios of assets in $\mathcal{M}'$. Indeed we have

Theorem 5.2 *Let* $\mathcal{E} = \text{span}\{k_{\mathsf{e}}, k_{\mathsf{q}}\}$ *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

$$\mathcal{P}[z^{\mathcal{E}}] \leq \mathcal{P}[z]$$

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

Denote by R_{e} and R_{q} the returns of k_{e} and k_{q}, respectively. We prove

Theorem 5.3 *The mimimum divergence portfolio in* $\mathcal{M}'$ *is given by*

$$z = R_{\mathsf{e}} + (1 - \beta)(R_{\mathsf{q}} - R_{\mathsf{e}})$$

where

$$\beta = -\frac{g(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{e}})}{g(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{q}} - R_{\mathsf{e}})}$$

with

$$g = \nabla^2 K(z).$$

In the case when there is a risk-free asset $\mathbf{1}$ in $\mathcal{M}$ with return R_{f} we obtain a generalized CAPM expression

$$\mathbb{E}[z] = R_{\mathsf{f}} + \beta(\mathbb{E}[R_{\mathsf{q}}] - R_{\mathsf{f}}). \tag{269}$$

Those expressions extend the classical CAPM formula. For instance, if we suppose that the returns of traded assets are distributed accordingly a q-Gaussian distributions it holds that

$$g|_z = \nabla^2 K(z(\cdot, \boldsymbol{\vartheta})) = \Sigma_q$$

for every $z \in \mathcal{M}'$, where the q-variance matrix Σ_q is defined in Section 5.2.

Corollary 5.1 *Suppose that the traded financial assets in $\mathcal{M}'$ are distributed according to a q-Gaussian distributions. Hence the minimum divergence portfolio given by*

$$z = R_{\mathsf{e}} + (1-\beta)(R_{\mathsf{q}} - R_{\mathsf{e}})$$

where

$$\beta = -\frac{g(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{e}})}{g(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{q}} - R_{\mathsf{e}})}$$

with

$$g = \Sigma_q,$$

the q-variance matrix defined in (332)-(334).

In Section 5.5 we prove that similar beta pricing equations are still valid if one replaces the returns of the expectation and pricing kernels by the returns of two assets in the mean-divergence efficient frontier that are orthogonal with respect to g. This is the case of assets with zero correlation in the classical setting.

Theorem 5.5 in Section 5.5 establishes a generalized CAPM equation based on the maximization of an utility function describing the preferences of a risk averse agent that is strictly decreasing with respect to the risk measure. For such utility functions, it is possible to prove that the market equilibrium portfolio lies on the mean-divergence efficient frontier.

5.4 Geometry of statistical divergences

Motivated by the discussion above, we fix a statistical divergence D on $\mathcal{M}$. Denoting by $\langle \cdot, \cdot \rangle$ the inner product in $\mathcal{M}$ induced from $\mathcal{H}$, we can assume that D is a Bregman divergence of the form

$$D(z|w) = K(z) - K(w) - \langle \nabla K(w), z - w \rangle \tag{270}$$

for some convex function $K : \mathcal{M} \to \mathbb{R}$. Here ∇K is the Fréchet differential of K on $\mathcal{M}$, what corresponds to the usual gradient in the case when $\mathcal{M}$ is finite-dimensional and $\langle \cdot, \cdot \rangle$

is the Euclidean inner product.

A trivial example is the Euclidean divergence given by the L^2-norm itself

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

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

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

or, more generally, the relative ϕ-entropy

$$D_\phi(z(\cdot, \boldsymbol{\vartheta})|w(\cdot, \boldsymbol{\vartheta}')) = \mathbb{E}_{\hat{p}(\cdot, \boldsymbol{\vartheta})}\big[\log_\phi p(\cdot, \boldsymbol{\vartheta})/p_0 - \log_\phi p(\cdot, \boldsymbol{\vartheta}')/p_0\big],$$

where $\hat{p}(s, \boldsymbol{\vartheta})$ is the escort distribution (24), (25) given by

$$\hat{p}(s, \boldsymbol{\vartheta}) = \frac{1}{h(\boldsymbol{\vartheta})}\psi\big(\langle T(z), \boldsymbol{\vartheta}\rangle - K(\boldsymbol{\vartheta})\big)p_0(s)$$

with

$$h(\boldsymbol{\vartheta}) = \int \phi\big(p(s, \boldsymbol{\vartheta}/p_0(s)\big) \,\mathrm{d}s$$

and

$$\psi(t) = \phi\big(\exp_\phi(t)\big).$$

Note that D_{KL} corresponds to D_ϕ for the particular choice of $\phi(t) = t$.

The Hessian of the cumulant function K defines a Riemannian metric g in $\mathcal{M}$ whose contravariant version g^* is the Hessian of the Legendre transform K^* of K. In the particular case of $\phi(t) = t$ that corresponds to the exponential family of distributions we have

$$g|_{z(s)} = \text{var}[z],$$

the variance taken with respect to the probability density

$$p(s, \boldsymbol{\vartheta}) = \exp(\langle T(s), \boldsymbol{\vartheta}\rangle - K(s, \boldsymbol{\vartheta}))\, p_0(\boldsymbol{\vartheta})$$

In general, the metrics g and g^* define a dually flat structure with affine connections whose geodesics are Euclidean lines in terms of the coordinates $\boldsymbol{\vartheta}, \boldsymbol{\eta}$ in $\mathcal{M}$. We refer the reader to (1), (5) for a comprehensive account of those concepts in terms of Information Geometry.

Projections and risk measures

One of the fundamental results in Information Geometry is the Pythagorean Theorem that can be stated as follows

Theorem 5.4 (Theorem 1.2 and Theorem 1.3, (1)) *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*

$$D(z|o) = D(w|o) + D(z|w). \tag{271}$$

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

$$D^*(z|o) = D^*(w|o) + D^*(z|w), \tag{272}$$

where D^* is the dual Bregman divergence

$$D^*(z|w) = K^*(z) - K^*(w) - \langle \nabla K^*(w), z - w \rangle. \tag{273}$$

Now, let $\mathcal{M}'$ be a finite-dimensional (more generally, a closed) subspace of $\mathcal{M}$ and consider the following minimization problem: given $z \in \mathcal{M}$ to find $z' \in \mathcal{M}'$ such that

$$D(z|z') = \min_{w \in \mathcal{M}'} D(z|w). \tag{274}$$

Fixed an arbitrary reference point $o \in \mathcal{M}'$ it follows from Theorem 5.4 that

$$D(z|o) = D(z'|o) + D(z|z'). \tag{275}$$

Suppose that the minimization problem (274) has a unique solution z' that we denote by

$$z' = \pi_{\mathcal{M}'}(z), \tag{276}$$

the projection of z onto $\mathcal{M}'$. Note that in the case of the Euclidean divergence

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

we have

$$|z|^2 = \langle z, \mathbf{1} \rangle^2 + |z^\perp|^2 = \mathbb{E}[z]^2 + \text{var}[z], \tag{277}$$

where $z^\perp = z - \langle z, \mathbf{1} \rangle \mathbf{1}$. Hence the projection of z onto the subspace $\mathcal{M}'$ orthogonal to the vector **1** is given by

$$\mathcal{P}_{\mathcal{M}'}(z) = \mathrm{var}[z].$$

The vector **1** corresponds to an asset whose payoffs are the same for every state of the world, that is, it represents a risk-less asset. We are not assuming that **1** belongs to the finite-dimensional asset span $\mathcal{M}'$.

Comparing (275) and (277) we define expected value and risk premium respectively by

$$\mathbb{E}_{\mathcal{M}'}[z] = \sqrt{D(z'|o)} \tag{278}$$

and

$$\mathcal{P}_{\mathcal{M}'}(z) = D(z'|z). \tag{279}$$

Note that the risk premium is the variance in the particular case of normally distributed asset returns.

