Decision errors regarding portfolio optimization: the ameliorating role of conservative selection via the concentration of probability phenomenon

Introduction

Portfolio optimization based on stochastic dominance (SD) is a non-parametric generalization of the standard mean-variance approach, for the optimal selection of portfolio weights' regarding investment strategies outside the realm of satiation and/or elliptical returns' distributions.[1]

In the relevant literature (see for example Constantinides et al. (2020) (3)) portfolio selection is usually performed via the optimization of an empirical criterion under the constraint that the choice set is comprised by portfolios that empirically dominate an exogenous benchmark portfolio. The empirically optimal portfolio by construction dominates the benchmark in the sample, it is however susceptible to the decision error of False Dominance (FD) classification in the population.

Under general sampling schemes this decision error becomes asymptotically negligible. Controlling the probability of this error for fixed sample size is however important in applications, especially when the sample size is not particularly large compared to the dimensionality of the base assets considered. In several relevant applications a heuristic used in the underlying empirical optimization seems to improve the out-of-sample properties of the optimal portfolio. The optimization problem is augmented by a restriction on the distance of the portfolio sought compared to the benchmark. Is this heuristic theoretically justified?

Stochastic dominance and portfolio optimization framework

$(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a stationary process with values in some subset of $\mathbb{R}^d$. The random vector $\boldsymbol{X}_t$ represents the one period stationary returns of d financial assets, and $\mathcal{X} \subset \mathbb{R}^d$ is the pointwise bounded from below support of its' latent joint distribution. Boundedness from above is considered plausible for moderate observation frequencies. The researcher has at her disposal an observable sample from the process, $(\boldsymbol{X}_t)_{t=1,\cdots,T}$; $\mathbb{P}_T$ denotes the empirical distribution of the sample.

[1] This post is heavily based on the results and formulations of the paper here.

A portfolio on $\boldsymbol{X}_0$ is any real linear function on $\mathbb{R}^d$; the elements of its representing vector are the portfolio weights. Alternative portfolios are evaluated inside the expected utility paradigm, using utility functions $u : \mathcal{X} \to \mathbb{R}$ that are increasing, continuous, and concave. Those utilities form the population of the functional set $\mathcal{U}_2$.

The analysis involves a set of portfolios $\boldsymbol{\Lambda} \subseteq \mathbb{R}^d$. In what follows $\boldsymbol{\lambda}$, $\boldsymbol{\tau}$ denote respectively a typical element of $\boldsymbol{\Lambda}$ and a distinguished benchmark portfolio inside $\boldsymbol{\Lambda}$.

The above enable the definition of a stochastic dominance relation on the sets of prospects, via $\mathcal{U}_2$: in the stationary framework considered, $\boldsymbol{\lambda}$ is said to dominate $\boldsymbol{\tau}$ w.r.t. the utility class $\mathcal{U}_2$ iff $D(u, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbb{P}) := \mathbb{E}_{\mathbb{P}}(u(\boldsymbol{\lambda}'\boldsymbol{X}_0)) - \mathbb{E}_{\mathbb{P}}(u(\boldsymbol{\tau}'\boldsymbol{X}_0)) \geq, \forall u \in \mathcal{U}_2$-here $\mathbb{E}_{\mathbb{P}}$ denotes integration w.r.t. $\mathbb{P}$. Thus, $\boldsymbol{\lambda}$ is preferred over $\boldsymbol{\tau}$ by every utility in the considered class, this is what is known in the literature as the second order stochastic dominance of $\boldsymbol{\lambda}$ over the benchmark $\boldsymbol{\tau}$; $\boldsymbol{\lambda} \underset{\mathbb{P},2}{\succeq} \boldsymbol{\tau}$. $\boldsymbol{\Lambda}_{\mathbb{P}}^{\succeq} := \left\{ \boldsymbol{\lambda} \in \boldsymbol{\Lambda}; \boldsymbol{\lambda} \underset{\mathbb{P},2}{\succeq} \boldsymbol{\tau} \right\}$ is the non-empty convex set of portfolios that dominate the benchmark in the population. Non-emptyness holds due to reflexivity of the dominance relation; the benchmark actually dominates itself, and convexity follows from the concavity of the utility functions at hand, the linearity of the portfolio formation and the monotonicity of the integral. Substituting the latent $\mathbb{P}$ with $\mathbb{P}_T$ in $\underset{\mathbb{P},2}{\succeq}$, the empirical analogue $\boldsymbol{\Lambda}_{\mathbb{P}_T}^{\succeq}$ is obtained.

