Review of "The Statistical Limit of Arbitrage"

An Xie Sida Li Yichen Ji
{anx, listar2000, jiyichen}@uchicago.edu

1 Introduction

1.1 Background in Financial Economics

In the evolving landscape of financial economics, understanding the dynamics and limitations of arbitrage has been a central pursuit, providing insights into the fundamental principles that govern financial markets ([9], [10], [6], [4]). [1] navigates the intriguing boundary where financial econometrics meets the practical challenges faced by arbitrageurs in real markets. It delves into the critical premise underlying most asset pricing theories, notably the Arbitrage Pricing Theory (APT), which posits that investment opportunities offering disproportionately high rewards relative to risk – essentially, near-arbitrage opportunities –should not persist in an efficient market. The rationale is straightforward: such opportunities, if existed, would be rapidly exploited by arbitrageurs, leading to their elimination. However, a significant departure from this theoretical ideal is the acknowledgment of the arbitrageurs' real-world predicament – a lack of precise knowledge of the data-generating process (DGP) for returns, compelling them to rely on a statistical analysis of historical data to identify potential arbitrage opportunities. This reliance introduces a layer of statistical uncertainty, particularly pronounced in the noisy and high-dimensional world of stock returns, which establishes a "statistical limit" to arbitrage.

From a statistical perspective, [1] scrutinizes a classic topic within asset pricing, arbitrage pricing theory (APT), and introduces a statistical arbitrage strategy inspired by the non-parametric empirical Bayes framework. This strategy optimizes portfolio weights in proportion to the posterior mean of pricing errors (alpha), aiming for optimal economic performance – a method that resonates with the classical normal mean problem discussed in [8]. The introduction outlines the transition from the theoretical construct, where arbitrage opportunities are swiftly leveraged by arbitrageurs endowed with perfect knowledge, to the empirical reality, marked by their struggle against the backdrop of statistical uncertainty. This conceptual shift not only lays the groundwork for reevaluating the bounds within which alphas can exist in equilibrium – broadening them compared to the traditional APT assumptions – but also heralds a nuanced approach that integrates empirical Bayes methods to refine these estimates, thereby offering a novel perspective on the statistical constraints that shape arbitrage strategies in financial markets.

1.2 Linear Factor Model Setup

[1] considers a setting in which returns follow a general linear factor model (we can think of it as the hypothetical return DGP)

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

where N is the number of assets, K is the number of risk factors, T is the sample size e.g. in months, r_t denotes the $N \times 1$ vector of excess returns, β is an $N \times K$ matrix of factor exposures, α is an $N \times 1$ vector of pricing errors (the component of returns not explained by the risk factors, which are our target of interest), v_t is a $K \times 1$ vector of factor innovations (unexpected changes in the risk factors) with covariance matrix Σ_v, γ is a $K \times 1$ vector of risk premia (the returns expected from bearing each unit of risk associated with the factors), and u_t is a vector of idiosyncratic returns (the asset-specific risks not captured by the common factors), independent of v_t with a diagonal covariance matrix Σ_u.

More specifically, the alphas represent potential arbitrage opportunities, as they signify returns in excess of what is predicted by exposure to the common risk factors. In an efficient market, we would expect these alphas to be zero, as any predictable, systematic returns above the risk premia would be quickly exploited by arbitrageurs, driving prices to adjust and eliminate the discrepancy. However, the presence of alpha signals, even if small and sparse, indicates the existence of statistical arbitrage opportunities.

1.3 Feasible Near-Arbitrage Opportunities

To better distinguish between (economic) theory and (statistical) reality, we first briefly introduce the concept of rational expectations hypothesis in economics [7]. The idea of rational expectations refers to the assumption that economic agents, including investors and arbitrageurs, use all available information efficiently and correctly to forecast future events. In the context of financial markets, this means that arbitrageurs would have perfect knowledge of the DGP for asset returns, enabling them to accurately predict α and exploit any arbitrage opportunities that arise. The rational expectations framework assumes that markets are efficient, with prices reflecting all available information, and thus, no systematic and exploitable arbitrage opportunities should exist because they would be immediately corrected by the market participants' actions.

However, the scenario outlined in [1] diverges from what rational expectations would imply, particularly regarding the asymptotic limit of alpha. In this paper's setting, arbitrageurs do not possess perfect knowledge of the return-generating process. Instead, they must rely on historical data to learn about potential alphas, which introduces statistical uncertainty into the process. This uncertainty arises from the challenge of estimating the true alpha values in a noisy and high-dimensional environment, where the number of assets (N) and the sample size (T) affect the precision of these estimates.

