Portfolio optimization based on divergence measures

Chalabi, Yohan and Wuertz, Diethelm

November 2012

Portfolio Optimization Based on Divergence Measures

Yohan Chalabi* Diethelm Wuertz

Institute for Theoretical Physics, ETH Zurich, Switzerland

Computational Science and Engineering, ETH Zurich, Switzerland

November 2012

A new portfolio selection framework is introduced where the investor seeks the allocation that is as close as possible to his "ideal" portfolio. To build such a portfolio selection framework, the ϕ-divergence measure from information theory is used. There are many advantages to using the ϕ-divergence measure. First, the allocation is made such that it is in agreement with the historical data set. Second, the divergence measure is a convex function, which enables the use of fast optimization algorithms. Third, the objective value of the minimum portfolio divergence measure provides an indication distance from the ideal portfolio. A statistical test can therefore be constructed from the value of the objective function. Fourth, with adequate choices of both the target distribution and the divergence measure, the objective function of the ϕ-portfolios reduces to the expected utility function.

---

*Corresponding author. Email address: chalabi@phys.ethz.ch. Postal address: Institut für Theoretische Physik, HIT G 31.5, Wolfgang-Pauli-Str. 27, 8093 Zürich, Switzerland.

1 Introduction

The portfolio selection problem consists of finding a repartitioning of a wealth W into N financial instruments with allocation $\mathbf{w} = (w_1, w_2, \ldots, w_N)^{\mathrm{T}}$ such that it produces a feasible portfolio that is most adequate for the investor. For an investor concerned with risk, the criterion of adequacy would be based on a portfolio pseudo-invariant transform, $Y = T_P(\mathbf{w}, V_t, I_t)$, which maps the portfolio selection, $\mathbf{w}$, the instrument values, V_t, and the information known at t, I_t, to a random variate defined on the probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, which models the risk perception of the investors. This is a rather general formulation of the portfolio selection problem. A more pragmatic approach of the portfolio selection problem is the risk-reward paradigm, which originated in the Nobel-memorial-awarded mean-variance (MV) portfolio selection of Markowitz [1952]. It consists of finding the weights that produce the smallest variance of the financial returns under specific mean return constraints. The mean-variance paradigm is usually represented as the minimization of any *Risk* measure (the variance, in the MV framework), with respect to a *Reward* target (the mean return, in MV framework). It can be represented by

$$\arg \min_{\mathbf{w} \in \mathcal{W}} \quad Risk \tag{1}$$

$$\text{subject to} \quad Reward = c, \tag{2}$$

where $\mathcal{W}$ is the set of weights that produces feasible portfolios for the investor, and c is the target value of the *Reward* quantity. Its dual formulation consists of maximizing the *Reward* measure given a target *Risk* value. In the MV framework, the risk measure is represented by the covariance matrix of the assets returns. However, this representation suffers from the fact that it does equally represent risk on both sides of the financial distribution returns. Over the years, new risk measures have emerged, which consider the downside of the return distribution. Examples are the semi-variance introduced by Markowitz [1959], the lower partial risk measure introduced by Fishburn [1977] and further developed by Sortino and Van Der Meer [1991]. Roy [1952] introduced another method of portfolio allocation, known as the safety first principle. It consists of finding the allocation that has the smallest probability of ruin. Over the years, this concept has involved into the value-at-risk (VaR), advocated by Jorion [1997], and the conditional value-at-risk (CVaR), introduced by Rockafellar and Uryasev [2000]. Another famous portfolio selection framework is that of Black and Litterman [1992], where the investor can include his personal views, on the evolution of the market, in the portfolio criteria. The incorporation of views into the portfolio selection has been extended, by Meucci [2008], into the entropy pooling approach.

Although many portfolio selection models have been introduced since that of Markowitz [1952], the general idea of optimizing a direct quantity of the portfolio has not changed. The investor is assumed to seek the most optimal value of his adequacy function. Nowadays, adequacy functions

can be so complicated that general optimization routines might need to be used because the adequacy function has many local minima. Although an investor might obtain weights from these complex optimization problems, in practice, he will not be able to replicate exactly the same allocation due to operational constraints. For example, he might not be able to buy as many contracts of one instrument as requested by his optimization results. The investor is thus left to approximate the solution he has obtained from the optimization, and build a portfolio that is close to his optimal solution. However, if the investor used a complex objective function, the small change might yield a portfolio that is less favorable than that of other local minima.

With regard to this problem, the contribution of this work is to go beyond the traditional portfolio optimization idea. A portfolio selection framework is introduced where the investor seeks the allocation that is as close as possible to his "ideal" portfolio, while remaining in agreement with the historical data set. To build such a portfolio selection framework, the ϕ-divergence measure, from information theory, is used. There are many advantages to using the ϕ-divergence measure. First, as will be seen in the remainder of this article, the allocation is built such that it is in agreement with the historical data set. Second, the divergence measure is a convex function and enables the use of fast optimization algorithms. Third, the value of the objective value in the minimum portfolio divergence measure provides an indication of how far removed one is from the ideal portfolio. One can therefore construct a statistical test from the value of the objective function. This contrasts with current portfolio optimization schemes. Fourth, with adequate choices of both the target distribution and the divergence measure, the objective function of the ϕ-portfolios reduces to the expected utility function.

The remainder of this paper is organized as follows. Section 2 reviews the definitions of the ϕ-divergence measures, and also their properties. The main result is presented in Section 3 where a portfolio selection framework based on ϕ-divergence measures is constructed. This is followed by a discussion of the dual representation in Section 4. This plays a key role in the numerical resolution of the ϕ-portfolios. Section 5 studies the numerical challenges inherent in the estimation of a portfolio that is as close as possible to a reference portfolio. As will be seen, this challenge reduces to solving a minimax problem. The impact of the choice of divergence measure on the optimized weights is studied in Section 6. Since the portfolio framework is to be built upon the divergence measure, the value of the divergence measure can be used as an indication on how far a given portfolio is from the ideal portfolio. In Section 7, the value of the objective function is used to calculate statistical tests of divergence from the ideal portfolio. Section 8 highlights the connection between the minimum ϕ-portfolio framework and the expected utility theory.

2 Divergence measures

The idea of distance between probability measures goes back to Mahalanobis [1936]. Later, Shannon [1948] introduced an information measure, $I(X, Y)$, defined as the divergence of the joint distribution, P_{XY}, of random processes, X and Y, from the product, $P_X P_Y$, of the marginal distributions;

$$D(P_{XY}, P_X P_Y) = \int \ln \frac{dP_{XY}}{d(P_X P_Y)} dP_{XY}.$$

Kullback and Leibler [1951] subsequently took the above measure from information theory and used it in probability theory. They used it to measure a difference between any two distributions, P_1 and P_2. It is nowadays known as the Kullback-Leibler divergence, or as the relative entropy. This divergence measure has been extended, by Csiszár [1963], Morimoto [1963], and Ali and Silvey [1966], to accommodate a whole class of divergences. Let P and G be two probability distributions over a sample space, Ω, such that P is absolutely continuous with respect to G. The Csiszár ϕ-divergence of P with respect to G is

$$D_\phi(P, G) = \int_\chi \phi(\frac{dP}{dG}) dG,$$

where $\phi \in \Phi$ is the class of all convex functions $\phi(x)$ defined for $x > 0$, and satisfying $\phi(1) = 0$. ϕ is a convex function twice continuously differentiable in a neighborhood of $x = 1$, with nonzero second derivative at the point $x = 1$. The limiting cases of ϕ are defined as: $0\phi(\frac{0}{0}) = 0$, $\phi(0) = \lim_{t \to 0} \phi(t)$, $0\phi(\frac{a}{0}) = lim_{t \to 0} t\phi(\frac{a}{0}) = a \lim_{u \to \infty} \frac{\phi(u)}{u}$, $f(0) = \lim_{t \to 0} f(t)$, and $0f(\frac{a}{0}) = \lim_{t \to 0} tf(\frac{a}{t}) = a \lim_{t \to \infty} \frac{f(u)}{u}$.

When P and G are continuous with respect to a reference measure, μ, on Ω, then the ϕ-divergence can be expressed in terms of the probability densities $f = dF/d\mu$ and $g = dG/d\mu$;

$$D(P, G) = \int_{-\infty}^{\infty} \phi\left(\frac{f}{g}\right) g d\mu. \tag{3}$$

The fundamental properties of the ϕ-divergence measure are that: it is always positive, $D(P, G) \geq 0$, being zero only if $P = G$; it increases when the two distributions disparity increases; and it has the conjugate divergence, $\phi(G, P) = \tilde{\phi}(P, G)$, where $\tilde{\phi}(x) = x\phi(\frac{1}{x})$.

