Thousands of Alpha Tests*

Stefano Giglio[†]
Yale School of Management
NBER and CEPR

Yuan Liao[‡]
Department of Economics
Rutgers University

Dacheng Xiu[§]
Booth School of Business
University of Chicago

This Version: October 5, 2018

Abstract

Data snooping is a major concern in empirical asset pricing. By exploiting the "blessings of dimensionality" we develop a new framework to rigorously perform multiple hypothesis testing in linear asset pricing models, while limiting the occurrence of false positive results typically associated with data-snooping. We first develop alpha test statistics that are asymptotically valid, weakly dependent in the cross-section, and robust to the possibility of omitted factors. We then combine them in a multiple-testing procedure that ensures that the rate of false discoveries is ex-ante bounded below a pre-specified 5% level. We also show that this method can detect all positive alphas with reasonable strengths. Our procedure is designed for high-dimensional settings and works even when the number of tests is large relative to the sample size, as in many finance applications. We illustrate the empirical relevance of our methodology in the context of hedge fund performance (alpha) evaluation. We find that our procedure is able to select – among more than 3,000 available funds – a subset of funds that displays superior in-sample and out-of-sample performance compared to the funds selected by standard methods.

Key words: Data Snooping, Multiple Testing, Alpha Testing, Factor Models, Hedge Fund Performance, False Discovery Rate, Machine Learning

*We benefited tremendously from discussions with Campbell Harvey, Chris Hansen, Bryan Kelly, and Olivier Scaillet.
[†]Address: 165 Whitney Avenue, New Haven, CT 06520, USA. E-mail address: `stefano.giglio@yale.edu`.
[‡]Address: 75 Hamilton St. New Brunswick, NJ 08901, USA. Email: `yuan.liao@rutgers.edu`.
[§]Address: 5807 S Woodlawn Avenue, Chicago, IL 60637, USA. E-mail address: `dacheng.xiu@chicagobooth.edu`.

1 Introduction

Multiple testing is pervasive in empirical finance. It takes place, for example, when trying to identify which among hundreds of factors add explanatory power for the cross-section of returns, relative to an existing model. It also appears when trying to identify which funds are able to produce positive alpha (i.e., have "skill"), among thousands of existing funds – a central question in asset management. In all these examples, the standard approach is to perform many individual statistical tests on the alphas of the factors or funds relative to the benchmark model, and then make a selection based on the significance of these individual tests.

With multiple testing comes the concern – closely related to data snooping – that as more and more tests are performed, an increasing number of them will be positive purely due to chance. Even if each test individually has a low probability of being due to chance alone, a potentially large fraction of the tests that ex post appear positive will be "false discoveries". A high "false discovery rate" (that is, when a large fraction of the tests that appear significant ex-post are expected to be due to chance) decreases the confidence we have in the testing procedure; in the extreme case, if the false discovery rate of a procedure approaches 100%, the significance of the individual tests becomes completely uninformative. To complicate things, the probability of false discoveries associated with any selection procedure is hard to quantify ex-ante, because it depends on the true (unknown) parameters of the model. For example, in the case of the evaluation of fund alphas, the false discovery rate depends on the true underlying distribution of alphas across funds, which is unobservable.

The existing literature in asset pricing is aware of these data-snooping concerns with multiple testing, and has taken in response two alternative approaches. One has been to abandon the multiple testing problem altogether: for example, rather than trying to identify *which* funds or factors have alphas, an alternative is to ask whether *any* fund beats the benchmark, or whether funds *on average* beat the benchmark. This approach can overcome the multiple-testing problem, since it replaces a multitude of null hypotheses (one per fund) with one joint null hypothesis; but it throws the baby out with the bathwater, as it cannot tell us which of the funds actually produces alpha. The second approach, proposed by the pioneering work of Barras et al. (2010), Bajgrowicz and Scaillet (2012), Harvey et al. (2015), etc, applies statistical methods that directly control the false discovery rate. While the recent statistical advances on false discovery control have been successful in many fields (like biology and medicine), their general applicability to finance is still not well understood. The main issue at play is that many of the assumptions on which these methods are based are clearly violated in many finance settings.

In this paper we propose a rigorous framework to address the data snooping issue that arises in a specific, but fundamental, finance setting: testing for multiple alphas in linear asset pricing models. We build upon the multiple-testing procedures proposed in the statistical literature, but extend them to the context of asset pricing, a task that requires additional estimation steps and new asymptotic theory.

The key idea on which our method is based is the concept of *false discovery rate (FDR) control*, introduced by Benjamini and Hochberg (1995) (hereinafter B-H), and advocated for the context of asset pricing by Harvey et al. (2015). The idea of the FDR control procedure is to optimally set different significance thresholds across the different individual tests in such a way that the false discovery rate of the procedure is bounded below a pre-specified level, for example 5%. The standard B-H procedure takes as input the set of t-statistics corresponding to N independent tests and compares them with the appropriate thresholds to decide which of the N hypotheses to reject (in the case of funds, for example, to choose which alphas are significantly positive).

Unfortunately, standard FDR control procedures like B-H cannot be applied directly to the context of asset pricing models, because of several fundamental challenges that our procedure aims to overcome. First of all, returns of hedge funds in excess of the standard benchmarks appear to be highly cross-sectionally correlated, suggesting that fund managers may trade common factors that are not observable. This means that there are plausibly omitted priced factors in the benchmark that can bias the resulting alpha estimates, and at the same time produce strong cross-sectional dependence in the excess returns over the benchmarks. Both issues invalidate the standard FDR control methods, which rely on consistency of the estimators and independence among the test statistics.

A second challenge is that the B-H procedure is designed for a fixed number of test hypotheses N: the theory assumes that length of the time series $T \to \infty$, but N does not grow asymptotically. In most finance application, however, the number of tests to be run is overwhelmingly large relative to the available sample size. For example, in our baseline empirical analysis we have around 270 months to evaluate more than 3,000 hedge funds. Even worse, many funds have short histories or missing records, which further diminishes the effective sample sizes. In this "large N, large T" context, many of the statistical tests for alpha face limitations. For example, GLS estimators and the GRS statistic become invalid, because it is infeasible to estimate the inverse of the covariance matrix of the residuals. The B-H procedure is no exception. While its properties are affected by the large dimensionality of the cross-section, little work has been done to explore them. This is something we directly tackle in the paper.

The fixed-N assumption of the B-H test poses a third challenge related to asset pricing tests. To evaluate alpha of a factor or a fund relative to a benchmark model, the risk premia of the benchmark factors need to be estimated. When the factors are nontradable, their estimates need to be obtained from cross-sectional regressions like Fama-MacBeth. But when N is fixed, the risk premia estimates are biased, since in any finite sample the orthogonality of the betas (the regressors in the cross-sectional regression) and the alphas (the residuals of the cross-sectional regression) is not guaranteed, even if they are orthogonal in population.[1] The bias in the risk premia spills over into biased estimates of

[1]Having $T \to \infty$ is of no help, because the risk premia are slopes in a cross-sectional regression for which the relevant sample size is N.

the alphas, and ultimately biases the FDR control test.

Our procedure tackles these challenges by exploiting the "blessing" of dimensionality – the other side of the "curse" of dimensionality (Donoho et al. (2000)) – that obtains as both $N, T \to \infty$. Our FDR control test proceeds as follows. We first estimate latent factors that drive the comovement of returns in excess of any observable factors, and augment the benchmark model with these estimated latent factors. We then use the cross-section of returns to estimate the risk premia of all factors (tradable, non-tradable, and latent) and all the individual alphas. Next, we compute asymptotic t-statistics that are weakly dependent in the cross-section, and use them in the B-H procedure. Finally, to further enhance the power of the FDR control procedure, we apply a screening method recently developed in the machine learning literature prior to the B-H step.[2] In the paper, we develop general asymptotic theory for our FDR control test in the high-dimensional asset-pricing setting.

The high dimensionality of the data on which our procedure relies helps us address the challenges described above in several ways. It allows us to consistently estimate any latent or omitted factors driving returns, something possible only as $N \to \infty$ (see for example Giglio and Xiu (2017)). This both eliminates the omitted factor bias and removes this source of common variation in the residuals. The high dimensionality also removes the bias in the cross-sectional estimation of risk premia of nontradable factors, since as we show in the paper the bias vanishes as $N \to \infty$. Finally, exploiting the large dimensionality, we develop asymptotically valid alpha test statistics that are weakly dependent in the cross-section. And of course, a procedure designed for large cross-sections allows us to tackle real-world situations, in which N is much larger than T.

We illustrate this procedure using the Lipper TASS data of hedge fund returns. We show empirically that hedge fund returns are highly correlated in the cross-section, even after controlling for the standard models, like the Fung-Hsieh 7-factor model or the Carhart 4-factor model. This is perhaps not surprising, as it is to be expected that many hedge fund strategies may be highly correlated; but it is something that we need to explicitly take into account when constructing our FDR control test. We show that our procedure – which bounds the false discovery rate to 5% – selects a significantly smaller number of funds based on their alpha compared to a standard set of individual t-tests. We therefore identify a variety of funds that consistently beats the benchmarks and delivers excess returns that are likely due to managerial skills instead of pure luck. In addition, the funds that are selected by our procedure yield superior out-of-sample performance compared to funds that are selected by the standard multiple testing procedure.

Our paper sits at the confluence of studies of asset pricing anomalies and fund managerial performance. All of these literatures perform multiple testing of alphas relative to a benchmark model.

[2] The screening step removes funds with ex-post particularly negative alphas from the set of candidate hypotheses, thus helping contrast the increase in N. The same idea based on screening was previously used in related literature by Hansen (2005); Chernozhukov et al. (2013b) and Romano and Wolf (2018). Our theory shows that asymptotically the new alpha-screening B-H procedure has a larger power on detecting positive alphas, and can consistently identify all positive alphas with reasonably strong signals.

Studies of anomalies interpret the alphas as either risk premia from a new factor or evidence of mispricing, whereas studies of fund alphas typically interpret them as reflecting managerial skills.

Data snooping has been a central topic in statistics ever since the early 1950s. Earlier work mainly focus on using Bonferroni-type procedures to control the family wise error rate (FWER), see, e.g., Simes (1986), Holm (1979). These procedures guard against any single false discovery and hence is overly conservative in particular when testing many hypotheses. Instead of targeting the question whether any error was made, the approach of FDR control, developed by Benjamini and Hochberg (1995), takes into account the number of erroneous rejections and control the expected proportion of errors among the rejected hypotheses. While this seminal work relies on the independence assumption among test statistics, following-up studies such as Benjamini and Yekutieli (2001) and Storey et al. (2004) demonstrate the robustness of this procedure to certain forms of dependence. Nevertheless, the literature, e.g., Schwartzman and Lin (2011), Fan et al. (2012), have recognized the drawbacks of the standard FDR approach in the presence of dependence, including overly conservativeness, high variance of the estimated FDR, etc, and proposed alternative procedures, e.g., Leek and Storey (2008), Romano and Wolf (2005), and Fan and Han (2016). Our procedure shares the spirit with the recent statistical literature, but differs substantially because our solution is specifically designed for empirical asset pricing models.

Data snooping bias has long been recognized as a serious concern in empirical asset pricing. Lo and MacKinlay (1990) point out and quantify such a bias in tests of asset pricing models with portfolios formed by sorting on some empirically motivated characteristics. White (2000) propose a reality check approach that summarizes in a single statistic the significance of the best-performing model after accounting for data-snooping in many individual tests. Sullivan et al. (1999) apply this reality check approach to investigate the profitability of technical trading rules. Harvey et al. (2015) and Harvey and Liu (2016) shed light on the data snooping issue in the proliferation of factors, and suggest various procedures, such as the use of reality check, FDR, and higher thresholds of significance level, to guard against false discoveries and promote reproducible research.

Data mining is pervasive in machine learning. Armed with flexible models and efficient algorithms, machine learning is well suited for prediction with countless potential covariates. However, because chance can lead to suspicious associations between some covariates and the outcome of interest, variable selection mistakes inevitably arise, to the extent that the best predictive model rarely coincides with the true data generating process. There is a burgeoning body of research that applies machine learning methods to push the frontiers of empirical asset pricing, see, e.g., Kozak et al. (2017), Freyberger et al. (2017), Giglio and Xiu (2017), Feng et al. (2017), Kelly et al. (2017), Gu et al. (2018). These papers employ variable selection or dimension reduction techniques to analyze a large number of covariates. These procedures, however, do not directly control the number of false discoveries. Our paper proposes a rigorous FDR procedure that allows for a large number of alpha tests relative to the sample size.

Finally, our empirical results directly speak to a long literature dedicated to evaluating the performance of the hedge fund industry. Contrary to the case of mutual funds, for which net alpha is estimated to be zero or negative for the vast majority of funds (with some exceptions, evidenced in the recent work of Berk and Van Binsbergen (2015)), there is more evidence that hedge funds are able to generate alpha. An important first step in this empirical exercise has been the exploration of hedge fund strategies and their risk exposures (see Fung and Hsieh (1997, 2001, 2004); Agarwal and Naik (2000, 2004); Agarwal et al. (2009); Patton and Ramadorai (2013); Bali et al. (2014)); we use many of the benchmarks proposed in this literature as observable factors in our analysis. At the same time, the literature has explored whether hedge funds are able to produce alpha in excess of these benchmarks, using different statistical methodologies (Liang (1999); Ackermann et al. (1999); Liang (2001); Mitchell and Pulvino (2001); Baquero et al. (2005); Kosowski et al. (2007); Fung et al. (2008); Jagannathan et al. (2010); Aggarwal and Jorion (2010); Bali et al. (2011)).

Section 2 discusses the detailed procedure for FDR control. Section 3 presents Monte Carlo simulations, followed by an empirical study in Section 4. Section 5 concludes. The appendix provides the asymptotic theory and technical details.

2 Methodology

Our framework is based on a combination of three key ingredients, each essential to execute multiple testing correctly in the asset pricing context: Fama-MacBeth regressions, principal component analysis (PCA), and FDR control. Our approach proceeds as follows. In a first step, we use time-series regressions to estimate fund exposures to (observable) fund benchmarks. Since these benchmarks do not fully capture the common comovement of fund returns, hiding for example unobservable risk exposures, we further apply PCA to the residuals to recover the missing commonalities. This results in a model where, effectively, both observable and estimated latent factors coexist.

Next, we implement the cross-sectional regressions like Fama-MacBeth to estimate the risk premia of the factors and the alphas relative to the augmented benchmark model that includes both observable and estimated latent factors. Finally, we build t-statistics for these alphas, and apply the FDR cpmtrol to determine which alphas are significantly positive. In what follows, we describe each ingredient in details; we then derive the statistical properties of our procedure, and show formally that it indeed achieves the desired false discovery rate in the multiple test of alpha, and identifies all positive alphas with reasonable strengths.

2.1 Model Setup

We begin with a description of the model. We assume the $N \times 1$ vector of excess returns r_t follows a linear factor model:

$$r_t = \alpha + \beta\lambda + \beta(f_t - \mathbb{E}(f_t)) + u_t, \tag{1}$$

where f_t is a $K \times 1$ vector of factors and u_t is the idiosyncratic component. The parameter λ is a $K \times 1$ vector of factor risk premia, which is identical to the expected return of f_t only if f_t is tradable.[3]

The objective is to find individual funds with truly positive alphas. To do so, we formulate a collection of null hypotheses, one for each fund:

$$\mathbb{H}_0^i : \alpha_i \leq 0, \quad i = 1, \ldots, N. \tag{2}$$

Importantly, the alpha testing problem we consider is fundamentally different from the standard GRS test, in which the null hypothesis is a single statement that

$$\mathbb{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_N = 0. \tag{3}$$

The former is a multiple testing problem that addresses which funds have significantly positive alphas. In contrast, the latter addresses whether there exists (at least one) fund whose alpha is significantly different from zero. While the latter is the natural way to test asset pricing models (which implies that all alphas should be zero), it is not the right one if the objective is to identify which funds are able to generate positive alpha.

Simultanteous testing of multiple hypotheses – like the test we propose – is prone to a false discovery problem, also referred to as data snooping bias: the possibility that many of the tests will look significant by pure chance, even if their true alpha is zero. To understand why, recall that for each 5%-level test, there is a 5% chance that the corresponding null hypothesis is falsely rejected. This is the so-called Type-I error. In other words, there is a 5% chance that a fund with no alpha realizes a significant test statistics and is therefore falsely recognized as one with real alpha.

This error exacerbates substantially when testing many hypotheses. For example, suppose there are 1,000 funds available, with only 10% of them having positive alphas. Conducting 1,000 tests independently would yield $1,000 \times (1 - 10\%) \times 5\% = 45$ false positive alphas, in addition to $1,000 \times 10\% = 100$ true positive alphas (assuming ideally a zero Type-II error). Consequently, among the $100 + 45 = 145$ "skilled" fund managers we find, almost 1/3 of them are purely due to luck. The false discovery rate of the test is the 1/3 number that reflects how many of the significant tests are expected to be false discoveries.

The multiple testing problem is one of the central concerns in statistics and machine learning. Since the 1950s, the literature has proposed various alternatives for assessing and correcting data-snooping bias. One of the classical approaches is to control the probability of *one or more* false rejections, i.e., the family-wise error rate (FWER), instead of the Type-I Error. One such approach is the Bonferroni procedure, which suggests rejecting the null of individual hypothesis at the $5\%/N$ level, where N is

[3] Throughout we impose an unconditional factor model in which both α and β are time-invariant. This is a practical trade-off between efficiency and robustness (to model misspecification) in light of the limited sample size in our empirical analysis. That said, it is straightforward to extend our theory to conditional models à la Ang and Kristensen (2012).

the total number of tests. However, this method is overly conservative when the number of hypotheses N is large relative to the sample size T, in that the level of the test shrinks to zero asymptotically. That is because to ensure that the probability that even just one of the N tests is a false discovery stays below a certain level, say 5%, it requires the procedure to adopt a higher and higher threshold as the number of tests run N increases; this will result in an unfeasibly high bar for the t-statistic of each test.

A more suitable procedure in this scenario is to control the *false discovery rate* (FDR) instead, i.e., the expected fraction of false rejections; this is the purpose of the original B-H procedure, that has been the most popular since it was introduced and has been widely used across disciplines. We now turn to describing the B-H procedure and showing under what conditions it can be applied in an asset pricing context.

2.2 Controlling the False Discovery Rate

We start by setting up some notation. Suppose t_i is a test statistic for the null $\mathbb{H}_0^i$ (often taken as the t-statistic), and a corresponding test which rejects the null whenever $t_i > c_i$ under a prespecified cutoff c_i. Let $\mathcal{H}_0 \subset \{1, ..., N\}$ denote the set of indices for which the corresponding null hypotheses are true. In addition, let $\mathcal{R}$ be the total number of rejections in a sample, and $\mathcal{F}$ be the number of false rejections in that sample:

$$\begin{aligned}\mathcal{F} &= \sum_{i=1}^{N} 1\{i \leq N : t_i > c_i \text{ and } i \in \mathcal{H}_0\},\\ \mathcal{R} &= \sum_{i=1}^{N} 1\{i \leq N : t_i > c_i\}.\end{aligned}$$

Both $\mathcal{F}$ and $\mathcal{R}$ are random variables. Note that, in a specific sample, we can obviously observe $\mathcal{R}$, but we cannot observe $\mathcal{F}$. However, we can design a test to effectively limit how large $\mathcal{F}$ is in expectation relative to $\mathcal{R}$. More formally: we write the *false discovery proportion* (FDP) and its expectation, FDR, as

$$\text{FDP} = \left(\frac{\mathcal{F}}{\max\{\mathcal{R}, 1\}}\right), \quad \text{FDR} = \mathbb{E}(\text{FDP}).$$

For comparison, we can also write the per test error rate $\mathbb{E}(\mathcal{F})/N$ and the FWER $\mathbb{P}(\mathcal{F} \geq 1)$. The naive procedure that tests each individual hypothesis at a predetermined level $\tau \in (0, 1)$ guarantees that $\mathbb{E}(\mathcal{F})/N \leq \tau$. But note that it does not guarantee any limits on the false discovery rate, which can be much larger than τ. The Bonferroni procedure, instead, tests each hypothesis at a level τ/N. This guarantees that $\mathbb{P}(\mathcal{F} \geq 1) \leq \tau$, and implies false discovery rates below τ, at the cost of reducing the power of the test in detecting the true alphas (in the limit, if a test is so strict that it never rejects, the false discovery rate is zero! But that test will have no power.).

The FDR control procedure (B-H procedure) strikes a balance between these two approaches. It accepts a certain number of false discoveries as the price to pay to gain power in detecting true rejections. Analogously to standard individual tests (that control the size of type-I error), this procedure controls the size of the FDR: it ensures that satisfies FDR $\leq \tau$.

We now describe the details of the B-H procedure.

Algorithm 1 (B-H procedure)**.**

S1. Sort the collection of p-values, $\{p_i : i = 1, \ldots, N\}$, of the individual test statistics $\{t_i\}$. Denote $p_{(1)} \leq \ldots \leq p_{(N)}$ as the sorted p-values.

S2. For $i = 1, \ldots, N$, reject $\mathbb{H}_0^i$ if $p_i \leq p_{(\widehat{k})}$, where $\widehat{k} = \max\{i \leq N : p_{(i)} \leq \tau i/N\}$.

Although Benjamini and Hochberg (1995) establish the validity of the procedure under certain conditions, a crucial assumption they make – typically violated in the asset pricing context – is that the test statistics $\{t_i : i \leq N\}$ are independent.[4] Violation of this assumption induces two fundamental problems: it deteriorates the power of the FDR test, and it magnifies the variance of the FDP, see, e.g., Schwartzman and Lin (2011), Fan et al. (2012). To apply the B-H procedure, therefore, we must construct test statistics that are weakly dependent or nearly independent.

