Queen's Economics Department Working Paper No. 1533

Norm Constrained Empirical Portfolio Optimization with Stochastic Dominance: Robust Optimization Non-Asymptotics

Stelios Arvanitis
Department of Economics, AUEB

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

2-2025

Norm Constrained Empirical Portfolio Optimization with Stochastic Dominance: Robust Optimization Non-Asymptotics

*Stelios Arvanitis**

Abstract

The present note provides an initial theoretical explanation of the way norm regularizations may provide a means of controlling the non-asymptotic probability of False Dominance classification for empirically optimal portfolios satisfying empirical Stochastic Dominance restrictions in an iid setting. It does so via a dual characterization of the norm-constrained problem, as a problem of Distributional Robust Optimization. This enables the use of concentration inequalities involving the Wasserstein distance from the empirical distribution, to obtain an upper bound for the non-asymptotic probability of False Dominance classification. This leads to information about the minimal sample size required for this probability to be dominated by a predetermined significance level.

JEL Codes: C44, C58, G11.
MSC2020: 62C, 62P20, 62E17.

Keywords:Portfolio optimization, Stochastic dominance, ℓ_p regularization, Wasserstein distance, Distributionally robust optimization, Concentration inequality, False dominance classification.

1 Introduction

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

*Department of Economics, Queen's University, Ontario, Canada, and Department of Economics, Athens University of Economics and Business, Greece. Email: stelios.a@queensu.ca

A recent literature in operations research and econometrics have addressed the analytical challenges for numerical optimization and statistical inference, largely making SD portfolio optimization empirically applicable. Indicative such applications include [1]; [2]; [3]; [4]. There 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 a 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 portfolios considered. In some of the applications mentioned above, 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. The studies [2], [3], and [4] demonstrate enhancements in both in-sample and out-of-sample performance for the optimized portfolios. This finding aligns with similar observations noted within the portfolio selection literature, which do not necessarily lie inside the stochastic dominance framework; refer to [5], [6], and [7] for further examples.

The present note provides an initial theoretical explanation of the way suchlike restrictions may provide a means of controlling the fixed sample size probability of FD classification in an iid setting. It does so first, via the conditional on particular events characterization of the norm-constrained problem, as a problem of Distributional Robust Optimization (DRO). There strong convex duality results (see [8]) enable the representation of the original problem as a problem of conservative optimization over a Wasserstein ball centered at the empirical distribution, as long as the criterion has a Lipschitz continuity property. Second, using this characterization along with concentration inequalities involving the Wasserstein distance from the empirical distribution, an upper bound for the non-asymptotic probability of FD classification is obtained, leading among others to considerations about the minimal sample size required for this probability to be dominated by a predetermined significance level.

The structure of the note is the following: the second section analyzes the SD framework for portfolio optimization and its' regularized formulation. The third section derives the DRO characterization, the non-asymptotic bounds for the probability of FD, and briefly discusses some paths for future research.

2 Stochastic dominance and portfolio optimization framework

$(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a process with values in some subset of $\mathbb{R}^d$. The random vector $\boldsymbol{X}_t$ represents the one period returns of d financial assets, $\mathbb{P}$ denotes their latent time-invariant joint distribution, and $\mathcal{X} \subset \mathbb{R}^d$ is the pointwise bounded from below support of $\mathbb{P}$. 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.

A portfolio on $\boldsymbol{X}_0$ is any element of the dual space of $\mathbb{R}^d$; 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. These populate the closed (in the topology of uniform convergence on compacta), convex set $\mathcal{U}_2$ while $\mathcal{U}_2^{\star}$ denotes the set of Russell-Seo utilities-see [9], i.e. the utilities of the form $u(x) = -(z-x)_+$, $z \in \mathcal{X}$, those constitute the extreme points of $\mathcal{U}_2$; $\mathcal{U}_2^{\star} - \{0\}$ does not contain the constant utility at zero-corresponding to the threshold that equals the pointwise infimum of $\mathcal{X}$. $\mathcal{U}_2^{=}$ is the set of constant utilities, subsequently, $\mathcal{U}_2 - \mathcal{U}_2^{=}$ analogously denotes $\mathcal{U}_2$ without constant utilities.

The analysis involves a set of portfolios $\boldsymbol{\Lambda} \subseteq \mathbb{R}^d$. It is considered convex and compact. 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 $D(u, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbb{P}) := \mathbb{E}_{\mathbb{P}}(u(\boldsymbol{\lambda}'\boldsymbol{X}_0)) - \mathbb{E}_{\mathbb{P}}(u(\boldsymbol{\tau}'\boldsymbol{X}_0)) \geq 0$, $\forall u \in \mathcal{U}_2$-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 $\boldsymbol{\tau}$; $\boldsymbol{\lambda} \underset{\mathbb{P},2}{\succeq} \boldsymbol{\tau}$. The definition remains invariant if $\mathcal{U}_2$ is replaced by $\mathcal{U}_2^{\star}$-see [9]. Furthermore, $\boldsymbol{\Lambda}_{\mathbb{P}}^{\succeq} := \left\{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}; \boldsymbol{\lambda} \underset{\mathbb{P},2}{\succeq} \boldsymbol{\tau}\right\}$ 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, 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 $\mathbb{P}_T \rightsquigarrow \mathbb{P}$ where $\rightsquigarrow$ denotes convergence in distribution and the existence of some $\epsilon > 0$ such that $\sup_T \mathbb{E}_{\mathbb{P}_T}(\|\boldsymbol{X}_0\|_2^{1+\epsilon}) < +\infty$, where $\|\cdot\|_2$ denotes the Euclidean norm. This is due to that $\mathcal{U}_2^\star$ is uniformly Lipschitz, and due to uniform integrability. The weak convergence is easily establishable in contexts of stationarity and ergodicity for $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$; boundedness for the $1+\epsilon$ moments of the normed process would follow in those stationary and ergodic frameworks along which a financially plausible assumption of $\mathbb{E}_{\mathbb{P}}(\|\boldsymbol{X}_0\|_2^{1+\epsilon}) < +\infty$-see for example Paragraph 3.1 and Theorem 1 of [10]. 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 trivial in cases where a constant criterion is involved-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.

2.1 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 $\max_{i=1,\dots,d} |\lambda_i - \tau_i|$ 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 multiplier. Existence of the regularized optimization portfolio is ensured by the concavity and continuity of u, the linearity of $\mathbb{E}_{\mathbb{P}_T}$, the convexity and continuity of $\|\cdot\|$, and the convexity of the set $\boldsymbol{\Lambda}_{\mathbb{P}_T}^{\succeq}$, which holds due to the concavity of the members of $\mathcal{U}_2$, and the Measurable Maximum Theorem-see Theorem 18.19.3 in [11]. Optimal selection of the multiplier 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.

