Portfolio Optimization in a Big Data Context

Thierry Bazier-Matte

Summer 2015

The goal of this document is twofold. First, given a risk averse utility function and a sample of the market observations (returns and states), we want to prescribe an optimal linear investment policy depending on the risk aversion of the investor. Second, we also wish to return to the investor probabilist bounds on the certainty equivalent of the prescribed investment policy.

1 Assumptions and Definitions

In the following, $\boldsymbol{A}$ (capital boldface) are assumed to represent a real subset of any dimension, A (capital case) represents random variables (or distributions) and a (lower case) represents deterministic variables or realizations. $\mathscr{R}$ represents the real set.

Let $M = (X, R)$ the *market* be a random variable with support $\boldsymbol{M} = \boldsymbol{X} \times \boldsymbol{R} \subseteq \mathscr{R}^{p+1}$, ie. numerically qualifiable, with $(x, r) = m \sim M$ a *market observation*, consisting in one part *state* $x \in \mathscr{R}^p$ and another part *outcome* $r \in \mathscr{R}$. Typically x is a vector of observations from various variable of interests, such as financial or economical news, etc. Scalar r in this article shall represent the return from a financial asset of interest. Finally, let $M_n = \{M, \dots, M\}$ be a *random set* of n (unrealized) observations (with support $\boldsymbol{M}^n$). Therefore $\mu_n \sim M_n$ represents an iid sample of n market observations.

We shall study linear investment decisions q such that $p = q^T x$ part of the portfolio is allocated to the risky asset with return r, and $1 - p$ part of the portfolio is allocated to the riskless asset, providing a total return $rp + R_f(1 - p)$.

Assumption. We suppose that observed returns r are constrained by $|r| \leq \bar{r}$ and that observed states x are constrained by $\|x\|_2 \leq X_{\max}$.

Assumption. We also suppose that the investor's utility function u is such that u can be rescaled to a standard utility $\bar{u}$ with transformation $u(r) = k\bar{u}(r) + l$, where $k > 0$ and $\bar{u}$

has the following properties:

- $\bar{u}$ is concave and non-decreasing;
- Let X be the random returns of the portfolio. Then $CE[X] \leq E[X] - \epsilon Var(X)$.;
- $\bar{u}$ is Lipschitz continuous, with constant γ, ie. $|\bar{u}(r_1) - \bar{u}(r_2)| \leq \gamma |r_1 - r_2|$;
- $\bar{u}$ is bounded above, ie. there exists a constant M such that $\bar{u}(r) \leq M$ for any r;
- $\bar{u}(0) = 0$;
- $\lim_{r \to 0^+} \bar{u}(r) = 1$.

Examples of standard utility functions are $\bar{u}(r) = \min(\beta x, x)$ with $\beta > 1$ and $\bar{u}(r) = \beta^{-1}(e^{\beta x} - 1)$ with $\beta > 0$. [Pente non contrôlée!]

Definition. Let $\ell : \boldsymbol{M} \times \boldsymbol{Q} \to \mathscr{R}$ be a *loss function* defined by

$$\ell(m, q) = \ell(x, r, q) = -u(r\, q^T x + R_f(1 - q^T x)),$$

where R_f is the risk free return rate. We also define the *cost function* $c : \mathscr{R} \times \boldsymbol{R} \to \mathscr{R}$ as

$$c(p, r) = -u(pr + (1 - p)R_f),$$

so that $\ell(x, r, q) = c(q^T x, r)$.

Definition. The *empirical risk* $\hat{R} : \boldsymbol{M}^n \times \boldsymbol{Q} \to \mathscr{R}$ associated with decision q and market sample μ_n is given by

$$\hat{R}_{\mu_n}(q) = n^{-1} \sum_{i=1}^{n} \ell(m_i, q).$$

Definition. The *empirical decision algorithm* $\hat{A}_n : \boldsymbol{M}^n \to \boldsymbol{Q}$ associated with market sample μ_n is the optimal value of the problem

$$\text{minimize} \quad \hat{R}_{\mu_n}(q) + \lambda \|q\|_2^2.$$

From now on, $\hat{q}_n := \hat{A}_n(\mu_n)$ the empirical decision associated with market sample μ_n and $\hat{Q}_n := A_n(S_n)$ the random empirical decision, ie. $\hat{q}_n \sim \hat{Q}_n$.

