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: February 13, 2022

Abstract

In the context of a linear asset pricing model, we document a statistical limit to arbitrage due to the fact that arbitrageurs are incapable of learning a large cross-section of alphas with sufficient precision given a limited time span of data. Consequently, the optimal Sharpe ratio of arbitrage portfolios developed under rational expectation in the classical arbitrage pricing theory (APT) is overly exaggerated, even as the sample size increases and the investment opportunity set expands. 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 more than 50 years of monthly US individual equity returns shows that all strategies we consider achieve a moderately low Sharpe ratio out of sample, in spite of a considerably higher yet infeasible one, suggesting the empirical relevance of the statistical limit of arbitrage and the empirical success of APT.

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.

†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

Most arbitrageurs in practice conduct some statistical analysis prior to the execution of investment strategies. Nonetheless, the statistical limit is absent from the theory of limits to arbitrage. This is partially because statistical challenges facing economic agents are often benign in the sense that the statistical uncertainty can be safely ignored, provided a sufficiently large sample size, to the extent that rational expectation retains its relevance, at least approximately. Nevertheless, in the case of arbitrage pricing theory, arbitrageurs are presented with an expansive set of investment opportunities. The scenario in which they can learn about the statistical properties of an increasing number of assets with infinite precision is at best extreme.

We are interested in a more realistic scenario, where the sample size increases with and is even outnumbered by the cross-section of investment opportunities.[1] Because of this high-dimensionality challenge, arbitrageurs in our setting are incapable of knowing the true alphas, even in certain limiting experiments. In fact, only when the sample size is larger than or comparable to the dimensionality of the investment opportunity sets can the assumption of rational expectation becomes relevant. In otherwise more realistic limiting experiments, the learning effect induces a limit to arbitrage.

To characterize the effect of learning, we assume that returns follow the same linear factor model as in APT and that arbitrageurs in this model are only allowed to employ a feasible trading strategy that relies on historical data to make inference on 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 in the rational expectation setting. This, in turn, provides a new no-near-feasible-arbitrage condition that accounts for the statistical limit of learning.

The difficulty of the learning problem hinges on the data generating process (DGP) of alpha signals. While our theory does not rely on specific cross-sectional distributions of alpha signals, we use a simple model to demonstrate how the optimal Sharpe ratio varies with the strength and sparsity of alphas. In accordance with our intuition, when alphas (signals) are strong and not too rare relative to the dimensionality and the sample size, the inference problem is simply reduced to the classical rational expectation result in the limit. Nevertheless, when alpha is weaker and more rare, its inference becomes more challenging, inducing a substantial gap between the optimal feasible Sharpe ratio and the classical infeasible limit.

Furthermore, we demonstrate how arbitrageurs can construct a feasible trading strategy that achieves the theoretically optimal Sharpe ratio, uniformly over DGPs of alphas, regardless of the strength and sparsity of alphas. A uniformly valid trading strategy is desirable because arbitrageurs do not know which DGP is a correct description of the observed data. This result also suggests that the aforementioned optimal Sharpe ratio bound is in fact sharp. The optimal strategy is designed

[1] In fact, the available sample size in most asset pricing settings is rather limited compared to the number of assets in the investment universe. Tens of years of monthly time series are often seen in the literature, and the sample size is at best in the order of hundreds. In contrast, there are thousands of stocks in US markets alone, not to mention opportunities in global markets or alternative asset classes.

by carefully examining the empirical distribution of alpha estimates and assigning weights based on their relative magnitudes and associated uncertainty. Significant alpha estimates are relatively straightforward to deal with, whose weights are proportional to their signal strength. The weaker alphas are more difficult to exploit, and simply ignoring them would lead to a suboptimal trading strategy.

We also propose an estimator of the infeasible Sharpe ratio, which can be interpreted as the perceived Sharpe ratio by rational expectation investors. While this Sharpe ratio can be estimated, it cannot be realized by any portfolio with weights constructed using historical data.

Next, we examine alternative strategies that exploit multiple testing, shrinkage, and selection techniques to build arbitrage portfolios. With alphas estimated from cross-sectional regressions, one strategy adopts a multiple-testing (BH) procedure as in Benjamini and Hochberg (1995) on the alpha t-statistics 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 the 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 against the less potent alphas. In contrast, the ridge approach is equivalent to the plain cross-sectional regression based alphas, which 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, almost 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 uses observed characteristics as betas. 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 9.01% over the sample period from 1965 to 2020. Our results are similar to existing estimates from 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) for latent factor models. The immediate implication of these numbers is that there exists a considerable amount of idiosyncratic noise in the cross-section of individual equities, to the extent that learning about alpha becomes an arduous statistical task.

Second, we plot the cross-sectional distribution of the t-statistics corresponding to alpha estimates of all individual stocks based on their full record in our sample. Among 12,415 test statistics we obtain, we find that only 4.69% (0.41%) of the t-statistics are greater than 2.0 (3.0) in absolute values. No t-statistics are greater than 4.80 in magnitude. Even without controlling for multiple testing, these estimates suggest that non-zero alphas are rather rare and weak.

Third, we find the optimal feasible arbitrage portfolio with different methods achieve a moderately low annualized Sharpe ratio, about 0.5, whereas the perceived Sharpe ratios over time are considerably higher – beyond 2.0 – for almost all sample periods. The perceived Sharpe ratio is what investors could estimate in light of the classical arbitrage pricing theory, but it is not realizable by constructing a feasible arbitrage portfolio. The large gap between realizable and perceived Sharpe ratios suggests the empirical relevance of the statistical limit of arbitrage. Existing literature on testing the APT focus on the perceived Sharpe ratio, which leads to more powerful rest statistics and rejections of the APT. In contrast, we argue that the fact that the feasible Sharpe ratio is small suggests the empirical success of APT.

Fourth, among all feasible strategies, the cross-sectional regression approach achieves the best performance, followed by the uniformly valid optimal strategy we develop. This is not surprising given that earlier empirical evidence suggests that alphas are rare and weak, so that the alpha generating process squares well with assumptions that make cross-sectional regression approach optimal. The BH approach performs the worst, because it is overly conservative that it eliminates almost all weak signals, which other strategies exploit to outperform.

