The Statistical Limit of Arbitrage[*]

Rui Da[†]
University of Chicago

Stefan Nagel[‡]
University of Chicago, NBER, and CEPR

Dacheng Xiu[§]
University of Chicago

This version: July 7, 2022

Abstract

When alphas are weak and rare, and arbitrageurs have to learn about alphas from historical data, there is a gap between Sharpe ratio that is feasible for them to achieve and the infeasible Sharpe ratio that could be obtained with perfect knowledge of the true return distribution. This statistical limit to arbitrage widens the bounds within which true alphas can survive in equilibrium relative to the arbitrage pricing theory (APT) in which arbitrageurs are endowed with perfect knowledge of the return distribution. We derive the optimal Sharpe ratio achievable by any feasible arbitrage strategy, and illustrate in a simple model how this Sharpe ratio varies with the strength and sparsity of alpha signals, which characterize the difficulty of arbitrageurs' learning problem. Furthermore, we design an "all-weather" arbitrage strategy that achieves this optimal Sharpe ratio regardless of the conditions of alpha signals. We also show how arbitrageurs can adopt multiple-testing, LASSO, and Ridge methods to achieve optimality under distinct conditions of alpha signals, respectively. Our empirical analysis of equity returns shows that all strategies we consider achieve a moderately low Sharpe ratio out of sample, in spite of a considerably higher infeasible Sharpe ratio, consistent with absence of feasible near-arbitrage opportunities and relevance of statistical limits to arbitrage.

Keywords: Learning about Alphas, Rational Expectation, Portfolio Choice, Rare and Weak Signal, False Negatives, testing APT, Machine Learning

---

[*] This version is preliminary and incomplete. Please do not circulate. We benefited tremendously from discussions with seminar and conference participants at Princeton University, Oxford University, EPFL, University of Liverpool, Sao Paulo School of Economics, Peking University, Tsinghua University, Shanghai University of Finance and Economics, University of Gothenburg, and Stockholm Business School.

[†] Address: 5807 S Woodlawn Avenue, Chicago, IL 60637, USA. E-mail address: `rui.da@chicagobooth.edu`.

[‡] Address: 5807 S Woodlawn Avenue, Chicago, IL 60637, USA. E-mail address: `stefan.nagel@chicagobooth.edu`.

[§] Address: 5807 S Woodlawn Avenue, Chicago, IL 60637, USA. E-mail address: `dacheng.xiu@chicagobooth.edu`.

1 Introduction

It is a fundamental underlying principle of most asset pricing theories, including the Arbitrage Pricing Theory (APT), that investment opportunities with extremely high ratios of reward to risk do not exist in financial markets. Implicitly, these theories rest on the premise that such near-arbitrage opportunities would attract arbitrageurs who exploit and thereby eliminate these opportunities. An important assumption in these theories is that the statistical distribution of returns is known to arbitrageurs. Therefore, near-arbitrage opportunities in the true distribution of returns are ruled out.

In practice, however, sophisticated investors searching for near-arbitrage opportunities do not know this distribution. Instead, they commonly conduct statistical analyses to learn about the existence of such opportunities from historical returns data. As a consequence, they face statistical uncertainty. In some settings, such as in some derivatives pricing applications, for instance, the statistical uncertainty may be sufficiently small that it is not a significant impediment to arbitrageur activity. But in noisy, high-dimensional settings such as the cross-section of stock returns, statistical uncertainty can be substantial and it can can constitute a statistical limit to arbitrage.

To analyze the effects of arbitrageur learning, we consider a setting in which returns follow a statistical linear factor model. Near-arbitrage opportunities are characterized by high Sharpe ratios. To exploit such opportunities, arbitrageurs need knowledge of factor model alphas, but they must learn about these from historical realizations of alpha signals. We derive the optimal Sharpe ratio achievable by any feasible arbitrage trading strategies, which is strictly dominated by the infeasible optimal Sharpe ratio that arbitrageurs could achieve if they were endowed with perfect knowledge of alphas. This, in turn, provides a new no-near-feasible-arbitrage bound on the Sharpe ratio that accounts for the statistical limit to arbitrage.

The difficulty of the learning problem hinges on the data generating process (DGP) of alpha signals. While our theory generally does not rely on specific cross-sectional distributions of alpha signals, we use simple special cases to demonstrate how the optimal Sharpe ratio varies with the strength and sparsity of alphas. When alphas are strong and not too rare relative to the dimensionality of the cross-section and the sample size, arbitrageurs can learn the distribution of alpha perfectly in the limit. But when alpha is weaker and more rare, its inference becomes more challenging and a gap arises between the optimal feasible Sharpe ratio and the infeasible Sharpe ratio that requires perfect knowledge of alphas. For instance, the infeasible Sharpe ratio may explode in the asymptotically, while the feasible Sharpe ratio stays bounded.

The existence of this statistical limit to arbitrage implies a widening of the bounds in which mispricing can survive in equilibrium compared with a situation in which arbitrageurs know the DGP and its parameters. Some mispricing may survive it is clouded by too much statistical uncertainty. Empirically therefore, the feasible, not the infeasible, Sharpe ratio tells us about the minimum reward-to-risk compensation that arbitrageurs require.

We further demonstrate how arbitrageurs can construct a feasible trading strategy that achieves the theoretically optimal feasible Sharpe ratio, uniformly over DGPs of alphas, regardless of the strength and sparsity of alphas. This means that the feasible Sharpe ratio bound is in fact sharp. A uniformly valid trading strategy is desirable because in reality arbitrageurs do not know which DGP is a correct description of the observed data. The optimal strategy estimates the empirical distribution of alpha signals and assigns weights based on the relative magnitudes and associated uncertainty of the alpha estimates. Assets with with high alpha t-statistics get portfolio weights proportional to their signal strength. Weaker alphas are more difficult to exploit, yet simply ignoring them would lead to a suboptimal trading strategy. The optimal strategy constructs portfolio weights for weak signals by locally smoothing alpha signals cross-sectionally.

To empirically contrast feasible and infeasible Sharpe ratios, we also propose an estimator of the infeasible Sharpe ratio that a hypothetical arbitrageur endowed with perfect knowledge of DGP parameters would perceive. While this Sharpe ratio can be estimated consistently, it cannot be realized by any feasible portfolio with weights constructed using historical data. The infeasible Sharpe ratio often serves as the building block for tests of APT, see, e.g., Gibbons et al. (1989), Gagliardini et al. (2016), Fan et al. (2015), and Pesaran and Yamagata (2017). While such tests are powerful and may lead to discoveries of alpha signals, they are not relevant for arbitrageurs who are confined to feasible trading strategies. Our effort in constructing the optimal feasible arbitrage portfolio and evaluating its economic performance directly responds to Shanken's call (Shanken (1992)): "... practical content is given to the notion of 'approximate arbitrage,' by characterizing the investment opportunities that are available as a consequence of the observed expected return deviation ... Far more will be learned, I believe, by examining the extent to which we can approximate an arbitrage with existing assets."

Next, we examine whether alternative strategies that exploit multiple testing, shrinkage, and selection techniques to build arbitrage portfolios can attain the optimal feasible Sharpe ratio. With alphas estimated from cross-sectional regressions, one strategy adopts a multiple-testing (BH) procedure as in Benjamini and Hochberg (1995) on the individual p-values of t-statistics for alpha, in order to guard against potential false discoveries among significant alphas, before building the optimal portfolio weights using selected alphas. Other strategies use either LASSO or Ridge penalties to regularize portfolio weights based on alpha estimates. Such strategies amount to imposing a prior distribution on the alphas. We illustrate with a simple example that these strategies can achieve optimal Sharpe ratio under distinct alpha assumptions. In particular, BH procedure achieves optimal performance only when few true alpha signals are substantially strong. Its failure to achieve optimality is precisely due to its conservativeness nature against the less potent alphas. In contrast, the ridge-based portfolio is equivalent to that constructed by alpha estimates from plain cross-sectional regressions. This approach can achieve optimality when almost all true alphas are either uniformly strong or uniformly weak. The LASSO approach attempts to strike a balance between the aforementioned two methods, with a small gap to achieving the theoretically optimal Sharpe ratio.

Finally, we demonstrate the empirical implications of the statistical limits of arbitrage by examining 56 years of monthly individual equity returns in US stock market from 1965 to 2020. The average number of stocks over this period exceeds 4000. We construct residuals via cross-sectional regressions from a multi-factor model that directly uses observed characteristics as risk exposures. These characteristics include market beta (Fama and MacBeth (1973)), size (Banz (1981)), operating profits/book equity (Fama and French (2006)), book equity/market equity (Fama and French (2006)), asset growth (Cooper et al. (2008)), momentum (Jegadeesh and Titman (1993)), short-term reversal (Jegadeesh (1990)), industry momentum (Moskowitz and Grinblatt (1999)), illiquidity (Amihud (2002)), leverage (Bhandari (1988)), return seasonality (Heston and Sadka (2008)), sales growth (Lakonishok et al. (1994)), accruals (Sloan (1996)), dividend yield (Litzenberger and Ramaswamy (1979)), tangibility (Hahn and Lee (2009)), and idiosyncratic risk (Ang et al. (2006)), as well as 11 Global industry Classification Standard (GICS) sectors. These characteristics and industry dummies capture similar equity factors in the MSCI Barra model widely-used among practitioners.

A few interesting findings emerge. First, the cross-sectional R^2s are rather low, with a time-series average 8.25% over our sample period from Jan 1965 to Dec 2020. These R^2s are in similar magnitudes compared to existing estimates in the literature, e.g., 7.8% average R^2s from May 1964 to Dec 2009 reported in Lewellen (2015) using 15 factors that largely overlap with ours, but lower than 12-14% over 1987 - 2016 reported in Gu et al. (2021) based on latent factor models. This indicates that there exists a considerable amount of idiosyncratic noise in the cross-section of individual equities, which makes learning about alphas an arduous statistical task.

Second, we obtain the t-statistics corresponding to alpha estimates of all individual stocks based on their full record in our sample. Among 12,415 test statistics in total, only 6.35% (0.63%) of the t-statistics are greater than 2.0 (3.0) in absolute values. Only 0.505% of these t-statistics translate to Sharpe ratios greater than 1.0 in magnitude. Even without controlling for multiple testing, these estimates suggest that non-zero alphas are rather rare and weak.

Third, we find that the optimal feasible arbitrage portfolio with different methods achieve a moderately low annualized Sharpe ratio, about 0.5, whereas the infeasible Sharpe ratios over time are considerably higher—beyond 2.5—on average, and can reach as high as 7.5 for some sample periods. The estimated infeasible Sharpe ratio is an estimate of what arbitrageurs could attain if they had perfect knowledge of DGP parameters, but it is not attainable by constructing a feasible arbitrage portfolio. The large gap between feasible and infeasible Sharpe ratios suggests the empirical relevance of the statistical limit of arbitrage. Moreover, the fact that the feasible Sharpe ratio is small suggests the empirical success of APT.

Fourth, among all feasible strategies, BH and our optimal strategy achieve the best performance, around a Sharpe ratio of 0.5, followed by CSR (0.450) and LASSO (0.384), though the differences are not substantial.However, the BH approach is overly conservative that it eliminates almost all weak signals and trade less than 10 stocks each month, with zero trading activities for over half of the entire sample. CSR and out strategy exploit weak signals. CSR trades all stocks, but receive a slightly lower Sharpe ratio, potentially due to misallocation of portfolio weights to fake signals. Our optimal strategy trade almost all stocks, but with weights adaptive to the signal strength. LASSO is most disappointing, but this is due to the uncertainty in optimal tuning parameters. The resulting number of traded stocks per month varies considerably from none to all.

Our paper builds on a large literature on the arbitrage pricing theory (APT) developed by Ross (1976) and later refined by Huberman (1982), Chamberlain and Rothschild (1983), and Ingersoll (1984). As in these papers, we rely on asymptotic arguments that do not rely on assumptions about investor preferences, but these results should be seen as an asymptotic approximation for a more realistic setting with a finite number of assets in which weak preference restrictions rule out Sharpe ratios far above the Sharpe ratios of diversified factor portfolios. The statistical limits to arbitrage that we highlight in this paper relax this Sharpe ratio bound compared with an economy in which arbitrageurs are endowed with perfect knowledge of DGP parameters. In this regard, our paper is also related to another large strand of literature on the limit of arbitrage, see Gromb and Vayanos (2010) for a comprehensive review. Complementary to the existing literature, the arbitrage limit in our setting stems from statistical uncertainty, instead of being induced from risk, costs, frictions, and other constraints rational expectation investors are facing.

Kozak et al. (2018) argue that the absence of near-arbitrage opportunities enforces the expected returns to approximately line up linearly with common factor covariances, even in a world in which belief distortions affect asset prices. Our study focuses on the deviations of expected returns from this approximate linear relation and how statistical limits to arbitrage allow bigger deviations. A closely related paper to ours is Kim et al. (2020), which proposes a characteristics-based factor model to construct arbitrage portfolios. Their model does not preclude arbitrage opportunities with a theoretically infinite Sharpe ratio, whereas our framework rules out such a possibility. Relatedly, Uppal and Zaffaroni (2018) propose a methodology to construct robust portfolios that can be decomposed into alpha (arbitrage) portfolios and beta (factor) portfolios. Our setting is considerably different in that alphas cannot possibly be recovered with certainty even when the sample size is large.

Our paper is also related to the evolving literature on applications of statistical and machine learning in asset pricing, and in particular on the topic of testing the APT, e.g., Gibbons et al. (1989), Gagliardini et al. (2016), and Fan et al. (2015), as well as on testing for alphas, e.g., Barras et al. (2010), Harvey and Liu (2020), and Giglio et al. (2021). The first literature focus on testing a null that all alphas are equal to zero. This is certainly an interesting null hypothesis, but as we emphasize in this paper, the APT does allow for alphas as long as they do not induce an explosive feasible Sharpe ratio. The second literature focuses on detecting strong alphas, in which widely used multiple testing methods, such as the BH method by Benjamini and Hochberg (1995), or its extensions can be applied to control the false discovery rate (FDR). In contrast, we allow for rare and weak alpha signals such that any procedure aiming to control the FDR is too conservative with too few or no discoveries.[1] Our objective here is not on model testing or signal detection. Rather, we strive for the optimal economic performance of arbitrage portfolios. We show that even if signals

---

[1] Donoho and Jin (2004) adopt the so-called higher criticism approach, dating back to Tukey (1976), to detect rare and weak signals in a stylized multiple testing problem.

were so weak that they are undetected by multiple testing methods, they may lead to a portfolio with a considerable Sharpe ratio.

There has been a long-standing critique of rational expectation models in macroeconomics and finance in which economic agents are not confronted with statistical uncertainty over structure parameters, see Hansen (2007). Bayesian learning is one way to expose model agents to statistical uncertainty. Pastor and Veronesi (2009) survey the literature on learning in financial markets. In many settings, e.g., Collin-Dufresne et al. (2016), learning can be sufficiently slow such that its effects persist in empirically realistic sample sizes, even though convergence to rational expectations takes place in the long-run. An exception is Martin and Nagel (2021) where learning effects persist because investors face a high-dimensional inference problem about the process generating firm cash flows. Similarly, arbitrageurs in our model attempt to make inference on a high-dimensional parameter vector with a potentially insufficient sample size, but they learn about returns, not firms' underlying cash flows. We examine different sequences of DGPs and in most scenarios, our learning system does not converge to a rational expectations limit.[2]

Last but not least, our paper is related to Chen et al. (2021b) and Chen et al. (2021a) in that they also account for the distinction between beliefs of economic agents and the DGP revealed by empirical evidence. They model belief distortions as a change of measure in moment conditions, use statistical measures of divergence relative to rational expectation to bound the set of subjective probabilities, and seek robust inference with this form of misspecification. We develop an optimal arbitrage strategy that is optimal with respect to a large class of DGPs. This Sharpe ratio is optimal for arbitrageurs inside the economic model, which is in contrast with the (infeasible) optimal Sharpe ratio from an outside econometrician's point of view, in the spirit of Hansen (2014). We provide estimators for both and compare them empirically.

Our paper proceeds as follows. Section 2 develops a statistical limit of arbitrage. Section 2.5 constructs a feasible trading strategy that achieves the optimal Sharpe ratio. Section 2.7 analyzes two alternative trading strategies. Section 3 provides simulation evidence, followed by an empirical analysis in Section 4. Section 5 concludes. The appendix provides technical details.

2 Statistical Limit of Arbitrage

We start by revisiting the arbitrage pricing framework developed by Ross (1976). This setting is ideal for explaining the statistical limit of arbitrage because the arbitrage pricing theory is largely developed based on a reduced-form statistical model for asset returns. This stylized model is sufficiently sophisticated to deliver theoretical insight, and is sufficiently relevant to guide empirical investment decisions.

---

[2] Our analysis is related to a large literature in econometrics and statistics that discuss uniform validity of asymptotic approximations, see, e.g., Staiger and Stock (1997), Imbens and Manski (2004), Leeb and Pötscher (2005), Andrews et al. (2020).

2.1 Factor Model Setup

To be more concrete, the factor economy has N assets in the investment universe. The $N \times 1$ vector of excess returns r_t follows a reduced-form linear factor model, for $t = 1, 2, \dots, T$:

$$r_t = \alpha + \beta\gamma + \beta v_t + u_t, \tag{1}$$

where β is an $N \times K$ matrix of factor exposures (with the first column being a vector of 1s), α is an $N \times 1$ vector of pricing errors, v_t is a $K \times 1$ vector of factor innovations with covariance matrix Σ_v, γ is a $K \times 1$ vector of risk premia (and zero beta rate), and u_t is a vector of idiosyncratic returns, independent of v_t, with a diagonal covariance matrix Σ_u. While it is customary to consider the approximate factor model, allowing for off-diagonal entries in the covariance matrix Σ_u would introduce additional statistical obstacles due to the estimation of large covariance matrix for inference on alpha and for building optimal portfolios. For the purpose of illustrating the economic insight of limits on arbitrage, we instead focus on the statistical learning problem of alpha signals, which is sufficiently involved and is of first-order importance for building arbitrage portfolios, as will become clear below.

Throughout we will consider asymptotic limits as N and T increase while K is fixed. To facilitate our asymptotic analysis along the cross-sectional dimension, N, we regard high dimensional objects such as α, β, and Σ_u as random variables drawn from some cross-sectional distributions, whereas γ and Σ_v are regarded as deterministic parameters, since their dimensions are fixed. We assume that α has mean zero, and is cross-sectionally independent of β, and that β has full column rank and is pervasive. These conditions are essential for identification of γ in a model that allows for pricing errors.

We formalize the conditions below.

Assumption 1. *For each $N \geq 1$, the following conditions hold:*

- *(a)* $\|\beta\|_{\mathrm{MAX}} \lesssim_{\mathrm{P}} 1$, *and* $\lambda_{\min}(\beta^\intercal\beta) \gtrsim_{\mathrm{P}} N$.[3]
- *(b)* v_t *is i.i.d. across* t, $\mathrm{E}(v_t) = 0$, *and its covariance matrix* Σ_v *satisfies* $1 \lesssim \lambda_{\min}(\Sigma_v) \leq \lambda_{\max}(\Sigma_v) \lesssim 1$.
- *(c)* $\mathrm{E}(\alpha) = 0$, *and* α_i *is i.i.d. with density* $p_\alpha(\cdot)$ *and satisfies* $\mathrm{E}\big(\max_{i:i\leq N} \alpha_i^2\big) = o(1)$.
- *(d)* $\mathrm{E}(u_t) = 0$, *and* $\sigma_i^2 = (\Sigma_u)_{i,i}$ *is i.i.d. across* i, *independent with* (α, v_t, β), *and satisfies* $1 \lesssim_{\mathrm{P}} \lambda_{\min}(\Sigma_u) \leq \lambda_{\max}(\Sigma_u) \lesssim_{\mathrm{P}} 1$.
- *(e) The pricing errors* α, *factors* v_t, *factor loadings* β, *and idiosyncratic error* u_t *are mutually independent.*

---

[3] For a matrix A, we use $\|A\|$ and $\|A\|_{\mathrm{MAX}} = \max_{i,j} |a_{ij}|$ to denote the operator norm (or $\mathbb{L}_2$ norm) and the $\mathbb{L}_\infty$ norm of A on the vector space. We use C to denote a generic constant that may change from line to line. We use $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ to denote the minimum and maximum eigenvalues of A. We also use the notation $a \lesssim b$ to denote $a \leq Cb$ for some constant $C > 0$ and $a \lesssim_{\mathrm{P}} b$ to denote $a = O_{\mathrm{P}}(b)$.

Assumption 1 (a) and (b) are commonly seen in the literature of factor models. In particular, the assumption on $\lambda_{\min}(\beta^\intercal\beta)$ requires that all factors are pervasive.[4] (c) and (d) suggest that the signals in our model are weak, in that as N increases their magnitudes shrink towards 0, whereas volatilities are bounded from above and from below. Condition (c) on α implies that $\mathrm{Var}(\alpha_i) = \mathrm{E}(\alpha_i^2) = o(1)$. As will become clear (from footnote 8), a diminishing variance on α is necessary for precluding near-arbitrage opportunities in Ross' APT.

There are at least three variations of the factor model (1), depending on what econometricians assume to be observable. The most common setup in academic finance literature imposes that factors are observable as in e.g., Fama and French (1993).[5] The second setting, which has gained more popularity since its debut in Connor and Korajczyk (1986), assumes that factors are latent. The third setting, arguably most prevalent among practitioners, is the MSCI Barra model originally proposed by Rosenberg (1974), where factor exposures, i.e., characteristics, are assumed observable. The advantage of the last model lies in the fact that estimating a large number of (potentially) time-varying stock-level factor exposures is statistically inefficient and computationally expensive, as opposed to directly specifying risk exposures as (linear functions of) observable characteristics.[6]

Our core theoretical results below (e.g., Theorem 1) directly apply to all three cases aforementioned. In our empirical analysis we will adopt the third framework most convenient for modeling individual stocks. This makes our analysis highly relevant for practitioners.

2.2 Feasible Near-Arbitrage Opportunities

Building upon the insight of Ross (1976), Huberman (1982) and Ingersoll (1984) established the concept of near-arbitrage, which can be formalized in a more general setting as below:

Definition 1. *A portfolio strategy w at time t is said to generate a near-arbitrage under a sequence of data-generating processes, such as (1), defined in a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathrm{P})$, if it satisfies $w \in \mathcal{F}_t$, and along some diverging subsequence,[7] with probability approaching one,*

$$\mathrm{Var}(w^\intercal r_{t+1}|\mathcal{F}_t) \to 0, \quad \mathrm{E}(w^\intercal r_{t+1}|\mathcal{F}_t) \geq \delta > 0.$$

---

[4] See, e.g., Assumption I.1 of Giglio and Xiu (2021). While our theoretical results may extend to certain weak factor settings, this is not our emphasis here.

[5] This is different from saying factor innovations, v_t, are observable. The setting of observable factors typically involves another equation that $f_t = \mu + v_t$, where μ are the population means of the observed factors f_t, which are not necessarily identical to the factor risk premia, γ.

[6] Strictly speaking, the MSCI Barra model is cast in a conditional version of (1):

$$r_t = \alpha_{t-1} + \beta_{t-1}\gamma_{t-1} + \beta_{t-1}v_t + u_t, \tag{2}$$

where β_t is a vector of observed characteristics and γ_{t-1} is a vector of time-varying risk premia. Analyzing this conditional model will not yield additional economic insight relative to the unconditional model with respect to the theoretical limit of arbitrage. This model is overly parametrized that parameters are not identifiable without additional restrictions. Some examples of parsimonious conditional factor models include Connor et al. (2012), Gagliardini et al. (2016), and Kelly et al. (2019).

[7] We adopt the same subsequence definition as that used in Ingersoll (1984). The subsequence typically depends on the count of investment opportunities, i.e., N, though we do not need make this explicit in this definition. For simplicity of notation and without ambiguity, we omit the dependence of w on N and t.

Intuitively, no near-arbitrage means there exist no sequence of portfolios that earn positive expected returns with vanishing risks. Under conditions similar to those in Assumption 1, Ingersoll (1984) established that a sufficient and necessary condition for the absence of near-arbitrage is that

$$S^\star = \sqrt{\alpha^\intercal \Sigma_u^{-1} \alpha} \lesssim_{\mathrm{P}} 1. \tag{3}$$

$S^\star$ is the theoretically optimal Sharpe ratio arbitrageurs can achieve in this economy without exposure to factor risks. This result suggests that moderate mispricing in the form of nonzero alphas is permitted in an economy without near-arbitrage opportunities, but there cannot be too many alphas that are too large, to the extent that $S^\star$ explodes along some diverging sequence.[8]

To achieve this optimal Sharpe ratio, arbitrageurs should hold a portfolio with weights given by $w^\star = \Sigma_u^{-1}\alpha$, according to Ingersoll (1984).[9] Under the rational expectation assumption, arbitrageurs (agents in this model) know the true (population) parameters: α and Σ_u. In reality, however, the true parameters are blind to arbitrageurs as they can only learn these parameters from a finite sample of data. This learning effect is sometimes harmless since it can be expected that when the sample size is large enough, the true parameters are (asymptotically) revealed, and hence the predictions under rational expectation hold approximately. Fundamentally, this phenomenon is due to the assumption that the learning problem in the limiting experiment becomes increasingly simpler as the sample size increases.

In the current context, the difficulty of the learning problem also hinges on the set of investment opportunities, N. As N increases, it becomes increasingly difficult for arbitrageurs to determine which among all assets truly have nonzero alphas for a given sample size, T. If the learning problem remains difficult as N and T increase, the learning effect persists, which could lead to distinct limiting implications as opposed to the rational expectation case. It turns out that the rational expectation limit $S^\star$ is only relevant for rather restrictive scenarios. In more realistic settings, e.g., N is much larger than T, the optimal Sharpe ratio arbitrageurs can achieve without factor exposures is far smaller than $S^\star$ because of their inability to make error-free inference. Therefore, the condition (3) could be excessively restrictive in such scenarios.

To illustrate this intuition, we consider a simple and specific example.

Example 1. *Suppose the cross-section of alphas is drawn from the following distribution:*

$$\alpha_i \overset{i.i.d.}{\sim} \begin{cases} \mu & \textit{with prob. } \rho/2 \\ -\mu & \textit{with prob. } \rho/2 \\ 0 & \textit{with prob. } 1-\rho \end{cases}, \quad 1 \le i \le N, \tag{4}$$

where $\mu \ge 0$ and $0 \le \rho \le 1$, and they potentially vary with N and T. In addition, we also assume

---

[8] It is obvious that $\mathrm{E}(N\alpha_i^2) = \mathrm{E}(\alpha^\intercal\alpha) \le \mathrm{E}(\alpha^\intercal \lambda_{\min}(\Sigma_u^{-1})\alpha) \le C\mathrm{E}(\alpha^\intercal\Sigma_u^{-1}\alpha) \le C$, which implies $\mathrm{E}(\alpha_i^2) = o(1)$.

[9] In Ingersoll (1984), α is defined to be the cross-sectional projection of the expected returns onto β in the population model such that $\alpha^\intercal\Sigma_u^{-1}\beta = 0$. In this paper, we assume instead that α is random, satisfying $\mathrm{E}(\alpha^\intercal\beta) = 0$, and hence in our setting $w^\star = \mathbb{M}_\beta\Sigma_u^{-1}\alpha$, which achieves the Sharpe ratio $S^\star$ (asymptotically as N increases).

$\beta = 0$, $\Sigma_u = \sigma^2 \mathbb{I}_N$, *for some* $\sigma > 0$.

In this example, μ dictates the strength of alphas, ρ describes how rare alphas are, whereas σ is a nuisance parameter. By modeling parameters μ and ρ as functions of the sample size and dimensions of the investment set, we can accurately characterize the difficulty of the finite sample problem arbitrageurs are facing.[10] To emphasize the role of signal strength and count, we impose in this example that all assets share the same alpha distribution and the same idiosyncratic variance.

Now suppose, more specifically, that the magnitude of (μ, ρ) satisfies

$$\mu \sim T^{-1/2} \quad \text{and} \quad \rho \sim N^{-1/2}. \tag{5}$$

This condition (5) implies that the signal strength vanishes as the sample size increases and the signal count decays as the investment universe expands. That is, only a small portion of assets have a nonzero yet small alpha. We assume σ is a fixed constant, since in reality idiosyncratic risks never vanish, whereas alphas can be small driven by competition among arbitrageurs. This model rests on an uncommon territory in the existing literature of asset pricing: weak and rare alphas. In fact, the classical no near-arbitrage condition (3) imposes, implicitly, weakness or rareness on alphas; otherwise, if alphas are strong and dense, $\alpha^\intercal \alpha$ would explode rather rapidly. Even in the current setting, in light of the fact that $\mathrm{E}(\alpha^\intercal \alpha) = \rho\mu^2 N$, we still have $\alpha^\intercal \alpha \xrightarrow{p} \infty$ as long as $N^{1/2}/T \to \infty$. In this case, a near-arbitrage opportunity arises according to (3), with a strategy $w = \sigma^{-2}\alpha$.

However, the statistical obstacle prevents arbitrageurs from having this "free lunch." In general, it is only possible to recover any element of alpha up to some estimation error of magnitude $T^{-1/2}$.[11] Since the true alpha is of the same order of magnitude as its level of statistical uncertainty by design, i.e., $\mu \sim T^{-1/2}$, it is impossible for arbitrageurs to determine precisely which assets among all have nonzero alpha.

For illustration purpose, suppose that arbitrageurs adopt the strategy $\widehat{w} = \sigma^{-2}\widehat{\alpha}$, replacing α in w with $\widehat{\alpha} = \bar{r} = \alpha + \bar{u}$.[12] Out of sample, this portfolio's conditional expected return and conditional variance can be written as:

$$\mathrm{E}\left(\sigma^{-2}\left(\alpha + \bar{u}\right)^\intercal \left(\alpha + u_t\right) | \mathcal{F}_{t-1}\right) = \sigma^{-2}(\alpha^\intercal \alpha + \bar{u}^\intercal \alpha),$$

---

[10] Adopting a drifting sequence for parameters is a common trick in econometrics to provide more accurate finite sample approximations. As Bekker (1994) put, "in evaluating the results, it is important to keep in mind that the parameter sequence is designed to make the asymptotic distribution fit the finite sample distribution better. It is completely irrelevant whether or not further sampling will lead to samples conforming to this sequence or not."

[11] Giglio et al. (2021) develop the asymptotic normality result for alpha estimates via a Fama-MacBeth procedure in various scenarios, in which factors are (partially) observable or latent whereas β is unknown. The CLTs in these scenarios share the same form: for any $1 \le i \le N$,

$$\sqrt{T}(\widehat{\alpha}_i - \alpha_i) \xrightarrow{d} \mathcal{N}(0, \sigma_i^2(1 + \gamma^\intercal(\Sigma_v)^{-1}\gamma)), \tag{6}$$

where σ_i^2 is the ith entry of Σ_u. In the case that β is observable (but factors are not), we can show that the CLT has a similar form except that the scalar $(1 + \gamma^\intercal(\Sigma_v)^{-1}\gamma)$ disappears.

[12] For any time series of random vector a_t, we use $\bar{a}$ to denote its sample average. As we will point out later in the paper, this strategy $\widehat{w}$, which we will denote by $\widehat{w}^{\mathrm{CSR}}$, fails to achieve the optimal Sharpe ratio in all scenarios. We will discuss the optimal strategy in Section 2.5.

$$\mathrm{Var}\left(\sigma^{-2}\left(\alpha+\bar{u}\right)^\intercal\left(\alpha+u_t\right)|\mathcal{F}_{t-1}\right)=\sigma^{-2}(\alpha^\intercal\alpha+2\alpha^\intercal\bar{u}+\bar{u}^\intercal\bar{u}),$$

where u_t denotes a future return at t, independent of $\bar{u}$ which is part of the information set up to $t-1$, $\mathcal{F}_{t-1}$, but share the same distribution as $\{u_s\}_{s\leq t-1}$. The resulting squared Sharpe ratio is given by:

$$S^2=\frac{\sigma^{-2}(\alpha^\intercal\alpha+\bar{u}^\intercal\alpha)}{\sigma^{-2}(\alpha^\intercal\alpha+2\alpha^\intercal\bar{u}+\bar{u}^\intercal\bar{u})}\lesssim_{\mathrm{P}} T^{-1}\to 0. \tag{7}$$

Indeed, as we will show later, in this setting, the optimal Sharpe ratio, denoted by S^{OPT}, among all *feasible* trading strategies arbitrageurs adopt, vanishes asymptotically as $N,T\to\infty$, even though the *infeasible* optimal Sharpe ratio $S^\star\to\infty$. The gap between S^{OPT} and $S^\star$, as shown by this example, is enormous.