Consider a choice $\boldsymbol{\lambda}_{\mathbb{P}_T} \in \boldsymbol{\Lambda}_{\mathbb{P}_T}^{\succeq}$. Controlling the probability of FD for $\boldsymbol{\lambda}_{\mathbb{P}_T}$, i.e. $\mathbb{P}(\boldsymbol{\lambda} \underset{\mathbb{P}_T,2}{\succeq} \boldsymbol{\tau}) \,/\, \boldsymbol{\lambda} \underset{\mathbb{P},2}{\not\succeq} \boldsymbol{\tau})$, can be of particular empirical interest, as FD can lead to suboptimal portfolio choices. This can asymptotically-as $T \to \infty$-vanish as long as the probabilistic properties of the sampling scheme ensure that $\mathbb{P}_T \rightsquigarrow \mathbb{P}$ plus some moment uniform existence conditions. The question of controlling this probability is also of interest for fixed-and potentially realistically large enough T. This is what is investigated in the subsequent analysis.

Every choice $\boldsymbol{\lambda}_{\mathbb{P}} \in \boldsymbol{\Lambda}_{\mathbb{P}}^{\succeq}$ can be represented as a solution-albeit in some cases trivial-of the optimization problem $\max_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}_{\mathbb{P}}^{\succeq}} \mathbb{E}_{\mathbb{P}}(u(\boldsymbol{\lambda}'\boldsymbol{X}_0))$ for some $u \in \mathcal{U}_2$. More importantly for a given non-constant $u \in \mathcal{U}_2$ any solution, say $\boldsymbol{\lambda}(u, \mathbb{P})$, to the optimization problem $\max_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}_{\mathbb{P}}^{\succeq}} \mathbb{E}_{\mathbb{P}}(u(\boldsymbol{\lambda}'\boldsymbol{X}_0))$ can be of economic interest; any such latent portfolio is perceivable as the best a risk averter investor with preferences represented by the particular u can achieve in terms of expected utility, if she insists on working with portfolios that would be weakly preferred by every risk averter to the benchmark. This is a problem of portfolio optimization augmented with stochastic dominance (second order) SD conditions.

Latency of $\mathbb{P}$ implies generally latency of $\boldsymbol{\lambda}(u, \mathbb{P})$. The latter can be statistically approximated by its empirical analogue; $\boldsymbol{\lambda}(u, \mathbb{P}_T)$, i.e. the solution to the empirical portfolio optimization augmented with empirical stochastic dominance conditions $\max_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}_{\mathbb{P}_T}^{\succeq}} \mathbb{E}_{\mathbb{P}_T}(u(\boldsymbol{\lambda}'\boldsymbol{x}))$. Hence the analysis that follows considers an arbitrary yet fixed u and asks whether there is a modification of the optimization problem that enables the non-asymptotic investigation of the probability of FD for its solutions.

Regularized formulation of portfolio optimization

A modification used in practice augments the expected utility criterion with an additive regularization term that depends on the ℓ^p distance between the portfolio sought and the benchmark. The intuition is that when the (Lagrange) multiplier of the aforementioned distance is chosen optimally, then in order for a portfolio that lies "away" from the benchmark to solve the optimization problem, it would have to "strongly" satisfy the empirical dominance conditions at least in some neighborhood of u.

As mentioned above, the ℓ_p-distance from the benchmark portfolio weights is considered here, $\|\boldsymbol{\lambda}-\boldsymbol{\tau}\|_p := (\sum_{i=1}^{d}|\lambda_i-\tau_i|^p)^{1/p}$, for the case where $p \geq 1$, and $\max_{i=1,\dots,d}|\lambda_i-\boldsymbol{\tau}_i|$ for $p=+\infty$. The regularized optimization portfolio is then defined by:

$$\boldsymbol{\lambda}(u,\mathbb{P}_T,p,\xi_T) \in \arg\max_{\boldsymbol{\Lambda}_{\mathbb{P}_T^{\succeq}}}(\mathbb{E}_{\mathbb{P}_T}(u(\boldsymbol{\lambda}'\boldsymbol{x})) - \xi_T\|\boldsymbol{\lambda}-\boldsymbol{\tau}\|_p), \tag{1}$$

where the random variable $\xi_T \geq 0$ assumes the role of the regularization (Lagrange) multiplier. Its optimal selection is expected to influence the non-asymptotic properties of the probability of FD. The modified problem additionally thus depends on both the choice of the multiplier ξ_T and the norm order p.

Results

The issue of the derivation of non-asymptotic properties for the portfolio solutions of the empirical regularized problem is considered here, with a view towards the fixed T properties of the probability of FD.

