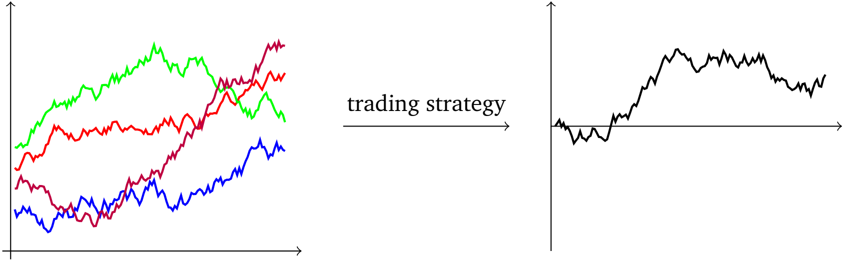
Portfolios generated by optimal transport

Leonard Wong

University of Southern California

USC Math Finance Colloquium
Oct 2, 2017

Motivation: performance attribution



market prices and information

portfolio value

Motivation

- Portfolio value is an integral of the trading strategy η against the price process μ:

$$V(t) = \int_0^t \sum_i \eta_i d\mu_i$$

- Path-dependent, no convenient simplifications in general

Stochastic portfolio theory

- Started by E. R. Fernholz (1999, 2002)
- Use market portfolio as the numeraire
- Functionally generated portfolios:

$$\log V(t) - \log V(0) = \text{market diversity} + \text{volatility}$$

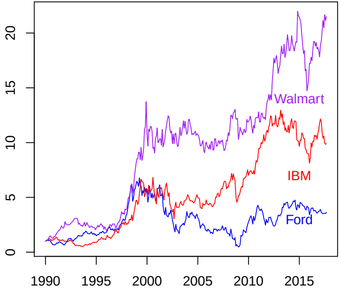
- Conditions to outperform the market: relative arbitrage

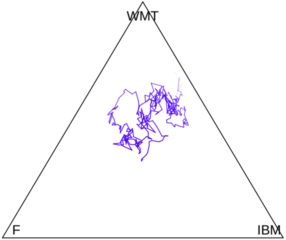
Simplex model in discrete time

- $X_i(t) > 0$ market capitalization of stock i at time t
- Market weight = price relative to market:

$$\mu_i(t) = \frac{X_i(t)}{X_1(t) + \cdots + X_n(t)}$$

- $\mu(t) \in \Delta_n$ unit simplex





Self-financing trading strategy

Two representations:

- $\eta(t) =$ number of shares

$$V(t) - V(0) = \sum_{s=0}^{t-1} \eta(s) \cdot (\mu(s+1) - \mu(s)) \quad \text{(additive)}$$

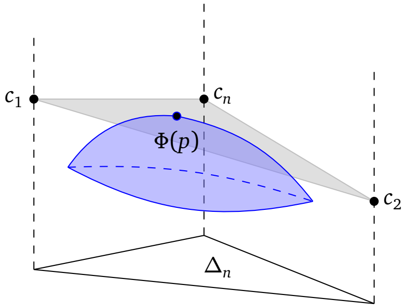
- $\pi(t) =$ portfolio weights

$$\frac{V(t)}{V(0)} = \prod_{s=0}^{t-1} \pi(s) \cdot \frac{\mu(s+1)}{\mu(s)} \qquad \text{(multiplicative)}$$

Here $V(t)$ is portfolio value relative to market.

Multiplicatively generated portfolio

Take a (smooth) function $\varphi : \Delta_n \to (0, \infty)$ such that $\Phi = e^{\varphi}$ is concave. We call φ exponentially concave.



- Define the portfolio map $\pi : \Delta_n \to \overline{\Delta}_n$ by

$$\pi_i(p) = \frac{c_i p_i}{c_1 p_1 + \cdots + c_n p_n} = p_i \left(1 + D_{e_i - p} \varphi(p)\right).$$

- The strategy: $\pi(t) = \pi(\mu(t))$.

Some literature

- Original papers:
  - R. Fernholz (1999, 2001, 2002)
- Applications to relative arbitrage:
  - Banner, D. Fernholz, R. Fernholz, Karatzas, Kadaras, Pal, Ruf, Vervuurt, ...
- Optimal transport and information geometry:
  - Pal and Wong (2013, 2015, 2016), Vervuurt (2016), Pal (2016), Wong (2017)
- Optimization, machine learning and universal portfolio:
  - Wong (2015, 2016), Kom Samo and Vervuurt (2016), Cuchiero, Schachermayer and Wong (2016)
- Extensions:
  - Strong (2014), Schied, Speiser and Voloshchenko (2016), Karatzas and Ruf (2016), Ruf and Xie (2017)

Logarithmic divergence (Pal and Wong (2015))

The L-divergence of φ is

$$D_L[q \mid p] = \log(1 + \nabla\varphi(p)\cdot(q-p)) - (\varphi(q) - \varphi(p)), \quad p, q \in \Delta_n.$$

- $D_L[q \mid p] \geq 0$. If e^{φ} is strictly concave then $= 0 \Rightarrow p = q$.
- $D_L[q \mid p] \not\equiv D[p \mid q]$. Distance-like but is not a metric.