We say a strategy is *feasible* if it only uses observable data, combined with necessary statistical inference. We formalize the definition of a feasible portfolio strategy below:

Definition 2. *A portfolio strategy $\widehat{w}$ is said to be feasible at time t, if it is a function of observables from $t-T+1$ to t, where T is the sample size.*

This Sharpe ratio gap is driven by the fact that the performance of a feasible portfolio depends on the difficulty of the learning problem. The learning problem in the above example remains difficult as N and T increase, to the extent that the learning effect does not diminish in the limit and the rational expectation limit becomes irrelevant. As we show below, the learning problem in practice is rather difficult, hence the optimal arbitrage Sharpe ratio achievable is expected to be much smaller than $S^\star$.

2.3 Upper Bound on Feasible Sharpe Ratios

We now demonstrate the impact of the feasibility constraint on the optimal arbitrage portfolio. For any feasible strategy $\widehat{w}$, its (conditional) Sharpe ratio can be written as:

$$S(\widehat{w}):=\mathrm{E}(\widehat{w}^\intercal r_{t+1}|\mathcal{F}_t)/\mathrm{Var}(\widehat{w}^\intercal r_{t+1}|\mathcal{F}_t)^{1/2}.$$

The next theorem provides an upper bound on $S(\widehat{w})$:

Theorem 1. *Suppose that r_t follows (1) and that Assumption 1 holds. For any feasible portfolio weight $\widehat{w}$, its Sharpe ratio, $S(\widehat{w})$, satisfies, as $N\to\infty$:*

$$S(\widehat{w})\leq\left(S(\mathcal{G})^2+\gamma^\intercal\Sigma_v^{-1}\gamma\right)^{1/2}+o_{\mathrm{P}}(1),\quad \textit{with}\quad S(\mathcal{G})^2:=\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\mathrm{E}(\alpha|\mathcal{G}), \tag{8}$$

where $\mathcal{G}$ is the information set (i.e., σ-algebra) generated by $\{(r_s,\beta,v_s,\Sigma_u): t-T+1\leq s\leq t\}$.

Intuitively, $S(\mathcal{G})^2$ and $\gamma^\intercal\Sigma_v^{-1}\gamma$ are squared Sharpe ratios earned from arbitrage portfolios and factor portfolios, respectively. Furthermore, $S(\mathcal{G})$ is an upper bound for Sharpe ratios of all feasible portfolio strategies that have no factor exposures.

Theorem 1 shows that it is $\mathrm{E}(\alpha|\mathcal{G})$, the posterior estimate of the pricing errors, α, that dictates the optimal feasible Sharpe ratio for arbitrageurs, rather than α themselves. In fact, it holds by the definition of $S(\mathcal{G})$ that

$$\mathrm{E}\big(S(\mathcal{G})^2\big) \leq \mathrm{E}\big(\alpha^\intercal \Sigma_u^{-1} \alpha\big),$$

with the equality holds only when $\mathrm{E}(\alpha|\mathcal{G}) = \alpha$ almost surely, where the right-hand side corresponds to the infeasible scenario in which arbitrageurs can learn α perfectly using their information set, which echoes (3), the result given by Huberman (1982).[13] In light of Definitions 1 and 2, we immediately obtain a sufficient condition of the absence of near-arbitrage with feasible strategies:

Corollary 1. *Suppose the same assumptions as in Theorem 1 hold. For any given return-generating process satisfying (1), there exists no feasible strategy $\widehat{w}$ that leads to a near-arbitrage, if*

$$S(\mathcal{G}) \lesssim_{\mathrm{P}} 1, \quad \textit{as} \quad N \to \infty. \tag{9}$$

The form of $S(\mathcal{G})$ in Theorem 1 appears that arbitrageurs rely on the information set $\mathcal{G}$, which embodies perfect knowledge of factors, v_t, and their exposures, β, in addition to past asset returns, r_t. Moreover, arbitrageurs appear to have perfect knowledge of the (diagonal) covariance matrix of idiosyncratic errors, Σ_u. In fact, the upper bound in (8) still holds if arbitrageurs are endowed with less information, because for any information sets $\mathcal{G}'$ and $\mathcal{G}$ such that $\mathcal{G}' \subseteq \mathcal{G}$, we have $\mathrm{E}(S(\mathcal{G}')^2) \leq \mathrm{E}(S(\mathcal{G})^2)$. Furthermore, we will show in Section 2.5 that $S(\mathcal{G})$ is in fact achievable by a feasible strategy we construct, which only assumes knowledge of β and r_t – the setting in which factor exposures are observable. This implies that the no near-arbitrage bound in (9) is sufficient and necessary.

The reason that Σ_u plays no significant role is that in our model idiosyncratic variances do not vanish as N and T increase, unlike alphas. This assumption makes sense empirically, because alphas are small and (potentially) rare, driven by competition among arbitrageurs, whereas idiosyncratic risks never diminish. Consequently, detecting alphas is more challenging as opposed to estimating idiosyncratic variances, and hence the latter plays a secondary (and negligible) role as opposed to the former in the limit of arbitrage.

To avoid technical details of the main text, we simplify model (1) by imposing that $\Sigma_u = \sigma^2 \mathbb{I}_N$ for some fixed constant $\sigma^2 > 0$ in the following-up sections, while leaving discussion of the general inhomogeneous case to the appendix, where we have Assumption A2 in place.

---

[13] For ease of discussion, we assume alpha is random. This difference with Huberman (1982) by itself does not affect any economic or statistical conclusions we draw in this paper.

2.4 Explicit Formula of the Sharpe Ratio Bound

To gain more insight on $S(\mathcal{G})$, we seek a more explicit expression in this section. For that purpose, we need impose a normality assumption on the idiosyncratic error.

Proposition 1. *Suppose that r_t follows (1) and Assumption 1 holds. We also assume that for all (i,t), $u_{i,t} \sim \mathcal{N}(0,\sigma^2)$. Then it holds that*

$$\mathrm{E}(\alpha_i|\mathcal{G}) = \psi(\widehat{z}_i),$$

where $\widehat{z}_i = T^{1/2}(\alpha_i + \bar{u}_i)/\sigma$, $\bar{u}$ is the average based on a sample of size T,

$$\psi(a) = \frac{\mathrm{E}\left(\alpha_i \phi(a - T^{1/2}\alpha_i/\sigma)\right)}{\mathrm{E}\left(\phi(a - T^{1/2}\alpha_i/\sigma)\right)},$$

$\phi(\cdot)$ is the normal pdf function, and $\mathrm{E}(\cdot)$ is the expectation taken with respect to the cross-sectional distributions of α_i. If it further holds that $\mathrm{E}(\alpha_i^2 \mathbb{1}_{\{|\alpha_i|\geq c_N\}}) \leq c_N N^{-1}$ for some sequence $c_N \to 0$, then we have

$$S(\mathcal{G}) = S^{\mathrm{OPT}} + o_{\mathrm{P}}(1), \quad \textit{with} \quad S^{\mathrm{OPT}} = \left(N\sigma^{-2}\int \psi(a)^2 p(a) da\right)^{1/2},$$

where $p(a) = \mathrm{E}\left(\phi(a - T^{1/2}\alpha_i/\sigma)\right)$ is the probability distribution function of $\widehat{z}_i$.

The first part of Proposition 1 provides a closed-form formula for $\mathrm{E}(\alpha_i|\mathcal{G})$. Under the stated conditions, $\widehat{\alpha}_i = \alpha_i + \bar{u}$, and in turn $\widehat{z}_i$, are sufficient statistics, so that $\mathrm{E}(\alpha_i|\mathcal{G}) = \mathrm{E}(\alpha_i|\widehat{z}_i)$. The latter expectation can then be evaluated with the help of the normality assumption on $u_{i,t}$. The second part of this proposition aims to simplify $S(\mathcal{G})$ using $\mathrm{E}(\alpha_i|\mathcal{G})$, according to Theorem 1, which needs an additional condition on the tail behavior of the cross-sectional distribution of α_i. The tail condition is stronger than Assumption 1(c).[14] It is a technical condition used to simplify the expression of $S(\mathcal{G})$.

To shed more light on this result, we compare this optimal Sharpe ratio S^{OPT} with $S^\star$ of Huberman (1982) using Example 1.

Corollary 2. *Suppose that the same assumptions as in Proposition 1 hold. In addition, we assume alpha follows (4) as in Example 1. Then we have $S^\star = \sigma^{-1}\mu(\rho N)^{1/2} + o_{\mathrm{P}}(1)$. Further, assuming that $\sigma^{-1}\mu(\rho N)^{1/2}$ does not vanish, then it holds that $S^{\mathrm{OPT}} \leq (1-\epsilon)\sigma^{-1}\mu(\rho N)^{1/2}$ for some $\epsilon > 0$, if and only if*

$$T^{1/2}\mu/\sigma - \sqrt{-2\log\rho} \lesssim 1. \tag{10}$$

Corollary 2 suggests that when $T^{1/2}\mu/\sigma$ is large, the constraint (10) is more likely violated, in which case $S^{\mathrm{OPT}} \approx \mathrm{E}(S^\star)$, that is, in the limit, the learning effect does not play any role, so

---

[14] Obviously, we have $\mathrm{E}(\max_i \alpha_i^2) \leq \mathrm{E}\left(\max_i \alpha_i^2 \mathbb{1}_{\{|\alpha_i|\geq c_N\}} + \max_i \alpha_i^2 \mathbb{1}_{\{|\alpha_i|< c_N\}}\right) \leq \mathrm{E}\left(\sum_i \alpha_i^2 \mathbb{1}_{\{|\alpha_i|\geq c_N\}}\right) + c_N^2 = o(1)$.

that arbitrageurs in this scenario achieve the same optimal Sharpe ratio as in Huberman (1982). Furthermore, the rareness parameter ρ does not make much difference if $T^{1/2}\mu/\sigma$ gets sufficiently large. That said, if ρ approaches to zero so fast to the extent that $\sqrt{-2\log\rho}$ dominates $T^{1/2}\mu/\sigma$, that is, alpha is extremely rare and sufficiently weak, the learning problem becomes rather challenging and hence S^{OPT} is dominated by $\mathrm{E}(S^{\star})$ in the limit, resulting in a strictly smaller Sharpe ratio than the infeasible Sharpe ratio in the classical case.

To give a concrete example of Corollary 2, consider an alternative DGP assumption as opposed to (5):[15]

$$\mu \sim N^{-\lambda} \quad \text{and} \quad \rho > 0 \quad \text{is fixed.} \tag{11}$$

In this scenario, $\mathrm{E}\left((S^{\star})^2\right) = \sigma^{-2}\mu^2\rho N \sim N^{1-2\lambda}$, which explodes unless $\lambda > 1/2$. If further assuming that $N/T \to \psi > 0$, then the left-hand-side of condition (10) is of order $N^{1/2-\lambda} \vee 1$, so that (10) holds if and only if $\lambda \geq 1/2$. Therefore, $\lambda < 1/2$ is not consistent with absence of (feasible) near arbitrage in that the infeasible Sharpe ratio explodes, while in the mean time the feasible Sharpe ratio approximately equals the infeasible Sharpe ratio (by Corollary 2) and hence explodes. If $\lambda > 1/2$, the infeasible Sharpe ratio (and hence the feasible one) vanishes, which does not seem like an economically plausible case. If we think that arbitrageur activity is required to prevent substantial mispricing, then a setting where mispricing disappears asymptotically even if the frictions faced by arbitrageurs are very large is not plausible. This suggests that under this DGP, the only economically plausible case with absence of near-arbitrage is $\lambda = 1/2$. That is, λ can be thought as determined in equilibrium, in which there are substantial asset demand distortions such that mispricing in the absence of arbitrageur action would be non-negligible asymptotically, and arbitrageurs are aggressive enough so that near-arbitrage opportunities do not exist asymptotically.

2.5 Constructing the Optimal Arbitrage Portfolio

In our previous discussion, we have shown in Theorem 1 that the optimal Sharpe ratio for any feasible strategy is bounded by $S(\mathcal{G})$. In Proposition 1, we have shown that $S(\mathcal{G}) \approx S^{\mathrm{OPT}}$ under additional assumptions. Corollary 2 further demonstrates that the optimal Sharpe ratio can vary with sequences of DGPs. In light of this, the optimal strategy may depend on the unobserved DGP as well.

In fact, we now show that arbitrageurs can construct a uniformly valid estimator of the optimal portfolio weights, which achieves S^{OPT} over a large class of data generating precesses, without perfect knowledge of the true DGP. This further implies that the no feasible near-arbitrage bound we develop is Sharp.

We demonstrate this in the setting where factors are latent but factor exposures are observable, since this is the case we analyze empirically. Moreover, we maintain the assumption that $\Sigma_u = \sigma^2\mathbb{I}_N$ for simplicity of the main text, leaving the general case to the appendix. Since all stocks share the

---

[15] It is easy to show that the setup (11) satisfies all assumptions of Proposition 1 for all fixed $\lambda > 0$.

common volatility σ^2, it does not affect the relative weights of different stocks in our portfolio. As a result, σ^2 becomes a nuisance parameter that does not influence the Sharpe ratio of the constructed portfolio.

Algorithm 1 (Constructing the Optimal Arbitrage Portfolio)**.**
Inputs: r_t, $t \in \mathcal{T} = \{t-T+1, \ldots, t\}$ *and* β.

S1. We split the observed sample $\mathcal{T}$ *into:*

$$S' = \{t - \lfloor T^{1/2} \rfloor + 1, \ldots, t\} \quad \textit{and} \quad S = \mathcal{T} - S',$$

and we construct cross-sectional regression estimates of alpha $\breve{\alpha}$ *and* $\breve{\alpha}'$*, and the t-statistics* $\breve{z}_i = |S|^{1/2}\breve{\alpha}_i$ *for each* $i = 1, 2, \ldots, N$*, using subsamples* S *and* S'*:*[16]

$$\breve{\alpha} = |S|^{-1} \sum_{s \in S} \mathbb{M}_\beta r_s, \quad \textit{and} \quad \breve{\alpha}' = |S'|^{-1} \sum_{s \in S'} \mathbb{M}_\beta r_s,$$

where $\mathbb{M}_\beta = \mathbb{I}_N - \beta(\beta^\intercal \beta)^{-1}\beta^\intercal$ *and* $\mathbb{I}_N$ *denotes the* $N \times N$ *identity matrix.*

S2. We choose the arbitrage portfolio weights as

$$\widehat{w}^{\mathrm{OPT}} = \mathbb{M}_\beta \breve{w}, \quad \textit{with} \quad \breve{w}_i = \begin{cases} \widehat{f}(\lfloor \breve{z}_i / k_N \rfloor), & |\breve{z}_i| \le k_N^{-2/3}, \\ \breve{\alpha}_i, & |\breve{z}_i| > k_N^{-2/3}. \end{cases}$$

For any set of integers (l)*, we choose* $k_N \sim (\log N)^{-1}$ *and define*

$$\widehat{f}(l) = \frac{1}{|B(l)|} \sum_{i \in B(l)} \breve{\alpha}'_i, \quad \textit{where} \quad B(l) = \{i \le N : l \le \breve{z}_i / k_N < l+1\}.$$

Outputs: $\widehat{w}^{\mathrm{OPT}}$.

As we have discussed in footnote 9, the optimal strategy in the case that arbitrageurs know the true DGP is given by

$$w^\star = \mathbb{M}_\beta \Sigma_u^{-1} \alpha. \tag{12}$$

Intuitively, part of the construction in (12), $\Sigma_u^{-1}\alpha$, is the optimal allocation to the ex-factor returns, $\alpha + u_t = r_t - \beta(\gamma + v_t)$, in a simple mean-variance analysis. Multiplying by $\mathbb{M}_\beta$ in (12) simply eliminates factor exposures in r_t. Correspondingly, Step S1 of Algorithm 1 provides estimates of α_i in two separate samples. Sample splitting is a convenient approach to avoiding dependence of these alpha estimates. Step S2 first constructs a nonparametric estimate of the function $f(\alpha) = \mathrm{E}(\alpha | \{r_s, \beta\}_{s \in \mathcal{T}})$, with which the optimal weights on ex-factor returns are constructed as $\breve{w}$. This, in turn, leads to the optimal weight estimates, $\widehat{w}^{\mathrm{OPT}}$, on original input asset returns.

---

[16] For any set S, we use $|S|$ to denote the number of elements in S.

An essential step towards uniform optimality is the construction of $\breve{w}$, in which we deal with strong and weak signals separately. The strong signals (those with t-statistics greater than $k_N^{-2/3}$) are singled out, for which we can obtain relatively precise estimates of their optimal weights. For weaker signals, we consolidate information therein to obtain an estimate of the conditional expectation of their signal strength, using which we obtain their optimal portfolio weights. This strategy outperforms the alternatives, which either directly use estimated alphas as if these estimates are not susceptible to errors even when they are rather weak, or simply ignore the contribution of all weaker signals.

The following theorem demonstrates the optimality of $\widehat{w}^{\mathrm{OPT}}$:

Theorem 2. *Let $\mathbb{P}$ denote the collection of all data-generating processes under which r_t follows (1), and Assumption 1 holds. In addition, suppose that $N^d \leq T$ for some $d > 0$. We denote the Sharpe ratio generated by the portfolio strategy $\widehat{w}^{\mathrm{OPT}}$ as $\widehat{S}^{\mathrm{OPT}} := \mathrm{E}(r_{t+1}^{\intercal}\widehat{w}^{\mathrm{OPT}}|\mathcal{F}_t)/\mathrm{Var}(r_{t+1}^{\intercal}\widehat{w}^{\mathrm{OPT}}|\mathcal{F}_t)^{1/2}$. Then it holds that $\widehat{w}^{\mathrm{OPT}}$ achieves, asymptotically, the upper bound S^{OPT} uniformly over all sequences of data-generating processes. That is, for any $\epsilon > 0$,*

$$\lim_{N,T\to\infty} \sup_{\mathrm{P}\in\mathbb{P}} \mathrm{P}\Big(\big|\widehat{S}^{\mathrm{OPT}} - S^{\mathrm{OPT}}\big| \geq \epsilon S^{\mathrm{OPT}} + \epsilon\Big) = 0.$$

Theorem 2 concludes that in the context of a linear factor model, arbitrageurs can construct this strategy, without any knowledge besides past returns and risk exposures (beta), to achieve the maximal Sharpe ratio over all feasible trading strategies that have zero exposure to factor risks. This Sharpe ratio precisely characterizes the limit of feasible arbitrages in economic terms. Its gap to $\mathrm{E}(S^{\star})$, the Sharpe ratio under rational expectation, is determined by the difficulty of the learning problem.

With Theorem 2, we establish the necessity for the no near-arbitrage condition given by (9).

Corollary 3. *Suppose the same assumptions as in Theorem 2 hold. The portfolio weights by $\widehat{w}^{\mathrm{OPT}}$ yields a near-arbitrage strategy under any sequences of data-generating processes for which condition (9) does not hold.*

We have shown that arbitrageurs can construct an optimal strategy that realizes S^{OPT}. Now suppose that the equilibrium "cost" of implementing an arbitrage is C in an economy with statistical limit of arbitrage. In equilibrium, $S^{\mathrm{OPT}} = C$, otherwise arbitrageurs can trade until it is no longer profitable to do so. We can thereby interpret $\widehat{S}^{\mathrm{OPT}}$ as an empirical estimate of the arbitrage cost, which we will estimate empirically.

2.6 Estimating Optimal Infeasible Sharpe Ratio

We are also interested in estimating the optimal infeasible Sharpe ratio, $S^{\star}$, which can be perceived as the optimal Sharpe ratio from an outside econometrician's point of view. Existing literature on testing APT often construct test statistics in the spirit of Gibbons et al. (1989), which are effectively based on $S^{\star}$, see, e.g., Pesaran and Yamagata (2017) and Fan et al. (2015). While such tests are powerful and may lead to discoveries of alpha signals, they are not relevant for arbitrageurs in that arbitrageurs may not construct a feasible portfolio to capture these statistical discoveries.

We now construct an estimator for $(S^{\star})^2$. The most natural estimator of the infeasible Sharpe ratio is given by:

$$(\widetilde{S}^{\star})^2 = \bar{r}^{\intercal}\mathbb{M}_{\beta}\widehat{\Sigma}_u^{-1}\mathbb{M}_{\beta}\bar{r}, \tag{13}$$

where $\bar{r} = T^{-1}\sum_{t\in\mathcal{T}} r_t$, $\widehat{\sigma}_i^2 = T^{-1}\sum_{t\in\mathcal{T}}(r_{i,t}-\bar{r}_i)^2$, and $\widehat{\Sigma}_u = \operatorname{diag}(\widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \ldots, \widehat{\sigma}_N^2)$.

Unfortunately, this estimator has a non-vanishing asymptotic bias for certain data generating processes we consider, as we will show later. To fix this issue, we propose a new estimator that is uniformly consistent:

$$(\widehat{S}^{\star})^2 = \bar{r}^{\intercal}\mathbb{M}_{\beta}\widehat{\Sigma}_u^{-1}\mathbb{M}_{\beta}\bar{r} - T^{-2}\sum_{t\in\mathcal{T}} r_t^{\intercal}\mathbb{M}_{\beta}\widehat{\Sigma}_u^{-1}\mathbb{M}_{\beta}r_t. \tag{14}$$

The second estimator again takes the form of a summation over individual squared Sharpe ratios, but it eliminates the term that will be dominated by the estimation bias under some data generating processes. The next proposition summarizes the asymptotic properties of both estimators.

Proposition 2. *Suppose r_t follows (1), and Assumptions 1 and A2 hold. Then we have*

$$\left|\widehat{S}^{\star} - S^{\star}\right| / (1+S^{\star}) = o_{\mathrm{P}}(T^{-1/4}), \quad \left|\widetilde{S}^{\star} - \left((S^{\star})^2 + NT^{-1}\right)^{1/2}\right| \Big/ (1+S^{\star}) = o_{\mathrm{P}}(T^{-1/4}).$$

As shown by this proposition, the estimation error is relative when $S^{\star}$ dominates one asymptotically, and absolute if $S^{\star}$ is dominated by one. This is necessary because we simultaneously consider a large class of models, some of which have an exploding or a shrinking $S^{\star}$, and because $S^{\star}$ also influences the convergence rate.

2.7 Alternative Strategies for Arbitrage Portfolios

Algorithm 1 suggests a relatively sophisticated procedure that distinguishes weaker and strong signals using t-statistics before constructing, separately, the optimal weights for these signals. In this section, we study several alternative methods, neither of which can achieve optimality uniformly across all DGPs we consider, but they are simpler and somewhat prevalent in practice. The contrast among these strategies helps illustrate their pros and cons in different scenarios.

2.7.1 Cross-Sectional Regression

The conventional approach to estimating alphas is through the cross-sectional regression:

$$\widehat{\alpha} = (\beta^{\intercal}\beta)^{-1}\beta^{\intercal}\bar{r},$$

with which the arbitrage portfolio weights can be constructed directly as:

$$\widehat{w}^{\mathrm{CSR}} = \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \widehat{\alpha}. \tag{15}$$

This choice of portfolio weight is the sample analog of the optimal weight given by (12). Proposition 3 below compares the asymptotic behavior of the expected Sharpe ratio of this arbitrage portfolio with the optimal Sharpe ratio, S^{OPT}, in Example 1. For convenience, we adopt a simplified volatility estimator: $\widehat{\Sigma}_u = \widehat{\sigma}^2 \mathbb{I}_N$, where $\widehat{\sigma}^2$ is averaged over all volatility estimates, because in this example, all assets share the same volatility. This further simplifies the analysis because the scaling factor, $\widehat{\sigma}^2$, is cancelled out, and hence does not play any role in this portfolio's Sharpe ratio.

Proposition 3. *Suppose that r_t follows (1) and Assumption 1 holds. In addition, we assume alpha follows (4) as in Example 1. The Sharpe ratio of the arbitrage portfolio, whose weights are given by $\widehat{w}^{\mathrm{CSR}} = \widehat{\sigma}^{-2} \mathbb{M}_\beta \widehat{\alpha}$, satisfies $\widehat{S}^{\mathrm{CSR}} - S^{\mathrm{CSR}} = o_{\mathrm{P}}(1)$, where*

$$\widehat{S}^{\mathrm{CSR}} = \mathrm{E}(r_{t+1}^\intercal \widehat{w}^{\mathrm{CSR}} | \mathcal{F}_t) / \mathrm{Var}(r_{t+1}^\intercal \widehat{w}^{\mathrm{CSR}} | \mathcal{F}_t)^{1/2}, \qquad S^{\mathrm{CSR}} = \frac{N^{1/2} \rho \mu^2 \sigma^{-2}}{(T^{-1} + \rho \mu^2 \sigma^{-2})^{1/2}}.$$

Further, assuming S^{OPT} does not vanish, then as $N, T \to \infty$, we have $S^{\mathrm{CSR}} \leq (1-\epsilon) S^{\mathrm{OPT}}$ for some fixed $\epsilon > 0$, if and only if

$$C \leq T \mu^2 \sigma^{-2} \leq C' \rho^{-1}. \tag{16}$$

for some constants C and C'.

Proposition 3 suggests that arbitrageurs using this cross-sectional regression strategy cannot always achieve the optimal feasible Sharpe ratio. In fact, this strategy is dominated by the optimal strategy when signals are both very strong ($C \leq T\mu^2\sigma^{-2}$) and there are not too many strong signals ($T\mu^2\sigma^{-2} \leq C'\rho^{-1}$). Intuitively, the CSR approach treats all signals equally, without distinguishing insignificant signals from the significant ones. Consequently, even fake signals could get non-zero weights allocated. This strategy thereby works well when the strong signals are abundant (i.e., ρ is relatively large) or when all signals are weak (so that they do not differ too much from fake ones). The latter case is more interesting, as it also suggests that simply ignoring weaker signals is not optimal.

The CSR approach is a simple benchmark as it does not rely on any advanced statistical techniques to detect signals or distinguish their strength. The strategy we discuss next is more advanced, in that it controls false discoveries among selected strong signals using the B-H procedure.

2.7.2 False Discovery Rate Control

From the statistical point of view, we can formalize the search for alpha as a multiple testing problem. Say, there are N assets potentially with nonzero α, and for each i, we can define a null hypothesis:

$\mathbb{H}_0^i : \alpha_i = 0$, so that alpha detection becomes a multiple testing problem. With multiple testing comes the concern of data snooping, meaning that a large fraction of tests that appear positive are in fact due to chance. One sensible approach is to control the false discovery rate (FDR), instead of the size of individual tests, a proposal advocated by Barras et al. (2010), Bajgrowicz and Scaillet (2012), and Harvey et al. (2016) in different asset pricing contexts.

The B-H procedure proposed by Benjamini and Hochberg (1995) is often adopted to control FDR in multiple testing problems. Giglio et al. (2021) have proved its validity in a general factor model setting for alpha detection. Below we describe the algorithm for constructing alpha estimates, which will be used as inputs to the construction of an arbitrage portfolio.

Algorithm 2 (The B-H based Alpha Selection)**.** *Let $\widehat{\alpha}$ be the estimator of α via the cross-sectional regression, and $\{p_i : i = 1, \dots, N\}$ be the p-values of the corresponding t-test statistics.*

- *S1. Sort in ascending order the collection of p-values, with the sorted p-values given by $p_{(1)} \leq \dots \leq p_{(N)}$.*
- *S2. For $i = 1, \dots, N$, reject $\mathbb{H}_0^i : \alpha_i = 0$, if $p_i \leq p_{(\widehat{k})}$, where $\widehat{k} = \max\{i \leq N : p_{(i)} \leq \tau i/N\}$, for any pre-determined level τ, say, 5%.*

We can then adjust our alpha estimates using

$$\widehat{\alpha}_i^{\mathrm{BH}}(\tau) = \widehat{\alpha}_i \mathbb{1}_{\{p_i \leq p_{(\widehat{k})}\}}. \tag{17}$$

The B-H procedure guarantees (in expectation) that at least a fraction $(1 - \tau)$ of selected assets have nonzero alphas, regardless of the actual percentage of alphas in the data generating process. Effectively, it imposes a hard-thresholding procedure on the alpha estimates, replacing less significant alphas by zero. Similar to (15), the optimal portfolio weights are thus given by:

$$\widehat{w}^{\mathrm{BH}} = \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \widehat{\alpha}^{\mathrm{BH}}(\tau). \tag{18}$$

Controlling the false discovery rate on top of the CSR estimates is intuitively appealing, but doing so incurs a potential loss of power and hence less investment opportunities. Our focus is on optimal portfolio construction instead of false discovery control. The next proposition shows that in the context of Example 1, arbitrageurs who adopt the B-H based alpha estimator cannot achieve optimal portfolio for a large class of DGP sequences. In fact, we can prove a richer result. Even if arbitrageurs knew that all assets with non-zero alphas have an equally strong alpha with the same idiosyncratic volatility (which is true in this example), and that they adopted the following estimates for alpha instead of (17) to exploit this fact:

$$\overline{\alpha}_i^{\mathrm{BH}}(\tau) = sgn(\widehat{\alpha}_i) \overline{\alpha} \mathbb{1}_{\{p_i \leq p_{(\widehat{k})}\}}, \quad \overline{\alpha} = \sum_i |\widehat{\alpha}_i| \mathbb{1}_{\{p_i \leq p_{(\widehat{k})}\}} / \sum_i \mathbb{1}_{\{p_i \leq p_{(\widehat{k})}\}}, \tag{19}$$

still they would not be able to achieve the optimal performance.

Proposition 4. *Suppose that* r_t *follows (1) and Assumption 1 holds. In addition, we assume alpha follows (4) as in Example 1. The Sharpe ratio of the arbitrage portfolio with weights given by* $\widehat{w}^{\mathrm{BH}} = \widehat{\sigma}^{-2}\mathbb{M}_\beta\widehat{\alpha}^{\mathrm{BH}}(\tau)$ *and* $\overline{w}^{\mathrm{BH}} = \widehat{\sigma}^{-2}\mathbb{M}_\beta\overline{\alpha}^{\mathrm{BH}}(\tau)$ *satisfies* $\overline{S}^{\mathrm{BH}} - \sqrt{1-\tau}S^{\mathrm{BH}} = o_{\mathrm{P}}(1)$ *and* $\widehat{S}^{\mathrm{BH}} \leq S^{\mathrm{BH}} + o_{\mathrm{P}}(1)$*, where*[17]

$$\widehat{S}^{\mathrm{BH}} = \mathrm{E}(r_{t+1}^\intercal \widehat{w}^{\mathrm{BH}}|\mathcal{F}_t)/\mathrm{Var}(r_{t+1}^\intercal \widehat{w}^{\mathrm{BH}}|\mathcal{F}_t)^{1/2}, \quad \overline{S}^{\mathrm{BH}} = \mathrm{E}(r_{t+1}^\intercal \overline{w}^{\mathrm{BH}}|\mathcal{F}_t)/\mathrm{Var}(r_{t+1}^\intercal \overline{w}^{\mathrm{BH}}|\mathcal{F}_t)^{1/2}$$

are corresponding Sharpe ratios, and

$$S^{\mathrm{BH}} = \mu\sigma^{-1}\sqrt{\rho N \Phi(T^{1/2}\mu/\sigma - z^*)},$$

where $\Phi(\cdot)$ *is the normal cumulative distribution function, and* z^* *is the positive solution of the equation*

$$2(1-\tau(1-\rho))\Phi(-z) = \tau\rho\Phi(T^{1/2}\mu/\sigma - z). \tag{20}$$

Suppose further that S^{OPT} *does not vanish, and that* $CN^{-1+d} \leq \rho \leq CN^{-d}$ *for some fixed* $d > 0$*. Then it follows that, as* $N, T \to \infty$*,* $S^{\mathrm{BH}} \leq (1-\epsilon)S^{\mathrm{OPT}}$ *for some fixed* $\epsilon > 0$*, if and only if, for some fixed* $\epsilon' > 0$*,*

$$T^{1/2}\mu/\sigma \leq (1-\epsilon')\sqrt{-\lambda_\tau \log\rho}, \tag{21}$$

where $\lambda_\tau \in (2/3, 2)$ *only depends on* τ *and* $\lambda_\tau \to 2$ *as* $\tau \to 0$*.*

As Proposition 4 shows, (21) indicates that if the signal-to-noise ratio is not sufficiently strong, the B-H procedure is unlikely to reach S^{OPT}. This is because it ignores many individually impotent signals, which would hurt the portfolio performance, even though B-H remains a preferable approach to selecting truly significant alphas and controlling false discoveries. In contrast, the optimal arbitrage portfolio exploits information embedded in all alpha estimates, including false positives, beyond the set of significant ones selected via B-H procedure. This result also demonstrates a clear distinction between two objectives: alpha testing and portfolio construction, the objectives of which do not always align.