Some further notation will be useful: for $\mathbb{Q}$ an arbitrary distribution on $\mathbb{R}^d$, and q defined via $\frac{1}{p}+\frac{1}{q}=1$, the first Wasserstein distance between $\mathbb{Q}$ and the empirical distribution $\mathbb{P}_T$ is $\mathcal{W}(\mathbb{P}_T,\mathbb{Q};p) := \min_{\gamma\in\Gamma(\mathbb{P}_T,\mathbb{Q})}\int_{\mathbb{R}^d\times\mathbb{R}^d}\|z-z^\star\|_q\, d\gamma(z,z^\star)$, where $\Gamma(\mathbb{P}_T,\mathbb{Q})$ denotes the set of probability distributions on $\mathbb{R}^d\times\mathbb{R}^d$ that have respective "marginals" $\mathbb{P}_T$, $\mathbb{Q}$, and also have finite first moment. For $\epsilon>0$, $\mathcal{M}_p(\mathbb{P}_T,\epsilon) := \{\mathbb{P} : \mathcal{W}(\mathbb{P}_T,\mathbb{Q};p)\leq\epsilon\}$ is the Wasserstein closed ball centered at $\mathbb{P}_T$ with radius ϵ.

The notion of the Wasserstein distance plays a very important role in modern probability; it is directly connected to the problem of optimal transport which has several interesting interpretations in economics (see Galichon (2018) (4)). It formulates geometries for spaces of probability distributions that have surprising connections to entropy and the geometry of the underlying spaces upon which those distributions are defined. We will explore such issues in subsequent posts. For now it is sufficient that they are helpful in defining balls comprised of probability measures centered at $\mathbb{P}_T$.

Conservative formulation

A characterization of the regularized problem in (1) as a distributionally robust optimization (DRO) problem is obtained here. It is based on strong convex duality results for robust optimization (see for example Lemma 1 of Gao, Chen, and Kleywegt (2017) (5)).

The event defined as

$$\mathcal{E}_2 := \left\{ \forall v \in \mathcal{U}_2^\star,\ \forall \boldsymbol{\lambda}(v, \mathbb{P}_T, p, \xi_T) : \inf_{\mathbb{Q} \in \mathcal{M}_p(\mathbb{P}_T, \xi_T)} D(v, \boldsymbol{\tau}, \boldsymbol{\lambda}(v, \mathbb{P}_T, \xi_T), \mathbb{Q}) > 0 \right\},$$

is comprised by every sample realization for which every choice of the objective utility results to non-trivial empirical solutions; there $\mathcal{U}_2^\star$ is a subset of $\mathcal{U}_2$ comprised of non-constant utilities. Then it can be proven:

Proposition 1. *(Distributional Robustness). Suppose that u has a unital Lipschitz coefficient. Then, conditional on $\mathcal{E}_2$, Problem (1) is equivalent to:*

$$\sup_{\boldsymbol{\Lambda}^{\succeq}_{\xi_T}} \inf_{\mathbb{Q} \in \mathcal{M}_p(\mathbb{P}_T, \xi_T)} \mathbb{E}_{\mathbb{Q}}(u(\boldsymbol{\lambda}'\boldsymbol{x})); \tag{2}$$

$$\text{where } \boldsymbol{\Lambda}^{\succeq}_{\xi_T} := \left\{ \boldsymbol{\lambda} \in \Lambda : \inf_{\mathbb{Q} \in \mathcal{M}_p(\mathbb{P}_T, \xi_T)} D(v, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbb{Q}) > 0,\ \forall v \in \mathcal{U}_2^\star - \{0\} \right\}. \tag{3}$$

Lipschitz coefficient unitarity holds for example in the case of portfolio choice via maximization of expected return; then u is the identity. More generally, given that utility rescaling does not affect preferences and optimal choice, if u is non-trivial and has a bounded derivative, then the Lipschitz coefficient can be always be "chosen" equal to one. Hence this restriction on the properties of the objective is in many cases empirically innocuous.

The regularized version of the objective function is then equal, due to duality, to a robust expected value of $u(\cdot'\boldsymbol{x})$; actually this equals the most conservative expectation over the distributions inside the ball centered at the empirical distribution, with radius formed by the Lagrange multiplier.

$\mathcal{E}_2$ implies then that the regularization also permeates the totality of the non-trivial SD conditions. It implies that $\boldsymbol{\lambda}$ dominates $\boldsymbol{\tau}$ only if dominance holds w.r.t. the SD conditions formed by every distribution in the above-mentioned ball; regularization of the objective also implies conservativeness in the formulation of the SD conditions, at least for every sample realization associated with the particular event. But how is this related to the probability of FD classifications?

