Internet Appendix for "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: July 5, 2020

Abstract

This appendix contains additional theoretical results and mathematical proofs, and a description of the data cleaning steps.

Contents

*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`.

A Asymptotic Theory

We present the formal asymptotic theory in this section. To begin with, we define the following notation. Suppose $A = (a_{ij})$ is an $n \times m$ matrix. We use $\psi_1(A) \geq ... \geq \psi_K(A)$ to denote the first K ordered singular values of a matrix A if $K \leq \min\{m, n\}$. We use $\psi_{\min}(A)$ and $\psi_{\max}(A)$ to denote its minimum and maximum eigenvalues. 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, we define $\|A\|_\infty = \max_{i \leq n} \sum_{j=1}^m |a_{ij}|$, and $\|A\|_n$ denotes the matrix nuclear norm. Finally, we define $\mathbb{M}_{1_N} = \mathrm{I}_N - 1_N 1_N'/N$, where $1_N = (1, ..., 1)'$ is a $N \times 1$ vector of ones.

A.1 Technical Assumptions

We start by describing and discussing the technical assumptions used for our asymptototic theory. We consider a general setting, where the DGP is given by (14), in which $(f_t, u_t, \alpha_i : i \leq N, t \leq T)$ are stochastic.

Assumption A.1. *There are constants $c, C > 0$, such that the following statements hold:*

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

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

(iii) Let $H_o = \frac{1}{T}\sum_{t=1}^T v_{l,t}v_{o,t}'(\frac{1}{T}\sum_{t=1}^T v_{o,t}v_{o,t}')^{-1}$. *Then for all* $s_t \in \{v_{l,t}, v_{l,t} - H_o v_{o,t}\}$,

$$c < \psi_K(\frac{1}{T}\sum_{t=1}^T s_t s_t') \leq ... \leq \psi_1(\frac{1}{T}\sum_{t=1}^T s_t s_t') < C.$$

In addition, for $S = (s_1, ..., s_T)$, the nonzero singular values of $\beta_l S$ are distinct.

Assumption A.1 is adopted by 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. In particular, condition (iii) is with respect to both the latent factors $s_t = v_{l,t}$ and the "transformed latent factors" $l_t = v_{l,t} - H_o v_{o,t}$, which is the essential latent factors in the case of both observed and latent factors are present. In this case, the essential latent factors are obtained by subtracting the effect observed factors: $H_o v_{o,t}$.

Assumption A.2. *The following statements hold:*

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

(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,*

$\max_{j\leq N}\sum_{i=1}^N |\mathbb{E}(u_{it}u_{jt}|f_t)| < C$, $\max_{j\leq N}\sum_{i=1}^N 1\{|\mathbb{E}(u_{it}u_{jt}|f_t)| \geq (\log N)^{-3}\} \leq CN^c$ *for some* $c > 0$, *and* $\max_{i,j\leq N}\sum_{k=1}^N |\operatorname{Cov}(u_{it}u_{kt}, u_{jt}u_{kt})| < C$.

Assumption A.2 imposes restrictions on the dependence structure of the DGP. We maintain 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}. This assumption is reasonable in that the idiosyncratic components should capture the remaining shocks and possible local factors after conditioning on the common risk factors.

Assumption A.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*

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

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* $w_t = \frac{1}{\sqrt{N}}\beta' u_t$.

(iii) There is $0 < L < 1$, *and a sequence* $B_{NT} > c$ *satisfying* $B_{NT}^2 \log(NT)^7 \leq T^L$, *such that*

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

where B_{NT} *may diverge.*

(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 $\frac{1}{N}\sum_{j=1}^N(\beta_j - \bar{\beta})(\beta_j - \bar{\beta})'$ *are bounded within* $[c, C]$.

Condition (ii) imposes that interchanging "max" with "$\mathbb{E}$" on $\xi_{i,j,k,l,t}$ imposes an additional term no larger than $O(T\log^2 N)$. It is a technical condition for applying concentration inequalities from Chernozhukov et al. (2013b) to establish

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

a key step to bound $\max_{i\le N}|\widehat{\alpha}_i-\alpha_i|$. In addition, condition (iv) imposes that $\mathbb{E}\|\frac{1}{\sqrt{N}}\beta' u_t\|^4<C$, which is reasonable given the cross-sectional weak correlations among u_{it}.