Our paper is built on a large strand of literature dating back to the arbitrage pricing theory (APT) by Ross (1976), which was later refined by Huberman (1982), Chamberlain and Rothschild (1983), and Ingersoll (1984). An important contribution from these papers is a powerful foundation for asset pricing, which does not rely on assumptions on economic agents about their preferences and beliefs such as those behind the CAPM. One purpose of our paper is to revisit the APT. We show that the optimal Sharpe ratio discussed in the prior literature is overly exaggerated and their no near-arbitrage condition can further be relaxed due to the incapability of arbitrageurs in identifying investment opportunities with infinite precision. 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 with common factor covariances, even in a world in which belief distortions affect asset prices. Our study instead focuses on the approximation error by common factor covariances in their SDF, or to put it differently, the alpha component of asset returns in a reduced-form factor model. We derive a new Sharpe ratio bound on near-arbitrage opportunities and provide a feasible trading strategy that optimally exploits these opportunities without exposure to any factor risk. Another 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 alpha and beta portfolios without violation of the conventional near-arbitrage condition. Nonetheless, our setting is considerably different in which 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 arbitrage Sharpe ratio. The second literature focus 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 overwhelmingly conservative to the extent of no or few discoveries.[2] 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 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, see Hansen (2007), in which economic agents are not confronted with statistical uncertainty over structure parameters. In fact, as surveyed by Pastor and Veronesi (2009), learning effect in financial markets is ubiquitous. Befuddled by parameter uncertainty, rational economic agents can learn about unknown parameters by updating their beliefs according to Bayes' rule. In many existing work, e.g., Collin-Dufresne et al. (2016), learning is slow due to a limited sample size, hence its effect persists, though in the limit these systems under learning will converge to stationary rational expectation limits. An exception is Martin and Nagel (2021), in which learning effects persist because investors face a high-dimensional inference problem. Similarly, arbitrageurs in our model attempt to make inference on a high-dimensional parameter with a potentially insufficient sample size. Nonetheless, our setup is substantially different from existing learning framework from several perspectives. We examine different sequences of DGPs to precisely characterize the finite sample behavior of arbitrageurs' trading strategies.[3] In most scenarios, our learning system does

[2] 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.

[3] Our analysis is related to a large literature in econometrics and statistics that discuss uniform validity of asymptotic

not converge to a rational expectation limit. Moreover, our analysis does not rely on assumptions about investors' preferences or beliefs.

Our paper proceeds as follows. Section 2 develops a statistical limit of arbitrage. Section 3 constructs a feasible trading strategy that achieves the optimal Sharpe ratio. Section 5 analyzes two alternative trading strategies. Section 6 provides simulation evidence, followed by an empirical analysis in Section 7. Section **??** 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. While this model is a bit stylized, it is sufficiently sophisticated to deliver theoretical insight, and is sufficiently relevant to guide empirical investment decisions.

2.1 A Toy Model

To fix the idea, we consider a toy model in which an N-dimensional vector of returns is pure constant alpha plus noise, i.e., $\alpha + u_t$, where $u_{i,t} \sim \mathcal{N}(0, \sigma^2)$ for some scalar $\sigma > 0$. Suppose arbitrageurs do not observe the true data generating process and they will make inference on α before building the optimal portfolio. With a sample size T, their estimated alphas satisfy: $\widehat{\alpha} = \alpha + \bar{u}$, where $\bar{u} \sim \mathcal{N}(0, \sigma^2/T)$. Suppose further that arbitrageurs know σ, then their optimal portfolio weights are given by $\widehat{w} = \sigma^{-2}\widehat{\alpha}$.

Out of sample, this portfolio's expected return and variance can be written as:

$$\mathrm{E}\left(\sigma^{-2}\left(\alpha+\bar{u}\right)^{\intercal}\left(\alpha+u\right)\right) = \sigma^{-2}\alpha^{\intercal}\alpha, \quad \mathrm{Var}\left(\sigma^{-2}\left(\alpha+\bar{u}\right)^{\intercal}\left(\alpha+u\right)\right) = \sigma^{-2}(1+T^{-1})\alpha^{\intercal}\alpha + NT^{-1},$$

where u denotes a random variable independent of $\bar{u}$, but share the same distribution as u_t. The resulting squared Sharpe ratio is given by:

$$S^2 = \frac{(\sigma^{-2}\alpha^{\intercal}\alpha)^2}{\sigma^{-2}(1+T^{-1})\alpha^{\intercal}\alpha + NT^{-1}} < \sigma^{-2}\alpha^{\intercal}\alpha := (S^{\star})^2, \tag{1}$$

where the right-hand side is an infeasible squared Sharpe ratio which requires perfect knowledge of α. The gap between $(S^{\star})^2$ and S^2,

$$\frac{(S^{\star})^2}{S^2} - 1 = \frac{1}{T} + \frac{N}{T(S^{\star})^2}.$$

depending on the magnitude of the last term on the right-hand side, can be considerably large.

approximations, see, e.g., Staiger and Stock (1997), Imbens and Manski (2004), Leeb and Pötscher (2005), Andrews et al. (2020).

Now suppose that whenever they find such arbitrage trading dominates its implication cost, arbitrageurs implement this strategy. In equilibrium, the feasible Sharpe ratio should be equal to the implementation cost, $S^2 \leq C^\star$. We can solve (1) for $S^\star$ and obtain

$$(S^\star)^2 \leq \frac{C^\star}{2}\left\{1 + \frac{1}{T} + \sqrt{1 + \frac{1}{T} + \frac{4N}{C^\star T}}\right\}.$$

This inequality suggests that the presence of estimation error widens the bounds within which the true squared Sharpe ratio can "live" in equilibrium. In addition, even though T is large, the bounds can be arbitrarily wide, depending on the order of N/T.

This toy model sheds light on when the limit of arbitrage could possibly matter in equilibrium. Besides, the toy model also has implications on the econometric analysis of factor models. When APT is tested in the literature, very often the null hypothesis is that all alphas are zero. Such a null hypothesis is overly restrictive, and more importantly, does not directly translate to violations of the APT because APT only implies $\alpha^\intercal\alpha < \infty$.

We follow Shanken (1992)'s suggestion, but instead of using the back of the envelop calculation he conducted, to build the optimal arbitrage portfolio and evaluate its performance against our priors to gauge the relevance of APT. Nonetheless, the toy model relies on rather restrictive assumptions to the extent that we cannot rely on this model for empirical analysis. The next section builds on the same intuition, but derive methodologies and implications from a general and more realistic factor model.

2.2 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 $t = 1, 2, \ldots, T$:

$$r_t = \alpha + \beta\gamma + \beta v_t + u_t, \tag{2}$$

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.[4]

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 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 as-

[4]While it is possible to extend our model to the approximate factor setting, allowing for off-diagonal entries in the covariance matrix Σ_u will not provide additional insight with respect to the limit of arbitrage we focus on in this paper.

sumptions are essential for identification of γ in a model that allows for pricing errors.[5] Assumption A1 in the appendix summarizes these conditions in details.