It is possible that $\boldsymbol{\Lambda}$ already includes (explicit or implicit) pre-existing norm constraints; then the analysis studies the effect on the probability of FD of tightening of the existing norm constraints.

Algorithmically, given the $\mathcal{U}_2^{\star}$ formulation of the dominance relation by [9], and if u is the identity, something that corresponds to portfolio mean maximization under dominance constraints, a usual practical specification of Problem (1) is:

$$\begin{aligned} \max_{\boldsymbol{\lambda}\in\Lambda} \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\lambda}'\boldsymbol{X}_t - \xi_T \|\boldsymbol{\lambda} - \boldsymbol{\tau}\|_p, \\ \text{s.t. } \frac{1}{T}\sum_{t=1}^{T}(z - \boldsymbol{\tau}'\boldsymbol{X}_t)_+ - \frac{1}{T}\sum_{t=1}^{T}(z - \boldsymbol{\lambda}'\boldsymbol{X}_t)_+ \geq 0,\ \forall z \in \mathcal{X}^{\star}, \end{aligned} \tag{2}$$

where now $\mathcal{X}^{\star}$ is a finite discretization of the support $\mathcal{X}$. If Λ is convex, and $p = 1$, or $p = \infty$, then Problem (2) is reducible to an $O(|\mathcal{X}^{\star}| \times T + d)$-LP formulation via the linear relaxations technology of [12]. Whenever $1 < p < +\infty$, the aforementioned relaxations can be combined with piecewise linear approximations of the ℓ_p norm, in which case the problem contains $O(|\mathcal{X}^{\star}| \times T + d \times S)$ variables, where S denotes the number of segments that support the linear pieces. For a general concave u, the problem can likewise be formulated as a convex optimization problem with linearizable restrictions-see [13] for a survey of LP relaxations of stochastic dominance problems. In any case, the solution produced by the LP relaxations, is characterized as any element of $\arg\max_{\boldsymbol{\Lambda}^{\star}_{\mathbb{P}_T^{\succeq}}} (\mathbb{E}_{\mathbb{P}_T}(u(\boldsymbol{\lambda}'\boldsymbol{x})) - \xi_T \|\boldsymbol{\lambda} - \boldsymbol{\tau}\|_p)$, where now $\boldsymbol{\Lambda}^{\star}_{\mathbb{P}_T^{\succeq}}$ is the set of $\boldsymbol{\lambda}$ that satisfy the resulting $O(|\mathcal{X}^{\star}| \times T)$ relaxed empirical dominance inequalities.

3 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.

In the first part of this section the regularized problem is translated to a problem of distributionally robust optimization (DRO) using a dual formulation of regularized problems involving Lipschitz criteria to problems of conservative optimization in Wasserstein neighborhoods of $\mathbb{P}_T$. There, the Lagrange multiplier ξ_T plays an important role in the DRO formulation below, as it determines the radius of pessimism and at least locally the slacks for the dominance constraints.

In the second part this conservative representation is exploited in order to non-asymptotically bound the probability of FD characterization via concentration inequalities.

Some further notation will be also useful: for $\mathbb{Q}$ an arbitrary distribution having finite first moment on $\mathbb{R}^d$, and $q(p)$ such that $\frac{1}{p}+\frac{1}{q}=1$, the first Wasserstein distance between $\mathbb{Q}$ and the empirical distribution $\mathbb{P}_T$ is defined by $\mathcal{W}(\mathbb{P}_T,\mathbb{Q};p):=\min_{\gamma\in\Gamma(\mathbb{P}_T,\mathbb{Q})}\int_{\mathbb{R}^d\times\mathbb{R}^d}\left\|z-z^\star\right\|_q d\gamma(z,z^\star)$, where $\Gamma(\mathbb{P}_T,\mathbb{Q})$ denotes the set of Borel probability distributions on $\mathbb{R}^d\times\mathbb{R}^d$ that have respective "marginals" $\mathbb{P}_T$, $\mathbb{Q}$, and also have finite q moments (see [8]). $\mathcal{W}$ metrizes weak convergence (see [14]). For $\epsilon>0$, $\mathcal{M}_p(\mathbb{P}_T,\epsilon):=\{\mathbb{Q}:\mathcal{W}(\mathbb{P}_T,\mathbb{Q};p)\leq\epsilon\}$ is the Wasserstein closed ball centered at $\mathbb{P}_T$ with radius ϵ. Also, $d(p):=d^l$, $l:=\max(\frac{1}{2}-\frac{1}{p},\frac{1}{p}-\frac{1}{2})=\begin{cases}\frac{1}{p}-\frac{1}{2}, & p\leq 2\\ \frac{1}{2}-\frac{1}{p}, & p>2.\end{cases}$

3.1 DRO 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 [8]).

The event $\mathcal{E}_1:=\left\{\exists\boldsymbol{\lambda}(u,\mathbb{P}_T,p,\xi_T):\inf_{\mathbb{Q}\in\mathcal{M}_p(\mathbb{P}_T,\xi_T)}D(u,\boldsymbol{\tau},\boldsymbol{\lambda}(u,\mathbb{P}_T,p,\xi_T),\mathbb{Q})>0\right\}$ characterizes the samples for which the empirical regularized program has non trivial solutions. Furthermore, for every sample realization appearing inside $\mathcal{E}_2:=\left\{\forall v\in\mathcal{U}_2^\star-\{0\},\forall\boldsymbol{\lambda}(v,\mathbb{P}_T,p,\xi_T):\inf_{\mathbb{Q}\in\mathcal{M}_p(\mathbb{P}_T,\xi_T)}D(v,\boldsymbol{\tau},\boldsymbol{\lambda}(v,\mathbb{P}_T,p,\xi_T),\mathbb{Q})>0\right\}$, every choice of the objective utility results to non-trivial empirical solutions. It is noted that $\mathcal{E}_2$ is a subset of $\mathcal{E}_1$ due to the Russell-Seo utility representation-see [9]. Then:

Proposition 1. *(Distributional Robustness). Suppose that u has a unital Lipschitz coefficient. Then:*