The above conditions allow for heavier tails than those of the sub-Gaussian distributions in the DGP, as our results only require moment conditions. That said, it is possible to further extend our assumptions to allow for even heavier tails, provided the use of Huber's loss function ((Huber, 1964)) and more robust estimators, see, e.g., Fan et al. (2016).

Assumption A.4 (Growing number of positive alphas)**.** *There is a growing sequence $L_{NT}\to\infty$, such that the true α satisfies*

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

Assumption A.4 requires there should be a growing number of true alternatives. This is needed to control the rate of false rejections and the same assumption is adopted by Liu and Shao (2014).

In a different context, Song and Zhao (2018) require the alphas of stock returns to be "sparse" in the sense that many entries should be nearly zero. However, this is not the case for hedge funds; our empirical studies indicate the presence of many nonzero alphas. For this reason, we do not require such a sparse structure.

Below we present the required assumption for the missing data case. We adopt the same notation as in Section 2.4.1. Let X be a general low rank matrix given by equation (16). Let (U_X, U_X^c) be the left singular-vectors of X, where columns of U_X and U_X^c correspond to the nonzero and zero singular values; let (V_X, V_X^c) be the right singular-vectors of X similarly defined. In addition, for any $N\times T$ matrix A, let

$$\mathcal{P}(A)=U_X^c U_X^{c\prime} A V_X^c V_X^{c\prime},\quad \mathcal{M}(A)=A-\mathcal{P}(A).$$

Here $\mathcal{M}(\cdot)$ can be regarded as the projection matrix onto the columns of U_X and V_X, and $\mathcal{P}(\cdot)$ is the projection onto its orthogonal space. Define the restricted low-rank set as, for some $c>0$,

$$\mathcal{G}(c_1,c_2)=\{N\times T \text{ matrix } A:\|\mathcal{P}(A)\|_n\le c\|\mathcal{M}(A)\|_n,\quad \frac{1}{NT}\|A\|_F^2\ge\frac{c_2}{\sqrt{NT}}\}.$$

Assumption A.5 (missing data)**.** *Let $\mathcal{N}_t=\{i: r_{it}$ is observed$\}$, $\mathcal{T}_i=\{t: r_{it}$ is observed$\}$, and $\omega_{it}=1\{r_{it}$ is observed$\}$. We assume the following conditions hold:*

(i) For any $(c_1,c_2)>0$ there is a constant $\kappa_c>0$ so that uniformly for all $A=(A_{it})_{N\times T}\in\mathcal{G}(c_1,c_2)$,

$$\sum_{i=1}^N\sum_{t=1}^T\omega_{it}A_{it}^2\ge\kappa_c\sum_{i=1}^N\sum_{t=1}^T A_{it}^2-O_P(N+T).$$

(ii) $\min_{i\le N}|\mathcal{T}_i|>c_0T$ *and* $\min_{t\le T}|\mathcal{N}_t|>c_0N$ *for some* $c_0>0$.

(iii) $\max_{t\leq T}\|\frac{1}{|\mathcal{N}_t|}\sum_{j\in\mathcal{N}_t}\beta_j\beta_j' - \frac{1}{N}\sum_{j\leq N}\beta_j\beta_j'\| = o(1)$.

(iv) $\{(v_t, \alpha_i, u_{it}) : i \leq N, t \leq T\}$ *is independent of* $\{\omega_{it} : i \leq N, t \leq T\}$*, and* $\mathbb{E}\omega_{it}$ *does not vary across both* (i, t).

(v) $\min_i \psi_{\min}(\frac{1}{T}\sum_{t=1}^{T}\omega_{it}v_t v_t') > c_0$.

(vi) $\frac{1}{N}\sum_{ij}|\mathrm{Cov}(\omega_{jt}, \omega_{it})| < C$, $\max_{it}\frac{1}{N}\sum_{j,k}|\mathrm{Cov}(u_{jt}u_{it}, u_{kt}u_{it})| < C$.