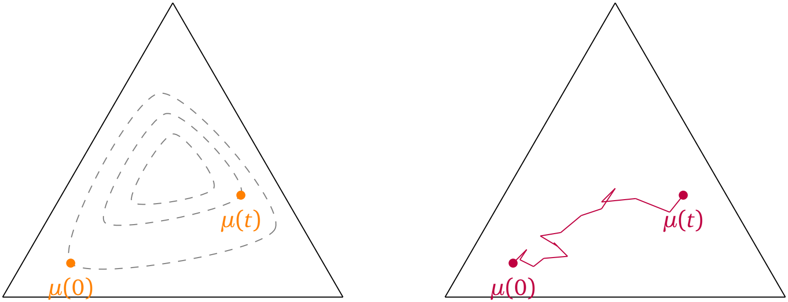
Example: Take $\varphi(p) = \sum_{i=1}^{n} \pi_i \log p_i$, $\pi \in \overline{\Delta}_n$ fixed:

- $\boldsymbol{\pi}(p) \equiv \pi$ (constant-weighted portfolio)
- $D_L[q \mid p] = \log\left(\sum_{i=1}^{n} \pi_i \frac{q_i}{p_i}\right) - \sum_{i=1}^{n} \pi_i \log \frac{q_i}{p_i}$, also known as: diversification return, excess growth rate, cumulant generating function (free energy)

Multiplicative pathwise decomposition

Theorem (Fernholz (1999), Pal and Wong (2015))

$$\log V(t) - \log V(0) = \varphi(\mu(t)) - \varphi(\mu(0)) + \sum_{s=0}^{t-1} D_L\left[\mu(s+1) \mid \mu(s)\right]$$



change in market diversity

accumulated market volatility

Additively generated portfolio (Karatzas and Ruf (2016))

Take a (smooth) concave function $\varphi : \Delta_n \to (0, \infty)$.

- The number of shares is

$$\eta_i(t) = D_{e_i - \mu(t)} \varphi(\mu(t)) + V(t).$$

- The Bregman divergence (Bregman (1967)) of φ is

$$D_B[q \mid p] = \nabla \varphi(p) \cdot (q - p) - (\varphi(q) - \varphi(p)), \quad p, q \in \Delta_n.$$

Theorem (Additive pathwise decomposition)

$$V(t) - V(0) = \varphi(\mu(t)) - \varphi(\mu(0)) + \sum_{s=0}^{t-1} D_B[\mu(s+1) \mid \mu(s)]$$

Our discrete time set up leads to the Bregman divergence.

Unified framework

Definition

We say that the trading strategy is generated by φ, with scale function g, if there exists a divergence $D[\cdot \mid \cdot]$ on Δ_n such that

$$g(V(t)) - g(V(0)) = \varphi(\mu(t)) - \varphi(\mu(0)) + \sum_{s=0}^{t-1} D[\mu(s+1) \mid \mu(s)]$$

for all market paths.

Examples:

- Multiplicative generation: $g(x) = \log x$, $D[\cdot|\cdot]$ is L-divergence
- Additive generation: $g(x) = x$, $D[\cdot|\cdot]$ is Bregman divergence

We will characterize *all* possibilities.

General concept of divergence

Definition

A divergence on Δ_n is a functional $D[\cdot \mid \cdot] : \Delta_n \times \Delta_n \to [0, \infty)$:

(i) $D[q \mid p] = 0 \Rightarrow p = q$.

(ii) When $q = p + \Delta p \approx p$, a quadratic approximation holds:

$$D[p + \Delta p \mid p] = \frac{1}{2} \sum_{i,j=1}^{n} g_{ij}(p) \Delta p_i \Delta p_j + o(|\Delta p|^3),$$

where g is strictly positive definite.