The other side of a testing problem is "power", i.e., the ability to detect false null hypotheses. For each fixed hypothesis $\mathbb{H}_0^i$, the critical value of the t-statistic is based on an asymptotic distribution that $\alpha_i = 0$, although negative values of α_i also conform with the null hypotheses. This enlarged critical value makes the popular FDR control test (B-H procedure) often overly conservative; this is particularly true when the dimension N is very large, because the B-H compares the i th largest p-value $p_{(i)}$ with the critical value $\tau i/N$ where the critical value is very small when all hypotheses are being considered. So the power of the B-H procedure adversely depends on the number of candidate hypotheses. We tackle this problem by using a simple yet powerful dimension reduction technique – the screening method. The idea is that when some of the alphas are "overwhelmingly negative" (which we call "deep in the null"), they should be simply removed from the set of candidate hypotheses. Based on this idea, we propose to reduce the set of funds to

$$\widehat{\mathcal{I}} = \left\{ i \leq N : t_i > -\sqrt{\log(\log N)} \right\}.$$

Our theory (presented in the appendix) shows that with probability approaching one, for any i such that $\widehat{\alpha}_i \notin \widehat{\mathcal{I}}$, the true $\alpha_i < 0$, indicating that we can simply accept this null hypothesis without further considering it in the B-H procedure. Hence we can focus on a much smaller set $\widehat{\mathcal{I}}$ to conduct the FDR control procedure.

Algorithm 2 (Alpha-screening B-H procedure)**.** Let $|\widehat{\mathcal{I}}|$ denote the number of elements in $\mathcal{I}$.

[4] To remedy this Benjamini and Yekutieli (2001) revise the use of $\widehat{k}$ in S2 by $\widehat{k} = \max\{i \leq N : p_{(i)} \leq \tau i/(NC_N)\}$. where $C_N = \sum_{i=1}^{N} i^{-1}$, which they show guarantees FDR control under certain form of dependence. Nonetheless, because of $C_N \approx \log(N) + 0.5$, $C_{1,000} \approx 7.4$, the method remains too conservative, limiting the power of the procedure.

S1. Sort the p-values, $p_{(1)} \leq \ldots \leq p_{(|\widehat{\mathcal{I}}|)}$ for $\{p_i : i \in \mathcal{I}\}$.

S2. For $i \in \mathcal{I}$, reject $\mathbb{H}_0^i$ if $p_i \leq p_{(\widehat{k})}$, where $\widehat{k} = \max\{i \leq |\widehat{\mathcal{I}}| : p_{(i)} \leq \tau i/|\widehat{\mathcal{I}}|\}$. Accept all other $\mathbb{H}_0^i$.

The critical value now becomes $\tau i/|\widehat{\mathcal{I}}|$ which is potentially much smaller. We formally show in the appendix that asymptotically the new alpha-screening B-H procedure has a larger power on detecting positive alphas, and can consistently identify *all* positive alphas with reasonably strong signals.

In a setting similar to ours, Harvey and Liu (2018) propose to increase statistical power by dropping funds that appear for few periods. Alternatively, Barras et al. (2010) apply a simple adjustment proposed by Storey (2002) to improve the power of the B-H procedure. Specifically, they suggest replacing $i\tau/N$ in the cutoff value by $i\tau/N_0$, where N_0 is the number of true null hypotheses that can be estimated using $\widehat{N}_0 = (1-\lambda)^{-1}\sum_{i=1}^{N}\{p_i > \lambda\}$, where $\lambda \in (0,1)$ is a user-specified parameter. The intuition behind this adjustment is that under the zero-alpha nulls, the p-values are uniformly distributed on $(0,1)$, therefore one would expect $N_0(1-\lambda)$ of the p-values to lie within the interval $(\lambda, 1)$. Replacing N by $N_0 < N$ thereby increases the power of the procedure. In our context, however, this adjustment is not applicable because our null hypotheses are inequalities, under which the p-values are no longer uniformly distributed. The deviation from the uniform distribution becomes very severe when many alphas are very negative, which would substantially overestimate N_0, eventually resulting in conservative discoveries. In contrast, our proposal is specifically designed to deal with the scenario of large negative alphas, hence is effective in this setting.

It is interesting to think about our screening procedure in relation to the problem of selection bias in hedge fund reporting (which affects our empirical application). A well known problem with standard hedge fund datasets (see for example Agarwal et al. (2013)) is that it is likely that "bad" funds (either funds with particularly negative alpha or realized performance) will simply not report to the dataset. This is of course an important issue for understanding the *average* alpha (denoted by α_0) of hedge funds (in our context, α_0 will be biased upwards). But when the objective is to identify funds with skill, this bias is much less relevant. The fact that funds with truly negative alpha, and funds that would have anyway displayed a negative t-statistic, are excluded from consideration, has the same effect as our screening step: it increases the power of the methodology to identify good funds among those that do report.

2.3 Issues with Classical Alpha Tests

There are two main issues that need to be solved in order to properly test for the alphas of asset or fund returns (either using multiple nulls or a single null on all alphas): the estimation of the risk premia of the benchmark factors, and the choice of the benchmark itself, with the possibility of omitted factors.

As discussed in detail in Cochrane (2009), when the benchmark includes non-tradable factors, estimating (1) requires two-pass Fama-MacBeth regressions. The first stage estimates β using time series regressions of individual fund returns onto the benchmark factors, and the second stage involves a cross-sectional regression of average returns onto the estimated β, where the residuals of this regression yield estimates of alpha. We can write this estimator explicitly as:

$$\widehat{\alpha} = \bar{r} - \widehat{\beta}\widehat{\lambda}, \tag{4}$$

where $\bar{r}$ is the $N \times 1$ time series average return, $\widehat{\beta}$ is the time series estimates, and $\widehat{\lambda}$ is the slope estimated in the cross-sectional regression.

The classical setting assumes a fixed dimension N, so that the asymptotic theory is developed under $T \to \infty$ only. In this case, the asymptotic covariance matrix of $\widehat{\alpha}$ (see Cochrane (2009)) is:

$$\mathrm{Cov}(\widehat{\alpha}) = \frac{1}{T}\left(\mathrm{I}_N - \beta(\beta'\beta)^{-1}\beta'\right)\Sigma_u\left(\mathrm{I}_N - \beta(\beta'\beta)^{-1}\beta'\right). \tag{5}$$

Here Σ_u is the covariance of the idiosyncratic components. A closer scrutiny of the derivation of this formula suggests that even in the absence of latent factors, it actually only holds under the GRS null hypothesis $\mathbb{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_N = 0$. When we examine the Fama-MacBeth estimator of alpha, we can see that more generally,

$$\widehat{\alpha} - \alpha = -\beta(\beta'\mathbb{M}_{1_N}\beta)^{-1}\beta'\mathbb{M}_{1_N}\alpha + O_P(T^{-1/2}), \tag{6}$$

where $\mathbb{M}_{1_N} = \mathrm{I}_N - N^{-1}1_N 1_N'$, which indicates that there exists a dominant bias term that prohibits a consistent estimator of α. We give a formal statement on the inconsistency for fixed N in Proposition 1 in the appendix. Importantly, $\widehat{\alpha}_i$ is inconsistent even if $\alpha_i = 0$ for any i, as long as $\beta_i \neq 0$. The bias arises from the cross-sectional correlation between betas and alphas. This is not surprising since the cross-sectional regression requires an exact orthogonality condition between residual and regressors, which is not satisfied if the alphas are not zero. Consequently, when the dimension N is fixed, the two-pass regression does not apply to the tests of multiple hypothesis. Since we do not want to exclude that some of the alphas are actually nonzero (since we are doing the hypothesis tests one by one), this creates fundamental obstacles to the multiple testing problem.

A potential solution to this problem is to estimate and test alphas using only time series regressions. This approach appears viable but only in the case in which all factors are excess returns; this assumption is often violated in practice, for example in the case of one of the main models used to evaluate hedge fund performance (the 7-factor model of Fung and Hsieh (2004)).

A second concern relates to the choice of the benchmark and the possibility that some important factors are omitted, thus attributing to alpha what truly is just exposure to the omitted risk factors. More explicitly, consider a specific example of (1):

$$r_t = \alpha + \begin{bmatrix} \beta_o & \beta_l \end{bmatrix} \begin{bmatrix} f_{o,t} \\ f_{l,t} \end{bmatrix} + u_t = \underbrace{\alpha + \beta_l \mathbb{E} f_{l,t}}_{\text{“alpha”}} + \beta_o f_{o,t} + \underbrace{\beta_l(f_{l,t} - \mathbb{E} f_{l,t}) + u_t}_{\text{“idiosyncratic” error}} \tag{7}$$

where $f_{o,t}$ is the observed benchmark model and $f_{l,t}$ is the vector of omitted factors missing from the benchmark. To make things simple in this example (but both conditions are not required in our general specification below), assume that both f_l and f_o are excess returns and that they are uncorrelated.

The "alpha" computed relative to the benchmark model that just includes f_o (and thus omits f_l) includes the risk premium associated with the missing factor f_l. As long as the latent factors in f_l contribute to the total risk premia, then a bias $\beta_l \mathbb{E} f_{l,t}$ would arise in the estimated "alpha."

Moreover, even if f_l is not priced, it plays the role of "idiosyncratic" error. Since the idiosyncratic error covariance matrix appears in the asymptotic covariance matrix of the alpha estimates, the presence of f_l in the residuals produces strong correlation among the alpha test statistics, which invalidates the independence assumption of the B-H procedure.

In practice, it is difficult to obtain a benchmark model that possibly summarizes all commonalities in the trading strategies of the universe of hedge funds. Indeed, close scrutiny of time series regressions of various factor models in our empirical analysis suggests the existence of common factors in the residuals. Again, we will exploit the blessings of dimensionality to correct for this potentially unobserved commonality.

2.4 Estimating Alpha

In this section, we explain how our new test statistics overcome these obstacles by exploiting the blessings of dimensionality. Our framework is developed for a model more general than (7):

$$r_t = \alpha + \begin{bmatrix} \beta_o & \beta_l \end{bmatrix} \begin{bmatrix} \lambda_o \\ \lambda_l \end{bmatrix} + \begin{bmatrix} \beta_o & \beta_l \end{bmatrix} \begin{bmatrix} f_{o,t} - \mathbb{E} f_{o,t} \\ f_{l,t} - \mathbb{E} f_{l,t} \end{bmatrix} + u_t, \tag{8}$$

where $f_{o,t}$ is a $K_o \times 1$ vector of observable factors, and $f_{l,t}$ is a $K_l \times 1$ vector of latent factors, respectively. Both factors can be non-tradable.

To better explain the intuition, we start with two special cases. For convenience, we introduce some additional notation. In what follows, we use capital letter A to denote the matrix $(a_1, a_2, \ldots, a_T)$, where a_t is a time series of vectors. We use $\mathbb{M}_B = \mathrm{I}_p - B(B'B)^{-1}B$ to denote the annihilator matrix for any $p \times q$ matrix B. Let F be the $K_o \times T$ matrix of $\{f_{o,t} : t \le T\}$, V be the $K \times T$ matrix of $\{f_t - \mathbb{E} f_t : t \le T\}$, R be the $N \times T$ matrix of $\{r_t : t \le T\}$ and U be the $N \times T$ matrix of $\{u_t : t \le T\}$.

2.4.1 Observable Factors Only

When all factors are observable, we can directly estimate α using the classical two-pass regression:

Algorithm 3 (Observable Factors Only)**.**

S1a. Run time series regressions and obtain the OLS estimator $\widehat{\beta}$.

$$\widehat{\beta} = (R\mathbb{M}_{1_T} F')(F\mathbb{M}_{1_T} F')^{-1}. \tag{9}$$

S2. Run a cross-sectional regression of $\bar{r}$ on the estimated $\widehat{\beta}$ and a constant regressor 1_N to obtain the slopes $\widehat{\lambda}$:

$$\widehat{\lambda} = (\widehat{\beta}'\mathbb{M}_{1_N}\widehat{\beta})^{-1}(\widehat{\beta}'\mathbb{M}_{1_N}\bar{r}). \tag{10}$$

S3. Estimate α by subtracting the estimated risk premia from average returns:

$$\widehat{\alpha} = \bar{r} - \widehat{\beta}\widehat{\lambda}. \tag{11}$$

As a side note, by including an intercept term in the cross-sectional regression, S2 allows for a possibly nonzero cross-sectional mean for α, α_0. Its estimator can be written explicitly as $\widehat{\alpha}_0 = N^{-1}1_N'\widehat{\alpha}$. It is also interesting to test if α_0 is non-negative or not, which addresses whether on average hedge fund alphas are positive.

2.4.2 Latent Factors Only

Now suppose that all factors are latent, hence first-step time series regressions are not applicable. This latent factor model is in fact quite general in that we can always assume all factors being latent without using any observable factors, and estimate them from the data. In this case, we follow Giglio and Xiu (2017) and proceed by rewriting (8) into a statistical factor model:

$$\bar{R} = \beta\bar{V} + \bar{U},$$

where $\bar{A} = A\mathbb{M}_{1_T}$ for $A = R$, V, and U.

Simply replacing S1a in Algorithm 3 by S1b below leads to a new algorithm for estimating α in this scenario:

Algorithm 4 (Latent Factors Only)**.**

S1b. Let $S_R = \frac{1}{T}\bar{R}\bar{R}'$ be the $N \times N$ sample covariance of S_R. Conduct the principal components analysis (PCA) of $\bar{R}$: set

$$\widehat{\beta} = \sqrt{N}(b_1, \ldots, b_K),$$

where $b_1, \ldots, b_K$ are the K eigenvectors of S_R, corresponding to its largest K eignvalues.

S2. & S3. are the same as in Algorithm 3.

This procedure therefore uses the principal components of returns as factors and uses them as a benchmark to estimate the alphas. Note that Algorithm 4 requires the number of latent factors as an input, which can be estimated using a variety of procedures in the literature, such as those based on information criterions (Bai and Ng (2002)), or based on eigenvalue ratios (Ahn and Horenstein (2013)), etc. Alternatively, we can treat the number of latent factors as an tuning parameter, which can be selected based on eigenvalue scree plot. We adopt this procedure in practice for convenience.

2.4.3 General Case

To estimate α in the general case (8), we combine S1a and S1b, and then proceed with S2 and S3 as in Algorithm 3. Specifically, we first obtain $\widehat{\beta}_o$ from time series regressions using observable factors alone, and then obtain $\widehat{\beta}_l$ by applying SVD on the matrix of residuals from time series regressions. The estimated $\widehat{\beta}_o$ and $\widehat{\beta}_l$ are stack together as $\widehat{\beta}$. The algorithm is summarized as follows.

Algorithm 5 (Estimating α in Model (8))**.**

S1. a. Run time series regressions and obtain the OLS estimator $\widehat{\beta}_o$ and residual matrix Z:

$$\widehat{\beta}_o = (R\mathbb{M}_{1_T}F_o')(F_o\mathbb{M}_{1_T}F_o')^{-1}, \quad Z = \bar{R} - \widehat{\beta}_o\bar{F}_o, \tag{12}$$

where $F_o = (f_{o,1}, f_{o,2}, \dots, f_{o,T})$.

b. Let $S_Z = \frac{1}{T}ZZ'$ be the $N \times N$ sample covariance matrix of Z. Let

$$\widehat{\beta}_l = \sqrt{N}(b_1, \dots, b_{K_l}),$$

where $b_1, \dots, b_{K_l}$ are the K_l eigenvectors of S_Z, corresponding to its largest K_l eigenvalues.

The resulting $\widehat{\beta}$ is given by

$$\widehat{\beta} = (\widehat{\beta}_o, \widehat{\beta}_l).$$

S2. & S3. The same as S2 & S3 in Algorithm 3.

It is worth mentioning that $\widehat{\beta}_o$ is a consistent estimator of β_o only if f_o and f_l are uncorrelated. But in our general setting, this condition is not imposed, so $\widehat{\beta}_o$ is actually inconsistent due to the omitted variable (latent factors) bias. However, one of our theoretical contributions is to show that the presence of such bias does not affect the inference for alphas, thanks to the invariance of alpha to the rotation of the factors. Formally, note that $\widehat{\beta}_o \xrightarrow{P} \beta_o + \beta_l w$ for some matrix w, where $\beta_l w$ denotes the omitted variable bias. Hence the probability limit of $\widehat{\beta}_o$ is still inside the space spanned by $\beta = (\beta_o, \beta_l)$, while the probability limit of the PCA estimator $\widehat{\beta}_l$ is also inside the space spanned by β. As a result, we show that there is a rotation matrix H so that

$$\widehat{\beta} = (\widehat{\beta}_o, \widehat{\beta}_l) \xrightarrow{P} \beta H.$$

The resulting alpha estimate remains consistent because it is invariant to rotations and is thus not affected by the omitted variable bias.

2.5 Constructing Valid Test Statistics for FDR Control

After estimating the alphas, we now build test statistics. One of our theoretical contributions is that we formally show in Theorem 1 of Appendix A.2, the two-pass regression yields: as $N, T \to \infty$ for each $i \leq N$,

$$\begin{aligned} \sigma_{i,NT}^{-1}(\widehat{\alpha}_i - \alpha_i) \quad &\xrightarrow{d} \quad \mathcal{N}(0,1), \\ \sigma_{i,NT}^2 \quad &= \quad \frac{1}{T}\operatorname{Var}(u_{it}(1 - v_t\Sigma_f^{-2}\lambda)) + \frac{1}{N}\operatorname{Var}(\alpha_i)\frac{1}{N}\beta_i' S_\beta^{-1}\beta_i, \end{aligned} \tag{13}$$

where $v_t := f_t - \mathbb{E}f_t$, $\Sigma_f := \operatorname{Cov}(f_t)$ and $S_\beta = \frac{1}{N}\beta'\mathbb{M}_{1_N}\beta$. This formula holds true for all three cases: (observable factors only, latent factors only and the general case). The asymptotic result

(13) can be used for inference about each individual alpha. Note that the variance $\sigma^2_{i,NT}$ consists of two components: in addition to the $1/T$ term that arises from time series estimations, the second term $\frac{1}{N}\operatorname{Var}(\alpha_i)\frac{1}{N}\beta_i' S_\beta^{-1}\beta_i$ directly reflects the statistical estimation errors from the cross-sectional regression.

When $T\log N = o(N)$, the second component in the expansion of $\sigma^2_{i,NT}$ vanishes, and in this case we have the following asymptotic expansion of $\widehat{\alpha}$:

$$\sqrt{T}(\widehat{\alpha}_i - \alpha_i) \;=\; \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it}(1 - v_t'\Sigma_f^{-1}\lambda) + o_P(1/\sqrt{\log N}),$$

therefore the estimated alphas are cross-sectionally weakly dependent.

Let $\operatorname{se}(\widehat{\alpha}_i)$ be the standard error of $\widehat{\alpha}_i$, with which we can calculate the t-statistics:

$$t_i = \frac{\widehat{\alpha}_i}{\operatorname{se}(\widehat{\alpha}_i)}, \quad i = 1, \ldots, N.$$

Using these t-statistics, we then apply the proposed alpha-screening B-H procedure (Algorithm 2) to select the positive alphas.

Algorithm 6 below provides standard errors in the general case.

Algorithm 6 (Construction of the Test Statistics)**.**

S1. & S2. & S3. are the same as those in Algorithm 5.

S4. Calculate the standard error as

$$\operatorname{se}(\widehat{\alpha}_i) \;=\; \frac{1}{\sqrt{T}}\widehat{\sigma}_i, \quad \widehat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{it}^2(1 - \widehat{v}_t'\widehat{\Sigma}_f^{-1}\widehat{\lambda})^2, \tag{14}$$

where $\widehat{u}_{it} = z_{it} - \widehat{\beta}_{l,i}'\widehat{v}_{l,t}$ is the residual, and

$$\widehat{v}_t = \begin{pmatrix} f_{o,t} - \bar{f}_{o,t} \\ \widehat{v}_{l,t} \end{pmatrix}, \quad \widehat{v}_{l,t} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\beta}_{l,i}(z_{it} - \bar{z}_i), \quad \widehat{\Sigma}_f = \frac{1}{T}\sum_{t=1}^{T}\widehat{v}_t\widehat{v}_t'.$$

The asymptotic covariance matrix of the vector of t-statistics $(t_1, \ldots, t_N)'$ is given by $\Sigma_{\mathcal{T}} := (\Sigma_{\mathcal{T},ij})_{N\times N}$, where, writing $\sigma_i^2 = \mathbb{E}u_{it}^2(1 - v_t'\Sigma_f^{-1}\lambda)$,

$$\Sigma_{\mathcal{T},ij} = \frac{1}{\sigma_i\sigma_j}\mathbb{E}(1 - v_t'\Sigma_f^{-1}\lambda)^2\mathbb{E}(u_{it}u_{jt}).$$

Note that the idiosyncratic error u_{it}'s are weakly dependent in the cross-section, so that $\Sigma_{\mathcal{T},ij} \approx 0$, $i \neq j$. A large cross-sectional dimension (relative to the sample size) plays a critical role in the approximation. Otherwise, the extra term of order $O(1/N)$ in the asymptotic variance of $\widehat{\alpha}_i$ leads to cross-sectional dependence among the t-statistics and eventually invalidates the FDR control.

In the appendix, we show that both B-H procedures based on these alpha tests have desired FDR control. Our theory shows that asymptotically the new alpha-screening B-H procedure has a larger power on detecting positive alphas, and can in fact consistently identify *all* positive alphas with reasonably strong signals.

Finally, we point out that Chordia et al. (2017) adopt a bootstrap approach for the FDR control, which constructs bootstrap distributions of the alphas and their t-statistics under the null hypothesis that all alphas are zero. This amounts to imposing the null of the GRS test, which is invalid for multiple testing. As we have explained earlier, FDR control is only relevant for testing many individual alphas. Each null hypothesis should only impose a zero alpha on the fund being tested. Therefore, a valid bootstrap distribution differs fund-by-fund, imposing a zero alpha for each fund separately, while allowing other funds (not being tested) to have nonzero alphas. This is however, computationally infeasible to implement for over 3,000 funds. We thereby adopt the asymptotic alternative instead in this paper.

2.6 Dealing with Missing Data

It is not uncommon in finance applications to deal with unbalanced panels. For example, consider the problem of hedge fund performance evaluation: many hedge funds last for short periods of time, then liquidate (others, instead, simply fail to report consistently to the datasets). It is therefore important that the test we propose works in presence of missing data. In this section, we describe how to use an EM (Expectation-Maximization) algorithm for PCA robust to missing data (Stock and Watson (2002)) within our procedure.[5]

For each fund $i \leq N$, suppose the fund's return data are observable at times in $\mathcal{T}_i := \{t_{i,1}, ..., t_{i,T_i}\}$, where $T_i \leq T$ and missing observations are present when $T_i < T$. We let $F_{o,i}$ be the $K_o \times T_i$ matrix of $\{f_{o,t} : t \in \mathcal{T}_i\}$, and R_i be the $T_i \times 1$ vector of $\{r_{it} : t \in \mathcal{T}_i\}$. We run the following algorithm in the general case.