Definition. The *true risk* $R_{\text{true}} : \boldsymbol{Q} \to \mathscr{R}$ associated with decision q is given by

$$R_{\text{true}}(q) = E_M[\ell(m, q)].$$

Definition. The *optimal decision* $q^\star$ is the optimal value of the problem

$$\text{minimize} \quad R_{\text{true}}(q).$$

Definition. The *optimal regularized decision* ${q^\star}_\lambda$ is the optimal value of the regularized problem

$$\text{minimize} \quad R_{\text{true}}(q) + \lambda \|q\|_2^2.$$

2 [Stability Definitions and Theorems]

We adapt in this section a number of theorems and definitions from Bousquet 2002 [Add reference].

2.1 Algorithmic Stability

Definition. Let $\hat{q}_n = \hat{A}_n(\mu_n)$ and $\hat{q}_{n\setminus i} = \hat{A}_n(\mu_{n\setminus i})$, where μ_n and $\mu_{n\setminus i}$ only differs in their i^{th} observation, which has been redrawn from M in the case of $\mu_{n\setminus i}$. The algorithm $\hat{A}_n$ is said to have *uniform stability* α_n if, for any $m \sim M$,

$$|\ell(m, \hat{q}_n) - \ell(m, \hat{q}_{n\setminus i})| \leq \alpha_n.$$

Definition. A loss function ℓ is σ-admissible if its cost function c is convex with respect to p the investment decision and the following holds for any p_1, p_2 and r:

$$|c(p_1, r) - c(p_2, r)| \leq \sigma|p_1 - p_2|.$$

Remark. The loss function as defined above is σ-admissible with $\sigma = k\gamma(r + R_f)$. See Claim 1.

Theorem 1. *If ℓ is σ-admissible and if, for any $x \in \boldsymbol{X}$, $\|x\|_2^2 \leq X_{\max}^2$, then $\hat{A}_n$ has uniform stability with*

$$\alpha_n = \frac{\sigma^2 X_{\max}^2}{2\lambda n}.$$

Proof. See Bousquet, Theorem 22. □

We therefore conclude that $\hat{A}_n$ has uniform stability with

$$\alpha_n = \frac{k^2\gamma^2(\bar{r} + R_f)^2 X_{\max}^2}{2\lambda n}.$$

2.2 Out of Sample Bound

Theorem 2. *If $\hat{A}_n$ has uniform stability α_n and the loss function is such that for any $m \sim M$ and any $\hat{q}_n = \hat{A}_n(\mu_n)$, $0 \leq \ell(m, \hat{q}_n) \leq B_n$, then for any $\delta \in (0, 1)$, the following bound holds with probability at least $1 - \delta$ over the random sample draw $\mu_n \sim M_n$:*

$$|R_{\text{true}}(\hat{q}_n) - \hat{R}(\hat{q}_n)| \leq 2\alpha_n + (4n\alpha_n + B_n)\sqrt{\frac{\log(2/\delta)}{2n}}.$$

The bound B_n from Theorem 2, can be explicitly stated: the highest loss occurs when investment decision is at its highest and the return at its lowest. With the definition of $\bar{p}$ from Claim 2, we find that

$$\begin{aligned} c(p,r) &\leq c(\bar{p}, -\bar{r}) \\ &= -k\bar{u}(-\bar{p}\bar{r} + (1-\bar{p})R_f) - l \\ &\leq |l| + k|\bar{u}(-\bar{p}\bar{r} + (1-\bar{p})R_f)| \\ &\leq |l| + k\gamma|R_f - \bar{p}(\bar{r} + R_f)|. \end{aligned}$$

[Can we do better? Here, we've taken the maximum slope of $\bar{u}$, ie. γ to derive the results, where as we should check the slope at this particular point.]

2.3 True Optimal Bound

We now derive a measure of performance where we compare the empiric risk $\hat{R}(\hat{q})$ with the optimal risk $R_{\text{true}}(q^\star)$.