Condition (i) is the so-called "restricted strong convexity" condition, which is needed for matrix completions. Additionally, for technical reasons, in the presence of latent factors, we require the missing be both at random and homogeneous. Especially, the homogeneous missing substantially simplifies the technical arguments for the effects of the nuclear-norm regularized estimations. The inference for matrix completions using the penalized regression has been a challenging problem because the nuclear-norm estimator is known to be biased. As a technical contribution to the literature, in Algorithm 6 we provide a new inference procedure for debiasing the regularization, and achieve asymptotically normal estimators for the latent factors and alphas in the context of asset pricing. In proving the theoretical properties of these estimators, we apply an auxiliary leave-one-out argument recently used in Chen et al. (2019), which crucially requires the assumption of homogeneous missing. While we conjecture that the theoretical results might still hold in the presence of heterogeneous missing, as evidenced by our simulation studies, we leave the theoretical treatment for that case as an important open question.

A.1.1 When Observed Factors are Tradable

In this section, we consider cases when observed factors are all tradable. 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 $\lambda_o = \mathbb{E}f_{o,t}$.

In this case we propose the following algorithm:

Algorithm A.1 (Estimating α with tradable observable factors)**.**

S1. The same as S1 in Algorithm 6.

S2. Estimate the risk premia for latent factors.

$$\widehat{\lambda}_l = (\widehat{\beta}_l'\mathbb{M}_{1_N}\widehat{\beta}_l)^{-1}\widehat{\beta}_l'\mathbb{M}_{1_N}\bar{M}, \quad \bar{M} = (\bar{r}_i - \widehat{\beta}_{o,i}'\bar{f}_{o,i})_{N\times 1}.$$

S3. Estimate and de-bias the estimates of α:

When there are latent factors

$$\widehat{\alpha}_i \;=\; \bar{r}_i - \widehat{\beta}'_{o,i}\bar{f}_{o,i} - \widehat{\beta}'_{l,i}\widehat{\lambda}_l + \widehat{A}_i, \quad i = 1, ..., N,$$

where, writing $\widehat{\xi}'_i = e'_i - \widehat{\beta}'_{l,i}(\widehat{\beta}'_l \mathbb{M}_{1_N}\widehat{\beta}_l)^{-1}\widehat{\beta}'_l\mathbb{M}_{1_N}$, $\widehat{g}_i = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\widehat{v}'_{l,t}\widehat{\beta}_{l,i}$,

$$\widehat{A}_i = \widehat{\beta}'_{l,i}(\widehat{H}_{o,i} - \widehat{H}_o)\bar{f}_{o,i} - \widehat{\xi}'_i\widehat{g}.$$

When there are no latent factors

$$\widehat{\alpha}_i \;=\; \bar{r}_i - \widehat{\beta}'_{o,i}\bar{f}_{o,i}, \quad i = 1, ..., N.$$

Note that S2 is the key difference between Algorithm A.1 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 A.1, we only need to run cross-sectional OLS on the latent factor betas to estimate the risk premia for the latent factors.

For completeness, the algorithm also includes the observed factors-only case. Fund-by-fund time series regressions can be applied directly to estimate α_i. However, when testing $\alpha_i \leq 0$, the problem of conservativeness associated with testing inequality nulls is still present. As such we shall still apply the alpha-screening step for dimension reductions.

A.2 Main Theoretical Results

We now present the asymptotic distributions for estimated alphas. They arise from the following five scenarios:

(i) observable factors only (Algorithm 3);

(ii) latent factors only (Algorithm 4);

(iii) mixture of observable and latent factors (Algorithm 5);

(iv) mixture of observable and latent factors with an additional condition that observable factors are tradable (Algorithm A.1);

(v) observable factors only and they are all tradable.

Theorems A.1, A.3 and A.4 below apply to estimators that are obtained in any of these scenarios.