There are at least three variations of the factor model (2), 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).[6] 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 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.[7]

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

2.3 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 (2), 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 subsequence $m \to \infty$,[8] with high probability,*

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

Intuitively, no near-arbitrage means there exist no sequence of portfolios that earn positive expected returns with vanishing risks. Under fairly general assumptions, Ingersoll (1984) established that a sufficient and necessary condition for the absence of near-arbitrage is that with high proba-

[5]See, e.g., Assumption I.1 of Giglio and Xiu (2021).

[6]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.

[7]Strictly speaking, the MSCI Barra model is cast in a conditional version of (2):

$$r_t = \alpha_{t-1} + \beta_{t-1}\gamma_{t-1} + \beta_{t-1}v_t + u_t, \tag{3}$$

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).

[8]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 m and t.

bility, for some constant $C > 0$,[9]

$$S^\star = \sqrt{\alpha^\intercal \Sigma_u^{-1} \alpha} \leq C. \tag{4}$$

The left-hand side 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.

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).[10] 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 them from a finite sample of data of size T. 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 simpler as the sample size increases.

In the current context, the difficulty of the learning problem is also affected by 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 difficulty of learning does not diminish as N and T increase, the learning effect will persist, leading to different limiting implications. It turns out that the rational expectation limit $S^\star$ is only relevant for rather restrictive scenarios, e.g., T is comparable to N, in which we will show a feasible strategy exists to achieve $S^\star$ approximately. 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 (4) 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:*

$$\alpha_i \overset{i.i.d.}{\sim} \begin{cases} \mu & \text{with prob. } \rho/2 \\ -\mu & \text{with prob. } \rho/2 \\ 0 & \text{with prob. } 1-\rho \end{cases}, \quad 1 \leq i \leq N, \tag{5}$$

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

[9] 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.

[10] 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 $w^\star = \mathbb{M}_\beta \Sigma_u^{-1} \alpha$.

$\Sigma_u = \sigma^2 \mathbb{I}_N$, *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.[11] 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

$$\mu \sim T^{-1/2} \quad \text{and} \quad \rho \sim N^{-1/2}. \tag{6}$$

This condition (6) 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 (4) 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 $N^{1/2}/T \to \infty$. Therefore, a near-arbitrage opportunity arises according to (4).

However, the statistical obstacle prevents arbitrageurs from having this "free lunch." Under fairly general assumptions, it is only possible to recover any element of alpha up to some estimation error of magnitude $T^{-1/2}$.[12] Since the true alpha is of the same order of magnitude as its level of statistical uncertainty $T^{-1/2}$, it is impossible for arbitrageurs to determine precisely which assets among all have nonzero alpha. Indeed, as we will show later, in this model 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:

[11] 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."

[12] 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 $1 \le i \le N$,

$$\sqrt{T}(\widehat{\alpha}_i - \alpha_i) \xrightarrow{d} \mathcal{N}(0, \sigma_i^2(1 + \gamma^\intercal (\Sigma_v)^{-1}\gamma)), \tag{7}$$

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.

Definition 2. *A portfolio strategy $\widehat{w}$ is said to be feasible at time t, if it is a deterministic 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. $S^{\star}$ is optimal and achievable only when T is comparable to N, in which case learning is not too difficult for arbitrageurs, to the extent that the learning effect diminishes in the limit and the rational expectation limit becomes relevant. In practice, T is typically far smaller than N, the optimal arbitrage Sharpe ratio achievable is thus much smaller, compared to $S^{\star}$.

2.4 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 (2) and Assumption A1 in the appendix holds. For any feasible portfolio weight $\widehat{w}$, its Sharpe ratio, $S(\widehat{w})$, satisfies:*

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

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 originated from trading 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 (4), 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:

[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.

Corollary 1. *Suppose the same assumptions as in Theorem 1 hold. For any given return-generating process satisfying (2), there exists no feasible strategy $\widehat{w}$ that leads to a near-arbitrage, if*

$$\mathrm{P}\big(S(\mathcal{G}) \leq C\big) \to 1, \quad \textit{as} \quad N, T \to \infty, \tag{9}$$

for some constant C.

The form of $S(\mathcal{G})$ in Theorem 1 appears to suggest that arbitrageurs rely on the information set $\mathcal{G}$ that embodies 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, this upper bound 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 3 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.

We end the current discussion with a more explicit expression for $S(\mathcal{G})$:

Proposition 1. *Suppose that r_t follows (2) and Assumptions A1 and A2 hold. Then it holds that*

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

where

$$\psi(a,\sigma_i,T) = \frac{\mathrm{E}\left(\alpha_i\phi(a - T^{1/2}\alpha_i/\sigma_i)\middle|\sigma_i\right)}{\mathrm{E}\left(\phi(a - T^{1/2}\alpha_i/\sigma_i)\middle|\sigma_i\right)}, \quad p(a,\sigma_i,T) = \mathrm{E}\left(\phi(a - T^{1/2}\alpha_i/\sigma_i)\middle|\sigma_i\right),$$

$\phi(\cdot)$ is the normal pdf function, $\mathrm{E}(\cdot)$ is the expectation taken with respect to the cross-sectional distributions of α and σ, and σ_i^2 is the ith entry of Σ_u.

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 r_t follows (2) and Assumption A1 holds. In addition, we assume alpha follows (5) as in Example 1. Then we have $S^\star = \mathrm{E}(S^\star) + o_{\mathrm{P}}(1)$. Further, assuming that $\mathrm{E}(S^\star)$ does not vanish, then it holds that $S^{\mathrm{OPT}} \leq (1-\epsilon)\mathrm{E}(S^\star)$ for some $\epsilon > 0$, if and only if*

$$T^{1/2}\mu/\sigma - \sqrt{-2\log\rho} \leq C, \tag{10}$$

for some constant C.

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 that our arbitrageurs 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 rare and in the mean time not sufficiently strong, the learning problem becomes rather challenging and hence S^{OPT} would be dominated by $\mathrm{E}(S^{\star})$ in the limit, resulting in a distinct optimal Sharpe ratio as opposed to the classical case.

3 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, we expect that the optimal strategy depends on the unobserved DGP as well, creating a challenge for arbitrageurs who cannot observe the true DGP.

It turns out that arbitrageurs can construct a uniformly valid estimator of the optimal portfolio weights, which achieves S^{OPT} over a large class of return generating precesses. 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 do not assume any knowledge of unknown variables, such as Σ_u. Instead we estimate them when necessary.