The CSR and the B-H approaches represent two typical strategies in practice. The former trades all signals without distinguishing their strength, whereas the latter only trades the stronger signals. Neither approach always achieves optimality.

2.7.3 Shrinkage Approaches

The analysis above suggests that we can construct the optimal portfolio out of the ex-factor returns, while imposing regularization on portfolio weights, before rewriting the regularized portfolio weights in terms of raw returns (multiplying by $\mathbb{M}_\beta$). Regularizing portfolio weights amounts to imposing priors directly on the alpha estimates. To see this, we adopt a shrinkage approach, when constructing

---

[17] If $\widehat{w}^{\mathrm{BH}} = 0$, i.e., no asset is selected, we set $\widehat{S}^{\mathrm{BH}} = 0$ by convention.

arbitrage portfolios on residual returns:

$$\arg\max_{w}\{w^\intercal \widehat{\alpha} - \frac{1}{2} w^\intercal \widehat{\Sigma}_u w - p_\lambda(w)\},$$

where $p_\lambda(w) = \lambda \|w\|_1$ or $\lambda \|w\|_2^2$, for some $\lambda > 0$. Since $\widehat{\Sigma}_u$ is diagonal, the closed-form solution is $\psi_q(\widehat{\alpha}, \widehat{\Sigma}_u, \lambda)$, where $q = 1$ corresponds to the LASSO penalty and $q = 2$ the ridge, and for $i = 1, 2, \ldots, N$,

$$\left(\psi_1(\widehat{\alpha}, \widehat{\Sigma}_u, \lambda)\right)_i = (\widehat{\sigma}_i)^{-2} sgn(\widehat{\alpha}_i)(|\widehat{\alpha}_i| - \lambda)_+, \quad \left(\psi_2(\widehat{\alpha}, \widehat{\Sigma}_u, \lambda)\right)_i = \left((\widehat{\sigma}_i)^2 + \lambda\right)^{-1} \widehat{\alpha}_i.$$

This leads to the optimal portfolio weight on r_t:[18]

$$\widehat{w}^q = \mathbb{M}_\beta \psi_q(\widehat{\alpha}, \widehat{\Sigma}_u, \lambda), \quad q = 1, 2.$$

Depending on the magnitude of λ, the LASSO approach replaces all smaller signals by zero and shrinks the larger signals by λ in absolute terms. In other words, the LASSO approach is the soft-thresholding alternative to the B-H method. In contrast, the ridge penalty shrinks all signals proportionally with a shrinkage factor depending on $\widehat{\sigma}_i^2$. Like the above analysis, when specialized to example (1), we can adopt $\widehat{\Sigma}_u = \widehat{\sigma}^2 \mathbb{I}_N$, in which case ridge becomes equivalent to CSR! This "embedded" shrinkage effect of CSR explains why it performs well in the case of small signals. Proposition 5, along with Proposition 3, demonstrate that neither LASSO nor ridge can achieve optimal Sharpe ratio in all DGPs even with the optimal tuning parameter λ.

Proposition 5. *Suppose that r_t follows (1) and Assumption 1 holds. In addition, we assume alpha follows (4) as in Example 1. The Sharpe ratio of the arbitrage portfolio with weights given by $\widehat{w}^q$, denoted as $\widehat{S}^q$ for $q = 1, 2$, satisfies $\widehat{S}^1 - S^{\text{LASSO}} = o_{\text{P}}(1)$ and $\widehat{S}^2 - S^{\text{CSR}} = o_{\text{P}}(1)$, where*

$$S^{\text{LASSO}} = \rho\mu\sigma^{-1}N^{1/2} \frac{\int_{-\infty}^{\infty} sgn(x)(T^{-1/2}\sigma|x| - \lambda)_+ \phi(T^{1/2}\sigma^{-1}\mu - x)dx}{\sqrt{\int_{-\infty}^{\infty} \left((T^{-1/2}\sigma|x| - \lambda)_+\right)^2 \left((1-\rho)\phi(x) + \rho\phi(T^{1/2}\sigma^{-1}\mu - x)\right)dx}},$$

and S^{CSR} is defined in Proposition 3.

Suppose further that S^{OPT} does not vanish, and that $CN^{-1+d} \le \rho \le CN^{-d}$ for some fixed $d > 0$. Then it follows that, as $N, T \to \infty$, $S^{\text{LASSO}} \le (1-\epsilon)S^{\text{OPT}}$ for some fixed $\epsilon > 0$ under all sequences of λ, if and only if

$$T^{1/2}\mu/\sigma \ge C, \quad \text{and} \quad \frac{T\mu^2/\sigma^2 + 2\log\rho}{\sqrt{-\log\rho}} \le C.$$

---

[18] An alternative strategy is to impose sparsity directly on the portfolio weights with respect to raw returns. While this approach might be appealing from the transaction cost point of view, it does not associate with an explicit prior on alpha, hence is more difficult to interpret.

3 Simulation Evidence

This section demonstrates the empirical relevance of our theory via simulations and examines the finite sample performance of the proposed portfolio strategies.

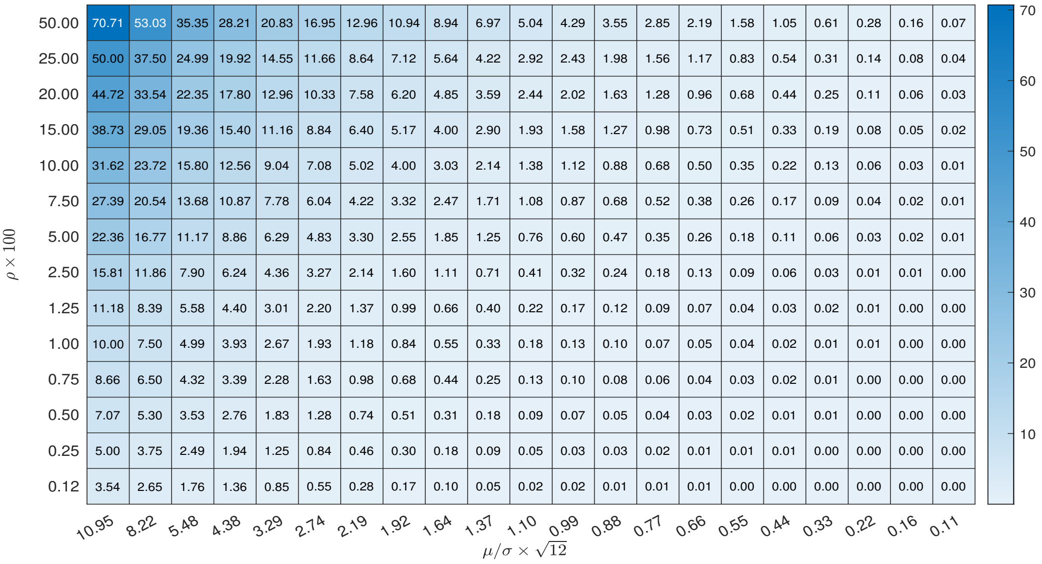
3.1 Numerical Illustration of Theoretical Predictions

We start by examining the theoretical predictions of Corollary 2 and Propositions 3 - 5. For simplicity and clarity, we simulate a one-factor (CAPM) model of returns given by (1). We choose the factor risk premium as 5% per year and set the annualized volatility at 25%. We model the cross-section of betas using a normal distribution with mean 1 and variance 1. Since we focus on the arbitrage portfolio, the parameters about the factor component (including the number of factors) are inconsequential, because factors, if any, are eliminated by $\mathbb{M}_{\beta}$ in the first step when constructing these trading strategies. In addition, we adopt model (4) in Example 1 for the cross-sectional distribution of alpha, and fix the idiosyncratic volatilities of all assets at σ, since it is α/σ that determines the signal strength and that there is no need of varying both α and σ in the cross section.

Figure 1 reports the Sharpe ratio, S^{OPT}, of optimal feasible arbitrage portfolios for a range of μ/σ and ρ values in the case of $N = 1,000$ and $T = 20$ years. Recall that according to model (4), a ρ percentage of assets have alphas with a Sharpe ratio μ/σ. That is, ρ characterizes the rareness of the alpha signal, whereas μ/σ captures its strength. We intentionally choose a wide range of μ/σ (with annualized Sharpe ratios from 0.11 to 10.95) and ρ (from 0.12% to 50%) to shed light on the dependence landscape of Sharpe ratios on signal weakness and rareness, despite that some of the resulting portfolio Sharpe ratios (the top left conner of Figure 1) are unrealistically high. Note that when $\mu/\sigma \times \sqrt{12}$ hits 0.44, its corresponding t-statistic based on a 20-year sample exceeds 1.96, the typical t-hurdle for a standard student-t test.

The pattern of Sharpe ratios agrees with our intuition and theoretical predictions. For any fixed ρ, as the alpha signal weakens (i.e., μ/σ decreases), the optimal Sharpe ratio drops. The same is true if we decrease the signal count (i.e., ρ vanishes), for any fixed value of μ/σ. The arbitrageur's learning problem is the easiest when signal is strong and count is large (top left conner), and the most challenging towards the right bottom corner, where the optimal Sharpe ratios drop to near 0.

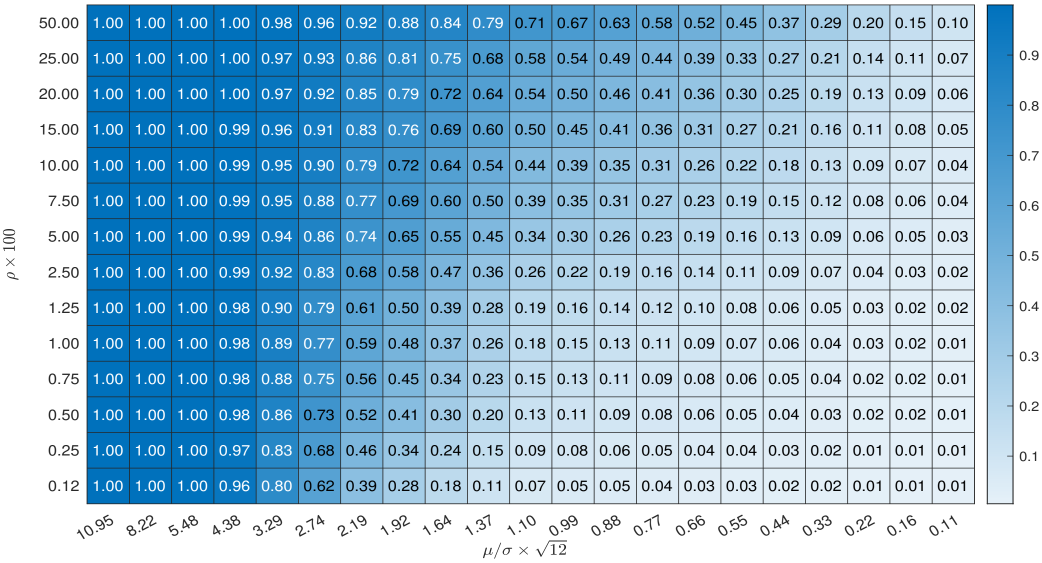
The reported Sharpe ratios on Figure 1 are only a fraction of the corresponding (infeasible) Sharpe ratios, $S^{\star} = \sqrt{\alpha^{\intercal}(\Sigma_u)^{-1}\alpha} = \mu/\sigma\sqrt{\rho N}$, as shown by Figure 2. The pattern we see from Figure 2 agrees with theoretical predictions of Corollary 2. When the annualized Sharpe ratio $\mu/\sigma \times \sqrt{12}$ is larger than 2.74, regardless of the values of ρ, the signal-to-noise ratio of the learning problem is sufficiently strong that the statistical limit to arbitrage does not matter much, and hence $S^{\mathrm{OPT}}/S^{\star}$ is close to 1. Nonetheless, this regime is irrelevant in practice, since it is mostly associated with unrealistically high Sharpe ratios (see Figure 1). In contrast, as μ/σ diminishes, the gap between $S^{\star}$ and S^{OPT} widens. In almost all empirically relevant scenarios, $S^{\star}$ is largely exaggerated.



$\rho \times 100$	10.95	8.22	5.48	4.38	3.29	2.74	2.19	1.92	1.64	1.37	1.10	0.99	0.88	0.77	0.66	0.55	0.44	0.33	0.22	0.16	0.11
50.00	70.71	53.03	35.35	28.21	20.83	16.95	12.96	10.94	8.94	6.97	5.04	4.29	3.55	2.85	2.19	1.58	1.05	0.61	0.28	0.16	0.07
25.00	50.00	37.50	24.99	19.92	14.55	11.66	8.64	7.12	5.64	4.22	2.92	2.43	1.98	1.56	1.17	0.83	0.54	0.31	0.14	0.08	0.04
20.00	44.72	33.54	22.35	17.80	12.96	10.33	7.58	6.20	4.85	3.59	2.44	2.02	1.63	1.28	0.96	0.68	0.44	0.25	0.11	0.06	0.03
15.00	38.73	29.05	19.36	15.40	11.16	8.84	6.40	5.17	4.00	2.90	1.93	1.58	1.27	0.98	0.73	0.51	0.33	0.19	0.08	0.05	0.02
10.00	31.62	23.72	15.80	12.56	9.04	7.08	5.02	4.00	3.03	2.14	1.38	1.12	0.88	0.68	0.50	0.35	0.22	0.13	0.06	0.03	0.01
7.50	27.39	20.54	13.68	10.87	7.78	6.04	4.22	3.32	2.47	1.71	1.08	0.87	0.68	0.52	0.38	0.26	0.17	0.09	0.04	0.02	0.01
5.00	22.36	16.77	11.17	8.86	6.29	4.83	3.30	2.55	1.85	1.25	0.76	0.60	0.47	0.35	0.26	0.18	0.11	0.06	0.03	0.02	0.01
2.50	15.81	11.86	7.90	6.24	4.36	3.27	2.14	1.60	1.11	0.71	0.41	0.32	0.24	0.18	0.13	0.09	0.06	0.03	0.01	0.01	0.00
1.25	11.18	8.39	5.58	4.40	3.01	2.20	1.37	0.99	0.66	0.40	0.22	0.17	0.12	0.09	0.07	0.04	0.03	0.02	0.01	0.00	0.00
1.00	10.00	7.50	4.99	3.93	2.67	1.93	1.18	0.84	0.55	0.33	0.18	0.13	0.10	0.07	0.05	0.04	0.02	0.01	0.01	0.00	0.00
0.75	8.66	6.50	4.32	3.39	2.28	1.63	0.98	0.68	0.44	0.25	0.13	0.10	0.08	0.06	0.04	0.03	0.02	0.01	0.00	0.00	0.00
0.50	7.07	5.30	3.53	2.76	1.83	1.28	0.74	0.51	0.31	0.18	0.09	0.07	0.05	0.04	0.03	0.02	0.01	0.01	0.00	0.00	0.00
0.25	5.00	3.75	2.49	1.94	1.25	0.84	0.46	0.30	0.18	0.09	0.05	0.03	0.03	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.00
0.12	3.54	2.65	1.76	1.36	0.85	0.55	0.28	0.17	0.10	0.05	0.02	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00

Figure 1: Optimal Sharpe Ratios (S^{OPT}) of Feasible Arbitrage Portfolios

Note: The figure reports optimal Sharpe ratios of feasible arbitrage portfolios in model (4), in which a $100 \times \rho\%$ of assets have alphas that correspond to an annualized Sharpe ratio $\mu/\sigma \times \sqrt{12}$.



$\rho \times 100$	10.95	8.22	5.48	4.38	3.29	2.74	2.19	1.92	1.64	1.37	1.10	0.99	0.88	0.77	0.66	0.55	0.44	0.33	0.22	0.16	0.11
50.00	1.00	1.00	1.00	1.00	0.98	0.96	0.92	0.88	0.84	0.79	0.71	0.67	0.63	0.58	0.52	0.45	0.37	0.29	0.20	0.15	0.10
25.00	1.00	1.00	1.00	1.00	0.97	0.93	0.86	0.81	0.75	0.68	0.58	0.54	0.49	0.44	0.39	0.33	0.27	0.21	0.14	0.11	0.07
20.00	1.00	1.00	1.00	1.00	0.97	0.92	0.85	0.79	0.72	0.64	0.54	0.50	0.46	0.41	0.36	0.30	0.25	0.19	0.13	0.09	0.06
15.00	1.00	1.00	1.00	0.99	0.96	0.91	0.83	0.76	0.69	0.60	0.50	0.45	0.41	0.36	0.31	0.27	0.21	0.16	0.11	0.08	0.05
10.00	1.00	1.00	1.00	0.99	0.95	0.90	0.79	0.72	0.64	0.54	0.44	0.39	0.35	0.31	0.26	0.22	0.18	0.13	0.09	0.07	0.04
7.50	1.00	1.00	1.00	0.99	0.95	0.88	0.77	0.69	0.60	0.50	0.39	0.35	0.31	0.27	0.23	0.19	0.15	0.12	0.08	0.06	0.04
5.00	1.00	1.00	1.00	0.99	0.94	0.86	0.74	0.65	0.55	0.45	0.34	0.30	0.26	0.23	0.19	0.16	0.13	0.09	0.06	0.05	0.03
2.50	1.00	1.00	1.00	0.99	0.92	0.83	0.68	0.58	0.47	0.36	0.26	0.22	0.19	0.16	0.14	0.11	0.09	0.07	0.04	0.03	0.02
1.25	1.00	1.00	1.00	0.98	0.90	0.79	0.61	0.50	0.39	0.28	0.19	0.16	0.14	0.12	0.10	0.08	0.06	0.05	0.03	0.02	0.02
1.00	1.00	1.00	1.00	0.98	0.89	0.77	0.59	0.48	0.37	0.26	0.18	0.15	0.13	0.11	0.09	0.07	0.06	0.04	0.03	0.02	0.01
0.75	1.00	1.00	1.00	0.98	0.88	0.75	0.56	0.45	0.34	0.23	0.15	0.13	0.11	0.09	0.08	0.06	0.05	0.04	0.02	0.02	0.01
0.50	1.00	1.00	1.00	0.98	0.86	0.73	0.52	0.41	0.30	0.20	0.13	0.11	0.09	0.08	0.06	0.05	0.04	0.03	0.02	0.02	0.01
0.25	1.00	1.00	1.00	0.97	0.83	0.68	0.46	0.34	0.24	0.15	0.09	0.08	0.06	0.05	0.04	0.04	0.03	0.02	0.01	0.01	0.01
0.12	1.00	1.00	1.00	0.96	0.80	0.62	0.39	0.28	0.18	0.11	0.07	0.05	0.05	0.04	0.03	0.03	0.02	0.02	0.01	0.01	0.01

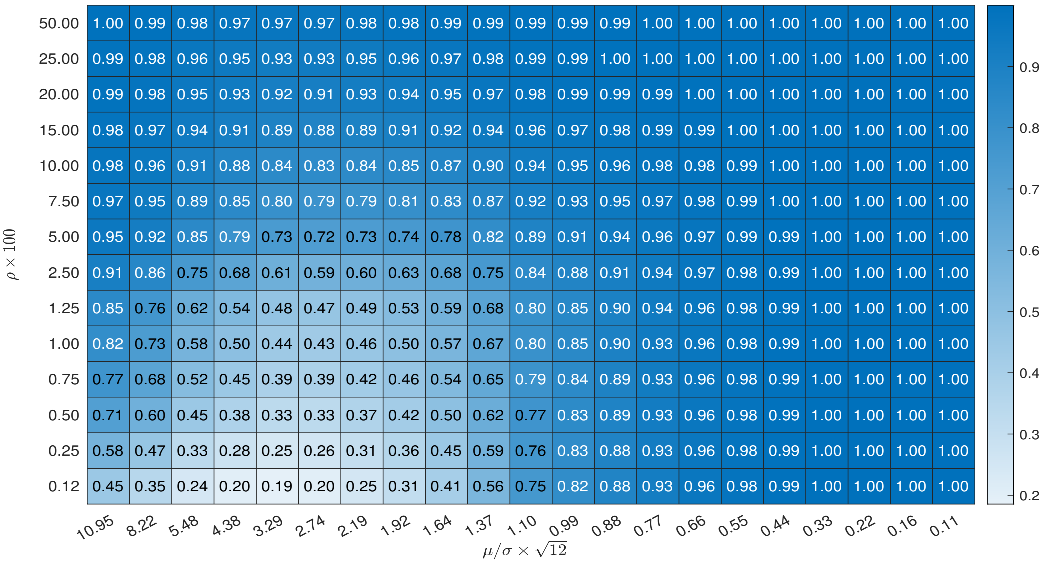
Figure 2: Ratios between S^{OPT} and $S^{\star}$

Note: The figure reports the ratios of optimal Sharpe ratios between feasible and infeasible arbitrage portfolios. The simulation setting is based on model (4), in which a $100 \times \rho\%$ of assets have alphas that correspond to an annualized Sharpe ratio $\mu/\sigma \times \sqrt{12}$.

We now turn to the comparison of Sharpe ratios of optimal feasible arbitrage portfolios with those achieved by alternative strategies. Figure 3 compares with the cross-section regression approach in

Section 2.7.1, Figure 4 with the B-H based procedure given by Section 2.7.2, and Figure 5 with LASSO given by Section 2.7.3, respectively.

According to Proposition 3, the optimal portfolio dominates the cross-sectional regression based portfolio if (16) holds. This dominance regime is bounded by a vertical line (as implied by the first inequality) and a cubic curve (as implied by the second inequality), which is visible from Figure 3 (black numbers on the heatmap). As $\mu/\sigma \times \sqrt{12}$ approaches 1.0 (a vertical line) from the right or the upper left corner, the gap between the two Sharpe ratios shrinks. Intuitively, when a large number of signals are clearly separable from the null (top left corner), the statistical inference becomes simpler so that the cross-sectional regression estimator of α is sufficient for building optimal portfolios. On the other hand, as the signal strength vanishes (the right vertical boundary), the relative performance of the regression approach improves because it is equivalent to a ridge penalized regression that works well when all signals are weak and almost indistinguishable from noise. Figure 1 shows that the DGPs with respect to parameters for which the cross-sectional regression approach is strongly dominated by our optimal strategy are associated with realistic Sharpe ratios.

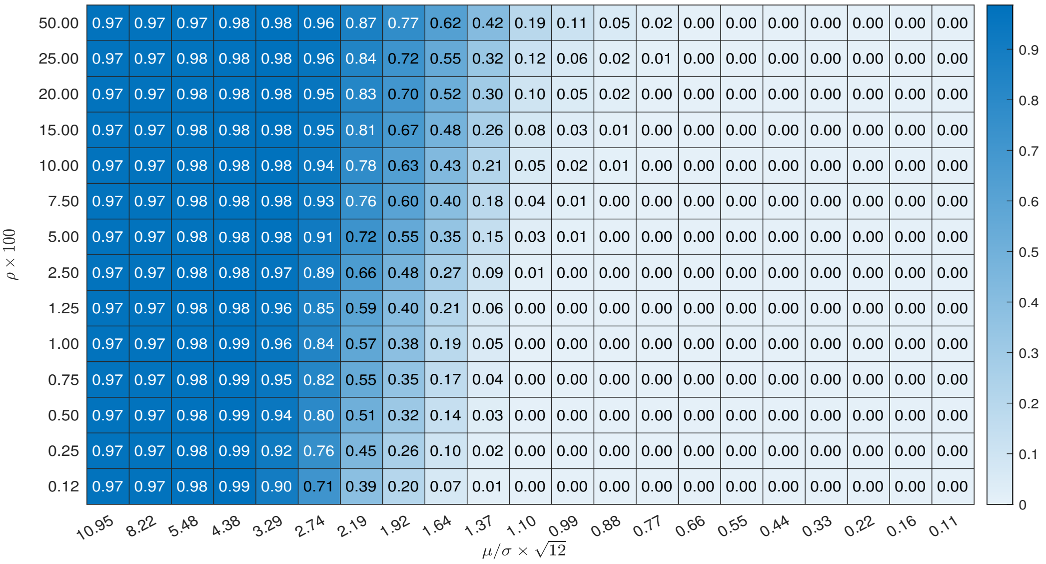


$\rho \times 100$	10.95	8.22	5.48	4.38	3.29	2.74	2.19	1.92	1.64	1.37	1.10	0.99	0.88	0.77	0.66	0.55	0.44	0.33	0.22	0.16	0.11
50.00	1.00	0.99	0.98	0.97	0.97	0.97	0.98	0.98	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
25.00	0.99	0.98	0.96	0.95	0.93	0.93	0.95	0.96	0.97	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20.00	0.99	0.98	0.95	0.93	0.92	0.91	0.93	0.94	0.95	0.97	0.98	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
15.00	0.98	0.97	0.94	0.91	0.89	0.88	0.89	0.91	0.92	0.94	0.96	0.97	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
10.00	0.98	0.96	0.91	0.88	0.84	0.83	0.84	0.85	0.87	0.90	0.94	0.95	0.96	0.98	0.98	0.99	1.00	1.00	1.00	1.00	1.00
7.50	0.97	0.95	0.89	0.85	0.80	0.79	0.79	0.81	0.83	0.87	0.92	0.93	0.95	0.97	0.98	0.99	1.00	1.00	1.00	1.00	1.00
5.00	0.95	0.92	0.85	0.79	0.73	0.72	0.73	0.74	0.78	0.82	0.89	0.91	0.94	0.96	0.97	0.99	0.99	1.00	1.00	1.00	1.00
2.50	0.91	0.86	0.75	0.68	0.61	0.59	0.60	0.63	0.68	0.75	0.84	0.88	0.91	0.94	0.97	0.98	0.99	1.00	1.00	1.00	1.00
1.25	0.85	0.76	0.62	0.54	0.48	0.47	0.49	0.53	0.59	0.68	0.80	0.85	0.90	0.94	0.96	0.98	0.99	1.00	1.00	1.00	1.00
1.00	0.82	0.73	0.58	0.50	0.44	0.43	0.46	0.50	0.57	0.67	0.80	0.85	0.90	0.93	0.96	0.98	0.99	1.00	1.00	1.00	1.00
0.75	0.77	0.68	0.52	0.45	0.39	0.39	0.42	0.46	0.54	0.65	0.79	0.84	0.89	0.93	0.96	0.98	0.99	1.00	1.00	1.00	1.00
0.50	0.71	0.60	0.45	0.38	0.33	0.33	0.37	0.42	0.50	0.62	0.77	0.83	0.89	0.93	0.96	0.98	0.99	1.00	1.00	1.00	1.00
0.25	0.58	0.47	0.33	0.28	0.25	0.26	0.31	0.36	0.45	0.59	0.76	0.83	0.88	0.93	0.96	0.98	0.99	1.00	1.00	1.00	1.00
0.12	0.45	0.35	0.24	0.20	0.19	0.20	0.25	0.31	0.41	0.56	0.75	0.82	0.88	0.93	0.96	0.98	0.99	1.00	1.00	1.00	1.00

Figure 3: Ratios between S^{CSR} and S^{OPT}

Note: The figure reports the ratios between the Sharpe ratios of the OLS based portfolio and the feasible optimal arbitrage portfolio. The simulation setting is based on model (4), in which a $100 \times \rho\%$ of assets have alphas that correspond to an annualized Sharpe ratio $\mu/\sigma \times \sqrt{12}$.

Similarly, the B-H procedure cannot achieve the optimal Sharpe ratio, as shown by Figure 4. According to Proposition 4, the gap between the optimal Sharpe ratio and the B-H approach largely depends on the signal strength. As long as $T\mu^2\sigma^{-2} \to 0$, the inequality (21) holds (since $\rho < 1$), the B-H procedure achieves the optimality. These scenarios correspond to the white values on Figure 4, where the border of the dominant region is located near the vertical line at $\mu/\sigma\sqrt{12} = 2.19$. Intuitively, the B-H is effective in singling out strong signals, so it leads to almost optimal portfolios as long as all signals are strong. However, when signals are weak, the B-H procedure, which amounts to hard-thresholding, performs worse than the cross-sectional regression, since in this case the embedded ridge regularization in the latter is more appropriate than hard-thresholding. As shown by Figure 1, even if alphas are individually weak, their empirical relevance should not be ignored because their collective contribution to the portfolio's Sharpe ratio can be highly non-trivial. The B-H approach is overly conservative compared to alternatives in this parameter regime.



$\rho \times 100$	10.95	8.22	5.48	4.38	3.29	2.74	2.19	1.92	1.64	1.37	1.10	0.99	0.88	0.77	0.66	0.55	0.44	0.33	0.22	0.16	0.11
50.00	0.97	0.97	0.97	0.98	0.98	0.96	0.87	0.77	0.62	0.42	0.19	0.11	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25.00	0.97	0.97	0.98	0.98	0.98	0.96	0.84	0.72	0.55	0.32	0.12	0.06	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20.00	0.97	0.97	0.98	0.98	0.98	0.95	0.83	0.70	0.52	0.30	0.10	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15.00	0.97	0.97	0.98	0.98	0.98	0.95	0.81	0.67	0.48	0.26	0.08	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10.00	0.97	0.97	0.98	0.98	0.98	0.94	0.78	0.63	0.43	0.21	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7.50	0.97	0.97	0.98	0.98	0.98	0.93	0.76	0.60	0.40	0.18	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5.00	0.97	0.97	0.98	0.98	0.98	0.91	0.72	0.55	0.35	0.15	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2.50	0.97	0.97	0.98	0.98	0.97	0.89	0.66	0.48	0.27	0.09	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.25	0.97	0.97	0.98	0.98	0.96	0.85	0.59	0.40	0.21	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.00	0.97	0.97	0.98	0.99	0.96	0.84	0.57	0.38	0.19	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.75	0.97	0.97	0.98	0.99	0.95	0.82	0.55	0.35	0.17	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.50	0.97	0.97	0.98	0.99	0.94	0.80	0.51	0.32	0.14	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.25	0.97	0.97	0.98	0.99	0.92	0.76	0.45	0.26	0.10	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.12	0.97	0.97	0.98	0.99	0.90	0.71	0.39	0.20	0.07	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

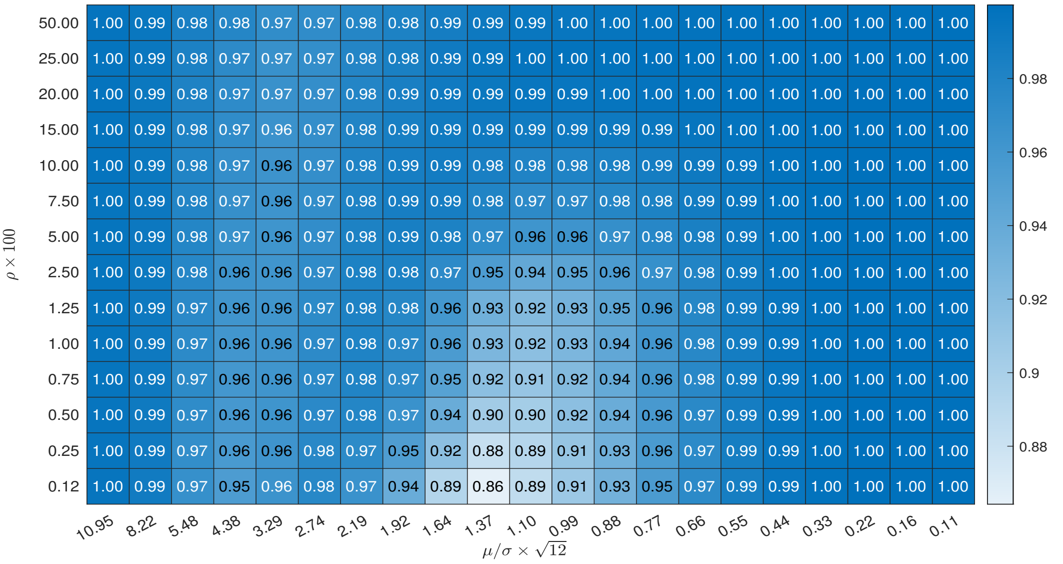
Figure 4: Ratios between S^{BH} and S^{OPT}

Note: The figure reports the ratios between the Sharpe ratios of the multiple testing based portfolio (via B-H procedure) and the feasible optimal arbitrage portfolio. The simulation setting is based on model (4), in which a $100 \times \rho\%$ of assets have alphas that correspond to an annualized Sharpe ratio $\mu/\sigma \times \sqrt{12}$.