It follows from (278) that the distribution of probabilities $(p_1, \ldots, p_k)$ for payoffs $z' = (z'(s_1), \ldots, z'(s_L)) \in \mathcal{M}'$ is determined, up to a scaling, by the Bregman divergence D. For any $i = 1, \ldots, L$, let $\mathsf{e}_i \in \mathcal{M}'$ the asset with payoff 1 at the state of the world s_i and 0 for any s_j with $j \neq i$. Then, up to a scaling,

$$p_i = \sqrt{D(\mathsf{e}_i|o)}, \quad i = 1, \ldots, L.$$

Expectation and price kernels

From now on, we restrict ourselves to the N-dimensional subspace $\mathcal{M}'$ where, as we have seen, the expected values could be determined up to scaling in terms of the divergence D.

We are going to prove the existence of payoff vectors k_{e} and k_{q} in $\mathcal{M}'$ representing the expectation and price functionals defined respectively by

$$\mathsf{e}(z) = \mathbb{E}_{\mathcal{M}'}[z] \tag{280}$$

and

$$Q(z) = \mathbb{E}_{\mathcal{M}'}[mz], \tag{281}$$

for any $z \in \mathcal{M}'$, where m is a stochastic discount factor. In sum, we search for a *expectation kernel* k_{e} in $\mathcal{M}'$ such that

$$(k_{\mathsf{e}}, z)_{T_z\mathcal{M}'} = \mathbb{E}_{\mathcal{M}'}[z] = \sum_{i=1}^{L} p_i z(s_i) \tag{282}$$

for some inner product $(\cdot, \cdot)_{T_z\mathcal{M}'}$ in the tangent space $T_z\mathcal{M}'$ of $\mathcal{M}'$ at the point $z \in \mathcal{M}'$.

In the same way the *pricing kernel* $k_{\mathsf{q}} \in \mathcal{M}'$ must satisfy

$$(k_{\mathsf{q}}, z)_{T_z\mathcal{M}'} = \mathbb{E}_{\mathcal{M}'}[mz] = \sum_{i=1}^{L} p_i m(s_i) z(s_i) \tag{283}$$

Those expressions indicate that we can also determine the inner product $(\cdot,\cdot)_{T_z\mathcal{M}'}$ at $T_z\mathcal{M}'$ only in terms of the divergence D.

In order to do this, we restart the minimization problem (274) above, this time projecting from $\mathcal{M}'$ to the (one or two-dimensional) span $\mathcal{E}$ of $\{k_{\mathsf{e}}, k_{\mathsf{q}}\}$ in $\mathcal{M}'$. More precisely, given $z \in \mathcal{M}'$ the problem now is to find $z^{\mathcal{E}} \in \mathcal{E}$ such that

$$D(z|z^{\mathcal{E}}) = \min_{w\in\mathcal{E}} D(z|w). \tag{284}$$

Fixed an arbitrary point $o \in \mathcal{M}'$ it follows again from Theorem 5.4 that

$$D(z|o) = D(z^{\mathcal{E}}|o) + D(z|z^{\mathcal{E}}), \tag{285}$$

that is,

$$D(z|o) = \mathbb{E}_{\mathcal{M}'}[z]^2 + \mathcal{P}(z), \quad z \in \mathcal{M}'. \tag{286}$$

where

$$\mathcal{P}(z) = D(z|z^{\mathcal{E}}). \tag{287}$$

This last relation implies that $\mathcal{E}$ is the mean-divergence frontier for payoffs in $\mathcal{M}'$, in the sense that given $z \in \mathcal{M}'$ it holds that

$$\mathbb{E}_{\mathcal{M}'}[z] = \mathbb{E}_{\mathcal{M}'}[z^{\mathcal{E}}] \tag{288}$$

and

$$\mathcal{P}(z^{\mathcal{E}}) \le \mathcal{P}(z) \tag{289}$$

with equality if and only if $z = z^{\mathcal{E}}$. We denote $z^{\mathcal{E}} = \pi_{\mathcal{E}}(z)$.

We now deduce an infinitesimal version of the condition (284) in the case when D is a Bregman divergence with cumulating function K. Fixed $z \in \mathcal{M}'$ and a curve $w(t)$ in $\mathcal{E}$ with $w(0) = z^{\mathcal{E}}$ we have

$$\begin{aligned} 0 = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} D(z|w(t)) &= \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \big(K(z) - K(w(t)) - \langle \nabla K(z), z - w(t)\rangle\big) \\ &= -\langle \nabla K(z^{\mathcal{E}}), w'(0)\rangle + \langle \nabla K(z^{\mathcal{E}}), w'(0)\rangle - w'(0)^\top \nabla^2 K(z^{\mathcal{E}})(z - z^{\mathcal{E}}), \end{aligned}$$

where the superscript $a^\top$ denotes the transpose of a. We conclude that

$$w'(0)^\top \nabla^2 K(z^{\mathcal{E}})(z - z^{\mathcal{E}}) = 0 \tag{290}$$

for an arbitrary $w'(0) \in T_{z^{\mathcal{E}}}\mathcal{E}$. Therefore, denoting by ε the returnless asset

$$\varepsilon = z - z^{\mathcal{E}} \tag{291}$$

and denoting

$$g|_{z^{\mathcal{E}}} = \nabla^2 K(z^{\mathcal{E}}), \tag{292}$$

a inner product in $T_{z^{\mathcal{E}}}\mathcal{E}$ due to the convexity of K, we conclude that

$$\varepsilon \perp_{g|_{z^{\mathcal{E}}}} T_{z^{\mathcal{E}}}\mathcal{E} = \mathcal{E}. \tag{293}$$

where in the last equality we have used that $\mathcal{E}$ is a vector space. In particular, it follows that

$$0 = g|_{z^{\mathcal{E}}}(\varepsilon, k_{\mathsf{e}}) = g|_{z^{\mathcal{E}}}(z, k_{\mathsf{e}}) - g|_{z^{\mathcal{E}}}(z^{\mathcal{E}}, k_{\mathsf{e}})$$

and

$$0 = g|_{z^{\mathcal{E}}}(\varepsilon, k_{\mathsf{q}}) = g|_{z^{\mathcal{E}}}(z, k_{\mathsf{q}}) - g|_{z^{\mathcal{E}}}(z^{\mathcal{E}}, k_{\mathsf{q}})$$

Therefore

$$g|_{z^{\mathcal{E}}}(z, k_{\mathsf{e}}) = g|_{z^{\mathcal{E}}}(z^{\mathcal{E}}, k_{\mathsf{e}}) \tag{294}$$

and

$$g|_{z^{\mathcal{E}}}(z, k_{\mathsf{q}}) = g|_{z^{\mathcal{E}}}(z^{\mathcal{E}}, k_{\mathsf{q}}) \tag{295}$$

for any z with projection $z^{\mathcal{E}}$ (that is, z of the form $z^{\mathcal{E}} + t\varepsilon$). Expressions (294) and (295) mean that expected values and prices should be the same for payoffs in $\mathcal{M}'$ with same projection on $\mathcal{E}$.

Comparing (282), (288) and (294) suggests that $g|_{z^{\mathcal{E}}}$ in (292) is the natural choice for a inner product $(\cdot,\cdot)_{T_z\mathcal{M}'}$ in the Hilbert tangent space $T_z\mathcal{M}'$ for any $z \in \mathcal{M}'$ with projection $z^{\mathcal{E}}$. This means that the resulting Riemannian metric in $\mathcal{M}'$ is invariant with respect to translations in directions ε perpendicular to $\mathcal{E}$. In other terms, we have a Riemannian product $\mathcal{M} = \mathcal{E} \times \mathcal{F}$, where $\mathcal{F}$ is ruled by Euclidean lines of the form $z^{\mathcal{E}} + t\varepsilon$, $t \in \mathbb{R}$, with $z^{\mathcal{E}} \in \mathcal{E}$ and $\varepsilon \perp_{g|_{z^{\mathcal{E}}}} T_{z^{\mathcal{E}}}\mathcal{E}$. In particular, the Riemannian curvature of $\mathcal{M}$ is determined by the Gaussian curvature of $\mathcal{E}$ in the case when k_{e} and k_{q} are linearly independent and $\mathcal{M}$ is a two-dimensional vector space.