A.2.1 Expansion of Estimated Alphas

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

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

where $\sigma_{i,NT}^2 = \frac{1}{T}\operatorname{Var}(u_{it}(1 - v_t \Sigma_f^{-1}\lambda)) + \frac{1}{N}\chi_i$. *In scenarios (i)-(iii),* $\chi_i = \frac{1}{N}\operatorname{Var}(\alpha_i)\beta_i' S_\beta^{-1}\beta_i$ *and* $S_\beta = \frac{1}{N}\beta' \mathbb{M}_{1_N}\beta$; *in scenario (iv) that observable factors are tradable,* $\chi_i = \frac{1}{N}\operatorname{Var}(\alpha_i)\beta_{l,i}' S_{\beta,l}^{-1}\beta_{l,i}$, *and* $S_{\beta,l} = \frac{1}{N}\beta_l' \mathbb{M}_{1_N}\beta_l$; *in scenario (v) that only tradable observable factors are present,* $\chi_i = 0$.

Theorem A.1 is derived from a more general joint asymptotic expansion for the $N \times 1$ vector $\widehat{\alpha}$, given in Proposition B.1: in scenarios (i)-(iv):

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

In the above expansion, the first term is $O_P(T^{-1/2})$, and the second term is $O_P(N^{-1/2})$. The presence of the second term is the key reason of inconsistency $\widehat{\alpha}$ in the low dimension setting, see detailed discussion in Section A.3.3. This term vanishes as $N \to \infty$. However, if N grows too slowly, it could result in strong cross-sectional correlations among the estimated alphas due to the common component η_N, which would adversely affect the FDR control.

For this reason, in what follows, we require $T \log N = o(N)$, so that the term, $\beta\eta_N$, is negligible, and that the asymptotic distribution of $\widehat{\alpha}$ is characterized by $\frac{1}{T}\sum_t u_{it}(1 - v_t'\Sigma_f^{-1}\lambda)$. The t-statistics are therefore weakly correlated in the cross section.

A.2.2 Matrix Completion for Unbalanced Panel

In the presence of missing data, scenarios (ii)(iii)(iv) require estimating latent factors. Then a key step for matrix completion is to solve the following regularized regression:

$$\widehat{X} = \arg\min_M \|(Z - X) \circ \Omega\|^2 + \lambda_{NT}\|X\|_n, \tag{A.2}$$

for a given Z and λ_{NT}. We begin by introducing the following *singular value thresholding operator*: let $Y = UDV'$ be the singular value decomposition of a given matrix Y. Define

$$S_\nu(Y) := UD_\nu V',$$

where D_ν is defined by replacing the diagonal entry D_{ii} of D by $\max\{D_{ii} - \nu, 0\}$. Then as shown by Ma et al. (2011), the Karush-Kuhn-Tucker condition for $\widehat{X}$ is: for any $\tau > 0$,

$$\widehat{X} = S_\nu(\widehat{X} - \tau\Omega \circ (\widehat{X} - Z)), \quad \nu = \tau\lambda_{NT}/2.$$

This fact suggests a simple iterative algorithm to solve for $\widehat{X}$.

Algorithm A.2 (Solving the low-rank regularized problem for $\widehat{X}$)**.**

S1. Fix the "step size" $\tau \in (0,1)$. Let $\nu = \tau\lambda_{NT}/2$. Initialize X_0 and set $k = 0$.

S2. Let $X_{k+1} = S_\nu(X_k - \tau\Omega \circ (X_k - Z))$. Set k to $k + 1$.

S3. Repeat S2 until convergence.

This algorithm requires two tuning parameters (τ, λ_{NT}). As for λ_{NT}, let W be an $N \times T$ matrix whose columns are generated as $\mathcal{N}(0, \Sigma_u)$ independently across (i,t), where Σ_u is an $N \times N$ diagonal matrix of estimated individual variances of u_{it}. Let $Q(\|\Omega \circ W\|_2; 1-\delta)$ be the $1-\delta$ th quantile of $\|\Omega \circ W\|_2$, where $\|.\|_2$ denotes the matrix spectral norm. We follow the suggestion of Chernozhukov et al. (2018) by choosing $\lambda_{NT} = 2(1+c)Q(\|\Omega \circ W\|; 1-\delta)$. In practice, we set $\tau = 0.9$, $c = 0.1$, and $\delta = 0.05$.