We have the following inequality:

$$|R_{\text{true}}(q^\star) - \hat{R}(\hat{q}_n)| \leq |R_{\text{true}}(\hat{q}) - \hat{R}(\hat{q})| + |R_{\text{true}}(q^\star) - R_{\text{true}}(\hat{q})|.$$

The first term has already been identified in the previous equation (REFERENCE), while the second is bounded by

$$|R_{\text{true}}(q^\star) - R_{\text{true}}(\hat{q})| \leq (\bar{r} + R_f)X_{\max}\|q^\star - \hat{q}_n\|_2 + R_f.$$

See Claim 3 in appendix for proof.

Therefore, what we're looking for is a probabilistic bound on the variance of $\hat{q}_n = \hat{A}_n(\mu_n)$ [because, using CLT, $\hat{q}_n$ converges in distribution to ${q^\star}_\lambda$ and therefore its variance can tell us how close we are from its expected value ${q^\star}_\lambda$].

Appendix

Claim 1. *The following inequality holds:*

$$|c(p_1, r) - c(p_2, r)| \leq k\gamma(\bar{r} + R_f)|p_1 - p_2|.$$

Proof. Using the Lipschitz property of $\bar{u}$ we have:

$$\begin{aligned}
|c(p_1, r) - c(p_2, r)| &= |u(p_1 r + (1 - p_1)R_f) - u(p_2 r + (1 - p_2)R_f)| \\
&= k|\bar{u}(p_1 r + (1 - p_1)R_f) - \bar{u}(p_2 r + (1 - p_2)R_f)| \\
&\leq kc|p_1 r + (1 - p_1)R_f - p_2 r - (1 - p_2)R_f| \\
&= k\gamma|r - R_f||p_1 - p_2| \\
&\leq k\gamma(\bar{r} + R_f)|p_1 - p_2|.
\end{aligned}$$

□

Claim 2. *The ℓ_2 norm of decision vectors $\hat{q}_n$ obtained from algorithm $\hat{A}_n$ are bounded by $\bar{p} = k\gamma X_{\max}(\bar{r} - R_f)/(2\lambda)$.*

Proof. Let μ_n be a sample of the market. The empirical decision algorithm

$$\text{minimize} \quad n^{-1}\sum_{i=1}^{n} \ell(m_i, q) + \lambda\|q\|_2^2$$

is equivalent to

$$\begin{aligned}
\text{minimize} \quad & n^{-1}\sum_{i=1}^{n} \ell(m_i, sq) + \lambda s^2 \\
\text{subject to} \quad & s \geq 0 \\
& \|q\|_2 = 1,
\end{aligned}$$

where the optimization variables are now on the direction (q) and the scale (s). Therefore, for any direction q, we can define a convex function $g(s)$ which becomes the objective:

$$\begin{aligned}
\text{minimize} \quad & g(s) \\
\text{subject to} \quad & s \geq 0,
\end{aligned}$$

where

$$g(s) = n^{-1}\sum_{i=1}^{n} \ell(m_i, sq) + \lambda s^2.$$

Convexity of g follows from convexity of ℓ (see [Claim ??]).

Because g is convex, we can consider two cases: either the minimum is realized at the boundary, ie. $s^\star = 0$, or there exists an optimal value $s^\star > 0$ such that $g'(s^\star) = 0$. To derive a bound on $s^\star$, we can seek a value $\bar{s}$ such that for any q, $g'(\bar{s}) > 0$ and therefore $\bar{s} > s^\star$.

To do so, we first note that

$$\begin{aligned} g'(s) &= \nabla_s \left[n^{-1} \sum_{i=1}^{n} \ell(m_i, sq) + \lambda s^2 \right] \\ &= 2\lambda s - k n^{-1} \sum_{i=1}^{n} \nabla_s \bar{u}(s r_i q^T x_i + R_f(1 - s q^T x_i)) \\ &= 2\lambda s - k n^{-1} \sum_{i=1}^{n} (r_i - R_f) q^T x_i \, \bar{u}'(s r_i q^T x_i + R_f(1 - s q^T x_i)). \end{aligned}$$