Algorithm 1 (Constructing the Optimal Arbitrage Portfolio)**.**

S1. We split the observed sample $\mathcal{T} = \{t-T+1,\dots,t\}$ into:

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

and we construct cross-sectional regression estimates of alpha $\breve{\alpha}$ and $\breve{\alpha}'$, volatility estimates $\breve{\sigma}$, and the t-statistics $\breve{z}_i = |S|^{1/2}\breve{\alpha}_i/\breve{\sigma}_i$ for each $i = 1, 2, \dots, N$, using subsamples S and S':[14]

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

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

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

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

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

$$\widehat{f}(l,m) = \frac{1}{m^2 k_N^3} \frac{1}{|B(l,m)|} \sum_{i \in B(l,m)} \widehat{\alpha}_i',$$

where

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

As we have discussed in footnote 10, the optimal strategy in the case that arbitrageurs know the true DGP is given by

$$w^\star = \mathbb{M}_\beta \Sigma_u^{-1} \alpha. \tag{11}$$

Part of the above construction, $\Sigma_u^{-1}\alpha$, is the optimal allocation to the ex-factor returns, $\alpha + u_t = r_t - \beta(\gamma + v_t)$, in the conventional setting. Multiplying by $\mathbb{M}_\beta$ in (13) simply eliminates factor exposures in r_t. Effectively, Step S1 of Algorithm 1 provides estimates of $\breve{\alpha}_i$ and $\breve{\sigma}_i^2$. Step S2 first constructs a nonparametric estimate of the function $f(\alpha, \sigma^2) = \mathrm{E}(\alpha/\sigma^2 | \{r_s, \beta\}_{s \in \mathcal{T}})$, with which the optimal weights on de-factor returns are estimated by $\breve{w}$. This, in turn, leads to the optimal weight estimates, $\widehat{w}^{\mathrm{OPT}}$, on original asset returns. An essential step towards uniform inference 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 alphas, using which we obtain their optimal portfolio weights. This strategy outperforms the alternatives that directly use estimated alphas as if these estimates are not susceptible to errors, or simply ignore the contribution from the 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 (2), and Assumptions A1 and A3 hold. 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 $\mathrm{P} \in \mathbb{P}$. That is, for any $\epsilon > 0$,*

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

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 optimal Sharpe ratio of any 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 can establish the necessity for the no near-arbitrage condition given by (9).

Corollary 3. *Suppose the assumptions 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 $C^\star$ in an economy with statistical limit of arbitrage. In equilibrium, $S^{\mathrm{OPT}} = C^\star$, 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 $C^\star$.

4 Estimating Optimal Infeasible Sharpe Ratio

We are also interested in estimating the maximal investment opportunities in the data generating process, as measured by $S^\star$. Existing literature on testing APT often construct a test statistics in the spirit of Gibbons et al. (1989), based on the infeasible optimal Sharpe ratio of arbitrage portfolios, $S^\star$, see, e.g., Pesaran and Yamagata (2017) and Fan et al. (2015).

While such tests are powerful, they do not account for the statistical limit facing arbitrageurs, who are incapable of constructing a feasible portfolio to realize this Sharpe ratio.

To illustrate the gap between $S^\star$ and S^{OPT}, we now construct an estimator for $(S^\star)^2$.

$$(\widehat{S}^\star)^2 = \sum_{i \le N} \widehat{\sigma}_i^{-2} T^{-2} \sum_{t=1}^{T} \sum_{1 \le t' \le T : t' \ne t} r_{i,t} \mathbb{M}_\beta r_{i,t'}. \tag{12}$$

The estimator takes the form of the sum over individual squared Sharpe ratios, but it is adjusted to remove squared returns, which would bias the estimates for certain data generating processes. The next proposition establishes its validity.

Proposition 2. *Suppose r_t follows (2), and Assumptions A1 and A3 hold. Then we have*

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

As shown by this proposition, the estimation error is relative when $S^\star$ is large and dominates 1 asymptotically, and absolute if $S^\star$ is small and dominated by 1. This is necessary as we simultaneously consider a large class of DGPs, some of which have an exploding or a shrinking $S^\star$, which also plays a role in the convergence rate.

5 Alternative Strategies for Arbitrage Portfolios

Algorithm 1 suggests a relatively complicated procedure that distinguishes weaker and strong signals using t-statistics before consolidating information about weaker signals to build the optimal portfolio. 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 prevalent in practice. The contrast among these strategies helps illustrate the necessity of consolidating information about weaker signals altogether in an optimal portfolio strategy.

5.1 Cross-Sectional Regression

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

$$\widehat{\alpha} = \left(\widehat{\beta}^{\intercal}\widehat{\beta}\right)^{-1}\widehat{\beta}^{\intercal}\bar{r}, \quad \bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t.$$

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

$$\widehat{w}^{\mathrm{CSR}} = \mathbb{M}_{\beta}\widehat{\Sigma}_u^{-1}\widehat{\alpha}. \tag{13}$$

This choice of portfolio weight is the sample analog of the optimal weight given by (11). Proposition 3 describes the asymptotic behavior of the expected Sharpe ratio of this arbitrage portfolio in the setting of Example 1. To enhance its finite sample performance and simplifies the proof, we assume that $\widehat{\Sigma}_u = \widehat{\sigma}^2\mathbb{I}_N$, since in this example, all assets share the same volatility, which can be estimated altogether. This further simplifies the analysis because a scaling factor does not play any role in the optimal Sharpe ratio.

Proposition 3. *Suppose that r_t follows (2) and Assumption A1 holds. In addition, we assume alpha follows (5) 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

$$C \leq T\mu^2\sigma^{-2} \leq C'\rho^{-1}. \tag{14}$$

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 very strong ($C \leq T\mu^2\sigma^{-2}$) and in the mean time 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 fake signals (zero alphas) from the true ones. This strategy 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 controls false discoveries among selected strong signals using the B-H procedure.

5.2 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: $\mathbb{H}_0^i : \alpha_i = 0$, hence detecting for alpha 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. 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 $\{p_i : i = 1, \ldots, N\}$ 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 $p_{(1)} \leq \cdots \leq p_{(N)}$.