Below we present the asymptotic results for the estimator given by all the five scenarios.

Theorem A.2. *Consider the case of unbalanced panel and all the five scenarios of observing factors. Suppose conditions of Theorem A.1 hold. Also, in the presence of latent factors (Scenarios (ii)(iii)(iv)), additionally assume Assumption A.5. Then uniformly in $i \le N$, when $T \log N = o(N)$ (which can be relaxed for scenario (iv)), the following results hold:*

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

A.2.3 FDR/FDP Control

Given the asymptotic properties of the estimated alphas achieved in Theorems A.1 and A.2, we are ready to establish the FDR/FDP control properties of the (alpha-screening) B-H procedure.

Theorem A.3. *In addition to conditions in Theorem A.1, suppose $T(\log N) = o(N)$, and Assumption A.4 holds. For the alpha estimators that arise from all five scenarios, the following results hold:*

(a) The B-H procedure satisfies:

$$\mathrm{FDR}_{B-H} \le \tau + o(1), \quad \mathrm{FDP}_{B-H} \le \tau + o_P(1).$$

As for the alpha-screening procedure, define $\sigma_i^2 = \mathrm{Var}(u_{it})\mathbb{E}(1-v_t'\Sigma_f^{-1}\lambda)^2$ and $\xi_{NT} = \log\log T\sqrt{\frac{\log N}{T}}$. Additionally assume $|\{i : -\xi_{NT}\sigma_i(1+\epsilon) < \alpha_i \le 0\}| \le |\{i : \alpha_i > -\xi_{NT}\sigma_i(1-\epsilon)\}|$ for some $\epsilon > 0$, where $|.|$ denotes the number of elements in the set. Then

$$\mathrm{FDR}_{\textit{screening B-H}} \le \tau + o(1), \quad \mathrm{FDP}_{\textit{screening B-H}} \le \tau + o_P(1).$$

(b) *Both the B-H and alpha-screening B-H procedures satisfy:*

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

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

(c) *Suppose* $\tau < \frac{1}{2}$. *Let* $\mathcal{G}_{B\text{-}H}$ *and* $\mathcal{G}_{screening\ B\text{-}H}$ *respectively be the numbers of correctly rejected alternatives by the B-H and screening B-H. We have:*

$$\mathbb{E}\frac{\mathcal{G}_{B\text{-}H}}{N} \leq \mathbb{E}\frac{\mathcal{G}_{screening\ B\text{-}H}}{N}.$$

In addition, define events:

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

Asymptotically, we have

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

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

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

For the screening approach, the additional condition says that the set $\{i : -\xi_{NT}\sigma_i(1+\epsilon) < \alpha_i \leq 0\}$ should not contain as many elements as $\{i : \alpha_i > -\xi_{NT}\sigma_i(1-\epsilon)\}$ does. This condition is equivalent to: the set $M := \{i : -\xi_{NT}\sigma_i(1+\epsilon) < \alpha_i < -\xi_{NT}\sigma_i(1-\epsilon)\}$ contains less alphas than the number of alternative alphas, which is a plausible assumption. We can take $\epsilon > 0$ be arbitrarily small so that $\xi_{NT}\sigma_i(1+\epsilon) \approx \xi_{NT}\sigma_i(1-\epsilon)$; also $\xi_{NT} \to 0$. So M is a set restricting on a very small range of negative α_i, and is often an empty set, while the number of alternative alphas grows to infinity. So it is indeed reasonable to assume M is smaller than the alternative set. This is a technical condition, needed to approximate the random sets $\widehat{\mathcal{I}}$ and $\widehat{\mathcal{I}} \cap \mathcal{H}_0$ by non-random sets $\{i : \alpha_i > -\xi_{NT}\sigma_i(1-\epsilon)\}$ and $\{i : -\xi_{NT}\sigma_i(1+\epsilon) < \alpha_i \leq 0\}$. The only requirement on the approximation is that the inequality $|\widehat{\mathcal{I}} \cap \mathcal{H}_0| \leq |\widehat{\mathcal{I}}|$ is preserved.