Table 1 lists some divergence measures that have arisen since the introduction of relative entropy as reported in [Pardo, 2006]. These divergence measures were originally used for hypothesis testing, and later used for estimating the parameters of both continuous and discrete models. Statistical applications, and the underlying theory, of these divergences measures can be found in the monographs of Read and Cressie [1988], Vajda [1989], and Pardo [2006], and also in the papers of Morales et al. [1997, 2003].

Liese and Vajda [2006] elegantly demonstrated, by a new generalized Taylor expansion, that all ϕ-divergences can be represented as an average statistical information with a measure of

Table 1: Examples of ϕ-divergence functions as reported in [Pardo, 2006].

ϕ-function	Divergence type
$x \log x - x + 1$	Kullback-Leibler (1959)
$-\log x + x - 1$	Minimum discrimination information
$(x-1)\log x$	J-divergence
$\frac{1}{2}(x-1)^2$	Pearson (1900), Kagan (1963)
$\frac{(x-1)^2}{(x+1)^2}$	Balakrishnan and Sanghvi (1968)
$\frac{-x^s + s(x-1)+1}{1-s}, \quad s \neq 1$	Rathie and Kannappan (1972)
$\frac{1-x}{2} - \left(\frac{x}{2}(1+x^{-r})\right)^{-1/r} + 1$	Harmonic mean (Mathai and Rathie (1975))
$\frac{(1-x)^2}{2(a+(1-a)x)}, \quad 0 \leq a \leq 1$	Rukhin (1994)
$\frac{ax \log x - (ax+1-a)\log(ax+1-a)}{a(1-a)}$	Lin (1991)
$\frac{x^{\lambda+1} - x - \lambda(x-1)}{\lambda(\lambda+1)}$	Cressie and Read (1984)
$|1-x^a|^{1/a}, \quad 0 < a < 1$	Matusita (1964)
$|1-x|^a, \quad a > 1$	χ-divergence of order a (Vajda 1973)
$|1-x|$	Total variation (Saks 1937)

a priori probabilities dependent on the ϕ function. This suggests a Bayesian interpretation of parameter estimators based on ϕ-divergence measures. The estimators seek the parameter values that build an empirical distribution that is the closest to the reference distribution, while remaining in agreement with the data set.

An important family of ϕ-divergence measures, is the one studied by Cressie and Read [1984]. It is frequently used in statistics because it includes a large set of divergence measures. The power ϕ-divergence can be expressed as

$$\phi_\gamma(x) = \frac{x^\gamma - 1 - \gamma(x-1)}{\gamma(\gamma-1)}$$

for $-\infty < \gamma < \infty$. The power-divergence measures are undefined for $\gamma = 1$ and $\gamma = 0$. However, in the continuous limits, ϕ_γ, becomes $\lim_{\gamma \to 1} \phi_\gamma(x) = x \log(x) - x + 1$, and $\lim_{\gamma \to 0} \phi_\gamma(x) = -\log(x) + x - 1$. The power divergence measure encompasses several well-known divergence measures. When $\gamma = \frac{1}{2}$, it corresponds to the Hellinger divergence measure, $2H^2(P,G)$; when $\gamma = 2$, it is the Pearson divergence measure, $\chi^2(P,G)/2$; for $\gamma = 1$ it equates to the Kullback-Leibler divergence, I(P,G); when $\gamma = -1$, it becomes the reverse Pearson (or Neyman) divergence, $\chi^2(G,P)/2$; and with $\gamma = 0$, it is the Leibler (or reverse Kullback) divergence.

The idea of using information theory in finance is not new. For example, the entropy measures have been used by Stutzer [1996] and Kitamura and Stutzer [1997] to obtain a predictive distribution that has minimum relative entropy with respect to a prior distribution. Their

method inspired Foster and Whiteman [1999] and Robertson et al. [2005] to assign probabilities to scenarios in order to forecast financial distributions. Meucci [2008] also used relative entropy in the context of portfolio optimization, where the relative entropy is used to combine investor's views to the empirical views. The use of entropy measures in finance differs from the approach taken here. The only exception found was the paper of Yamada and Rajasekera [1993], which used the relative entropy to find an allocation of a portfolio such that it is as close as possible to an existing portfolio.

3 Phi-portfolios

The optimal portfolio allocation, $\mathbf{w}^* \in \mathcal{W}$, defined on some set of feasible weights for the investors, $\mathcal{W}$, is the allocation that minimizes the ϕ-divergence between $P_{\mathbf{w}}$, the empirical portfolio distribution with allocation $\mathbf{w}$, and the target portfolio distribution, P_r. That is,

$$\mathbf{w}^* = \arg \min_{\mathbf{w} \in \mathcal{W}} \int_a^b \phi \left(\frac{dP_{\mathbf{w}}}{dP_r} \right) dP_r, \tag{4}$$

where $a = P_r^{-1}(\alpha)$ and $b = P_r^{-1}(\beta)$ are, respectively, the lower and upper quantiles of the reference distribution, P_r, over which the divergence measure should be estimated. Note that the range over which the divergence measure should be estimated has been restricted. This enables the focus to be brought upon specific parts of the target distribution; for example, when the interest is in getting as close as possible to only the lower part of the target distribution. This circumstance would be expressed as the lower and upper quantile bounds of Eq. (4), with $\{\alpha = 0, \beta = \frac{1}{2}\}$. Hereinafter, the name ϕ-portfolio will be used to denote the allocation that is most appropriate to the investor in the sense of the ϕ-divergence measure in Eq. (4).

In the general case, a multivariate distribution would be used for the reference portfolio, P_r. However, here, to simplify the approach, the problem is reduced to the univariate distribution of the target distribution, i.e., the expected distribution of the returns of the final portfolio. Note that the inter-connectivity among the components of the portfolio could be modeled as additional constraints. For example, the risk budget constraints could be used to introduce, in the portfolio selection, the notion of risk relation among the components, in terms of the covariance. Moreover, a notion of the temporality of financial returns could be included by the use of maximum drawdown. Risk budgeting would be an adequate approach to model risk, while remaining simple in the optimization.

4 Dual phi-portfolios

The estimation of the ϕ-divergence, as expressed in Eq. (3), requires the estimation of the empirical portfolio distribution, $P_{\mathbf{w}}$. In practice, the empirical distribution can be estimated by

non-parametric estimators such as the kernel density estimators. However, doing so introduces the problem of the bias originating from the density estimator on the divergence measure. For example, in the case of kernel density estimators, the bias introduced by the choice of the kernel bandwidth must be taken into account. With regard to this, several attempts have been made to lower the influence of the empirical density estimation in the ϕ-divergence measure. Lindsay [1994] suggests applying the same kernel estimator to the target distribution in order to reduce the biasing effect of the kernel estimator. However, as pointed out by Keziou [2003], Liese and Vajda [2006] and Broniatowski and Keziou [2006], the ϕ-divergence estimator can be expressed in a dual representation that does not require an empirical density estimator. This dual representation is based on the properties of convex functions. For all a and b in the domain of a continuously differentiable convex function f, it holds that

$$f(b) \geq f(a) + f'(a)(b - a), \tag{5}$$

with $f(b) = f(a)$ if and only if $a = b$. Geometrically, this means that a convex function always lies above its tangents. This is known as the first-order convex condition.

Given a set of weights, $\mathbf{w}$, the ϕ-portfolio measure, $D_\phi(P_\mathbf{w}, P_r)$, can be expressed in terms of an auxiliary distribution, P_θ, that shares the same support as both the reference distribution and the empirical distribution. By substituting $\frac{dP_\theta}{dP_r}$ for a and $\frac{dP_\mathbf{w}}{dP_r}$ for b, in Eq. (5), and by integrating with respect to P_r over the interval (a, b), yields

$$D_\phi(P_\mathbf{w}, P_r) \geq \int_a^b \phi\left(\frac{dP_\theta}{dP_r}\right) dP_r + \int_a^b \phi'\left(\frac{dP_\theta}{dP_r}\right) dP_\mathbf{w} - \int_a^b \phi'\left(\frac{dP_\theta}{dP_r}\right) dP_\theta. \tag{6}$$

It directly follows that the inequality in Eq. (6) reduces to an equality when $P_\theta = P_\mathbf{w}$. The dual representation of the ϕ-portfolio measure for an auxiliary probability P_θ is

$$D_\phi(P_\mathbf{w}, P_r) = \sup_{P_\theta \in \mathbb{P}} d_\phi(P_\mathbf{w}, P_r, P_\theta),$$

where

$$d_\phi(P_\mathbf{w}, P_r, P_\theta) = \int_a^b \phi\left(\frac{dP_\theta}{dP_r}\right) dP_r + \int_a^b \phi'\left(\frac{dP_\theta}{dP_r}\right) dP_\mathbf{w} - \int_a^b \phi'\left(\frac{dP_\theta}{dP_r}\right) dP_\theta. \tag{7}$$

The expected value of the portfolio distribution with allocation $\mathbf{w}$ in Eq. (7), is $\mathbb{E}_{P_\mathbf{w}}[x] = \int_a^b x \, dP_\mathbf{w}$. This can be replaced by its empirical value, since $\mathbb{E}_{P_\mathbf{w}} \to \frac{1}{n} \sum P_\mathbf{w}$ for large n. The dual form for continuous distributions thus becomes

$$d_\phi(P_\mathbf{w}, P_r, P_\theta) = \frac{1}{n} \sum_i \phi'\left(\frac{f_\theta(x_i)}{f_r(x_i)}\right) + \int_a^b \left[\phi\left(\frac{f_\theta(x)}{f_r(x)}\right) f_r(x) - \phi'\left(\frac{f_\theta(x)}{f_r(x)}\right) f_\theta(x)\right] dx, \tag{8}$$

for all indices i such that $a \leq x_i \leq b$ and n is the number of points that fulfill the index condition.