1. *Given $\mathcal{E}_1$, there exists some non-empty open subset $\mathcal{U}_2(u)$ of $\mathcal{U}_2$, for which Problem (1) is equivalent to:*

$$\max_{\mathbf{\Lambda}^{\succeq}_{\mathbb{P}_T}\cap\mathbf{\Lambda}^{\succeq}_{u}} \inf_{\mathbb{Q}\in\mathcal{M}_p(\mathbb{P}_T,\xi_T)} \mathbb{E}_{\mathbb{Q}}(u(\boldsymbol{\lambda}'\boldsymbol{x})); \tag{3}$$

$$\mathbf{\Lambda}^{\succeq}_{u} := \left\{\boldsymbol{\lambda}\in\Lambda : \inf_{\mathbb{Q}\in\mathcal{M}_p(\mathbb{P}_T,\xi_T)} D(v,\boldsymbol{\tau},\boldsymbol{\lambda},\mathbb{Q}) > 0,\ \forall v\in\mathcal{U}_2(u)\right\}. \tag{4}$$

2. *Given $\mathcal{E}_2$, Problem (1) is equivalent to:*

$$\sup_{\mathbf{\Lambda}^{\succeq}_{\xi_T}} \inf_{\mathbb{Q}\in\mathcal{M}_p(\mathbb{P}_T,\xi_T)} \mathbb{E}_{\mathbb{Q}}(u(\boldsymbol{\lambda}'\boldsymbol{x})); \tag{5}$$

where $$\mathbf{\Lambda}^{\succeq}_{\xi_T} := \left\{\boldsymbol{\lambda}\in\Lambda : \inf_{\mathbb{Q}\in\mathcal{M}_p(\mathbb{P}_T,\xi_T)} D(v,\boldsymbol{\tau},\boldsymbol{\lambda},\mathbb{Q}) > 0,\ \forall v\in\mathcal{U}_2^{\star}-\{0\}\right\}. \tag{6}$$

Proof. Theorem 1 of [8] implies that

$$\begin{array}{c} G_{\mathbb{P}_T}(\boldsymbol{\tau}+(\boldsymbol{\lambda}-\boldsymbol{\tau})) - \xi_T(c)\left\|\boldsymbol{\lambda}-\boldsymbol{\tau}\right\|_p = \inf_{\mathbb{Q}\in\mathcal{M}_p(\mathbb{P}_T,\xi_T)} G_{\mathbb{Q}}(\boldsymbol{\tau}+(\boldsymbol{\lambda}-\boldsymbol{\tau})) \\ D(u,\boldsymbol{\tau},\boldsymbol{\tau}+(\boldsymbol{\lambda}-\boldsymbol{\tau}),\mathbb{P}_T) = \inf_{\mathbb{Q}\in\mathcal{M}_p(\mathbb{P}_T,\xi_T)} D(u,\boldsymbol{\tau},\boldsymbol{\tau}+(\boldsymbol{\lambda}-\boldsymbol{\tau}),\mathbb{Q}) + \xi_T\left\|\boldsymbol{\lambda}-\boldsymbol{\tau}\right\|_p \end{array},$$

for $G_{\mathbb{P}}(\boldsymbol{\lambda}) := \mathbb{E}_{\mathbb{P}}(u(\boldsymbol{\lambda}'\boldsymbol{X}))$. This and translation in $\mathbb{R}^d$ by $\boldsymbol{\tau}$, directly imply that Problem (1) is equivalent to $\max_{\mathbf{\Lambda}} \inf_{\mathbb{Q}\in\mathcal{M}_p(\mathbb{P}_T,\xi_T)} G_{\mathbb{Q}}(\boldsymbol{\lambda})$. Using a scaling and translation argument that is allowed for von Neyman-Morgenstern preferences, $\mathcal{U}_2$ can be chosen uniformly bounded due to the uniform Lipschitz property of $\mathcal{U}_2^{\star}$ and the Russell-Seo representation-see [9]. This and Lemma 2.7.5 of [15] along with the compactness of the support, imply that $\mathcal{U}_2$ is totally bounded, and thereby equicontinuous. This and uniform integrability then imply that $D(u,\boldsymbol{\tau},\boldsymbol{\lambda},\mathbb{P}) - \xi_T\left\|\boldsymbol{\lambda}-\boldsymbol{\tau}\right\|_p$ is jointly continuous in $(u,\boldsymbol{\lambda})$. Since there exists a non trivial solution to (1), it must satisfy $D(u,\boldsymbol{\tau},\boldsymbol{\lambda},\mathbb{P}_T) \geq \xi_T\left\|\boldsymbol{\lambda}-\boldsymbol{\tau}\right\|_p$. This then implies the existence of $\mathcal{U}_2(u)$. $\mathcal{U}_2(u)$ is open as a union of open sets; it is obtained as the inverse image of $(0,+\infty)$ w.r.t. the regularized criterion evaluated at the solution, on $\mathcal{U}_2$. Finally, similarly to Theorem 1 of [10], for the portfolio defined as $\boldsymbol{\gamma} = \boldsymbol{\gamma}_{(\mathcal{U},\boldsymbol{\tau})}((1-\delta^{\star})w_u + \delta^{\star}w)$, with $\boldsymbol{\gamma}_{(\mathcal{U},\boldsymbol{\tau})}(w) := \int_{\mathcal{U}_2^{\star}-\{0\}} \boldsymbol{\lambda}(u,\mathbb{P}_T,p,\xi_T)dw(v)$, where w lies in the set of non-degenerate Borel measures on $\mathcal{U}_2$, and w_u the degenerate measure at u, $\delta^{\star}\in(0,1)$, we have that $\boldsymbol{\gamma}\in\mathbf{\Lambda}^{\succeq}_{\xi_T}$ due to the concavity of the utilities involved; then by the definition of the portfolios involved, and due the Lipschitz continuity of $D(u, \boldsymbol{\tau}, \cdot, \mathbb{P}_T)$ and of the ℓ_p norm, setting $\boldsymbol{\lambda}_T := \boldsymbol{\lambda}(u, \mathbb{P}_T, p, \xi_T)$ it is obtained that

$$\begin{aligned} 0 &\leq \mathbb{E}_{\mathbb{P}_T}(u(\boldsymbol{\lambda}_T' - \xi_T \left\|\boldsymbol{\lambda}_T - \boldsymbol{\tau}\right\|_p - \mathbb{E}_{\mathbb{P}_T}(u(\boldsymbol{\gamma}'\boldsymbol{X})) + \xi_T \left\|\boldsymbol{\gamma} - \boldsymbol{\tau}\right\|_p \\ &\leq \delta^\star(\sup_{\mathbb{Q}\in\mathcal{M}_p(\mathbb{P}_T,\xi_T)} \mathbb{E}_{\mathbb{Q}}(\left\|\boldsymbol{x}\right\|_2) + \xi_T d^{\frac{1}{p}-\frac{1}{2}})\mathrm{diam}(\boldsymbol{\Lambda}), \end{aligned}$$