From now on, we assume that $\mathcal{E}$ is two-dimensional. Hence, setting

$$(\cdot,\cdot)_{T_z\mathcal{M}'} = g|_{z^{\mathcal{E}}} \tag{296}$$

for $z = z^{\mathcal{E}} + t\varepsilon$, $t \in \mathbb{R}$, we determine k_{e} solving the linear system

$$(k_{\mathsf{e}}, z)_{T_z\mathcal{M}'} = g|_{z^{\mathcal{E}}}(k_{\mathsf{e}}, z) = \mathbb{E}_{\mathcal{M}'}[z] = \sum_{j=1}^{L} p_j z(s_j).$$

Denoting the components of $g|_{z^{\mathsf{e}}}$ by g_{ij} the expectation kernel has components given by the solution of the system

$$\sum_{j=1}^{N} g_{ij}(k_{\mathsf{e}})_j = (k_{\mathsf{e}}, z_i)_{T_z\mathcal{M}'} = \sum_{j=1}^{L} p_j z_i(s_j) = \mathbb{E}_{\mathcal{M}'}[z_i] = \sqrt{D(z_i|o)}$$

for every $i = 1. \dots, N$. Therefore, we consider the solution of the system of equations

$$\sum_{j=1}^{N} g_{ij}(k_{\mathsf{e}})_j = \sqrt{D(z_i|o)},$$

that is given by

$$k_{\mathsf{e}} = \big(g|_{z^{\mathcal{E}}}\big)^{-1}\big(\sqrt{D(z_1|o)}, \dots, \sqrt{D(z_N|o)}\big) \tag{297}$$

where

$$g|_{z^{\mathcal{E}}} = \nabla^2 K(z^{\mathcal{E}}). \tag{298}$$

The kernel pricing is determined in a similar way. Hence we have

$$k_{\mathsf{q}} = \big(g|_{z^{\mathcal{E}}}\big)^{-1}\big(Q(z_1), \dots, Q(z_N)\big), \tag{299}$$

where $Q(z_i)$ is the current market price of the basic asset z_i, that is,

$$Q(z_i) = \sum_{j=1}^{k} m(s_j)\sqrt{D(\mathsf{e}_j|o)}z_i(s_j).$$

Note that if K is quadratic, as is the case for the Euclidean divergence D_{euc}, we have

$$\nabla^2 K(z^{\mathcal{E}}) = \nabla^2 K(o),$$

where $o \in \mathcal{M}'$ is the arbitrarily fixed reference point. In this case,

$$g = \nabla^2 K(o)$$

and $\mathcal{M}'$ is a flat Riemannian manifold. In the general case we have an expansion of the form

$$\nabla^2 K(z^{\mathcal{E}}) \simeq \nabla^2 K(o) + o(|z|^2),$$

where the quadratic remainder encodes the Riemannian curvature of $\mathcal{M}'$ and its covariant derivatives. In statistical terms, these curvature terms can be associated to the contribution of higher moments of the underlying probability distributions. For instance, considering a local coordinate system given by the principal directions of $\nabla^2 K(z^{\mathcal{E}})$, one has that the Gaussian cuvature of $\mathcal{E}$ at $z^{\mathcal{E}}$ is given by

$$\mathcal{K}|_{z^{\mathcal{E}}} = -\frac{1}{2\sqrt{\lambda_1\lambda_2}}\left(\partial_1\left(\frac{\partial_1\lambda_2}{\sqrt{\lambda_1\lambda_2}}\right) + \partial_2\left(\frac{\partial_2\lambda_1}{\sqrt{\lambda_1\lambda_2}}\right)\right), \tag{300}$$

where $\lambda_1, \lambda_2 > 0$ are the eigenvalues of $\nabla^2 K(z^{\mathcal{E}})$. Note that the curvature involves third and fourth moments of the distributions $p(s, \boldsymbol{\vartheta})\,\mathrm{d}s$.

5.5 Minimum divergence portfolio

We have proved that $\mathcal{E} = \mathrm{span}\{k_{\mathsf{q}}, k_{\mathsf{e}}\}$ is the mean-divergence frontier in $\mathcal{M}'$. Now we address the problem of minimizing the risk measure

$$\mathcal{P}(z) = D(z|o)$$

among points in $z \in \mathcal{E}$ only, that is,

$$\min_{z\in\mathcal{E}} D(z|o), \tag{301}$$

where $o \in \mathcal{M}'$ is an arbitrarily fixed reference point. Any point $z \in \mathcal{E}$ is of the form

$$z = ak_{\mathsf{q}} + bk_{\mathsf{e}},$$

for some $a, b \in \mathbb{R}$. The price of this portfolio is

$$Q(z) = \mathbb{E}_{\mathcal{M}'}[mz] = a\mathbb{E}_{\mathcal{M}'}[mk_{\mathsf{q}}] + b\mathbb{E}_{\mathcal{M}'}[mk_{\mathsf{e}}] = aQ(k_{\mathsf{q}}) + bQ(k_{\mathsf{e}}) \tag{302}$$

Fixing the constraint that the price of the portfolio is $Q(z) = 1$, we denote

$$\beta = aQ(k_{\mathsf{q}})$$

and therefore

$$1 - \beta = bQ(k_{\mathsf{e}}).$$

Therefore the portfolios with unit price are parameterized by

$$z = \beta\frac{k_{\mathsf{q}}}{Q(k_{\mathsf{q}})} + (1-\beta)\frac{k_{\mathsf{e}}}{Q(k_{\mathsf{e}})} = \beta R_{\mathsf{q}} + (1-\beta)R_{\mathsf{e}} = R_{\mathsf{e}} + \beta(R_{\mathsf{q}} - R_{\mathsf{e}}) \tag{303}$$

with $\beta \in \mathbb{R}$. Here R_{q} and R_{e} are the returns of k_{q} and k_{e}, respectively. Then, minimizing the risk premium among payoffs in $\mathcal{E}$ with unit price turns out to be equivalent to the one-dimensional minimization problem

$$\min_{\beta} D(R_{\mathsf{e}} + \beta(R_{\mathsf{q}} - R_{\mathsf{e}})|o),$$

whose first order necessary condition is

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}\beta} D(R_{\mathsf{e}} + \beta(R_{\mathsf{q}} - R_{\mathsf{e}})|o) \\ &= -(R_{\mathsf{q}} - R_{\mathsf{e}})^{\top} \nabla^2 K(R_{\mathsf{e}} + \beta(R_{\mathsf{q}} - R_{\mathsf{e}}))(R_{\mathsf{e}} + \beta(R_{\mathsf{q}} - R_{\mathsf{e}})). \end{aligned}$$

We conclude that the optimal portfolio with unit price is determined by

$$\beta = -\frac{(R_{\mathsf{q}} - R_{\mathsf{e}})^{\top} \nabla^2 K(R_{\mathsf{e}} + \beta(R_{\mathsf{q}} - R_{\mathsf{e}}))R_{\mathsf{e}}}{(R_{\mathsf{q}} - R_{\mathsf{e}})^{\top} \nabla^2 K(R_{\mathsf{e}} + \beta(R_{\mathsf{q}} - R_{\mathsf{e}}))(R_{\mathsf{q}} - R_{\mathsf{e}})}. \tag{304}$$

Considering the approximation

$$\nabla^2 K(R_{\mathsf{e}} + \beta(R_{\mathsf{q}} - R_{\mathsf{e}})) \simeq \nabla^2 K(o),$$

we fix an approximate value of β that determines the choice of optimal portfolio $z_0 \in \mathcal{E}$ with unit price by

$$\beta_0 := -\frac{(R_{\mathsf{q}} - R_{\mathsf{e}})^{\top} \nabla^2 K(o) R_{\mathsf{e}}}{(R_{\mathsf{q}} - R_{\mathsf{e}})^{\top} \nabla^2 K(o)(R_{\mathsf{q}} - R_{\mathsf{e}})} \tag{305}$$

Note that the expected return of this portfolio is

$$\mathbb{E}_{\mathcal{M}'}[z_0] = \mathbb{E}_{\mathcal{M}'}[R_{\mathsf{e}}] + \beta_0\, \mathbb{E}[R_{\mathsf{q}} - R_{\mathsf{e}}] \tag{306}$$

We have in the case when the risk-free asset $\mathbf{1}$ with risk-less return R_{f} is an element in $\mathcal{M}$ that

$$k_{\mathsf{e}} = \pi_{\mathcal{M}'}(\mathbf{1})$$

and $R_{\mathsf{e}} = \mathbb{E}_{\mathcal{M}'}(\pi_{\mathcal{M}'}(\mathbf{1})) = R_{\mathsf{f}}$. We have in this case

$$\mathbb{E}_{\mathcal{M}'}[z_0] = R_{\mathsf{f}} + \beta_0\,(\mathbb{E}[R_{\mathsf{q}}] - R_{\mathsf{f}}), \tag{307}$$

what is similar to the classical beta pricing equation in the case of the Euclidean divergence where $\nabla^2 K(o) = \mathrm{var}$.