5 Minimum phi-portfolios

For the dual form of the divergence measure in Section 4, the optimal weights, $\mathbf{w}^*$, in terms of the minimum ϕ-portfolio are

$$\mathbf{w}^* = \arg \min_{\mathbf{w} \in \mathcal{W}} \sup_{\theta} d_\phi(P_{\mathbf{w}}, P_r, P_\theta).$$

A simplistic approach to solving this optimization problem would be to perform inner and outer optimizations that solves the minimization and maximization problems respectively. Such an approach would be time consuming and might be numerically unstable. However, by definition, the divergence measure is a continuous convex function. Moreover, the dual form is a concave function with bounded constraints. The minimum and supremum of the objective function are attained when the derivatives with respect to the weights, $\mathbf{w}$, and the auxiliary distribution parameters, θ, are equal to zero. This problem is the equivalent of finding the saddle point of the objective function. Many different methods have been introduced to find the saddle point of constrained functions. Here, a primal-dual interior point solution to the above minimax problem is introduced.

The general form of the minimax problem can be formulated as:

$$
\begin{aligned}
\min_x \max_y \quad & f(x, y) \\
\text{subject to} \quad & c_i(x) = 0, \qquad i \in \mathcal{E} \\
& c_i(x) \le 0, \qquad i \in \mathcal{I} \\
& Ax = a \\
& h_j(y) = 0, \qquad j \in \mathcal{E} \\
& h_j(y) \le 0, \qquad j \in \mathcal{I} \\
& By = b,
\end{aligned}
$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $f(x, y)$ is convex in x and concave in y, the constraints h and c are both convex, $\mathcal{E}$ is the index set of equality constraints, $\mathcal{I}$ is the index set of inequality constraints, and A, a, B, and b describe the linear constraints of x and y respectively. The modified KKT conditions to solve this problem are

$$
R = \begin{bmatrix}
\delta_x f(x, y) + Dc^T \lambda + A^T \nu \\
-\Gamma_\lambda c - (1/t)\mathbf{I}_c \\
Ax - a \\
\delta_y f(x, y) - Dh^T \xi - B^T \mu \\
-\Gamma_\xi h(y) - (1/t)\mathbf{I}_h \\
By - b
\end{bmatrix} = 0,
$$

where λ, ν, ξ, μ are the KKT multipliers, Γ_v is the diagonal operator that maps a vector v to the diagonal matrix $\mathrm{diag}[v]$, $\mathbf{I}_c$ and $\mathbf{I}_h$ are diagonal matrices with zeros when $i \in \mathcal{E}$ and ones when $i \in \mathcal{I}$, and

$$c = \begin{bmatrix} c_1(x) \\ \vdots \\ c_n(x) \end{bmatrix}, \ Dc = \begin{bmatrix} \nabla c_1(x) \\ \vdots \\ \nabla c_n(x) \end{bmatrix}, \ h = \begin{bmatrix} h_1(y) \\ \vdots \\ h_m(y) \end{bmatrix}, \text{ and } Dh = \begin{bmatrix} \nabla h_1(y) \\ \vdots \\ \nabla h_m(y) \end{bmatrix}.$$

The solution of the modified KKT conditions can be found by Newton steps. Each iteration are solutions of the linear system of the form

$$\begin{bmatrix} H_{xx}^2 & Dc & A^T & H_{xy}^2 & O & O \\ -\Gamma_\lambda Dc & -\Gamma_c & O & O & O & O \\ A & O & O & O & O & O \\ H_{yx}^2 & O & O & H_{yy}^2 & -Dh & -B^T \\ O & O & O & -\Gamma_\nu Dh & -\Gamma_h & O \\ O & O & O & B & O & O \end{bmatrix} \begin{bmatrix} \delta x \\ \delta \lambda \\ \delta \nu \\ \delta y \\ \delta \xi \\ \delta \mu \end{bmatrix} = -R,$$

where $H_{xx}^2 = \delta_{xx}^2 f(x,y) + \sum_{i=1}^n \lambda_i \nabla^2 c_i(x)$, $H_{yy}^2 = \delta_{yy}^2 f(x,y) - \sum_{i=1}^m \xi_i \nabla^2 h_i(y)$, $H_{xy}^2 = \delta_{xy}^2 f(x,y)$ and $H_{yx}^2 = \delta_{yx}^2 f(x,y)$. The sequence of Newton steps are iterated until a convergence criterion has been satisfied.

6 Choice of divergence measures

This section seeks to understand the effect of the divergence measure on the weights allocation. Indeed, the choice of the divergence measure can produce different optimal weights. Figure 1 displays the divergence surface between the normal distribution with mean 0 and standard deviation 1, and normal distributions of varying mean and standard deviation. The power divergence family was used, with different sets of exponent values as shown in the figure. It is clear that the divergence has different shapes, depending upon the power parameter. This is a strong indication that portfolio estimators based on different divergence measures will yield different weight parameter estimates.

The supremum of the dual form of the divergence measure is used to estimate the divergence measure itself. Recall that the supremum of the dual form of the power divergence is obtained when its derivatives are equal to zero, where

$$d_\gamma(P_{\mathbf{w}}, P_r, P_\theta) = \frac{1}{n(\gamma - 1)} \sum_{i=1}^n \left[\left(\frac{f_\theta}{f_r} \right)^{\gamma-1} - 1 \right] - \frac{1}{\gamma} \int_a^b f_r \left[\left(\frac{f_\theta}{f_r} \right)^\gamma - 1 \right] du,$$

$$\delta_\theta d_\gamma(P_{\mathbf{w}}, P_r, P_\theta) = \frac{1}{n} \sum_{i=1}^n \frac{\nabla f_\theta(x_i)}{f_r(x_i)} \left(\frac{f_\theta(x_i)}{f_r(x_i)} \right)^{\gamma-1} - \int_a^b \nabla f_\theta \left(\frac{f_\theta}{f_r} \right)^{\gamma-1} du.$$
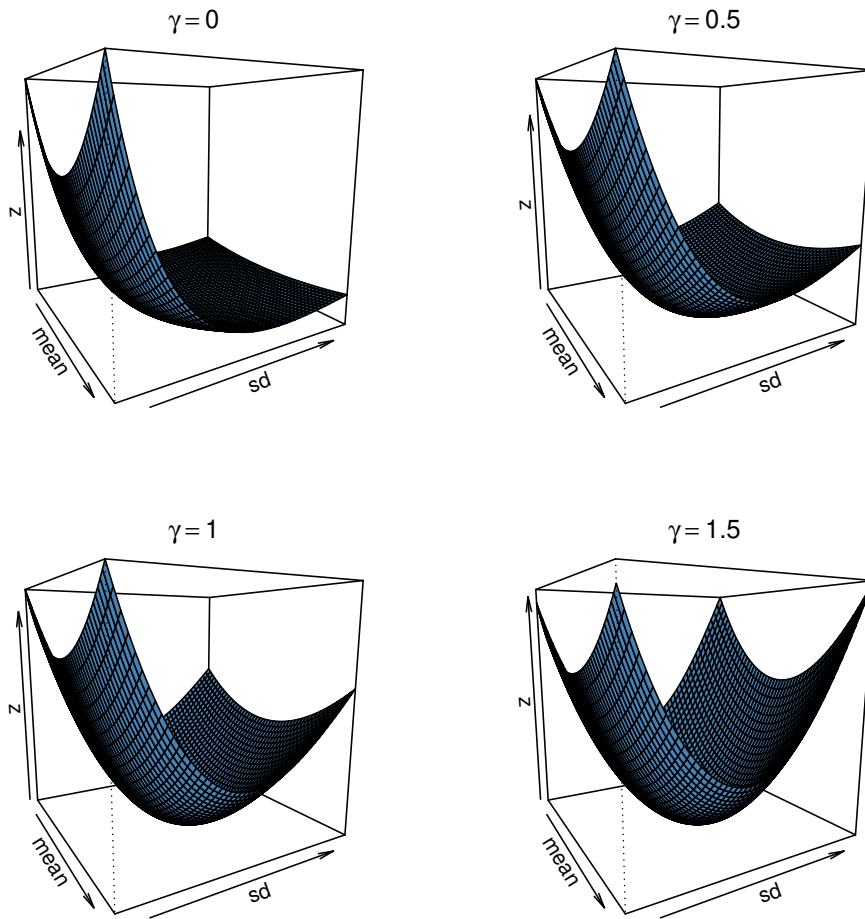
Figure 1: Examples of shapes of the power divergence measure with different γ values.