and (5) holds since $\delta^\star$ can be chosen arbitrarily small. □

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 equal to one. The regularized version of the objective function is equal, due to duality, to a robust expected value of $u(\cdot'\boldsymbol{x})$; actually this is the most conservative expectation over the Wasserstein ball centered at the ecdf, with radius formed by the Lagrange multiplier. The dual representation of $\mathcal{W}$ (see [16]) implies that $\mathcal{M}_p(\mathbb{P}_T, \xi_T)$ is convex; then Sion's Minimax Theorem-see [17]-implies that $\max_{\boldsymbol{\Lambda}^{\succeq}_{\mathbb{P}_T}} \inf_{\mathbb{Q}\in\mathcal{M}_p} \mathbb{E}_{\mathbb{Q}}(u(\boldsymbol{\lambda}'\boldsymbol{X})) = \inf_{\mathbb{Q}\in\mathcal{M}_p} \max_{\boldsymbol{\Lambda}^{\succeq}_{\mathbb{P}_T}} \mathbb{E}_{\mathbb{Q}}(u(\boldsymbol{\lambda}'\boldsymbol{X}))$. This conforms to the solution characterization as the most conservative (over the aforementioned ball) maximizer of expected utility that stochastically dominates the benchmark.

Due to the equi-continuity properties of $\mathcal{U}_2$, and if there exists an empirically optimal portfolio that strictly dominates the benchmark, i.e. $\mathcal{E}_1$ holds, the empirical problem has a representation where the regularization term, permeates as a positive slack to a neighborhood of SD conditions around $D(u, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbb{P}_T)$; each of these conditions then, by the same duality property, has a conservative characterization given in (4). Hence, regularizing the objective implies a local regularization for the dominance conditions. Under the stronger $\mathcal{E}_2$, the representation uses the regularization term as a positive slack on the totality of non-trivial SD conditions. This implies that there exist elements of $\boldsymbol{\lambda}(u, \mathbb{P}_T, p, \xi_T)$ that satisfy enhanced versions of the SD inequalities, resulting into strong properties regarding the probability of FD classifications, as the result in the following section reveals. It is noted that due to the Russell-Seo representation-see [9], the linearity of D w.r.t. u and the properties of the infimum, under $\mathcal{E}_2$, the regularization permeates every SD condition involving a non-constant utility.

3.2 Non-asymptotic bounds for the False Dominance probability

In order to derive non asymptotic bounds for the probability of selecting a portfolio that does not dominate the benchmark in the population (FD), the event

$\boldsymbol{\Lambda}^{\succeq}_{\mathbb{P}_T} - \boldsymbol{\Lambda}^{\succeq}_{\mathbb{P}} \neq \emptyset$ is also utilized: it corresponds to the samples for which FD classification actually occurs, when regularization is not used inside the SD inequalities. Furthermore, 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. Then the following result is obtained via the use of concentration inequalities involving the Wasserstein distance from the empirical distribution:

Proposition 2 (False Dominance Classification and Opportunity Loss)**.** *Suppose that $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is iid, that $d > 2$, that for some $\tau > 0$, $\mathbb{E}_{\mathbb{P}}(\exp\left(\tau \|\boldsymbol{x}\|_2^2\right)) < +\infty$, and that u has unital Lipschitz coefficient. If $\boldsymbol{\lambda}_T := \boldsymbol{\lambda}(u, \mathbb{P}_T, p, \xi_T)$, and $\mathcal{E} := \boldsymbol{\Lambda}^{\succeq}_{\xi_T} - \boldsymbol{\Lambda}^{\succeq}_{\mathbb{P}} \neq \emptyset \cap \mathcal{E}_2$, then,*

1. *for any $T \geq 1$, and if $\xi_T \inf_{\mathcal{E}} \|\boldsymbol{\lambda}_T - \boldsymbol{\tau}\|_p > 2d(p) \sup_{\boldsymbol{\lambda}} \|\boldsymbol{\lambda}\|_2 C(d)T^{-\frac{1}{d}}$,*

$$\mathbb{P}(\mathcal{E}) \leq \exp\left(-\frac{\left(1 - C(d)T^{-\frac{d+1}{d}}\right)^2 TM^2(\xi_T, \Lambda, p, \mathcal{U}_2^\star, \mathcal{E})}{2\inf_{\tau>0} h^2(\tau)}\right), \tag{7}$$

   where $M(\xi_T, \Lambda, p, \mathcal{U}_2^\star, \mathcal{E}) := \frac{\xi_T \inf_{\mathcal{E}} \|\boldsymbol{\lambda}_T - \boldsymbol{\tau}\|_p - \sup_{\mathcal{E}} \inf_{\mathcal{U}_2^\star} D(v, \boldsymbol{\tau}, \boldsymbol{\lambda}_T, \mathbb{P})}{2d(p)\sup_{\boldsymbol{\lambda}}\|\boldsymbol{\lambda}\|_2}$.

   Consequently, if the event $\mathcal{E}_2$ holds w.h.p. and $T\xi_T^2 \inf_{\mathcal{E}} \|\boldsymbol{\lambda}_T - \boldsymbol{\tau}\|_p^2 \to \infty$, then the probability of FD classification for any with asymptotically positive probability $\boldsymbol{\lambda}_T \in \boldsymbol{\Lambda}^{\succeq}_{\xi_T} - \boldsymbol{\Lambda}_{\mathbb{P}}$, converges to zero.

2. *There exists some $T_0 \geq 1$ such that for any $T \geq T_0$, if the regularization coefficient satisfies $\xi_T \inf_{\mathcal{E}} \|\boldsymbol{\lambda}_T - \boldsymbol{\tau}\|_p > 2d(p) \sup_{\boldsymbol{\Lambda}} \|\boldsymbol{\lambda}\|_2 \max\left\{C(d)T^{-\frac{1}{d}}, \sqrt{2}\inf_{\tau>0} h(\tau)\sqrt{\frac{\ln T}{T}}\right\}$, then,*

$$\mathbb{P}(\boldsymbol{\Lambda}^{\succeq}_{\mathbb{P}_T} - \boldsymbol{\Lambda}^{\succeq}_{\mathbb{P}} \neq \emptyset) > \mathbb{P}(\mathcal{E}). \tag{8}$$