Algorithm 7 (Estimating α in Model (8) in the presence of missing data)**.**

S1. a. Run time series regressions

$$\widehat{\beta}_{o,i} = (F_{o,i}\mathbb{M}_{1_{T_i}}F'_{o,i})^{-1}(F_{o,i}\mathbb{M}_{1_{T_i}}R_i).$$

Obtain residual $Z_i = \mathbb{M}_{1_{T_i}}(R_i - F'_{o,i}\widehat{\beta}_{o,i}) = (z_{i,t} : t \in \mathcal{T}_i), \quad i = 1, ..., N.$

b. Set $k = 0$. Let $(\widehat{\beta}^0_{l,i}, \widehat{f}^0_{l,t})$ be some initial values for the latent betas and factors.

c. Define $Z^{k+1} = (z^{k+1}_{it})_{N\times T}$, where

$$z^{k+1}_{it} = \begin{cases} z_{it}, & \text{if } r_{it} \text{ is not missing} \\ \widehat{\beta}^{k\prime}_{l,i}\widehat{f}^k_{l,t}, & \text{if } r_{it} \text{ is missing} \end{cases}$$

[5] The implicit assumption is that the data is missing at random; the procedure cannot solve the problem of selective reporting by hedge funds, which we have discussed in Section 2.2.

Run step S1b in Algorithm 5 with Z^{k+1} in place of Z, obtain $\widehat{\beta}_l^{k+1}$. Let

$$\widehat{F}_l^{k+1} = \frac{1}{N}\widehat{\beta}_l^{k+1\prime} Z^{k+1}, \quad \widehat{F}_l^{k+1} = (\widehat{f}_{l,1}^{k+1}, ..., \widehat{f}_{l,T}^{k+1}).$$

Set $k = k + 1$.

d. Repeat c until convergence. Let $\widehat{F}_l$ and $\widehat{\beta}_l$ be the final estimators. The resulting $\widehat{\beta}$ is given by

$$\widehat{\beta} = (\widehat{\beta}_o, \widehat{\beta}_l).$$

S2. & S3. The same as S2 & S3 in Algorithm 3.

S4. Calculate $\mathrm{se}(\widehat{\alpha}_i)$ as in S4 in Algorithm 6, with T_i in place of T.

In essence, the EM algorithm replaces the missing values of z_{it} by $\widehat{\beta}_{l,i}^{k\prime}\widehat{f}_{l,t}^{k}$, i.e., its expected value conditional on the observed data, estimated using the parameters from the previous iteration. This allows us to iteratively estimate betas and latent factors using the PCA.

3 Simulations

In this section, we examine the finite sample performance of the asymptotic approximations developed in Appendix A.2. With respect to the data-generating process, we consider a 7-factor model for hedge fund returns, with factors calibrated to match the Fung and Hsieh factors used in our empirical study. We then calibrate the cross-sectional means and variances of betas, and idiosyncratic component of each fund, so that the summary statistics (e.g., time series R^2s, volatilities) of the simulated fund returns resemble their empirical counterparts. We adopt a benchmark model with 4 factors, so that the remaining 3 factors are omitted.

Throughout, we fix $T = 300$ and $N = 3,000$ to mimic the scenario we encounter in the empirical study. We vary the simulated cross-sectional distribution of alphas to check the performance of FDR control under various portions of true null hypotheses. More specifically, the alphas are simulated from a mixture of two Gaussians $\mathcal{N}(-2\sigma, \sigma^2)$ and $\mathcal{N}(2\sigma, \sigma^2)$, plus a point mass at zero. We vary their mixture probabilities are p_1, p_2, and $1 - p_1 - p_2$, respectively, whereas σ is calibrated to the cross-sectional standard error of the estimated alphas.

Figure 1 provides histograms of the standardized alpha estimates for the first fund based on each of the three estimators described in Algorithms 3, 4, and 5. Because the 4-factor benchmark model is misspecified, the estimator based on Algorithms 3 is inconsistent, as verified from the left panel. Algorithm 5 is designed to take into account the 3 omitted factors in the regression residual. Not surprisingly it works well. The estimator based on Algorithm 4 ignores the 4 observable factors, but it estimates a 7-latent factor model, which also corrects the omitted factor bias, its histogram thereby matches the asymptotic distribution.

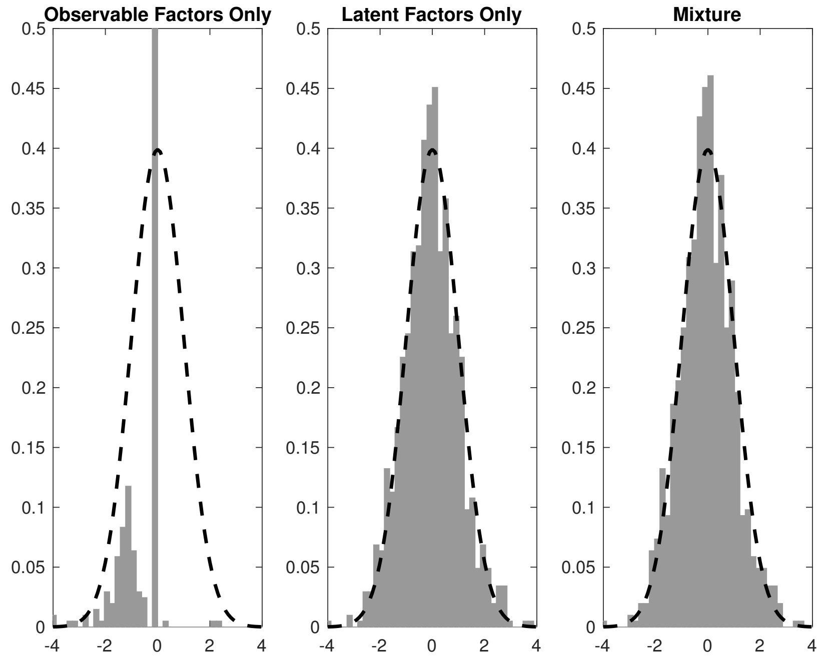

Figure 1: Histograms of the Standardized Individual Alpha Tests

Note: The figure plots the histograms of the standardized alpha estimates for one fixed fund using Algorithms 3 (observable factors only), 4 (latent factors only), and 5 (mixture), respectively. The benchmark model is a 4-factor model, whereas the true data generating process is a 7-factor model. The number of Monte Carlo repetitions is 1,000. p_1 and p_2 are fixed at 0.1.

Next, in Table 1 we compare the false discovery rates of the B-H procedure using Algorithm 5, its modified version with alpha-screening, the B-H procedures which either relies on latent factors alone (Algorithm 4), or ignores omitted factors (Algorithm 3), as well as the naive approach that only controls size of the individual tests. First of all, we find it critical to use alpha tests that take into account omitted factors, as shown from the column under "B-H + observable." Secondly, using mixed factor model is slightly better than the model based on latent factors only. This is because more information helps improve the finite sample performance, which of course only holds if the included factors are correctly specified. Third, alpha-screening is substantially less conservative than the usual B-H procedure, because the values under the third column are larger than those of the fourth column. Finally, without any B-H type control, the false discovery rate can reach as high as above 30% among the experiments we consider.

Overall, the alpha-screening B-H procedure outperforms the rest. So we adopt it along with Algorithm 5 to estimate alphas in the following-up empirical analysis.

Table 1: False Discovery Rates in Simulations

p_1	p_2	A-S B-H + Mixture	B-H + Mixture	B-H + Latent	B-H + Observable	no FDR + Mixture
0.1	0.1	5.49	4.63	8.04	23.46	31.77
0.1	0.3	3.92	3.38	5.92	12.27	10.55
0.1	0.5	2.55	2.24	3.79	6.84	4.57
0.3	0.1	4.97	3.42	7.91	18.65	25.29
0.3	0.3	3.16	2.20	4.75	8.49	7.18
0.5	0.1	4.33	2.23	5.99	13.38	18.08

Note: The table reports the false discovery rates in simulation settings with different choices of mixture probabilities p_1 and p_2, for a variety of estimators including the alpha-screening (A-S) B-H procedure based on mixed factor model (Algorithms 2 and 5), the B-H procedure using mixed factor model (Algorithms 1 and 5), latent factor model (Algorithms 1 and 4) , and observable factor model (Algorithms 1 and 3), respectively, as well as the approach that ignores FDR control despite the use of the mixed factor model for alpha estimates (Algorithm 5). The number of Monte Carlo repetitions is 1,000.

4 Empirical Analysis: Hedge Fund Alphas

4.1 Hedge Fund Returns Data

We apply our methodology to the Lipper TASS hedge funds dataset, covering the time period 1994-2016. The dataset contains a panel of returns and asset under management (AUM) for around 19,000 funds. The Lipper TASS dataset is subject to a number of potential biases. In what follows, we follow closely the bias correction procedures of Sinclair (2018), who kindly shared his code with us; these are in turn mostly based on the data-cleaning procedures detailed in Getmansky et al. (2015). We describe briefly the main concerns with the data and the bias corrections (see Sinclair (2018) for more details).

The main concern with Lipper TASS – as well as most other hedge fund data sources – is that reporting to the dataset is voluntary, which induces a selection issue in the funds that appear in the dataset. In addition, funds are able to backfill returns for periods before they start reporting to TASS, which can introduce further bias in the data. Finally, funds can also revise returns they had previously entered in the dataset.

Some of these biases can be partially addressed by using snapshots of the TASS data, which are available at irregular intervals. To correct for the backfill bias, we only use returns data for the periods after the funds start reporting to TASS. Unfortunately, as reported in Sinclair (2018), the variable that records the data in which a fund is added to TASS is mostly missing after 2011. For these funds, we use observations after the date of the first snapshot in which the fund appears as alive. In addition, we try to mitigate the problems of revisions of the returns series by always using data from the earliest snapshots available.

We further clean the data as recommended by Getmansky et al. (2015) and Sinclair (2018). In addition to what described above, we compute when possible the returns using the changes in NAV; when NAV is not available, we use reported returns. We remove funds that do not report monthly, and funds that do not report net-of-fee returns. We also remove funds that do not consistently report AUM: we require funds to report AUM at least 95% of the time. The motivation for this requirement is twofold: first, funds that strategically list their AUM in some periods but not in others are likely to also be funds that manipulate their reported returns; second, because we want to use the AUM information to focus only on large enough funds, as described below. We also remove suspicious returns: monthly returns above 200% or below -100%, stale returns (equal to the past two monthly returns), cases where AUM is reported as zero, and cases where within 6 months funds display a 10,000% increase immediately reversed, as these are likely just data entry errors. Finally, as the literature has noted (e.g., Aggarwal and Jorion (2009)), hedge fund dataset sometimes report duplicate series (for example, multiple share classes or cases in which multiple feeder funds channel capital to one investing master fund). To prevent this, we screen for cases in which two funds have return correlations of 99% or more for periods of at least 12 months (the 99% correlation cutoff was also used in Aggarwal and Jorion (2009)), and in that case we remove one of the two funds (we keep the one with the longest available time series of returns in case the two funds have data for different periods).

We impose two further constraints on the funds, again following the existing literature. First, we require funds to have reported returns and AUM to the dataset for at least 12 months (much of the literature requires a minimum of 24 or 36 months, but we prefer to err on the conservative side by including some extra funds; results are robust to using a stricter requirement). Second, we require funds to have at least $20m of AUM (as Kosowski et al. (2007) do), and drop them after they fall below this amount. This ensures that we focus our analysis on larger funds, which are also less likely to manipulate reporting to TASS; the results are robust to using different cutoffs for AUM.

Overall, we are left with 3,100 funds in our dataset. Figure 2 reports the histogram of average monthly excess returns; the average excess return is positive, at 12bp per month; the median average excess return is 28bp per month. The distribution of average excess returns is noticeably spread out, with some funds obtaining average excess returns around +8% or -8% over the sample period; the distribution appears left-skewed.

4.2 Benchmark Models

We consider several alternative benchmark models. First, we consider standard asset pricing benchmarks like the CAPM and the Fama-French 4-factor model (market, size, value and momentum factors). We also consider a well-known proposed specifically to benchmark hedge funds: the Fung and Hsieh (2004) 7-factor model, that includes market, size, a bond factor, a credit risk factor, and three trend-following factors (related to bonds, currencies, and commodities). Finally, we also consider two option-based factors (an out-of-the-money call and an out-of-the-money put factor) from Agarwal and Naik (2004). Given the discussion in the previous sections, we also use latent factors to construct the

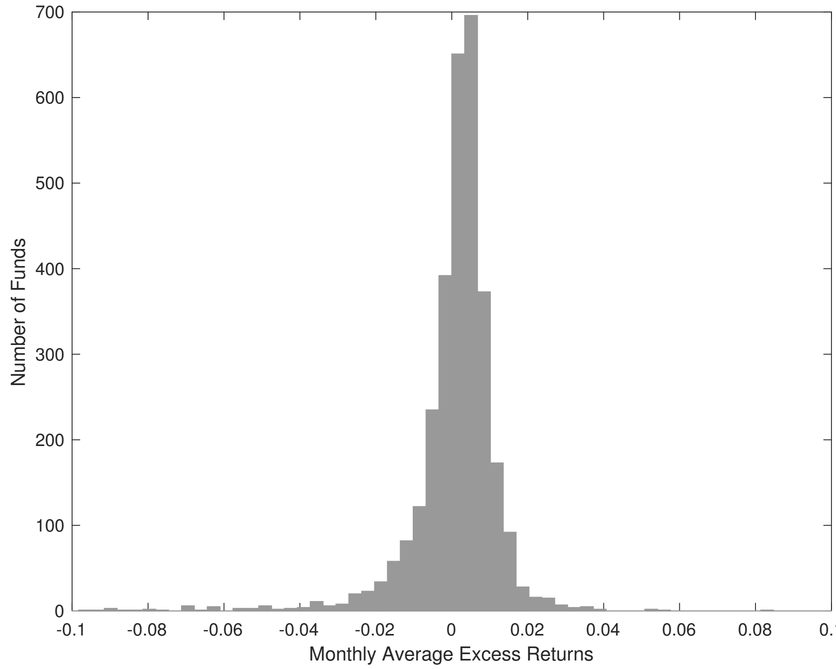

Figure 2: Histogram of Average Hedge Fund Returns

Note: In this figure, we show the histogram of average monthly returns for the 3,100 funds in our dataset over the time period 1994-2016.

benchmark, estimated from the hedge funds' principal components.

4.3 In-Sample Results

The objective of this paper is to provide a procedure to select a set of funds based on their alpha, controlling the false discovery rate of the test. The natural comparison for our procedure is the use of individual t-test on the funds' alphas, which, as discussed in the previous sections, does not keep the false discovery rate below a certain proportion (we use one-sided tests in both cases, since the objective is to find funds with positive alphas against the null that alpha is less than or equal to zero).

We therefore begin by studying what the selection procedure based on individual t-tests would imply when applied to the funds in our dataset. In this section, we use the full sample available (1994-2016); the next section, in which we do out-of-sample analysis, looks at various subsamples. Each row of the table corresponds to a different benchmark model; the first column of the table shows how many factors (observable or latent) appear in each benchmark.

The second column of Table 1 reports the number of funds (out of a total of 3,100) whose alpha appears significantly positive using multiple individual t-tests. Using the CAPM, for example, 1,342 funds (43% of the total) appear to have a statistically significantly positive alpha at the 5% confidence

Table 2: Standard multiple tests vs. FDR control test: in-sample results

	# factors	Individual test	FDR	% Excluded
CAPM	1	1342	1079	20%
FF4	4	1367	1121	18%
FH	7	1200	927	23%
FH + Option factors	9	1041	731	30%
All observable	11	1112	817	27%
5 Latent factors	5	1296	1043	20%
10 Latent factors	10	1137	835	27%
All observable + 5 Latent	16	917	602	34%

Note: The table reports the results of the multiple alpha tests for the 3,100 hedge funds in our sample, using different methodologies and benchmark models. Each row corresponds to a different benchmark against which funds are evaluated. The first column reports the total number of factors in the benchmark model. The second column shows how many funds have a significant alpha using standard significance levels for each individual test. The third column reports how many alphas are significant according to our FDR control test. The last column shows what fraction of the tests appear significant using the standard significance levels, but are deemed not significant by the FDR control test. The benchmarks could include only observable factors, only latent factors, or a mix of the two. FF4 includes Market, SMB, HML and Momentum. FH (Fung and Hsieh) is the 7-factor model proposed by Fung and Hsieh (2004). The Option factors are OTM call and put factors from Agarwal and Naik (2004). Sample periods is 1994-2016.

level. The distribution of t-stats from individual tests is broadly in line with that reported in the literature (for example, in Kosowski et al. (2007)). This number generally decreases as more factors are added to the benchmark – either observable or latent – since exposure to these factors accounts for part of the estimated average excess return.

Among the models we consider in the table, some include latent factors, either exclusively or together with some observable factors. The estimation of the latent factors and their risk premia proceeds as described in the previous sections. From an econometric point of view, it is important to use latent factors both to avoid the possibility of omitted priced factors (following the same intuition as in Giglio and Xiu (2017) and to ensure that the residuals from the model have low correlation (which is an assumption needed for all the standard multiple tests to work correctly).

To get a sense of the factor structure of hedge fund returns, Figure 3 shows the first 15 eigenvalues of the excess returns. There clearly are important common components driving hedge fund returns; the pattern of eigenvalues suggests the presence of 5 to 10 latent factors. We choose these two numbers when constructing our latent-factors-based benchmarks in Table 2. We also plot the eigenvalues of the residual covariance matrices of a variety of benchmark models, including the CAPM, the FH7, and the model with all 11 observable factors. It is evident that observable factors indeed help capture certain common variations in the cross-section, because the largest few eigenvalues shrink substantially. The largest gain comes from the market factor, which shrinks the largest eigenvalue by about a half. The marginal contribution by the remaining 10 factors is less significant. Moreover, there remains moderate common variation in the residuals of the 11-factor model. Not surprisingly, with as few as 5 latent factors eliminated, the residuals from the latent factor model are less correlated than those of the 11-factor model.

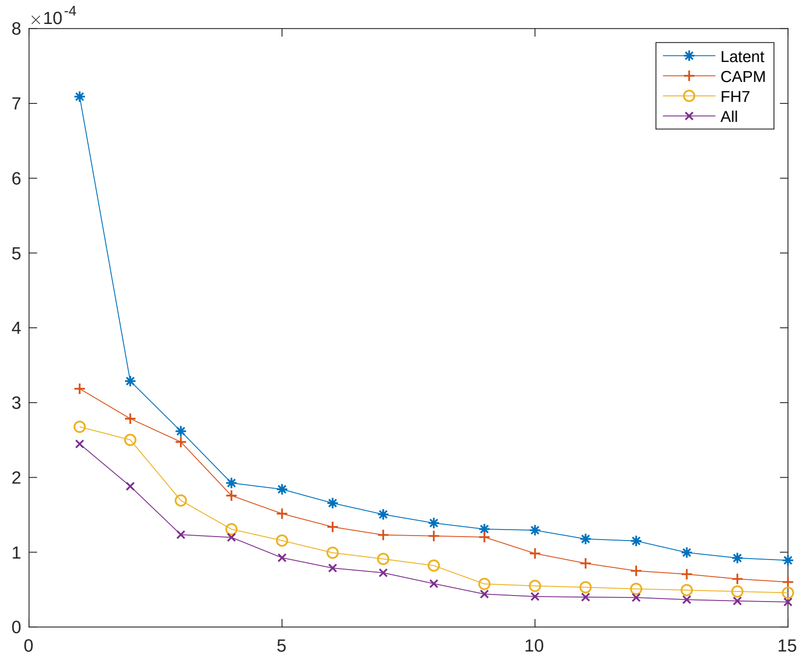

Figure 3: Scree Plot of Eigenvalues

Note: The figure reports the first 15 eigenvalues of the covariance matrix of excess returns, denoted as "Latent", for the 3,100 hedge funds in our panel, sorted from highest to lowest, as well as eigenvalues of the residual covariance matrices relative to benchmark models such as "CAPM", "FH7", and the model "All" with all 11 observed factors. Sample period covers 1994-2016.

The third column of the table reports the number of funds selected using the FDR control procedure (with the false discovery rate set at 5%). Comparing columns 2 and 3 of the table, we can see that, as expected, the FDR control results in a smaller number of significant funds compared to the multiple individual testing. The last column shows the fraction of funds that are selected by individual tests but are excluded by the FDR control. This is a substantial amount, of about one quarter to one third of the funds. In-sample, therefore, the FDR control test is substantially more selective than the multiple individual test.

4.4 Out-of-Sample Analysis

We now turn to the study of the out-of-sample performance of the two procedures. We proceed as follows. Consider any time period t: using all information available up to t, we construct the t-statistics for the alphas of all funds alive at time t, and select funds based on the multiple individual test and on our FDR control test. We compute alphas relative to two benchmarks: one that uses the standard Fung and Hsieh 7-factor model, and one that uses 5 latent factors. Given that one of the benchmark models we include has 7 factors, we select funds among those that have at least 7 months of available historical data up to time t.

We then look at how many funds survive for at least 7 months after t, among those that are selected at t. For those that do survive at least 7 months, we compute the alphas and t-statistics between $t + 1$ and the end of the sample. We then report both the short-term closure rate (i.e., how many funds disappear within the first 6 months after t) as well as the performance of those that survive at least 7 months.[6] Finally, we repeat this exercise for different cut-off points t, and show all of them for robustness.

Table 3 reports the results of the out-of-sample analysis. The Table has a top and a bottom part. The top part uses the Fung and Hsieh 7-factor model as benchmark, the bottom part uses the 5-latent-factors model. Results are very similar under the two benchmarks, so here we focus the description to the top part.

Each row of the table corresponds to a different choice for the cutoff point t, reported in the first column. For example, the first cutoff is at the end of 1999 (we leave a few years before the first cutoff and after the last cutoff so that we have enough data to estimate the alphas and evaluate the funds). The second column shows how many funds were alive at that point in time (and have at least 7 months of available historical data at that point). The number of available funds increased over time up to the financial crisis, then dropped in the following years.

The rest of the table has 5 panels, each divided in three columns. The first column in each panel refers to the funds selected by the FDR control test. The second panel refers to the funds that were *not* selected by the FDR control test, but that were selected by the individual t-test. Finally, the third column refers to all other available funds, that were not selected by either procedure. This way, the total funds alive at each time t is divided in three non-overlapping groups.