The presence of statistical uncertainty means that even if alphas exist, arbitrageurs might not be able to detect and exploit them reliably. This leads to a "statistical limit" to arbitrage, where the feasibility of identifying and leveraging arbitrage opportunities is constrained by the arbitrageurs' inability to perfectly learn the true alphas from historical data. Consequently, the asymptotic limit of alpha in this scenario is different from what rational expectations would imply because it acknowledges the practical limitations faced by arbitrageurs, including incomplete information, model misspecification risks, and the impact of learning under uncertainty. These factors can result in mispricings persisting in equilibrium, contrary to the rational expectations model, where such mispricings would be swiftly eliminated by fully informed market participants.

Therefore, we say a strategy is feasible if the portfolio weights/allocation only uses observable data, that is, a feasible strategy at t needs to be adapted to the filtration (information set) generated by a function of observables from $t - T + 1$ to t. Furthermore, [1] formalizes the concept of near-arbitrage in a more general setting: 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 $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathrm{P}\right)$, if it satisfies $w \in \mathcal{F}_t$, and along some diverging sub-sequence, with probability approaching one,

$$\operatorname{Var}\left(w^\top r_{t+1} \mid \mathcal{F}_t\right) \to 0, \quad \mathrm{E}\left(w^\top r_{t+1} \mid \mathcal{F}_t\right) \geq \delta > 0. \tag{2}$$

Intuitively, no near-arbitrage means there exists no sequence of portfolios that could obtain positive expected returns with vanishing risks.

2 Connection with Empirical Bayes

2.1 The Optimal Feasible Sharpe Ratio and Weight

To draw the connection between the optimal feasible portfolio and empirical Bayes method, especially the empirical Bayes estimator based on Tweedie's formula ([8]), [1] first shows that any feasible portfolio weight $\hat{w}$ has a Sharpe ratio upper bound $S(\mathcal{G})$ with

$$S(\mathcal{G}) = \sqrt{\mathrm{E}[\alpha|\mathcal{G}]^T \Sigma_u^{-1} \mathrm{E}[\alpha|\mathcal{G}]} \tag{3}$$

where $\mathcal{G} = \{(r_s, \beta, v_s, \Sigma_u)) : t - T + 1 \leq s \leq t\}$ denotes the information available at t for the past T timestamps. Recall that a "feasible" portfolio weight $\hat{w}$ can only use information in $\mathcal{G}$ without having access to the DGP.

It is shown by [5] that the optimal feasible portfolio weight with Sharpe ratio reaching $S(\mathcal{G})$ is $w^* = \mathbb{M}_\beta \Sigma_u^{-1} \mathrm{E}[\alpha|\mathcal{G}]$ where $\mathbb{M}_\beta$ is the projection ("hat") matrix $\mathbb{M}_\beta = \mathbb{I}_N - \beta(\beta^T\beta)^{-1}\beta^T$. Therefore, to construct a portfolio with weights close to w^*, we need to figure out how to evaluate the expectation $\mathrm{E}[\alpha|\mathcal{G}]$.

2.2 Additional Assumptions and Normal Means Model

In order to obtain a better characterization of $\mathrm{E}[\alpha|\mathcal{G}]$, the paper imposes two additional assumptions on the model:

(a) $s_i := \frac{\alpha_i}{\sigma_i}$ is independent of σ_i and satisfies $E\left(s_i^2 \mathbf{1}_{\{|s_i| \geq c_N\}}\right) \leq cN^{-1}$ for some sequence $c_N \to 0$.

(b) $\varepsilon_{i,t}$ follows a standard normal distribution.

Based on the assumptions, we first notice that $\hat{\alpha}$, i.e. the estimated α by regressing on $\mathcal{G}$, satisfies $\hat{\alpha}_i/\sigma_i \sim N(s_i, 1)$[1]. Conditioning on σ_i^2 and assuming $s_i \sim G$ from some unknown prior G, the setting resembles the normal means problem encountered in empirical Bayes problems.

To be more concrete, we now can rewrite

$$\mathrm{E}[\alpha_i|\mathcal{G}] = \mathrm{E}[\alpha_i|\hat{\alpha}_i, \sigma^2] = \sigma_i \mathrm{E}[s_i|\tilde{z}_i] \tag{4}$$

where in the first step, our earlier assumptions make sure that $\hat{\alpha}_i, \sigma^2$ are sufficient in summarizing information from $\mathcal{G}$. We then define $\tilde{z}_i = \sqrt{T}(\frac{\hat{\alpha}_i}{\sigma_i})$ to further refine the dependence. It is easy to see that $\tilde{z}_i \sim N(\sqrt{T}s_i, 1)$ and we are ready to tackle the posterior expectation $\mathrm{E}[s_i|\tilde{z}_i]$ using Tweedie's formula.