3. *Finally, if $\boldsymbol{\lambda}^\star$ denotes the optimal solution to the population problem, then for any $T \geq 1$, and if $\xi_T > C(d)T^{-\frac{1}{d}}$,*

$$\begin{aligned}\mathbb{P}(\left|\mathbb{E}_{\mathbb{P}_T}(u(\boldsymbol{\lambda}'_T\boldsymbol{x})) - \mathbb{E}_{\mathbb{P}}(u\left(\boldsymbol{\lambda}^{\star\prime}\boldsymbol{x}\right))\right| &> d(p)\mathbb{E}_{\mathbb{P}_T}\|\boldsymbol{X}_0\|_2 \|\boldsymbol{\lambda}_T - \boldsymbol{\lambda}^\star\|_p + \|\boldsymbol{\lambda}^\star\|_2 \xi_T) \\ &\leq \exp\left(-\frac{\left(1 - C(d)T^{-\frac{d+1}{d}}\right)^2 T\xi_T^2}{2\inf_{\tau>0} h^2(\tau)}\right).\end{aligned} \tag{9}$$

Proof. Given $\mathcal{E}_2$, the event $\Lambda^{\succeq}_{\xi_T} - \Lambda^{\succeq}_{\mathbb{P}} \neq \emptyset$ is equivalent to the inequalities' system comprised of $\inf_{\mathcal{U}_2^\star} \inf_{\mathbb{P}\in\mathcal{M}_p(\mathbb{P}_T,\xi_T)} D(u,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P}) \geq 0$ and $\inf_{\mathcal{U}_2^\star} D(u,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P}) < 0$. Due to the Kantorovich-Rubinstein representation of $\mathcal{W}$ (see [16]), Proposition 1.2, and the equivalence between the ℓ_p norms in $\mathbb{R}^d$,

$$
\begin{aligned}
&\mathbb{P}\left(\boldsymbol{\Lambda}^{\succeq}_{\xi_T} - \boldsymbol{\Lambda}^{\succeq}_{\mathbb{P}} \neq \emptyset \cap\ \mathcal{E}_2\right)\\
&\leq \mathbb{P}\left(\inf_{\mathcal{U}_2^\star}\inf_{\mathbb{Q}\in\mathcal{M}_p(\mathbb{P}_T,\xi_T)} D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{Q}) - \inf_{\mathcal{U}_2^\star} D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P}) \geq -\inf_{\mathcal{U}_2^\star} D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P}) \cap\ \mathcal{E}_2\right)\\
&\leq \mathbb{P}\left(\inf_{\mathcal{U}_2^\star}(D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P}_T) - \xi_T\|\boldsymbol{\lambda}_T-\boldsymbol{\tau}\|_p) - \inf_{\mathcal{U}_2^\star} D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P}) \geq -\inf_{\mathcal{U}_2^\star} D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P}) \cap\ \mathcal{E}_2\right)\\
&\leq \mathbb{P}\left(\inf_{\mathcal{U}_2^\star} D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P}_T) - \inf_{\mathcal{U}_2^\star} D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P}) \geq \xi_T\|\boldsymbol{\lambda}_T-\boldsymbol{\tau}\|_p - \inf_{\mathcal{U}_2^\star} D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P}) \cap\ \mathcal{E}_2\right)\\
&\leq \mathbb{P}\left(\sup_{\mathcal{U}_2^\star}|D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P}_T) - D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P})| \geq \xi_T\|\boldsymbol{\lambda}_T-\boldsymbol{\tau}\|_p - \inf_{\mathcal{U}_2^\star} D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P}) \cap\ \mathcal{E}_2\right)\\
&\leq \mathbb{P}\left(2\sup_{\boldsymbol{\Lambda}}\|\boldsymbol{\lambda}\|_2\, d^{\max\left(0,\frac{1}{p}-\frac{1}{2}\right)}\mathcal{W}\left(\mathbb{P}_T,\mathbb{P};p\right) \geq \xi_T\|\boldsymbol{\lambda}_T-\boldsymbol{\tau}\|_p - \inf_{\mathcal{U}_2^\star} D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P}) \cap\ \mathcal{E}_2\right)\\
&\leq \mathbb{P}\left(\mathcal{W}\left(\mathbb{P}_T,\mathbb{P};p\right) \geq \frac{\xi_T\|\boldsymbol{\lambda}_T-\boldsymbol{\tau}\|_p - \inf_{\mathcal{U}_2^\star} D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P})}{2d(p)\sup_{\boldsymbol{\lambda}}\|\boldsymbol{\lambda}\|_2} \cap\ \mathcal{E}_2\right)\\
&\leq \mathbb{P}\left(\mathcal{W}\left(\mathbb{P}_T,\mathbb{P};p\right) \geq \frac{\xi_T\inf_{\mathcal{E}}\|\boldsymbol{\lambda}_T-\boldsymbol{\tau}\|_p - \sup_{\mathcal{E}}\inf_{\mathcal{U}_2^\star} D(v,\boldsymbol{\tau},\boldsymbol{\lambda}_T,\mathbb{P})}{2d(p)\sup_{\boldsymbol{\lambda}}\|\boldsymbol{\lambda}\|_2}\right),
\end{aligned}
$$

where the final inequality in the previous display follows from the monotonicity of $\mathbb{P}$. Due to the Bobkov-Godge equivalence (see Theorem 1.3 of [18]-see also Corollary 2.4 and Particular Case 2.5 of [19]), and Theorem 1.1 of [20], the existence of the exponential squared moment for $\|\boldsymbol{X}_0\|_2$, and relation (14) of [21], for any $t^\star > 0$, we have that

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

Setting in the above $t^\star := t - \mathbb{E}\left(\mathcal{W}\left(\mathbb{P}_T,\mathbb{P};\frac{1}{2}\right)\right)$ for $t > \mathbb{E}\left(\mathcal{W}\left(\mathbb{P}_T,\mathbb{P};\frac{1}{2}\right)\right)$, and observing that $\left(t - \mathbb{E}\left(\mathcal{W}\left(\mathbb{P}_T,\mathbb{P};\frac{1}{2}\right)\right)\right) \geq \left(1 - C(d)T^{-\frac{d+1}{d}}\right)^2 t^2$, whenever $t > \mathbb{E}\left(\mathcal{W}\left(\mathbb{P}_T,\mathbb{P};\frac{1}{2}\right)\right)$, we obtain from the proof of Theorem 2.3 of [22] (where we identify 4 as an upper bound for the constant c_p), as well as from the proof of Theorem 3.1 of [22], where first, we identify $c \leq 1 - \frac{\log_3(d)}{d}$ due to the equivalence between the Euclidean and the max norm, and the form of the packing number for closed max-balls, and, second, let $m \to \infty$ in the case where $d - \log_3(d) < \log_3(T)$ due to the existence of the exponential squared moment, that if $t > C(d)T^{-\frac{1}{d}}$, then

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

Setting in the previous $t := M(\xi_T, \Lambda, p, \mathcal{U}_2^\star, \mathcal{E})$ and observing that when the event $\boldsymbol{\Lambda}_{\xi_T}^{\succeq} - \boldsymbol{\Lambda}_{\mathbb{P}}^{\succeq} \neq \emptyset \cap \mathcal{E}_2$ occurs, the non stochastic term $\sup_{\mathcal{E}} \inf_{\mathcal{U}_2^\star} D(v, \boldsymbol{\tau}, \boldsymbol{\lambda}_T, \mathbb{P})$ is negative, establishes (7).