The first panel shows the number of funds in each of the three groups. For example, at the end of 1999, out of a total of 166 funds, 60 are selected by the FDR control test, 17 are excluded by the FDR control but are selected by the individual t-stat test, and 89 funds are not selected by any of the two procedures. The second panel reports the total AUM in each group, in \$bn. Interestingly, up to the financial crisis, the AUM of funds in which the FDR control test grows dramatically, to around \$200bn, and then drops as the hedge fund industry shrank. The AUM of funds excluded by the FDR but included by the standard individual t-test is generally small, but it increases dramatically in relative terms after the financial crisis.

The third panel shows the fraction of funds that close less than 7 months after the cutoff date. The number is generally small, but it rises significantly around the financial crisis. Note that the rate

[6] Again, the choice of looking at funds surviving at least 7 months is due to the fact that the benchmark models have up to 7 factors.

Table 3: Out of sample Results

Model: FH7		Panel 1: # Selected			Panel 2: AUM ($bn)			Panel 3: Closure rate			Panel 4: α (bp/month)			Panel 5: T-stats		
Cutoff	# Funds	FDR	Indiv.	None	FDR	Indiv.	None	FDR	Indiv.	None	FDR	Indiv.	None	FDR	Indiv.	None
200112	254	107	18	129	25.3	1.9	16.0	0.02	0.00	0.08	30	12	25	1.82	0.17	0.89
200212	488	125	46	317	33.8	8.0	47.8	0.06	0.15	0.08	47	36	40	3.60	1.66	1.52
200312	560	272	40	248	91.5	8.1	39.0	0.06	0.08	0.09	41	58	31	2.73	1.88	1.35
200412	661	314	41	306	139.4	11.6	67.2	0.08	0.20	0.15	43	61	37	3.24	1.64	1.45
200512	844	452	55	337	202.1	10.5	60.8	0.07	0.13	0.09	49	31	51	2.92	1.40	1.70
200612	966	555	60	351	254.7	11.2	74.8	0.08	0.10	0.15	38	64	49	1.89	1.57	1.13
200712	1020	518	72	430	249.5	13.9	109.3	0.16	0.21	0.23	30	9	-6	1.89	0.78	0.56
200812	1069	137	142	790	62.0	51.5	144.4	0.12	0.15	0.20	31	30	28	2.08	1.04	0.86
200912	1001	224	110	667	83.9	25.1	106.7	0.05	0.08	0.19	21	14	12	1.31	0.89	0.61
201012	810	235	72	503	78.5	18.0	93.0	0.07	0.10	0.17	33	18	6	1.87	1.03	0.89
201112	909	187	102	620	68.7	17.7	104.4	0.09	0.09	0.15	29	13	12	1.65	0.98	0.75
201212	753	227	83	443	87.9	17.6	60.3	0.07	0.06	0.18	34	35	15	1.54	1.11	0.52
Model: 5 latent																
200112	254	123	11	120	29.0	1.0	13.3	0.02	0.00	0.08	30	25	37	2.21	1.65	1.37
200212	488	108	55	325	30.0	10.1	51.4	0.07	0.11	0.07	40	25	43	3.41	1.56	1.72
200312	560	322	30	208	97.3	7.3	39.1	0.06	0.10	0.09	40	38	27	2.63	1.55	1.45
200412	661	314	44	303	138.9	12.9	72.0	0.09	0.11	0.16	38	31	26	2.97	1.27	1.05
200512	844	478	42	324	208.4	12.0	53.6	0.06	0.07	0.11	39	49	33	2.44	1.23	1.24
200612	966	542	49	375	250.3	14.8	78.1	0.07	0.08	0.15	51	33	57	2.55	1.27	1.40
200712	1020	355	147	518	172.7	37.9	108.3	0.10	0.20	0.25	57	80	22	2.01	1.61	0.82
200812	1069	193	128	748	93.5	20.3	117.0	0.12	0.13	0.21	40	49	39	2.27	1.35	1.45
200912	1001	174	104	723	70.2	33.1	113.6	0.06	0.11	0.17	30	33	14	1.96	1.35	0.96
201012	810	179	94	537	79.4	20.0	90.9	0.07	0.09	0.16	52	45	19	2.70	2.07	1.19
201112	909	186	95	628	76.8	21.0	88.8	0.08	0.07	0.16	50	47	29	2.49	1.84	1.07
201212	753	266	61	426	92.4	11.8	58.0	0.07	0.05	0.19	37	43	21	1.46	0.91	0.64

Note: The table reports out of sample performance of the funds selected using different methodologies. The top part of the table uses the Fung and Hsieh (2004) model as benchmark; the bottom part uses a model with 5 latent factors as benchmark. Each row in the table corresponds to a different cutoff point t for the out-of sample analysis: for each t, the data before t is used to estimate the alphas and select the funds, and the data after t is used to evaluate the performance. The first column reports how many funds were active at the time t when the selection occurs. The rest of the table has 5 panels. Each panel is divided in three columns. The first column refers to the funds that are selected at t by the FDR control test. The second panel refers to the funds selected at t by the standard multiple individual tests, but not by the FDR control test. The third column refers to funds that were not selected by either method. Panel 1 reports the number of funds in each group. Panel 2 reports the total AUM of the funds in each group. Panel 3 reports the fraction of funds in each group that close within 6 months after t. Panel 4 reports the average out-of-sample alpha for those funds in each group that survive more than 6 months after t. Panel 5 reports the corresponding average out-of-sample t-stat. Sample period is 1994-2016.

of closures is substantially lower for funds selected by FDR than for funds excluded by it. On average, 7% of funds selected by FDR close within the next 6 months, as do 12% of those selected by the individual t-test but excluded by FDR, and 14% of those not selected by either test. This is a first important sign that the FDR control is able to select better funds than standard individual t-tests.

The fourth and fifth panel of Table 3 report the average out-of-sample alphas and t-stats of those funds that survive at least 7 months after each cutoff. The results here are strongly in favor of the FDR control procedure. In 10 out of 14 choices of the cutoff, the average out-of-sample alpha of the FDR-selected funds is larger than the alpha of those excluded but selected by the standard test; the average t-stat is actually larger in all 14 cases period. So the FDR-selected funds perform out-of-sample, in the almost totality of cases, better than all remaining funds (both those that are selected by the individual tests and those not selected).

Altogether, these results indicate that the FDR control test is able to select funds that are much less likely to close in the short term, and, among those that survive, much better performing out of sample than the ones it excludes.

5 Conclusion

This paper presents a rigorous framework to address the data-snooping concerns that arise when applying multiple testing in the asset pricing context. In situations in which many tests are performed, many "false discoveries" should be expected: cases in which the significance of some of the tests is obtained by pure chance. The rate of false discoveries for a multiple testing procedure is hard to evaluate ex-ante; and it can grow unboundedly when standard statistical tests are used as the number of tests performed increases.

Statistical theory has proposed different methods that aim to control and mitigate this data-snooping problem, like the so-called "false discovery rate" (FDR) control test of Benjamini and Hochberg (1995). But these methods do not work in the standard asset pricing context, in which some of the main assumptions for the procedures are violated. In the paper, we show that the FDR control test can be extended and generalized to be valid under a much broader range of assumptions, specifically those that appear crucial when thinking about testing for alphas in linear factor models.

Our paper exploits the "blessing of dimensionality" to build a FDR control test that is valid when the benchmark includes non-tradable factors whose risk-premia need to be estimated, and is robust to the presence of omitted priced factors from the benchmark and to correlation of excess returns in the cross-section and in the time series. In addition, contrary to existing multiple-testing methods, our test is built explicitly to handle large cross-sections; this makes it particularly suitable for many finance applications, in which the size of the cross-section N can be large relative to the sample size T.

We illustrate this procedure by applying it to the evaluation of hedge fund performance, a typical example where multiple testing issues arise. We show empirically that hedge fund returns are highly correlated in the cross-section, even after controlling for the standard models. We show that our procedure – which allows for such correlation and bounds the false discovery rate to 5% – selects a significantly smaller number of funds based on their alpha compared to a standard set of individual t-tests on the fund alphas, and produces superior out of sample results.

There is a burgeon strand of literature on the applications of machine learning techniques to high dimensional problems in asset pricing, in which data snooping leads to potentially numerous false discoveries. Our paper serves as an example that rigorously accounts for the data snooping bias. There remain many other settings in which our high-dimensional multiple-testing framework can be applied: for example, the evaluation of multiple potential new factors against an existing asset pricing model. We leave the study of these applications to future research.

References

Ackermann, C., R. McEnally, and D. Ravenscraft (1999). The performance of hedge funds: Risk, return, and incentives. *The Journal of Finance 54*(3), 833–874.

Agarwal, V., G. Bakshi, and J. Huij (2009). Do higher-moment equity risks explain hedge fund returns? Technical report, Georgia State University.

Agarwal, V., V. Fos, and W. Jiang (2013). Inferring reporting-related biases in hedge fund databases from hedge fund equity holdings. *Management Science 59*(6), 1271–1289.

Agarwal, V. and N. Y. Naik (2000). Multi-period performance persistence analysis of hedge funds. *Journal of financial and quantitative analysis 35*(3), 327–342.

Agarwal, V. and N. Y. Naik (2004). Risks and portfolio decisions involving hedge funds. *The Review of Financial Studies 17*(1), 63–98.

Aggarwal, R. K. and P. Jorion (2009). The risk of emerging hedge fund managers. *The Journal of Investing 18*(1), 100–107.

Aggarwal, R. K. and P. Jorion (2010). The performance of emerging hedge funds and managers. *Journal of Financial Economics 96*(2), 238–256.

Ahn, S. and A. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*, 1203–1227.

Ang, A. and D. Kristensen (2012). Testing conditional factor models. *Journal of Financial Economics 106*(1), 132–156.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*, 135–171.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

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.

Bali, T. G., S. J. Brown, and M. O. Caglayan (2011). Do hedge funds' exposures to risk factors predict their future returns? *Journal of Financial Economics 101*(1), 36–68.

Bali, T. G., S. J. Brown, and M. O. Caglayan (2014). Macroeconomic risk and hedge fund returns. *Journal of Financial Economics 114*(1), 1–19.

Baquero, G., J. Ter Horst, and M. Verbeek (2005). Survival, look-ahead bias, and persistence in hedge fund performance. *Journal of Financial and Quantitative analysis 40*(3), 493–517.

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.

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.

Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics 29*(4), 1165–1188.

Berk, J. B. and J. H. Van Binsbergen (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics 118*(1), 1–20.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013a). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics 41*(6), 2786–2819.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013b). Testing many moment inequalities. Technical report, MIT.

Chordia, T., A. Goyal, and A. Saretto (2017). p-hacking: Evidence from two million trading strategies. Technical report, Emory University.

Cochrane, J. H. (2009). *Asset pricing: Revised edition*. Princeton university press.

Donoho, D. L. et al. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture*, 1–32.

Fan, J. and X. Han (2016). Estimation of the false discovery proportion with unknown dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*(4), 1143–1164.

Fan, J., X. Han, and W. Gu (2012). Estimating False Discovery Proportion Under Arbitrary Covariance Dependence. *Journal of the American Statistical Association 107*(499), 1019–1035.

Fan, J., Y. Ke, and Y. Liao (2016). Robust factor models with explanatory proxies. Technical report, Princeton University.

Feng, G., S. Giglio, and D. Xiu (2017). Taming the factor zoo. Technical report, University of Chicago.

Freyberger, J., A. Neuhierl, and M. Weber (2017). Dissecting characteristics nonparametrically. Technical report, University of Wisconsin-Madison.

Fung, W. and D. A. Hsieh (1997). Empirical characteristics of dynamic trading strategies: The case of hedge funds. *The Review of Financial Studies 10*(2), 275–302.

Fung, W. and D. A. Hsieh (2001). The risk in hedge fund strategies: Theory and evidence from trend followers. *The Review of Financial Studies 14*(2), 313–341.

Fung, W. and D. A. Hsieh (2004). Hedge fund benchmarks: A risk-based approach. *Financial Analysts Journal*, 65–80.

Fung, W., D. A. Hsieh, N. Y. Naik, and T. Ramadorai (2008). Hedge funds: Performance, risk, and capital formation. *The Journal of Finance 63*(4), 1777–1803.

Getmansky, M., P. A. Lee, and A. W. Lo (2015). Hedge funds: A dynamic industry in transition. *Annual Review of Financial Economics 7*, 483–577.

Giglio, S. and D. Xiu (2017). Asset pricing with omitted factors. Technical report, Yale University and University of Chicago.

Gu, S., B. Kelly, and D. Xiu (2018). Empirical asset pricing via machine learning. Technical report, University of Chicago.

Hansen, P. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics 23*, 365–380.

Harvey, C. R. and Y. Liu (2016). Lucky factors. Technical report, Duke University.

Harvey, C. R. and Y. Liu (2018). False (and missed) discoveries in financial economics. Technical report, Duke University.

Harvey, C. R., Y. Liu, and H. Zhu (2015). ... and the cross-section of expected returns. *Review of Financial Studies 29*(1), 5–68.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics 6*(2), 65–70.

Huber, P. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics 35*(1), 73–101.

Jagannathan, R., A. Malakhov, and D. Novikov (2010). Do hot hands exist among hedge fund managers? an empirical evaluation. *The Journal of Finance 65*(1), 217–255.

Kelly, B., S. Pruitt, and Y. Su (2017). Some characteristics are risk exposures, and the rest are irrelevant. Technical report, University of Chicago.

Kosowski, R., N. Y. Naik, and M. Teo (2007). Do hedge funds deliver alpha? a bayesian and bootstrap analysis. *Journal of Financial Economics 84*(1), 229–264.

Kozak, S., S. Nagel, and S. Santosh (2017). Shrinking the cross section. Technical report, University of Michigan.

Leek, J. T. and J. D. Storey (2008). A general framework for multiple testing dependence. *Proceedings of the National Academy of Sciences 105*(48), 18718–18723.

Liang, B. (1999). On the performance of hedge funds. *Financial Analysts Journal*, 72–85.

Liang, B. (2001). Hedge fund performance: 1990-1999. *Financial Analysts Journal*, 11–18.

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

Lo, A. W. and A. C. MacKinlay (1990). Data-snooping biases in tests of financial asset pricing models. *Review of financial studies 3*(3), 431–467.

Mitchell, M. and T. Pulvino (2001). Characteristics of risk and return in risk arbitrage. *the Journal of Finance 56*(6), 2135–2175.

Patton, A. J. and T. Ramadorai (2013). On the high-frequency dynamics of hedge fund risk exposures. *The Journal of Finance 68*(2), 597–635.

Romano, J. P. and M. Wolf (2005, July). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

Romano, J. P. and M. Wolf (2018). Multiple testing of one-sided hypotheses: combining bonferroni and the bootstrap. In *International Conference of the Thailand Econometrics Society*, pp. 78–94. Springer.

Schwartzman, A. and X. Lin (2011). The effect of correlation in false discovery rate estimation. *Biometrika 98*(1), 199–214.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika 73*(3), 751–754.

Sinclair, A. J. (2018). The allocative role of prime brokers. Technical report, The University of Hong Kong.

Song, Y. and Q. Zhao (2018). Performance evaluation with latent factors. Technical report, University of Pennsylvania.

Stock, J. and M. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*(2), 147–162.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(3), 479–498.

Storey, J. D., J. E. Taylor, and D. Siegmund (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66*(1), 187–205.

Sullivan, R., A. Timmermann, and H. White (1999). Data-snooping, technical trading rule performance, and the bootstrap. *The Journal of Finance 54*(5), 1647–1691.

White, H. (2000). A reality check for data snooping. *Econometrica 68*(5), 1097–1126.

Appendix

A Asymptotic Theory

We present the formal asymptotic theory below. We define the following notation. Let $A = (a_{ij})$ be an $n \times m$ matrix, and $\psi_1(A) \geq ... \geq \psi_K(A)$ denote the first K ordered singular values of a matrix A if $K \leq \min\{m, n\}$. Let $\|A\| = \sqrt{tr(A'A)}$, which is also known as the "Frobenius norm" for A. In particular, if A is a vector, then $\|A\|$ equals its Euclidean norm. In addition, if $n \to \infty$ and m fixed constant, we write $\|A\|_\infty = \max_{i \leq n} \sum_{j=1}^m |a_{ij}|$. Write $\mathbb{M}_{1_N} = \mathrm{I}_N - 1_N 1_N'/N$ with $1_N = (1, ..., 1)'$ be the $N \times 1$ vector of ones.

A.1 Technical Assumptions

We start by listing and discussing the technical assumptions used for our asympttotic theory. We assume $(f_t, u_t, \alpha_i : i \leq N, t \leq T)$ are stochastic.

Assumption 1. *There are constants $c, C > 0$, so that:*

(i) (pervasiveness)

$$c < \psi_K(\frac{1}{N}\beta_l'\beta_l) \leq ... \leq \psi_1(\frac{1}{N}\beta_l'\beta_l) < C.$$

(i) (idiosyncrasy)

$$\psi_1(\mathrm{Cov}(u_t)) < C.$$

Assumption 1 is adopted from Stock and Watson (2002) and many other works on estimating latent factors. This assumption ensures that the factors are asymptotically identified (up to a rotation) and that $\mathrm{Cov}(r_t)$ has K growing eigenvalues whose rate is $O(N)$, while its remaining $N - K$ eigenvalues do not grow with the dimensionality. This distinguishing behavior among the eigenvalues provides us the intuitions of estimating the betas for latent factors using PCA.

Assumption 2. *The following statements hold:*

(i) $\{f_t, u_t : t \leq T\}$ are independent and identically distributed, and $\mathbb{E}(u_t|f_1, ..., f_T) = 0$.

(ii) $\{\alpha_i : i \leq N\}$ are mutually independent, and also independent of $\{f_t, u_t : t \leq T\}$.

(iii) Weak cross-sectional dependence: There is a constant $C > 0$ so that almost surely , $\max_{j \leq N} \sum_{i=1}^N |\mathbb{E}(u_{it}u_{jt}|f_1, ..., f_T)| < C$, and $\max_{j \leq N} \sum_{i=1}^N 1\{|\mathbb{E}(u_{it}u_{jt}|f_1, ..., f_T)| \geq (\log N)^{-3}\} \leq CN^c$ for some $c > 0$ and $\max_{i,j \leq N} \sum_{k=1}^N |\mathrm{Cov}(u_{it}u_{kt}, u_{jt}u_{kt})| < C$.

Assumption 2 regulates the dependence structures of the data generating process. We consider serially independent data. We maintain the serial independence to keep the technical tools relatively simple. Allowing for serially weakly dependent data is possible, by imposing extra mixing conditions for the time series. Condition (iii) requires cross-sectional weak correlations among the idiosyncratic components u_{it}. The intuition of the cross-sectional weak correlations is that the idiosyncratic components should capture the remaining shocks and possible local factors after conditioning on the common risk factors. Technically, as the cross-sectiona dimension increases, we show that the correlations among the constructed t-statistics based on the estimated alphas are mainly driven by those of u_{it}. Hence this condition ensures the t-statistics are weakly correlated, making a valid procedure for the FDR control.

Assumption 3 (Moment bounds). *There are $C > c > 0$, such that*

(i) $\mathbb{E}\|f_t\|^4 + \max_{i\leq N} \mathbb{E}u_{it}^8 < C$.

(ii) For any $k, l \leq \dim(f_t)$, we have

$$\frac{\mathbb{E}\max_{i,j,d\leq N,t\leq T} \xi^4_{i,j,d,k,l,t}}{\max_{i,j,d\leq N,t\leq T} \mathbb{E}\xi^4_{i,j,d,k,l,t}} \leq (\log N)^2 TC$$

where $\xi_{i,j,d,k,l,t} \in \{u_{it}u_{jt}, u_{it}, u_{it}w_t, u_{it}^2 w_{kt}^2, u_{it}f_{kt}, u_{it}^2 f_{kt}, u_{it}^2 f_{kt} f_{lt}, u_{it}^2 u_{jt} u_{dt}\}$ *and for* $w_t = \frac{1}{\sqrt{N}}\beta' u_t$.

(iii) There is $0 < L < 1$, and a sequence $B_{NT} > c$ satisfying $B_{NT}^2 \log(NT)^7 \leq T^L$, ($B_{NT}$ may either diverge or not), such that

$$\mathbb{E} \max_{i\leq N,t\leq T} u_{it}^4 + \mathbb{E} \max_{i\leq N,t\leq T} u_{it}^4 \|f_t\|^4 < B_{NT}^4.$$

(iv) $\|\Sigma_f^{-1}\| < C$ *and* $\min_{i\leq N} \mathbb{E}u_{it}^2(1 - v_t'\Sigma_f^{-1}\lambda)^2 > c$, $\mathbb{E}\|\frac{1}{\sqrt{N}}\beta' u_t\|^4 < C$

(v) All eigenvalues of $\frac{1}{N}\sum_{j=1}^N (\beta_j - \bar{\beta})(\beta_j - \bar{\beta})'$ *belong to* $[c, C]$.

Conditions (ii) require that interchanging "max" with "$\mathbb{E}$" on $\xi_{i,j,k,l,t}$ increases a quantity no more than $O(T \log^2 N)$. Technically, it is a required condition to apply concentration inequalities from Chernozhukov et al. (2013b) for establishing

$$\max_{i,j,d\leq N} |\frac{1}{T}\sum_t \xi_{i,j,d,k,l,t} - \mathbb{E}\xi_{i,j,d,k,l,t}| = O_P(\sqrt{\frac{\log N}{T}}),$$

which is a key step to bound $\max_{i\leq N} |\widehat{\alpha}_i - \alpha_i|$. In addition, assuming $\mathbb{E}\|\frac{1}{\sqrt{N}}\beta' u_t\|^4 < C$ in (iv) is not stringent given the cross-sectional weak correlations among u_{it}. Above all, the above conditions allow the underlying distribution of the data generating process to have heavier tails than those of the sub-Gaussian distributions, in the sense that in our context it is sufficient to have finitely many moments. On the other hand, we note that in the recent literature on factor models, e.g., Fan et al. (2016); Song and Zhao (2018), researchers apply more robust estimators based on Huber's loss function (Huber, 1964) that allow more general tail distributions.[7] In addition, Song and Zhao (2018) do not allow non-tradable observed factors and require the alphas be "sparse" in the sense that many components

[7] Song and Zhao (2018) apply the maximum likelihood estimation to extract the latent factor space, which is in fact more sensitive to outliers than PCA. A completely robust procedure should be based on Huber's estimation in all steps of the algorithms: including both time series and cross-sectional regressions and the PCA steps. We shall leave it for future research.

should be nearly zero. However, unlike stock returns, this is not the case for hedge funds; our empirical studies indicate the presence of many nonzero alphas. In contrast, we do not require such a sparse structure, and allow arbitrary cross-sectional structures of the true alphas.