2.3 Tweedie Formula and Construction of Optimal Portfolio

Recall that, in general, if the parameter $\mu \sim G$ for some unknown prior G, and the observation has likelihood $z|\mu \sim N(\mu, 1)$, the Gaussian version of the Tweedie's formula ([2]) gives us

$$\mathrm{E}[\mu|z] = z + \frac{d}{dz} \log f_G(z) \tag{5}$$

where $f_G(z)$ is the marginal of z that depends on prior G. In our setting, we thus have

$$\mathrm{E}[s_i|\tilde{z}_i] = \psi(\tilde{z}_i) = \frac{1}{\sqrt{T}}\tilde{z}_i + \frac{1}{\sqrt{T}}\frac{d}{dz} \log p(\tilde{z}_i) \tag{6}$$

Since the prior on s_i is unknown, we instead take an empirical Bayes approach to estimate the marginal $p(\tilde{z}_i)$ by Gaussian kernel density estimation:

$$\hat{p}(a) = \frac{1}{Nk_N} \sum_i \phi\left(\frac{\hat{z}_i - a}{k_N}\right) \tag{7}$$

[1]we originally have $\hat{\alpha}_i = \alpha_i + \bar{u}$, then independence between s_i and σ_i as well as normality of $\epsilon_{i,t}$ gives the result.

where $k_N = (\log N)^{-1}$ is the bandwidth and ϕ is the standard normal pdf. Plugging in $\hat{p}$ and adding some correction factors, we have

$$\mathrm{E}[s_i|\tilde{z}_i] \approx \hat{\psi}(\tilde{z}_i) = \frac{1}{\sqrt{T}}\tilde{z}_i + \frac{1+k_N^2}{\sqrt{T}}\frac{d}{dz}\log\hat{p}(\tilde{z}_i) \tag{8}$$

Finally, taking the estimator back to the optimal feasible portfolio weight in **2.1**, we can construct a portfolio weight (that approximates the optimal weight) with

$$\hat{w}^{OPT} = \mathbb{M}_\beta \Sigma_u^{-1}\hat{\psi}(\tilde{z}_i) = \mathbb{M}_\beta \breve{w} \tag{9}$$

where each entry $\breve{w}_i = \hat{\psi}(\tilde{z}_i)/\hat{\sigma}_i$ (notice that here we estimate $\hat{\sigma}_i$ as a plug-in for σ_i. This makes the above construction "all-weather" even when the signal strength is unknown).

3 Main Findings

3.1 Optimality of the proposed algorithm

This section assumes that the true β is known. The proposed estimator makes heuristic sense: when we know the true alphas, the optimal portfolio weight is $\mathbb{M}_\beta\Sigma_u^{-1}\alpha$, where the "hat" matrix $\mathbb{M}_\beta$ eliminates the systematic risk exposure. When the true α is not observed, the posterior mean

$$w^{\mathrm{OPT}} = \mathbb{M}_\beta\Sigma_u^{-1}\mathrm{E}[\alpha|\mathcal{G}] \tag{10}$$

achieves the best feasible Sharpe ratio given the data generation procedure. The proposed estimator is a finite sample version of w^{OPT}, which produces a Sharpe ratio that is close to S^{OPT}, as summarized in theorem 2. Below is a brief rephrase of the theorem:

Theorem. *Suppose $T \sim N^d$ with $d \in (\frac{1}{2}, 1)$. Denote $\hat{S}^{OPT}$ as the one-step-ahead Sharpe ratio generated by algorithm 1. Under the assumptions on DGP and the factor model,*

$$\lim_{T,N\to\infty}\sup_{P\in\mathbb{P}}\mathbb{P}(|S^{OPT} - \hat{S}^{OPT}| \geq \epsilon S^{OPT} + \epsilon) = 0 \tag{11}$$

This says that the absolute deviation between $\hat{S}^{OPT}$ and S^{OPT} shrinks below any fraction or constant threshold, whichever dominates, in the limit.