Generalized beta pricing

Recall that we are assuming that $\mathcal{E}$ has dimension two. It is then convenient to restate the results above using two linearly independent assets other than k_{e} and k_{q}. We fix such assets, say k_λ and k_ν, with respective returns

$$r_\lambda = R_{\mathsf{e}} + \lambda(R_{\mathsf{q}} - R_{\mathsf{e}}) \tag{308}$$

and

$$r_\nu = R_{\mathsf{e}} + \nu(R_{\mathsf{q}} - R_{\mathsf{e}}) \tag{309}$$

in such a way that

$$g|_o(r_\lambda, r_\nu) = 0. \tag{310}$$

Hence, ν is given by

$$\nu = -\frac{g|_o(R_{\mathsf{e}}, R_{\mathsf{e}}) + \lambda g|_o(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{e}})}{g|_o(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{e}}) + \lambda g|_o(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{q}} - R_{\mathsf{e}})} \tag{311}$$

Note that ν is well-defined if and only if $\lambda \neq \beta_0$ in (305), that is, if k_λ is not the (approximate) minimum divergence portfolio in $\mathcal{E}$.

Given an asset $z \in \mathcal{M}'$ with unit price we have the decomposition

$$z = z^{\mathcal{E}} + \varepsilon$$

where

$$z^{\mathcal{E}} = ak_\lambda + bk_\nu$$

with $\varepsilon \perp T_{z^{\mathcal{E}}}\mathcal{E}$ and $\mathsf{e}(\varepsilon) = \mathbb{E}_{\mathcal{M}'}[\varepsilon] = 0$. It follows that

$$\begin{aligned} \mathbb{E}[z] &= a\mathbb{E}_{\mathcal{M}'}[k_\lambda] + b\mathbb{E}_{\mathcal{M}'}[k_\nu] = a\mathsf{q}(k_\lambda)\mathbb{E}_{\mathcal{M}'}[r_\lambda] + b\mathsf{q}(k_\mu)\mathbb{E}_{\mathcal{M}'}[r_\nu] \\ &=: \mathbb{E}_{\mathcal{M}'}[r_\nu] + \beta(\mathbb{E}_{\mathcal{M}'}[r_\lambda] - \mathbb{E}_{\mathcal{M}'}[r_\nu]) \end{aligned}$$

with $\beta = a\mathsf{q}(k_\lambda)$. Denoting by r the return of z one obtains

$$r = z = aQ(k_\lambda)r_\lambda + bQ(k_\nu)r_\nu + \varepsilon = r_\nu + \beta(r_\lambda - r_\nu) + \varepsilon$$

from what follows that

$$\begin{aligned} g|_{z^{\mathcal{E}}}(r, r_\lambda) &= g|_{z^{\mathcal{E}}}(r_\nu, r_\lambda) + \beta g|_{z^{\mathcal{E}}}(r_\lambda - r_\nu, r_\lambda) + g|_{z^{\mathcal{E}}}(\varepsilon, r_\lambda) \\ &= g|_o(r_\nu, r_\lambda) + o(|z|^2) + \beta g|_{z^{\mathcal{E}}}(r_\lambda - r_\nu, r_\lambda) \\ &= \beta g|_{z^{\mathcal{E}}}(r_\lambda - r_\nu, r_\lambda) + o(|z|^2). \end{aligned}$$

We conclude that

$$\beta = \frac{g|_{z^{\mathcal{E}}}(r, r_\lambda)}{g|_{z^{\mathcal{E}}}(r_\lambda - r_\nu, r_\lambda)} + o(|z|^2) = \frac{g|_o(r, r_\lambda)}{g|_o(r_\lambda - r_\nu, r_\lambda)} + o(|z|^2) = \frac{g|_o(r, r_\lambda)}{g|_o(r_\lambda, r_\lambda)} + o(|z|^2),$$

where $g|_o = \nabla^2 K(o)$. In sum, we have obtained a generalized beta pricing equation

$$\mathbb{E}_{\mathcal{M}'}[z] = \mathbb{E}_{\mathcal{M}'}[r_\nu] + \beta(\mathbb{E}_{\mathcal{M}'}[r_\lambda] - \mathbb{E}[r_\nu]) \tag{312}$$

for assets in $z \in \mathcal{M}'$, where the generalized beta coefficient is approximated (up to quadratic remainder terms) by

$$\beta = \frac{g|_o(r, r_\lambda)}{g|_o(r_\lambda, r_\lambda)}. \tag{313}$$

If the risk-free asset $\mathbf{1}$ with return R_{f} lies in the space of contingent claims $\mathcal{M}$, we fix $r_\nu = \pi_{\mathcal{M}'}(\mathbf{1})$. With this choice, (312) reduces to

$$\mathbb{E}_{\mathcal{M}'}[z] = R_{\mathsf{f}} + \beta(\mathbb{E}_{\mathcal{M}'}[r_\lambda] - R_{\mathsf{f}}), \tag{314}$$

a generalized beta pricing equation written in terms of an asset k_λ instead of the pricing kernel k_{q} as in (307).

Efficient market portfolio and generalized CAPM

As in the classical CAPM, we can take r_λ 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. In this case, both (312) and (314) define a generalized security market line (16), (18).

Suppose that every agent in the market has consumption preferences given by a time-separable utility function of the form

$$u(c^{(0)}, c^{(1)}) = u_0(c^{(0)}) + u_1(\mathbb{E}_{\mathcal{M}'}[c^{(1)}], g|_{\pi_{\mathcal{M}'}(c^{(1)})}(\pi_{\mathcal{M}'}(c^{(1)}), \pi_{\mathcal{M}'}(c^{(1)}))) \tag{315}$$

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 $t = 1$. Here $\pi_{\mathcal{M}'}$ stands for the projection onto $\mathcal{M}'$.

The optimal agent's consumption plan is a solution of the constrained optimization problem

$$\max_{c^{(0)}, c^{(1)}, \boldsymbol{\alpha}} u(c^{(0)}, c^{(1)})$$

subject to the constraints

$$c^{(0)} \leq w^{(0)} - \boldsymbol{\alpha} \cdot mz,$$
$$c^{(1)} \leq w^{(1)} + \boldsymbol{\alpha} \cdot z$$

where z is a portfolio of risky traded assets in $\mathcal{M}'$, $\boldsymbol{\alpha}$ is the portfolio allocation vector and $w^{(0)}$ and $w^{(1)}$ are, respectively, the agent's endowments at time $t = 0$ and $t = 1$ which we have denoted earlier by $\mathsf{e}^{(0)}$ and $\mathsf{e}^{(1)}$, respectively. If we suppose for the sake of simplicity that we have an interior optimal solution then the first-order condition reads as

$$mz = \frac{\partial_{c^{(1)}} u}{\partial_{c^{(0)}} u} z,$$

where the ratio on the right hand side is the marginal rate of substitution for the utility function u, see (21). Taking expected values on both sides one gets

$$Q(z) = \mathbb{E}_{\mathcal{M}'}\left[\frac{\partial_{c^{(1)}} u}{\partial_{c^{(0)}} u} z\right]$$

for the optimal agent's consumption plan $c^{(1)}$.