Assumption 4. *Growing number of positive alphas: there is a growing sequence $L_{NT} \to \infty$, for the true α,*

$$\sum_{i=1}^{N} 1\{\alpha_i \geq L_{NT}\sqrt{\frac{\log N}{T}}\} \to \infty.$$

Assumption 4 requires there should be growing amount of true alternatives, so that the number of true null hypotheses is not "overwhelmingly many". This is needed to control the rate of false rejections and the same assumption is assumed in Liu and Shao (2014).

A.2 Main Theoretical Results

We now present the asymptotic distributions for estimated alphas. Theorems 1 and 2 apply to estimators that are obtained in any of the four factor scenarios: (i) observable factors only (Algorithm 3), (i) latent factors only (Algorithm 4), (iii) the general case (Algorithm 5), and (iv) mix of observable and latent factors with additional condition that observable factors are tradable (Algorithm 8).

Theorem 1. *Suppose $T, N \to \infty$, $(\log N)^c = o(T)$, for some $c > 7$ and Assumptions 1-3 hold. Then for any $i \leq N$,*

$$\sigma_{i,NT}^{-1}(\widehat{\alpha}_i - \alpha_i) \xrightarrow{d} \mathcal{N}(0, 1)$$

where $\sigma_{i,NT}^2 = \frac{1}{T}\operatorname{Var}(u_{it}(1 - v_t\Sigma_f^{-2}\lambda)) + \frac{1}{N}\chi_i$. In scenarios (i)-(iii), $\chi_i = \frac{1}{N}\operatorname{Var}(\alpha_i)\beta_i' S_\beta^{-1}\beta_i$; in scenario (iv) that observable factors are tradable, $\chi_i = \frac{1}{N}\operatorname{Var}(\alpha_i)\beta_{l,i}' S_{\beta,l}^{-1}\beta_{l,i}$.

Theorem 1 arises from a more general joint asymptotic expansion for the $N \times 1$ vector $\widehat{\alpha}$, given in Proposition 2:

$$\widehat{\alpha} - \alpha \approx \frac{1}{T}\sum_t u_t(1 - v_t'\Sigma_f^{-1}\lambda) - \beta\eta_N, \quad \eta_N := \frac{1}{N}S_\beta^{-1}\beta'\mathbb{M}_{1_N}\alpha.$$

For each element of the above expansion, the first term is $O_P(T^{-1/2})$ while the second term is $O_P(N^{-1/2})$ but depends on a common component η_N. The presence of the second term is the key reason of inconsistency of the low dimension setting, but vanishes as $N \to \infty$. However, if N grows slowly, it is first-order not negligible, whose presence could bring strong cross-sectional correlations among the estimated alphas due to the common component η_N and would adversely affect the FDR control. Thus we require $T \log N = O(N)$ to make η_N be negligible so that the asymptotic distribution is only determined by $\frac{1}{T}\sum_t u_{it}(1 - v_t'\Sigma_f^{-1}\lambda)$, and is therefore weakly cross-sectionally correlated. Assuming $T \log N = O(N)$ makes our main results work in effect under a very high dimension, and is satisfied as the number of funds could easily reach over three thousands with monthly data of less than three hundred.

The theorem below gives the main results for the FDR control procedure. Also note that the alpha-screening method focuses only on $\widehat{\mathcal{I}} = \left\{ i \leq N : \widehat{\alpha}_i > -\operatorname{se}(\widehat{\alpha}_i)\sqrt{\log(\log N)} \right\}$, which we call "screening B-H".

Theorem 2. *In addition to conditions in Theorem 1, suppose $T(\log N) = o(N)$, and Assumption 4 hold. Then both the B-H procedure (B-H) and alpha-screening B-H procedure (screening B-H) satisfy:*

(i)

$$\limsup_{N,T\to\infty} \mathrm{FDR} \leq \tau.$$

(ii)

$$\mathbb{P}(\mathbb{H}_0^i \textit{ is corrected rejected, for all } i \in \mathcal{H}) \to 1.$$

In addition, as for the screening B-H procedure, we have:

(iii) Define events:

$$\begin{aligned} A_{B\text{-}H} &= \{\textit{all false } \mathbb{H}_0^i \textit{ are correctly rejected by B-H}\} \\ A_{screening\ B\text{-}H} &= \{\textit{all false } \mathbb{H}_0^i \textit{ are correctly rejected by screening B-H}\}. \end{aligned}$$

Then as long as $\tau < \frac{1}{2}$, asymptotically, we have

$$\mathbb{P}(A_{screening\ B\text{-}H}) \geq \mathbb{P}(A_{B\text{-}H}).$$

(iv) Recall that $\widehat{\mathcal{I}} = \{i \leq N : t_i > -\sqrt{\log(\log N)}\}$, we have

$$\mathbb{P}(\mathbb{H}_0^i : \alpha_i \leq 0 \textit{ is true for all } i \notin \widehat{\mathcal{I}}) \to 1.$$

Note that the usual B-H procedure requires the t-statistics be computed based on the "sample average" and its standard errors (Liu and Shao, 2014), while in this context, $\widehat{\alpha}_i$ is only approximated the sample average: $\sqrt{T}(\widehat{\alpha}_i - \alpha_i) = \frac{1}{\sqrt{T}}\sum_t u_{it}(1 - v_t'\Sigma_f\lambda)\sigma_i^{-1} + \Delta_i$ where $\max_{i\leq N}|\Delta_i| = o_P(1/\sqrt{\log N})$ when $T\log N = O(N)$. This theorem shows that the additional approximation error does not affect the "size" asymptotically. In addition, as for the "power" property for detecting the significant alphas, note that Assumption 4 ensures that for the true vector of α, there is a set $\mathcal{H} \subset \{1, ..., N\}$ so that

$$\mathcal{H} := \{i \leq N : \alpha_i \geq L_{NT}\sqrt{\frac{\log N}{T}}\}$$

and $|\mathcal{H}| \to \infty$. Apparently $\mathbb{H}_0^i$ is false for all $i \in \mathcal{H}$. Theorem 2 shows that all false hypotheses indexed in $\mathcal{H}$ can be correctedly rejected asymptotically, that is, we can correctly detect all positive alphas whose magnitudes are larger than $\sqrt{\frac{\log N}{T}}$. In addition, to compare the power of the regular B-H procedure and the B-H with alpha-screening, note that the latter focuses only on $\widehat{\mathcal{I}} = \left\{ i \leq N : \widehat{\alpha}_i > -\operatorname{se}(\widehat{\alpha}_i)\sqrt{\log(\log N)} \right\}$ which successfully screens out alphas that are all "deep in the null" (meaning that they are "overwhelmingly negative"). As we explained in the main text, the intuition of using the alpha screening step is that when many of the true alphas are deep in the null, including them when conducting multiple testing could make the B-H overly conservative, this is because the B-H compared the i th largest p-value $p_{(i)}$ with $i\tau/N$ where N is the number of hypotheses to be tested and can be very large when all hypotheses are being considered. So the power of the B-H procedure adversely depends on the number of candidate hypotheses. In sharp contrast, the alpha screening focuses on a potentially much smaller subset $\widehat{\mathcal{I}}$, and thus gains additional power. The same idea based on screening was previously used in the literature to boost the power for reality check by Hansen (2005); Chernozhukov et al. (2013b), and was also applied to multiple testing for FWER control by Romano and Wolf (2018). Finally, we should not worry that skilled funds might also be screened off because (iv) shows that $\alpha_i \leq 0$ for all $i \notin \widehat{\mathcal{I}}$.

To summarize the results in Theorem 2, asymptotically

(i) The FDR rate can be controlled under the pre-determined level $\tau \in (0, 1)$.

(ii) Our procedure can correctly identify all the alphas satisfying

$$\alpha_i \geq L_{NT}\sqrt{\frac{\log N}{T}}$$

for sequence $L_{NT} \to \infty$ that grows arbitrarily slowly.

(iii) The detecting power of the alpha-screening is larger than or equal to that of the regular B-H.

(iv) Finally, unlike the B-H that tests all the alphas, the screening B-H procedure only tests alphas that are in $\widehat{\mathcal{I}}$. Our theorem shows that it is safe to only focus on $\widehat{\mathcal{I}}$, because those alphas that are not inside $\widehat{\mathcal{I}}$ all satisfy $\alpha_i \leq 0$ (asymptotically).

A.3 When Observed Factors are Tradable

In this section, we consider a special case of the mixed factors (observed + latent factors), in which observed factors are all tradable. In this case, the observed factors' risk premia are equal to the factors' time series expectations. As a result, a simpler algorithm can be employed to estimate alphas. Consider the model

$$r_{it} = \alpha_i + \beta_{l,i}'\lambda_l + \beta_{o,i}'f_{o,t} + \beta_{l,i}'(f_{l,t} - \mathbb{E}f_{l,t}) + u_{it} \tag{A.1}$$

where $f_{o,t}$ and $f_{l,t}$ respectively denote the observed and latent factors, and λ_l is the risk premia for the latent factors. We assume $f_{o,t}$ are tradable so the risk premia for the observable factors satisfies $\lambda_o = \mathbb{E}f_{o,t}$.

We assume $K_l = \dim(f_{l,t}) > 0$; otherwise a simple time series regression would suffice. Consider the following algorithm.

Algorithm 8 (Estimating α in Model (A.1))**.**

S1. The same as S1 in Algorithm 6.

S2. Let $\bar{f}_o = \frac{1}{T}\sum_t f_{o,t}$,

$$\begin{aligned} \widehat{\lambda}_l &= (\widehat{\beta}_l'\mathbb{M}_{1_N}\widehat{\beta}_l)^{-1}\widehat{\beta}_l'\mathbb{M}_{1_N}(\bar{r} - \widehat{\beta}_o\bar{f}_o) \\ \widehat{\alpha}_i &= \bar{r}_i - \widehat{\beta}_{o,i}'\bar{f}_o - \widehat{\beta}_{l,i}'\widehat{\lambda}_l, \quad i = 1, ..., N. \end{aligned}$$

S3. Calculate the standard error as

$$\mathrm{se}(\widehat{\alpha}_i) = \frac{1}{\sqrt{T}}\widehat{\sigma}_i, \quad \widehat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{it}^2(1 - \widehat{v}_{l,t}'\widehat{\Sigma}_{f,l}^{-1}\widehat{\lambda}_l)^2 \tag{A.2}$$

where $\widehat{u}_{it} = z_{it} - \widehat{\beta}_{l,i}'\widehat{v}_{l,t}$, $\widehat{v}_{l,t} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\beta}_{l,i}(z_{it} - \bar{z}_i)$, and $\widehat{\Sigma}_{f,l} = \frac{1}{T}\sum_{t=1}^{T}\widehat{v}_{l,t}\widehat{v}_{l,t}'$.

Note that S2 is the key difference between Algorithm 8 and Algorithm 6. Algorithm 6 S2 runs the cross-sectional regression on all the estimated betas $(\widehat{\beta}_o, \widehat{\beta}_l)$ to estimate the risk premia for both observed and latent factors. In contrast, when the observed factors are tradable, their risk premia can be simply estimated by taking the factor time series averages. Hence in S2 of Algorithm 8, we only need to run cross-sectional OLS on the latent factor betas to estimate the risk premia for the latent factors.

A.4 Inference on α_0.

Let $\alpha_0 = \frac{1}{N}\sum_i \mathbb{E}\alpha_i$. Here we provide the asymptotic distribution for the estimator for α_0, given by $\widehat{\alpha}_0 = \frac{1}{N}\sum_{i=1}^{N}\widehat{\alpha}_i$. The presented result can be used for inferences about α_0.

Theorem 3. *Let $\sigma_\alpha^2 > 0$ denote the cross-sectional variance. Suppose $\mathbb{E}\alpha_i^4 < C$ and β is deterministic. Suppose $\liminf(1 - \bar{\beta}'(\frac{1}{N}\beta'\beta)^{-1}\bar{\beta})^2 > 0$ and Assumptions 1-4 hold. In all the four scenarios of factors, for $N = o(T^2)$,*

$$\sqrt{N}\frac{\widehat{\alpha}_0 - \alpha_0}{s_0} \xrightarrow{d} \mathcal{N}(0,1), \quad s_0^2 = (1 - \widehat{\bar{\beta}}'(\frac{1}{N}\widehat{\beta}'\widehat{\beta})^{-1}\widehat{\bar{\beta}})^{-1}\widehat{\sigma}_\alpha^2,$$

where $\widehat{\bar{\beta}} = \frac{1}{N}\sum_i \widehat{\beta}_i$ and $\widehat{\sigma}_\alpha^2 = \frac{1}{N}\sum_i(\widehat{\alpha}_i - \widehat{\alpha}_0)^2$.

A.5 Inconsistency in the Low Dimensional Setting

When the dimension N is fixed and only observable factors (but not all tradable) are considered, researchers frequently use two-pass regressions to estimate the alphas: (i) run time series regressions to estimate individual betas; (ii) run cross-sectional regressions of the averaged returns on the estimated betas to estimate the risk premia and alphas. This procedure works in GRS's asset pricing applications, where the goal is to test the null hypothesis: $\mathbb{H}_0$: all alphas are zero. On the other hand, testing for $\mathbb{H}_0$ is not of direct interest in studying hedge funds. As we shall formally show below, when the dimension N is fixed, the two-pass regression method fails to consistently estimate *any* alpha, so cannot be used in the FDR control or any multiple testing problems.

Proposition 1 (Inconsistent Estimation of α)**.** *Consider the case factors are observable. Suppose $N < C$ for some $C > 0$, and $T \to \infty$. Suppose α is stochastic and β is deterministic, satisfying $\alpha_1, ..., \alpha_N$ are iid,* $\mathrm{Var}(\alpha_i) > 0$*, and $S_\beta = \frac{1}{N}\sum_{j=1}^N(\beta_j - \bar\beta)(\beta_j - \bar\beta)$ is positive definite. We have: for each $i \le N$, as long as $\beta_i \neq 0$, then there is a random variable X_i so that* $\mathrm{Var}(X_i) > 0$ *and*

$$\widehat{\alpha}_i \xrightarrow{P} \alpha_i + X_i.$$

In fact, $X_i = -\beta_i'\eta_N$ with $\eta_N = \frac{1}{N}S_\beta^{-1}\beta'\mathbb{M}_{1_N}\alpha$.

Proof. When N is bounded, (B.12) still holds:

$$\widehat{\alpha} - \alpha = \bar{u} - \frac{1}{T}\sum_t u_t v_t' S_f^{-1}\widehat{\lambda} + \bar{u}\bar{v}' S_f^{-1}\widehat{\lambda} - \beta S_\beta^{-1}\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha - \beta\sum_{d=1}^7 A_d$$

Now $\widehat{\beta} - \beta = O_P(\frac{1}{\sqrt{T}})$, $\bar{u} = O_P(\frac{1}{\sqrt{T}})$ and $\frac{1}{T}\sum_t u_t v_t' = O_P(\frac{1}{\sqrt{T}})$. So $A_d = o_P(1)$ for all d. So

$$\widehat{\alpha}_j - \alpha_j = X_i + o_P(1),$$

where $X_i = -\beta_i' S_\beta^{-1}\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha$. Then $\mathrm{Var}(X_i) = \frac{1}{N}\beta_i' S_\beta^{-1}\beta_i \mathrm{Var}(\alpha_i) > 0$ so long as $\beta_i \neq 0$. Q.E.D.

In the usual factor pricing literature, the two-pass regression is consistent for alphas when N is fixed, because the consistency and asymptotic distributions are established under the null hypothesis: $\mathbb{H}_0$: all alphas are zero. Under such null, $\mathrm{Var}(\alpha_i) = 0$ so $X_i = 0$ in the above proposition. However, as long as there are at least one alpha that is nonzero so that $\mathrm{Var}(\alpha_i) > 0$, we have $X_i \neq 0$, then the estimated $\widehat{\alpha}_i$ would be inconsistent.

B Technical Proofs

Recall that $v_t = f_t - \mathbb{E}f_t$. Throughout the proofs, we shall use Δ to represent a generic $N \times d$ matrix of "estimation errors", which may vary from case by case; here $d \in \{K, K_o, K_l\}$ is a fixed dimension that does not grow.

B.1 Proof of Theorem 1

By Proposition 2, $\widehat{\alpha}_i - \alpha_i = \frac{1}{T}\sum_t u_{it}(1 - v_t'\Sigma_f^{-1}\lambda) - \frac{1}{N}\beta_i' S_\beta^{-1}\beta'\mathbb{M}_{1_N}\alpha + O_P(\frac{\log N}{T} + \frac{1}{N})$. Now let $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$, we have for $\zeta_{i,T} = \frac{1}{\sqrt{T}}\sum_t u_{it}(1 - v_t'\Sigma_f^{-1}\lambda)$ and $\zeta_{i,N} = -\frac{1}{\sqrt{N}}\beta_i' S_\beta^{-1}\beta'\mathbb{M}_{1_N}\alpha$

$$\delta_{NT}(\widehat{\alpha}_i - \alpha_i) = \frac{\delta_{NT}}{\sqrt{T}}\zeta_{i,T} + \frac{\delta_{NT}}{\sqrt{N}}\zeta_{i,N} + o_P(1).$$

Then $\zeta_{i,T} \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(u_{it}(1 - v_t\Sigma_f^{-2}\lambda)))$ and $\zeta_{i,N} \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(\alpha_i)\beta_i' S_\beta^{-1}\beta_i)$. In addition, $\mathrm{Cov}(\zeta_{i,T}, \zeta_{i,N}) = 0$, thus $(\zeta_{i,T}, \zeta_{i,N})$ jointly converges to a bivariate normal distribution. Based on this, we can apply the same argument of the proof of Theorem 3 in Bai (2003) to conclude that

$$\frac{\widehat{\alpha}_i - \alpha_i}{(\frac{1}{T}\mathrm{Var}(u_{it}(1 - v_t\Sigma_f^{-1}\lambda)) + \frac{1}{N}\mathrm{Var}(\alpha_i)\beta_i' S_\beta^{-1}\beta_i)^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1).$$

B.2 Proof of Theorem 2

We use $\widehat{\alpha}$, $\mathrm{se}(\widehat{\alpha}_i)$ and t_i to denote the estimated α, its standard error and t-statistics. The proof extends that of Liu and Shao (2014) to our context that (i) $\sqrt{T}(\widehat{\alpha}-\alpha)$ is only approximately equal to $\frac{1}{\sqrt{T}}\sum_t u_t(1-v_t'\Sigma_f^{-1}\lambda)$, up to a term $\|\Delta\|_\infty = o_P(1)$ when $T\log N = o(N)$; (ii) The power comparison between the usual B-H and the screening B-H.

By Assumption 4, there is $\mathcal{H}\subset\{1,...,N\}$ so that $|\mathcal{H}|\to\infty$ and

$$\sqrt{T}\sigma_i^{-1}\alpha_i \geq 4\sqrt{\log N}, \forall i\in\mathcal{H}. \tag{B.3}$$

Next, let $\mathcal{H}_0$ denote the index set of all the true null hypotheses. Also, let $\Psi(x) := 1-\Phi(x)$. Our major goal is to bound the number of false rejections

$$\mathcal{F} = \sum_{i\in\mathcal{H}_0} 1\{t_i \geq t_{(\widehat{k})}\}.$$

The main inequality to use is: uniformly for $x\in[0,t^*]$, where $t^* = \Psi^{-1}(\tau|\mathcal{H}|/N)$,

$$\frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} 1\{t_i\geq x\} \leq \Psi(x)(1+o_P(1)) \tag{B.4}$$

The remaining proof is divided into the following steps.

Step 1. Show the inequality (B.4). This inequality is essentially the Gaussian approximation to the "empirical measure" of the t-statistics for those true null hypotheses, whose proof requires weak dependence among the t-statistics. The proof simply extends that of Liu and Shao (2014) to allowing approximation errors Δ_i.

Proof. (a) Write $z_i = \frac{1}{\sqrt{T}}\sum_t X_{it}/s_i$ where $X_{it} = u_{it}(1-v_t'\Sigma_f^{-1}\lambda)$. When $T\log N = o(N)$, $\alpha_i\leq 0$ we have $t_i \leq (\widehat{\alpha}_i-\alpha_i)/\mathrm{se}(\widehat{\alpha}_i) = \frac{1}{\sqrt{T}}\sum_t X_{it}/s_i + \Delta_i$ where $\max_i|\Delta_i| = o_P(1/\sqrt{\log N})$ by Proposition 2. Hence

$$\frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} 1\{t_i\geq x\} \leq \frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} 1\{z_i\geq x-\|\Delta\|_\infty\}.$$

The right hand side does not depend on α because z_i is centered and independent of α.

(b) The same argument as that of Liu and Shao (2014) shows, uniformly for $x\leq\Psi^{-1}(\tau|\mathcal{H}|/(2N))$,

$$\frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} 1\{z_i\geq x\} \leq \Psi(x)(1+o_P(1)) \tag{B.5}$$

where $o_P(1)$ is independent of x,α. On the other hand, there is $\eta_x\in[0,\|\Delta\|_\infty]$ so that for some universe constant $C>0$, uniformly for $0<x\leq t^*$

$$\begin{aligned}
|\Psi(x)-\Psi(x-\|\Delta\|_\infty)| &\leq \phi(x+\eta_x)\|\Delta\|_\infty \leq \phi(x)\|\Delta\|_\infty\frac{\phi(x+\eta_x)}{\phi(x)} \\
&\leq C\phi(x)\|\delta\|_\infty\exp(C\eta_x(\eta_x+t^*)) \\
&\leq Cx\Psi(x)\|\Delta\|_\infty(1+o(1)) \leq Ct^*\Psi(x)\|\Delta\|_\infty(1+o(1))
\end{aligned}$$

$$\leq \quad o(1)\Psi(x)$$

where $o(1)$ is a uniform term because $\eta_x t^* \leq \|\Delta\|_\infty t^* \leq o_P(1/\sqrt{\log N})\sqrt{2\log N} = o(1)$; the fact that $t^* \leq \sqrt{2\log N}$ is to be shown in step 2 below. This proves $\Psi(x) = \Psi(x - \|\Delta\|_\infty)(1 + o(1))$. Also,

$\Psi(x - \|\Delta\|_\infty) = \Psi(x)(1+o(1)) \geq \Psi(t^*)(1+o(1)) \geq (1+o(1))\tau|\mathcal{H}|/N \geq \tau|\mathcal{H}|/(2N)$. So $x - \|\Delta\|_\infty \leq \Psi^{-1}(\tau|\mathcal{H}|/(2N))$. Hence apply (B.5),

$$\frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} 1\{t_i \geq x\} \leq \frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} 1\{z_i \geq x - \|\Delta\|_\infty\} \leq \Psi(x - \|\Delta\|_\infty)(1 + o_P(1)) = \Psi(x)(1 + o_P(1)).$$

Step 2. An equivalent statement for rejections: $t_i \geq t_{(\widehat{k})}$ if and only if $t_i \geq \widehat{t}$, where

$$\widehat{t} := \inf\{x \in \mathbb{R} : \Psi(x) \leq \tau\frac{1}{N}\max\{\sum_{i=1}^N 1\{t_i \geq x\}, 1\}\}.$$

Proof. The proof of this step is to show that $t_{(\widehat{k}+1)} \leq \widehat{t} \leq t_{(\widehat{k})}$, and is the same as that of Lemma 1 of Storey et al. (2004). So we omit it to avoid repetitions.