Furthermore, due to Theorem 2 of [23] and the Berry-Esseen Theorem, and for any v, for which $D(v, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbb{P}) < 0$, we obtain that there exist constants $C_1, C_2, c > 0$, independent of T, and a positive asymptotically negligible sequence $\gamma(T)$ that satisfies $\sqrt{T}\gamma(T) \to \infty$, such that eventually,

$$\begin{gathered}\tfrac{C_1}{\sqrt{T}(1+T^{-1})} - \tfrac{C_2\gamma(T)}{T(1+T^{-1})} - o\left(T^{-\frac{1}{2}}\right) + \exp\left(-cTD^2(v, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbb{P})\right) \\ \leq \mathbb{P}\left(D(v, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbb{P}_T) - D(v, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbb{P}) \geq -D(v, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbb{P})\right) \leq \mathbb{P}\left(\boldsymbol{\Lambda}_{\mathbb{P}_T}^{\succeq} - \boldsymbol{\Lambda}_{\mathbb{P}}^{\succeq} \neq \emptyset\right).\end{gathered}$$

Then, if (7) holds, and since under $\mathcal{E}$ the $\sup_{\mathcal{E}} \inf_{\mathcal{U}_2^\star} D(v, \boldsymbol{\tau}, \boldsymbol{\lambda}_T, \mathbb{P})$ is negative, if furthermore $\xi_T \inf_{\mathcal{E}} \|\boldsymbol{\lambda}_T - \boldsymbol{\tau}\|_p > 2\sqrt{2}d(p) \inf_{\tau>0} h(\tau) \sup_{\boldsymbol{\Lambda}} \|\boldsymbol{\lambda}\|_2 \sqrt{\frac{\frac{1}{2}\ln T + \ln(1+T^{-1}) - \ln C_1}{T}}$, the second result in (8) follows. For the final result notice that due to the triangle inequality, the Lipschitz continuity property of v, the boundedness of the support, the Cauchy-Schwarz inequality, and the equivalence between the ℓ_p norms in $\mathbb{R}^d$,

$$\begin{gathered}\left|\mathbb{E}_{\mathbb{P}_T}\left(u\left(\boldsymbol{\lambda}_T' \boldsymbol{X}_0\right)\right) - \mathbb{E}_{\mathbb{P}}\left(u\left(\boldsymbol{\lambda}^{\star\prime} \boldsymbol{X}_0\right)\right)\right| \\ \leq \left|\mathbb{E}_{\mathbb{P}_T}\left(u\left(\boldsymbol{\lambda}_T' \boldsymbol{X}_0\right)\right) - \mathbb{E}_{\mathbb{P}_T}\left(u\left(\boldsymbol{\lambda}^{\star\prime} \boldsymbol{X}_0\right)\right)\right| + \left|\mathbb{E}_{\mathbb{P}_T}\left(u\left(\boldsymbol{\lambda}^{\star\prime} \boldsymbol{X}_0\right)\right) - \mathbb{E}_{\mathbb{P}}\left(u\left(\boldsymbol{\lambda}^{\star\prime} \boldsymbol{X}_0\right)\right)\right|,\end{gathered}$$

with

$$\left|\mathbb{E}_{\mathbb{P}_T}\left(v\left(\boldsymbol{\lambda}_T' \boldsymbol{X}_0\right)\right) - \mathbb{E}_{\mathbb{P}_T}\left(v\left(\boldsymbol{\lambda}^{\star\prime} \boldsymbol{X}_0\right)\right)\right| \leq d(p)\, \mathbb{E}_{\mathbb{P}_T}(\|\boldsymbol{X}_0\|_2)\, \|\boldsymbol{\lambda}_T - \boldsymbol{\lambda}^\star\|_p,$$

and

$$\left|\mathbb{E}_{\mathbb{P}_T}\left(u\left(\boldsymbol{\lambda}^{\star\prime} \boldsymbol{X}_0\right)\right) - \mathbb{E}_{\mathbb{P}}\left(u\left(\boldsymbol{\lambda}^{\star\prime} \boldsymbol{X}_0\right)\right)\right| \leq \|\boldsymbol{\lambda}^\star\|_2\, \mathcal{W}\left(\mathbb{P}_T, \mathbb{P}; \tfrac{1}{2}\right),$$

and thereby

$$\begin{gathered}\mathbb{P}\left(\left|\mathbb{E}_{\mathbb{P}_T}\left(u\left(\boldsymbol{\lambda}_T' \boldsymbol{X}_0\right)\right) - \mathbb{E}_{\mathbb{P}}\left(u\left(\boldsymbol{\lambda}^{\star\prime} \boldsymbol{X}_0\right)\right)\right| > d(p)\, \mathbb{E}_{\mathbb{P}_T}\left(\|\boldsymbol{X}_0\|_2\right) \|\boldsymbol{\lambda}_T - \boldsymbol{\lambda}^\star\|_p + \|\boldsymbol{\lambda}^\star\|_2\, \xi_T\right) \\ \leq \mathbb{P}\left(\mathcal{W}\left(\mathbb{P}_T, \mathbb{P}; \tfrac{1}{2}\right) > \xi_T\right),\end{gathered}$$

and the result in (9) follows by setting $t := \xi_T$ in (3.2). $\square$

The results rely first on the iidness of the sample, an assumption that is compatible with our data frequency; it can be extended to m-dependent processes as well as to a class of Markov processes that possess contractive transition kernels (see [21]).