Non-asymptotic bounds for the false dominance probability

For $\tau > 0$ let $h(\tau) := \frac{1+\ln \mathbb{E}_{\mathbb{P}}[\exp(\tau \|\boldsymbol{x}\|_2^2)]}{\tau}$, and for $m > 2$ let $C(d) := 2 \times 3^{d - \log_3(d)} \mathbb{I}(d - \log_3(d) < \frac{\log_3(T)}{2}) + 4\mathbb{E}_{\mathbb{P}}[\|\boldsymbol{x}\|_2^m]\mathbb{I}(d - \log_3(d) \geq \frac{\log_3(T)}{2})$, where $\mathbb{I}$ denotes the indicator function. Also, $d(p) := d^l$, $l := \max(\frac{1}{2} - \frac{1}{p}, \frac{1}{p} - \frac{1}{2}) = \begin{cases} \frac{1}{p} - \frac{1}{2}, & p \leq 2 \\ \frac{1}{2} - \frac{1}{p}, & p > 2. \end{cases}$ The following result is then obtained:

Proposition 2 (False Dominance Classification)**.** *Suppose that $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is iid, that $d>2$, that for some $\tau>0$, $\mathbb{E}_{\mathbb{P}}(\exp\left(\tau\left\|\boldsymbol{x}\right\|_2^2\right))<+\infty$, and that u has unital Lipschitz coefficient. Then, for any $T\geq 1$, and if $\xi_T>2C(d)d(p)\sup_{\boldsymbol{\Lambda}}\left\|\boldsymbol{\lambda}\right\|_2 T^{-\frac{1}{d}}$*

$$\mathbb{P}(\boldsymbol{\Lambda}^{\succeq}_{\xi_T}-\boldsymbol{\Lambda}^{\succeq}_{\mathbb{P}}\neq\emptyset\,/\,\mathcal{E}_2)\leq\exp\left(-\frac{\left(1-C(d)T^{-\frac{d+1}{d}}\right)^2 T\xi_T^2}{8d^2\left(p\right)\inf_{\tau>0}h^2(\tau)\sup_{\boldsymbol{\Lambda}}\left\|\boldsymbol{\lambda}\right\|_2^2}\right). \tag{4}$$

Consequently, if the event $\mathcal{E}_2$ holds w.h.p. and $T\xi_T^2\to\infty$, then the probability of FD classification for the elements of $\boldsymbol{\Lambda}^{\succeq}_{\xi_T}$ converges to zero.

The result relies first on the iidness of the sample; it can be extended to m-dependent processes as well as to a class of Markov processes (see Boissard (2011) (1)). It also relies on the existence of some square-exponential moment for $\left\|\boldsymbol{X}_0\right\|_2^2$. This is equivalent to the existence of the moment generating function of $\left\|\boldsymbol{X}_0\right\|_2^2$ in a neighborhood of zero, a condition that fails whenever $\left\|\boldsymbol{X}_0\right\|$ follows a distribution with the right-tail behavior of the log-normal distribution. The exponential moment existence holds whenever X is bounded, or more generally whenever its squared elements follow sub-Gaussian distributions (see indicatively Chapter 2 of Vershynin (2018) (7)). The maximal moment parameter τ can be estimated via the ratio $\frac{(\kappa+1)\mathbb{E}_{\mathbb{P}_T}\left(\left\|\boldsymbol{X}_0\right\|_2^{2\kappa}\right)}{\mathbb{E}_{\mathbb{P}_T}\left(\left\|\boldsymbol{X}_0\right\|_2^{2\kappa+2}\right)}$, due to the power series representation of the exponential moment and the properties of the ratio test for real series. Given this the optimization resulting to $\inf_{\tau>0}h^2(\tau)$ can be empirically approximated. The choice of some non-optimal τ can also be considered at the cost of a potentially less efficient probability bound, and a larger regularization parameter.

The probability bound declines exponentially fast in $T\xi_T^2$, and holds for all T as long as the regularization parameter dominates a sequence of order $\exp(-\frac{\ln T}{d})$; this declines slowly when the base asset dimensionality is large. This low rate of asymptotic negligibility for the multiplier can be circumvented at either the cost of some positive large multiplicative constant in front of the probability bound, or at the cost that the results hold eventually for large enough T that also depends on the multiplier (see for example Bolley et al. (2007) (2)).