Given step 2, our goal becomes to bound $F = \sum_{i\in\mathcal{H}_0} 1\{i \leq N : t_i \geq \widehat{t}\}$. To use inequality (B.4), we then aim to prove that $x = \widehat{t} \leq t^*$. To do so, note that

$$c\Psi(\widehat{t}) = \tau\frac{1}{N}\max\{\sum_{i=1}^N 1\{t_i \geq \widehat{t}\}, 1\}, \tag{B.6}$$

Hence proving $\widehat{t} \leq t^*$ is equivalent to proving $\Psi(\widehat{t}) \geq \Psi(t^*)$, that is

$$\sum_{i=1}^N 1\{t_i \geq \widehat{t}\} \geq |\mathcal{H}|. \tag{B.7}$$

In words, the number of rejections (if there is any) is at least $|\mathcal{H}|$. This is to be done in the following steps.

Step 3. Prove $\mathbb{P}(\forall j \in \mathcal{H}, t_j \geq \sqrt{2\log N}) \to 1$. Intuitively, it means the t-statistics of "large" true alphas are also large. In then implies

$$\sum_{i=1}^N 1\{t_i \geq \sqrt{2\log N}\} \geq |\mathcal{H}|.$$

Proof. By Proposition 2, for $z_i = \frac{1}{\sqrt{T}}\sum_t u_{it}(1 - v_t'\Sigma_f^{-1}\lambda)/s_i$, $(\widehat{\alpha}_i - \alpha_i)/\operatorname{se}(\widehat{\alpha}_i) = z_i + \Delta_i$. So

$$t_i \geq \alpha_i/\operatorname{se}(\widehat{\alpha}_i) - |z_i| - \Delta_i.$$

Next, $\sqrt{T}\max_i|\operatorname{se}(\widehat{\alpha})\sqrt{T} - \sigma_i| \leq O_P(\sqrt{\log N} + \sqrt{T/N})$ by (B.19). So for all α_i satisfying $\sqrt{T}\sigma_i^{-1}\alpha_i \geq L_n\sqrt{\log N}$ with $L_n \to \infty$, and $T = o(N)$,

$$\alpha_i/\operatorname{se}(\widehat{\alpha}_i) \geq \sqrt{T}\sigma_i^{-1}\alpha_i - O_P(\sqrt{\log N} + \sqrt{T/N}) \geq L_n\sqrt{\log N}/2.$$

Now note that $\sqrt{T}\sigma_i^{-1}\alpha_i \geq L_n\sqrt{\log N}$ for all $i \in \mathcal{H}$. So by Lemma 1, uniformly for these i,

$$t_i \geq L_n\sqrt{\log N}/2 - \sqrt{3\log N} - o_P(1) \geq \sqrt{2\log N}.$$

Step 4. The number of rejections (if there is any) is at least $|\mathcal{H}|$. It is equivalent to (B.7).

Proof. Because $|\mathcal{H}| \to \infty$, $\Phi(x) \leq 0.5\exp(-x^2/2)$, we have $t^* = \Psi^{-1}(\tau|\mathcal{H}|/N) \leq \sqrt{2\log N}$. Then by step 3, $\Psi(t^*) \leq \frac{1}{N}\sum_{i=1}^N 1\{t_i \geq \sqrt{2\log N}\}\tau \leq \frac{1}{N}\sum_{i=1}^N 1\{t_i \geq t^*\}\tau$. So by the definition of $\widehat{t}$, we have $\widehat{t} \leq t^*$ and thus $\Psi(\widehat{t}) \geq \tau|\mathcal{H}|/N$. In addition, by the definition of $\widehat{t}$, we have

$$\Psi(\widehat{t}) = \tau\frac{1}{N}\sum_{i=1}^N 1\{t_i \geq \widehat{t}\} \geq \tau\frac{|\mathcal{H}|}{N} \tag{B.8}$$

Step 5. The FDR control.

Proof. In step 4 $\widehat{t} \leq t^*$ with probability converging to one, then by (B.4), $\mathcal{F} \leq \Psi(\widehat{t})|\mathcal{H}_0| + o_P(1)|\mathcal{H}_0|$. Also by (B.6),

$$\mathcal{R} = \sum_{i=1}^N 1\{t_i \geq \widehat{t}\} = \Psi(\widehat{t})N/\tau.$$

It then gives, for some $X = o_P(1)$, and $|X| \leq 1$ almost surely, $\frac{\mathcal{F}}{\mathcal{R}} \leq \tau\frac{|\mathcal{H}_0|}{N} + X$, on the event $\widehat{t} \leq t^*$. The above inequality is proved conditioning on the event $R \geq 1$. Together, for any $\epsilon > 0$,

$$\begin{aligned} \text{FDR} &\leq \mathbb{E}(\tau\frac{|\mathcal{H}_0|}{N} + X|\mathcal{R} \geq 1, \widehat{t} \leq t^*) + \mathbb{P}(\widehat{t} > t^*|\mathcal{R} \geq 1) \\ &\leq \tau\frac{|\mathcal{H}_0|}{N} + \epsilon + \mathbb{P}(|X| \geq \epsilon|\mathcal{R} \geq 1) + o(1). \end{aligned}$$

Since ϵ is chosen arbitrarily, FDR $\leq \tau\frac{|\mathcal{H}_0|}{N} + o(1)$.

Step 6. The power property.

Proof. Note that in the proof of Steps 3 and 4 we have proved

$$\mathbb{P}(t_i \geq \sqrt{2\log N} \geq t^* \geq \widehat{t}, \quad \forall i \in \mathcal{H}) \to 1.$$

Note that $t_i \geq \widehat{t}$ if and only if $\mathbb{H}_0^i$ is rejected. This proves the desired power property that

$$\mathbb{P}(\mathbb{H}_0^i \text{ is false and rejected, for all } i \in \mathcal{H}) \to 1.$$

Step 7. FDR and power properties for the screening B-H.

Proof. Steps 1-6 proves the properties for the B-H procedure. The proof is immediately adaptive to the screening B-H procedure. The only additional proof we need is to make sure that $\mathcal{H} \subset \widehat{\mathcal{I}}$ with probability approaching one. In fact, for any $i \in \mathcal{H}$, we note $\alpha_i \geq L_{NT}\sqrt{\frac{\log N}{T}}$. So

$$\widehat{\alpha}_i/\operatorname{se}(\widehat{\alpha}_i) > \sqrt{T}\sigma_i^{-1}\alpha_i - O_P(\sqrt{\log N} + \sqrt{T/N}) \geq L_n\sqrt{\log N}/2 > 0$$

Hence $i \in \widehat{\mathcal{I}}$. The rest of the proof for the screening B-H is the same as that of B-H. Next, to prove (iv), we aim to show $\mathbb{P}(\widehat{\mathcal{I}}^c \subset \mathcal{H}_0) \to 1$ where $\mathcal{H}_0$ denotes the collection of all true null hypotheses. In fact, for any $i \notin \widehat{\mathcal{I}}$, we have $\widehat{\alpha}_i/\operatorname{se}(\widehat{\alpha}_i) \leq -\sqrt{\log\log N}$. Thus

$$\alpha_i/\operatorname{se}(\widehat{\alpha}_i) \leq -\sqrt{\log\log N} + \sqrt{M\log N} + O_P(\sqrt{T/N}) < 0$$

Hence it is true that $\alpha_i < 0$ and thus $i \in \mathcal{H}_0$.

Finally, to prove $\mathbb{P}(A_{\text{screening B-H}}) \geq \mathbb{P}(A_{\text{B-H}})$, let $\widehat{k}_{\text{screening B-H}}$ and $\widehat{k}_{\text{B-H}}$ respectively denote the cut-off for the screening B-H and B-H. Thus

$$p_{(\widehat{k}_{\text{B-H}})} \leq \frac{\tau \widehat{k}_{\text{B-H}}}{N} \leq \frac{\tau \widehat{k}_{\text{B-H}}}{|\widehat{\mathcal{I}}|}.$$

Let j be the index of $(\widehat{k}_{\text{B-H}})$ so that $p_{(\widehat{k}_{\text{B-H}})} = p_j$. Suppose it is true that $j \in \widehat{\mathcal{I}}$, then by the alpha-screening method, $\widehat{k}_{\text{B-H}} \leq \widehat{k}_{\text{screening B-H}}$. Hence on the event $A_{\text{B-H}}$,

$$\max_{\mathbb{H}_0^i \text{ is false}} p_i \leq p_{(\widehat{k}_{\text{B-H}})} \leq p_{(\widehat{k}_{\text{screening B-H}})}.$$

Because we proved $\widehat{\mathcal{I}}^c \subset \mathcal{H}_0$ asymptotically, thus if $\mathbb{H}_0^i$ is false, $i \in \widehat{\mathcal{I}}$. Now for all $i \in \widehat{\mathcal{I}}$, it is rejected if and only if $p_i \leq p_{(\widehat{k}_{\text{screening B-H}})}$. The above inequality then implies that on the event $A_{\text{B-H}}$, the event $A_{\text{screening B-H}}$ also holds.

It remains to prove that $j \in \widehat{\mathcal{I}}$. Note that $p_j = p_{(\widehat{k}_{\text{B-H}})} \leq \frac{\tau \widehat{k}_{\text{B-H}}}{N} \leq \tau < 1/2$, then for one-sided test, $t_j > 0 > -\sqrt{\log\log N}$, so indeed $j \in \widehat{\mathcal{I}}$.

Q.E.D.

B.3 Asymptotic Expansions for the Estimated Alpha

The following proposition gives the asymptotic expansion for the estimated alphas. It applies to estimators that are obtained in any of the four factor scenarios: (i) observable factors only (Algorithm 3), (i) latent factors only (Algorithm 4), (iii) the general case (mixed of observable and latent factors, Algorithm 3), and (iv) mixed of observable and latent factors with additional condition that observable factors are tradable (Algorithm 8)

Proposition 2. *Under the conditions of Theorem 1, (i)*

$$\widehat{\alpha} - \alpha = \frac{1}{T}\sum_t u_t(1 - v_t'\Sigma_f^{-1}\lambda) - \frac{1}{N}\zeta \mathbb{M}_{1_N}\alpha + \Delta,$$

with $\|\Delta\|_\infty = O_P(\frac{\log N}{T} + \frac{1}{N})$. *Here* $\zeta = \beta S_\beta^{-1}\beta'$ *for scenarios (i)-(iii) and* $\zeta = \beta_l S_{\beta,l}^{-1}\beta_l'$ *for scenario (iv).*

(ii) Uiformly in $i \leq N$, *when* $T\log N = o(N)$,

$$\frac{\widehat{\alpha}_i - \alpha_i}{\text{se}(\widehat{\alpha}_i)} = \sqrt{T}\frac{\frac{1}{T}\sum_t u_{it}(1 - v_t'\Sigma_f\lambda)}{\sigma_i} + o_P(1/\sqrt{\log N})$$
$$= \sqrt{T}\frac{\frac{1}{T}\sum_t u_{it}(1 - v_t'\Sigma_f\lambda)}{s_i} + o_P(1/\sqrt{\log N})$$

where $\sigma_i^2 = \mathbb{E}u_{it}^2(1 - v_t'\Sigma_f^{-1}\lambda)^2$, $s_i^2 = \frac{1}{T}\sum_t u_{it}^2(1 - v_t'\Sigma_f^{-1}\lambda)^2$, *Here* $\widehat{\alpha}_i$ *and* $\text{se}(\widehat{\alpha}_i)$ *denote the estimated alpha and its standard error.*

Proof. For notational simplicity, we shall simply work with the case $\dim(f_t) = 1$. We use $C > 0$ to be generic constant.

(i) **Scenario I.** In the known factor case, let $\widehat{\beta}$ be the $N \times K$ matrix of $\widehat{\beta}_i$. Then we have

$$\widehat{\beta} - \beta = (\frac{1}{T}\sum_t u_t v_t' - \bar{u}\bar{v}')S_f^{-1} \tag{B.9}$$

where $S_f = \frac{1}{T}\sum_t (f_t - \bar{f})(f_t - \bar{f})'$. It is easy to show $\frac{1}{N}\|\widehat{\beta} - \beta\|^2 = O_P(\frac{1}{T})$.

Step 1. Expand $\widehat{\lambda} - \lambda$. Note that $\bar{r} - \mathbb{E}r_t = \beta\bar{v} + \bar{u}$, and $\widehat{\lambda} = \widehat{S}_\beta^{-1}\frac{1}{N}\widehat{\beta}'\mathbb{M}_{1_N}\bar{r}$, so

$$\widehat{\lambda} - \lambda = \bar{v} + \frac{1}{N}S_\beta^{-1}\beta'\mathbb{M}_{1_N}\alpha + \sum_{d=1}^{7} A_d,$$

where

$$\begin{aligned}
A_1 &= \frac{1}{N}\widehat{S}_\beta^{-1}(\widehat{\beta} - \beta)'\mathbb{M}_{1_N}\alpha, \quad A_2 = \frac{1}{N}\widehat{S}_\beta^{-1}(\widehat{\beta} - \beta)'\mathbb{M}_{1_N}(\beta - \widehat{\beta})\lambda \\
A_3 &= \frac{1}{N}\widehat{S}_\beta^{-1}\beta'\mathbb{M}_{1_N}(\beta - \widehat{\beta})\lambda, \quad A_4 = \frac{1}{N}\widehat{S}_\beta^{-1}\widehat{\beta}'\mathbb{M}_{1_N}(\beta - \widehat{\beta})\bar{v} \\
A_5 &= \frac{1}{N}\widehat{S}_\beta^{-1}(\widehat{\beta} - \beta)'\mathbb{M}_{1_N}\bar{u}, \quad A_6 = \frac{1}{N}\widehat{S}_\beta^{-1}\beta'\mathbb{M}_{1_N}\bar{u} \\
A_7 &= (\frac{1}{N}\widehat{S}_\beta^{-1} - \frac{1}{N}S_\beta^{-1})\beta'\mathbb{M}_{1_N}\alpha
\end{aligned}$$

We now show $\|A_d\| = O_P(\frac{1}{\sqrt{NT}})$ for all d. Note that $\psi_1(\mathrm{Var}(u_t|f_t)) < C$ almost surely,

$$\mathbb{E}\|\alpha'\frac{1}{T}\sum_t u_t f_t'\|^2 = \frac{1}{T}\mathbb{E}f_t^2\alpha'\,\mathrm{Var}(u_t|f_t)\alpha \le C\mathbb{E}f_t^2\|\alpha\|^2 \le NT^{-1}C.$$

$$\mathbb{P}(\|\frac{1}{N}\alpha'\frac{1}{T}\sum_t u_t f_t'\|\sqrt{NT} > C_\delta) \le \mathbb{E}\|\frac{1}{N}\alpha'\frac{1}{T}\sum_t u_t f_t'\|^2 NT/C_\delta^2 < \delta. \tag{B.10}$$

Similarly, $|\alpha'\bar{u}|$, $\|\beta'\frac{1}{T}\sum_t u_t f_t'\|$, $\|\beta'\bar{u}\|$, $\|1_N'\frac{1}{T}\sum_t u_t f_t'\|$, and $|1_N'\bar{u}|$ are all $O_P(N^{1/2}T^{-1/2})$. . Thus it is straightforward to prove all the following terms are $O_P(\frac{1}{\sqrt{NT}})$: $\|\frac{1}{N}(\widehat{\beta} - \beta)'\mathbb{M}_{1_N}\zeta\|$ for $\zeta \in \{\alpha, \beta, 1_N\}$ and $(\widehat{S}_\beta^{-1} - S_\beta^{-1})\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha$. This implies $\|A_d\| = O_P(\frac{1}{\sqrt{NT}})$ for all d. In other words,

$$\widehat{\lambda} - \lambda = \bar{v} + \frac{1}{N}S_\beta^{-1}\beta'\mathbb{M}_{1_N}\alpha + O_P(\frac{1}{\sqrt{NT}}). \tag{B.11}$$

It also implies $\widehat{\lambda} = O_P(1)$ and $\widehat{\lambda} - \lambda = O_P(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{N}})$.

Step 2. Expand $\widehat{\alpha} - \alpha$. Note that $\widehat{\alpha} = \bar{r} - \widehat{\beta}\widehat{\lambda}$, we have $\widehat{\alpha} - \alpha = \beta\bar{v} + \bar{u} - \beta(\widehat{\lambda} - \lambda) + (\beta - \widehat{\beta})\widehat{\lambda}$. Substitute in (B.9) (B.11),

$$\widehat{\alpha} - \alpha = \bar{u} - \frac{1}{T}\sum_t u_t v_t' S_f^{-1}\widehat{\lambda} + \bar{u}\bar{v}'S_f^{-1}\widehat{\lambda} - \beta\frac{1}{N}S_\beta^{-1}\beta'\mathbb{M}_{1_N}\alpha - \beta\sum_{d=1}^{7} A_d \tag{B.12}$$

By Lemma 2, $\|\bar{u}\bar{v}'S_f^{-1}\widehat{\lambda}\|_\infty = O_P(\sqrt{\log N}/T)$. In addition, by step 1,

$$\|\beta\sum_{d=1}^{7} A_d\|_\infty = O_P(1)\|\sum_{d=1}^{7} A_d\| = O_P(\frac{1}{\sqrt{NT}}).$$

Also $\|\frac{1}{T}\sum_t u_t v_t'(S_f^{-1}\widehat{\lambda}-\Sigma_f^{-1}\lambda)\|_\infty \leq \|\frac{1}{T}\sum_t u_t v_t\|_\infty \|S_f^{-1}\widehat{\lambda}-\Sigma_f^{-1}\lambda\| K \leq O_P(\sqrt{\log N}/T)$. So for $\|\Delta\|_\infty = O_P(\sqrt{\log N}/T + 1/N) = o_P(T^{-1/2})$,

$$\widehat{\alpha}-\alpha = \frac{1}{T}\sum_t u_t(1-v_t'\Sigma_f^{-1}\lambda) - \beta\frac{1}{N}S_\beta^{-1}\beta'\mathbb{M}_{1_N}\alpha + \Delta.$$

Scenario II. In the latent factor case, we proceed as follows.

Step 1. Expand $\widehat{\beta}$. Recall that V is the $K_l \times K_l$ diagonal matrix of the first K_l eigenvalues of S/N. Let $\dot{v}_t = v_t - \bar{v}$ and $\dot{u}_t = u_t - \bar{u}$, and

$$H = \frac{1}{NT}\sum_t \dot{v}_t\dot{v}_t'\beta'\widehat{\beta}D^{-1} + \frac{1}{NT}\sum_t \dot{v}_t\dot{u}_t'\widehat{\beta}D^{-1}.$$

Note that there are three small differences here compared to Bai (2003). First, here we expand the estimated betas while he expanded the estimated factors. They are symmetric, so can be analogously derived; secondly, Bai (2003) defined H using just the first term. In contrast, we have a second term in the definition, which introduces just tiny differences because it is $o_P(1)$ and dominated by the first term. Doing so makes the technical argument slightly more convenient, because one of the terms in the expansions in Bai (2003) now is "absorbed" in the second term in H. Finally, we use "demeaned data" $\dot{v}_t$ and $\dot{u}_t$, which also introduce further terms in the expansions below (term G). Above all, we can use the same argument to reach $\|D^{-1}\| + \|H\| = O_P(1)$. The same proof as in Bai (2003) shows the following equality holds

$$\widehat{\beta}-\beta H = \frac{1}{NT}\sum_t u_t v_t'\beta'\widehat{\beta}D^{-1} + \frac{1}{NT}\sum_t (u_t u_t' - \mathbb{E}u_t u_t')\widehat{\beta}D^{-1} + \frac{1}{N}(\mathbb{E}u_t u_t')\widehat{\beta}D^{-1} - G \tag{B.13}$$

where

$$G = \bar{u}\bar{v}'\frac{1}{N}\beta'\widehat{\beta}D^{-1} + \frac{1}{N}\bar{u}\bar{u}'\widehat{\beta}D^{-1}.$$

Note that $\|\frac{1}{\sqrt{N}}G\| = O_P(T^{-1})$, $\psi_1(\mathbb{E}u_t u_t') = O(1)$,

$\|\frac{1}{T}\sum_t u_t v_t'\| = O_P(\sqrt{N/T})$ and $\|\frac{1}{T}\sum_t (u_t u_t' - \mathbb{E}u_t u_t')\| = O_P(N/\sqrt{T})$. Also, the columns of $\widehat{\beta}/\sqrt{N}$ are eigenvector, so $\|\widehat{\beta}\| = O_P(\sqrt{N})$. Hence we have $\frac{1}{\sqrt{N}}\|\widehat{\beta}-\beta H\| = O_P(T^{-1/2}+N^{-1})$.