As noted by Toma and Broniatowski [2011], the summation corresponds to the weighted likelihood score function. The weight function is

$$\left(\frac{f_\theta(x_i)}{f_r(x_i)}\right)^{\gamma-1}.$$

The effect of the choice of the power divergence parameter, γ, upon the optimized weights, can be understood by studying this weight function. Note that for $\gamma = 1$, there are no weights, and so the estimator reduces to the maximum likelihood estimator.

Two cases must be distinguished. The first case is when the reference distribution can be described by the data set. The second case is when the data set does not converge in distribution to the reference model.

When the portfolio data set converges in distribution to the reference portfolio distribution, the auxiliary distribution in the dual form converges to the reference distribution as well; i.e., $f_\theta \to f_r$. The weighting function, $[f_\theta(x)/f_r(x)]^{\gamma-1}$, converges to unity, and thus all points are equally weighted. Figure 2 illustrates this effect; the weight function is close to one for most of the distribution range, irrespective of the gamma factor.

Regarding portfolio allocation, when the portfolio data set can converge in distribution to the reference distribution, the choice of the divergence measure does therefore not change the weights allocation. To illustrate this, two time series were simulated: both from the normal distribution with a mean of 0; one having a standard deviation of 1 and the other having a standard deviation of 10. The reference distribution was taken to be the normal distribution with mean 0 and standard deviation 5. The optimized portfolio should therefore be the equally weighted portfolio. Figure 3 displays the weight plot for a range of power exponents between 0 and 1. The weights remain constant over the range of power exponents.

However, when the data is significantly different from the reference distribution, the likelihood score function is weighted by the likelihood ratio. At the supremum of the dual divergence measure, the auxiliary distribution, f_θ, converges in distribution to the portfolio distribution, f_r, as seen in Section 4. The value of the weight function depends on the power exponent γ, if it is less or greater than 1, and on the likelihood ratio, $\frac{f_\theta(x_i)}{f_r(x_i)}$, if the likely value of the auxiliary distribution is less or greater than the one of the reference distribution at x_i. Consider the points with index i where $f_\theta(x_i) > f_r(x_i)$; that is when the reference distribution has thinner tails than the empirical portfolio. When $\gamma > 1$ ($\gamma < 1$), the score function is then upweighted (downweighted) at these points. The effect is inverted for values at index i where $f_\theta(x_i) < f_r(x_i)$.

To illustrate this, three cases are considered. The first case is when the reference distribution scale is larger than the data set scale. The second case is when the scale parameter of the reference distribution is smaller than that of the data set. The third case is when the reference distribution has a larger mean than does the data set.

Figure 4 displays the results for $\gamma < 1$. When the reference distribution has larger scale than
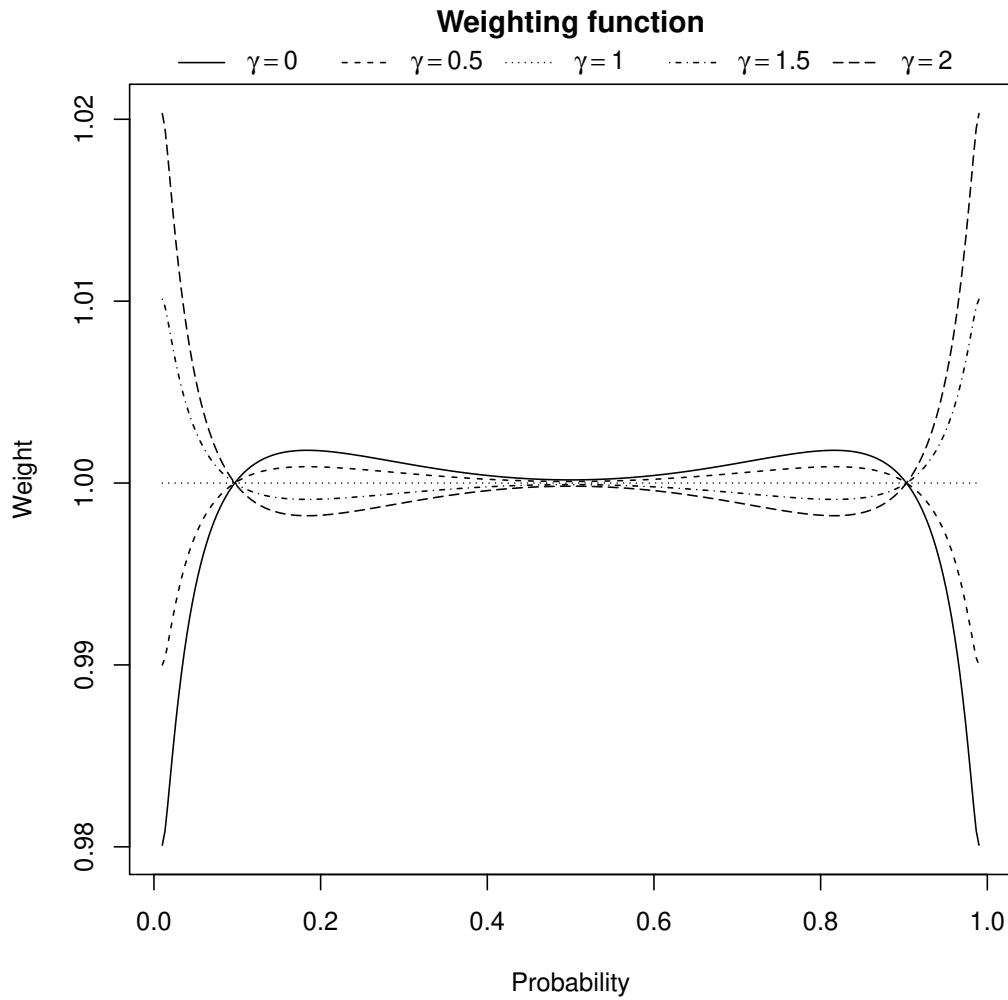
Figure 2: Weighing function of the likelihood score function when the empirical portfolio converges in distribution to the target distribution for different γ values.
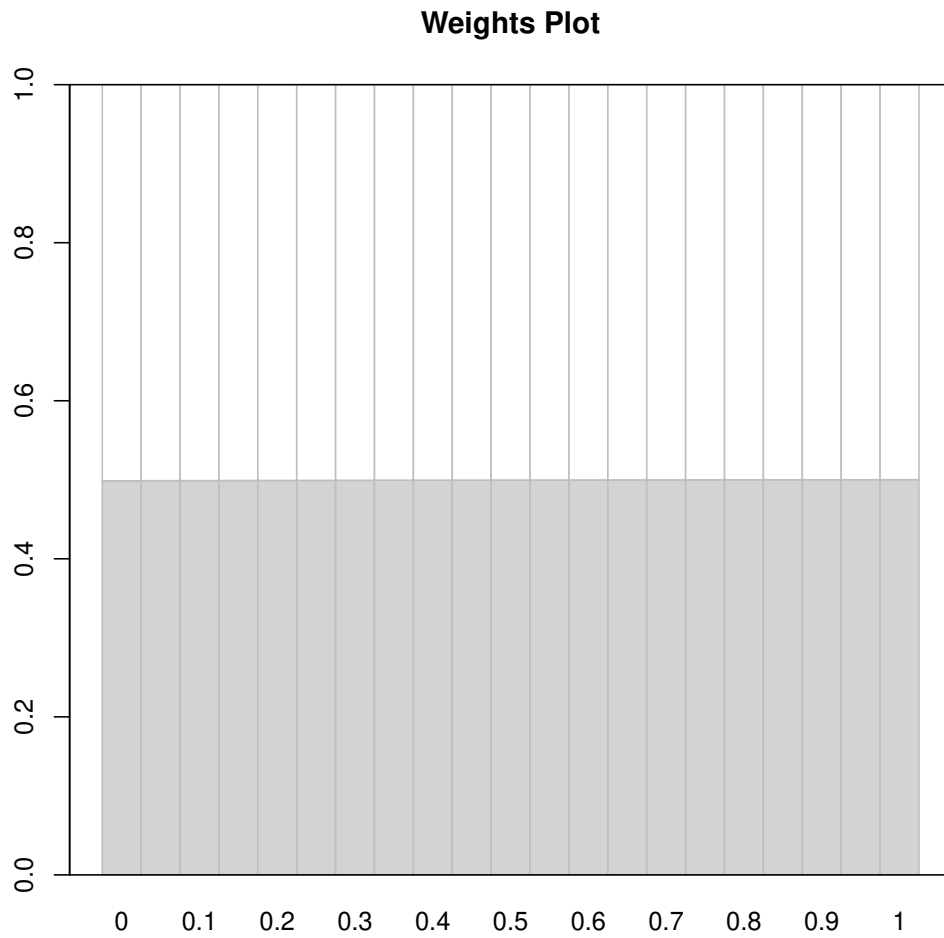
Weights Plot

Figure 3: Weights plot for a range of power exponents between 0 and 1 for a two time series portfolio where the equally weighted portfolio distribution is set as the target distribution.

the data set, the divergence portfolio privileges the allocation that produces the best fit in the tails. This effect is increased when reducing γ toward 0. In the case when the target distribution has a smaller scale than the reference distribution, reducing γ towards zero has the effect of increasing the weighting around the center of the distribution.