Second, they rely on the existence of some square-exponential moment for $\|\boldsymbol{X}_0\|_2^2$. This is equivalent to the existence of the moment generating function of $\|\boldsymbol{X}_0\|_2^2$ in a neighborhood of zero, a condition that fails whenever $\|\boldsymbol{X}_0\|$ 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 [24]). The maximal moment parameter τ can be estimated via the ratio $\frac{(\kappa+1)\mathbb{E}_{\mathbb{P}_T}\left(\|\boldsymbol{X}_0\|_2^{2\kappa}\right)}{\mathbb{E}_{\mathbb{P}_T}\left(\|\boldsymbol{X}_0\|_2^{2\kappa+2}\right)}$, 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 bounds in the first and third cases of the theorem decline exponentially fast in $TM^2(\xi_T, \Lambda, p, \mathcal{U}_2^\star, \mathcal{E}_2)$ and $T\xi_T^2$ respectively, and hold for all T as long as the worst case value, over the adverse event $\mathcal{E}$, of the penalization term $-\xi_T\|\boldsymbol{\lambda}_T - \boldsymbol{\tau}\|_p$, and respectively the penalization coefficient, dominates in absolute value 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 penalization term 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 regularizer (see for example [20]). The requirement of existence of squared exponential moments, can also be circumvented at the cost of loss of exponentially decreasing probability bounds; the existence of regular (polynomial) moments of some order would imply eventual polynomially decreasing probability bounds in in $TM^2(\xi_T, \Lambda, p, \mathcal{U}_2^\star, \mathcal{E}_2)$ and $T\xi_T^2$ respectively, see also [20]. It is possible that the form of the bounds as well as the requirements for the penalization related quantities can be refined using results pertaining to infinitesimally Hilbertian structures, like Corollary 4.11 in [25].

The first result estimates the probability of the adverse event $\mathcal{E}$; this corresponds to the existence of empirically enhanced optimal portfolios that are non-dominant in the population. It says that if the regularization term $\xi_T \inf_{\mathcal{E}} \|\boldsymbol{\lambda}_T - \boldsymbol{\tau}\|_p$ is strictly greater than a term of the form $CT^{\frac{1}{d}}$, then the logarithmic probability that there exist empirically enhanced portfolios that are non dominant in the population, is bounded above by the term $-\frac{(1-C(d)T^{-\frac{d+1}{d}})^2 TM^2(\xi_T,\Lambda,p,\mathcal{U}_2^\star,\mathcal{E})}{2\inf_{\tau>0} h^2(\tau)}$. It is noted that the $\xi_T \inf_{\mathcal{E}} \|\boldsymbol{\lambda}_T -$

$\boldsymbol{\tau}\|_p$ term is not linear in the regularization multiplier, since by construction $\boldsymbol{\lambda}_T \in \boldsymbol{\Lambda}^{\succeq}_{\xi_T}$. Nevertheless the term $\inf_{\mathcal{E}} \|\boldsymbol{\lambda}_T - \boldsymbol{\tau}\|_p$ can be approximated via resampling, something that implies that it is statistically feasible to do inference on the validity of the inequality. The bound depends on the aforementioned worst case value over the event of the penalization term, the base assets dimensionality, the size of the portfolio space, through $d(p)$ on the choice of the ℓ_p norm, the squared exponential moment parameter, and the supremum over the elements of $\mathcal{E}$ of $\inf_{\mathcal{U}_2^\star} D(v, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbb{P})$; e.g, if $\xi_T \inf_{\mathcal{E}} \|\boldsymbol{\lambda}_T - \boldsymbol{\tau}\|_p = 2.1 d(p) \sup_{\boldsymbol{\lambda}} \|\boldsymbol{\lambda}\|_2 C(d) T^{-\frac{1}{d}}$, 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 adverse case regularization term is greater than the maximum between $\sqrt{-2\ln(\alpha)} \inf_{\tau>0} h(\tau) \left|1 - C(d) T^{-\frac{d+1}{d}}\right|^{-1} T^{-\frac{1}{2}}$ and $2d(p) \sup_{\boldsymbol{\lambda}} \|\boldsymbol{\lambda}\|_2 C(d) T^{-\frac{1}{d}}$. Tracing the proof shows that analogous bounds would hold for all $\boldsymbol{\Lambda}^{\succeq}_{\mathbb{P}_T}$, as long as the weaker $\mathcal{E}_1$ holds, yet failure of population dominance only happens inside $\mathcal{U}_2^\star(u)$ with $\mathbb{P}$ probability 1.

For an analogous result, regarding the solutions of the LP relaxed formulation formed as $\boldsymbol{\Lambda}^\star_{\xi_T} := \arg\max_{\boldsymbol{\Lambda}^\star_{\mathbb{P}^{\succeq}_T}} (\mathbb{E}_{\mathbb{P}_T}(u(\boldsymbol{\lambda}'\boldsymbol{x})) - \xi_T \|\boldsymbol{\lambda} - \boldsymbol{\tau}\|_p)$, arguing analogously to the first part of the proof of the proposition, it is not difficult to see that the logarithmic probability of the event $\mathcal{E}^\star := \boldsymbol{\Lambda}^\star_{\xi_T} - \boldsymbol{\Lambda}^{\succeq}_{\mathbb{P}} \neq \emptyset \cap \mathcal{E}_2$ is bounded above by the term $-\frac{(1-C(d)T^{-\frac{d+1}{d}})^2 T \hat{M}^2(\xi_T, \Lambda, p, \mathcal{U}_2^\star, \mathcal{E})}{2\inf_{\tau>0} h^2(\tau)}$, where now $\hat{M}(\xi_T, \Lambda, p, \mathcal{U}_2^\star, \mathcal{E}_2) := M(\xi_T, \Lambda, p, \mathcal{U}_2^\star, \mathcal{E}_2) - \frac{\sup_{\mathcal{E}^\star}(\inf_{\mathcal{U}_2^\star(\mathcal{X}^\star)} D(v, \boldsymbol{\tau}, \boldsymbol{\lambda}_T, \mathbb{P}) - \inf_{\mathcal{U}_2^\star} D(v, \boldsymbol{\tau}, \boldsymbol{\lambda}_T, \mathbb{P}))}{2d(p) \sup_{\boldsymbol{\Lambda}} \|\boldsymbol{\lambda}\|_2}$, for $\mathcal{U}_2^\star(\mathcal{X}^\star) := \{-(z-x)_+,\ z \in \mathcal{X}^\star\}$, and $\boldsymbol{\lambda}_T \in \boldsymbol{\Lambda}^\star_{\xi_T}$. The numerator of the second term above represents a conservative estimate of the approximation error due to the LP-discretization of the support $\mathcal{X}$; under the event $\mathcal{E}^\star$ this term is non-negative, thus augmenting the resulting probability bound compared to the one regarding the probability of $\mathcal{E}$.

The second result says that for the appropriate choice of the regularization multiplier the probability of FD for the members of the enhanced set, will eventually become smaller than the probability of false dominance for the whole of $\Lambda^{\succeq}_{\mathbb{F}_T}$, whenever failure of population dominance happens due to the complement of $\mathcal{U}_2^\star(u)$. It shows the significance of the positive slacks implied by the regularization parameters on the SD conditions, for the small sample mitigation of the particular decision error.