Last but not least, Figure 5 presents the result for LASSO. This approach involves a tuning parameter, which calls for a cross-validation procedure. We adopt an infeasible and theoretically optimal tuning parameter, λ, which maximizes S^{LASSO}, making this approach a stronger competitor. Even though Proposition 5 suggests that LASSO is not uniformly optimal, it performs quite well, achieving the optimal Sharpe ratio in almost all regimes. Intuitively, when signals are very strong, LASSO behaves like a hard-thresholding selector, as shrinkage does not play too much a role. When signals are rather weak, LASSO behaves like Ridge, because shrinking these signals does not change the fact that they are almost indistinguishable from noise.

3.2 Comparison of Portfolio Strategies in Finite Sample

We now compare the finite sample performance of our portfolio estimators over different DGPs. For any given parameter value $(\mu/\sigma, \rho)$ in a DGP, we estimate the portfolio weights, $\widehat{w}^{\mathrm{OPT}}$, using our Algorithm 1, and calculate the resulting (theoretical) Sharpe ratio: $\widehat{w}^{\mathrm{OPT}\intercal}\mu/\sqrt{\widehat{w}^{\mathrm{OPT}\intercal}\Sigma_u^{-1}\widehat{w}^{\mathrm{OPT}}}$.



$\rho \times 100$ \ $\mu/\sigma \times \sqrt{12}$	10.95	8.22	5.48	4.38	3.29	2.74	2.19	1.92	1.64	1.37	1.10	0.99	0.88	0.77	0.66	0.55	0.44	0.33	0.22	0.16	0.11
50.00	1.00	0.99	0.98	0.98	0.97	0.97	0.98	0.98	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
25.00	1.00	0.99	0.98	0.97	0.97	0.97	0.98	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20.00	1.00	0.99	0.98	0.97	0.97	0.97	0.98	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
15.00	1.00	0.99	0.98	0.97	0.96	0.97	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10.00	1.00	0.99	0.98	0.97	0.96	0.97	0.98	0.99	0.99	0.98	0.98	0.98	0.98	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00
7.50	1.00	0.99	0.98	0.97	0.96	0.97	0.98	0.99	0.99	0.98	0.97	0.97	0.98	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00
5.00	1.00	0.99	0.98	0.97	0.96	0.97	0.98	0.99	0.98	0.97	0.96	0.96	0.97	0.98	0.98	0.99	1.00	1.00	1.00	1.00	1.00
2.50	1.00	0.99	0.98	0.96	0.96	0.97	0.98	0.98	0.97	0.95	0.94	0.95	0.96	0.97	0.98	0.99	1.00	1.00	1.00	1.00	1.00
1.25	1.00	0.99	0.97	0.96	0.96	0.97	0.98	0.98	0.96	0.93	0.92	0.93	0.95	0.96	0.98	0.99	0.99	1.00	1.00	1.00	1.00
1.00	1.00	0.99	0.97	0.96	0.96	0.97	0.98	0.97	0.96	0.93	0.92	0.93	0.94	0.96	0.98	0.99	0.99	1.00	1.00	1.00	1.00
0.75	1.00	0.99	0.97	0.96	0.96	0.97	0.98	0.97	0.95	0.92	0.91	0.92	0.94	0.96	0.98	0.99	0.99	1.00	1.00	1.00	1.00
0.50	1.00	0.99	0.97	0.96	0.96	0.97	0.98	0.97	0.94	0.90	0.90	0.92	0.94	0.96	0.97	0.99	0.99	1.00	1.00	1.00	1.00
0.25	1.00	0.99	0.97	0.96	0.96	0.98	0.97	0.95	0.92	0.88	0.89	0.91	0.93	0.96	0.97	0.99	0.99	1.00	1.00	1.00	1.00
0.12	1.00	0.99	0.97	0.95	0.96	0.98	0.97	0.94	0.89	0.86	0.89	0.91	0.93	0.95	0.97	0.99	0.99	1.00	1.00	1.00	1.00

Figure 5: Ratios between S^{LASSO} and S^{OPT}

Note: The figure reports the ratios between the Sharpe ratios of the LASSO based portfolio and the feasible optimal arbitrage portfolio. The simulation setting is based on model (4), in which a $100 \times \rho\%$ of assets have alphas that correspond to an annualized Sharpe ratio $\mu/\sigma \times \sqrt{12}$. The tuning parameter λ is selected to maximize S^{LASSO}.

We then calculate the average Sharpe ratio over all Monte Carlo repetitions. Our approach requires a tuning parameter k_n. For robustness, we report results based on three parameter values $(0.5k_n, k_n, 2k_n)$ with $k_n = 0.25$. We repeat this exercise for the CSR, B-H, and LASSO methods for comparison.

In light of Theorem 2, a sensible choice of the estimation error can be written as:

$$\text{Err}^{\text{A}}(\mu/\sigma, \rho) = |\widehat{S}^{\text{A}} - S^{\text{OPT}}|/(1 + S^{\text{OPT}}),$$

where A denotes OPT, CSR, BH, or LASSO, and the dependence of $\widehat{S}^{\text{A}}$ and S^{OPT} on μ/σ and ρ is omitted. When S^{OPT} is large (i.e., $>> 1$), this error is in percentages relative to S^{OPT}; when S^{OPT} is small (i.e., $o_{\text{P}}(1)$), the error is measured in terms of the absolute difference. The error is defined this way because S^{OPT} itself can diverge or diminish depending on different parameters in the simulated DGPs.

Table 1 reports the maximal error over all values of μ/σ and ρ given in Section 3.1. The results show that OPT has a smaller error in almost all cases for all tuning parameters than CSR, BH, or LASSO. As T increases from 10 years to 40 years, the maximum error drops from 0.377 to 0.263 in the case of $N = 1,000$ for $k_n = 0.25$, whereas CSR, BH and LASSO stay above 0.44. The maximal error for CSR is achieved at the lower left conner of Figure 1, where signals are strong but rare; for BH, the worst performance occurs around the upper right corner, where many weak signals exist; for LASSO, the worse is near the bottom but in the middle, where signals are neither too strong nor too weak.

	$N = 1,000$, Monthly			$N = 3,000$, Monthly			$N = 1,000$, Daily		
	$T = 10$	$T = 20$	$T = 40$	$T = 10$	$T = 20$	$T = 40$	$T = 10$	$T = 20$	$T = 40$
OPT	0.385	0.332	0.289	0.442	0.367	0.320	0.449	0.440	0.408
	0.377	0.309	0.263	0.437	0.333	0.282	0.411	0.382	0.356
	0.381	0.282	0.233	0.446	0.318	0.247	0.370	0.334	0.303
CSR	0.540	0.489	0.441	0.618	0.570	0.515	0.537	0.485	0.427
BH	0.742	0.703	0.651	0.814	0.789	0.748	0.760	0.715	0.657
LASSO	0.537	0.488	0.440	0.615	0.568	0.512	0.536	0.483	0.426

Table 1: Sharpe Ratio Comparison in Simulations

Note: This table reports the maximum error, defined by $\sup_{\mu/\sigma,\rho} \mathrm{Err}^{\mathrm{A}}(\mu/\sigma, \rho)$, where A denotes either OPT, or CSR, or BH, over all values of μ/σ and ρ in Figure 1, for several choices of N, T (in years), and data frequencies. The first three rows correspond to the OPT approach with three different values of tuning parameters, $0.5k_n$, k_n, and $2k_n$, respectively, where $k_n = 0.25$. The BH approach controls false discovery rate at a level 5%. The LASSO approach uses the optimal (infeasible) tuning parameter that optimizes S^{LASSO}.

3.3 Finite Sample Performance of the Infeasible Sharpe Ratio Estimator

Finally, Figure 6 reports the estimation error $\left|\widehat{S}^\star - S^\star\right| / \left(1 + S^\star\right)$ in simulations. The result confirms the consistency result given by Proposition 2. The error is relative when $S^\star$ is large or moderate ($>> 1$). We find the relative error is around 1% towards the left top corner. For DGPs near the bottom right corner of Figure 6, $S^\star$ vanishes as shown by Figures 1 and 2, the error becomes absolute ($S^\star << 1$) and is moderately small given the sample size and the cross-sectional dimension.

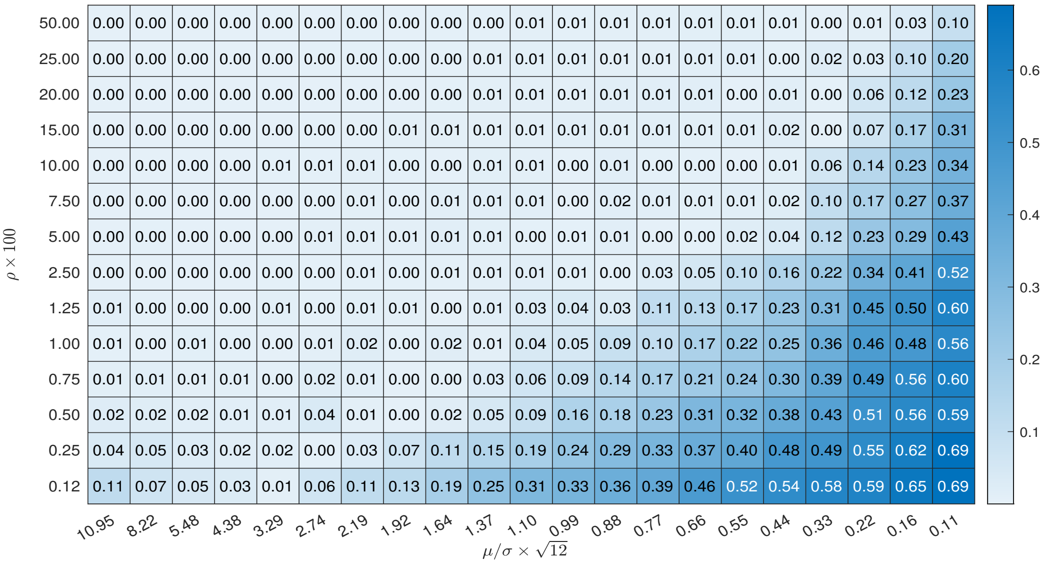
4 Empirical Analysis of US Equities

To demonstrate the empirical relevance of the statistical limit of arbitrage, we study US monthly equity returns from January 1965 to December 2020.

4.1 Data Preprocessing

We adopt a multi-factor model with 16 characteristics and 11 GICS sectors, which are selected to incorporate empirical insight from existing asset pricing literature and industry practice. The selected characteristics include market beta, size, operating profits/book equity, book equity/market equity, asset growth, momentum, short-term reversal, industry momentum, illiquidity, leverage, return seasonality, sales growth, accruals, dividend yield, tangibility, and idiosyncratic risk, which are downloaded directly from the website openassetpricing.com, see Chen and Zimmermann (2020) for construction details.

We download the monthly return data for individual equities from CRSP. We take a number of steps to preprocess the data. First, we single out delisted stocks, and attach delisting returns as their last returns (on the delisting months). Next, we merge the returns data with the aforementioned



$\rho \times 100$ \ $\mu/\sigma \times \sqrt{12}$	10.95	8.22	5.48	4.38	3.29	2.74	2.19	1.92	1.64	1.37	1.10	0.99	0.88	0.77	0.66	0.55	0.44	0.33	0.22	0.16	0.11
50.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.03	0.10
25.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.02	0.03	0.10	0.20
20.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.06	0.12	0.23
15.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.00	0.07	0.17	0.31
10.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.06	0.14	0.23	0.34
7.50	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.02	0.01	0.01	0.01	0.02	0.10	0.17	0.27	0.37
5.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.02	0.04	0.12	0.23	0.29	0.43
2.50	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.00	0.03	0.05	0.10	0.16	0.22	0.34	0.41	0.52
1.25	0.01	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.01	0.03	0.04	0.03	0.11	0.13	0.17	0.23	0.31	0.45	0.50	0.60
1.00	0.01	0.00	0.01	0.00	0.00	0.01	0.02	0.00	0.02	0.01	0.04	0.05	0.09	0.10	0.17	0.22	0.25	0.36	0.46	0.48	0.56
0.75	0.01	0.01	0.01	0.01	0.00	0.02	0.01	0.00	0.00	0.03	0.06	0.09	0.14	0.17	0.21	0.24	0.30	0.39	0.49	0.56	0.60
0.50	0.02	0.02	0.02	0.01	0.01	0.04	0.01	0.00	0.02	0.05	0.09	0.16	0.18	0.23	0.31	0.32	0.38	0.43	0.51	0.56	0.59
0.25	0.04	0.05	0.03	0.02	0.02	0.00	0.03	0.07	0.11	0.15	0.19	0.24	0.29	0.33	0.37	0.40	0.48	0.49	0.55	0.62	0.69
0.12	0.11	0.07	0.05	0.03	0.01	0.06	0.11	0.13	0.19	0.25	0.31	0.33	0.36	0.39	0.46	0.52	0.54	0.58	0.59	0.65	0.69

Figure 6: Comparison between $\widehat{S}^\star$ and $S^\star$

Note: The figure reports the error between $\widehat{S}^\star$ and $S^\star$ defined as $\left|\widehat{S}^\star - S^\star\right|/\left(1+S^\star\right)$. The simulation setting is based on model (4), in which a $100 \times \rho\%$ of assets have αs that correspond to an annualized Sharpe ratio $\mu/\sigma \times \sqrt{12}$. In this experiment, $N = 1,000$ and $T = 20$ years.

characteristics database using permnos. The total number of unique permnos on average per month is 6,536. We then apply the usual filters (share codes 10 and 11 and exchange codes 1, 2, and 3) to the database, to eliminate (part of) the sampling periods for stocks that fail to meet these criteria. The remaining average number of stocks per month is 4,756. For stocks whose returns are missing for more than 3 months, we eliminate the missing periods, otherwise we fill the missing returns by zeros.

We now deal with missing characteristics. We start by removing all characteristics data for any stocks since their delisting months. We then fill missing GICS codes with the corresponding stocks' most recent records prior to their missing dates. Stocks without any GICS codes over the entire sample period are eliminated. If the GICS codes become available later in the sample for some stocks, their sample prior to the first dates when GICS become available are eliminated, which mainly occurs prior to 1990. With GICS information, we adopt a two-step procedure to fill in other missing characteristics. For any missing value in a stock's characteristic, we fill it with the sector-wise median of this characteristic each month. If a characteristic's values are not available for an entire sector in a certain month, we fill them with this characteristic's cross-sectional median over all stocks in this month. After data preprocessing, the final average number of stocks per month is reduced to 4,067.

The resulting panel is not balanced, because we do not fill in missing data before a stock's IPO or after its delisting. Our approach to filling missing data thereby avoids forward-looking bias. An empirical issue we have not discussed so far is time-varying exposures and missing data. We adapt Algorithm 1 and estimators (13) and (14) to this case in Appendix B.

4.2 Model Performance

At the end of each month, we run cross-sectional regressions of next month returns onto the 27 cross-sectional predictors (including the intercept). We do so using all stocks in the current month's cross sections. Following Gu et al. (2020), the 16 characteristics are rank-normalized within each cross-section, alleviating the impact of extreme outliers in characteristics, though this barely changes any follow-up results.

Figure 7 plots the time series of the cross-sectional regression R^2s over time. The R^2 has been on the decline since the beginning of the sample till 1990s. This coincides with the period when the number of stocks in the US equity markets increases. The R^2s are moderately low, with an average of 8.25%. The low R^2s suggest that a substantial portion of cross-sectional variation of individual equity returns is idiosyncratic noise. Therefore, learning alphas from residuals of the factor model is an incredibly difficult statistical task.

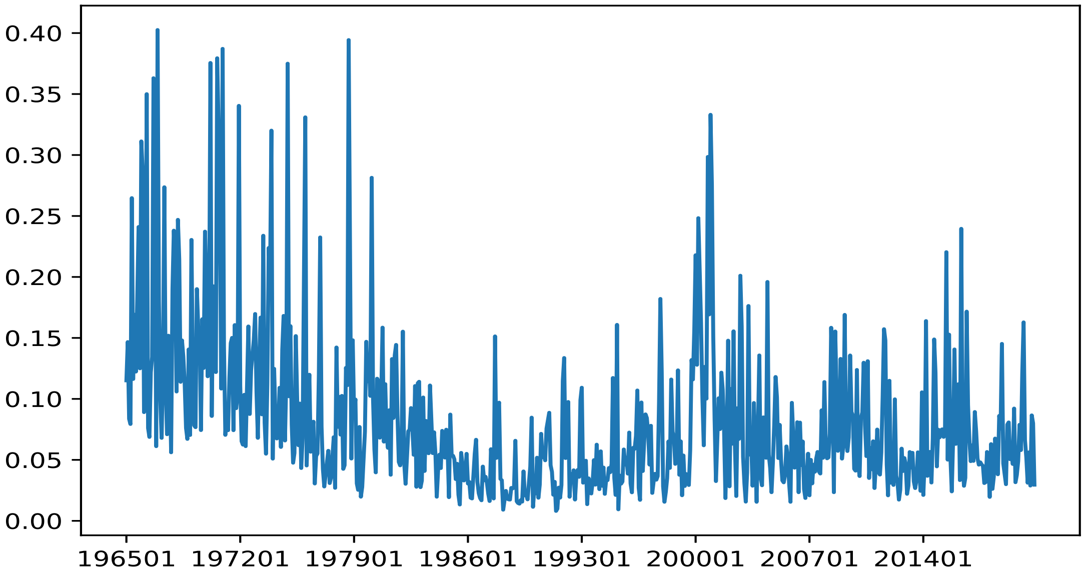


Figure 7: Time-series of the Cross-sectional R^2s

4.3 Rare and Weak Alphas

We now study the statistical properties of alphas using the full sample data. For each stock, we collect its regression residuals and take their average as an estimate for its alpha. We impose that all residuals have at least 60 observations. This ensures enough sample size for inference on alpha, although the distribution of alphas' t-statistics turns out not sensitive to this requirement. Figure 8 provides histograms of the t-statistics and Sharpe ratios for alphas of all 12,415 stocks in our sample that meet this criterion. Because these stocks have different sample sizes, the histograms of the Sharpe ratios are not simply the scaled version of the histogram of the t-statistics.

Only 6.35% of the t-statistics exceed 2.0 in magnitude, and more than 0.63% exceed 3.0. This suggests that truly significant alphas are extremely rare. Moreover, the largest Sharpe ratio of all individual stocks' alphas is rather modest, about 1.699. Only 0.505% of the alphas have a Sharpe ratio greater than 1.0. These summary statistics suggest that rare and weak alpha is perhaps the most relevant scenario in practice.

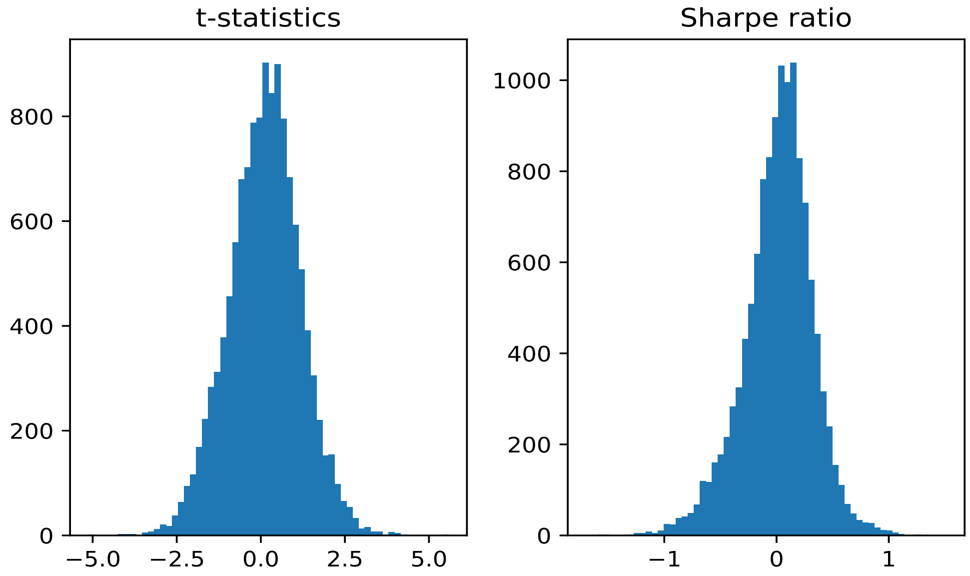


Figure 8: Histograms of the t-Statistics and Sharpe Ratios of Estimated Alphas

Note: The figure provides the histograms of the t-statistics (left) and Sharpe ratios (right) of estimated alphas for all tickers in our sample with at least 60 months of data. The total number of tickers available is 12,482.

4.4 Performance of Arbitrage Portfolios

Throughout we assume alphas do not vary over time. If alphas are driven by some observable characteristics, then it is possible to construct a factor using these characteristics via cross-sectional regressions, which turns "alpha" into risk premia. In this regard, alphas are meaningless without reference to a specific factor model. Extracting more "factors" out of alphas would lead to even smaller arbitrage profits.

We now compare arbitrage portfolios based on various strategies, including the optimal strategy, the cross-sectional regression (CSR) approach, the multiple-testing based procedure (BH), and LASSO approach. The ridge approach is omitted, since it is equivalent to the CSR.

Specifically, at the end of each month, we build optimal portfolio weights using these strategies. We only invest in stocks with a continuous record for at least 96 months. We rebalance these portfolios at the end of each month, with weights recalculated using a 120-month rolling window.

Both Lasso and the optimal strategy require a tuning parameter. Out of the 10-year rolling window, we leave the last 2 years as the validation sample for tuning parameter selection. As expected, optimal tuning parameter is difficult to select, which undermines the performance of both strategies.

All these strategies yield similar Sharpe ratios. BH and OPT tie for the top of the chart, yielding 0.497 and 0.496, respectively, followed by CSR that scores 0.450. The LASSO approach only obtains 0.384. The Sharpe ratios of different strategies are not influenced by risk aversion, though the cumulative returns are. To compare cumulative returns, we normalize all strategies to have the same (ex-post) volatility. The resulting time-series of normalized cumulative returns are shown in Figure 9.

Closely examining these strategies reveals more insight. BH is highly conservative. Out of 46 years of out-of sample trading months (1975/01 - 2020/12), 289 months have no trading activities. The largest number of stocks selected for trading in a month is 10, and the average over all non-zero periods is 2.43. In contrast, CSR trade all stocks that meet our trading criteria, with an average of 2,366 stocks per month. OPT almost does so, with an average of 2,359. The number of stocks traded by LASSO is rather volatile, varying between none and all stocks from month to month, with an average of 757.6 per month. This is likely caused by the noise in the tuning process.

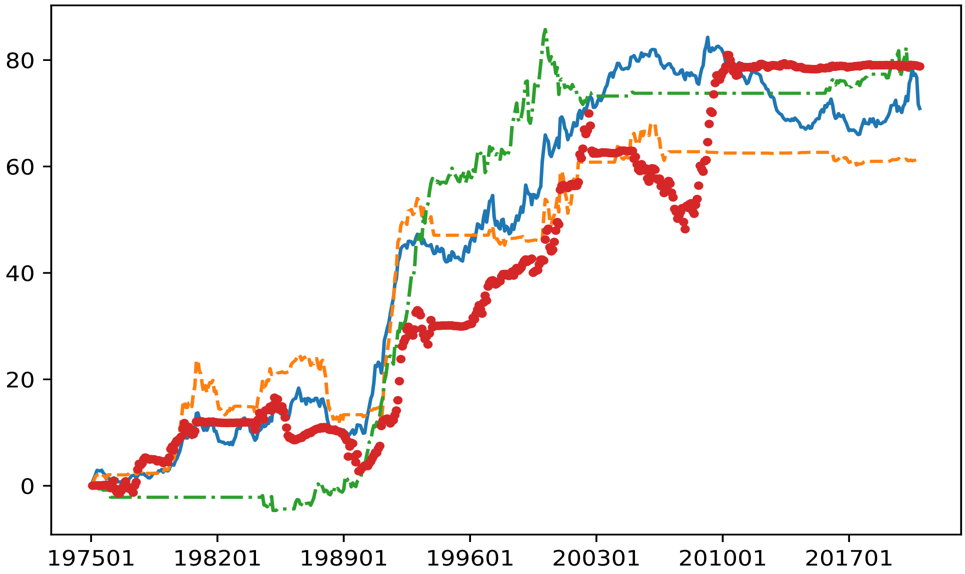


Figure 9: Normalized Cumulative Returns of Arbitrage Portfolios

Note: This figure compares the cumulative returns of OPT (red dotted), CSR (blue solid), BH (green dot-dashed), and LASSO (orange dashed) strategies. We normalize all returns by their realized volatilities calculated by the square root of the sum of the squared returns over the entire sample, only for comparison purpose.

We also calculate the perceived Sharpe ratios using (13), and provide a time-series plot of $\widehat{S}^\star$ in Figure 10. We also compare it with the biased estimates $\widetilde{S}^\star$ using (14). We observe a huge gap between the estimated perceived Sharpe ratios using these formulae. As predicted by Proposition 2, $\widetilde{S}^\star$ overestimates $S^\star$, though it guarantees positive values. Our estimate $\widetilde{S}^\star$ is averaged around

2.55 (we truncate negative estimates by 0), but can sometimes exceed 7.5. These estimates are far greater than the feasible Sharpe ratios we obtain for any of these strategies. That said, even the infeasible Sharpe ratios can be as low as 0 for certain periods of the sample. The feasible portfolio returns seem in agreement with the prediction. For instance, OPT, LASSO, and BH's cumulative returns are almost flat post 2010, whereas CSR has negative returns.

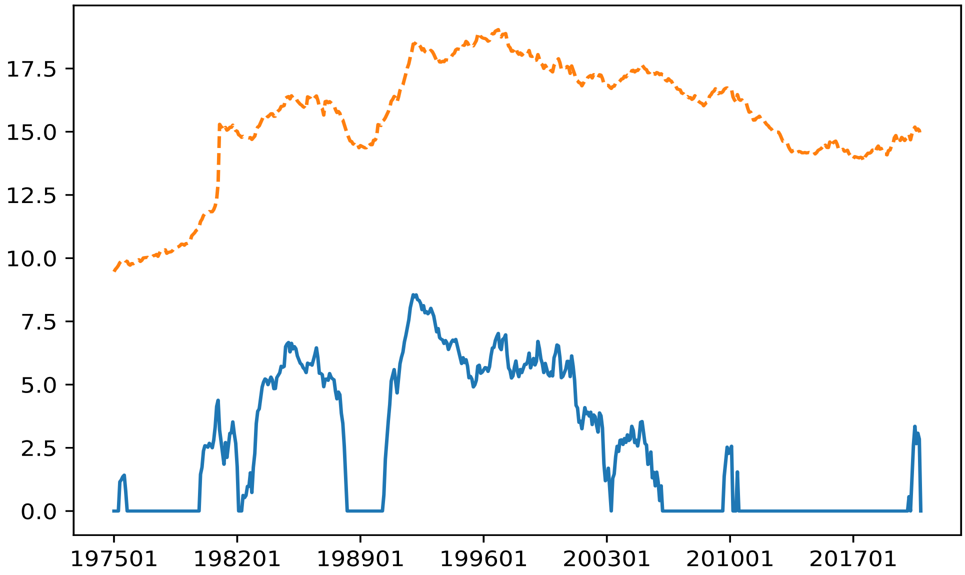


Figure 10: Time Series of Sharpe Ratios

Note: The figure compares the optimal feasible (blue solid) and infeasible (orange dashed) Sharpe ratios, estimated based on a rolling window of 120 months.

5 Conclusion

More broadly, existing literature have documented impressive Sharpe ratios on various machine learning based trading strategies. Such strategies often rely on ad-hoc model design (e.g., a neural network with a specific architecture) and tuning parameters selection. In this regard, the empirical analysis can at best provide a "lower bound" on the performance of machine learning strategies in investment. Our paper provides a theoretical framework to understand the "upper bound" on the performance of any strategy in the specific context of arbitrage pricing theory, tying together this statistical limit with economic rationale.

The empirical message should be confined within the context of monthly rebalancing strategies via linear factor models. The gap between feasible and infeasible Sharpe ratios will further increase if arbitrageurs face additional statistical challenges, e.g., model misspecification, omitted factors, weak factors, large non-sparse idiosyncratic covariance matrix, etc. Consequently, the empirical gap should remain for any arbitrageurs, including those who engage in higher frequency trading or use more complex nonlinear models.

On a side note, our theoretical and empirical analyses also have implications on the econometric analysis in asset pricing. Examining the economic performance of asset pricing models is as important as and complementary to statistical tests. The criteria of a good statistical test are primarily statistical in nature, such as Type I and Type II errors, false discovery rate, etc; whereas in practice, it is the economic performance that agents in the economy fundamentally care about. There is often a wedge between these two objectives. For instance, a statistical procedure that guards against false discovery rate may be overly conservative for investment purpose; rejection by a powerful test statistic may not necessarily lead to the practical irrelevance of an economic theory.

References

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5*(1), 31–56.

Andrews, D. W. and X. Cheng (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica 80*(5), 2153–2211.

Andrews, D. W. K., X. Cheng, and P. Guggenberger (2020). Generic results for establishing the asymptotic size of confidence sets and tests. *Journal of Econometrics 218*(2), 496–531.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *Journal of Finance 61*(1), 259–299.

Bajgrowicz, P. and O. Scaillet (2012, December). Technical trading revisited False discoveries, persistence tests, and transaction costs. *Journal of Financial Economics 106*(3), 473–491.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics 9*(1), 3–18.

Barras, L., O. Scaillet, and R. Wermers (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *Journal of Finance 65*(1), 179–216.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica 62*(3), 657–681.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological) 57*(1), 289–300.

Bhandari, L. C. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *Journal of Finance 43*(2), 507–528.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica 51*, 1281–1304.

Chen, A. Y. and T. Zimmermann (2020). Open source cross-sectional asset pricing. *Available at SSRN*.

Chen, X., L. P. Hansen, and P. G. Hansen (2021a). Robust identification of investor beliefs. *Proceedings of the National Academy of Sciences 117*(52), 33130–33140.

Chen, X., L. P. Hansen, and P. G. Hansen (2021b). Robust inference for moment condition models without rational expectations. *Journal of Econometrics, forthcoming*.

Collin-Dufresne, P., M. Johannes, and L. A. Lochstoer (2016). Parameter learning in general equilibrium: The asset pricing implications. *American Economic Review 106*(3), 664–698.

Connor, G., M. Hagmann, and O. Linton (2012). Efficient semiparametric estimation of the fama-french model and extensions. *Econometrica 80*(2), 713–754.

Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics 15*(3), 373–394.

Cooper, M. J., H. Gulen, and M. J. Schill (2008). Asset growth and the cross-section of stock returns. *Journal of Finance 63*(4), 1609–1651.

Donoho, D. and J. Jin (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Annals of Statistics 32*(3), 962–994.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (2006). Profitability, investment and average returns. *Journal of Financial Economics 82*(3), 491–518.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*, 607–636.

Fan, J., Y. Liao, and J. Yao (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica 83*(4), 1497–1541.

Gagliardini, P., E. Ossola, and O. Scaillet (2016). Time-varying risk premium in large cross-sectional equity datasets. *Econometrica 84*(3), 985–1046.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica 29*, 1121–1152.