3.2 Benchmark algorithms

The author discussed the dominance of the proposed algorithm above three alternative approaches $A \in$ {cross-sectional regression (CSR), false discovery rate control (B-H), shrinkage (LASSO)} by deriving their closed-forms Sharpe ratio S^A under the setting of Example 1 and a few additional simplifying assumptions. Cross-sectional regression uses $\hat{\alpha} = (\beta^T \beta)^{-1} \beta^T \bar{r}$ and the plug in estimator $w^{\text{CSR}} = \mathbb{M}_\beta \hat{\Sigma_u^{-1}} \hat{\alpha}$. This strategy is strongly dominated (i.e., $\frac{S^{CSR}}{S^{OPT}}$ is small) by our empirical Bayes approach when the signal is strong (in example 1, that means $\frac{\mu}{\sigma}$ is large) but sparse (ρ is small). On top of the CSR, the false discovery rate control method selects assets with alphas that reject the null with Bonferroni's correction. We split the data set into S and S': on S, we filter out the assets with weak signals, while on S', we fit α and w on those remaining assets and assign 0 weight to filtered-out assets. This procedure is overly conservative and performs worse than CSR when the signal is weak, suggesting that simply ignoring weak signals is sub-optimal. The shrinkage approach uses a penalization $p_\lambda(w)$, which (softly) filters weak signals when p uses a 1-norm (LASSO) and is a proportional shrinkage when p uses a 2-norm (ridge). Ridge is equivalent to CSR under the given settings, and using an optimal (infeasible) λ, LASSO performs almost as well as empirical Bayes for most of the ρ and $\frac{\mu}{\sigma}$.

3.3 Simulations Evidence and Empirical Analysis

The author simulated the path r_t under a one-factor CAPM model under the setting of example (1) in the paper [3]. The proposed algorithm uses a bandwidth K_N when estimating the marginal density of the t-statistics. For three different values of K_N, the author simulated $\hat{S}^{OPT}$. The realized Sharpe ratio is also simulated for the three benchmark approaches. For each choice of $(\rho, \mu/\sigma)$, compute the error

$$\text{Err}^A\left(\rho, \frac{\mu}{\sigma}\right) = \frac{|\hat{S}^A - S^{OPT}|}{1 + S^{OPT}} \tag{12}$$

for each of the algorithms $A \in \{\text{CSR, B-H, LASSO, OPT}\}$. The author reported the maximal error over all $(\rho, \frac{\mu}{\sigma})$. The proposed empirical Bayes estimate has the smallest error in almost all cases.

The author conducted an empirical analysis of US equities. A multi-factor model with 16 characteristics and 11 GICS is employed, where the factors are selected based on literature. The author evaluated the model performance by plotting the cross-sectional R^2. The average R^2 is 8.25%, which is rather low, suggesting that the estimation of α is hard. In addition, the author collected the averaged residuals as alphas to investigate the strength and rarity arbitrage opportunities. It turns out that the t-statistics is usually low (only 6.35% exceeding 2.0), and that the magnitude of the largest alpha is modest (at 1.699). This says that rare and weak alpha is the most common scenario.

The realized Sharpe ratio is simulated by recalculating the weight at the end of each month using a 10-year rolling window of returns. The four strategies yield a similar Sharpe ratio, with B-H and OPT slightly higher (at around 0.5).

References

[1] Rui Da, Stefan Nagel, and Dacheng Xiu. *The statistical limit of arbitrage*. Tech. rep. Technical Report, Chicago Booth, 2022.

[2] Bradley Efron. "Tweedie's formula and selection bias". In: *Journal of the American Statistical Association* 106.496 (2011), pp. 1602–1614.

[3] Eugene F Fama and Kenneth R French. "The capital asset pricing model: Theory and evidence". In: *Journal of economic perspectives* 18.3 (2004), pp. 25–46.

[4] Denis Gromb and Dimitri Vayanos. "Limits of arbitrage". In: *Annu. Rev. Financ. Econ.* 2.1 (2010), pp. 251–275.

[5] Jonathan E Ingersoll Jr. "Some results in the theory of arbitrage pricing". In: *The Journal of Finance* 39.4 (1984), pp. 1021–1039.

[6] FY Eric C Lam and KC John Wei. "Limits-to-arbitrage, investment frictions, and the asset growth anomaly". In: *Journal of Financial Economics* 102.1 (2011), pp. 127–149.

[7] John F Muth. "Rational expectations and the theory of price movements". In: *Econometrica: Journal of the Econometric Society* (1961), pp. 315–335.

[8] Herbert E Robbins. "An empirical Bayes approach to statistics". In: *Breakthroughs in Statistics: Foundations and basic theory*. Springer, 1992, pp. 388–394.

[9] Richard Roll and Stephen A Ross. "An empirical investigation of the arbitrage pricing theory". In: *The journal of finance* 35.5 (1980), pp. 1073–1103.

[10] Andrei Shleifer and Robert W Vishny. "The limits of arbitrage". In: *The Journal of finance* 52.1 (1997), pp. 35–55.