The third result corresponds to an oracle inequality that relates the opportunity loss entailed by the empirical regularized problem, with the ℓ_p-deviation between the empirical and the population solution and the regularization factor. If the empirical solution is consistent, it implies asymptotic negligibility for the opportunity loss w.h.p.

If instead of the regularized problem, the norm-constrained Lagrangean formulation is used,

$$\max_{\boldsymbol{\Lambda}_p(c)\cap\boldsymbol{\Lambda}^{\succeq}_{\mathbb{P}_T}} \mathbb{E}_{\mathbb{P}_T}(u(\boldsymbol{\lambda}'\boldsymbol{x})); \tag{10}$$

$$\boldsymbol{\Lambda}_p(c) := \left\{\boldsymbol{\lambda}\in\boldsymbol{\Lambda} : \|\boldsymbol{\lambda}-\boldsymbol{\tau}\|_p \leq c\right\}, \tag{11}$$

for $c > 0$, then by the Lagrangian Duality Theorem, see Ch. 8, Theorem 1 of [26], the above results would transfer in this framework, as long as the term ξ_T is replaced by the worst case over the relevant event optimal Lagrange multiplier, and likewise the term $\inf_{\mathcal{E}} \|\boldsymbol{\lambda}_T-\boldsymbol{\tau}\|_p$ is replaced by c.

3.3 Discussion

Similar results would be analogously derivable to the above for other stochastic dominance relations, if the relevant set of utility functions that define the relation is equi-continuous w.r.t. the topology of locally uniform convergence. This could accommodate cases of preferences with non-global disposition towards risk like appropriate sets of S-shaped utilities associated with the Prospect Stochastic Dominance relation-see [27] or [28]. Moreover, it is observed that in these instances, the numerical expression of the related optimization problems may not entirely simplify to standard LP programs because they may include components of MILP formulations. In such cases, the associated problems could be solved via cutting plane methods (see for example [29]), or via the use of meta-heuristic methods like the SMA algorithm (see for example [30]). The even more complicated issue of deriving the probabilistic properties of solutions emerging from adopting suchlike optimization methods to stochastic dominance problems is an interesting path for further research.

The iid framework employed in Proposition 2 can be restrictive for economic data of high to moderate observation frequencies. The extension of such-like results to the more econometrically plausible framework of ergodicity and strong mixing could be benefited by analogous extensions of the Wasserstein distance empirical concentration inequalities to such settings. Bakry-Émery positive curvature-dimension conditions, as they relate to transportation inequalities and concentration phenomena, could aid in such considerations; refer to [31] for further insights. Optimality of the probability bounds is of simultaneous interest for the non-asymptotic control of the probability of FD in realistically large samples.

References

[1] E. Hodder, J. C. Jackwerth, and O. Kolokolova. Improved portfolio choice using second-order stochastic dominance. *Review of Finance*, 19(4):1623â1647, 2015.

[2] T. Post and I.R. Longarela. Risk arbitrage opportunities for stock index options. *Operations Research*, 69(1):100–113, 2021.

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

[4] S. Arvanitis and T. Post. Generalized stochastic arbitrage opportunities. *Management Science*, 70(7):4629–4648, 2024.

[5] Ravi Jagannathan and Tongshu Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The journal of finance*, 58(4):1651–1683, 2003.

[6] Victor DeMiguel, Lorenzo Garlappi, Francisco J Nogales, and Raman Uppal. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management science*, 55(5):798–812, 2009.

[7] Philipp J Kremer, Sangkyun Lee, Małgorzata Bogdan, and Sandra Paterlini. Sparse portfolio selection via the sorted 1-norm. *Journal of Banking & Finance*, 110:105687, 2020.

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

[9] W.R. Russell and T.K. Seo. Representative sets for stochastic dominance rules. 1989.

[10] S. Arvanitis and T. Post. Stochastic arbitrage opportunities: Set estimation and statistical testing. *Mathematics*, 12(4):608, 2024.

[11] C. D. Aliprantis and K. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, 2006.

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

[13] T. Post and M. Kopa. Portfolio choice based on third-degree stochastic dominance. *Management Science*, 63(10):3381–3392, 2017.

[14] H. Rahimian and S. Mehrotra. Distributionally robust optimization: A review. *arXiv preprint*, 2019.

[15] A.W. Vaart and J.A. Wellner. *Weak convergence*. Springer, New York, NY, 1996.

[16] L.V. Kantorovich. *On methods of analysis of some extremal problems in planning production*. RAND, 1958.

[17] M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.

[18] S.G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic sobolev inequalities. *Journal of Functional Analysis*, 163(1):1–28, 1999.

[19] F. Bolley and C. Villani. Weighted csiszár-kullback-pinsker inequalities and applications to transportation inequalities. *Annales de la Faculté des sciences de Toulouse: Mathématiques*, 14(3):331–352, 2005.

[20] F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probability Theory and Related Fields*, 137(3):541–593, 2007.

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

[22] J. Lei. Convergence and concentration of empirical measures under wasserstein distance in unbounded functional spaces. *Bernoulli*, 26(1):767–798, 2020.

[23] Y. Maesono. Lower bounds for the normal approximation of a sum of independent random variables. *Australian Journal of Statistics*, 31(3):475–485, 1989.

[24] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press, 2018.

[25] Massimo Fornasier, Giuseppe Savaré, and Giacomo Enrico Sodini. Density of subalgebras of lipschitz functions in metric sobolev spaces and applications to wasserstein sobolev spaces. *Journal of Functional Analysis*, 285(11):110153, 2023.

[26] D.G. Luenberger. *Optimization by vector space methods.* John Wiley Sons, 1997.

[27] M. Levy and H. Levy. Prospect theory: Much ado about nothing? In *Handbook of the Fundamentals of Financial Decision Making: Part I*, pages 129–144. 2013.

[28] S. Arvanitis, O. Scaillet, and N. Topaloglou. Spanning analysis of stock market anomalies under prospect stochastic dominance. *Management Science*, 70(9):6002–6025, 2024.

[29] H. Marchand, A. Martin, R. Weismantel, and L. Wolsey. Cutting planes in integer and mixed integer programming. *Discrete Applied Mathematics*, 123(1-3):397–446, 2002.

[30] F. S. Gharehchopogh, A. Ucan, T. Ibrikci, B. Arasteh, and G. Isik. Slime mould algorithm: A comprehensive survey of its variants and applications. *Archives of Computational Methods in Engineering*, 30(4):2683–2723, 2023.

[31] C. Villani. *Optimal transport: old and new*, volume 338. Springer, Berlin, 2009.