S2. For $i = 1, \ldots, N$, 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

$$\widehat{\alpha}_i^{\mathrm{BH}}(\tau) = \widehat{\alpha}_i \mathbb{1}_{\{p_i \leq p_{(\widehat{k})}\}}. \tag{15}$$

The B-H procedure guarantees (in expectation) the selection of a group of assets among which at least a fraction of $(1 - \tau)$ 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 (13), the optimal portfolio weights are thus given by:

$$\widehat{w}^{\mathrm{BH}} = \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \widehat{\alpha}^{\mathrm{BH}}(\tau). \tag{16}$$

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 (which is true in this example) with the same idiosyncratic volatility, so that they adopted the following estimates for alpha instead of (15):

$$\overline{\alpha}_i^{\mathrm{BH}}(\tau) = sgn(\widehat{\alpha}_i)\overline{\alpha}\mathbb{1}_{\{p_i \leq p_{(\widehat{k})}\}}, \quad \overline{\alpha} = \sum_i |\widehat{\alpha}_i|\mathbb{1}_{\{p_i \leq p_{(\widehat{k})}\}} \Big/ \sum_i \mathbb{1}_{\{p_i \leq p_{(\widehat{k})}\}}, \tag{17}$$

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

Proposition 4. *Suppose that r_t follows (2) and Assumption A1 holds. In addition, we assume alpha follows* (5) *as in Example 1. The Sharpe ratio of the arbitrage portfolio with weights given by $\widehat{w}^{\mathrm{BH}} = \widehat{\sigma}^{-2}\mathbb{M}_\beta\widehat{\alpha}_i^{\mathrm{BH}}(\tau)$ and $\overline{w}^{\mathrm{BH}} = \widehat{\sigma}^{-2}\mathbb{M}_\beta\overline{\alpha}_i^{\mathrm{BH}}(\tau)$ 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*[15]

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

and

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

Here $\Phi(\cdot)$ is the normal cdf, and z^* is the positive solution of the equation

$$2(1-\tau(1-\rho))\Phi(-z) = \tau\rho\Phi(T^{1/2}\mu/\sigma - z). \tag{18}$$

Suppose further that S^{OPT} does not vanish, and that $CN^{-1+\lambda} \leq \rho \leq CN^{-\lambda}$ for some fixed $\lambda > 0$. 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 $\epsilon' > 0$,

$$T^{1/2}\mu/\sigma \leq (1-\epsilon')\sqrt{-\lambda_\tau \log\rho}, \tag{19}$$

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

As Proposition 4 shows, (19) 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 while 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 treats all signals identically without distinguishing their strength, whereas the latter only focuses on the stronger signals. Neither of the two always achieves optimality.

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

5.3 Shrinkage Approaches

The analysis above suggests that we can construct the optimal portfolio out of the de-factored returns directly, while imposing priors to regularize the portfolio weights. This amounts to imposing such priors on the alpha estimates. To see this, suppose we adopt a shrinkage approach:

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

where $p_\lambda(w) = \lambda \left\|w\right\|_1$ or $\lambda \left\|w\right\|_2^2$, for some $\lambda > 0$. Since $\widehat{\Sigma}_u$ is diagonal, the closed-form optimal portfolio weight is thereby given by

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

where $q = 1$ corresponds to the Lasso penalty and $q = 2$ the ridge, and for $i = 1, 2, \ldots, N$,

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

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. The "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 (2) and Assumption A1 holds. In addition, we assume alpha follows* (5) *as in Example 1. The Sharpe ratio of the arbitrage portfolio with weights given by $\widehat{w}^q$ satisfies $\widehat{S}^1 - S^{\mathrm{LASSO}} = o_{\mathrm{P}}(1)$ and $\widehat{S}^2 - S^{\mathrm{CSR}} = o_{\mathrm{P}}(1)$, where*

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

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

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

$$T^{1/2}\mu/\sigma \geq C, \quad \textit{and} \quad \frac{T\mu^2/\sigma^2 + 2\log\rho}{\sqrt{-\log\rho}} \leq C.$$

6 Simulation Evidence

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

6.1 Numerical Illustration of Theoretical Predictions

We start by examining theoretical predictions. For simplicity and clarity, we simulate a one-factor (CAPM) model of returns given by (2). We choose the factor risk premiua 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 (5) 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 (5), 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 conner, 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.

We now turn to the comparison of Sharpe ratios of optimal feasible arbitrage portfolios with those

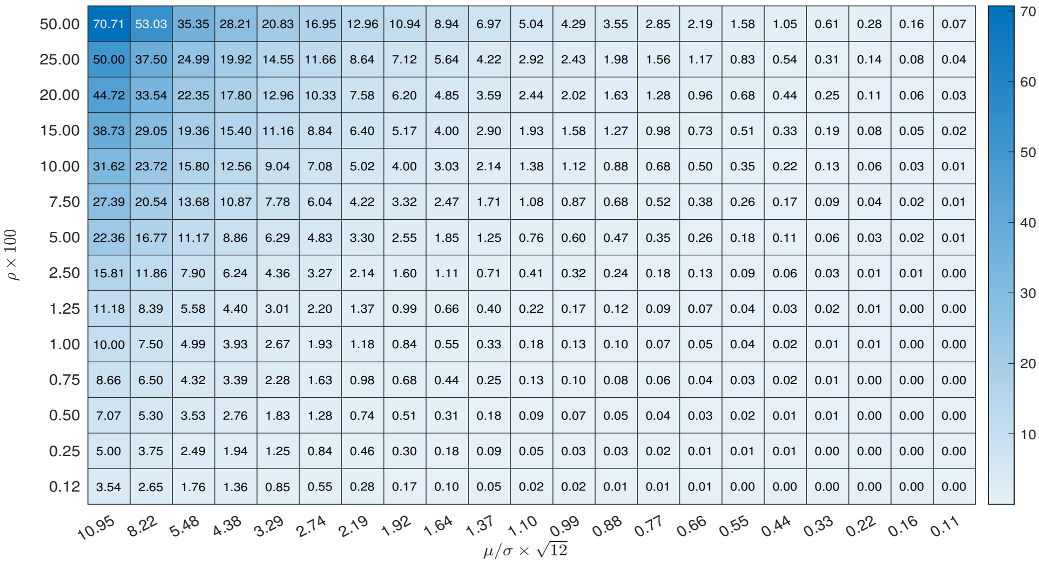

Figure 1: Optimal Sharpe Ratios (S^{OPT}) of Feasible Arbitrage Portfolios

Note: The figure reports optimal Sharpe ratios of feasible arbitrage portfolios in model (5), in which a $100 \times \rho\%$ of assets have alphas that correspond to an annualized Sharpe ratio $\mu/\sigma \times \sqrt{12}$.