Giglio, S., Y. Liao, and D. Xiu (2021). Thousands of alpha tests. *Review of Financial Studies 34*(7), 3456–3496.

Giglio, S. and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy 129*(7), 1947–1990.

Gromb, D. and D. Vayanos (2010). Limits of arbitrage. *Annual Review of Financial Economics 2*, 251–275.

Gu, S., B. Kellly, and D. Xiu (2020). Empirical asset pricing via machine learning. *Review of Financial Studies 33*(5), 2223–2273.

Gu, S., B. T. Kelly, and D. Xiu (2021). Autoencoder asset pricing models. *Journal of Econometrics 222*, 429–450.

Hahn, J. and H. Lee (2009). Financial constraints, debt capacity, and the cross-section of stock returns. *Journal of Finance 64*(2), 891–921.

Hansen, L. P. (2007). Beliefs, doubts, and learning: Valuing macroeconomic risk. *American Economic Review 97*(2), 1–30.

Hansen, L. P. (2014). Nobel lecture: Uncertainty outside and inside economic models. *Journal of Political Economy 122*(51), 945–987.

Harvey, C. R. and Y. Liu (2020). False (and missed) discoveries in financial economics. *Journal of Finance, forthcoming*.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *Review of Financial Studies 29*(1), 5–68.

Heston, S. L. and R. Sadka (2008). Seasonality in the cross-section of stock returns. *Journal of Financial Economics 87*(2), 418–445.

Huberman, G. (1982). A simple approach to arbitrage pricing theory. *Journal of Economic Thoery 28*(1), 183–191.

Imbens, G. W. and C. F. Manski (2004). Confidence intervals for partially identified parameters. *Econometrica 72*(6), 1845–1857.

Ingersoll, J. E. (1984). Some results in the theory of arbitrage pricing. *Journal of Finance 39*(4), 1021–1039.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance 45*(3), 881–898.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91.

Kelly, B., S. Pruitt, and Y. Su (2019). Some characteristics are risk exposures, and the rest are irrelevant. *Journal of Financial Economics, forthcoming*.

Kim, S., R. Korajczyk, and A. Neuhierl (2020). Arbitrage portfolios. *Review of Financial Studies, forthcoming*.

Kozak, S., S. Nagel, and S. Santosh (2018). Interpreting factor models. *Journal of Finance 73*(3), 1183–1223.

Lakonishok, J., A. Shleifer, and R. W. Vishny (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance 49*(5), 1541–1578.

Leeb, H. and B. M. Pötscher (2005). Model selection and inference: Facts and fiction. *Econometic Theory 21*(1), 21–59.

Lewellen, J. (2015). The cross-section of expected stock returns. *Critical Finance Review 4*(1), 1–44.

Litzenberger, R. H. and K. Ramaswamy (1979). The effects of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics 7*, 163–195.

Liu, W. and Q.-M. Shao (2014). Phase transition and regularized bootstrap in large-scale t-tests with false discovery rate control. *The Annals of Statistics 42*(5), 2003–2025.

Martin, I. and S. Nagel (2021). Market efficiency in the age of big data. *Journal of Financial Economics, forthcoming*.

Moskowitz, T. J. and M. Grinblatt (1999). Do industries explain momentum? *Journal of Finance 54*(4), 1249–1290.

Pastor, L. and P. Veronesi (2009). Learning in financial markets. *Annual Review of Financial Economics 1*(1), 361–381.

Pesaran, H. and T. Yamagata (2017). Testing for alpha in linear factor pricing models with a large number of securities. Technical report.

Rosenberg, B. (1974). Extra-market components of covariance in security returns. *Journal of Financial and Quantitative Analysis 9*(2), 263–274.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3), 341–360.

Shanken, J. (1992). The current state of the arbitrage pricing theory. *Journal of Finance 47*(4), 1569–1574.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review 71*(3), 289–315.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Tukey, J. W. (1976). The higher criticism. course notes, statistics 411. Technical report, Princeton University.

Uppal, R. and P. Zaffaroni (2018). Robust portfolio choice. Technical report, EDHEC Business School and Imperial College London.

Appendix A Assumptions

Finally, the assumption below specifies the relative rate of N and T in our limiting experiments.

Assumption A2. *For each $N \geq 1$, the following conditions holds:*

(a) $C^{-1}N^{d} \leq T \leq CN^{d'}$ for fixed constants $d > 1/3$ and $d' < 1$.

(b) The pdf of σ_i, $p_\sigma(x)$, has a support $(\underline{\sigma}, \bar{\sigma})$ which satisfies $C^{-1} \leq \underline{\sigma} \leq \bar{\sigma} \leq C$; the pdf also satisfies $C^{-1} \leq p_\sigma(x) \leq C$ for all $x \in (\underline{\sigma}, \bar{\sigma})$.

(c) $\varepsilon_{i,t}$ has a finite twelfth moment.

Appendix B Missing Data and Time-Varying Exposures

In practice, when applying Algorithm 1 to individual stocks, we have to deal with time-varying factor loadings in an unbalanced panel of individual equity returns. In this section, we provide an extension of Algorithm 1 to this case:

Algorithm 3 (Constructing the Optimal Arbitrage Portfolio with Missing Data)**.**

S1. For each i, the observed sample is $\mathcal{T}_i$, a subset of $\mathcal{T} = \{t - T + 1, \ldots, t\}$. We split $\mathcal{T}_i$ into:

$$S_i' = \mathcal{T}_i \cap \{t - \lfloor T^{1/2} \rfloor + 1, \ldots, t\} \quad \textit{and} \quad S_i = \mathcal{T}_i - S_i',$$

and we construct cross-sectional regression estimates of alpha $\breve{\alpha}$ and $\breve{\alpha}'$, volatility estimates $\breve{\sigma}$, and the t-statistics $\breve{z}_i = |S_i|^{1/2}\breve{\alpha}_i/\breve{\sigma}_i$ for each $i = 1, 2, \ldots, N$, using subsamples S and S':

$$\breve{\alpha}_i = |S_i|^{-1} \sum_{s \in S_i} (\mathbb{M}_{\beta_s^o} r_s^o)_i, \quad \breve{\alpha}_i' = |S_i'|^{-1} \sum_{s \in S_i'} (\mathbb{M}_{\beta_s^o} r_s^o)_i, \quad \textit{and} \quad \breve{\sigma}_i^2 = |S_i|^{-1} \sum_{s \in S_i} \big((\mathbb{M}_{\beta_s^o} r_s^o)_i - \breve{\alpha}_i \big)^2,$$

where β_s^o and r_s^o replace the missing values in β_s and r_s with zeros, $\mathbb{M}_{\beta_s^o} = \mathbb{I}_N - \beta_s^o(\beta_s^{o\intercal}\beta_s^o)^{-1}\beta_s^{o\intercal}$, and $\mathbb{I}_N$ denotes the $N \times N$ identity matrix.

S2. We choose the arbitrage portfolio weights as

$$\widehat{w}^{\mathrm{OPT}} = \mathbb{M}_\beta \breve{w}, \quad \textit{with} \quad \breve{w}_i = \begin{cases} |\mathcal{T}_i|^{-1} \widehat{f}(\lfloor \breve{z}_i / k_N \rfloor, \lfloor (T/|\mathcal{T}_i|)^{1/2} \breve{\sigma}_i / k_N^{3/2} \rfloor), & |\breve{z}_i| \leq k_N^{-2/3}, \\ (\breve{\sigma}_i)^{-2} \breve{\alpha}_i, & |\breve{z}_i| > k_N^{-2/3}. \end{cases}$$

For any set of integers (l, m), we choose $k_N \sim (\log N)^{-1}$ and define

$$\widehat{f}(l, m) = \frac{T}{m^2 k_N^3} \frac{1}{|B(l, m)|} \sum_{i \in B(l,m)} \breve{\alpha}_i',$$

where

$$B(l,m)=\left\{i\leq N: l\leq \breve{z}_i/k_N<l+1; m-1\leq (T/|\mathcal{T}_i|)^{1/2}\breve{\sigma}_i/k_N^{3/2}<m+2\right\}.$$

In the same vein, we can adapt estimators in (13) and (14) by:

$$(\widehat{S}^{\star})^2=\sum_{i\leq N}\widehat{\sigma}_i^{-2}|\mathcal{T}_i|^{-2}\sum_{t\in\mathcal{T}_i}\sum_{t'\in\mathcal{T}_i:t'\neq t} r_{i,t}\mathbb{M}_{\beta}r_{i,t'},\quad (\widetilde{S}^{\star})^2=\sum_{i\leq N}\widehat{\sigma}_i^{-2}|\mathcal{T}_i|^{-2}\sum_{t\in\mathcal{T}_i}\sum_{t'\in\mathcal{T}_i} r_{i,t}\mathbb{M}_{\beta}r_{i,t'}. \tag{B.1}$$

Appendix C Mathematical Proofs

C.1 Proof of Theorem 1 and Proposition 1, and Corollary 2

Proof of Theorem 1. To simplify the notation, we omit the dependence of β, Σ on N, and $\widehat{w}$ on N and T. All limits are taken as $N\to\infty$. The derivation applies to either fixed T or $T\to\infty$ together with N.

We first note that, given (1), conditioning on $\mathcal{G}$ is equivalent to conditioning on the information set generated by

$$\{(\alpha_i+u_{i,s},\beta_i,v_s,\sigma_i): t-T+1\leq s\leq t, i\leq N\}.$$

According to Assumption 1, $\{(\alpha_i,\alpha_i+u_{i,s},\sigma_i): t-T+1\leq s\leq t\}$ is indepedent of $\{(\alpha_j+u_{j,s},\beta_{j'},v_s,\sigma_j): t-T+1\leq s\leq t, j,j'\leq N, j\neq i\}$. Therefore, the $\mathcal{G}$-conditional distribution of α_i is the same as the $\mathcal{G}_i$-conditional distribution of α_i, where $\mathcal{G}_i$ is the information set generated by $\{(\alpha_i+u_{i,s},\sigma_i): t-T+1\leq s\leq t\}$. Since $\mathcal{G}_i$ is independent across i by Assumption 1, we conclude that, conditionally on $\mathcal{G}$, α_i remains independent across i.

Now define $\mathcal{E}=\mathrm{E}(\widehat{w}^{\intercal}r_{t+1}|\mathcal{F}_t)-\mathrm{E}(\widehat{w}^{\intercal}r_{t+1}|\mathcal{G})$. By the definition of $S(\widehat{w})$, we have

$$S(\widehat{w})=\mathrm{E}(\widehat{w}^{\intercal}r_{t+1}|\mathcal{G})/\mathrm{Var}(\widehat{w}^{\intercal}r_{t+1}|\mathcal{F}_t)^{1/2}+\mathcal{E}/\mathrm{Var}(\widehat{w}^{\intercal}r_{t+1}|\mathcal{F}_t)^{1/2}. \tag{C.2}$$

Since $\widehat{w}$ is $\mathcal{G}$-measurable, it follows that $\mathcal{E}=\widehat{w}^{\intercal}(\alpha-\mathrm{E}(\alpha|\mathcal{G}))$ and that $\mathrm{E}(\mathcal{E}^2|\mathcal{G})=\widehat{w}^{\intercal}\mathrm{Var}(\alpha|\mathcal{G})\widehat{w}$. Then, using Chebyshev's inequality, we have, for all positive fixed ϵ,

$$\mathrm{P}\left(|\mathcal{E}|/\|\widehat{w}\|\geq\epsilon\right)\leq \mathrm{E}(\mathcal{E}^2/\|\widehat{w}\|^2)/\epsilon^2=\mathrm{E}(\widehat{w}^{\intercal}\mathrm{Var}(\alpha|\mathcal{G})\widehat{w}/\|\widehat{w}\|^2)/\epsilon^2. \tag{C.3}$$

Because conditionally on $\mathcal{G}$, α_i is independent across i, we have $\mathrm{Var}(\alpha|\mathcal{G})_{i,j}=\delta_{i,j}\mathrm{Var}(\alpha_i|\mathcal{G})$. It thereby follows that

$$\mathrm{E}(\widehat{w}^{\intercal}\mathrm{Var}(\alpha|\mathcal{G})\widehat{w}/\|\widehat{w}\|^2)\leq \mathrm{E}\left(\max_{i\leq N}\mathrm{Var}(\alpha_i|\mathcal{G})\right)\leq \mathrm{E}\left(\max_{i\leq N}\alpha_i^2\right)=o(1), \tag{C.4}$$

where the last step comes from condition (c) of Assumption 1. Combining (C.3) and (C.4), and using $\mathrm{Var}(\widehat{w}^\intercal r_{t+1}|\mathcal{F}_t) = \widehat{w}^\intercal \Sigma \widehat{w} \geq \lambda_{\min}(\Sigma_u)\|\widehat{w}\|^2 \gtrsim_{\mathrm{P}} \|\widehat{w}\|^2$, we obtain

$$|\mathcal{E}|/\mathrm{Var}(\widehat{w}^\intercal r_{t+1}|\mathcal{F}_t)^{1/2} \lesssim_{\mathrm{P}} |\mathcal{E}|/\|\widehat{w}\| = o_{\mathrm{P}}(1). \tag{C.5}$$

(C.5) and (C.2) lead to

$$S(\widehat{w}) = \widehat{w}^\intercal \mathrm{E}(r_{t+1}|\mathcal{G})(\widehat{w}^\intercal \Sigma \widehat{w})^{-1/2} + o_{\mathrm{P}}(1).$$

Furthermore, applying Cauchy-Schwarz inequality, we obtain

$$|\widehat{w}^\intercal \mathrm{E}(r_{t+1}|\mathcal{G})|^2 \, (\widehat{w}^\intercal \Sigma \widehat{w})^{-1} \leq \mathrm{E}(r_{t+1}|\mathcal{G})^\intercal \Sigma^{-1} \mathrm{E}(r_{t+1}|\mathcal{G}). \tag{C.6}$$

On the other hand, it implies by Woodbury matrix identity and from the fact that $\Sigma = \beta\Sigma_v\beta^\intercal + \Sigma_u$,

$$\Sigma^{-1} = \Sigma_u^{-1} - \Sigma_u^{-1}\beta(\Sigma_v^{-1} + \beta^\intercal\Sigma_u^{-1}\beta)^{-1}\beta^\intercal\Sigma_u^{-1}. \tag{C.7}$$

By direct calculations, we have

$$\beta^\intercal\Sigma^{-1}\beta = ((\beta^\intercal\Sigma_u^{-1}\beta)^{-1} + \Sigma_v)^{-1}.$$

Let $H_1 = (\beta^\intercal\Sigma_u^{-1}\beta)^{-1}$ and $H_2 = \Sigma_v$, and using the fact that $(H_1 + H_2)^{-1} - H_2^{-1} = -(H_1 + H_2)^{-1}H_1H_2^{-1}$, we have

$$\beta^\intercal\Sigma^{-1}\beta - \Sigma_v^{-1} = -((\beta^\intercal\Sigma_u^{-1}\beta)^{-1} + \Sigma_v)^{-1}(\beta^\intercal\Sigma_u^{-1}\beta)^{-1}\Sigma_v^{-1}.$$

Therefore, using the fact that $\lambda_{\min}(\beta^\intercal\beta) \gtrsim_{\mathrm{P}} N$ and that $\lambda_{\max}(\Sigma_u) \lesssim_{\mathrm{P}} 1$ in light of condition (a) and (d) of Assumption 1, we have

$$\lambda_{\max}((\beta^\intercal\Sigma_u^{-1}\beta)^{-1}) = \lambda_{\min}^{-1}(\beta^\intercal\Sigma_u^{-1}\beta) \leq \lambda_{\min}^{-1}(\beta^\intercal\beta)\lambda_{\max}(\Sigma_u) \lesssim_{\mathrm{P}} N^{-1}. \tag{C.8}$$

Also, note that $\lambda_{\max}(\Sigma_v^{-1}) = \lambda_{\min}^{-1}(\Sigma_v) \lesssim 1$, and that

$$\lambda_{\max}(((\beta^\intercal\Sigma_u^{-1}\beta)^{-1} + \Sigma_v)^{-1}) = \lambda_{\min}^{-1}((\beta^\intercal\Sigma_u^{-1}\beta)^{-1} + \Sigma_v) \leq \lambda_{\min}^{-1}(\Sigma_v) \lesssim 1,$$

we have

$$\|\beta^\intercal\Sigma^{-1}\beta - \Sigma_v^{-1}\| \lesssim_{\mathrm{P}} N^{-1},$$

which in turn leads to

$$\gamma^\intercal\beta^\intercal\Sigma^{-1}\beta\gamma = \gamma^\intercal\Sigma_v^{-1}\gamma + o_{\mathrm{P}}(1). \tag{C.9}$$

Next, we show

$$\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma^{-1}\beta\gamma = o_{\mathrm{P}}(1). \tag{C.10}$$

Notice that $\mathrm{E}\left(\mathrm{E}(\alpha|\mathcal{G})|\Sigma,\beta\right) = \mathrm{E}\left(\alpha|\Sigma,\beta\right) = \mathrm{E}\left(\alpha\right) = 0$, and that, conditionally on (Σ,β), $\mathrm{E}(\alpha_i|\mathcal{G})$ is independent across i. Therefore,

$$\mathrm{E}\left(\left(\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma^{-1}\beta\gamma\right)^2|\Sigma,\beta\right)\leq\sum_{i\leq N}\mathrm{E}\left(\mathrm{E}(\alpha_i|\mathcal{G})^2|\Sigma,\beta\right)\max_{j\leq N}(\gamma^\intercal\beta^\intercal\Sigma^{-1})_j^2. \tag{C.11}$$

On the other hand, from (C.7), we obtain

$$\gamma^\intercal\beta^\intercal\Sigma^{-1}=\gamma^\intercal\Sigma_v^{-1}(\Sigma_v^{-1}+\beta^\intercal\Sigma_u^{-1}\beta)^{-1}\beta^\intercal\Sigma_u^{-1}.$$

Because of $\lambda_{\min}(\Sigma_v)\gtrsim 1$, $\|\beta\|_{\text{MAX}}\lesssim_{\mathrm{P}}1$, $\|\Sigma_u\|_{\text{MAX}}\leq\|\Sigma_u\|\lesssim_{\mathrm{P}}1$, and (C.8), we have

$$\|\gamma^\intercal\beta^\intercal\Sigma^{-1}\|_{\text{MAX}}\lesssim\|(\Sigma_v^{-1}+\beta^\intercal\Sigma_u^{-1}\beta)^{-1}\|\|\beta^\intercal\Sigma_u^{-1}\|_{\text{MAX}}\lesssim_{\mathrm{P}}\lambda_{\max}((\beta^\intercal\Sigma_u^{-1}\beta)^{-1})\lesssim_{\mathrm{P}}N^{-1}.$$

Hence, we have, for all positive fixed ϵ,

$$\mathrm{P}\left(|\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma^{-1}\beta\gamma|\geq\epsilon|\Sigma,\beta\right)\leq\mathrm{E}\left(\left(\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma^{-1}\beta\gamma\right)^2|\Sigma,\beta\right)/\epsilon^2=o_{\mathrm{P}}(1), \tag{C.12}$$

where the last equality comes from (C.11) and that $\mathrm{E}\left(\sum_{i\leq N}\mathrm{E}\left(\mathrm{E}(\alpha_i|\mathcal{G})^2|\Sigma,\beta\right)\right)\leq\sum_{i\leq N}\mathrm{E}(\alpha_i^2)=o(N)$ by condition (c) of Assumption 1. Since $\mathrm{P}\left(|\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma^{-1}\beta\gamma|\geq\epsilon|\Sigma,\beta\right)\leq 1$ are uniformly bounded for all N (by definition), we obtain by taking expectations on both sides of (C.12) that, for all positive fixed ϵ,

$$\mathrm{P}\left(|\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma^{-1}\beta\gamma|\geq\epsilon\right)=o(1),$$

which is equivalent to (C.10).

Finally, we derive

$$\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma^{-1}\mathrm{E}(\alpha|\mathcal{G})=\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\mathrm{E}(\alpha|\mathcal{G})+o_{\mathrm{P}}(1). \tag{C.13}$$

Following the same derivation for (C.11), we obtain

$$\mathrm{E}\left(\left\|\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\beta\right\|_{\mathrm{F}}^2|\Sigma,\beta\right)\leq\sum_{i\leq N}\mathrm{E}\left(\mathrm{E}(\alpha_i|\mathcal{G})^2|\Sigma,\beta\right)\max_j(\Sigma_u^{-1}\beta\beta^\intercal\Sigma_u^{-1})_{j,j}.$$

Because $\|\beta\|_{\text{MAX}}\lesssim_{\mathrm{P}}1$ and $\lambda_{\min}(\Sigma_u)\gtrsim_{\mathrm{P}}1$, we have

$$\max_j(\Sigma_u^{-1}\beta\beta^\intercal\Sigma_u^{-1})_{j,j}\lesssim\|\Sigma_u^{-1}\beta\|_{\text{MAX}}^2\lesssim_{\mathrm{P}}1.$$

Then given the above result that $\mathrm{E}\left(\sum_{i\leq N}\mathrm{E}\left(\mathrm{E}(\alpha_i|\mathcal{G})^2|\Sigma,\beta\right)\right)=o(N)$, we obtain that $\mathrm{E}\left(\left\|\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\beta\right\|_{\mathrm{F}}^2|\Sigma,\beta\right)=o_{\mathrm{P}}(N)$. Therefore, similar to the derivation of (C.10), we obtain

$$\left\|\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\beta\right\|_{\mathrm{F}}^2=o_{\mathrm{P}}(N).$$

On the other hand, using (C.8), we obtain

$$\|(\Sigma_v^{-1}+\beta^\intercal\Sigma_u^{-1}\beta)^{-1}\| = \lambda_{\min}^{-1}(\Sigma_v^{-1}+\beta^\intercal\Sigma_u^{-1}\beta) \le \lambda_{\max}((\beta^\intercal\Sigma_u^{-1}\beta)^{-1}) \lesssim_{\mathrm{P}} N^{-1}. \tag{C.14}$$

Then, using (C.14), we have

$$\begin{aligned}&\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\beta(\Sigma_v^{-1}+\beta^\intercal\Sigma_u^{-1}\beta)^{-1}\beta^\intercal\Sigma_u^{-1}\mathrm{E}(\alpha|\mathcal{G})\\ &\le \left\|\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\beta\right\|_{\mathrm{F}}^2 \|(\Sigma_v^{-1}+\beta^\intercal\Sigma_u^{-1}\beta)^{-1}\| = o_{\mathrm{P}}(1),\end{aligned}$$

and hence, in light of (C.7), we obtain (C.13).

Given that $\mathrm{E}(r_{t+1}|\mathcal{G}) = \mathrm{E}(\alpha|\mathcal{G}) + \beta\gamma$, it follows from (C.9), (C.10), and (C.13) that

$$\mathrm{E}(r_{t+1}|\mathcal{G})^\intercal\Sigma^{-1}\mathrm{E}(r_{t+1}|\mathcal{G}) = \mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\mathrm{E}(\alpha|\mathcal{G}) + \gamma^\intercal\Sigma_v^{-1}\gamma + o_{\mathrm{P}}(1).$$

In light of (C.6), we conclude the proof. ∎

Proof of Proposition 1. Step 1. We have established in the proof of Theorem 1 that the $\mathcal{G}$-conditional distribution of α_i is the same as the conditional distribution of α_i given $\{(\alpha_i + u_{i,s}) : t-T+1 \le s \le t\}$. Note that $u_{i,s}$ is centered normal, we have that the the conditional probability density of $\{r_{i,s}^* := \alpha_i + u_{i,s}, t-T+1 \le s \le t\}$ given α_i, denoted by $p(r_i^*|\alpha_i,\sigma_i)$, is

$$p(r_i^*|\alpha_i) = \prod_{t-T+1\le s\le t} \sigma^{-1}\phi\left(\frac{r_{i,s}^*-\alpha_i}{\sigma}\right) = \phi(T^{1/2}\sigma^{-1}(\bar{r}_i^*-\alpha_i))f(r_i^*).$$

Here $\bar{r}_i^* = T^{-1}\sum_{t-T+1\le s\le t} r_{i,s}^*$ and $f(r_i^*)$ is a function of r_i^* that does not depend on α_i. Hence, applying Bayes' theorem, we have

$$\begin{aligned}\mathrm{E}(\alpha_i|\mathcal{G}) &= \int xp(\alpha_i = x|r_i^*)dx\\ &= \int x\frac{p(r_i^*|\alpha_i=x)p_\alpha(x)}{\int p(r_i^*|\alpha_i=x')p_\alpha(x')dx'}dx\\ &= \int x\frac{\phi(\widehat{z}_i - T^{1/2}\sigma^{-1}x)p_\alpha(x)}{\int \phi(\widehat{z}_i - T^{1/2}\sigma^{-1}x')p_\alpha(x')dx'}dx = \psi(\widehat{z}_i),\end{aligned}$$

where $\widehat{z}_i = T^{1/2}\sigma^{-1}\bar{r}_i^*$ and $p_\alpha(\cdot)$ is the density of α_i. This concludes the proof of the first statement.

Step 2. Apparently, $\widehat{z}_i = T^{1/2}\sigma^{-1}(\alpha_i + \bar{u}_i)$ is i.i.d. across i, whose conditional distribution given α_i is normal, it follows that its unconditional density function $p(a) = \mathrm{E}\left(\phi(a - T^{1/2}\alpha_i/\sigma)\right)$. By direction calculation and the definition of S^{OPT} in the statement of the proposition, we have

$$\mathrm{E}(\mathrm{E}(\alpha_i|\mathcal{G})^\intercal\Sigma_u^{-1}\mathrm{E}(\alpha_i|\mathcal{G})) = \sum_i \sigma^{-2}\mathrm{E}\left(\mathrm{E}(\alpha_i|\mathcal{G})^2\right) = N\sigma^{-2}\int\psi(a)^2p(a)da = \left(S^{\mathrm{OPT}}\right)^2. \tag{C.15}$$

Now we study $S(\mathcal{G}) = \mathrm{E}(\alpha_i|\mathcal{G})^\intercal\Sigma_u^{-1}\mathrm{E}(\alpha_i|\mathcal{G}) = \sum_i \sigma^{-2}\mathrm{E}(\alpha_i|\mathcal{G})^2$. Using the fact that $a^2 - b^2 =$

$(a-b)^2 + 2b(a-b)$, we have

$$\begin{aligned}
&\mathrm{E}\Big(\big|\mathrm{E}(\alpha_i \mathbb{1}_{\{|\alpha_i|\le c_N\}}|\mathcal{G})^2 - \mathrm{E}(\alpha_i|\mathcal{G})^2\big| \Big) \\
\le\ & \mathrm{E}\left(\mathrm{E}(\alpha_i \mathbb{1}_{\{|\alpha_i|> c_N\}}|\mathcal{G})^2\right) + 2\mathrm{E}\left(\left|\mathrm{E}(\alpha_i|\mathcal{G})\mathrm{E}(\alpha_i \mathbb{1}_{\{|\alpha_i|> c_N\}}|\mathcal{G})\right|\right) \\
\le\ & \mathrm{E}(\alpha_i^2 \mathbb{1}_{\{|\alpha_i|> c_N\}}) + 2\sqrt{\mathrm{E}\left(\mathrm{E}(\alpha_i|\mathcal{G})^2\right)\mathrm{E}\left(\mathrm{E}(\alpha_i \mathbb{1}_{\{|\alpha_i|> c_N\}}|\mathcal{G})^2\right)} \\
\le\ & \mathrm{E}(\alpha_i^2 \mathbb{1}_{\{|\alpha_i|> c_N\}}) + 2\sqrt{\mathrm{E}\left(\mathrm{E}(\alpha_i|\mathcal{G})^2\right)\mathrm{E}(\alpha_i^2 \mathbb{1}_{\{|\alpha_i|> c_N\}})} \\
\le\ & c_N N^{-1} + \sqrt{\mathrm{E}\left(\mathrm{E}(\alpha_i|\mathcal{G})^2\right) c_N N^{-1}}, \qquad \text{(C.16)}
\end{aligned}$$

where the last step comes from the tail condition. Then we have

$$\begin{aligned}
&\mathrm{E}\Big(\Big| \sum_i \sigma^{-2}\mathrm{E}(\alpha_i \mathbb{1}_{\{|\alpha_i|\le c_N\}}|\mathcal{G})^2 - \sum_i \sigma^{-2}\mathrm{E}(\alpha_i|\mathcal{G})^2 \Big| \Big) \\
\le\ & \sigma^{-2}\sum_i \mathrm{E}\left(\left|\mathrm{E}(\alpha_i \mathbb{1}_{\{|\alpha_i|\le c_N\}}|\mathcal{G})^2 - \mathrm{E}(\alpha_i|\mathcal{G})^2\right|\right) \\
\le\ & c_N + \sqrt{\sum_i \mathrm{E}\left(\mathrm{E}(\alpha_i|\mathcal{G})^2\right) c_N} = o\left(1 + S^{\mathrm{OPT}}\right), \qquad \text{(C.17)}
\end{aligned}$$

where the second inequality is a direct result of (C.16), and the last estimate is given by (C.15). From (C.17) and (C.15), it follows, respectively, using Markov's inequality and triangle inequality that

$$\sum_i \sigma^{-2}\mathrm{E}(\alpha_i \mathbb{1}_{\{|\alpha_i|\le c_N\}}|\mathcal{G})^2 = \sigma^{-2}\sum_i \mathrm{E}(\alpha_i|\mathcal{G})^2 + o_{\mathrm{P}}\left(1 + S^{\mathrm{OPT}}\right), \qquad \text{(C.18)}$$

$$\mathrm{E}\Big(\sum_i \sigma^{-2}\mathrm{E}(\alpha_i \mathbb{1}_{\{|\alpha_i|\le c_N\}}|\mathcal{G})^2 \Big) = \left(S^{\mathrm{OPT}}\right)^2 + o\left(1 + S^{\mathrm{OPT}}\right). \qquad \text{(C.19)}$$

Further, we have

$$\begin{aligned}
\mathrm{Var}\Big(\sum_i \sigma^{-2}\mathrm{E}(\alpha_i \mathbb{1}_{\{|\alpha_i|\le c_N\}}|\mathcal{G})^2 \Big) &= \sigma^{-4}\sum_i \mathrm{Var}\left(\mathrm{E}(\alpha_i \mathbb{1}_{\{|\alpha_i|\le c_N\}}|\mathcal{G})^2\right) \\
&\le c_N^2\sigma^{-4}\sum_i \mathrm{E}\left(\mathrm{E}(\alpha_i \mathbb{1}_{\{|\alpha_i|\le c_N\}}|\mathcal{G})^2\right) = o(1 + \left(S^{\mathrm{OPT}}\right)^2). \quad \text{(C.20)}
\end{aligned}$$

For the first line, we use that $\mathrm{E}(\alpha_i \mathbb{1}_{\{|\alpha_i|\le c_N\}}|\mathcal{G})$ is independent across i. The second line is obvious as $|\alpha_i|\mathbb{1}_{\{|\alpha_i|\le c_N\}} \le c_N$. The last line comes from (C.19). Combining (C.19) and (C.20), we obtain

$$\sum_i \sigma^{-2}\mathrm{E}(\alpha_i \mathbb{1}_{\{|\alpha_i|\le c_N\}}|\mathcal{G})^2 = \left(S^{\mathrm{OPT}}\right)^2 + o\left(1 + S^{\mathrm{OPT}}\right) + o_{\mathrm{P}}\left(1 + \left(S^{\mathrm{OPT}}\right)^2\right)^{1/2}.$$

Along with (C.18), we obtain

$$\sum_i \sigma^{-2}\mathrm{E}(\alpha_i|\mathcal{G})^2 = \left(S^{\mathrm{OPT}}\right)^2 + o_{\mathrm{P}}\left(1 + S^{\mathrm{OPT}}\right).$$