When $\gamma > 1$, Fig. 5, the effect is inverted as is expected from the form of the weight function, which is to the power $\gamma - 1$. When the reference distribution has a larger scale than the data set does, the center of the distribution gains higher weighting as γ increases. When the reference distribution has a smaller scale than the data set does, the tail of the distribution has higher weight as γ increases.

Now to examine the effect of the power parameter γ in the context of portfolio optimization with a real data set; the Swiss Pension Fund Benchmarks[1] of Pictet is used. The family of Pictet LPP indices was created in 1985 with the introduction of new regulations in Switzerland governing the investment of pension fund assets. Since then it has established itself as the authoritative pension fund index for Switzerland. In 2000, a family of three indices, called LPP25, LPP40, and LPP60, where the increasing numbers denote increasingly risky profiles, were introduced to provide a suitable benchmark for Swiss pension funds. It is composed of 6 instruments. In this example, the objective is to find the asset allocation of these 6 assets such that the resulting portfolio is as close as possible to the normal distribution with the mean and standard deviation of the benchmark LPP60. Figure 6 displays the optimal weights obtained for different values of the power divergence.

Given that the allocation might change, depending on the choice of the divergence measure, it is of interest to see if the shape of the divergence measure between the empirical portfolio and the reference portfolios might change as well. In this regard, Fig. 7 displays the ternary plot of the divergence measure for the portfolio, with components SBI, SPI and SII, with respect to the reference portfolio. SBI is the Swiss bond index, SPI is the Swiss performance index and SII is the Swiss real estate index. The logarithm returns of these indices with daily records from 2000–01–03 to 2007–05–04[2] were considered. The reference distribution used is the Student t-distribution with its parameter-values fitted to the equally weighted portfolio with the SBI, SPI and SII. The four plots have been generated with different power divergence parameters as indicated in the figure. The shape of the divergence measure does not significantly change, although the divergence values themselves are quite different. This is important because it shows that the overall shape of the divergence measure is insensitive to the choice of divergence estimator. This provides an indication of the stability of the approach.

---

[1]The data set (LPP2005) is available in the R package **fEcofin**

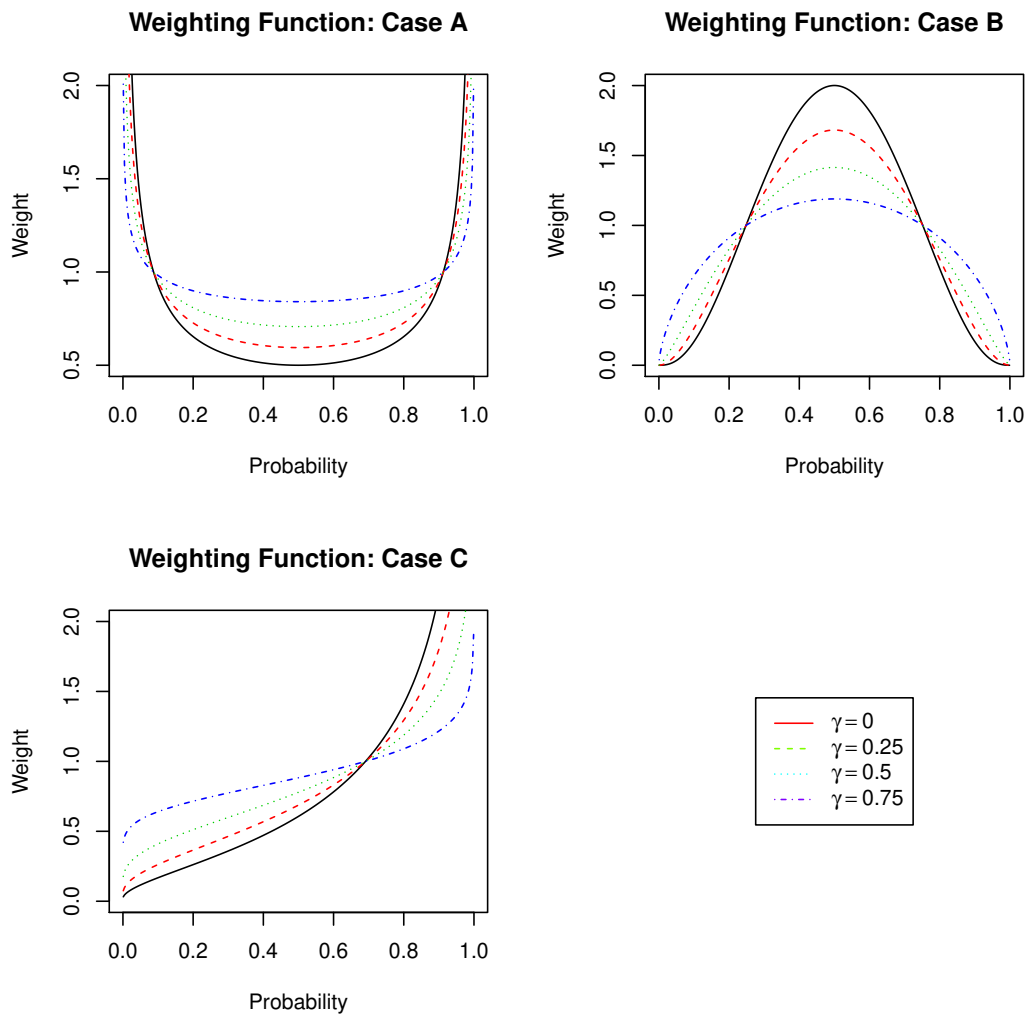[2]The data set (SWXLP) is available in the R package **fEcofin**