In addition, the usual B-H procedure focuses on the t-statistics formulated based on the "sample average" and its standard errors (Liu and Shao, 2014), while in our context, $\widehat{\alpha}_i$ is, approximately, 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.

As for the "power" property for detecting the significant alphas, note that Assumption A.4 ensures that for the true vector of α, there is a set $\mathcal{H} \subset \{1, ..., N\}$ so that

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

and $|\mathcal{H}| \to \infty$. Apparently $\mathbb{H}_0^i$ is false for all $i \in \mathcal{H}$. Theorem A.3 shows that we can correctly detect all positive alphas whose magnitudes are larger than $\sqrt{\frac{\log N}{T}}$.

To compare the power of the regular B-H procedure and the B-H with alpha-screening, we use the notation of "average power", denoted by $\mathbb{E}\mathcal{G}_{\text{B-H}}/N$ and $\mathbb{E}\mathcal{G}_{\text{screening B-H}}/N$, that is, the expected proportion of rejected false null hypothesis among the set of false null hypotheses. This definition is adopted from Benjamini and Liu (1999). In addition, we also adopt the notation of "family power", which is defined as the probability of rejecting all of the false null hypotheses, as in Lee and Whitmore (2002). For both definitions of power, the screening method improves the power of the usual B-H procedure.

We summarize the results in Theorem A.3:

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

(b) Our procedure can correctly identify all true alphas satisfying

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

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

(c) The alpha-screening B-H procedure more power than that of the regular B-H procedure.

(d) Unlike the B-H that tests all the alphas, the alpha-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 $\alpha_i \le 0$ (asymptotically).

A.2.4 Wild-Bootstrap

In this section, we prove that the wild-bootstrap algorithm delivers the desirable FDR control.

Theorem A.4. *Consider the case of unbalanced panel and all the five scenarios of observing factors. and $T \log N = o(N)$ (which can be relaxed for scenario (iv)). Then*

(a) uniformly in $i = 1, ..., N$,

$$\widehat{\alpha}_i^* = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i} \widehat{u}_{it}^*(1 - v_t'\Sigma_f^{-1}\lambda) + o_{P^*}(\frac{1}{\sqrt{T \log N}}),$$

where, $a_i, b_{1,it}$ and $b_{2,t}$ are as defined in Theorem A.2.

(b) Let $p_i^* = \frac{1}{B}\sum_{b=1}^B 1\{\widehat{\alpha}_i^* - \widehat{\alpha}_i\}$ *be the bootstrap p-value. Then* $\max_{i\leq N}|p_i^* - p_i| = o_P(1)$.

(c) Let $\mathrm{FDR}_{bootstrap}$ *and* $\mathrm{FDP}_{bootstrap}$ *be the* FDR *and* FDP *of applying the B-H procedure to the bootstrap p-values* p^**, we have*

$$\mathrm{FDR}_{bootstrap} \leq \tau + o(1), \quad \mathrm{FDP}_{bootstrap} \leq \tau + o_P(1).$$

A.3 Additional Theoretical Results

A.3.1 Identification of Alphas

We investigate the identification of α when both observable and latent factors are present. First, define

$$\begin{aligned}
\Gamma &= \mathbb{E}\left[(r_t - \mathbb{E}r_t)(f_{o,t} - \mathbb{E}f_{o,t})'\right]\mathrm{Cov}(f_{o,t})^{-1}, \\
Z_t &= r_t - \mathbb{E}r_t - \Gamma(f_{o,t} - \mathbb{E}f_{o,t}), \quad t = 1, ..., T.
\end{aligned}$$

Both are identified quantities given the observables $\{(r_t, f_{o,t}) : t = 1, ..., T\}$. In addition, define

$$\mathcal{T}(\beta) := \beta(\beta'\mathbb{M}_{1_N}\beta)^{-1}\beta'.$$

Note that $\mathcal{T}$ is rotation invariant, in the sense that $\mathcal{T}(\beta H) = \mathcal{T}(\beta)$ for any invertible matrix H. We show that α is identified by the following system of equations in the next theorem.