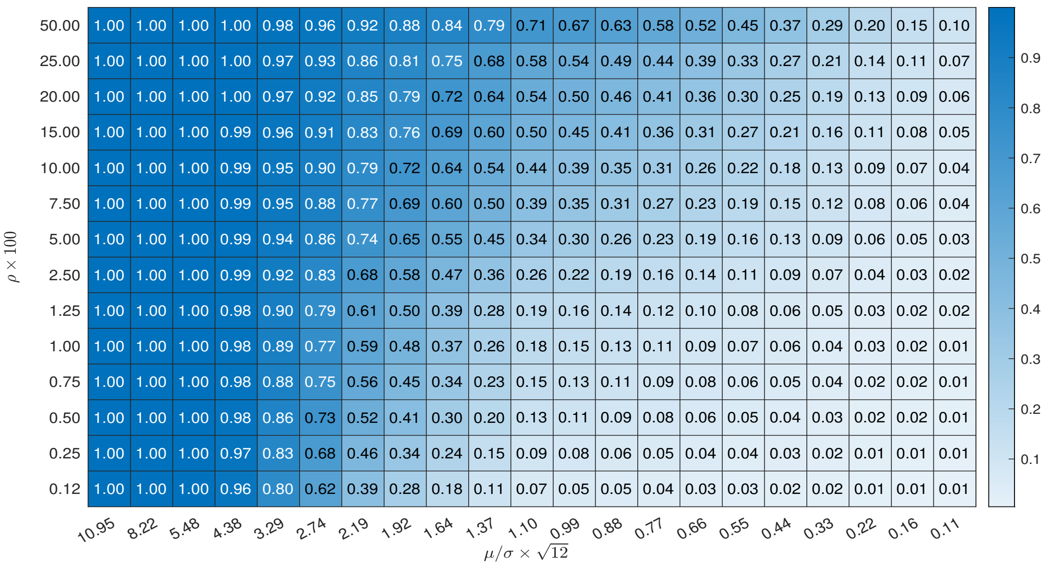

Figure 2: Ratios between S^{OPT} and $S^{\star}$

Note: The figure reports the ratios of optimal Sharpe ratios between feasible and infeasible arbitrage portfolios. The simulation setting is based on model (5), in which a $100 \times \rho\%$ of assets have alphas that correspond to an annualized Sharpe ratio $\mu/\sigma \times \sqrt{12}$.

achieved by alternative strategies. Figure 3 compares with the cross-section regression approach in Section 5.1, Figure 4 with the B-H based procedure given by Section 5.2, and Figure 5 with LASSO given by Section 5.3, respectively.

According to Proposition 3, the optimal portfolio dominates the cross-sectional regression based portfolio if (14) 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.

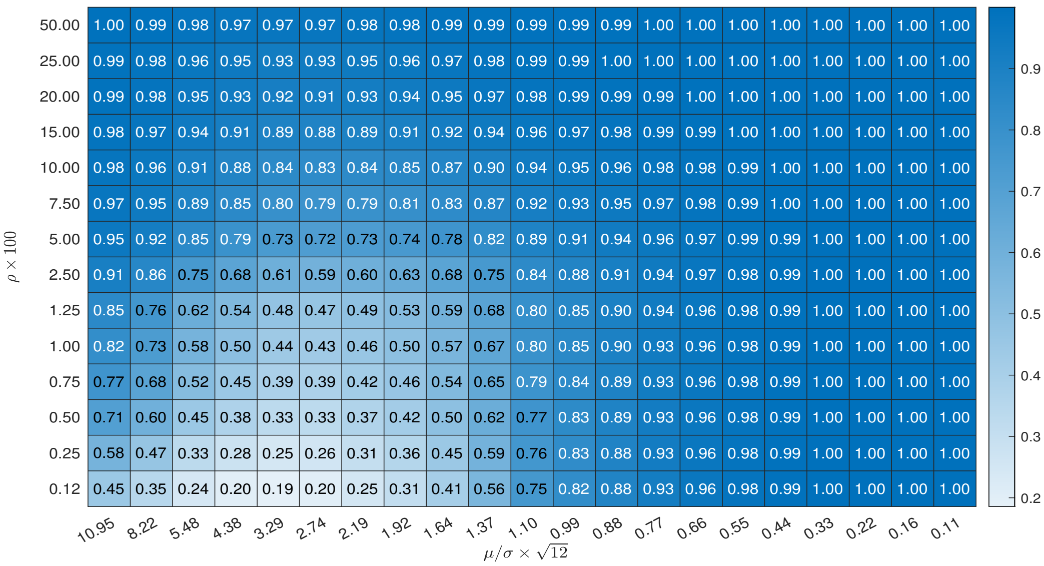

Figure 3: Ratios between S^{CSR} and S^{OPT}

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 (5), 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 (19) 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.

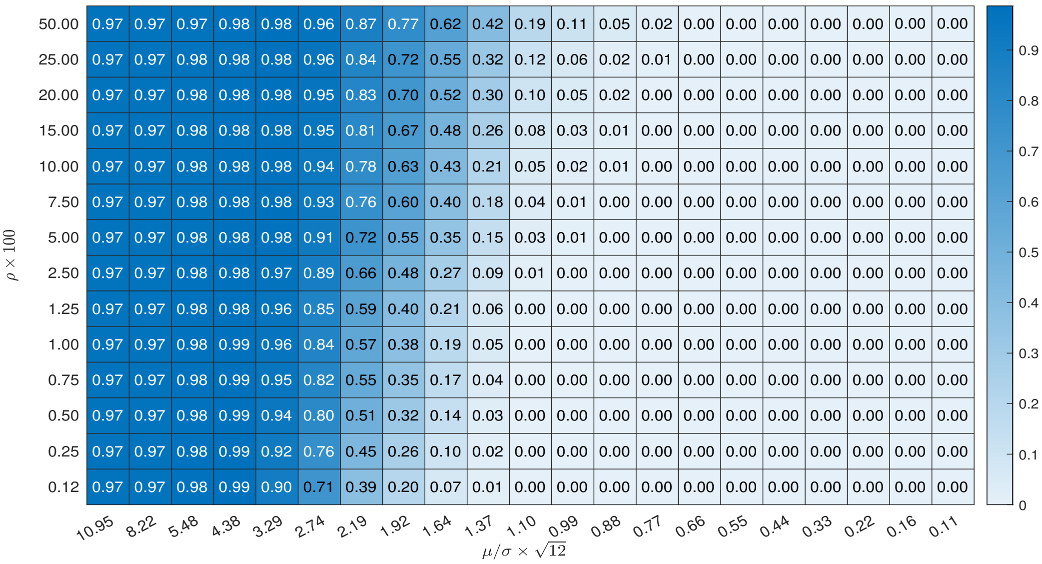

Figure 4: Ratios between S^{BH} and S^{OPT}

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 (5), 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 a ridge, because shrinking these signals does not change the fact that they are almost indistinguishable from noise.