We now prove that the tradable projection $\pi_{\mathcal{M}'}(c^{(1)})$ lies in the mean-divergence frontier $\mathcal{E}$. We consider the orthogonal decomposition

$$\pi_{\mathcal{M}'}(c^{(1)}) = \pi_{\mathcal{E}}(\pi_{\mathcal{M}'}(c^{(1)})) + \pi_{\mathcal{M}'}(c^{(1)})^{\perp}$$

where

$$g|_{\pi_{\mathcal{M}'}(c^{(1)})}(\pi_{\mathcal{M}'}(c^{(1)})^{\perp}, \mathcal{E}) = 0.$$

Then we define an alternative consumption plan by

$$\widetilde{c}^{(1)} = \pi_{\mathcal{E}}(\pi_{\mathcal{M}'}(c^{(1)})) + (c^{(1)} - \pi_{\mathcal{M}'}(c^{(1)})),$$

where the second summand is the non-tradable component of $c^{(1)}$. Suppose by contradition that $\pi_{\mathcal{M}'}(c^{(1)})^{\perp} > 0$. We conclude that

$$\widetilde{c}^{(1)} - w^{(1)} < c^{(1)} - w^{(1)}.$$

Moreover since $Q(\pi_{\mathcal{M}'}(c^{(1)})^{\perp}) = \mathbb{E}(\pi_{\mathcal{M}'}(c_1)^{\perp}) = 0$ we have

$$Q(\widetilde{c}^{(1)} - w^{(1)}) = Q(c^{(1)} - w^{(1)})$$

and

$$\mathbb{E}(\widetilde{c}^{(1)} - w^{(1)}) = \mathbb{E}(c^{(1)} - w^{(1)}).$$

We also have $\pi_{\mathcal{E}}(\pi_{\mathcal{M}'}(c^{(1)})) = \pi_{\mathcal{E}}(\pi_{\mathcal{M}'}(\widetilde{c}^{(1)}))$ and

$$g|_{\pi_{\mathcal{M}'}(\widetilde{c}^{(1)})}(\pi_{\mathcal{M}'}(\widetilde{c}^{(1)}), \pi_{\mathcal{M}'}(c^{(1)})) \leq g|_{\pi_{\mathcal{M}'}(c^{(1)})}(\pi_{\mathcal{M}'}(c^{(1)}), \pi_{\mathcal{M}'}(c^{(1)})).$$

Since the agent's preferences are described by an utility function that is strictly increasing with respect to the risk measure (the second variable), we conclude that $\widetilde{c}^{(1)}$ is strictly preferred to $c^{(1)}$. This contradicts the optimality of the consumption plan $c^{(1)}$. From this contradiction, we conclude that $\pi_{\mathcal{M}'}(c^{(1)}) \in \mathcal{E}$ for every agent. Since the market payoff z_{m} is by definition the sum over agents of the tradable components of agents' consumption plans, the market return lies on the mean-divergence frontier as well.

In view of the above, we now deduce a generalized CAPM equation.

Theorem 5.5 *The equilibrium prices for efficient assets z with returns r in a market with agents' preferences described by an utility function of the form (315) are given by*

$$\mathbb{E}[r] - \mathbb{E}[r_\nu] = \widetilde{\beta}(\mathbb{E}[r_{\mathsf{m}}] - \mathbb{E}[r_\nu]), \tag{316}$$

where r_{m} is the return of the market portfolio and

$$\widetilde{\beta} = \frac{g(r, r_{\mathsf{m}})}{g(r_{\mathsf{m}}, r_{\mathsf{m}})}. \tag{317}$$

and r_ν is the return of one of the assets spanning the efficient frontier fixed in (309) and (310).

Proof. Since $c^{(1)} - w^{(1)} = \boldsymbol{\alpha} \cdot z \in \mathcal{M}'$ we have $c^{(1)} - \pi_{\mathcal{M}'}(c^{(1)}) = w^{(1)} - \pi_{\mathcal{M}'}(w^{(1)})$. We have proved that the tradable component of the optimal consumption plan lies in the mean-divergence efficient frontier. Then for each agent, labeled by $\ell = 1, \ldots, M$, we have

$$c_\ell^{(0)} \leq w_\ell^{(0)} - Q(\boldsymbol{\alpha}_\ell) \cdot z - b_\ell,$$

where the portfolio $\boldsymbol{\alpha}_\ell \cdot z$ lies in the mean-divergence frontier and satisfies $g(\boldsymbol{\alpha}_\ell \cdot z, r_\nu) = 0$ for every $\ell = 1, \ldots, M$. The component b_ℓ stands for the total amount invested in the asset with return r_ν which plays the role of a risk-free asset in the classical CAPM. Hence

$$c_\ell^{(1)} \leq w_\ell^{(1)} + \boldsymbol{\alpha}_\ell \cdot z + b_\ell r_\nu.$$

Since $Q(\boldsymbol{\alpha}_\ell \cdot z) = \boldsymbol{\alpha}_\ell \cdot mz$ we have

$$\mathbb{E}_{\mathcal{M}'}[c_\ell^{(1)}] = \mathbb{E}_{\mathcal{M}'}[\pi_{\mathcal{M}'}(c_\ell^{(1)})] = (w_\ell^{(0)} - c_\ell^{(0)})\mathbb{E}[r_\nu] + \boldsymbol{\alpha}_\ell \cdot (\mathbb{E}[z] - m\mathbb{E}[r_\nu])$$

and using that $g(\boldsymbol{\alpha}_\ell \cdot z, r_\nu) = 0$ we have

$$g(\pi_{\mathcal{M}'}(c_\ell^{(1)}), \pi_{\mathcal{M}'}(c_\ell^{(1)})) = g(\boldsymbol{\alpha}_\ell \cdot z, \boldsymbol{\alpha}_\ell \cdot z) + b_\ell^2 g(r_\nu, r_\nu).$$

Differentiating $u_{(\ell)}$ (the utility function for the preferences of the ℓ-th agent) at an equilibrium portfolio with respect to the allocation parameter $\boldsymbol{\alpha}$ one obtains the vector equation

$$\partial_1 u_{(\ell)}(\mathbb{E}[z_i] - m_i\mathbb{E}[r_\nu]) + 2\partial_2 u_{(\ell)} \sum_{j=1}^{N} \boldsymbol{\alpha}_{(\ell)}^j g(z_i, z_j) = 0$$

from what follows that the optimal allocation for each agent is given by

$$\boldsymbol{\alpha}_{(\ell)}^j = -\frac{\partial_1 u_{(\ell)}}{\partial_2 u_{(\ell)}} \sum_{i=1}^{N} g(z_i, z_j)^{-1}(\mathbb{E}[z_i] - m_i\mathbb{E}[r_\nu])$$

Summing up on $\ell = 1, \ldots, M$, one gets

$$\gamma^{-1} \sum_{i=1}^{N} g(z_i, z_j)^{-1}(\mathbb{E}[z_i] - m_i \mathbb{E}[r_\nu]) = 1$$

where

$$\gamma = -\left(\sum_{\ell=1}^{M} \frac{\partial_1 u_{(\ell)}}{\partial_2 u_{(\ell)}} \right)^{-1}$$

Denoting $g_{ij} = g(z_i, z_j)$ one concludes that the market equilibrium price for each asset z_i is given by

$$m_i = \frac{1}{\mathbb{E}[r_\nu]} \left(\mathbb{E}[z_i] - \sum_{j=1}^{N} \gamma g_{ij} \mathbf{1}^j \right)$$

Hence we have

$$\mathbb{E}[r_i] = \frac{1}{m_i} \mathbb{E}[z_i] = \mathbb{E}[r_\nu] + \frac{\gamma}{\nu_i} \sum_{j=1}^{N} g_{ij} \mathbf{1}^j = \mathbb{E}[r_\nu] + \gamma g\Big(\frac{z_i}{\nu_i}, \sum_i z_i\Big) = \mathbb{E}[r_\nu] + \gamma g\Big(\frac{z_i}{\nu_i}, z_{\mathsf{m}}\Big).$$