Figure 4: Weighting function when: (A) the reference distribution scale is larger than the data set scale, (B) the scale parameter of the reference distribution is smaller than that of the data set, (C) the reference distribution has a larger mean than does the data set.
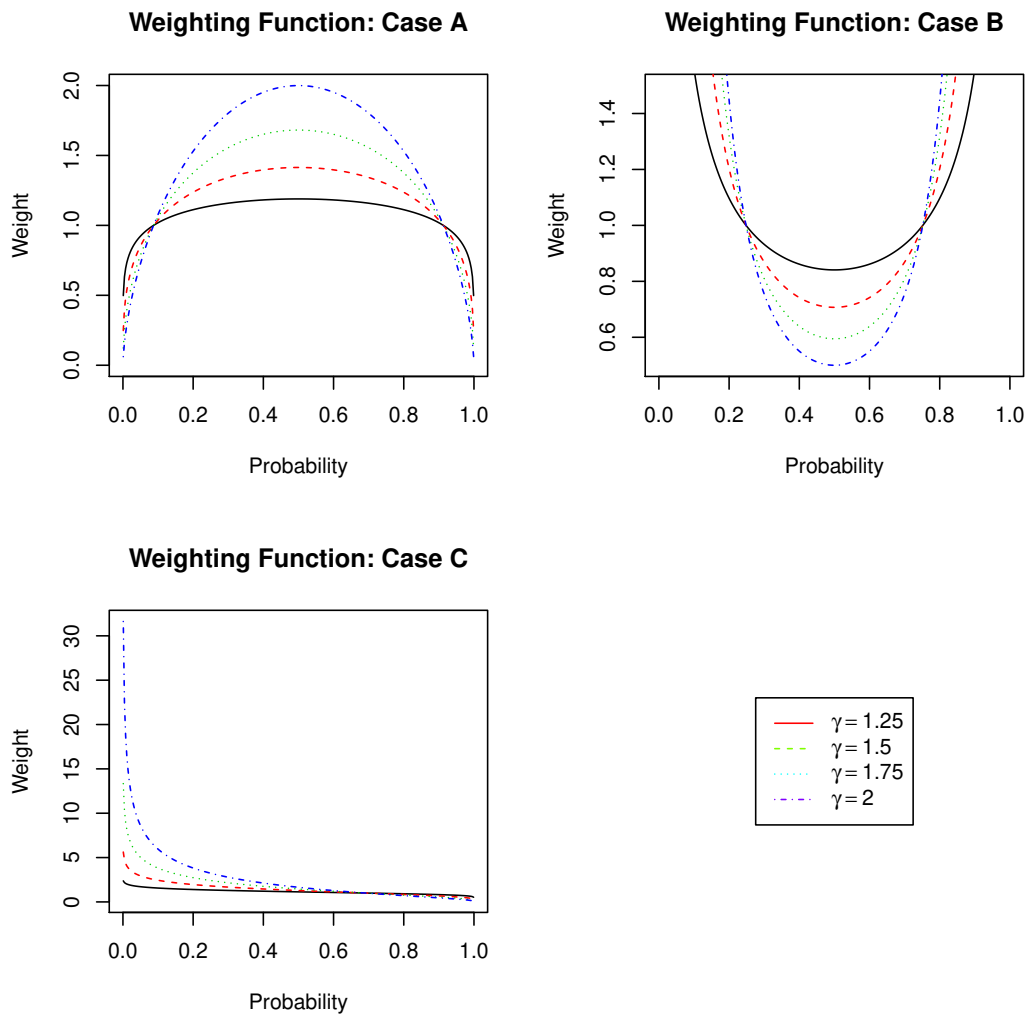
Figure 5: Weighting function when: (A) the reference distribution scale is larger than the data set scale, (B) the scale parameter of the reference distribution is smaller than that of the data set, (C) the reference distribution has a larger mean than does the data set.
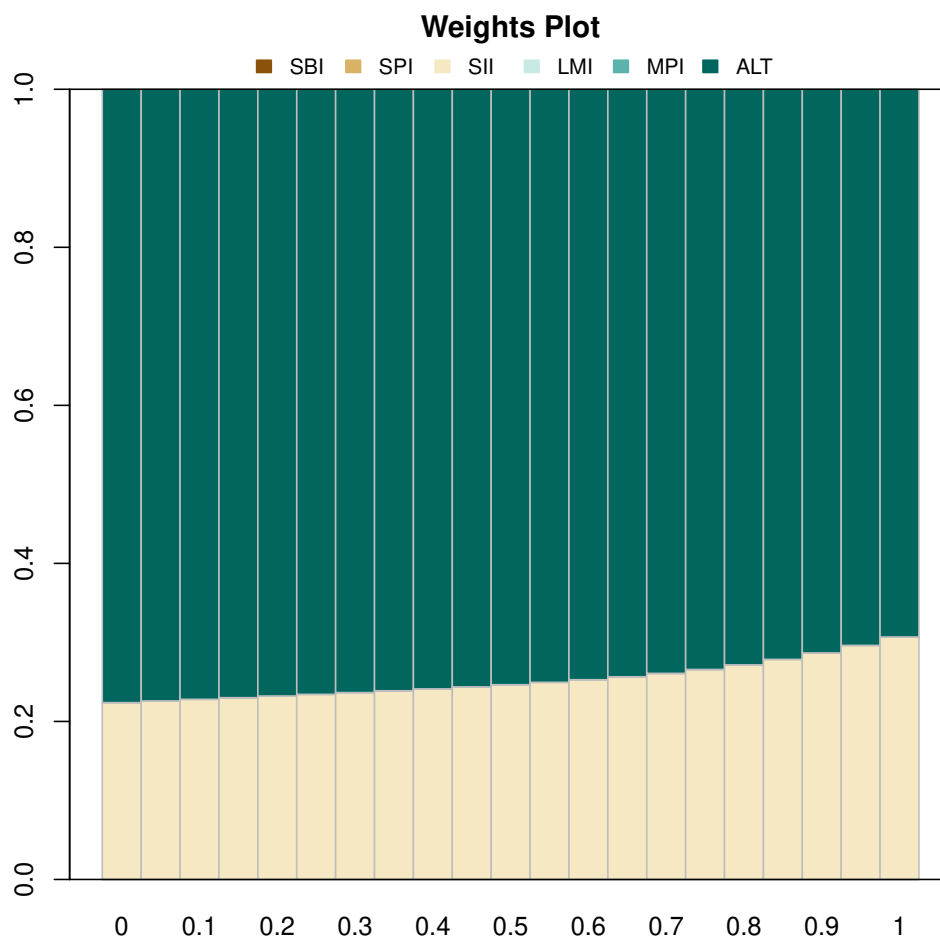
Figure 6: Weights plot for the LPP2005 data set. the reference distribution is the normal distribution fitted to the LPP60 benchmark.
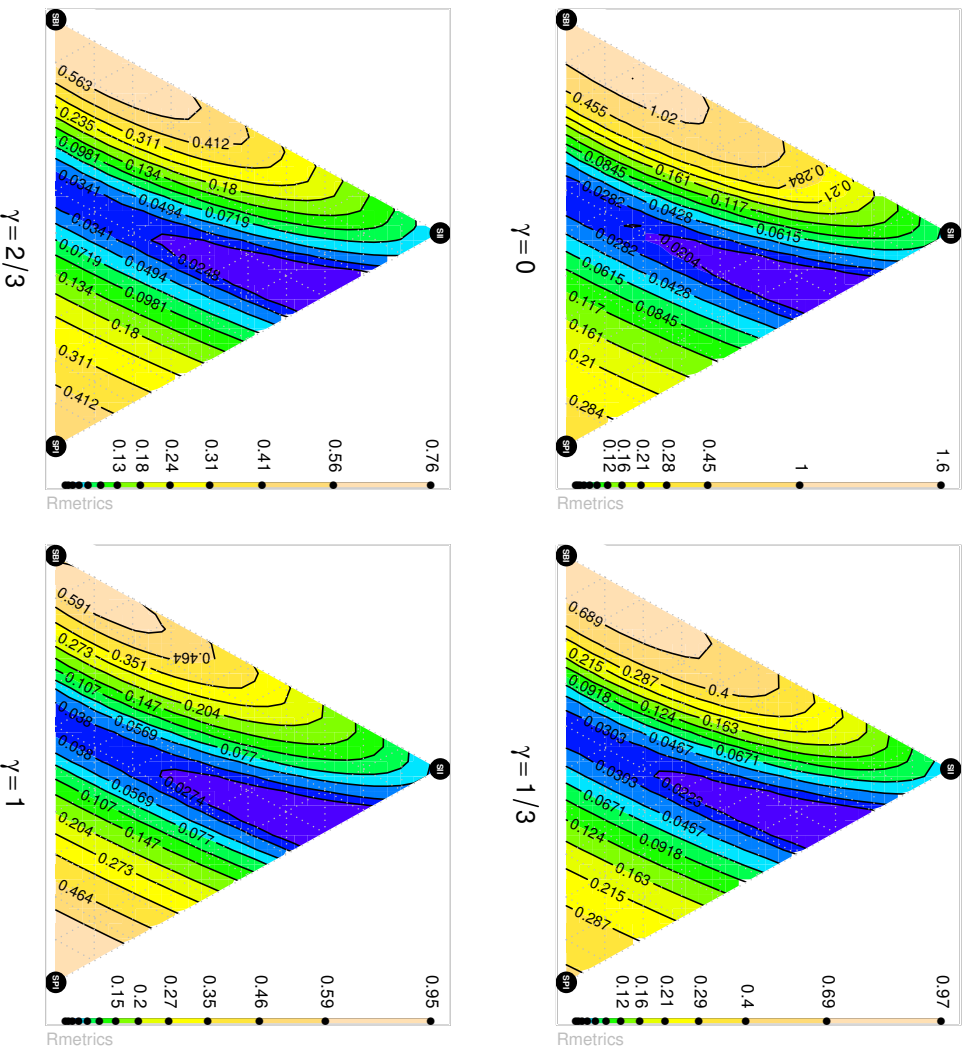
Phi–Portfolio Ternary Maps

Figure 7: Ternary plot to illustrate the shape of the divergence measure for the phi-portfolio with three assets portfolio with different γ values.

18

7 Statistical test

Recall that the ϕ-portfolio seeks the allocation that produces a portfolio that is as close as possible to the reference distribution. The closeness is expressed by the divergence measure. The value of the objective function in the ϕ-portfolio therefore carries valuable information. It says how close one can get to the reference model. This is not the case for traditional portfolio allocation.