6.2 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: $\widehat{w}^{\mathrm{OPT}\intercal}\mu/\sqrt{\widehat{w}^{\mathrm{OPT}\intercal}\Sigma_u^{-1}\widehat{w}^{\mathrm{OPT}}}$. We then calculate the average Sharpe ratio over all Monte Carlo repetitions. Our approach re-

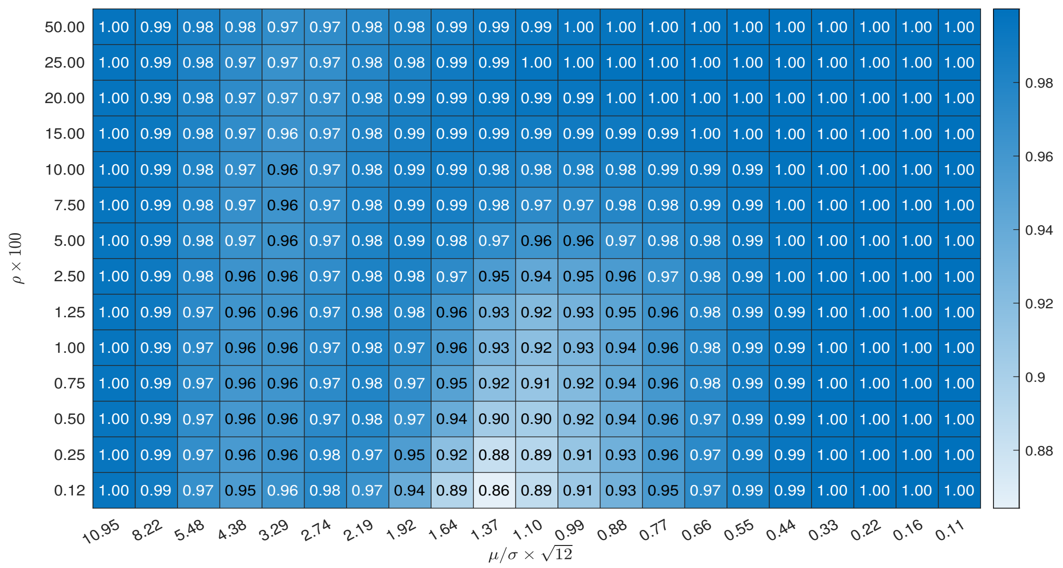

Figure 5: Ratios between S^{LASSO} and S^{OPT}

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 (5), 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}.

quires 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 6.1. 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		
	$T = 10$	$T = 20$	$T = 40$	$T = 10$	$T = 20$	$T = 40$	$T = 10$	$T = 20$	$T = 40$
	0.385	0.332	0.289	0.442	0.367	0.320	0.449	0.440	0.408
OPT	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}.

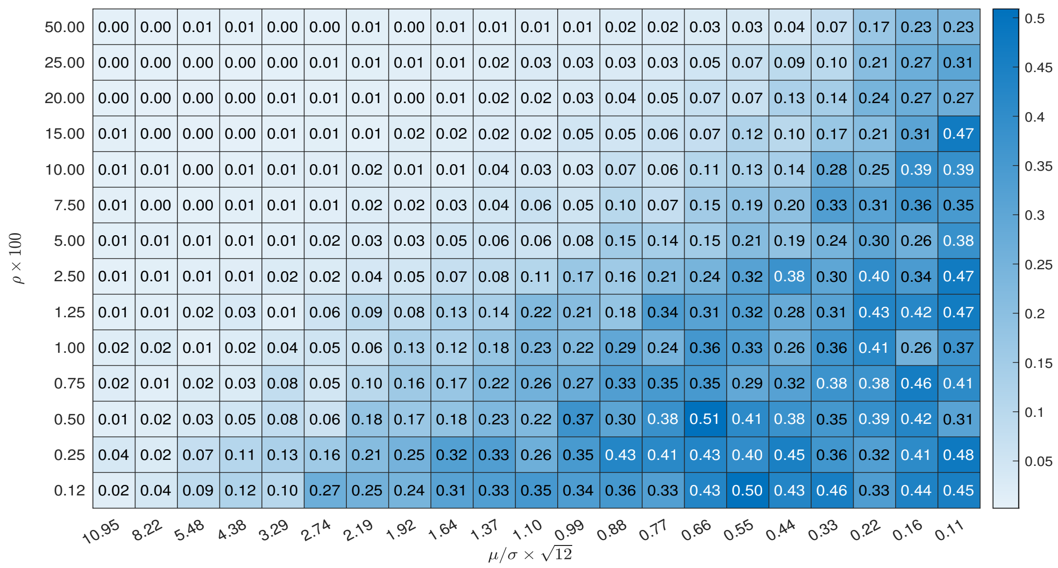

Figure 6: Ratios between $(\widehat{S}^{\star})^2$ and S^{OPT}

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 (5), in which a $100 \times \rho\%$ of assets have αs that correspond to an annualized Sharpe ratio $\mu/\sigma \times \sqrt{12}$.

7 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. We apply the usual filters (share codes 10 and 11 and exchange cod 1, 2, and 3) to the universe of stock returns downloaded from CRSP. The average number of stocks per month is 4,720.

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 only consider stock-month pairs without missing industry sectors. The average number of stocks per month that meet this criterion is reduced to 4,073. The missing of sector information mainly occurs prior to 1990. With information on industry sectors, we adopt a two-step procedure to fill in 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 there are no characteristic data for the entire sector in certain month, we fill them with this characteristic's cross-sectional median in this month.

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.

7.1 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 only for common stocks in these two cross-sections. Following Gu et al. (2020), the 16 characteristics are rank-normalized within each cross-section, which eliminate outliers in characteristics. The cross-section of returns are also winsorized at 99.5% and 0.5% quantiles, but only in cross-sectional regressions for the reason of robustness. The regression residuals for these winsorized returns are re-calculated using unwinsorized returns and regression coefficients.

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 9.03%. 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.

7.2 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 to have at least 60 observations. This ensures enough sample size for inference on alpha, although our results are not sensitive to this choice. 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.