Examples:

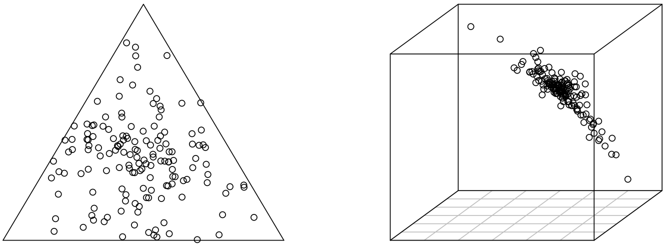
- Bregman divergence: $g(p) = -D^2 \varphi(p)$
- L-divergence: $g(p) = -D^2 \varphi(p) - (\nabla \varphi(p))(\nabla \varphi(p))^\top$

Connection with optimal transport

Multiplicatively generated portfolio: Pal and Wong (2015, 2016)

Additively generated portfolio:

- $\eta(t) = \text{portfolio} \longleftrightarrow \nabla\varphi(\mu(t)) = \text{gradient of concave function}$



The map $p \mapsto \nabla\varphi(p)$ where $\varphi(p) = -\sum_{i=1}^{n} p_i \log p_i$

Connection with optimal transport

- Additive (Vervuurt 2016): Consider the cost function

$$c(p,v) = p \cdot v, \quad p \in \Delta_n, v \in \mathbb{R}^n \text{ tangent to } \Delta_n$$

  Then, for any $p^{(1)}, \ldots, p^{(m)} \in \Delta_n$, the transport map $p^{(k)} \mapsto v^{(k)} = \nabla \varphi(p^{(k)})$ solves the optimal transport problem

$$\min_{\sigma \text{ permutation}} \sum_{k=1}^{n} c(p^{(k)}, v^{(\sigma(k))}).$$

  This follows from Rockafellar's theorem in convex analysis. (Holds true for transport of general Borel probability measures.)

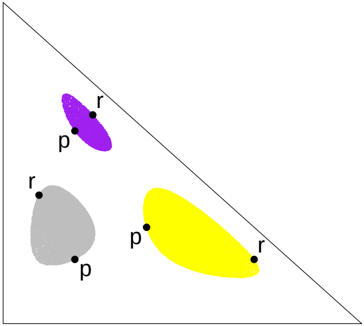
- Multiplicative (Pal and Wong (2015)): analogous results hold where
  - $c(p, q^*) = \log(p \cdot q^*), \quad p \in \Delta_n,\ q^* \in \Delta_n^*$
  - The transport map is $p \mapsto q^*$, $q_i^* = \frac{\pi_i(p)/p_i}{\sum_{j=1}^n \pi_j(p)/p_j}$

Connection with information geometry

Generalized Pythagorean relation: For $p, q, r \in \Delta_n$, when does

$$D[q \mid p] + D[r \mid q] \geq D[r \mid p]?$$

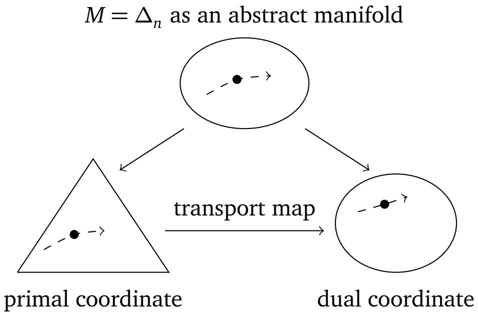
- Optimal trading frequency
- Many applications in statistics and machine learning



The set $\{q : \text{LHS} \leq \text{RHS}\}$ where $\varphi(p) = \frac{1}{3} \sum_{i=1}^{3} \log p_i$

Dual geometry

- Additive (Bregman): Amari and Nagaoka (1982)
- Multiplicative (L): Pal and Wong (2016)
- General framework: Eguchi (1983)



- Primal geodesic: a straight line in primal coordinates
- Dual geodesic: a straight line in dual coordinates

Dual geometry

We also define a Riemannnian metric by

$$\langle u, v \rangle_p = u^\top g(p) v, \quad u, v \text{ tangent vectors at } p,$$

where $g(p)$ is the Riemannian matrix for the divergence $D[\cdot \mid \cdot]$.

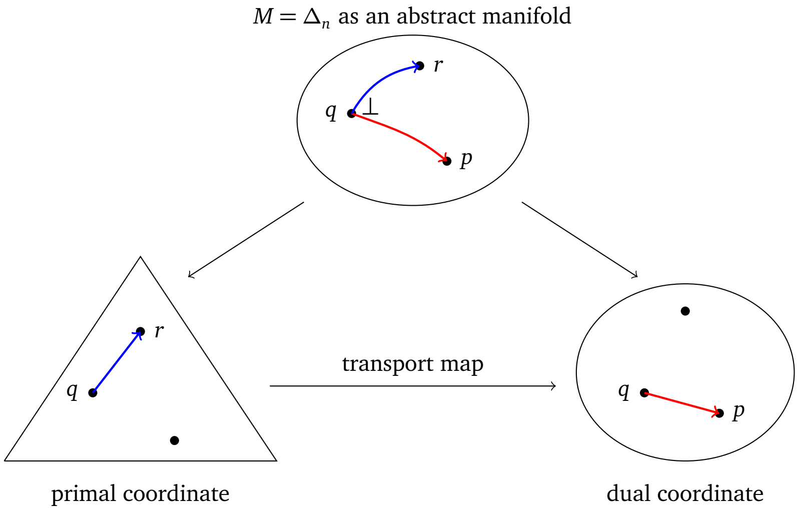
Theorem (Generalized Pythagorean theorem)