We conclude that

$$\mathbb{E}[r_i] - \mathbb{E}[r_\mu] = \gamma \mu_{\mathsf{m}} g(r_i, r_{\mathsf{m}})$$

where $\mu_{\mathsf{m}} = z_{\mathsf{m}}/r_{\mathsf{m}}$ is the value of the market payoff at $t = 0$. Denoting

$$\widetilde{\beta}_i = \frac{g(r_i, r_{\mathsf{m}})}{g(r_{\mathsf{m}}, r_{\mathsf{m}})} \tag{318}$$

one obtains

$$\mathbb{E}[r_i] - \mathbb{E}[r_\mu] = \gamma \widetilde{\beta}_i \mu_{\mathsf{m}} g(r_{\mathsf{m}}, r_{\mathsf{m}}).$$

In particular,

$$\mathbb{E}[r_{\mathsf{m}}] - \mathbb{E}[r_\nu] = \gamma \mu_{\mathsf{m}} g(r_{\mathsf{m}}, r_{\mathsf{m}}).$$

Therefore

$$\mathbb{E}[r_i] - \mathbb{E}[r_\nu] = \widetilde{\beta}_i (\mathbb{E}[r_{\mathsf{m}}] - \mathbb{E}[r_\nu]). \tag{319}$$

This finishes the proof. □

We finish pointing out that 316 is indeed an asset pricing formula. In fact denoting by P the price of the asset one has by definition that

$$r = \frac{Q - P}{P}$$

Hence,

$$\mathbb{E}[Q] = P(1 + \mathbb{E}[r]).$$

Therefore

$$P = \frac{\mathbb{E}[Q]}{1 + \mathbb{E}[r_\nu] + \widetilde{\beta}(\mathbb{E}[r_{\mathsf{m}}] - \mathbb{E}[r_\nu])}. \tag{320}$$

Taking, for instance, r_ν as the risk-free return, we conclude that the price is the expected future value of the asset discount at a rate *adjusted to risk*:

$$P = \frac{\mathbb{E}[Q]}{1 + r + \widetilde{\beta}(\mathbb{E}[r_{\mathsf{m}}] - r)}. \tag{321}$$

6 PRINCIPAL CURVES AND PORTFOLIOS

6.1 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 into a n-dimensional statistical manifold

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

where $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_n)$ 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 $\mathbf{f} : \Lambda \subset \mathbb{R} \to \mathcal{M}$, a projection of z on the trace of $\mathbf{f}$ is a point $\mathbf{f}(\lambda_*)$, for some $\lambda_* \in \Lambda$, such that

$$D(z|\mathbf{f}(\lambda_*)) = \inf_{\lambda \in \Lambda} D(z|\mathbf{f}(\lambda)). \tag{322}$$

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

$$\pi_{\mathbf{f}}(z) = \mathbf{f}(\lambda_*). \tag{323}$$

Fixed this notation, we propose the following variational notion of principal curve relatively to D

Definition 6.1 *A curve* $\mathbf{f} : \Lambda \to \mathcal{M}$ *is a* principal curve *in* $(\mathcal{M}, D)$ *if*

$$D(z|\pi_{\mathbf{f}}(z)) = \inf_{s \in (-\epsilon, \epsilon)} D(z|\pi_{\mathbf{f}_s}(z)) \tag{324}$$

for all one-parameter family of curves $\mathbf{f}_s : \Lambda \to \mathcal{M}$*,* $s \in (-\epsilon, \epsilon)$*, 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 Pytaghorean theorem, one of the fundamental results in Information Geometry that can be stated as follows (see also Theorem 5.4)

Theorem 6.1 (Theorem 1.2 and Theorem 1.3, (1)) *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*

$$D(z|o) = D(w|o) + D(z|w). \tag{325}$$

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

$$D^*(z|o) = D^*(w|o) + D^*(z|w), \tag{326}$$

where D^* is the dual divergence.

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

Theorem 6.2 *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 ∇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

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

where the derivative is computed at the critical value $\lambda = \lambda_*(s)$. Denote

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

Note that

$$\mathbf{c}'(s) = \frac{\partial \mathbf{F}}{\partial s}(s, \lambda_*(s)) + \frac{\partial \mathbf{F}}{\partial \lambda}(s, \lambda_*(s)) \frac{\mathrm{d}\lambda_*}{\mathrm{d}s}$$

Hence, we obtain

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

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

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

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

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

We conclude that

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

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

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

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

$$\nabla^2 D(z|\mathbf{f}(\lambda_*))\mathbf{f}'(\lambda_*) = \mu \mathbf{f}'(\lambda_*). \tag{327}$$

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

$$\mathcal{G}_{\mathbf{f}(\lambda_*)} = \nabla^2 D(z|\mathbf{f}(\lambda_*)) \tag{328}$$

associated to the divergence D. This finishes the proof. □

A result concerning principal *submanifolds* similar to Theorem 6.2 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 $\mathcal{M}$ locally parameterized by a smooth map $\mathbf{f} : \Lambda \subset \mathbb{R}^p \to \mathcal{M}$ 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

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

on which is based both the least-squares method and the principal component analysis. In their seminal work (14), Hastie and Stuetzle proved that an 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 $\nabla^2 K$.

Corollary 6.1 *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 6.2 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

$$D(z|w) = K(z) - K(w) - \langle \nabla K(w), z - w \rangle. \tag{329}$$

For details, we refer the reader to (1). □

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

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

or, more generally, the relative ϕ-entropy associated to ϕ-exponential distributions.

6.2 The space of financial assets

In this section, we summarize the results obtained in Chapters 4 and 5.

From now on, $\mathcal{M}$ stands for the linear span of financial assets traded in a security 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 $p(s; \boldsymbol{\vartheta})$. Recall that $\boldsymbol{\vartheta}$ is the distribution parameter of a family of probability distributions in a n-dimensional statistical manifold $\mathcal{S}$.

In what follows, we will recall the statistical manifold of ϕ-exponential probability densities defined in Section 3.2 by

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

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 ϕ-exponential defined as the inverse function of the ϕ-logarithm (24), (25)

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

where $\phi : (0, +\infty) \to (0, +\infty)$ 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 $q > 0$, what corresponds to set $\phi(t) = t^q$, Hence, the q-logarithm is defined by

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

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 $p(s, \boldsymbol{\vartheta})$ and $p(s, \boldsymbol{\vartheta}')$.

Deformed exponentials and portfolio selection

Setting $\phi(t) = t$ one gets the family of exponential distributions, in particular multivariate Gaussian distributions. For this family, R. Nock, B. Magdalou, E. Bryis and F. Nielsen (30), (31) 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 CARA utility functions the certainty equivalent and risk premium of risky assets are respectively given by

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

and

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

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

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

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

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

where Σ is the variance-covariance matrix of the returns on the assets.

In (10), C. Cavalcante, I. Guerreiro and A. F. Rodrigues extended this approach to ϕ-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 cumulanf function by

$$\boldsymbol{\alpha} = \frac{\nabla^2 K(\boldsymbol{\vartheta})^{-1}\mathbf{1}}{\mathbf{1}^\top \nabla^2 K(\boldsymbol{\vartheta})^{-1}\mathbf{1}}. \tag{330}$$

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 (32), the optimal allocation portfolio is given by

$$\boldsymbol{\alpha} = \frac{\Sigma_q^{-1}\mathbf{1}}{\mathbf{1}^\top \Sigma_q^{-1}\mathbf{1}} \tag{331}$$

where

$$\Sigma_q = \gamma_q C_{q,n}^{1-q} |\Sigma|^{\frac{1-q}{2}} \Sigma \tag{332}$$

with

$$\gamma_q = \frac{1}{2}\big((n+4) - (n+2)q\big) \tag{333}$$

and

$$K_{q,n} = \begin{cases} \frac{\Gamma(\frac{1}{q-1}-\frac{n}{2})\sqrt{\pi}}{\Gamma(\frac{1}{q-1})}\left(\frac{1}{q-1}\right)^{\frac{n}{2}}\left((n+4)-(n+2)q\right)\right)^{\frac{n}{2}}, & \text{for} \quad 1 < q < \frac{n+4}{n+2}, \\ \frac{\Gamma(\frac{2-q}{1-q})\sqrt{\pi}}{\Gamma(\frac{2-q}{q-1}+\frac{n}{2})}\left(\frac{1}{1-q}\right)^{\frac{n}{2}}\left((n+4)-(n+2)q\right)\right)^{\frac{n}{2}}, & \text{for} \quad q < 1. \end{cases} \tag{334}$$

Here $|\Sigma|$ is the determinant of Σ. We refer the reader to (32) for further details in q-multivariate Gaussian distributions. It is evident that one reobtains the Markowitz's portfolio for $q = 1$ in (331).