Cressie and Read [1984] established the asymptotic null distribution of the goodness-of-fit statistics within the power divergence family. They show that the t-test is linked to the χ^2 distribution. Broniatowski and Keziou [2009] proved that the asymptotic null distribution of the power divergence family remains a χ^2 distribution when the divergence measure is expressed in its dual form. An important result in [Broniatowski and Keziou, 2009, Theorem 3.2.b] is that at the supremum of the dual divergence, which occurs when the auxiliary distribution tends to the distribution of the sample data, the statistic

$$\frac{2n}{\phi''(1)}\hat{d}_\phi$$

convergences in distribution to the χ^2 distribution where $\hat{d}_\phi$ is the supremum of the dual divergence. In the case of the power divergence family, we have $\phi''(1) = 1$ and the statistics reduces to $2n\,\hat{d}_\gamma$.

The asymptotic distribution of ϕ-divergence statistical tests remains an open research topic. In this context, Jager and Wellner [2007] introduced confidence bands for the power divergence family. The interested reader is referred to their paper and their R package **phitest**. Lately, Toma and Leoni-Aubin [2010] and Toma and Broniatowski [2011] have introduced a more accurate approximation of the p-values of the dual divergence, based on the saddle point approximation of the distribution of the supremum dual divergence as introduced by Robinson et al. [2003] for M-estimators.

Figure 8 displays the histogram of the statistical tests of the divergence for 1000 time series with the same distribution as the reference distribution that being the normal distribution with mean 0 and standard deviation 1. It is clear from the plot that the statistical test converges in distribution to the χ^2 distribution.

As a second illustration, Fig. 9 shows how the p-values increase while the distribution of the data converges in distribution to the reference model. The data is distributed according to the Student t-distribution with mean 0, standard deviation 1 and different degrees of freedom. The normal distribution with mean 0 and standard deviation 1 is used as the reference distribution. It is clear that when Student's t-distribution's degrees of freedom increases, the p-values increase as well since the distribution of data set will converge to a normal distribution.
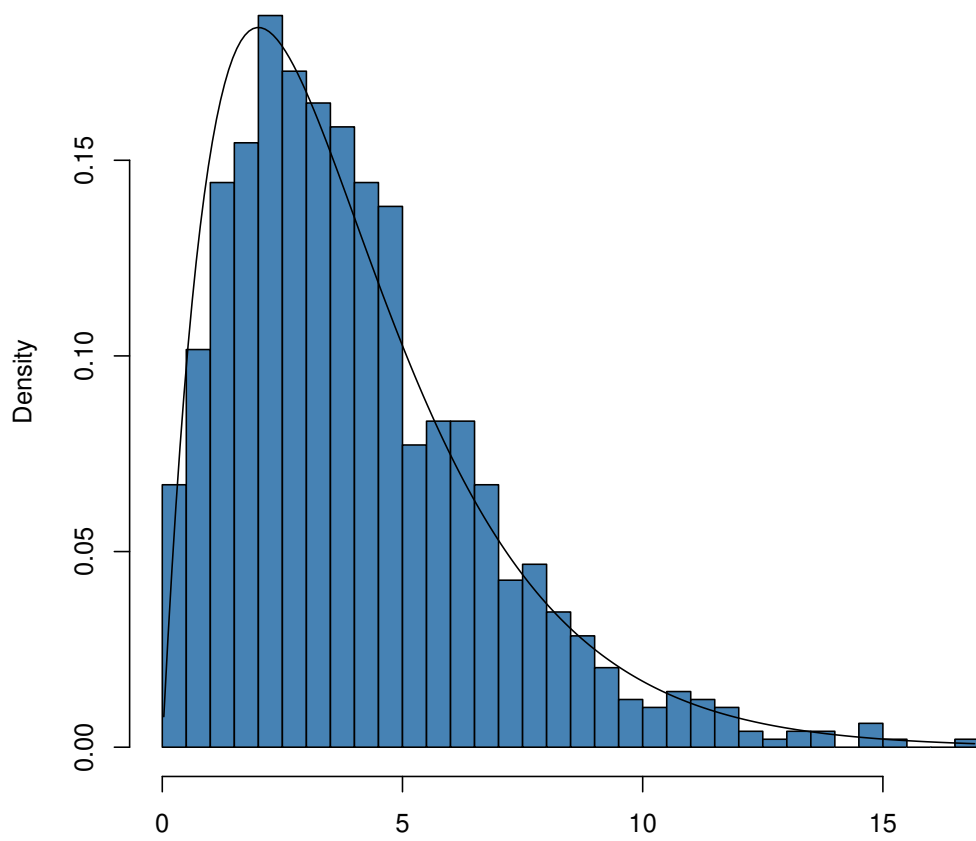
Figure 8: Histogram of the statistical test of divergence measures together with the χ^2 density distribution.
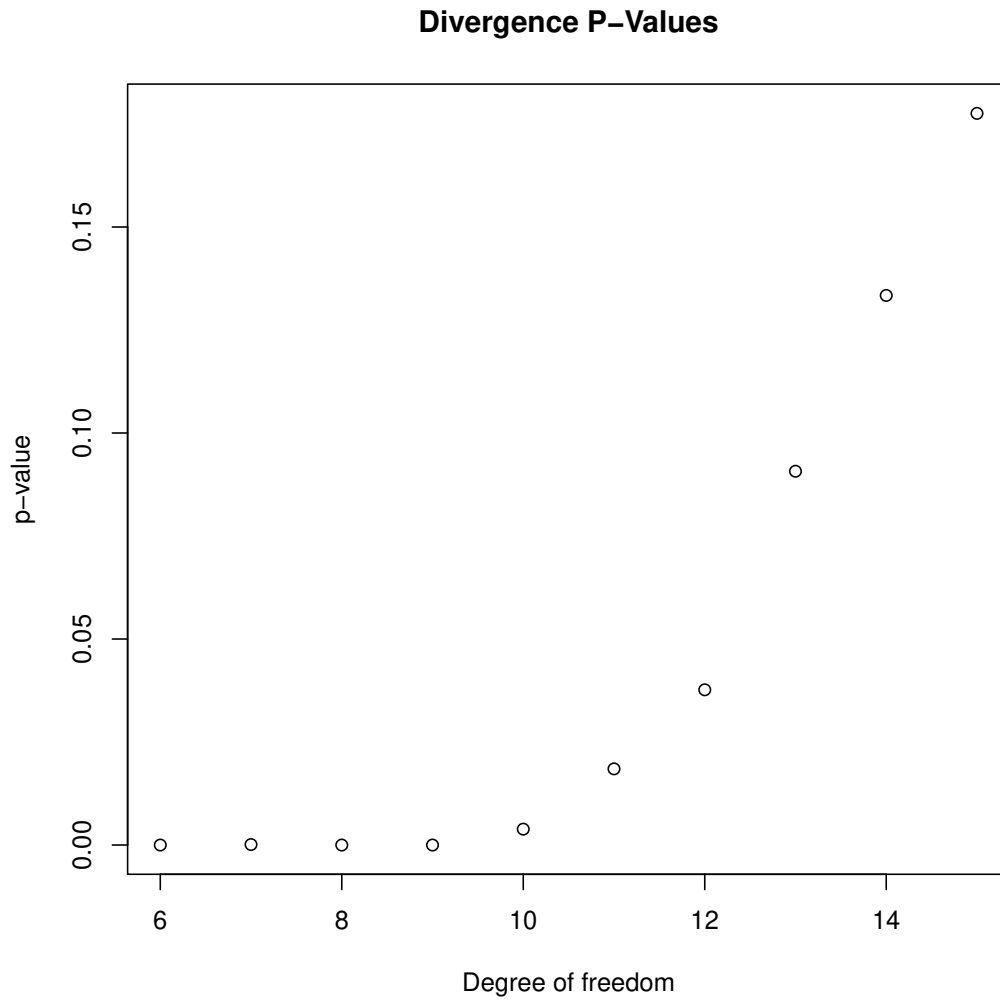
Divergence P−Values

Figure 9: p-values of the statistical test divergence measure when the generated data set converges in distribution to the reference distribution. Here, the normal distribution is used as the reference model, whereas the samples have Student's t-distribution with increasing degree of freedom.

8 Expected utility theory

In the expected utility framework, the decision maker constructs a utility function that represents how much the investment would be of interest for him relative to his wealth. The utility function is usually modeled by a concave function that reflects that a smaller gain would be of greater interest to an investor with a smaller wealth. A gain of \$1 is of greater interest to an investor who has \$10, than to an investor who has \$1000. The investor is assumed to be risk-averse and would therefore seek a portfolio allocation that is optimal in a global sense. the investor preference $\succ$ is modeled by the stochastic dominance. For two random outcomes, X and Y, that belong to a subset of the measurable probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, the preference $\succ_u$ is defined as

$$X \succ_u Y \iff \mathbb{E}[u(X)] > \mathbb{E}[u(Y)].$$

The functional form of the expected utility $\nu(.) = \mathbb{E}[u(.)]$ is known as the expected utility function. When the utility function expresses the risk that the investors are willing to take, then the expected utility functions give a way to measure investor's preferences.