Let $D[\cdot \mid \cdot]$ be the Bregman or L-divergence of φ which determines the transport map under respectively the cost $c(p, v) = p \cdot v$ or $c(p, q^*) = \log(p \cdot q^*)$. Then the generalized Pythagorean relation

$$D[q \mid p] + D[r \mid q] = D[r \mid p]$$

holds if and only if the primal geodesic from q to r and the dual geodesic from q to p meet g-orthogonally at q.

Generalized Pythagorean theorem



Characterization of functional generation

Theorem (W. (2017))

The scale function g admits functional portfolio generation if and only if

$$g(x) = c_1 x + c_2,$$

where $c_1 > 0$, $c_2 \in \mathbb{R}$, *or*

$$g(x) = \log(c_1 x + c_2) + c_3,$$

where $c_1 > 0$, $c_2 \geq 0$, $c_3 \in \mathbb{R}$.

Key idea: The pathwise decomposition implies that the scale function satisfies the nonlinear ODE

$$g'(x) g'''(x) = 2(g''(x))^2.$$

(α, C)-portfolio generation: $\alpha > 0$, $C \geq 0$

Definition

Let φ be smooth and α-exponentially concave, i.e., $e^{\alpha\varphi}$ is concave. It generates the trading strategy with

$$\eta_i^{\alpha,C}(t) = \alpha(C + V(t))D_{e_i - \mu(t)}\varphi(\mu(t)) + V(t), \quad i = 1, \ldots, n.$$

If $V(t) > 0$, the portfolio weight is given by

$$\pi^{\alpha,C}(t) = \frac{C + V(t)}{V(t)}\pi^{\alpha,0}(t) - \frac{C}{V(t)}\mu(t),$$

where $\pi^{\alpha,0}$ is generated multiplicatively by $\alpha\varphi$.

This allows us to generate different portfolios with the same φ:

- Multiplicative generation: $(\alpha, C) = (1, 0)$
- Additive generation: limit of $(\alpha, C) = (\alpha, \frac{1}{\alpha})$ as $\alpha \downarrow 0$

General pathwise decomposition

Definition ($L^{(\alpha)}$-divergence)

$$D_{L^{(\alpha)}}[q \mid p] = \frac{1}{\alpha} \log(1 + \alpha \nabla \varphi(p) \cdot (q - p)) - (\varphi(q) - \varphi(p))$$

For φ fixed, it interpolates between the L-divergence ($\alpha = 1$) and the Bregman divergence ($\alpha \downarrow 0$).

Theorem (W. (2017))

If the strategy is (α, C)-generated by φ and $V(\cdot) > -C$, we have

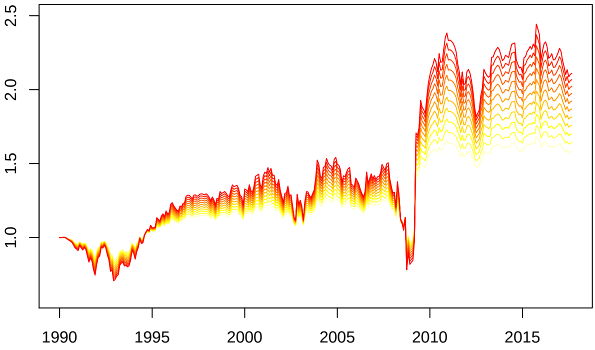
$$\frac{1}{\alpha} \log \frac{C + V(t)}{C + V(0)} = \varphi(\mu(t)) - \varphi(\mu(0)) + \sum_{s=0}^{t-1} D_{L^{(\alpha)}}[\mu(s+1) \mid \mu(s)].$$

Empirical example

Consider the earlier example of Ford, IBM and Walmart. For $0 \leq \alpha \leq 1$, let $\eta^{(\alpha)}$ be the trading strategy $(\alpha, \frac{1}{\alpha})$-generated by

$$\varphi(p) = \frac{1}{3}\sum_{i=1}^{n} \log p_i.$$

We also set $V(0) \equiv 1$. Simulated (relative) portfolio values:



Concluding remarks

- Unified framework of functional portfolio construction
- Approach motivated by optimal transport and divergence in information geometry
- Covers both:
  - Additive portfolio / Bregman divergence / quadratic cost
  - Multiplicative portfolio / L-divergence / logarithmic cost

  and identifies natural interpolations
- Future directions:
  - Further connections and results in optimal transport and information geometry
  - Portfolio optimization, practical applications of transport and geometry in finance and statistics