In view of (330), the authors have elaborated in (10) 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 (10) 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.

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 (16) and (19), S. LeRoy, J. Werner and D. 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 (11), 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

$$\mathcal{G}_z = \nabla^2 K(z), \tag{335}$$

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

$$\mathcal{G}_z \sim \nabla^2 K(o) + o(|z|^2), \tag{336}$$

where quadratic terms are determined in terms of the Riemann curvature of the Riemannian manifold $(\mathcal{M}, g)$, see (15).

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

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

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

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

where m is a stochastic discount factor. Here $Q : \mathcal{M} \to \mathbb{R}$ 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 (6), (16).

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 Pytaghorean Theorem for divergences (Theorem 6.1 above) that fixed a reference point $o \in \mathcal{M}$ one has

$$D(z|o) = D(z^{\mathcal{E}}|o) + D(z|z^{\mathcal{E}}), \quad z \in \mathcal{M}'. \tag{337}$$

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

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

expression (337) reduces to the Euclidean decomposition

$$|z|^2 = \mathbb{E}[z]^2 + \text{var}[z], \tag{338}$$

where

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

is the variance, the classical risk measure in Portfolio Theory (20), (6).

Motivated by the analogy between (337) and (338), the authors proposed in (11) 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 $p(s, \boldsymbol{\vartheta})$, this measure encodes higher moments of z instead only the variance. Moreover, one easily verifies that Π 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 (11) 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 Π. Indeed we have

Theorem 6.3 (Theorem 2 in (11)) *Let $\mathcal{E} = \mathrm{span}\{k_{\mathsf{e}}, k_{\mathsf{q}}\}$ 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 ϕ-exponentials that the efficient mean-divergence frontier for portfolio selection is spanned by two portfolios

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

where μ is the desired expected return of the portfolio.

6.3 Generalized beta pricing models and CAPM

Denote by R_{e} and R_{q} the returns of k_{e} and k_{q}, respectively. In (11), the authors have proved that the mimimum divergence portfolio in $\mathcal{M}$ is given by

$$z = R_{\mathsf{e}} + (1-\beta)(R_{\mathsf{q}} - R_{\mathsf{e}})$$

where

$$\beta = -\frac{g(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{e}})}{g(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{q}} - R_{\mathsf{e}})}$$

A similar expression holds replacing the basic assets k_{e} and k_{q} by two efficient assets k_λ and k_ν in $\mathcal{E}$ such that

$$\mathcal{G}(r_\lambda, r_\nu) = 0. \tag{339}$$

These zero-covariance pair of assets is given by

$$r_\lambda = R_{\mathsf{e}} + \lambda(R_{\mathsf{q}} - R_{\mathsf{e}})$$

and

$$r_\nu = R_{\mathsf{e}} + \nu(R_{\mathsf{q}} - R_{\mathsf{e}})$$

where ν is given by

$$\nu = -\frac{\mathcal{G}(R_{\mathsf{e}}, R_{\mathsf{e}}) + \lambda\mathcal{G}(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{e}})}{\mathcal{G}(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{e}}) + \lambda\mathcal{G}(R_{\mathsf{q}} - R_{\mathsf{e}}, R_{\mathsf{q}} - R_{\mathsf{e}})} \tag{340}$$

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

We have obtained in (11) a generalized beta pricing equation involving k_λ and k_ν

$$\mathbb{E}[z] = \mathbb{E}[r_\nu] + \beta(\mathbb{E}[r_\lambda] - \mathbb{E}[r_\nu]) \tag{341}$$

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

$$\beta = \frac{\mathcal{G}(r, r_\lambda)}{\mathcal{G}(r_\lambda, r_\lambda)}. \tag{342}$$

If there exists a risk-free asset $\mathbf{1}$ with return R_{f} in $\mathcal{M}$, we fix $r_\nu = \mathbf{1}$ reducing (341) to

$$\mathbb{E}[z] = R_{\mathsf{f}} + \beta(\mathbb{E}[r_\lambda] - R_{\mathsf{f}}). \tag{343}$$

As in the classical CAPM, we can take r_λ 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

$$u(c_0, c_1) = u_0(c_0) + u_1(\mathbb{E}[c_1], \mathcal{G}|_{c_1}(c_1, c_1)) \tag{344}$$

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 have obtained in (11) a generalized CAPM equation

$$\mathbb{E}[r] - R_{\mathsf{f}} = \widetilde{\beta}(\mathbb{E}[r_{\mathsf{m}}] - R_{\mathsf{f}}), \tag{345}$$

where r_{m} is the return of the market portfolio and

$$\widetilde{\beta} = \frac{\mathcal{G}(r, r_{\mathsf{m}})}{\mathcal{G}(r_{\mathsf{m}}, r_{\mathsf{m}})} \tag{346}$$

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 (341) and (345) define a generalized security market line (16), (18).

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 accordingly a q-Gaussian distributions it holds that

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

for every $z \in \mathcal{M}$, where the q-variance matrix Σ_q is defined in Section 5.2.

6.4 Generalized PCA and applications to Finance

The results we have quoted in Sections 6.2 and 6.3 indicate that the Hessian information matrix

$$\mathcal{G} = \nabla^2 D \tag{347}$$

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 $z_1, \ldots, z_N$ in $\mathbb{M}$ is determined by an allocation vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N) \in \mathbb{R}^N$

$$\boldsymbol{\alpha} D, \tag{348}$$

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

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

whereas its generalized covariance is given by

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

The matrix

$$\mathcal{G}_{ij} := \mathcal{G}(z_i, z_j) \tag{349}$$

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

$$\min_{\boldsymbol{\alpha}} \pi \tag{350}$$

subject to the constraint

$$\boldsymbol{\alpha}\boldsymbol{\alpha}^\top = 1. \tag{351}$$

Setting the Lagrangian

$$L = \pi - \lambda_1(\boldsymbol{\alpha}\boldsymbol{\alpha}^\top - 1),$$

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

$$\sum_{j=1}^{N} \mathcal{G}_{ij} F_j^{(1)} = \lambda_1 F_i^{(1)}, \tag{352}$$

that is, $F^{(1)}$ is an eigenvector of the generalized covariance matrix $\mathcal{G}$ relative to the eigenvalue λ_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 $F^{(1)}, \ldots, F^{(N)}$ correspondent to the eigenvalues

$$\lambda_1 > \ldots > \lambda_N > 0.$$

We then define a matrix R by

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

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

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

Next, we restrict ourselves to the projections of portfolios onto the (totally geodesic) affine subspace spanned by the first $p < N$ principal directions $F^{(1)}, \ldots, F^{(p)}$, taken as the most significant ones due to the fact 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

$$\boldsymbol{\alpha} D = \sum_{i=1}^{p} \beta_i F^{(i)} + \varepsilon, \tag{353}$$

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 μ_* and minimum generalized variance is determined by the weights

$$\beta_i = \left(\sum_{j=1}^{p} \frac{\mathbb{E}[F^{(j)}]}{\sqrt{\lambda_j}} \right)^{-1} \frac{\mathbb{E}[F^{(i)}]}{\lambda_i} \tag{354}$$

To prove this claim, we denote

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

and then we set the Lagrangian

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

with a constraint given by

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

The first order condition is

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

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

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

We conclude that

$$\beta_i = \left(\sum_{j=1}^{p} \frac{R_j}{\sqrt{\lambda_j}} \right)^{-1} \frac{R_i}{\sqrt{\lambda_i}} \tag{355}$$

as claimed. In sum, we have proved

Theorem 6.4 *The p-principal portfolio with minimum generalized variance is given by*

$$\beta_i = \left(\sum_{j=1}^{p} \frac{\mathbb{E}[F^{(j)}]}{\sqrt{\lambda_j}} \right)^{-1} \frac{\mathbb{E}[F^{(i)}]}{\lambda_i} \tag{356}$$

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

7 CONCLUSION AND DEVELOPMENTS

We present in this work a portfolio selection method when considering the returns as q-exponentially distributed. The proposed method is shown to generalize the classical Markowitz framework. The numerical evaluation show the proposal is able to provide better tracking of deep changes in the stock market, such as the ones present in crisis scenarios, and yet provide a higher return than the classical strategy provided by Markowitz's model.