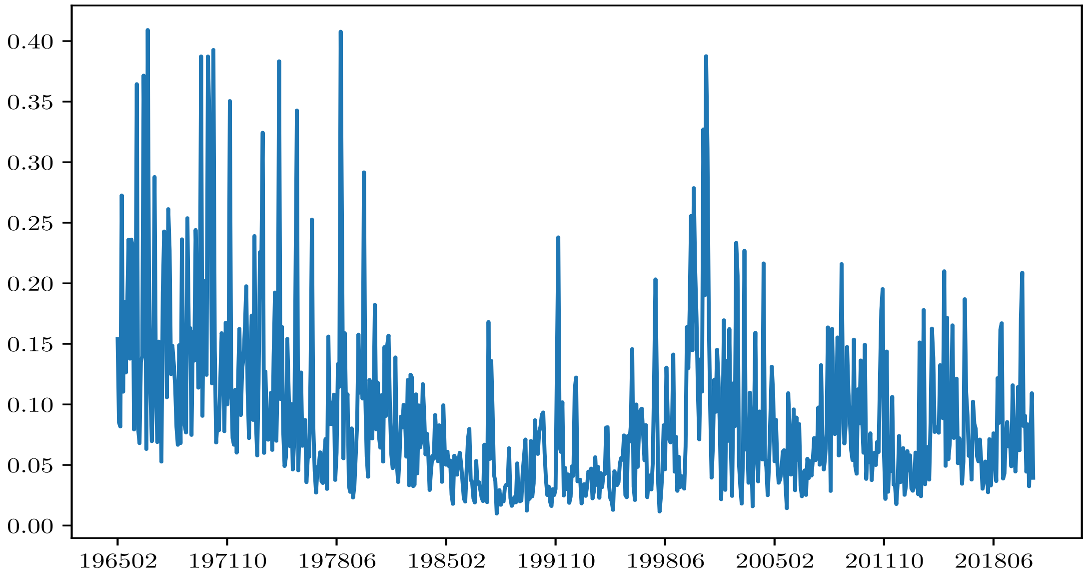

Figure 7: Time-series of the Cross-sectional R^2s

Only 4.69% of the t-statistics exceed 2.0 in magnitude, and more than 0.41% 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 1.60. Only 0.36% 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.

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 this 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.

7.3 Performance of Arbitrage Portfolios

Finally, we 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 calculate optimal portfolio weights using these strategies. The Sharpe ratios of different strategies are not influenced by risk version, though the cumulative returns are. So when compare cumulative returns, we normalize all strategies to have the same (ex-post) volatility for comparison purpose. We have also provided a time-series plot in Figure xxx of the perceived Sharpe ratio estimated using 12.

We observe a few important findings. First and foremost, there is clear gap between the perceived Sharpe ratio and the realizable Sharpe ratios by all strategies. The former is averaged around 2.0,

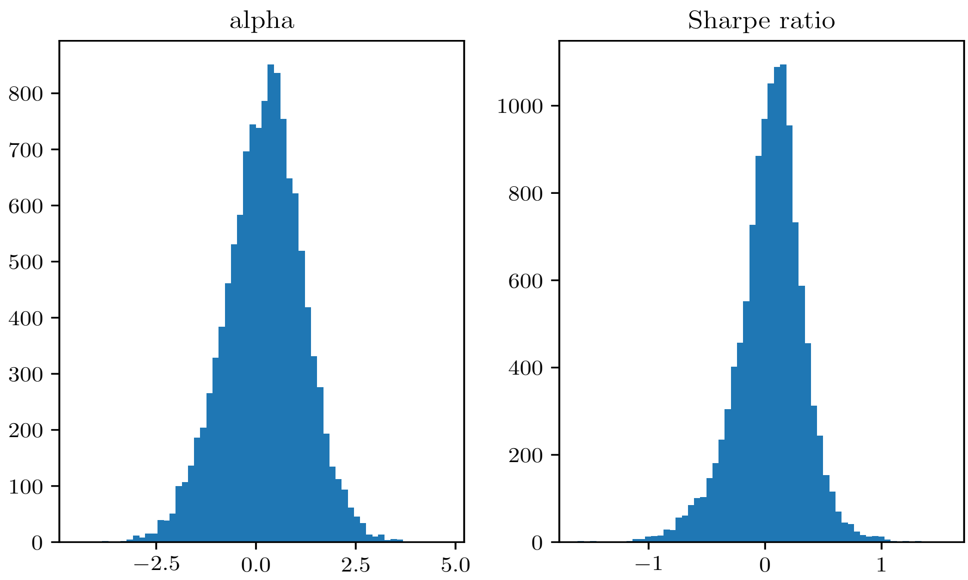

Figure 8: Histograms of the t-Statistics and Sharpe Ratios of Estimated Alphas

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 12,415.

and can sometimes exceed 5.0 or 6.0. The optimal strategy and CSR appear to dominate the BH and Lasso. The reason that BH underperform is due to its conservativeness, whereas for Lasso, selecting the optimal tuning parameter is difficult, which undermines its performance.

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. 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*.

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.

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.

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). Portfolio choice with model misspecification: A foundation for alpha and beta portfolios. Technical report, EDHEC Business School and Imperial College London.

Appendix A Mathematical Proofs

We start with assumptions needed for results in the main text.

Assumption A1. *For each $N \geq 1$, the following conditions hold:*

(a) The pricing errors α, factors v_t, factor loadings β, and idiosyncratic shocks u_t are mutually independent. $\mathrm{E}(\alpha) = 0$, $\mathrm{E}(v_t) = 0$, *and* $\mathrm{E}(u_t) = 0$. *γ and β satisfy $\|\gamma\| \leq C$, $\|\beta\|_{\mathrm{MAX}} \leq C$, and $\lambda_{\min}(\beta^\intercal\beta) \geq CN$ almost surely. v_t is i.i.d. across t and its covariance matrix Σ^v satisfies $C^{-1} \leq \lambda_{\min}(\Sigma^v) \leq \lambda_{\max}(\Sigma^v) \leq C$.*[16]

(b) α_i is i.i.d. across i, with probability density function (pdf) $p_\alpha(x)$; $|\alpha_i| \leq CN^\lambda$ for some fixed constant $\lambda < 0$ almost surely.

(c) u has the representation $u_{i,t} = \sigma_i\varepsilon_{i,t}$. σ_i is i.i.d. across i with pdf $p_\sigma(x)$; the support of $p_\sigma(x)$ is $(\underline{\sigma}, \bar{\sigma})$ satisfying $C^{-1} \leq \underline{\sigma} \leq \bar{\sigma} \leq C$. $\varepsilon_{i,t}$ is independent of σ_i, is i.i.d. across i and t, and has zero mean and unit variance.

Assumption A1 (a) is commonly seen in the literature of factor models. In particular, the assumption on $\lambda_{\min}(\beta^\intercal\beta)$ requires that all factors are pervasive. (b) and (c) suggest that the signals in our model are weak, in that as N increases their magnitudes shrink towards 0, whereas volatilities are bounded.

Assumption A2. *For each $N \geq 1$ and all (i, t), $\varepsilon_{i,t}$ is normally distributed.*

Assumption A2 imposes the normality assumption, which is only needed for a more explicit description of the arbitrage limit.

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

Assumption A3. *For each $N \geq 1$, it holds that: $C^{-1}N^\lambda \leq T \leq CN^{\lambda'}$ for fixed constants $\lambda > 1/3$ and $\lambda' < 1$; $C^{-1} \leq p_\sigma(x) \leq C$ for all $x \in (\underline{\sigma}, \bar{\sigma})$; $\varepsilon_{i,t}$ has a finite twelfth moment.*

[16] We use $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ to denote the minimum and maximum eigenvalues of A.