In light of the definition of $S(\mathcal{G})$, and the fact that

$$\left(\left(S^{\mathrm{OPT}}\right)^2 + o_{\mathrm{P}}\left(1 + S^{\mathrm{OPT}}\right)\right)^{1/2} = S^{\mathrm{OPT}} + o_{\mathrm{P}}(1),$$

we conclude the proof. ∎

Proof of Corollary 2. Because of the tail condition $\mathrm{E}(\alpha_i^2 \mathbb{1}_{\{|\alpha_i| \geq c_N\}}) \leq c_N N^{-1}$ for some sequence $c_N \to 0$, we have

$$\mathrm{E}\left|\alpha^{\intercal}\alpha - \sum_i \alpha_i^2 \mathbb{1}_{\{|\alpha_i| < c_N\}}\right| = \mathrm{E}\left|\sum_i \alpha_i^2 \mathbb{1}_{\{|\alpha_i| \geq c_N\}}\right| = o(1),$$

which, by Markov's inequality and triangle inequality, respectively, leads to

$$\alpha^{\intercal}\alpha = \sum_i \alpha_i^2 \mathbb{1}_{\{|\alpha_i| < c_N\}} + o_{\mathrm{P}}(1), \quad \mathrm{E}\left(\sum_i \alpha_i^2 \mathbb{1}_{\{|\alpha_i| < c_N\}}\right) = \mu^2 \rho N. \tag{C.21}$$

On the other hand, it holds that

$$\mathrm{Var}\left(\sum_i \alpha_i^2 \mathbb{1}_{\{|\alpha_i| < c_N\}}\right) \leq \sum_i \mathrm{E}(\alpha_i^4 \mathbb{1}_{\{|\alpha_i| < c_N\}}) \leq c_N^2 \sum_i \mathrm{E}(\alpha_i^2) = c_N^2 \mu^2 \rho N. \tag{C.22}$$

Combining (C.21) and (C.22), we obtain

$$\alpha^{\intercal}\alpha = \mu^2 \rho N + o_{\mathrm{P}}(1 + \mu\sqrt{\rho N}).$$

As a result, it holds that

$$S^{\star} = \sigma^{-1}\sqrt{\alpha^{\intercal}\alpha} = \sigma^{-1}\mu(\rho N)^{1/2} + o_{\mathrm{P}}(1). \tag{C.23}$$

Further, in light of the explicit distribution of α in Example 1, we have

$$\psi(a) = \frac{\mu\rho\phi(a - T^{1/2}\mu/\sigma) - \mu\rho\phi(a + T^{1/2}\mu/\sigma)}{(2-2\rho)\phi(a) + \rho\phi(a - T^{1/2}\mu/\sigma) + \rho\phi(a + T^{1/2}\mu/\sigma)}, \tag{C.24}$$

$$\left(S^{\mathrm{OPT}}\right)^2 = \frac{\mu\rho N}{2\sigma^2}\int \psi(a)(\phi(a - T^{1/2}\mu/\sigma) - \phi(a + T^{1/2}\mu/\sigma))da. \tag{C.25}$$

Suppose that for some constant $C > 0$, $T^{1/2}\mu\sigma^{-1} - \sqrt{-2\log\rho} \leq C < \infty$. Then we have

$$\sup_{a \geq C} \frac{\rho\phi(a)}{\phi(a - T^{1/2}\mu/\sigma)} \leq 1. \tag{C.26}$$

On the other hand, in light of (C.24) and (C.25), we have

$$\left(S^{\mathrm{OPT}}\right)^2 \leq \frac{\mu^2\rho N}{\sigma^2}\int \frac{\rho\phi(a)}{(2-2\rho)\phi(a - T^{1/2}\mu/\sigma) + \rho\phi(a)}\phi(a)da.$$

We hence obtain from (C.26) that, for N sufficiently large,

$$\left(S^{\mathrm{OPT}}\right)^2 \leq \frac{\mu^2\rho N}{\sigma^2}\left(1-\frac{1}{2}\Phi(-C)\right).$$

This proves the "if" part, given (C.23) and that $\mu^2\rho N/\sigma^2$ does not vanish. Now suppose $T^{1/2}\mu\sigma^{-1}-\sqrt{-2\log\rho}\to\infty$. Then, for all fixed $x>0$, we have

$$\inf_{a\leq x}\frac{\rho\phi(a)}{\phi(a-T^{1/2}\mu/\sigma)}\to\infty, \quad \inf_{a\leq x}\frac{\phi(a)}{\phi(a-2T^{1/2}\mu/\sigma)}\to\infty. \tag{C.27}$$

Given (C.24), we further have, for all fixed $x>0$,

$$\inf_{a:|a-T^{1/2}\mu/\sigma|\leq x}\psi(a)\to\mu, \qquad \sup_{a:|a+T^{1/2}\mu/\sigma|\leq x}\psi(a)\to-\mu. \tag{C.28}$$

Combining (C.27), (C.28), and (C.25), we obtain, for all fixed $x>0$,

$$\left(S^{\mathrm{OPT}}\right)^2 \geq \frac{\mu^2\rho N}{\sigma^2}(1-o(1)-\Phi(-x)).$$

The "only if" part is proved. ∎

C.2 Proof of Theorem 2

Without loss of generality, Σ_u is assumed to be identity matrix so that we do not need keep track of the nuisance parameter σ^2, which does not influence the relative weights of our portfolio. We introduce the following notation:

$$\widetilde{k}_N=k_N^{-2/3}, \quad \bar{k}_N=k_N^{3/2}, \quad \check{k}_N=k_N^{3/4}, \quad \epsilon_N=(\log N)^{-10},$$

$$a_l=lk_N, \quad L=\{l:-\widetilde{k}_N\leq a_l<a_{l+1}<\widetilde{k}_N\}, \quad B(l)=\{i:a_l\leq|S|^{1/2}\breve{\alpha}_i<a_{l+1}\}.$$

Step 1. Throughout the proof, we (re)use the following short-hand notation, introduced in the statement of Proposition 1, but with T therein replaced by $|S|$:

$$\phi(a,x)=\phi(a-|S|^{1/2}x), \quad p(a)=\mathrm{E}(\phi(a-|S|^{1/2}\alpha_i)).$$

Since $p(a)$ is the density of $|S|^{1/2}\breve{\alpha}_i$, it follows that

$$\mathrm{P}(i\in B(l))=\int_{a_l}^{a_{l+1}}p(a)da=\int_{a_l}^{a_{l+1}}\int_{-\infty}^{\infty}\phi(a,x)p_\alpha(x)dxda.$$

We also introduce some auxiliary quantities, constructed similarly:

$$\phi^*(a,x)=\phi(\sqrt{T/|S|}a-T^{1/2}x), \quad \psi^*(a)=\frac{\mathrm{E}(\alpha_i\phi^*(a,\alpha_i))}{\mathrm{E}(\phi^*(a,\alpha_i))}, \quad p^*(a)=\mathrm{E}(\phi^*(a,\alpha_i)).$$

They are primarily used to help prove that while our algorithm uses a smaller sample of size $|S|$ to estimate α, the optimal feasible Sharpe ratio our algorithm achieves would be as large as S^{OPT} (that is based on a sample of size T).

We also need the following definition:

$$\psi_i = \alpha_i \mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}} + \sum_{l \in L} \bar{\psi}_l \mathbb{1}_{\{i \in B(l)\}},$$

where

$$\bar{\psi}_l = \mathrm{E}(\alpha_i | i \in B(l)) = \mathrm{E}(\alpha_i \mathbb{1}_{\{i \in B(l)\}}) / \mathrm{P}(i \in B(l)).$$

Since α_i is i.i.d., $\bar{\psi}_l$ is a deterministic scaler that does not depend on i. Because $\{B(l) : l \in L\}$ are disjoint for distinct l and that $i \in B(l)$ with $l \in L$ leads to $|\breve{z}_i| \leq \widetilde{k}_N$, it holds that

$$\mathrm{E}(\psi^\intercal \psi) = \sum_{i \leq N} \mathrm{E}(\alpha_i^2 \mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}}) + \sum_{l \in L} \bar{\psi}_l^2 \mathrm{E}(|B(l)|). \tag{C.29}$$

The next two steps will establish

$$\mathrm{E}(\psi^\intercal \psi) = \left(S^{\text{OPT}}\right)^2 + o\left(1 + \left(S^{\text{OPT}}\right)^2\right). \tag{C.30}$$

Step 2. This step analyzes $\mathrm{E}\left(\alpha_i^2 \mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}}\right)$ of (C.29) and will show

$$\mathrm{E}\left(\alpha_i^2 \mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}}\right) = o(N^{-1}) + (1 + o(1)) \int_{|a| > \widetilde{k}_N} \psi^*(a)^2 p^*(a) da. \tag{C.31}$$

Using the tower property of conditional expectations, we can directly evaluate the following expectation (conditioning on α_i first):

$$\mathrm{E}\left(\alpha_i^2 \mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}}\right) = \int_{|a| > \widetilde{k}_N} \int x^2 \phi(a, x) p_\alpha(x) dx da. \tag{C.32}$$

On the other hand, since $p(a) = \int \phi(a, x) p_\alpha(x) dx$, it holds that, for any positive sequence $b_N \to 0$,

$$\begin{aligned}
&\left| |S| \int x^2 \phi(a, x) p_\alpha(x) dx - a^2 p(a) \right| \\
\leq\ & \left| \int_{||S|x^2 - a^2| \leq 3 b_N a^2} (|S| x^2 - a^2) \phi(a, x) p_\alpha(x) dx \right| + \sup_{x : ||S|x^2 - a^2| > 3 b_N a^2} ||S| x^2 - a^2| \phi(a, x) \\
\lesssim\ & 3 b_N a^2 p(a) + \sup_{x : ||S|x^2 - a^2| > 3 b_N a^2} ||S| x^2 - a^2| \exp(-(a - |S|^{1/2} x)^2 / 2).
\end{aligned}$$

Further, it holds that $||S|^{1/2} x - a| \leq b_N |a|$ leads to $||S| x^2 - a^2| \leq (|S|^{1/2} x - a)^2 + 2|a(|S|^{1/2} x - a)| \leq$

$b_N^2|a|^2 + 2b_N|a|^2 \leq 3b_N|a|^2$. As a result, for $|a|$ sufficiently large,

$$\begin{aligned}
&\sup_{x:||S|x^2-a^2|\geq 3b_N a^2} ||S|x^2 - a^2| \exp(-(a-|S|^{1/2}x)^2/2) \\
\leq &\sup_{x:||S|^{1/2}x-a|\geq b_N|a|} ||S|x^2 - a^2| \exp(-(a-|S|^{1/2}x)^2/2) \leq \sup_{y:|y|\geq b_N|a|} (y^2 + 2|ay|)\exp(-y^2/2).
\end{aligned}$$

Therefore, for any positive sequence $b_N \to 0$ and all a satisfying $|a|^{-1} \leq c_N b_N$, we have

$$\left| |S| \int x^2 \phi(a,x) p_\alpha(x) dx - a^2 p(a) \right| \lesssim b_N a^2 \exp(-b_N^2 a^2/2) + b_N a^2 p(a).$$

Then, by selecting b_N as $b_N = \widetilde{k}_N^{-1/5}$ and noting $T - |S| = O(\sqrt{T})$, we obtain that, for all a satisfying $|a| \geq C\widetilde{k}_N$,

$$\left| T \int x^2 \phi(a,x) p_\alpha(x) dx - a^2 p(a) \right| \lesssim c_N N^{-2} \epsilon_N + c_N a^2 p(a). \tag{C.33}$$

Furthermore, by the fact that $\mathrm{E}(\alpha_i^2 \mathbb{1}_{\{|\alpha_i|\geq 1\}}) \leq \mathrm{E}(\alpha_i^2 \mathbb{1}_{\{|\alpha_i|\geq c_N\}}) \leq c_N N^{-1}$ for some sequence $c_N \to 0$, we have

$$\int_{|x|\geq 1} p_\alpha(x) dx \leq \int_{|x|\geq 1} x^2 p_\alpha(x) dx \leq c_N N^{-1}.$$

Note also that when $|x| < 1$, we can find $C > 1$ such that $a \geq C\sqrt{T}$ implies $|a - \sqrt{S}x| \geq (C-1)\sqrt{T}$. Therefore, for $|x| < 1$, we have

$$\int_{|a|\geq CT^{1/2}} \phi(a,x) da \lesssim \int_{|a|\geq (C-1)T^{1/2}} \exp(-a^2/2) da \lesssim T^{-1/2} \exp(-T).$$

Using the above results as well as the fact that $N^d \leq T$ for some $d > 0$, we can bound

$$\begin{aligned}
\int_{|a|\geq CT^{1/2}} p(a) da &\leq \int_{|x|\geq 1} p_\alpha(x) dx + \int_{|x|<1} \int_{|a|\geq CT^{1/2}} \phi(a,x) da p_\alpha(x) dx \\
&\leq \int_{|x|\geq 1} p_\alpha(x) dx + \sup_{x:|x|<1} \int_{|a|\geq CT^{1/2}} \phi(a,x) da \leq c_N N^{-1}.
\end{aligned}$$

This in turn leads to

$$\int \mathbb{1}_{\{p(a)<N^{-1}T^{-1/2}\epsilon_N\}} p(a) da \leq c_N N^{-1} + \int_{|a|<CT^{1/2}} \mathbb{1}_{\{p(a)<N^{-1}T^{-1/2}\epsilon_N\}} p(a) da \leq c_N N^{-1}. \tag{C.34}$$

Using (C.34) and the fact that $\mathrm{E}(\alpha_i^2 \mathbb{1}_{\{|\alpha_i|\geq c_N\}}) \leq c_N N^{-1}$ for some sequence $c_N \to 0$, we have

$$\begin{aligned}
&\int \mathbb{1}_{\{p(a)<N^{-1}T^{-1/2}\epsilon_N\}} \int x^2 \phi(a,x) p_\alpha(x) dx da \\
\leq &\int \mathbb{1}_{\{p(a)<N^{-1}T^{-1/2}\epsilon_N\}} p(a) da + \int\int_{|x|\geq 1} x^2 \phi(a,x) p_\alpha(x) dx da \leq c_N N^{-1}.
\end{aligned} \tag{C.35}$$

Therefore, we have

$$
\begin{aligned}
\mathrm{E}(\alpha_i^2 \mathbb{1}_{\{|\check{z}_i|>\widetilde{k}_N\}}) &= \int_{|a|>\widetilde{k}_N} \mathbb{1}_{\{p(a)\geq N^{-1}T^{-1/2}\epsilon_N\}} \int x^2\phi(a,x)p_\alpha(x)dxda + o(N^{-1}) \\
&= (1+o(1)) \int_{|a|>\widetilde{k}_N} \mathbb{1}_{\{p(a)\geq N^{-1}T^{-1/2}\epsilon_N\}} \frac{a^2}{T} p(a)da + o(N^{-1}). \qquad \text{(C.36)}
\end{aligned}
$$

Here the first equality comes from (C.32) and (C.35), whereas the second line is a directly result of (C.33), that $\int \mathbb{1}_{\{p(a)\geq N^{-1}T^{-1/2}\epsilon_N\}}da \leq NT^{1/2}\epsilon_N^{-1}\int p(a)da \leq NT^{1/2}\epsilon_N^{-1}$.

Next, we analyze the right-hand-side of (C.31) and show it is close to the right-hand-side of (C.36). We note that, for all y, all $j \in \{0,2\}$, and all positive sequence $\lambda_N \leq c_N\epsilon_N$,

$$
\begin{aligned}
|y|^j|\phi(y+\lambda_N y) - \phi(y)| &\leq \sup_{y:|y|\leq\epsilon_N^{-1/4}} |y|^j|\phi(y+\lambda_N y)-\phi(y)| + 2\sup_{y:|y|>\epsilon_N^{-1/4}} |y|^j\phi(y) \\
&\lesssim \lambda_N\epsilon_N^{-1}\phi(y) + \epsilon_N^{-1/2}\exp(-\epsilon_N^{-1/2}/2) \leq c_N\phi(y) + c_N N^{-2}T^{-1/2}\epsilon_N.
\end{aligned}
$$

Then obviously, by taking $\lambda_N = \sqrt{T/|S|}-1$ and noting $T \geq N^d$ for some fixed $d>0$ by assumption, we have $\lambda_N \sim T^{-1/2} \leq c_N\epsilon_N$ and thereby $(y^2+1)\left|\phi\left(\sqrt{T/|S|}y\right) - \phi(y)\right| \leq c_N\phi(y) + c_N N^{-2}T^{-1/2}\epsilon_N$. Hence, denoting $y = a - |S|^{1/2}x$, we have for all a and for $j \in \{0,1\}$,

$$
\begin{aligned}
(a^2/T)^j|p^*(a)-p(a)| &\leq \int 2(y^2+x^2)^j\left|\phi(y)-\phi(\sqrt{T/|S|}y)\right|p_\alpha(x)dx \\
&\leq c_N\int(x^{2j}+1)(\phi(a,x)+N^{-2}T^{-1/2}\epsilon_N)p_\alpha(x)dx \\
&\leq c_N N^{-2}T^{-1/2}\epsilon_N + c_N p(a) + \mathbb{1}_{\{j=1\}}c_N\int_{|x|\geq 1} x^2\phi(a,x)p_\alpha(x)dx. \qquad \text{(C.37)}
\end{aligned}
$$

Further, similar to (C.35), we can write

$$
\begin{aligned}
&\int \mathbb{1}_{\{p(a)<N^{-1}T^{-1/2}\epsilon_N\}} \int x^2\phi^*(a,x)p_\alpha(x)dxda \\
\leq &\int \mathbb{1}_{\{p(a)<N^{-1}T^{-1/2}\epsilon_N\}}p^*(a)da + \int\int_{|x|\geq 1} x^2\phi^*(a,x)p_\alpha(x)dxda \leq c_N N^{-1}. \qquad \text{(C.38)}
\end{aligned}
$$

Here for the last inequality, to bound the first term, on top of (C.34), we also use (C.37) with $j=0$ and that $\int \mathbb{1}_{\{p(a)\geq N^{-1}T^{-1/2}\epsilon_N\}}da \leq NT^{1/2}\epsilon_N^{-1}$ (see below (C.36)). Moreover, noting that by definition $\psi^*(a) = \int x\phi^*(a,x)p_\alpha(x)dx/p^*(a)$, we use Cauchy-Schwarz inequality to obtain

$$
\psi^*(a)^2 \leq p^*(a)^{-1}\int x^2\phi^*(a,x)p_\alpha(x)dx. \qquad \text{(C.39)}
$$

Combining (C.38) and (C.39), we have

$$
\int \mathbb{1}_{\{p(a)<N^{-1}T^{-1/2}\epsilon_N\}}\psi^*(a)^2p^*(a)da \leq c_N N^{-1}. \qquad \text{(C.40)}
$$

We hence obtain

$$\int_{|a|>\widetilde{k}_N} \psi^*(a)^2 p^*(a)da = \int_{|a|>\widetilde{k}_N} \mathbb{1}_{\{p(a)\geq N^{-1}T^{-1/2}\epsilon_N\}} \psi^*(a)^2 p^*(a)da + o(N^{-1}). \tag{C.41}$$

Following the same reasoning behind (C.33), we have, for all a satisfying $|a| \geq C\widetilde{k}_N$,

$$\left| T^{1/2} \int x\phi^*(a,x)p_\alpha(x)dx - ap^*(a) \right| \leq c_N N^{-3/2}\epsilon_N + c_N |a| p^*(a). \tag{C.42}$$

We hence have, for all a satisfying $|a| \geq C\widetilde{k}_N$ and $p(a) \geq N^{-1}T^{-1/2}\epsilon_N$,

$$\left| \psi^*(a) - \frac{a}{T^{1/2}} \right| = \left| \frac{\int x\phi^*(a,x)p_\alpha(x)dx}{p^*(a)} - \frac{a}{T^{1/2}} \right| \leq c_N \frac{|a|}{T^{1/2}} + c_N N^{-1/2}.$$

Here, to obtain the last inequality, we combine (C.42) and the fact that $p^*(a) \geq CN^{-1}T^{-1/2}\epsilon_N$ due to (C.37) and $p(a) \geq N^{-1}T^{-1/2}\epsilon_N$. Therefore, we have

$$\begin{aligned} & \int_{|a|>\widetilde{k}_N} \mathbb{1}_{\{p(a)\geq N^{-1}T^{-1/2}\epsilon_N\}} \left| \psi^*(a) - \frac{a}{T^{1/2}} \right|^2 p^*(a)da \\ \leq \; & c_N \int_{|a|>\widetilde{k}_N} \mathbb{1}_{\{p(a)\geq N^{-1}T^{-1/2}\epsilon_N\}} \frac{a^2}{T} p^*(a)da + c_N N^{-1}. \end{aligned} \tag{C.43}$$

Here we also use that $\int p^*(a)da \leq 1$. On the other hand, it follows from Cauchy-Schwarz inequality that for all $d > 0$,

$$\begin{aligned} & \int_{|a|>\widetilde{k}_N} \mathbb{1}_{\{p(a)\geq N^{-1}T^{-1/2}\epsilon_N\}} \left| \psi^*(a)^2 - \frac{a^2}{T} \right| p^*(a)da \\ \leq \; & \int_{|a|>\widetilde{k}_N} \mathbb{1}_{\{p(a)\geq N^{-1}T^{-1/2}\epsilon_N\}} \left((1+d^{-1}) \left| \psi^*(a) - \frac{a}{T^{1/2}} \right|^2 + d\frac{a^2}{T} \right) p^*(a)da \end{aligned}$$

Then, using (C.41) and (C.43), we obtain

$$\int_{|a|>\widetilde{k}_N} \psi^*(a)^2 p^*(a)da = (1+o(1)) \int_{|a|>\widetilde{k}_N} \mathbb{1}_{\{p(a)\geq N^{-1}T^{-1/2}\epsilon_N\}} \frac{a^2}{T} p^*(a)da + o(N^{-1}). \tag{C.44}$$

On the other hand, combining (C.37) with $j = 1$, that $\int \mathbb{1}_{\{p(a)\geq N^{-1}T^{-1/2}\epsilon_N\}} da \leq NT^{1/2}\epsilon_N^{-1}$, and that $\int \int_{|x|\geq 1} x^2 \phi(a,x)p_\alpha(x)dxda \leq c_N N^{-1}$ as in (C.35), we can replace $p^*(a)$ in the right-hand-side of (C.44) with $p(a)$:

$$\int_{|a|>\widetilde{k}_N} \mathbb{1}_{\{p(a)\geq N^{-1}T^{-1/2}\epsilon_N\}} \frac{a^2}{T} p(a)da = (1+o(1)) \int_{|a|>\widetilde{k}_N} \mathbb{1}_{\{p(a)\geq N^{-1}T^{-1/2}\epsilon_N\}} \frac{a^2}{T} p^*(a)da + o(N^{-1}). \tag{C.45}$$

Given this result, the right-hand-sides of (C.36) and (C.44) are the same, and we obtain (C.31).

Step 3. Next, we analyze $\sum_{l\in L}\bar{\psi}_l^2\mathrm{E}(|B(l)|)$ of (C.29). We define

$$\pi(a)=\int x\phi(a,x)p_\alpha(x)dx, \quad \pi^*(a)=\int x\phi^*(a,x)p_\alpha(x)dx,$$

$$\bar{p}(l)=k_N^{-1}\mathrm{P}(i\in B(l)), \quad \bar{\pi}(l)=k_N^{-1}\int x\mathrm{P}(i\in B(l)|\alpha_i=x)p_\alpha(x)dx.$$

We note that, by definition,

$$\psi^*(a)=\pi^*(a)/p^*(a), \quad \bar{\psi}_l=\bar{\pi}(l)/\bar{p}(l). \tag{C.46}$$

We have that, for all $(a,a',\bar{a})$ satisfying $|a|\lesssim\tilde{k}_N$, $|a-a'|\lesssim k_N$, and $|a-\bar{a}|\lesssim\check{k}_N$,

$$\begin{aligned}
|p(\bar{a})-p^*(a)| &\leq \int|\phi(\bar{a},x)-\phi^*(a,x)|p_\alpha(x)dx\\
&= \int|\phi(\bar{a},x,1)-\phi(\tau a,x,\tau^{-1})|p_\alpha(x)dx\\
&\leq \int(c_N\phi(\tau a,x,\tau^{-1})+c_N\epsilon_N N^{-1})p_\alpha(x)dx=c_Np^*(a)+c_N\epsilon_N N^{-1}, \qquad \text{(C.47)}\\
|\pi(a')-\pi^*(a)| &\leq \left|\int x(\phi(a',x)-\phi^*(a,x))p_\alpha(x)dx\right|\\
&= \left|\int x(\phi(a',x,1)-\phi(\tau a,x,\tau^{-1}))p_\alpha(x)dx\right|\leq c_Nf_N(\tau a,\tau^{-1})+c_N\epsilon_N N^{-1}. \qquad \text{(C.48)}
\end{aligned}$$

Here the inequality of the third line comes from (D.127) of Lemma D1, whereas the inequality of the fifth line comes from (D.129) thereof. Both $f_N(a,y)$ and $\phi(a,x,y)$ are introduced in the statement of Lemma D1, and here we use the short-hand notation $\tau:=\sqrt{T/|S|}$. With the same reasoning, we also have that, for all $(a,a',\bar{a})$ satisfying $|a|\lesssim\tilde{k}_N$, $|a-a'|\lesssim k_N$, and $|a-\bar{a}|\lesssim\check{k}_N$,

$$\begin{aligned}
|p^*(\bar{a})-p^*(a)| &\leq c_Np^*(a)+c_N\epsilon_N N^{-1}, \qquad \text{(C.49)}\\
|\pi^*(a')-\pi^*(a)| &\leq c_Nf_N(\tau a,\tau^{-1})+c_N\epsilon_N N^{-1}. \qquad \text{(C.50)}
\end{aligned}$$

Using (C.47) and (C.48), we obtain that, for all $l\in L$,

$$\begin{aligned}
|\bar{p}(l)-p^*(a_l)| &\leq k_N^{-1}\int_{a_l}^{a_{l+1}}|p(a)-p^*(a_l)|da\leq c_Np^*(a_l)+c_N\epsilon_N N^{-1}, \qquad \text{(C.51)}\\
|\bar{\pi}(l)-\pi^*(a_l)| &\leq k_N^{-1}\int_{a_l}^{a_{l+1}}|\pi(a)-\pi^*(a_l)|da\leq c_Nf_N(\tau a_l,\tau^{-1})+c_N\epsilon_N N^{-1}. \qquad \text{(C.52)}
\end{aligned}$$

Further, we define $L':=\{l\in L: p^*(a_l)\geq\epsilon_N N^{-1}\}$. Then we obtain, for all $l\in L'$,

$$\begin{aligned}
|\bar{\psi}_l-\psi^*(a_l)| &= \left|\frac{\bar{\pi}(l)}{\bar{p}(l)}-\frac{\pi^*(a_l)}{\bar{p}(l)}\right|+\left|\frac{\pi^*(a_l)}{\bar{p}(l)}-\frac{\pi^*(a_l)}{p^*(a_l)}\right|\leq(1+c_N)\frac{|\bar{\pi}(l)-\pi^*(a_l)|}{p^*(a_l)}+c_N\frac{|\pi^*(a_l)|}{p^*(a_l)}\\
&\leq c_N\frac{\epsilon_N N^{-1}}{p^*(a_l)}+c_N\frac{f_N(\tau a_l,\tau^{-1})}{p^*(a_l)}+c_N\frac{|\pi^*(a_l)|}{p^*(a_l)}
\end{aligned}$$

$$= \quad c_N \frac{\epsilon_N N^{-1}}{p^*(a_l)} + c_N \sum_{j\in\{-1,0,1\}} |\psi^*(a_l + j\check{k}_N)|. \tag{C.53}$$

Here the first equality comes from (C.46), the first inequality comes from $p^*(a_l) \geq \epsilon_N N^{-1}$ by definition and (C.51), the second inequality is a result of (C.52), whereas the last equality holds by the definition of $f_N(a,y)$. Similarly, using (C.49) and (C.50) instead of (C.51) and (C.52), we obtain that, for all (a,a') satisfying $|a| \lesssim \tilde{k}_N$, $p^*(a) \geq \epsilon_N N^{-1}$, and $|a-a'| \lesssim k_N$,

$$\begin{aligned} |\psi^*(a') - \psi^*(a)| &\leq (1+c_N)\frac{|\bar{\pi}(a') - \pi^*(a)|}{p^*(a)} + c_N \frac{|\pi^*(a)|}{p^*(a)} \\ &\leq c_N \frac{\epsilon_N N^{-1}}{p^*(a)} + c_N \sum_{j\in\{-1,0,1\}} |\psi^*(a + j\check{k}_N)|. \end{aligned} \tag{C.54}$$

Given (C.53), we have, for all $l \in L'$,

$$\begin{aligned} |\bar{\psi}_l^2 - \psi^*(a_l)^2| &\leq (1+c_N^{-1/2})|\bar{\psi}_l - \psi^*(a_l)|^2 + c_N^{1/2}\psi^*(a_l)^2 \\ &\leq c_N^{1/2}\frac{\epsilon_N N^{-1}}{p^*(a_l)} + c_N^{1/2} \sum_{j\in\{-1,0,1\}} \psi^*(a_l + j\check{k}_N)^2, \end{aligned} \tag{C.55}$$

where for the second inequality we use $p^*(a_l) \geq \epsilon_N N^{-1}$ for all $l \in L'$. For the same reason, it follows from (C.54) that, for all (a,a') satisfying $|a| \lesssim \tilde{k}_N$, $p^*(a) \geq \epsilon_N N^{-1}$, and $|a-a'| \lesssim k_N$,

$$|\psi^*(a')^2 - \psi^*(a)^2| \leq c_N \frac{\epsilon_N N^{-1}}{p^*(a)} + c_N \sum_{j\in\{-1,0,1\}} \psi^*(a + j\check{k}_N)^2. \tag{C.56}$$

Next, it follows from Cauchy-Schwarz inequality that

$$\bar{\psi}_l^2 \leq \frac{\mathrm{E}(\alpha_i^2 \mathbb{1}_{\{i\in B(l)\}})}{\mathrm{P}(i\in B(l))} = \frac{\int x^2 \mathrm{P}(i\in B(l)|\alpha_i = x)p_\alpha(x)dx}{\mathrm{P}(i\in B(l))}. \tag{C.57}$$

We then obtain

$$\begin{aligned} &\sum_{l\in L-L'} \bar{\psi}_l^2 \mathrm{P}(i\in B(l)) \leq \sum_{l\in L-L'} \int x^2 \mathrm{P}(i\in B(l)|\alpha_i = x)p_\alpha(x)dx \\ = &\sum_{l\in L-L'} \int_{|x|\geq 1} x^2 \mathrm{P}(i\in B(l)|\alpha_i = x)p_\alpha(x)dx + \sum_{l\in L-L'} \int_{|x|<1} x^2 \mathrm{P}(i\in B(l)|\alpha_i = x)p_\alpha(x)dx \\ \leq &\ \mathrm{E}(\alpha_i^2 \mathbb{1}_{\{|\alpha_i|>1\}}) + \sum_{l\in L-L'} \mathrm{P}(i\in B(l)) \leq c_N N^{-1}. \end{aligned} \tag{C.58}$$

Here the first inequality comes from (C.57), the second is a result of the facts that $\sum_{l\in L-L'} \mathrm{P}(i \in B(l)|\alpha_i = x) \leq 1$, that $\int \mathrm{P}(i\in B(l)|\alpha_i = x)p_\alpha(x)dx = \mathrm{P}(i\in B(l))$, that $\mathrm{P}(i\in B(l)) \lesssim k_N \epsilon_N N^{-1}$ for $l \in L-L'$ due to the definition of L' and (C.52), and that $|L| \lesssim k_N^{-5/3}$. The last inequality comes from $\mathrm{E}(\alpha_i^2 \mathbb{1}_{\{|\alpha_i|\geq 1\}}) \leq \mathrm{E}(\alpha_i^2 \mathbb{1}_{\{|\alpha_i|\geq c_N\}}) \leq c_N N^{-1}$. For the same reason, we have

$$\begin{aligned}
\sum_{l\in L'} \int_{a_l}^{a_{l+1}} \psi^*(a)^2 p^*(a) da &\leq \sum_{l\in L'} \int_{a_l}^{a_{l+1}} \int x^2 \phi^*(a,x) p_\alpha(x) dx da \\
&\leq \mathrm{E}(\alpha_i^2 \mathbb{1}_{\{|\alpha_i|>1\}}) + \sum_{l\in L'} \int_{a_l}^{a_{l+1}} p^*(a) da \leq c_N N^{-1}, \qquad \text{(C.59)}
\end{aligned}$$

Here the first inequality comes from (C.39). For the last inequality we also use (C.49). Noting $\mathrm{E}(|B(l)|) = k_N N \bar{p}(l)$, we obtain

$$\begin{aligned}
&\sum_{l\in L'} \left| \bar{\psi}_l^2 \mathrm{E}(|B(l)|) - N\psi^*(a_l)^2 p^*(a_l) k_N \right| \\
\leq\ & N k_N \sum_{l\in L'} (\bar{\psi}_l^2 |\bar{p}(l) - p^*(a_l)| + |\bar{\psi}_l^2 - \psi^*(a_l)^2| p^*(a_l)) \\
\leq\ & c_N N k_N \sum_{l\in L'} \left(\bar{\psi}_l^2 p^*(a_l) + \epsilon_N N^{-1} + \sum_{j\in\{-1,0,1\}} \psi^*(a_l + j\check{k}_N)^2 p^*(a_l + j\check{k}_N) \right) \\
\leq\ & c_N + c_N N k_N \sum_{j\in\{-1,0,1\}} \sum_{l\in L'} \psi^*(a_l + j\check{k}_N)^2 p^*(a_l + j\check{k}_N), \qquad \text{(C.60)}
\end{aligned}$$

where the second inequality comes from (C.51) and (C.55). Similarly, using (C.49) and (C.56), we obtain

$$\begin{aligned}
&\sum_{l\in L'} \left| \psi^*(a_l)^2 p^*(a_l) k_N - \int_{a_l}^{a_{l+1}} \psi^*(a)^2 p^*(a) da \right| \\
\leq\ & \sum_{l\in L'} \int_{a_l}^{a_{l+1}} (\psi^*(a)^2 |p^*(a_l) - p^*(a)| + |\psi^*(a)^2 - \psi^*(a_l)^2| (p^*(a) + p^*(a_l))) da \\
\leq\ & c_N \sum_{l\in L'} \int_{a_l}^{a_{l+1}} \left(\psi^*(a)^2 p^*(a) + \epsilon_N N^{-1} + \sum_{j\in\{-1,0,1\}} \psi^*(a + j\check{k}_N)^2 p^*(a + j\check{k}_N) \right) da \\
\leq\ & c_N N^{-1} + c_N \sum_{j\in\{-1,0,1\}} \int_{|a|\leq\tilde{k}_N} \psi^*(a + j\check{k}_N)^2 p^*(a + j\check{k}_N) da \\
\leq\ & c_N N^{-1} + c_N N^{-1} \left(S^{\mathrm{OPT}}\right)^2.
\end{aligned}$$

In other words, we have

$$N \sum_{l\in L'} \psi^*(a_l)^2 p^*(a_l) k_N = N \int_{|a|\leq\tilde{k}_N} \psi^*(a)^2 p^*(a) da + o\left(1 + \left(S^{\mathrm{OPT}}\right)^2\right). \qquad \text{(C.61)}$$

Similarly, it holds that, for $j \in \{-1, 1\}$,

$$N \sum_{l\in L'} \psi^*(a_l + j\check{k}_N)^2 p^*(a_l + j\check{k}_N) k_N \lesssim 1 + \left(S^{\mathrm{OPT}}\right)^2. \qquad \text{(C.62)}$$

Substituting (C.61) and (C.62) into (C.60), we obtain

$$\sum_{l\in L'} \bar{\psi}_l^2 \mathrm{E}(|B(l)|) = N\sum_{l\in L'} \int_{a_l}^{a_{l+1}} \psi^*(a)^2 p^*(a)da + o\left(1+\left(S^{\mathrm{OPT}}\right)^2\right). \tag{C.63}$$

Combining (C.58), (C.59), and (C.63), we have

$$\begin{aligned}\sum_{l\in L} \bar{\psi}_l^2 \mathrm{E}(|B(l)|) &= N\sum_{l\in L} \int_{a_l}^{a_{l+1}} \psi^*(a)^2 p^*(a)da + o\left(1+\left(S^{\mathrm{OPT}}\right)^2\right) \\ &= \int_{|a|\le \tilde{k}_N} \psi^*(a)^2 p^*(a)da + o\left(1+\left(S^{\mathrm{OPT}}\right)^2\right). \qquad \text{(C.64)}\end{aligned}$$

Combining (C.29), (C.31), and (C.64), we prove (C.30).