As further directions for our research we envisage evaluating the performance of the proposed portfolio selection considering an estimator for the covariance matrix which may degrade the performance of the used strategies. In this sense, an analysis for the most relevant assets in large portfolio is also a problem that deserves attention since dimensionality may play a key role in the performance of the estimator and thus methods for reducing it, such as PCA, are of great interest. In this direction, we intend to run some numerical experiments in order to apply the generalized principal component analysis we have formulated.

Finally, the derivation in this thesis of a variant of Capital Asset Pricing Model (CAPM) for deformed exponentials suggests some natural follows-ups. For instance, we can apply the proposed definition of generalized beta coefficients to valuation problems in Corporate Finance, developing new tools to decision-making strategies.

BIBLIOGRAPHY

[1] AMARI, S.-I. **Information Geometry and its applications**. Tokyo: Springer, 2016.

[2] AMARI, S.-I.; OHARA, A.; MATSUZOE, H. Geometry of deformed exponential families: Invariant, dually-flat and conformal geometries. **Physica A**, Amsterdam, v. 391, p. 4308–4319, 2012.

[3] AMARI, S.-I.; CICHOCKI, A. Information geometry of divergence functions. **Bulletin of the Polish Academy of Sciences, Technical Sciences**, Krakow, v. 58, p. 183–195, 2010.

[4] AMARI, S.-I.; OHARA, A. Geometry of q-exponential family of probability distributions. **Entropy**, Basel, v. 13, p. 1170–1185, 2011.

[5] AY, N.; JOST, J.; LE, H. V. Schwachhöfer, **Information Geometry**. Heidelberg: Springer, 2017.

[6] BACK, K. **Asset pricing and portfolio choice theory**. Oxford: Oxford University Press, 2010.

[7] BORLAND, L. A theory of non-Gaussian option pricing. **Quant. Finance**, London, v. 6, p. 415–431, 2002.

[8] BORLAND, L.; BOUCHAUD, J.-P. A non-Gaussian option pricing model with skew. **Quant. Finance**, London, v. 4, p. 499–514, 2004.

[9] BORLAND, L. Option Pricing Formulas based on a non-Gaussian Stock Price Model. **Phys. Rev. Lett.**, Rhode Island, v. 89, p. 098–701, 2002.

[10] CAVALCANTE, C. C,; GUERREIRO, I.; RODRIGUES, A. F. P. Deformed exponentials and portfolio selection. **Int. J. Mod. Phys. C**, Australia, v. 29, p. 185–0029, 2018.

[11] CAVALCANTE, C. C,; GUERREIRO, I.; RODRIGUES, A. F. P. Principal Curves for Statistical Divergences and an Application to Finance, **Entropy**, Basel, v. 20, p. 333–347, 2018.

[12] DUFFIE, D. **Dynamic Asset Pricing Theory**. Princeton: Princeton University Press, 2001.

[13] FONTANA, C.; SCHWEIZER, M. Simplified mean-variance portfolio optimisation. **Math. Financ. Econ.**, Heidelberg, v. 6, p. 125–152, 2012.

[14] HASTIE, T.; STUETZLE, W. Principal curves. **Journal of the American Statistical Association**, New York, v. 84, p. 502–516, 1989.

[15] JOST, J. **Riemannian Geometry and Geometric Analysis**. Heidelberg: Springer, 2017.

[16] LeROY, S., WERNER, J. **Principles of financial economics**. Cambridge: Cambridge University Press, 2014.

[17] LINTNER, J. The aggreation of investor's diverse judgments and preferences in purely competitive security markets. **Journal of Financial and Quantitative Analysis**, Cambridge, v. 4, p. 347–400, 1969.

[18] LUENBERGER, D. **Investment science**. Oxford: Oxford University Press, New York, 1998.

[19] LUENBERGER, D. Projection pricing. **J. Optim. Theory Appl.**, Heidelberg, v. 109, p. 1–25, 2001.

[20] MARKOWITZ, H. Portfolio selection. **The Journal of Finance**, New York, v. 7, p. 77–91, 1952.

[21] MAS-COLLEL, A.; WHINSTON, M.; GREEN, J. **Microeconomic Theory**. Oxford: Oxford University Press, 1995.

[22] MORETTO, E; PASQUALI, S.; TRIVELLATO, B. Option pricing under deformed Gaussian distributions, **Physica A**, Amsterdam, v. 446, p. 246–263, 2016.

[23] MOSSIN, J. Equilibrium in a capital asset market. **Econometrica**, New Haven, v. 34, p. 768–783, 1966.

[24] NAUDTS, J. **Generalised Thermostatistics**. London: Springer, 2011.

[25] NAUDTS, J. Estimators, escort probabilities, and ϕ-exponential families in statistical physics. **J. Ineq. Pure Appl. Math.**, v. 5, p. 3–15, 2004.

[26] NAUDTS, J. On the maximum entropy principle in non-extensive thermostatistics, arXiv:cond-mat/0405508v1.

[27] NAUDTS, J. Generalised exponential families and associated entropy functions. **Entropy**, Basel. v. 10, n. 3, p. 131–149, 2008.

[28] NAUDTS, J. Deformed exponentials and logarithms in generalized thermostatistics, **Physica A**, Amsterdam, v. 316, p. 323-334, 2002.

[29] NAUDTS, J.; ANTHONIS, B. The exponential family in abstract information theory. *In*: Geometric science of information. **Lecture notes in computer science**. Heidelberg: v. 8085, p. 265–272, 2013.

[30] NOCK. R.; MAGDALOU, B.; BRiYS, E.; NIELSEN, F. On tracking portfolios with certainty equaivalents on a generalization of Markowitz model: the fool, the wise and the adaptive. *In:* **Proceedings of the 28th International Conference on Machine Learning**. Madison: Omnipress, 2011.

[31] NOCK. R.; MAGDALOU, B.; BRiYS, E.; NIELSEN, F. Mining Matrix Data with Bregman Matrix Divergences for Portfolio Selection. *In:* F. Nielsen and R. Bhatia (Org.), **Matrix Information Geometry**. Springer: Heidelberg, 2012.

[32] PLASTINO, A. Non-extensive statistical mechanics and generalized Fokker-Planck equation. **Phyica A**, Amsterdam v. 222, p. 347–354, 1995.

[33] PLASTINO, A.; VIGNAT, C. Central limit theorem and deformed exponentials, **Journal of Physics A: Mathematical and Theoretical**, London, v. 40, p. 969–978, 2007.

[34] SHARPE, W. Capital asset prices: A theory of market equilibrium under conditions of risk. **Journal of Finance**, New Jersey v. 19, p. 425–442, 1964.

[35] TSALLIS, C. Possible generalization of Boltzmann-Gibbs statistics. **J. Stat. Phys.**, New York, v. 52, p. 479–487, 1988.

[36] TSALLIS, C.; BUKMAN, D. J. Anomalous diffusion in the presence of external forces: Exact time-dependent solutions and their thermostatistical basis. **Phys. Rev. E**, Los Alamos, v. 54, p. 2197–2200, 1996.

[37] TSALLIS, C. **Introduction to Nonextensive Statistical Mechanics**. New York: Springer, 2009.

[38] TSALLIS, C.; ANTENEODO, C.; BORLAND, L.; OSORIO, R. Nonextensive statistical mechanics and economics. **Physica A**, Amsterdam, v. 324, p. 89–100, 2003.

[39] TSALLIS, C. Economics and Finance: q-Statistical stylized features galore. **Entropy**, Basel, v. 19, p. 457–475, 2017.

[40] TRIVELLATO, B. Deformed exponentials and applications to finance. **Entropy**, Basel, v. 15, p. 3471-3489, 2013.