Now, because $\|q\|_2 = 1$, we have $q^T x_i \leq \|x_i\|_2 \leq X_{\max}$. We also have $r_i - R_f \leq \bar{r} - R_f$ and $\bar{u}' \leq \gamma$, so that

$$k\gamma X_{\max}(\bar{r} - R_f) \geq n^{-1} \sum_{i=1}^{n} (r_i - R_f) q^T x_i \, \bar{u}'(s r_i q^T x_i + R_f(1 - s q^T x_i)).$$

Therefore, with

$$\bar{s} := \frac{k\gamma}{2\lambda} X_{\max}(\bar{r} - R_f),$$

for any $s > \bar{s}$,

$$g'(s) \geq 0.$$

The previous inequality therefore implies that the norm $\|\hat{q}_n\|_2$ is at most $\bar{s}$. □

Claim 3. *The following inequality holds:*

$$|R_{\text{true}}(q^\star) - R_{\text{true}}(\hat{q})| \leq (\bar{r} + R_f) X_{\max} \|q^\star - \hat{q}_n\|_2 + R_f.$$

Proof. The following development holds:

$$\begin{aligned} |R_{\text{true}}(q^\star) - R_{\text{true}}(\hat{q}_n)| &= |E_M[\ell(m, q^\star)] + E_M[-\ell(m, \hat{q}_n)] \\ &= |E_M[\ell(m, q^\star) - \ell(m, \hat{q}_n)]| \\ &\leq E_M[|\ell(m, q^\star) - \ell(m, \hat{q}_n)|]. \end{aligned}$$

Now, if we let $m = (x, r)$, using the Lipschitz property of u, with Lipschitz constant γ, and the Hölder's inequality, we also have the following identity:

$$\begin{aligned}
|\ell(m, q^\star) - \ell(m, \hat{q}_n)| &= |u(r\, q^{\star T} x + R_f(1 - q^{\star T} x)) - u(r\, \hat{q}_n^T x + R_f(1 - \hat{q}_n^T x))| \\
&\leq \gamma |r\, q^{\star T} x + R_f(1 - q^{\star T} x) - r\, \hat{q}_n^T x - R_f(1 - \hat{q}_n^T x)| \\
&= \gamma |r(q^\star - \hat{q}_n)^T x + R_f(1 - (q^\star - \hat{q}_n)^T x)| \\
&\leq \gamma |r(q^\star - \hat{q}_n)^T x| + R_f |1 - (q^\star - \hat{q}_n)^T x| \\
&\leq \gamma(\bar{r} X_{\max} \|q^\star - \hat{q}_n\|_2 + R_f(1 + X_{\max} \|q^\star - \hat{q}_n\|_2)) \\
&= \gamma((\bar{r} + R_f) X_{\max} \|q^\star - \hat{q}_n\|_2 + R_f).
\end{aligned}$$

Where we used the bounds on outcomes x and r. And so we are left with the claimed inequality. $\square$

Claim 4. *If the utility $u(r)$ is unbounded as $r \to \infty$ and the random return R is a linear transformation of the random market state, ie. $R = t^T X$, then $\|q^\star\|_2$ is unbounded.*

Proof. We first observe that, by definition, $q^\star$ is the minimal value of $E_M[\ell(m, q)]$, ie.

$$\begin{aligned}
q^\star &= \arg\max_q E_M[u(r\, q^T x + R_f(1 - q^T x)] \\
&\leq \arg\max_q E_M[r\, q^T x + R_f(1 - q^T x)] \\
&= \arg\max_q E_M[t^T x q^T x] + R_f E_M[1 - q^T x] \\
&= \arg\max_q t^T E_M[x x^T] q + R_f E_M[1 - q^T x] \\
&= \arg\max_q t^T \Sigma q + R_f E_M[1 - q^T x],
\end{aligned}$$

with Σ the covariance of X. It is therefore easy to see that if $(t^T \Sigma)_i$ is positive, we set $q_i = \infty$ and if $(t^T \Sigma)_i$ is negative, we set $q_i = -\infty$, so that the total expression tends to infinity. [The general idea is here, but the proof is far from perfect.] $\square$