This section draws the connection between the ϕ-portfolios and the expected utility theory. Recall that the ϕ-portfolio is defined as the allocation, $\mathbf{w}^*$, for which divergence between the portfolio distribution and a reference distribution is smallest:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \int \phi \left(\frac{f_{\mathbf{w}}(x)}{f_r(x)} \right) f_r(x) \, dx, \tag{9}$$

where ϕ is the divergence measure, $f_{\mathbf{w}}$ is the empirical distribution of the portfolio given by the set of weights, $\mathbf{w}$, and f_r is the reference distribution. When the uniform distribution is taken as the reference distribution, the optimal allocation minimizes the expected value of the composition of the divergence ϕ function and the distribution of the empirical portfolio distribution, $\phi \circ f_{\mathbf{w}}$. In this case, the optimal weights, $\mathbf{w}^*$, are

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} \int u(x) dx = \mathbb{E}[u(x)] = \nu(x), \tag{10}$$

where $u = -\phi \circ f_{\mathbf{w}}$. By construction, the function ν is both increasing and concave, and corresponds to the utility function defined above. The ϕ-portfolio is thus equivalent to the maximum expected utility portfolio when the reference distribution is the uniform distribution.

It is interesting to note that the Legendre-Fréchet transformation, used here to express to the divergence measure in its dual form, has also been used within the scope of the utility theory [Föllmer and Schied, 2002], where it has been shown that for special choice of the loss function, it reduces to the relative entropy.

The traditional risk-reward portfolio optimization approach can be linked to the utility theory when the risk measures are described as convex functions as shown by De Giorgi [2005]. The ϕ-portfolios thus also inherit this connection to the risk-reward paradigm.

9 Conclusion

This paper has introduced a portfolio allocation scheme based on the ϕ-divergence measure. The approach uses the dual representation of the divergence measure to avoid using a non-parametric estimation of the empirical distribution. The advantages of the approach are multifold.

As was seen, the scheme seeks the allocation that produces a portfolio that is the closest to a reference model while remaining as compatible as possible to the empirical data, which can be interpreted as a type of Bayesian estimator. This is in contrast to the traditional risk-reward paradigm where an attempt is made to optimize one quantity given constraints. It is believed to be a more viable approach because it reduces the risk estimation error induced by the risk-reward model. In the approach devised in this article, the weights selected are those that are in greatest accord with the empirical data. This could lead to a situation where the selected weights might have a slightly increased deviation, but a much better mean vector. This is also important in the context of robust estimators. As in the case of the power divergence measure, the power exponent can make the estimator more robust to outliers. Given the choice of the divergence measure, the ϕ-portfolios can be viewed as robust estimators.

Moreover, the value of the optimized ϕ-portfolio function provides information on how close the investor is to his reference model. It can be used as a statistical test; providing a warning that the reference model might have been too ambitious for the empirical data.

Another important aspect is that the divergence measure is convex and fast optimization routines can be used in practice. This paper has presented the primal-dual interior optimization for estimating the supremum of the dual representation of the divergence measure and to solve the minimum ϕ-portfolio problem.

The relationship between this new framework and the utility theory has been highlighted. It has been shown that the utility function becomes a particular case of the ϕ-portfolio when the uniform distribution is chosen as the reference distribution.

Divergence measure is an active research topic. Many problems remain open. There is no clear comparison of the properties for the different functions that have been introduced over the last few decades. Also, a complete asymptotic theory of ϕ-divergences has yet to be found. All these open problems constitute a favorable outlook for the practical use of ϕ-portfolios. An exhaustive comparison of ϕ-portfolios to common portfolio optimization schemes would be a valuable contribution to the field.

Another possible direction of research in portfolio allocation based on divergence measure would be the selection of assets such that the allocated portfolio is as far as possible from a worst-case reference distribution. Preliminary studies have been performed, in the context of the safety first principle, by Stutzer [2000]; Haley and Whiteman [2008]. A study based on general ϕ-divergence measures would be interesting.

Acknowledgments

References

S. Ali and S. Silvey. A general class of coefficients of divergence of one distribution from another. *J. R. Stat. Soc., Ser. B*, 28:131–142, 1966.

F. Black and R. Litterman. Global portfolio optimization. *Financial Analysts Journal*, pages 28–43, 1992.

M. Broniatowski and A. Keziou. Minimization of ϕ-divergences on sets of signed measures. *Stud. Sci. Math. Hung.*, 43(4):403–442, 2006.

M. Broniatowski and A. Keziou. Parametric estimation and tests through divergences and the duality technique. *Journal of Multivariate Analysis*, 100(1):16–36, 2009.

N. Cressie and T. R. Read. Multinomial goodness-of-fit tests. *J. R. Stat. Soc., Ser. B*, 46: 440–464, 1984.

I. Csiszár. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publ. Math. Inst. Hung. Acad. Sci., Ser. A*, 8:85–108, 1963.

E. De Giorgi. Reward-risk portfolio selection and stochastic dominance. *Journal of Banking & Finance*, 29(4):895–926, 2005.

P. Fishburn. Mean-risk analysis with risk associated with below-target returns. *American Economic Review*, 67(2):116–126, 1977.

H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002.

F. Foster and C. Whiteman. An application of Bayesian option pricing to the Soybean market. *American journal of agricultural economics*, 81(3):722–727, 1999.

M. R. Haley and C. H. Whiteman. Generalized safety first and a new twist on portfolio performance. *Econometric Reviews*, 27(4–6):457–483, 2008.

L. Jager and J. A. Wellner. Goodness-of-fit tests via phi-divergences. *Ann. Stat.*, 35(5):2018–2053, 2007.

P. Jorion. *Value at risk: the new benchmark for controlling market risk*, volume 2. McGraw-Hill New York, 1997.

A. Keziou. Dual representation of ϕ-divergences and applications. *Comptes Rendus Mathematique*, 336(10):857–862, 2003.

Y. Kitamura and M. Stutzer. An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, 65(4):861–874, July 1997.

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

F. Liese and I. Vajda. On divergences and informations in statistics and information theory. *Information Theory, IEEE Transactions on*, 52(10):4394–4412, Oct. 2006.

B. Lindsay. Efficiency versus robustness - the case for minimum Hellinger distance and related methods. *Annals of Statistics*, 22(2):1081–1114, June 1994.

P. Mahalanobis. On the generalized distance in statistics. In *Proceedings of the National Institute of Sciences of India*, volume 2, pages 49–55. New Delhi, 1936.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

H. Markowitz. *Portfolio Selection*. John Wiley & Sons, New York, 1959.

A. Meucci. Fully flexible views: theory and practice. *Risk*, 21(10):97–102, 2008.

D. Morales, L. Pardo, and I. Vajda. Some new statistics for testing hypotheses in parametric models. *Journal of Multivariate Analysis*, 62(1):137–168, July 1997.

D. Morales, L. Pardo, M. Pardo, and I. Vajda. Limit laws for disparities of spacings. *Journal of Nonparametric Statistics*, 15(3):325–342, June 2003.

T. Morimoto. Markov processes and the H-theorem. *J. Phys. Soc. Jap.*, 18:328–331, 1963.

L. Pardo. *Statistical Inference Based on Divergence Measures*. Taylor & Francis Group, LLC, 2006.

T. R. Read and N. A. Cressie. *Goodness-of-fit statistics for discrete multivariate data*. New York etc.: Springer-Verlag, 1988.

J. Robertson, E. Tallman, and C. Whiteman. Forecasting using relative entropy. *Journal of Money Credit and Banking*, 37(3):383–401, June 2005.

J. Robinson, E. Ronchetti, and G. Young. Saddlepoint approximations and tests based on multivariate M-estimates. *Ann. Stat.*, 31(4):1154–1169, 2003.

R. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2: 21–42, 2000.

A. D. Roy. Safety first and the holding of assets. *Econometrica*, 20(3):431–449, 1952.

C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3): 379–423, 1948.

F. Sortino and R. Van Der Meer. Downside risk. *The Journal of Portfolio Management*, 17(4): 27–31, 1991.

M. Stutzer. A simple nonparametric approach to derivative security valuation. *Journal of Finance*, 51(5):1633–1652, Dec. 1996.

M. Stutzer. A portfolio performance index. *Financial Analysts Journal*, 56(3):52, 2000.

A. Toma and M. Broniatowski. Dual divergence estimators and tests: Robustness results. *Journal of Multivariate Analysis*, 102(1):20–36, 2011.

A. Toma and S. Leoni-Aubin. Robust tests based on dual divergence estimators and saddlepoint approximations. *Journal of Multivariate Analysis*, 101(5):1143–1155, 2010.

I. Vajda. *Theory of statistical inference and information.* Dordrecht etc.: Kluwer Academic Publishers, 1989.

M. Yamada and J. Rajasekera. Portfolio re-balancing with the entropy criteria. Technical report, Report No. 310, QUICK Research Institute Corp, 1993.