Step 4. This step shows $\psi^\intercal\psi = \left(S^{\mathrm{OPT}}\right)^2 + o_{\mathrm{P}}\left(1+\left(S^{\mathrm{OPT}}\right)^2\right)$. We define

$$\psi_i' = \alpha_i \mathbb{1}_{\{|\breve{z}_i|>\tilde{k}_N\}}\mathbb{1}_{\{|\alpha_i|\le c_N\}} + \sum_{l\in L}\bar{\psi}_l'\mathbb{1}_{\{i\in B(l)\}}, \quad \text{with} \quad \bar{\psi}_l' = \mathrm{E}(\alpha_i \mathbb{1}_{\{|\alpha_i|\le c_N\}}|i\in B(l)).$$

We write

$$\begin{aligned}\mathrm{E}(|\psi_i'-\psi_i|^2) &= \mathrm{E}\left(\left|\alpha_i\mathbb{1}_{\{|\breve{z}_i|>\tilde{k}_N\}}\mathbb{1}_{\{|\alpha_i|>c_N\}} + \sum_{l\in L}\mathrm{E}(\alpha_i\mathbb{1}_{\{|\alpha_i|>c_N\}}|i\in B(l))\mathbb{1}_{\{i\in B(l)\}}\right|^2\right) \\ &\le \mathrm{E}\left(\alpha_i^2\mathbb{1}_{\{|\alpha_i|>c_N\}}\mathbb{1}_{\{|\breve{z}_i|>\tilde{k}_N\}} + \sum_{l\in L}\mathrm{E}(\alpha_i^2\mathbb{1}_{\{|\alpha_i|>c_N\}}|i\in B(l))\mathbb{1}_{\{i\in B(l)\}}\right) \\ &= \mathrm{E}(\alpha_i^2\mathbb{1}_{\{|\alpha_i|>c_N\}}) \le c_N N^{-1},\end{aligned}$$

where the last inequality holds by assumption. As a direct result, we have

$$N\mathrm{E}(|(\psi_i')^2-\psi_i^2|) \le N\mathrm{E}(|\psi_i'-\psi_i|^2 + 2\psi_i|\psi_i'-\psi_i|) \le c_N + c_N N^{1/2}\mathrm{E}(\psi_i^2)^{1/2} = o\left(1+S^{\mathrm{OPT}}\right). \tag{C.65}$$

Here the last inequality comes from (C.30). From (C.65) and (C.30), it follows, respectively, using Markov's inequality and triangle inequality that

$$\begin{aligned}(\psi')^\intercal\psi' &= \psi^\intercal\psi + o_{\mathrm{P}}\left(1+S^{\mathrm{OPT}}\right), &\text{(C.66)}\\ \mathrm{E}((\psi')^\intercal\psi') &= \left(S^{\mathrm{OPT}}\right)^2 + o\left(1+\left(S^{\mathrm{OPT}}\right)^2\right). &\text{(C.67)}\end{aligned}$$

Further, we have

$$\mathrm{Var}((\psi')^\intercal\psi') = N\mathrm{Var}((\psi_i')^2) \le N\mathrm{E}((\psi_i')^4) \le c_N^2 N\mathrm{E}((\psi_i')^2) = o\left(1+\left(S^{\mathrm{OPT}}\right)^2\right), \tag{C.68}$$

where the first inequality holds because ψ_i' is independent across i, the second inequality comes from that $|\psi_i'|\le c_N$ almost surely by definition, and the last inequality comes from (C.67). Combining

(C.67) and (C.68), we obtain

$$(\psi')^\intercal\psi' = \left(S^{\mathrm{OPT}}\right)^2 + o_{\mathrm{P}}\left(1 + \left(S^{\mathrm{OPT}}\right)^2\right),$$

Along with (C.66), we obtain

$$\psi^\intercal\psi = \left(S^{\mathrm{OPT}}\right)^2 + o_{\mathrm{P}}\left(1 + \left(S^{\mathrm{OPT}}\right)^2\right). \tag{C.69}$$

Step 5. Now we show that the Sharpe ratio of the strategy $\breve{w}$ we construct achieves $\left(S^{\mathrm{OPT}}\right)^2$ asymptotically. We note that

$$\breve{w}_i = \breve{\alpha}_i \mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}} + \sum_{l\in L}(w_l + \bar{w}_l)\mathbb{1}_{\{i\in B(l)\}},$$

$$\text{with}\quad w_l = \frac{\sum_j \alpha_j \mathbb{1}_{\{j\in B(l)\}}}{1\vee|B(l)|},\quad \bar{w}_l = \frac{\sum_j \bar{u}'_j \mathbb{1}_{\{j\in B(l)\}}}{1\vee|B(l)|},\quad \bar{u}'_j = |S'|^{-1}\sum_{s\in S'} u_{j,s}.$$

Therefore, since $\mathbb{1}_{\{i\in B(l)\}}$ for all $l\in L$ and $\mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}}$ are disjoint, we obtain

$$\sum_{i\le N}|\breve{w}_i - \psi_i|^2 = \sum_{i\le N}|\breve{\alpha}_i - \alpha_i|^2\mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}} + \sum_{l\in L}(w_l + \bar{w}_l - \bar{\psi}_l)^2|B(l)|. \tag{C.70}$$

Further, we have

$$\begin{aligned}
\sum_{i\le N}|\breve{\alpha}_i - \alpha_i|^2\mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}} &\lesssim_{\mathrm{P}} \frac{\log N}{T}\sum_{i\le N}\mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}},\\
|S|^{-1}\widetilde{k}_N^2\sum_{i\le N}\mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}} &\le \sum_{i\le N}|S|^{-1}|\breve{z}_i|^2\mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}} = \sum_{i\le N}\breve{\alpha}_i^2\mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}}\\
&\le \sum_{i\le N}(2\alpha_i^2 + 2(\breve{\alpha}_i - \alpha_i)^2)\mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}}.
\end{aligned}$$

Here in the first line we use $\max_{i\le N}|\breve{\alpha}_i - \alpha_i| \lesssim_{\mathrm{P}} T^{-1/2}(\log N)^{1/2}$ by the uniform bound on i.i.d. normally distributed random variables u_t. From these two results and that $T^{-1}\log N \le c_N|S|^{-1}\widetilde{k}_N^2$, it follows

$$\sum_{i\le N}|\breve{\alpha}_i - \alpha_i|^2\mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}} = o_{\mathrm{P}}\left(\sum_{i\le N}\alpha_i^2\mathbb{1}_{\{|\breve{z}_i| > \widetilde{k}_N\}}\right) = o_{\mathrm{P}}\left(1 + \left(S^{\mathrm{OPT}}\right)^2\right), \tag{C.71}$$

where the last equation comes from (C.69).

Now we analyze the second term in (C.70). We start by deriving an auxiliary bound on $\sum_{l\in L}\mathrm{E}(\alpha_i^2|i\in B(l))\mathbb{1}_{\{|B(l)|\ge 1\}}$. First, it holds by definition that

$$\sum_{l\in L}\mathrm{E}(\alpha_i^2\mathbb{1}_{\{|\alpha_i|\le\epsilon_N\}}|i\in B(l)) \le \epsilon_N^2|L| \le c_N. \tag{C.72}$$

On the other hand, we note for all x satisfying $|x| \geq \epsilon_N$,

$$\begin{aligned}\sup_{x:|x|\geq\epsilon_N} \mathrm{E}(\mathbb{1}_{\{|\breve{z}_i|\leq\widetilde{k}_N\}}|\alpha_i = x) &= \sup_{x:|x|\geq\epsilon_N} \int \phi(a,x)\mathbb{1}_{\{|a|\leq\widetilde{k}_N\}}da \leq \int_{|S|^{1/2}\epsilon_N-\widetilde{k}_N}^{\infty} \phi(x)dx \\ &\lesssim ||S|^{1/2}\epsilon_N - \widetilde{k}_N|^{-1} \exp\big(-(|S|^{1/2}\epsilon_N - \widetilde{k}_N)^2/2\big) \leq c_N. \quad \text{(C.73)}\end{aligned}$$

Since $\mathrm{E}(\mathbb{1}_{\{|\breve{z}_i|\leq\widetilde{k}_N\}}|\alpha_i = x) = 1 - \mathrm{E}(\mathbb{1}_{\{|\breve{z}_i|>\widetilde{k}_N\}}|\alpha_i = x)$, it follows that, for all x satisfying $|x| \geq \epsilon_N$,

$$\mathrm{E}(\mathbb{1}_{\{|\breve{z}_i|\leq\widetilde{k}_N\}}|\alpha_i = x) \leq c_N \mathrm{E}(\mathbb{1}_{\{|\breve{z}_i|>\widetilde{k}_N\}}|\alpha_i = x),$$

which in turn leads to

$$\mathrm{E}(\alpha_i^2\mathbb{1}_{\{|\alpha_i|\geq\epsilon_N\}}\mathbb{1}_{\{|\breve{z}_i|\leq\widetilde{k}_N\}}) \leq c_N \mathrm{E}(\alpha_i^2\mathbb{1}_{\{|\breve{z}_i|>\widetilde{k}_N\}}). \tag{C.74}$$

Then, we have

$$\begin{aligned}&\sum_{l\in L} \mathrm{E}(\alpha_i^2\mathbb{1}_{\{|\alpha_i|\geq\epsilon_N\}}|i\in B(l))\mathbb{1}_{\{|B(l)|\geq1\}} \lesssim_{\mathrm{P}} \sum_{l\in L} \mathrm{E}\left(\mathrm{E}(\alpha_i^2\mathbb{1}_{\{|\alpha_i|\geq\epsilon_N\}}|i\in B(l))|B(l)|\right) \\ = \; & N\sum_{l\in L} \mathrm{E}(\alpha_i^2\mathbb{1}_{\{|\alpha_i|\geq\epsilon_N\}}\mathbb{1}_{\{i\in B(l)\}}) = N\mathrm{E}(\alpha_i^2\mathbb{1}_{\{|\alpha_i|\geq\epsilon_N\}}\mathbb{1}_{\{|\breve{z}_i|\leq\widetilde{k}_N\}}) \leq c_N + c_N\left(S^{\mathrm{OPT}}\right)^2. \quad \text{(C.75)}\end{aligned}$$

The first inequality uses the fact that $\mathbb{1}_{\{|B(l)|\geq1\}} \leq |B(l)|$ and Markov's inequality, the last inequality comes from (C.74) and (C.30), and the first equality use the fact that $\mathrm{E}(|B(l)|) = \sum_i \mathrm{E}(\mathbb{1}_{\{i\in B(l)\}}) = N\mathrm{P}(i\in B(l))$. Combining (C.72) and (C.75), we obtain

$$\sum_{l\in L} \mathrm{E}(\alpha_i^2|i\in B(l))\mathbb{1}_{\{|B(l)|\geq1\}} \leq c_N + c_N\left(S^{\mathrm{OPT}}\right)^2. \tag{C.76}$$

Next, we note

$$\begin{aligned}\mathrm{E}(w_l|B(l)) &= \frac{\sum_j \mathrm{E}(\alpha_j|B(l))\mathbb{1}_{\{j\in B(l)\}}}{1\vee|B(l)|} = \frac{\mathrm{E}(\alpha_i|i\in B(l))\times|B(l)|}{1\vee|B(l)|} \\ &= \mathrm{E}(\alpha_i|i\in B(l))\mathbb{1}_{\{|B(l)|\geq1\}} = \bar{\psi}_l\mathbb{1}_{\{|B(l)|>0\}}.\end{aligned}$$

Here we use $\mathrm{E}(\alpha_j|B(l)) = \mathrm{E}(\alpha_i|i\in B(l))\mathbb{1}_{\{j\in B(l)\}} + \mathrm{E}(\alpha_i|i\notin B(l))\mathbb{1}_{\{j\notin B(l)\}}$ to obtain the second inequality. As a result,

$$\mathrm{E}((w_l - \bar{\psi}_l\mathbb{1}_{\{|B(l)|>0\}})^2|B(l)) = \mathrm{Var}(w_l|B(l)) = \frac{\mathrm{Var}(\alpha_i|i\in B(l))}{1\vee|B(l)|}\mathbb{1}_{\{|B(l)|>0\}} \leq \frac{\mathrm{E}(\alpha_i^2|i\in B(l))}{1\vee|B(l)|}. \tag{C.77}$$

On the other hand, noting that $\bar{u}'_j$ and $B(l)$ are independent because of the the sample-splitting, it holds that

$$\mathrm{E}(\bar{w}_l^2|B(l)) \leq \frac{|S'|^{-1}|B(l)|}{(1\vee|B(l)|)^2} \leq \frac{|S'|^{-1}}{1\vee|B(l)|}. \tag{C.78}$$

Therefore, we have

$$\begin{aligned}\sum_{l\in L}(w_l-\bar{\psi}_l)^2|B(l)| &\lesssim_{\mathrm{P}} \mathrm{E}\Big(\sum_{l\in L}(w_l-\bar{\psi}_l)^2|B(l)|\Big) && \text{(C.79)}\\ &\leq \sum_{l\in L}\mathrm{E}(\alpha_i^2|i\in B(l))\mathbb{1}_{\{|B(l)|\geq 1\}}\leq c_N+c_N\left(S^{\mathrm{OPT}}\right)^2, && \text{(C.80)}\\ \sum_{l\in L}\bar{w}_l^2|B(l)| &\lesssim_{\mathrm{P}} \mathrm{E}\Big(\sum_{l\in L}\bar{w}_l^2|B(l)|\Big)\leq |L||S'|^{-1}\leq c_N. && \text{(C.81)}\end{aligned}$$

Here the first inequality of (C.80) comes from Markov's inequality, the second is a result of (C.77), and the last follows from (C.76). The second last inequality in (C.81) is a result of (C.78). Substituting (C.71), (C.80), and (C.81) back into (C.70), we obtain

$$\|\breve{w}-\psi\|^2=\sum_{i\leq N}|\breve{w}_i-\psi_i|^2=o_{\mathrm{P}}\left(1+\left(S^{\mathrm{OPT}}\right)^2\right). \tag{C.82}$$

Then it follows

$$\begin{aligned}|(\breve{w}^\intercal-\psi^\intercal)\psi| &\leq \|\breve{w}-\psi\|\|\psi\|=o_{\mathrm{P}}\left(1+\left(S^{\mathrm{OPT}}\right)^2\right), && \text{(C.83)}\\ |\breve{w}^\intercal\breve{w}-\psi^\intercal\psi| &\leq \|\breve{w}-\psi\|^2+2\|\breve{w}-\psi\|\|\psi\|=o_{\mathrm{P}}\left(1+\left(S^{\mathrm{OPT}}\right)^2\right). && \text{(C.84)}\end{aligned}$$

Here for both (C.83) and (C.84), the first inequalities come from Cauchy-Schwarz, whereas the last inequalities come from (C.82) and (C.69). Next, we write

$$\begin{aligned}\breve{w}^\intercal(\alpha-\psi) &= \sum_{l\in L}(w_l+\bar{w}_l)\sum_{i\leq N}(\alpha_i-\bar{\psi}_l)\mathbb{1}_{\{i\in B(l)\}}=\sum_{l\in L}(w_l+\bar{w}_l)(w_l-\bar{\psi}_l)|B(l)|\\ &\leq \|\breve{w}\|\sqrt{\sum_{l\in L}(w_l-\bar{\psi}_l)^2|B(l)|}=o_{\mathrm{P}}\left(\|\breve{w}\|\left(1+S^{\mathrm{OPT}}\right)\right). && \text{(C.85)}\end{aligned}$$

Here the first two equalities hold by definition, the inequality comes from Cauchy-Schwarz, whereas the last equality is simply (C.80). Given (C.85), we obtain

$$\breve{w}^\intercal\alpha=o_{\mathrm{P}}\left(\|\breve{w}\|\left(1+S^{\mathrm{OPT}}\right)\right)+\breve{w}^\intercal\psi. \tag{C.86}$$

Using (C.86), we have

$$\frac{\breve{w}^\intercal\alpha}{\sqrt{\breve{w}^\intercal\breve{w}}}=o_{\mathrm{P}}\left(1+S^{\mathrm{OPT}}\right)+\frac{\breve{w}^\intercal\psi}{\sqrt{\breve{w}^\intercal\breve{w}}}. \tag{C.87}$$

Further, substituting (C.69) into (C.83) and (C.84), we obtain

$$\breve{w}^\intercal\breve{w}=\left(S^{\mathrm{OPT}}\right)^2+o_{\mathrm{P}}\left(1+\left(S^{\mathrm{OPT}}\right)^2\right),\quad \breve{w}^\intercal\psi=\left(S^{\mathrm{OPT}}\right)^2+o_{\mathrm{P}}\left(1+\left(S^{\mathrm{OPT}}\right)^2\right). \tag{C.88}$$

(C.88) directly leads to that, when S^{OPT} does not vanish,

$$\frac{\breve{w}^{\intercal}\psi}{\sqrt{\breve{w}^{\intercal}\breve{w}}} = S^{\mathrm{OPT}} + o_{\mathrm{P}}\left(S^{\mathrm{OPT}}\right). \tag{C.89}$$

On the other hand, when S^{OPT} does vanish, it holds that

$$\frac{\breve{w}^{\intercal}\psi}{\sqrt{\breve{w}^{\intercal}\breve{w}}} \leq \psi^{\intercal}\psi = o_{\mathrm{P}}(1), \tag{C.90}$$

where the first inequality follows from Cauchy-Schwarz inequality and the last equality comes from (C.69). Substituting (C.89) and (C.90) back into (C.87), and using the subsequence argument (see, e.g., Andrews and Cheng (2012)), we obtain

$$\frac{\breve{w}^{\intercal}\alpha}{\sqrt{\breve{w}^{\intercal}\breve{w}}} = S^{\mathrm{OPT}} + o_{\mathrm{P}}\left(1 + S^{\mathrm{OPT}}\right).$$

In other words, we have, for all $\mathrm{P} \in \mathbb{P}$,

$$\lim_{N,T\to\infty} \mathrm{P}\left(\left|\widehat{S}^{\mathrm{OPT}} - S^{\mathrm{OPT}}\right| \geq \epsilon S^{\mathrm{OPT}} + \epsilon\right) = 0. \tag{C.91}$$

Suppose the theorem does not hold, then there is a sequence of data-generating processes P_k with $\mathrm{P}_k \in \mathbb{P}$ for each $k \in \{1, 2, ...\}$ such that

$$\limsup_{N,T\to\infty} \lim_{k\to\infty} \mathrm{P}_k\left(\left|\widehat{S}^{\mathrm{OPT}} - S^{\mathrm{OPT}}\right| \geq \epsilon S^{\mathrm{OPT}} + \epsilon\right) > 0.$$

This contradicts (C.91), and the theorem is proved.

C.3 Proof of Proposition 3

Proof. From Assumption 1, it holds that

$$\alpha^{\intercal}\alpha = \mu^2\rho N + o_{\mathrm{P}}(\mu^2\rho N^{1/2}), \quad \|\alpha^{\intercal}\beta\| = O_{\mathrm{P}}(\mu(\rho N)^{1/2}), \quad \alpha^{\intercal}\bar{u} = O_{\mathrm{P}}(\mu(\rho N)^{1/2}T^{-1/2}), \tag{C.92}$$

$$\bar{u}^{\intercal}\bar{u} = T^{-1}N\varsigma^2 + O_{\mathrm{P}}(T^{-1}N^{1/2}), \quad \|\bar{u}^{\intercal}\beta\| = O_{\mathrm{P}}(T^{-1}N^{1/2}). \tag{C.93}$$

Here the first result of (C.92) is provided above (C.23); the second result of (C.92) comes from (**??**), $\mathrm{E}(\alpha) = 0$, the independence between α and β, and $\|\beta\|_{\mathrm{MAX}} \leq C$ almost surely; the last result of (C.92) also relies on the independence of $u_{i,t}$ across t; the first result of (C.93) comes from the independence of $u_{i,t}$ across (i,t); the second result of (C.93) is a result of the independence of $u_{i,t}$ across (i,t) and $\|\beta\|_{\mathrm{MAX}} \leq C$. Since by definition $\widehat{\alpha}_i = \alpha_i + \bar{u}_i - (\mathbb{P}_{\beta}\alpha)_i - (\mathbb{P}_{\beta}\bar{u})_i$, we obtain, using (C.92), (C.93), and $\lambda_{\min}(\beta^{\intercal}\beta) \geq CN$,

$$\alpha^{\intercal}\widehat{\alpha} = \mu^2\rho N + o_{\mathrm{P}}(\mu(\rho N)^{1/2}), \quad \widehat{\alpha}^{\intercal}\widehat{\alpha} = T^{-1}N\varsigma^2 + \mu^2\rho N + o_{\mathrm{P}}(T^{-1/2}N^{1/2} + \mu(\rho N)^{1/2}). \tag{C.94}$$

Because $\mathrm{E}\left(r_{t+1}^{\intercal}\widehat{w}^{\mathrm{CSR}}|\mathcal{F}_t\right) = \alpha^{\intercal}\widehat{\alpha}$ and $\mathrm{Var}\left(r_{t+1}^{\intercal}\widehat{w}^{\mathrm{CSR}}|\mathcal{F}_t\right) = \varsigma^2\widehat{\alpha}^{\intercal}\widehat{\alpha}$, we prove $\widehat{S}^{\mathrm{CSR}} = S^{\mathrm{CSR}} + o_{\mathrm{P}}(1)$ directly from (C.94).

Now we prove (16) is indeed the sufficient and necessary condition. By defintion we can write

$$\left(S^{\mathrm{OPT}}\right)^2 = N\mu^2\rho^2\varsigma^{-2}\int \frac{\exp(-T\mu^2\varsigma^{-2})\sinh(aT^{1/2}\mu\varsigma^{-1})^2}{(1-\rho)+\rho\exp\left(-\frac{1}{2}T\mu^2\varsigma^{-2}\right)\cosh(aT^{1/2}\mu\varsigma^{-1})}\phi(a)da. \tag{C.95}$$

Suppose $T^{1/2}\mu \leq C$. Then it holds

$$\sup_{a:|a|\leq 2\sqrt{\log N}} \cosh(aT^{1/2}\mu\varsigma^{-1}) = o(\rho^{-1}), \qquad \sup_{a:a>2\sqrt{\log N}} \Phi(-a) = o(N^{-1}). \tag{C.96}$$

Combining (C.95) and (C.96), we obtain

$$\begin{aligned}\left(S^{\mathrm{OPT}}\right)^2 &= (1+o(1))N\mu^2\rho^2\varsigma^{-2}\int \exp(-T\mu^2\varsigma^{-2})\sinh(aT^{1/2}\mu\varsigma^{-1})^2\phi(a)da + o(1) \\ &= (1+o(1))N\mu^2\rho^2\varsigma^{-2}\sinh(T\mu^2\varsigma^{-2}) + o(1).\end{aligned} \tag{C.97}$$

Now we suppose $T^{1/2}\mu \to \infty$. Using (C.24), we have, for all fixed $x > 0$,

$$\inf_{a:|a-T^{1/2}\mu|\leq x} \mu^{-1}\rho^{-1}\psi(a,\varsigma,T) \to \infty, \qquad \sup_{a:|a+T^{1/2}\mu|\leq x} \mu^{-1}\rho^{-1}\psi(a,\varsigma,T) \to -\infty, \tag{C.98}$$

$$\inf_{a:|a-T^{1/2}\mu|\leq x} \frac{\phi(a-T^{1/2}\mu/\varsigma)}{\phi(a+T^{1/2}\mu/\varsigma)} \to \infty, \qquad \inf_{a:|a+T^{1/2}\mu|\leq x} \frac{\phi(a+T^{1/2}\mu/\varsigma)}{\phi(a-T^{1/2}\mu/\varsigma)} \to \infty. \tag{C.99}$$

Combining (C.25), (C.98), and (C.99), we have, for some $\lambda_N \to \infty$,

$$\left(S^{\mathrm{OPT}}\right)^2 = \lambda_N TN\mu^4\rho^2\varsigma^{-4} + o(1). \tag{C.100}$$

Next, we suppose $T^{1/2}\mu \geq C\rho^{-1}$. Then it follows from (C.24) that, for all fixed x,

$$\inf_{a:|a-T^{1/2}\mu|\leq x} \psi(a,\varsigma,T) \to \mu, \qquad \sup_{a:|a+T^{1/2}\mu|\leq x} \mu^{-1}\rho^{-1}\psi(a,\varsigma,T) \to -\mu, \tag{C.101}$$

Given (C.25) and (C.101), we conclude

$$\left(S^{\mathrm{OPT}}\right)^2 \geq N\rho\mu^2\varsigma^{-2}(1+o(1)). \tag{C.102}$$

Comparing the definition of S^{CSR} with (C.97), (C.100), and (C.102), we obtain both the "if" and "only if" parts and conclude the proof. ∎

C.4 Proof of Proposition 4

Proof of Proposition 4, Part 1. In the current part, we prove $\widehat{S}^{\mathrm{BH}} - S^{\mathrm{BH}} = o_{\mathrm{P}}\left(S^{\mathrm{BH}}+1\right)$. We let $\widehat{z} = -\Phi^{-1}(p_{(\widehat{k})}/2)$ and $\widehat{z}' = -\Phi^{-1}(p_{(\widehat{k}+1)}/2)$, where we recall Φ is the standard normal cdf. In other words, $\widehat{z}$ and $\widehat{z}'$ are the t-statistics whose p-values, calculated based on standard normal distribution, are $p_{(\widehat{k})}$ and $p_{(\widehat{k}+1)}$. We also set $c_\lambda = \sqrt{(2-\lambda)\log N}$. We note, for all sequences $a_N \to \infty$,

$$a_N \Phi(-a_N)/\phi(a_N) \to 1. \tag{C.103}$$

As a result, it holds, for all fixed $0 < \lambda \leq 2$,

$$N\Phi(-c_\lambda) \to \infty. \tag{C.104}$$

We further define

$$H_0 = \{i \leq N : \alpha_i = 0\}, \quad H_+ = \{i \leq N : \alpha_i = \mu\}, \quad H_- = \{i \leq N : \alpha_i = -\mu\},$$

$$m_0(a) = \sum_{i \in H_0} \mathbb{1}_{\{|\widehat{z}_i| \geq a\}}, \quad m_+(a) = \sum_{i = H_+} \mathbb{1}_{\{|\widehat{z}_i| \geq a\}}, \quad m_-(a) = \sum_{i = H_-} \mathbb{1}_{\{|\widehat{z}_i| \geq a\}},$$

where $\widehat{z}_i = |S|^{1/2}\widehat{\alpha}_i/\widehat{\sigma}_i$ is the t-statistic of stock i. From the definitions of $\widehat{z}$ and $\widehat{z}'$, we obtain

$$\frac{2N\Phi(-\widehat{z})}{m_0(\widehat{z}) + m_+(\widehat{z}) + m_-(\widehat{z})} \leq \tau, \quad \frac{2N\Phi(-\widehat{z}')}{m_0(\widehat{z}') + m_+(\widehat{z}') + m_-(\widehat{z}')} > \tau, \tag{C.105}$$

$$m_0(\widehat{z}') + m_+(\widehat{z}') + m_-(\widehat{z}') = m_0(\widehat{z}) + m_+(\widehat{z}) + m_-(\widehat{z}) + 1. \tag{C.106}$$

From (**??**) and using $T \leq CN^\lambda$ for some fixed $\lambda < 1$, there exists a deterministic sequence $b_N \sim N^\lambda$ for some fixed $\lambda < 0$, such that

$$\mathrm{P}(\mathbb{1}_{\{|S|^{1/2}(\alpha_i + \bar{u}_i)\widetilde{\sigma}_i^{-1} \geq a - b_N\}} \leq \mathbb{1}_{\{|\widehat{z}_i| \geq a\}} \leq \mathbb{1}_{\{|S|^{1/2}(\alpha_i + \bar{u}_i)\widetilde{\sigma}_i^{-1} \geq a + b_N\}}, \forall a, \forall 1 \leq i \leq N) \to 1. \tag{C.107}$$