Step 2. Expand $\widehat{\lambda}$. We have

$$\widehat{\lambda} - H^{-1}\lambda = H^{-1}\bar{v} + \widehat{S}_\beta^{-1}\frac{1}{N}H'\beta'\mathbb{M}_{1_N}\alpha + \sum_{d=1}^4 A_{\lambda,d}$$

where

$$\begin{aligned}
A_{\lambda,1} &= \widehat{S}_\beta^{-1}\frac{1}{N}\widehat{\beta}'\mathbb{M}_{1_N}\bar{u} \\
A_{\lambda,2} &= \widehat{S}_\beta^{-1}\frac{1}{N}\widehat{\beta}'\mathbb{M}_{1_N}(\beta H - \widehat{\beta})H^{-1}\bar{v} \\
A_{\lambda,3} &= \widehat{S}_\beta^{-1}\frac{1}{N}\widehat{\beta}'\mathbb{M}_{1_N}(\beta H - \widehat{\beta})H^{-1}\lambda \\
A_{\lambda,4} &= \widehat{S}_\beta^{-1}\frac{1}{N}(\widehat{\beta}-\beta H)'\mathbb{M}_{1_N}\alpha.
\end{aligned}$$

We shall examine the terms on the right hand side one by one. First note that $\widehat{S}_\beta = H'S_\beta H + o_P(1)$ so $\widehat{S}_\beta^{-1} = O_P(1)$. For the first term, we proved $\|\beta'\mathbb{M}_{1_N}\bar{u}\| = O_P(N^{1/2}T^{-1/2})$ in part (i), so

$$A_{\lambda,1} = \widehat{S}_\beta^{-1}\frac{1}{N}(\widehat{\beta}-\beta H)'\mathbb{M}_{1_N}\bar{u} + \widehat{S}_\beta^{-1}\frac{1}{N}H'\beta'\mathbb{M}_{1_N}\bar{u} = O_P(\frac{1}{\sqrt{NT}}+\frac{1}{T}).$$

For $A_{\lambda,2}\sim A_{\lambda,4}$, note that the assumption $\max_{i,j\leq N}\sum_{k=1}^N|\operatorname{Cov}(u_{it}u_{kt},u_{jt}u_{kt})|<C$ implies $\max_j\psi_1(\operatorname{Var}(u_tu_{jt}))<C$, so

$$\begin{aligned}
\mathbb{E}\|\frac{1}{\sqrt{N}}\beta'\frac{1}{NT}\sum_t(u_tu_t'-\mathbb{E}u_tu_t')\|^2 &= \frac{1}{N}\sum_{j=1}^N\frac{1}{N^2T}\beta'\operatorname{Var}(u_tu_{jt})\beta \\
&\leq \max_j\psi_1(\operatorname{Var}(u_tu_{jt}))\frac{1}{N^2T}\|\beta\|^2 \leq \frac{C}{NT}. \\
\mathbb{E}\|\frac{1}{\sqrt{N}N}\beta'1_N1_N'\frac{1}{NT}\sum_t(u_tu_t'-\mathbb{E}u_tu_t')\|^2 &\leq \frac{C}{NT}.\quad\text{similarly}
\end{aligned}\tag{B.14}$$

Hence $\frac{1}{\sqrt{N}}\beta'\mathbb{M}_{1_N}\frac{1}{NT}\sum_t(u_tu_t'-\mathbb{E}u_tu_t') = O_P((NT)^{-1/2})$. Similarly, $\frac{1}{\sqrt{N}}\alpha'\mathbb{M}_{1_N}\frac{1}{NT}\sum_t(u_tu_t'-\mathbb{E}u_tu_t') = O_P((NT)^{-1/2})$. Note that $\|\mathbb{E}u_tu_t'\|<C$ by the assumption of weak cross-sectional correlation,

$$\begin{aligned}
\frac{1}{N}\beta'\mathbb{M}_{1_N}(\widehat{\beta}-\beta H) &= \frac{1}{N}\beta'\mathbb{M}_{1_N}\frac{1}{NT}\sum_t u_tv_t'\beta'\widehat{\beta}D^{-1} + \frac{1}{N}\beta'\mathbb{M}_{1_N}\frac{1}{NT}\sum_t(u_tu_t'-\mathbb{E}u_tu_t')\widehat{\beta}D^{-1} \\
&\quad +\frac{1}{N}\beta'\mathbb{M}_{1_N}\frac{1}{N}(\mathbb{E}u_tu_t')\widehat{\beta}D^{-1} - \frac{1}{N}\beta'\mathbb{M}_{1_N}G \\
&= O_P(\frac{1}{\sqrt{NT}}+\frac{1}{N}).
\end{aligned}$$

Similarly, $\frac{1}{N}\alpha'\mathbb{M}_{1_N}(\widehat{\beta}-\beta H) = O_P(\frac{1}{\sqrt{NT}}+\frac{1}{N})$. Thus $A_{\lambda,2} = O_P(\frac{1}{T}+\frac{1}{N})$. Similarly, both $A_{\lambda,3}$ and $A_{\lambda,4}$ are $O_P(\frac{1}{N}+\frac{1}{T})$. Together,

$$\widehat{\lambda}-H^{-1}\lambda = H^{-1}\bar{v} + \widehat{S}_\beta^{-1}\frac{1}{N}H'\beta'\mathbb{M}_{1_N}\alpha + O_P(\frac{1}{N}+\frac{1}{T}).\tag{B.15}$$

Step 3. Expand $\widehat{\alpha}-\alpha$. Substitute in the expansions (B.13) and (B.15) in step 2, 3,

$$\begin{aligned}
\widehat{\alpha}-\alpha &= \beta\bar{v}+\bar{u}-\beta H(\widehat{\lambda}-H^{-1}\lambda)+(\beta H-\widehat{\beta})\widehat{\lambda} \\
&= \bar{u}+\sum_{d=1}^4 G_d - \beta H\widehat{S}_\beta^{-1}\frac{1}{N}H'\beta'\mathbb{M}_{1_N}\alpha + O_P(\frac{1}{N}+\frac{1}{T}) \\
G_1 &= -\frac{1}{N}(\mathbb{E}u_tu_t')\widehat{\beta}D^{-1}\widehat{\lambda} \\
G_2 &= -\frac{1}{NT}\sum_t u_tv_t'\beta'\widehat{\beta}D^{-1}\widehat{\lambda} \\
G_3 &= -\frac{1}{NT}\sum_t(u_tu_t'-\mathbb{E}u_tu_t')\widehat{\beta}D^{-1}\widehat{\lambda} \\
G_4 &= (\bar{u}\bar{v}'\frac{1}{N}\beta'\widehat{\beta}D^{-1}+\frac{1}{N}\bar{u}\bar{u}'\widehat{\beta}D^{-1})\widehat{\lambda}.
\end{aligned}$$

First note that $\|\beta H\widehat{S}_\beta^{-1}\frac{1}{N}H'\beta'\mathbb{M}_{1_N}\alpha\|_\infty = O_P(1)\|\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha\| = O(N^{-1/2})$. For G_1, we shall obtain its rate later. For G_2, note that

$$\frac{1}{N}\beta'\widehat{\beta}D^{-1}\widehat{\lambda}-\Sigma_f^{-1}\lambda = \frac{1}{N}H'^{-1}(H'B'-\widehat{\beta}')\widehat{\beta}D^{-1}\widehat{\lambda}+H'^{-1}D^{-1}(\widehat{\lambda}-H^{-1}\lambda)+(HDH')^{-1}\lambda-\Sigma_f^{-1}\lambda$$

$$= O_P(\frac{1}{\sqrt{T}} + \frac{1}{N}) + [(HDH')^{-1} - \Sigma_f^{-1}]\lambda.$$

But $HDH' = O_P(\frac{1}{\sqrt{T}}) + \frac{1}{NT}\sum_t \dot{v}_t\dot{v}_t' H'^{-1}(H'B' - \widehat{\beta}')\widehat{\beta}H' + (\frac{1}{T}\sum_t \dot{v}_t\dot{v}_t' - \Sigma_f) + \Sigma_f = \Sigma_f + O_P(\frac{1}{\sqrt{T}})$.So

$$\|\frac{1}{NT}\sum_t u_t v_t'\beta'\widehat{\beta}D^{-1}\widehat{\lambda} - \frac{1}{T}\sum_t u_t v_t'\Sigma_f^{-1}\lambda\|_\infty = O_P(\sqrt{\frac{\log N}{T^2}} + \sqrt{\frac{\log N}{TN^2}}).$$

For G_3, note that by Lemma 2,

$$\begin{aligned}
\|\frac{1}{NT}\sum_t (u_t u_t' - \mathbb{E}u_t u_t')\widehat{\beta}D^{-1}\|_\infty &\le \max_i \|\frac{1}{\sqrt{NT}}\sum_t (u_{it}u_t' - \mathbb{E}u_{it}u_t')\|\frac{1}{\sqrt{N}}\|\widehat{\beta} - \beta H\|\|D^{-1}\| \\
&\quad + \|\frac{1}{NT}\sum_t (u_t u_t' - \mathbb{E}u_t u_t')\beta H D^{-1}\|_\infty \\
&= O_P(\sqrt{\frac{\log N}{NT}} + \sqrt{\frac{\log N}{T^2}}). \qquad \text{(B.16)}
\end{aligned}$$

As for G_4, note that for $G = \bar{u}\bar{v}'\frac{1}{N}\beta'\widehat{\beta}D^{-1} + \frac{1}{N}\bar{u}\bar{u}'\widehat{\beta}D^{-1}$

$$\|G\|_\infty \le \|\bar{u}\|_\infty \|\bar{v}'\frac{1}{N}\beta'\widehat{\beta} + \frac{1}{N}\bar{u}'\widehat{\beta}\|\|D^{-1}\| \le O_P(\sqrt{\frac{\log N}{T^2}}).$$

It remains to show that $\|G_1\|_\infty = O_P(1/N)$. To do so, we need to show $\|\widehat{\beta} - \beta H\|_\infty = O_P(\sqrt{\frac{\log N}{T}} + \frac{1}{N})$.

We use $\|A\|_1 = \max_i \sum_j |A_{ij}|$. Then by (B.16) and Lemma 2

$$\begin{aligned}
\|\widehat{\beta} - \beta H\|_\infty &\le \|\frac{1}{NT}\sum_t u_t v_t'\beta'\widehat{\beta}D^{-1} + \frac{1}{NT}\sum_t (u_t u_t' - \mathbb{E}u_t u_t')\widehat{\beta}D^{-1} + \frac{1}{N}(\mathbb{E}u_t u_t')\widehat{\beta}D^{-1} - G\|_\infty \\
&\le O_P(\sqrt{\frac{\log N}{T}}) + \frac{1}{N}\|(\mathbb{E}u_t u_t')\|_1\|\beta H D^{-1}\|_\infty + \frac{1}{N}\|(\mathbb{E}u_t u_t')\|_1\|\widehat{\beta} - \widehat{\beta}H\|_\infty\|D^{-1}\|.
\end{aligned}$$

Move the last term to the left hand side, and note that $\|(\mathbb{E}u_t u_t')\|_1 < C$ by the assumption.

$$\|\widehat{\beta} - \beta H\|_\infty = O_P(\sqrt{\frac{\log N}{T}} + \frac{1}{N}).$$

Then $\|\widehat{\beta}\|_\infty \le \|\widehat{\beta} - \beta H\|_\infty + \max_i \|b_i\| = O_P(1)$. So

$$\|G_1\|_\infty \le \frac{1}{N}\|(\mathbb{E}u_t u_t')\|_1\|\widehat{\beta}\|_\infty\|HD^{-1}\| = O_P(\frac{1}{N}).$$

Put together,

$$\widehat{\alpha} - \alpha = \frac{1}{T}\sum_t u_t(1 - v_t'\Sigma_f^{-1}\lambda) - \beta S_\beta^{-1}\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha + \Delta$$

where $\|\Delta\|_\infty = O_P(\sqrt{\frac{\log N}{T^2}} + \frac{1}{N})$.

Scenario III. In the mixed factor case, let $\widehat{\beta}_o$ be the $N \times K_o$ matrix of $\widehat{\beta}_{o,i}$ where $r = \dim(f_{o,t})$. Then we have $\widehat{\beta}_o - \beta_o = (\frac{1}{T}\sum_t u_t v_{o,t}' - \bar{u}\bar{v}_o')S_o^{-1} + \beta_l(\frac{1}{T}\sum_t f_{l,t}v_{o,t}' - \bar{f}_l\bar{v}_o')S_o^{-1}$ where $S_o = \frac{1}{T}\sum_t (f_{o,t} - \bar{f}_o)(f_{o,t} - \bar{f}_o)'$. So there is a matrix

$$A = \begin{pmatrix} I_r \\ \frac{1}{T}\sum_t \dot{f}_{l,t}\dot{v}_{o,t}'S_o^{-1} \end{pmatrix} = \begin{pmatrix} I_r \\ a \end{pmatrix}$$

so that

$$\widehat{\beta}_o - \beta A = \xi_1, \quad \xi_1 = (\frac{1}{T}\sum_t u_t v'_{o,t} - \bar{u}\bar{v}'_o)S_o^{-1}. \tag{B.17}$$

Step 1. Note that $\widehat{\beta}_o$ is a biased estimator for β_o, due to the correlations between $f_{o,t}$ and $f_{l,t}$. But the bias is $\beta_l \frac{1}{T}\sum_t \dot{f}_{l,t}\dot{v}'_{o,t}S_o^{-1}$, so is still inside the space spanned by $\beta = (\beta_o, \beta_l)$. As a result, in terms of estimating βA, $\widehat{\beta}_o$ is unbiased. In fact, we shall also show that $\widehat{\beta}_l$ also estimates "the subspace of β" without bias. As such, when we expand $\widehat{\beta}_l$, instead of centering at a rotation of β_l, we should center it at a rotation of β. We also have $Z_t = \beta l_t + \dot{u}_t - \xi_1 \dot{f}_{o,t}$, where $l_t = \dot{f}_t - A\dot{f}_{o,t}$. Therefore, we let V be the $K_l \times K_l$ diagonal matrix of the first K_l eigenvalues of $\frac{1}{T}\sum_t Z_t Z'_t$. Let $H_1 = \frac{1}{TN}\sum_t l_t(l'_t B' + \dot{u}'_t - \dot{f}'_{o,t}\xi'_1)\widehat{\beta}_l D^{-1}$. Then $\widehat{\beta}_l - \beta H_1 = \xi_2$ where $\xi_2 = \frac{1}{TN}\sum_t \dot{u}_t(l'_t B' + \dot{u}'_t - \dot{f}'_{o,t}\xi'_1)\widehat{\beta}_l D^{-1} - \xi_1 \frac{1}{TN}\sum_t \dot{f}_{o,t}(l'_t B' + \dot{u}'_t - \dot{f}'_{o,t}\xi'_1)\widehat{\beta}_l D^{-1}$. Let $\widehat{\beta} = (\widehat{\beta}_o, \widehat{\beta}_l)$, $H = (A, H_1)$ and $\xi_3 = (\xi_1, \xi_2)$. So

$$\widehat{\beta} = \beta H + \xi_3.$$

This implies $\frac{1}{N}\|\widehat{\beta} - \beta H\|^2 = O_P(\frac{1}{T} + \frac{1}{N^2})$

Step 2. Recall that $\bar{r} - \mathbb{E}r_t = \beta\bar{v} + \bar{u}$, and $\widehat{\lambda} = \widehat{S}_\beta^{-1}\frac{1}{N}\widehat{\beta}'\mathbb{M}_{1_N}\bar{r}$, where $\mathbb{M}_{1_N} = I - 1_N 1'_N/N$, So

$$\widehat{\lambda} - H^{-1}\lambda = H^{-1}\bar{v} + \widehat{S}_\beta^{-1}\frac{1}{N}H'\beta'\mathbb{M}_{1_N}\alpha + \sum_{d=1}^{4} A_{\lambda,d}$$

where

$$\begin{aligned}
A_{\lambda,1} &= \widehat{S}_\beta^{-1}\frac{1}{N}\widehat{\beta}'\mathbb{M}_{1_N}\bar{u} \\
A_{\lambda,2} &= \widehat{S}_\beta^{-1}\frac{1}{N}\widehat{\beta}'\mathbb{M}_{1_N}(\beta H - \widehat{\beta})H^{-1}\bar{v} \\
A_{\lambda,3} &= \widehat{S}_\beta^{-1}\frac{1}{N}\widehat{\beta}'\mathbb{M}_{1_N}(\beta H - \widehat{\beta})H^{-1}\lambda \\
A_{\lambda,4} &= \widehat{S}_\beta^{-1}\frac{1}{N}(\widehat{\beta} - \beta H)'\mathbb{M}_{1_N}\alpha.
\end{aligned}$$

To bound each term, note that (B.14) still applies. Even though ξ_3 now takes a different form than in the previous case, most of the proofs for the expansion in (B.15) still carries over. So we can avoid repeating ourselves but directly conclude that

$$\widehat{\lambda} - H^{-1}\lambda = H^{-1}\bar{v} + \widehat{S}_\beta^{-1}\frac{1}{N}H'\beta'\mathbb{M}_{1_N}\alpha + O_P(\frac{1}{N} + \frac{1}{T}).$$

Step 3. Similar to part (ii), we have

$$\begin{aligned}
\widehat{\alpha} - \alpha &= \beta\bar{v} + \bar{u} - \beta H(\widehat{\lambda} - H^{-1}\lambda) + (\beta H - \widehat{\beta})\widehat{\lambda} \\
&= \bar{u} - \beta H \widehat{S}_\beta^{-1}\frac{1}{N}H'\beta'\mathbb{M}_{1_N}\alpha - \xi_3 H^{-1}\lambda + \Delta
\end{aligned}$$

where Δ denotes a generic $N \times 1$ vector satisfying $\|\Delta\|_\infty = O_P(\frac{1}{N} + \frac{\log N}{T})$.

The main difference from the previous latent factor only case is to derive an expression for $\xi_3 H^{-1}$, which we now focus on. Note that by definition, for $C = \frac{1}{N}\beta'\widehat{\beta}_l D^{-1}$, $L = (l_t : t \leq T)$ be $K \times T$, $\dot{F}_o = (\dot{f}_{o,t} : t \leq T)$ be $K \times T$, and U be $N \times T$ matrix of u_{it}, we can write in a matrix form

$$\xi_3 \quad = \quad \left(\frac{1}{T} U \dot{F}_o' S_o^{-1}, \quad (\frac{1}{T} U - \frac{1}{T} U \dot{F}_o' S_o^{-1} \frac{1}{T} \dot{F}_o) L' C \right) + \Delta.$$

Write $L' = (0, L_1')$ where L_1 is $K_l \times T$ and $J = (\frac{1}{T}U - \frac{1}{T}U\dot{F}_o'S_o^{-1}\frac{1}{T}\dot{F}_o)L_1'(\frac{1}{T}L_1L_1')^{-1}$. It can be verified that

$$\begin{aligned} \xi_3 H^{-1} &=^{(1)} \left(\frac{1}{T} \sum_t u_t v_{o,t}' S_o^{-1} - Ja, J \right) + \Delta \\ &=^{(2)} \frac{1}{T} \sum_t u_t v_t' \Sigma_f^{-1} + \Delta. \end{aligned}$$

We now prove both equalities.

As for (1), note that $H_1 = \frac{1}{TN}\sum_t l_t l_t' \beta' \widehat{\beta}_l D^{-1} + O_P(\frac{1}{T} + \frac{1}{N})$,

$$l_t = \dot{f}_t - A\dot{f}_{o,t} = \begin{pmatrix} 0 \\ \dot{f}_{l,t} - a\dot{f}_{o,t} \end{pmatrix}, \quad C := \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \quad H_1 = \begin{pmatrix} 0 \\ \frac{1}{T}L_1L_1'm_2 \end{pmatrix} + O_P(\frac{1}{T} + \frac{1}{N})$$

So $L_1 = (\dot{f}_{l,t} - a\dot{f}_{o,t} : t \leq T)$, and $L'C = L_1'm_2$. Also note that $H = (A, H_1)$ so

$$\begin{aligned} \left(\frac{1}{T} U \dot{F}_o' S_o^{-1} - Ja, \quad J \right) A &= \frac{1}{T} U \dot{F}_o' S_o^{-1} \\ \left(\frac{1}{T} U \dot{F}_o' S_o^{-1} - Ja, \quad J \right) H_1 &= J \frac{1}{T} L_1 L_1' m_2 = (\frac{1}{T} U - \frac{1}{T} U \dot{F}_o' S_o^{-1} \frac{1}{T} \dot{F}_o) L_1' m_2 + O_P(\frac{1}{T} + \frac{1}{N}) \end{aligned}$$

So $\left(\frac{1}{T}U\dot{F}_o'S_o^{-1} - Ja, \quad J\right) H = \left(\frac{1}{T}U\dot{F}_o'S_o^{-1}, \quad J\frac{1}{T}L_1L_1'm_2\right) + O_P(\frac{1}{T} + \frac{1}{N}) = \xi_3 + O_P(\frac{1}{T} + \frac{1}{N})$. This proves $\zeta_3 = \left(\frac{1}{T}U\dot{F}_o'S_o^{-1} - Ja, \quad J\right) H + \Delta$ and thus (1).

As for (2), write $v_o = (v_{o,t} : t \leq T)$, $v_l = (v_{l,t} : t \leq T)$, both are "short" by "long" matrices.

$$S_f = \begin{pmatrix} S_o & S_{ol} \\ S_{ol}' & S_l \end{pmatrix} = \frac{1}{T} \begin{pmatrix} v_o v_o' & v_o v_l' \\ v_l v_o' & v_l v_l' \end{pmatrix}.$$

Let $W = S_l - S_{ol}'S_o^{-1}S_{ol}$. Using the matrix block inverse formula, $\frac{1}{T}\sum_t u_t v_t' \Sigma_f^{-1} = (a_1, a_2)$ where

$$\begin{aligned} a_1 &= \frac{1}{T} U v_o' S_o^{-1} + [\frac{1}{T} U v_o' S_o^{-1} S_{ol} - \frac{1}{T} U v_l'] W^{-1} S_{ol}' S_o^{-1} \\ a_2 &= -\frac{1}{T} U v_o' S_o^{-1} S_{ol} W^{-1} + \frac{1}{T} U v_l' W^{-1}. \end{aligned}$$

Note that $a = S_{ol}'S_o^{-1} + O_P(T^{-1/2})$, so $(\frac{1}{T}L_1L_1') = \frac{1}{T}(v_l' - av_o')(v_l - v_oa') = W + O_P(T^{-1/2})$. In the definition of J, $\dot{F}_o$ can be replaced with v_o up to Δ. So

$$J = \left[\frac{1}{T} U v_l' - \frac{1}{T} U v_o' S_o^{-1} S_{ol} \right] W^{-1} + \Delta = a_2 + \Delta$$

$$-Ja \quad = \quad \left[\frac{1}{T}Uv_o'S_o^{-1}S_{ol} - \frac{1}{T}Uv_l'\right]W^{-1}S_{ol}'S_o^{-1} + \Delta.$$

Then

$$\frac{1}{T}\sum_t u_t v_{o,t}'S_o^{-1} - Ja = \frac{1}{T}Uv_o'S_o^{-1} + \left[\frac{1}{T}Uv_o'S_o^{-1}S_{ol} - \frac{1}{T}Uv_l'\right]W^{-1}S_{ol}'S_o^{-1} + \Delta = a_1 + \Delta.$$

This proves (2). Together, in the mixed factor case, we also have

$$\widehat{\alpha} - \alpha = \frac{1}{T}\sum_t u_t(1 - v_t'\Sigma_f^{-1}\lambda) - BS_\beta^{-1}\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha + \Delta \tag{B.18}$$

where $\|\Delta\|_\infty = O_P(\frac{\log N}{T^2} + \frac{1}{N})$.