The result says that under $\mathcal{E}_2$, the probability that there exist empirically enhanced portfolios that are non dominant in the population, is bounded above by the exponential $\exp\left(-\frac{\left(1-C(d)T^{-\frac{d+1}{d}}\right)^2 T\xi_T^2}{8d^2(p)\inf_{\tau>0}h^2(\tau)\sup_{\boldsymbol{\Lambda}}\left\|\boldsymbol{\lambda}\right\|_2^2}\right)$. The bound depends on the regularization coefficient, the base assets dimensionality, the size of the portfolio space, through $d\left(p\right)$ on the choice of the ℓ_p norm, and the squared exponential moment parameter; e.g, if $d-\log_3(d)\geq\frac{\log_3(T)}{2}$ and $\xi_T=2cC(d)d(p)\sup_{\boldsymbol{\Lambda}}\left\|\boldsymbol{\lambda}\right\|_2 T^{-\frac{1}{d}}$, for $c>\max(1,\frac{1}{8d(p)\inf_{\tau>0}h(\tau)})$ and p slightly less than 2, then the result implies that the probability of FD error is eventually bounded above by $\exp(-c^{\star}T^{1-\frac{2}{d}})$ for some (estimable) positive constant $c^{\star}$.

Thus, for a given significance level $\alpha \in (0,1)$, and if $T \geq (-\frac{\ln(\alpha)}{c^{\star}})^{\frac{d}{d-2}}$, the probability of FD is thus bounded above by α. The same upper bound on the probability of FD holds whenever the regularization parameter is greater than the maximum between $\frac{\sqrt{-8\ln(\alpha)}d(p)\inf_{\tau>0}h(\tau)}{\left|1-C(d)T^{-\frac{d+1}{d}}\right|}\sup_{\mathbf{\Lambda}}\|\boldsymbol{\lambda}\|_2 T^{-\frac{1}{2}}$ and $2C(d)d(p)\sup_{\mathbf{\Lambda}}\|\boldsymbol{\lambda}\|_2 T^{-\frac{1}{d}}$.

Hence, through the inequality (4), and at least in the present restrictive framework, the particular regularization of the SD enhanced portfolio optimization problem provides with possibilities for control of the FD decision error, when either the regularization parameter is conveniently selected, and/or the sample size is large enough. This is an initial theoretical justification of the heuristic mentioned in the introduction.

Discussion

The proof of Proposition 2 crucially depends on an inequality of the form

$$\mathbb{P}\left(\mathcal{W}\left(\mathbb{P}_T,\mathbb{P};\frac{1}{2}\right) > t^{\star} + \mathbb{E}\left(\mathcal{W}\left(\mathbb{P}_T,\mathbb{P};\frac{1}{2}\right)\right)\right) \leq \exp\left(\frac{-Tt^{\star 2}}{2\inf_{\tau>0}h^2(\tau)}\right).$$

This is an example of a concentration inequality; it provides, a hopefully tight, non-asymptotic bound on the probability that the positive random variable $\mathcal{W}\left(\mathbb{P}_T,\mathbb{P};\frac{1}{2}\right)$ exceeds its expected value. Such inequalities are very useful since among others-as the aforementioned proposition suggests-they could provide with non-asymptotic control of the risk associated with procedures of statistical inference.

There is a thriving literature in mathematical analysis, geometry and probability theory (see for example Ohta and Takatsu (2011) (6) and the references therein), that surprisingly connects the validity of concentration inequalities with strong convexity properties of the entropy of the probability distributions at hand, information inequalities relating entropy with the Wasserstein distance, and generalized concepts of curvature related to the geometry of the spaces that shelter those distributions. The subsequent posts will investigate whether suchlike connections can be used to extend results like (4) to non iid frameworks, and can furthermore provide non-asymptotic enhancements to econometrically relevant statistical procedures like the predictive ability tests considered in a previous post. To be continued!

References

[1] Boissard, E., 2011. Simple bounds for the convergence of empirical and occupation measures in 1-Wasserstein distance. Electronic Journal of Probability, 16, 2296-2333.

[2] Bolley, F., Guillin, A. and Villani, C., 2007. Quantitative concentration inequalities for

empirical measures on non-compact spaces. Probability Theory and Related Fields, 137(3), pp.541-593.

[3] Constantinides, G.M., Czerwonko, M. and Perrakis, S., 2020. Mispriced index option portfolios. Financial Management, 49(2), pp.297-330.

[4] Galichon, A., 2018. Optimal transport methods in economics. Princeton University Press.

[5] Gao, R., Chen, X. and Kleywegt, A.J., 2017. Distributional robustness and regularization in statistical learning. arXiv preprint arXiv:1712.06050.

[6] Ohta, S.I. and Takatsu, A., 2011. Displacement convexity of generalized relative entropies. Advances in Mathematics, 228(3), pp.1742-1787.

[7] Vershynin, R, 2018, High-dimensional probability: An introduction with applications in data science (Vol. 47). Cambridge University Press.