Using (C.103), we have, for all deterministic positive sequences (a_N, b_N) satisfying $N\Phi(-a_N) \to \infty$ and $b_N \sim N^\lambda$ with fixed $\lambda < 0$,

$$\Phi(-a_N \pm b_N) = \Phi(-a_N)(1 + o(1)). \tag{C.108}$$

Given (C.107) and (C.108), and noting $(\alpha_i + \bar{u}_i)\widetilde{\sigma}_i^{-1}$ is independent across i, we have, for all deterministic positive sequences (a_N, a_N') satisfying $N\Phi(-a_N) \to \infty$ and $\rho N\Phi(T^{1/2}\mu/\varsigma - a_N') \to \infty$,

$$m_0(a_N) = 2N\Phi(-a_N)\left(1 + o_{\mathrm{P}}(1)\right), \tag{C.109}$$

$$m_\pm(a_N) = \frac{\rho}{2}N(\Phi(T^{1/2}\mu/\varsigma - a_N') + \Phi(-T^{1/2}\mu/\varsigma - a_N'))\left(1 + o_{\mathrm{P}}(1)\right). \tag{C.110}$$

On the other hand, from the definitions of (m_0, m_+, m_-), it follows

$$\begin{aligned} \mathrm{E}\left(r_{t+1}^\intercal \widehat{w}^{\mathrm{BH}} | \mathcal{F}_t\right) &= \mu m_+(\widehat{z}) + \mu m_-(\widehat{z}) - \alpha^\intercal \mathbb{P}_\beta \widehat{\alpha}^{\mathrm{BH}}(\tau), \\ \mathrm{Var}\left(r_{t+1}^\intercal \widehat{w}^{\mathrm{BH}} | \mathcal{F}_t\right) &= \varsigma^2(m_0(\widehat{z}) + m_+(\widehat{z}) + m_-(\widehat{z})) - \varsigma^2 \widehat{\alpha}^{\mathrm{BH}}(\tau)^\intercal \mathbb{P}_\beta \widehat{\alpha}^{\mathrm{BH}}(\tau). \end{aligned}$$

Further, using $\lambda_{\min}(\beta^\intercal\beta) \geq CN$ and $\|\beta\|_{\mathrm{MAX}} \leq C$ from Assumption 1, and the second part of (C.92), we have

$$\begin{aligned} \alpha^\intercal \mathbb{P}_\beta \widehat{\alpha}^{\mathrm{BH}}(\tau) &\leq \|\alpha^\intercal\beta\| \lambda_{\min}(\beta^\intercal\beta)^{-1}\|\beta\|_{\mathrm{MAX}}|\widehat{\alpha}^{\mathrm{BH}}(\tau)|_1 = o_{\mathrm{P}}(N^{-1/2}(m_0(\widehat{z})+m_+(\widehat{z})+m_-(\widehat{z}))), \\ \widehat{\alpha}^{\mathrm{BH}}(\tau)^\intercal \mathbb{P}_\beta \widehat{\alpha}^{\mathrm{BH}}(\tau) &\leq \left(\|\beta\|_{\mathrm{MAX}}\left|\widehat{\alpha}^{\mathrm{BH}}(\tau)\right|_1\right)^2 \lambda_{\min}(\beta^\intercal\beta)^{-1} = O_{\mathrm{P}}(N^{-1}(m_0(\widehat{z})+m_+(\widehat{z})+m_-(\widehat{z}))^2). \end{aligned}$$

Therefore, from the definitions of $\widehat{S}^{\mathrm{BH}}$, it holds that, under $m_0(\widehat{z})+m_+(\widehat{z})+m_-(\widehat{z}) = o_{\mathrm{P}}(N)$,

$$\widehat{S}^{\mathrm{BH}} = (1+o_{\mathrm{P}}(1))\mu\varsigma^{-1}\frac{m_+(\widehat{z})+m_-(\widehat{z})}{\sqrt{m_0(\widehat{z})+m_+(\widehat{z})+m_-(\widehat{z})}} + o_{\mathrm{P}}(1). \tag{C.111}$$

Now suppose $z^* \geq c_\lambda$ for all fixed $0<\lambda\leq 2$. Then by direct calculations it follows from (20) that, for all fixed $0<\lambda\leq 2$,

$$\frac{\rho\Phi(T^{1/2}\mu/\varsigma - c_\lambda)}{\Phi(-c_\lambda)} \to 0. \tag{C.112}$$

Using (C.104), (C.112), (C.109), (C.110) and the monotonicity of Φ, we obtain, for all fixed $0 < \lambda \leq 2$,

$$m_0(c_\lambda) = 2N\Phi(-c_\lambda)\left(1+o_{\mathrm{P}}(1)\right), \quad m_\pm(c_\lambda) = o_{\mathrm{P}}(N\Phi(-c_\lambda)).$$

Hence, from (C.105) and (C.106), we conclude, for all fixed $0<\lambda\leq 2$,

$$\mathrm{P}(\widehat{z} \geq c_\lambda) \to 1. \tag{C.113}$$

Combining (C.113), (C.112), and (C.110), we have, for all fixed $0<\lambda\leq 2$,

$$\mathrm{P}(m_0(\widehat{z})+m_+(\widehat{z})+m_-(\widehat{z}) \leq CN^\lambda) \to 1.$$

Then from (C.111) and that $\mu \leq CN^\lambda$ for some fixed $\lambda<0$, it follows $\widehat{S}^{\mathrm{BH}} = o_{\mathrm{P}}(1)$. On the other hand, from (C.112) it follows $\rho N\Phi(T^{1/2}\mu/\varsigma - z^*) \leq CN^\lambda$ for all fixed $0<\lambda\leq 2$. Hence $S^{\mathrm{BH}} = o(1)$ and we prove $\widehat{S}^{\mathrm{BH}} - S^{\mathrm{BH}} = o_{\mathrm{P}}\left(S^{\mathrm{BH}}+1\right)$ under $z^* \geq c_\lambda$.

Next, we suppose $z^* \leq c_\lambda$ for some fixed $0<\lambda\leq 2$. Then, using (20) and (C.103), it holds that for some fixed $0<\lambda\leq 2$,

$$\frac{\rho\Phi(T^{1/2}\mu/\varsigma - c_\lambda)}{\Phi(-c_\lambda)} \to \infty. \tag{C.114}$$

We combine (C.104), (C.114), (C.109), and (C.110) to conclude that, for some fixed $0<\lambda\leq 2$ and in probability,

$$\frac{m_\pm(c_\lambda)}{m_0(c_\lambda)} \to \infty.$$

It then follows from (C.105) and (C.106) that $\mathrm{P}(\widehat{z} \leq c_\lambda) \to 1$ for some fixed $0<\lambda\leq 2$. Given (C.104) and (C.114), we have, in probability

$$N\Phi(-\widehat{z}) \to \infty, \quad \rho N\Phi(T^{1/2}\mu/\varsigma - \widehat{z}) \to \infty. \tag{C.115}$$

Applying equation (13) of Liu and Shao (2014) to (C.115), and using (C.107) and (C.108), we obtain

$$m_0(\widehat{z}) = 2N\Phi(-\widehat{z})\left(1+o_{\mathrm{P}}(1)\right), \quad m_{\pm}(\widehat{z}) = \frac{\rho}{2}N(\Phi(T^{1/2}\mu/\varsigma - \widehat{z}) + \Phi(-T^{1/2}\mu/\varsigma - \widehat{z}))\left(1+o_{\mathrm{P}}(1)\right). \tag{C.116}$$

Since $\widehat{z}' \geq \widehat{z}$, (C.116) would still hold if all $\widehat{z}$ are replaced by $\widehat{z}'$. Hence, substituting (C.116) back into (C.105) and (C.106), and noting $\Phi(-T^{1/2}\mu/\varsigma - \widehat{z}) \leq \Phi(-\widehat{z})$, we have

$$\frac{2(1-\tau)\Phi(-\widehat{z})}{\tau\rho\Phi(T^{1/2}\mu/\varsigma - \widehat{z})} = 1 + o_{\mathrm{P}}(1). \tag{C.117}$$

Substituting (C.117) back into (C.116), we obtain

$$m_0(\widehat{z}) = \frac{\tau}{1-\tau}\rho N\Phi(T^{1/2}\mu/\varsigma - \widehat{z})\left(1+o_{\mathrm{P}}(1)\right), \quad m_{\pm}(\widehat{z}) = \frac{\rho}{2}N\Phi(T^{1/2}\mu/\varsigma - \widehat{z})\left(1+o_{\mathrm{P}}(1)\right). \tag{C.118}$$

Next, using (20) and (C.103), we note that $z^* \leq c_\lambda$ for some fixed $0 < \lambda \leq 2$ leads to that $T^{1/2}\mu \geq c_{\lambda'}$ for some fixed $\lambda' < 2$. As a result, using (C.103), and comparing (20) and (C.117), we have

$$\Phi(T^{1/2}\mu/\varsigma - \widehat{z}) = \Phi(T^{1/2}\mu/\varsigma - z^*)\left(1+o_{\mathrm{P}}(1)\right). \tag{C.119}$$

Given (C.111), (C.118), and (C.119), we obtain $\widehat{S}^{\mathrm{BH}} - S^{\mathrm{BH}} = o_{\mathrm{P}}\left(S^{\mathrm{BH}}+1\right)$ under $z^* \leq c_\lambda$. ∎

Proof of Proposition 4, Part 2. In this part, we demonstrate (21) is the correct sufficient and necessary condition. Clearly, for all sequences (a_N, a'_N) satisfying $a_N \geq 0$ and $a'_N \to \infty$, we have $\phi(a_N + a'_N)/\phi(a_N) \to 0$. Then using the expression of S^{OPT} given by (C.25), we obtain, under $T^{1/2}\mu \to \infty$,

$$\left(S^{\mathrm{OPT}}\right)^2 = (1+o(1))\mu^2\varsigma^{-2}\rho N \int \frac{\rho\phi(a - T^{1/2}\mu/\varsigma)}{(2-\rho)\phi(a) + \rho\phi(a - T^{1/2}\mu/\varsigma)}\phi(a - T^{1/2}\mu/\varsigma)da. \tag{C.120}$$

First suppose $z^* \leq T^{1/2}\mu\varsigma^{-1}$. Applying (C.103) to (20), we obtain $(z^*+1)^{-1}\phi(z^*) \sim \rho$, which leads to

$$z^* = \sqrt{-2\log\rho} + o\left(\sqrt{\log N}\right) \quad \text{and} \quad \sup_{a \leq z^*} \frac{\rho\phi(a - T^{1/2}\mu/\varsigma)}{\phi(a)} \leq C(z^*+1)^{-1} \to 0. \tag{C.121}$$

Given the first pat of (C.121), the condition (21) never holds, and we have $T^{1/2}\mu \to \infty$ so we can apply (C.120). From the second part of (C.121), we directly obtain

$$\int \frac{\rho\phi(a - T^{1/2}\mu/\varsigma)}{(2-\rho)\phi(a) + \rho\phi(a - T^{1/2}\mu/\varsigma)}\phi(a - T^{1/2}\mu/\varsigma)da \leq \Phi(T^{1/2}\mu/\varsigma - z^*) + o(\Phi(z^* - T^{1/2}\mu/\varsigma)). \tag{C.122}$$

Substituting (C.122) back into (C.120), and using $z^* \leq T^{1/2}\mu\varsigma^{-1}$, we have $\left(S^{\mathrm{OPT}}\right)^2 = (1 + o(1))\mu^2\varsigma^{-2}\rho N\Phi(T^{1/2}\mu/\varsigma - z^*)$. Comparing with the definition of S^{BH}, we conclude that $S^{\mathrm{BH}} \leq (1-\epsilon)\sqrt{1-\tau}S^{\mathrm{OPT}}$ does not hold either.

Now we suppose $z^* > T^{1/2}\mu\varsigma^{-1}$. Again applying (C.103) to (20), we obtain

$$\frac{z^* - T^{1/2}\mu\varsigma^{-1} + 1}{z^*} \frac{\phi(z^*)}{\phi(z^* - T^{1/2}\mu/\varsigma)} \sim \rho, \tag{C.123}$$

from which it follows

$$z^* = \frac{T^{1/2}\mu\varsigma^{-1}}{2} - \frac{\log\rho}{T^{1/2}\mu\varsigma^{-1}} + o\left(\sqrt{\log N}\right). \tag{C.124}$$

If $z^*/\sqrt{\log N} \to \infty$, then we have $S^{\mathrm{BH}} = o(1)$ (see after (C.113)). Given (C.124), we have $T^{1/2}\mu\varsigma^{-1} = o\left(\sqrt{\log N}\right)$, hence (21) always holds and is indeed the correct condition.

If $z^* - T^{1/2}\mu\varsigma^{-1} = o\left(\sqrt{\log N}\right)$, then (21) does not hold, and we have $T^{1/2}\mu \to \infty$ from (C.124). Further, by direct calculations we obtain from (C.103) and (C.123) that, for some $b_N \to \infty$ and all $a \le z^*$,

$$\Phi(T^{1/2}\mu/\varsigma - z^*) \ge b_N \frac{\phi(z^* - T^{1/2}\mu/\varsigma)}{T^{1/2}\mu/\varsigma} \quad \text{and} \quad \frac{\rho\phi(a - T^{1/2}\mu/\varsigma)}{\phi(a)} = o(\exp(T^{1/2}\mu/\varsigma(a - z^*))).$$

Therefore, $S^{\mathrm{BH}} \le (1-\epsilon)\sqrt{1-\tau}S^{\mathrm{OPT}}$ does not hold due to (C.120), and (21) is again the correct condition.

If $z^* - T^{1/2}\mu\varsigma^{-1} \sim \sqrt{\log N}$, then from (C.124) it holds $T^{1/2}\mu \to \infty$ and (C.120) is applicable. To study the integral in the r.h.s. of (C.120), here and only here we use notation $b := \lim_{N,T\to\infty}(z^* - T^{1/2}\mu/\varsigma)^{-1}T^{1/2}\mu/\varsigma$. We implicitly assume the limit exists, which is without loss of generality. From (C.103) and (20), and by direct calculations, we have the following two results. When $b > 1$, it holds

$$\Phi(T^{1/2}\mu/\varsigma - z^*)^{-1} \int \frac{\rho\phi(a - T^{1/2}\mu/\varsigma)}{(2-\rho)\phi(a) + \rho\phi(a - T^{1/2}\mu/\varsigma)} \phi(a - T^{1/2}\mu/\varsigma)da \to (1-\tau)\int_0^\infty \frac{dx}{1 + \frac{(1+b)\tau}{1-\tau}x^b}.$$

When $b \le 1$, it holds

$$\Phi(T^{1/2}\mu/\varsigma - z^*)^{-1} \int \frac{\rho\phi(a - T^{1/2}\mu/\varsigma)}{(2-\rho)\phi(a) + \rho\phi(a - T^{1/2}\mu/\varsigma)} \phi(a - T^{1/2}\mu/\varsigma)da \to \infty.$$

Then, from (C.120) we have $S^{\mathrm{BH}} \le (1-\epsilon)\sqrt{1-\tau}S^{\mathrm{OPT}}$ if and only if $b > 1$ and $\int_0^\infty (1 + \frac{(1+b)\tau}{1-\tau}x^b)^{-1}dx \ge 1 + \epsilon'$ for some fixed $\epsilon' > 0$. For any fixed b, the integral is strictly greater than one for sufficiently small τ. Substituting (C.124) into the definition of b, we obtain (21). We have analyzed all the cases and the proof concludes. ∎

C.5 Sketch of the proof of Proposition 5

The leading order of expected return is

$$\frac{\rho}{2}N\mu\mathrm{E}\left(\mathrm{sgn}(\hat{\alpha}_i)(|\hat{\alpha}_i| - \lambda)_+ | \alpha_i = \mu\right) - \frac{\rho}{2}N\mu\mathrm{E}\left(\mathrm{sgn}(\hat{\alpha}_i)(|\hat{\alpha}_i| - \lambda)_+ | \alpha_i = -\mu\right),$$

which, by symmetry, is just

$$\rho N \mu \mathrm{E}\left(\operatorname{sgn}(\hat{\alpha}_i)(|\hat{\alpha}_i| - \lambda)_+ | \alpha_i = \mu\right) . \tag{C.125}$$

The leading order of variance is

$$\rho N \sigma^2 \mathrm{E}((|\hat{\alpha}_i| - \lambda)_+^2 | \alpha_i = \mu) + (1-\rho) N \sigma^2 \mathrm{E}((|\hat{\alpha}_i| - \lambda)_+^2 | \alpha_i = 0). \tag{C.126}$$

Conditional on $\alpha_i = \mu$, the pdf of $T^{1/2}\sigma^{-1}\hat{\alpha}_i$ is $\mathcal{N}(T^{1/2}\sigma^{-1}\mu, 1)$. Hence I have

$$\begin{aligned}
\mathrm{E}\left(\operatorname{sgn}(\hat{\alpha}_i)(|\hat{\alpha}_i| - \lambda)_+ | \alpha_i = \mu\right) &= \int_{-\infty}^{\infty} \operatorname{sgn}(x)(|x| - \lambda)_+ \phi(T^{1/2}\sigma^{-1}(\mu - x)) d(T^{1/2}\sigma^{-1}x) \\
&= (1 + o(1)) \int_{\lambda}^{\infty} (x - \lambda) \phi(T^{1/2}\sigma^{-1}(\mu - x)) d(T^{1/2}\sigma^{-1}x); \\
\mathrm{E}((|\hat{\alpha}_i| - \lambda)_+^2 | \alpha_i = \mu) &= (1 + o(1)) \int_{\lambda}^{\infty} (x - \lambda)^2 \phi(T^{1/2}\sigma^{-1}(\mu - x)) d(T^{1/2}\sigma^{-1}x); \\
\mathrm{E}((|\hat{\alpha}_i| - \lambda)_+^2 | \alpha_i = 0) &= \int_{-\infty}^{\infty} (|x| - \lambda)_+^2 \phi(T^{1/2}\sigma^{-1}x) d(T^{1/2}\sigma^{-1}x) \\
&= 2 \int_{\lambda}^{\infty} (x - \lambda)^2 \phi(T^{1/2}\sigma^{-1}x) d(T^{1/2}\sigma^{-1}x).
\end{aligned}$$

Substituting this results back into (1) and (2), I obtain

$$S^{\text{Lasso}} = \frac{\rho N \mu \int_{\lambda}^{\infty} (x - \lambda) \phi(T^{1/2}\sigma^{-1}(\mu - x)) d(T^{1/2}\sigma^{-1}x)}{\sqrt{N\sigma^2 \int_{\lambda}^{\infty} (x - \lambda)^2 \big(\rho \phi(T^{1/2}\sigma^{-1}(\mu - x)) + 2(1-\rho)\phi(T^{1/2}\sigma^{-1}x)\big) d(T^{1/2}\sigma^{-1}x)}}.$$

Appendix D Proofs of Technical Lemmas

Lemma D1. *Suppose Assumption 1 holds. Then it holds that, for all* $(a, a', \bar{a}, y, y')$ *satisfying* $|a| \lesssim \widetilde{k}_N$, $|a - a'| \lesssim k_N$, $|\bar{a} - a| \lesssim \check{k}_N$, $C^{-1} \le y \le C$, *and* $|y - y'| \lesssim \bar{k}_N$,

$$\begin{aligned}
|\phi(\bar{a}, x, y') - \phi(a, x, y)| &\le c_N \phi(a, x, y) + c_N N^{-1} \epsilon_N, & \text{(D.127)} \\
\int \min\{\check{k}_N T^{1/2} x^2, |x|\} \phi(a, x, y') p_\alpha(x) dx &\le C f_N(a, y) + c_N N^{-1} \epsilon_N, & \text{(D.128)} \\
\left| \int x(\phi(a', x, y') - \phi(a, x, y)) p_\alpha(x) dx \right| &\le c_N f_N(a, y) + c_N N^{-1} \epsilon_N, & \text{(D.129)}
\end{aligned}$$

where $\phi(a, x, y) = \phi(a - |S|^{1/2} x / y)$, *and* $f_N(a, y)$ *is defined as*

$$f_N(a, y) = \left| \int x \phi(a + \check{k}_N, x, y) p_\alpha(x) dx \right| + \left| \int x \phi(a - \check{k}_N, x, y) p_\alpha(x) dx \right| .$$

Proof of Lemma D1. In the sequel, all the equations hold for all (a, y, y') satisfying the requirements as in the statement of this lemma. For simpler exposition, we assume $|xp_\alpha(x)|$ is upper bounded by $b_N < \infty$ for each N, i.e. we exclude point mass when $x \neq 0$. Eliminating this assumption only requires some heavier notation.

Step 1. For all $a \in \mathbb{R}$ and all $z \leq 0$, we define $h(z; a, y) \in [0, \infty) \cup \{\infty\}$ as

$$h(z; a, y) = \sup_{u \geq 0} u \quad \text{s.t.} \quad \int_z^u x\phi(a, x, y)p_\alpha(x)dx < 0;$$

in the case where the constraint $\int_z^u x\phi(a, x, y)p_\alpha(x)dx < 0$ never holds, we set $h(z; a, y) = 0$. We further define

$$z^*(a, y) = \sup_{z \leq 0} z, \quad \text{s.t.} \quad \max\{|z|, h(z; a, y)\} \geq T^{-1/2}\check{k}_N^{-1};$$

$$A_\alpha(a, y) = [z^*(a, y), h(z^*(a, y); a, y)], \quad A'_\alpha(a, y) = \mathbb{R} - A_\alpha(a, y).$$

Given the definition of h, for all (a, z, y) satisfying $a \in \mathbb{R}$, $z \leq 0$, and $C^{-1} \leq y \leq C$, it holds that, as long as $h(z; a, y) < \infty$,

$$\int_z^{h(z;a,y)} x\phi(a, x, y)p_\alpha(x)dx = 0. \tag{D.130}$$

On the other hand, if $h(z; a, y) = \infty$, then $\int_z^u x\phi(a, x, y)p_\alpha(x)dx < 0$ for all $u < \infty$. Because $|xp_\alpha(x)|$ is upper bounded, there exists $z' > z$ such that $\int_z^u x\phi(a, x, y)p_\alpha(x)dx < 0$ for all $u < \infty$, that is, $h(z'; a, y) = \infty$. Therefore, it follows from the definition of z^* that $h(z^*(a, y); a, y) < \infty$, for all (a, y) satisfying $a \in \mathbb{R}$, and $C^{-1} \leq y \leq C$. As a result, we obtain from (D.130) that

$$\int_{A_\alpha(a,y)} x\phi(a, x, y)p_\alpha(x)dx = 0. \tag{D.131}$$

Next, by definition it holds that, for all $a, \bar{a}, x, y, y' \in \mathbb{R}$,

$$\phi(\bar{a}, x, y') = \phi(a, x, y)\bar{\phi}(a, \bar{a}, x, y')\widetilde{\phi}(a, x, y, y'), \tag{D.132}$$

where

$$\begin{aligned} \bar{\phi}(a, \bar{a}, x, y') &= e^{(a^2 - \bar{a}^2)/2} e^{|S|^{1/2}(\bar{a} - a)x/y'}, \\ \widetilde{\phi}(a, x, y, y') &= e^{|S|x^2(y^{-2} - (y')^{-2})/2} e^{|S|^{1/2}ax((y')^{-1} - y^{-1})}. \end{aligned}$$

It follows from the requirements on (a, y, y') imposed in the statement of this lemma that, for all $(x, \bar{a})$ satisfying $T^{1/2}|x| \lesssim \widetilde{k}_N$ and $|\bar{a} - a| \lesssim \check{k}_N$,

$$\begin{aligned} |a^2 - \bar{a}^2| + |S|^{1/2}|\bar{a} - a||x|/y' &\leq c_N, \\ |S|x^2|y^{-2} - (y')^{-2}| + |S|^{1/2}a|x||(y')^{-1} - y^{-1}| &\leq C\widetilde{k}_N\bar{k}_N T^{1/2}|x| \leq c_N. \end{aligned}$$

Substituting the two lines above into the definitions of $\bar{\phi}$ and $\widetilde{\phi}$, we obtain that, for all $(x, \bar{a})$ satisfying $T^{1/2}|x| \lesssim \widetilde{k}_N$ and $|\bar{a} - a| \lesssim \check{k}_N$.

$$|\bar{\phi}(a, \bar{a}, x, y') - 1| \leq c_N, \quad |\widetilde{\phi}(a, x, y, y') - 1| \leq C\widetilde{k}_N \bar{k}_N T^{1/2}|x|. \tag{D.133}$$

Furthermore, for all $(x, \bar{a})$ satisfying $T^{1/2}|x|\widetilde{k}_N^{-1} \to \infty$ and $|\bar{a} - a| \lesssim \check{k}_N$, by direct calculations we have

$$\phi(\bar{a}, x, y') \leq c_N N^{-1}\epsilon_N, \quad \phi(a, x, y)\bar{\phi}(a, \bar{a}, x, y') \leq c_N N^{-1}\epsilon_N. \tag{D.134}$$

Combining (D.132), (D.133) and (D.134), we obtain that, for all $\bar{a}$ satisfying $|\bar{a} - a| \lesssim \check{k}_N$,

$$\begin{aligned} |\phi(\bar{a}, x, y') - \phi(a, x, y)| &\leq c_N\phi(a, x, y) + c_N N^{-1}\epsilon_N, \\ |\phi(\bar{a}, x, y') - \phi(a, x, y)\bar{\phi}(a, \bar{a}, x, y')| &\leq C\widetilde{k}_N \bar{k}_N T^{1/2}|x|\phi(a, x, y) + c_N N^{-1}\epsilon_N. \end{aligned} \tag{D.135}$$

The first inequality is exactly (D.127).

Step 2. First, noting $\int |x|p_\alpha(x)dx \leq c_N$ by condition (b) of Assumption 1, it follows from (D.127) that, for all $\bar{a}$ satisfying $|\bar{a} - a| \lesssim \check{k}_N$,

$$\left| \int_{A'_\alpha(a,y)} x(\phi(\bar{a}, x, y') - \phi(a, x, y))p_\alpha(x)dx \right| \leq c_N \int_{A'_\alpha(a,y)} |x|\phi(a, x, y)p_\alpha(x)dx + c_N N^{-1}\epsilon_N. \tag{D.136}$$

Next, we obtain from (D.135) that, again for all $\bar{a}$ satisfying $|\bar{a} - a| \lesssim \check{k}_N$,

$$\begin{aligned} &\left| \int_{A_\alpha(a,y)} x(\phi(a', x, y') - \phi(a, x, y)\bar{\phi}(a, \bar{a}, x, y'))p_\alpha(x)dx \right| \\ \leq\ & C\widetilde{k}_N \bar{k}_N T^{1/2} \int_{A_\alpha(a,y)} x^2\phi(a, x, y)p_\alpha(x)dx + c_N N^{-1}\epsilon_N. \end{aligned} \tag{D.137}$$

Further, because $T^{1/2}|x| \leq \check{k}_N^{-1}$ for all $x \in A_\alpha(a, y)$, it holds that, for all $\bar{a}$ satisfying $|\bar{a} - a| \lesssim \check{k}_N$,

$$e^{|S|^{1/2}(\bar{a}-a)x/y'} - 1 \sim T^{1/2}(\bar{a} - a)x.$$

As a result, we have, for all $\bar{a}$ satisfying $|\bar{a} - a| \lesssim \check{k}_N$,

$$\begin{aligned} &\int_{A_\alpha(a,y)} x\phi(a, x, y)(\bar{\phi}(a, \bar{a}, x, y') - e^{(a^2-\bar{a}^2)/2})p_\alpha(x)dx \\ \sim\ & T^{1/2}(\bar{a} - a) \int_{A_\alpha(a,y)} x^2\phi(a, x, y)p_\alpha(x)dx. \end{aligned} \tag{D.138}$$

Combining (D.137) and (D.138), using (D.131), and noting $\widetilde{k}_N \bar{k}_N \leq c_N \check{k}_N$ and $k_N \leq c_N \check{k}_N$, we obtain that

$$\begin{aligned}\int_{A_\alpha(a,y)} x\phi(a+\check{k}_N,x,y')p_\alpha(x)dx &\geq C\check{k}_N T^{1/2}\int_{A_\alpha(a,y)} x^2\phi(a,x,y)p_\alpha(x)dx - c_N N^{-1}\epsilon_N, \quad (\text{D.139})\\ \int_{A_\alpha(a,y)} x\phi(a-\check{k}_N,x,y')p_\alpha(x)dx &\leq -C\check{k}_N T^{1/2}\int_{A_\alpha(a,y)} x^2\phi(a,x,y)p_\alpha(x)dx + c_N N^{-1}\epsilon_N, \quad (\text{D.140})\end{aligned}$$

and that, for all $\bar{a}$ satisfying $|\bar{a}-a| \lesssim k_N$,

$$\left|\int_{A_\alpha(a,y)} x(\phi(\bar{a},x,y')-\phi(a,x,y))p_\alpha(x)dx\right| \leq c_N\check{k}_N T^{1/2}\int_{A_\alpha(a,y)} x^2\phi(a,x,y)p_\alpha(x)dx + c_N N^{-1}\epsilon_N. \tag{D.141}$$

Using (D.136) and (D.141), we immediately achieve that, for all $\bar{a}$ satisfying $|\bar{a}-a| \lesssim k_N$,

$$\begin{aligned}&\left|\int x(\phi(\bar{a},x,y')-\phi(a,x,y))p_\alpha(x)dx\right|\\ \leq\ & c_N\int(\check{k}_N T^{1/2}x^2\mathbb{1}_{A_\alpha(a,y)} + |x|\mathbb{1}_{A'_\alpha(a,y)})\phi(a,x,y)p_\alpha(x)dx + c_N N^{-1}\epsilon_N.\end{aligned} \tag{D.142}$$

Finally, we note that by definition at least one of equalities $z^*(a,y) = -T^{-1/2}\check{k}_N^{-1}$ and $h(z^*(a,y);a,y) = T^{-1/2}\check{k}_N^{-1}$ must hold. We first consider the case where $z^*(a,y) = -T^{-1/2}\check{k}_N^{-1}$. The definition of ϕ leads to

$$\int_{-\infty}^{z^*(a,y)} |x|\phi(a+\check{k}_N,x,y')p_\alpha(x)dx \leq c_N \sup_{x:x\leq -T^{-1/2}\check{k}_N^{-1}} \phi(a+\check{k}_N,x,y') \leq c_N N^{-1}\epsilon_N. \tag{D.143}$$

On the other hand, because $h(z^*(a,y);a,y) \geq 0$ by definition, it follows from (D.127) that

$$\int_{h(z^*(a,y);a,y)}^{\infty} x\phi(a+\check{k}_N,x,y')p_\alpha(x)dx \geq C\int_{h(z^*(a,y);a,y)}^{\infty} |x|\phi(a,x,y)p_\alpha(x)dx - c_N N^{-1}\epsilon_N. \tag{D.144}$$

Combining (D.143) and (D.144), we obtain

$$\int_{A'_\alpha(a,y)} x\phi(a+\check{k}_N,x,y')p_\alpha(x)dx \geq C\int_{A'_\alpha(a,y)} |x|\phi(a,x,y)p_\alpha(x)dx - c_N N^{-1}\epsilon_N.$$

Symmetric reasoning leads to that, when $h(z^*(a,y);a,y) = T^{-1/2}\check{k}_N^{-1}$,

$$\int_{A'_\alpha(a,y)} x\phi(a-\check{k}_N,x,y')p_\alpha(x)dx \leq -C\int_{A'_\alpha(a,y)} |x|\phi(a,x,y)p_\alpha(x)dx + c_N N^{-1}\epsilon_N.$$

Combining these two inequalties with (D.139) and (D.140), we obtain

$$f_N(a,y') \geq C\int(\check{k}_N T^{1/2}x^2\mathbb{1}_{A_\alpha(a,y)} + |x|\mathbb{1}_{A'_\alpha(a,y)})\phi(a,x,y)p_\alpha(x)dx - c_N N^{-1}\epsilon_N. \tag{D.145}$$

Renaming variables, (D.145) already gives (D.128). Comparing (D.145) with (D.142), we prove

(D.129). ∎