Scenario IV. In the mixed factor case with tradable observable factors, the proof is very similar to scenario III, so we omit details to avoid repetitions.

(ii) We shall work only on the latent factor case. The other cases are very similar.

Let $m_i := \frac{1}{\sqrt{T}}\sum_t u_{it}(1 - v_t'\Sigma_f^{-1}\lambda)$. When $T\log N = o(N)$,

$$\frac{\widehat{\alpha}_i - \alpha_i}{\operatorname{se}(\widehat{\alpha}_i)} = \frac{m_i + \Delta_{it}\sqrt{T}}{\sqrt{T}\operatorname{se}(\widehat{\alpha}_i)}.$$

Note that $\sqrt{T\log N}\|\Delta\|_\infty = o_P(1)$. It suffices to prove,

$$\sqrt{\log N}\max_i |m_i||\sigma_i - \sqrt{T}\operatorname{se}(\widehat{\alpha}_i)| = o_P(1) = \sqrt{\log N}\max_i |m_i||\sigma_i - s_i|.$$

By Lemma 2, $\max_i |m_i| = O_P(\sqrt{\log N})$. In addition, let $L = D^{-1}\widehat{\lambda}$. Then

$$\max_i |\sigma_i^2 - T\operatorname{se}(\widehat{\alpha}_i)^2| \le \max_i |\frac{1}{T}\sum_t \widehat{u}_{it}^2(1 - \widehat{v}_t'L)^2 - u_{it}^2(1 - v_t'\Sigma_f^{-1}\lambda)^2| + \max_i |s_i^2 - \sigma_i^2|.$$

The second term on the right is $O_P(\sqrt{\log N/T})$ by Lemma 2. We now focus on the first term. The first term is bounded by $Q_1 + Q_2 + Q_3$, where

$$\begin{aligned}
Q_1 &= \max_i |\frac{1}{T}\sum_t u_{it}^2(2 + \widehat{v}_t'L + v_t'\Sigma_f^{-1}\lambda)(\widehat{v}_t - H^{-1}v_t)'L| \\
Q_2 &= \max_i |\frac{1}{T}\sum_t u_{it}^2(2 + \widehat{v}_t'L + v_t'\Sigma_f^{-1}\lambda)v_t'(H^{-1'}L - \Sigma_f^{-1}\lambda)| \\
Q_3 &= \max_i |\frac{1}{T}\sum_t (\widehat{u}_{it} + u_{it})(\widehat{u}_{it} - u_{it})(1 - \widehat{v}_t'L)^2|
\end{aligned}$$

(a) Bound Q_1. Note that $\widehat{v}_t = \frac{1}{N}\widehat{\beta}'(r_t - \bar{r})$. So

$$\begin{aligned}
\widehat{v}_t - H^{-1}v_t &= \frac{1}{N}\widehat{\beta}'(\beta H - \widehat{\beta})H^{-1}v_t - \frac{1}{N}\widehat{\beta}'\beta\bar{v} + \frac{1}{N}\widehat{\beta}'u_t - \frac{1}{N}\widehat{\beta}'\bar{u} \\
&= \frac{1}{N}\widehat{\beta}'(\beta H - \widehat{\beta})H^{-1}v_t + \frac{1}{N}\widehat{\beta}'u_t + O_P(T^{-1/2})
\end{aligned}$$

where the last $O_P(T^{-1/2})$ is uniform in (i, t). Hence

$$Q_1 \quad \le \quad \max_i |\frac{1}{T}\sum_t v_t u_{it}^2(2 + \widehat{v}_t'v + v_t'\Sigma_f^{-1}\lambda)|\|\frac{1}{N}\widehat{\beta}'(\beta H - \widehat{\beta})H^{-1}L\|$$

$$
\begin{aligned}
&+\max_i|\frac{1}{T}\sum_t\frac{1}{N}\widehat{\beta}'u_tu_{it}^2(2+\widehat{v}_t'v+v_t'\Sigma_f^{-1}\lambda)L| \\
&+\max_i|\frac{1}{T}\sum_t u_{it}^2(2+v_t'\Sigma_f^{-1}\lambda)|O_P(T^{-1/2})+\max_i|\frac{1}{T}\sum_t u_{it}^2\widehat{v}_t|O_P(T^{-1/2}) \\
\leq\;& O_P(T^{-1}+N^{-1})\max_i|\frac{1}{T}\sum_t v_tu_{it}^2(2+v_t'\Sigma_f^{-1}\lambda)-\mathbb{E}v_tu_{it}^2(2+v_t'\Sigma_f^{-1}\lambda)| \\
&+O_P(T^{-1}+N^{-1})\max_i|\frac{1}{T}\sum_t v_t^2u_{it}^2-\mathbb{E}v_t^2u_{it}^2|+\max_i|\frac{1}{T}\sum_t u_{it}^2v_t-\mathbb{E}u_{it}^2v_t|O_P(T^{-1/2}) \\
&+O_P(T^{-1/2})\max_i|\frac{1}{T}\sum_t u_{it}^2(2+v_t'\Sigma_f^{-1}\lambda)-\mathbb{E}u_{it}^2(2+v_t'\Sigma_f^{-1}\lambda)| \\
&+|\widehat{\beta}\|_\infty\max_{ij}\left[|\frac{1}{T}\sum_t(u_{jt}u_{it}^2v_t-\mathbb{E}u_{jt}u_{it}^2v_t)|+|\frac{1}{T}\sum_t u_{jt}u_{it}^2(2+v_t'\Sigma_f^{-1}\lambda)-\mathbb{E}u_{jt}u_{it}^2(2+v_t'\Sigma_f^{-1}\lambda)|\right] \\
&+\|\widehat{\beta}\|_\infty\frac{1}{N}\max_i\left[\|\mathbb{E}u_tu_{it}^2(2+v_t'\Sigma_f^{-1}\lambda))\|_1+\|\mathbb{E}u_tu_{it}^2v_t\|_1\right] \\
&+O_P(T^{-1}+N^{-1})\max_i|\frac{1}{T}\sum_t v_tu_{it}^2(\widehat{v}_t-H^{-1}v_t)|+\max_i|\frac{1}{T}\sum_t u_{it}^2(\widehat{v}_t-H^{-1}v_t)|O_P(T^{-1/2}) \\
&+\max_i|\widehat{\beta}'\frac{1}{T}\sum_t\frac{1}{N}u_tu_{it}^2(\widehat{v}_t-H^{-1}v_t)| \\
&+O_P(T^{-1/2}+N^{-1}) \\
=\;& O_P(\sqrt{\frac{\log N}{T}}+\frac{1}{N})+O_P(1)\frac{1}{N}\max_i\sum_j|\mathbb{E}(u_{jt}u_{it}^2|v_t)| \\
&+O_P(T^{-1}+N^{-1})\max_i|\frac{1}{T}\sum_t v_tu_{it}^2(\widehat{v}_t-H^{-1}v_t)|+\max_i|\frac{1}{T}\sum_t u_{it}^2(\widehat{v}_t-H^{-1}v_t)|O_P(T^{-1/2}) \\
&+\max_i|\widehat{\beta}'\frac{1}{T}\sum_t\frac{1}{N}u_tu_{it}^2(\widehat{v}_t-H^{-1}v_t)| \\
=\;& O_P(\sqrt{\frac{\log N}{T}}+\frac{1}{N})+\frac{1}{N^2}\max_i|\beta'\frac{1}{T}\sum_t u_tu_{it}^2u_t'\beta|O_P(1)=O_P(\sqrt{\frac{\log N}{T}}+\frac{1}{N})
\end{aligned}
$$

where we bounded $\frac{1}{N^2}\max_i|\beta'\frac{1}{T}\sum_t u_tu_{it}^2u_t'\beta|$ as, for $w_t=u_t'\beta/\sqrt{N}$,

$$
\frac{1}{N}\max_i|\frac{1}{T}\sum_t u_{it}^2w_t^2|\leq\frac{1}{N}\max_i|\frac{1}{T}\sum_t(u_{it}^2w_t^2-\mathbb{E}u_{it}^2w_t^2)|+\frac{1}{N}\max_i|\mathbb{E}\frac{1}{T}\sum_t u_{it}^2w_t^2|=O_P(N^{-1}).
$$

(b) For Q_2, note that

$$
\|H^{-1'}L-\Sigma_f^{-1}\lambda\|\leq\|H^{-1'}D^{-1}H^{-1}-\Sigma_f^{-1}\|\|H\widehat{\lambda}\|+\|\Sigma_f^{-1}H\|\|\widehat{\lambda}-H^{-1}\lambda\|=O_P(N^{-1/2}+T^{-1/2}).
$$

So

$$
\begin{aligned}
Q_2\;&\leq\;\max_i\|\frac{1}{T}\sum_t u_{it}^2(2+\widehat{v}_t'L+v_t'\Sigma_f^{-1}\lambda)v_t'\|O_P(\frac{1}{\sqrt{T}}+\frac{1}{\sqrt{N}}) \\
&\leq\; O_P(\frac{1}{\sqrt{T}}+\frac{1}{\sqrt{N}})+\max_i\|\frac{1}{T}\sum_t u_{it}^2(\widehat{v}_t-H^{-1}v_t)\|O_P(\frac{1}{\sqrt{T}}+\frac{1}{\sqrt{N}}) \\
&\leq\; O_P(\frac{1}{\sqrt{T}}+\frac{1}{\sqrt{N}}).
\end{aligned}
$$

(c) For Q_3, note that $\widehat{u}_t-u_t=-\beta\bar{v}-\bar{u}-(\widehat{\beta}-\beta H)\widehat{v}_t-\beta H(\widehat{v}_t-H^{-1}v_t)$. First, we show

$\max_i |\frac{1}{T}\sum_t(\widehat{u}_{it}+u_{it})(1-\widehat{v}_t'L)^2| = O_P(1)$, due to:

$$\begin{aligned} \max_i |\frac{1}{T}\sum_t u_{it}(1-\widehat{v}_t'L)^2| &\leq O_P(1) + \max_i |\frac{1}{T}\sum_t u_{it}(\widehat{v}_t - H^{-1}v_t)^2| = O_P(1). \\ \max_i |\frac{1}{T}\sum_t \widehat{u}_{it}(1-\widehat{v}_t'L)^2| &\leq O_P(1) + \max_i |\frac{1}{T}\sum_t (\widehat{u}_{it}-u_{it})(1-\widehat{v}_t'L)^2| = O_P(1). \end{aligned}$$

Similarly, it can be shown $\max_i |\frac{1}{T}\sum_t(\widehat{u}_{it}+u_{it})(1-\widehat{v}_t'L)^2(\widehat{v}_t+w_t)| = O_P(1)$ where $w_t = \frac{1}{\sqrt{N}}B'u_t$. Next,

$$\begin{aligned} & \max_i |\frac{1}{T}\sum_t(\widehat{u}_{it}+u_{it})(1-\widehat{v}_t'L)^2(\widehat{v}_t - H^{-1}v_t)| \\ \leq\ & \max_i |\frac{1}{T}\sum_t(\widehat{u}_{it}+u_{it})(1-\widehat{v}_t'L)^2 v_t| O_P(N^{-1}+T^{-1}) \\ & + \max_i |\frac{1}{T}\sum_t(\widehat{u}_{it}+u_{it})(1-\widehat{v}_t'L)^2 \frac{1}{N}\widehat{\beta}'u_t| + O_P(T^{-1/2}) \\ \leq\ & O_P(\frac{1}{N}+\frac{1}{\sqrt{T}}) + \max_{ij} |\frac{1}{T}\sum_t(\widehat{u}_{it}+u_{it})(1-\widehat{v}_t'L)^2||u_{jt}|\|\widehat{\beta}-\beta H\|_\infty \\ \leq\ & O_P(\sqrt{\frac{\log N}{T}}+\frac{1}{N}). \end{aligned}$$

Then

$$\begin{aligned} Q_3 &= \max_i |\frac{1}{T}\sum_t(\widehat{u}_{it}+u_{it})(1-\widehat{v}_t'L)^2(\widehat{u}_{it}-u_{it})| \leq \sum_{d=1}^{9} A_d \\ A_1 &= \max_i |\frac{1}{T}\sum_t(\widehat{u}_{it}+u_{it})(1-\widehat{v}_t'L)^2 b_i\bar{v}| = O_P(T^{-1/2}) \\ A_2 &= \max_i |\frac{1}{T}\sum_t(\widehat{u}_{it}+u_{it})(1-\widehat{v}_t'L)^2| \max_i |\bar{u}_i| = O_P(\sqrt{\frac{\log N}{T}}) \\ A_3 &= \max_i |\frac{1}{T}\sum_t(\widehat{u}_{it}+u_{it})(1-\widehat{v}_t'L)^2\widehat{v}_t||\widehat{\beta}-\beta H|_\infty = O_P(\sqrt{\frac{\log N}{T}}+\frac{1}{N}) \\ A_4 &= \max_i |\frac{1}{T}\sum_t(\widehat{u}_{it}+u_{it})(1-\widehat{v}_t'L)^2(\widehat{v}_t-H^{-1}v_t)|\|\beta H\|_\infty = O_P(\sqrt{\frac{\log N}{T}}+\frac{1}{N}) \end{aligned}$$

So $Q_3 = O_P(\sqrt{\frac{\log N}{T}}+\frac{1}{N})$. Together, $Q_1+Q_2+Q_3 = O_P(\sqrt{\frac{\log N}{T}}+\frac{1}{\sqrt{N}})$. Thus

$$\max_i |\sigma_i^2 - T\,\mathrm{se}(\widehat{\alpha}_i)^2| \leq O_P(\sqrt{\frac{\log N}{T}}+\frac{1}{\sqrt{N}}). \tag{B.19}$$

Hence $\max_i |m_i||\sigma_i - \sqrt{T}\,\mathrm{se}(\widehat{\alpha}_i)| = O_P(\sqrt{\log N})O_P(\sqrt{\frac{\log N}{T}}+\frac{1}{\sqrt{N}}) = o_P(1/\sqrt{\log N})$.

Q.E.D.

Next, we prove the following lemma.

Lemma 1. *With probability going to one, and any constant $M>2$,*

$$\max_i \frac{|\frac{1}{\sqrt{T}}\sum_t u_{it}(1-v_t'\Sigma_f^{-1}\lambda)|}{\sigma_i} + \max_i \frac{|\frac{1}{\sqrt{T}}\sum_t u_{it}(1-v_t'\Sigma_f^{-1}\lambda)|}{s_i} \leq \sqrt{M\log N}.$$

where $s_i^2 = \frac{1}{T}\sum_t u_{it}^2(1-v_t'\Sigma_f^{-1}\lambda)^2$.

Proof. The proof simply applies Corollary 2.1 of Chernozhukov et al. (2013a). Let $X_{it} = u_{it}(1 - v_t'\Sigma_f^{-1}\lambda)$. Then under Assumption 3 (iii) and $\log(N)^c = o(T)$ for $c > 7$, Corollary 2.1 of Chernozhukov et al. (2013a) implies for some $c > 0$,

$$\sup_s \left| \mathbb{P}\left(\max_i \frac{|\frac{1}{\sqrt{T}}\sum_t X_{it}|}{\sigma_i} > s \right) - \mathbb{P}\left(\max_i |Y_i| > s \right) \right| \leq T^{-c}$$

where $Y_i \sim \mathcal{N}(0,1)$. In addition, $\mathbb{P}\left(\max_{i\leq N}|Y_i| > s\right) \leq 2N(1-\Phi(s)) \leq 2\exp(\log N - s^2/2) = o(1)$ for $s = \sqrt{M \log N}$ for any $M > 2$. Next, replacing σ_i with s_i, the result still holds, due to $\sigma_i > c$ and $\max_i |\sigma_i^2 - s_i^2| = o_P(1)$, by Lemma 2 . Q.E.D.

Lemma 2. *$\max_{i\leq N} \|\frac{1}{T}\sum_t u_{it}^m f_{kt}^n f_{qt}^v - \mathbb{E}u_{it}^m f_{kt}^n f_{qt}^v\| = O_P(\sqrt{\frac{\log N}{T}})$, for $m, n, v \in \{0,1,2\}$ for any $q, k \leq K$. Also, $\max_{ij} \|\frac{1}{T}\sum_t (u_{it}u_{jt} - \mathbb{E}u_{it}u_{jt})\| = O_P(\sqrt{\frac{\log N}{T}})$, $\max_{ijk} \|\frac{1}{T}\sum_t (u_{it}^2 u_{jt}u_{kt} - \mathbb{E}u_{it}^2 u_{jt}u_{kt})\| = O_P(\sqrt{\frac{\log N}{T}})$,*

and for $w_t = \frac{1}{\sqrt{N}}B'u_t$, $\max_{i\leq N} \|\frac{1}{T}\sum_t u_{it}^d w_t^d - \mathbb{E}u_{it}^d w_t^d\| = O_P(\sqrt{\frac{\log N}{T}})$ for $d \in \{1,2\}$

Proof. We apply Lemmas A.2 and A.3 of Chernozhukov et al. (2013b) to reach a concentration inequality: let $X_1, ..., X_T$ be independent in $\mathbb{R}^p$ where $p = N$ or N^2. Let $\sigma^2 = \max_i \mathbb{E}X_{it}^2$. Suppose $\mathbb{E}\max_{it} X_{it}^2 \log N \leq C\sigma^2 T$, then there is a universe constant $C > 0$, for any $x > 0$,

$$\max_{i\leq N} |\frac{1}{T}\sum_t X_{it} - \mathbb{E}X_{it}| \leq C\sigma\sqrt{\frac{\log N}{T}} + \frac{x}{T}$$

with probability at least $1 - \exp(-\frac{x^2}{3\sigma^2 T}) - CT\frac{\mathbb{E}\max_{i,t}|X_{it}|^4}{x^4}$. Now we set $x = \sigma\sqrt{T\log N}$. With the assumption that $(\log N)^4 = O(T)$, and $\mathbb{E}\max_{it} X_{it}^4 \leq \sigma^4(\log N)^2 TC$, we have, for any $\epsilon > 0$, there is C_ϵ, with probability at least $1-\epsilon$, $\max_{i\leq N} |\frac{1}{T}\sum_t X_{it}| \leq C_\epsilon \sigma\sqrt{\frac{\log N}{T}}$. The desired result then holds by respectively taking X_t as $u_{it}^m f_{kt}^n f_{qt}^v$, $u_{it}u_{jt}$ and $u_{it}w_t$.

B.4 Proof of Theorem 3

The proof is the same for all three cases of observing the factors. Let $\widehat{\epsilon}_i = \bar{r}_i - \widehat{\alpha}_0 - \widehat{\beta}_i'\widehat{\lambda}$, then $\sum_i \widehat{\epsilon}_i = 0$. Hence $\widehat{\alpha}_0 = \frac{1}{N}\sum_i \widehat{\alpha}_i$. From (B.18), we have

$$\widehat{\alpha}_0 - \bar{\alpha} = -\bar{\beta}' S_\beta^{-1} \frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha + \frac{1}{N}\sum_i \Delta_i.$$

In (B.18), we showed $\|\Delta\|_\infty = O_P(\frac{\log N}{T} + \frac{1}{N})$. In fact, the $\log N$ term arises from bounding uniform estimation errors, which can be avoided for $\frac{1}{N}\sum_i \Delta_i$. A more careful analysis could yield that $\frac{1}{N}\sum_i \Delta_i = O_P(\frac{1}{T} + \frac{1}{N})$. We omit details for brevity. For $1_N = (1, ..., 1)'$, $P_\beta = \beta(\beta'\beta)^{-1}\beta'$, $M_\beta = I - P_\beta$,

$$\widehat{\alpha}_0 - \alpha_0 - \frac{1}{N}\sum_i \Delta_i \quad = \quad \frac{1}{N}1_N'\alpha - \frac{1}{N}1_N'\beta(\frac{1}{N}\beta'\mathbb{M}_{1_N}\beta)^{-1}\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha - \alpha_0$$

$$
\begin{aligned}
&= \frac{1}{N}1_N'\alpha - (1_N'M_\beta 1_N)^{-1}1_N'P_\beta \mathbb{M}_{1_N}\alpha - \alpha_0 \\
&= \frac{1}{N}1_N'\alpha - (1_N'M_\beta 1_N)^{-1}1_N'P_\beta\alpha + (1_N'M_\beta 1_N)^{-1}1_N'P_\beta 1_N 1_N'\frac{1}{N}\alpha - \alpha_0 \\
&= (1_N'M_\beta 1_N)^{-1}1_N'M_\beta(\alpha - 1_N\alpha_0) \\
&= (1_N'M_\beta 1_N)^{-1}\sum_i(\alpha_i - \alpha_0)(1 - \bar{\beta}(\frac{1}{N}\beta'\beta)^{-1}\beta_i)
\end{aligned}
$$

where the second equality uses the Woodbury matrix identity for $(\frac{1}{N}\beta'\mathbb{M}_{1_N}B)^{-1}$. Hence by the Lindeburg It is easy to check that the triangular array Lindeburg condition holds, given $\mathbb{E}\alpha_i^4 < C$. Define

$$
\bar{\sigma}^2 = (\frac{1}{N}1_N'M_\beta 1_N)^{-1}\sigma_\alpha^2,
$$

then $\sqrt{N}\frac{\widehat{\alpha}-\alpha_0}{\bar{\sigma}} \xrightarrow{d} \mathcal{N}(0,1)$. The result then follows due to $s_0^2 - \bar{\sigma}^2 = o_P(1)$ and that $\bar{\sigma}^2 > 0$.