Theorem A.5. *Consider the case when both* $(f_{o,t}, f_{l,t})$ *are present. There are latent invertible matrices* Q, H*, and a latent* $\dim(g_t)$*-vector* h_t*, so that equations (A.3)- (A.6) hold, where*

$$Z_t = \beta_l h_t + u_t, \tag{A.3}$$

$$\beta H = (\Gamma, \beta_l Q), \tag{A.4}$$

$$\beta\lambda = \mathcal{T}(\beta H)\mathbb{M}_{1_N}\mathbb{E}r_t - \mathcal{T}(\beta H)\mathbb{M}_{1_N}\alpha, \tag{A.5}$$

$$\alpha = \mathbb{E}r_t - \beta\lambda. \tag{A.6}$$

In view of the relation between the above system of equations and Algorithm 5, we note the following observations:

1. The identified components (Γ, Z_t) are the population counterparts of $(\widehat{\beta}_o, Z)$ obtained in Step S1a.

2. Equation (A.3) shows that Z_t admits a factor structure, with β_l as the factor loadings. It is well known that in this case there is a rotation matrix Q, so that $\frac{1}{\sqrt{N}}\beta_l Q$ is identified as the first K_l eigenvectors $\mathbb{E}Z_tZ_t'$. Therefore, $\beta_l Q$ is the population counterpart of $\widehat{\beta}_l$ obtained in Step S1b.

3. Equation (A.4) shows that β is identified up to a rotation H, given that $(\Gamma, \beta_l Q)$ are both identified. In fact $(\Gamma, \beta_l Q)$ is the population counterpart of $\widehat{\beta}$ obtained in Step S1.

4. $(\alpha, \beta\lambda)$ are then identified (as $N \to \infty$) through equations (A.5), (A.6) given the identification of βH. In particular, $\mathcal{T}(\beta H)\mathbb{M}_{1_N}\mathbb{E}r_t$ is the population counterpart of

$$\widehat{\beta\lambda} = \mathcal{T}(\widehat{\beta})\mathbb{M}_{1_N}\bar{r},$$

whereas $\mathcal{T}(H\beta)\mathbb{M}_{1_N}\alpha$ in (A.5) converges to zero as $N \to \infty$.

A.3.2 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 $\widehat{\alpha}_0 = \frac{1}{N}\sum_{i=1}^N \widehat{\alpha}_i$. Additionally, let $\widehat{\sigma}_\alpha^2 = \frac{1}{N}\sum_i(\widehat{\alpha}_i - \widehat{\alpha}_0)^2$.

Theorem A.6. *Consider a general case where unbalanced panel is allowed. Let $\sigma_\alpha^2 > 0$ denote the cross-sectional variance. Suppose $\mathbb{E}\alpha_i^4 < C$ and β is deterministic. Assumptions A.1-A.5 hold, but the condition $T\log N = o(N)$ can be relaxed. Then*

$$\sqrt{N}\frac{\widehat{\alpha}_0 - \alpha_0}{s_0} \xrightarrow{d} \mathcal{N}(0,1), \quad \textit{where}$$

for scenarios (i)-(iii), suppose in addition $\liminf(1 - \bar{\beta}'(\frac{1}{N}\beta'\beta)^{-1}\bar{\beta})^2 > 0$ *and* $N = o(T^2)$,

$$s_0^2 = (1 - \widehat{\bar{\beta}}'(\frac{1}{N}\widehat{\beta}'\widehat{\beta})^{-1}\widehat{\bar{\beta}})^{-1}\widehat{\sigma}_\alpha^2, \quad \widehat{\bar{\beta}} = \frac{1}{N}\sum_i \widehat{\beta}_i;$$

for scenario (iv), suppose $\liminf(1 - \bar{\beta}_l'(\frac{1}{N}\beta_l'\beta_l)^{-1}\bar{\beta}_l)^2 > 0$ *and* $N = o(T^2)$,

$$s_0^2 = (1 - \widehat{\bar{\beta}}_l'(\frac{1}{N}\widehat{\beta}_l'\widehat{\beta}_l)^{-1}\widehat{\bar{\beta}}_l)^{-1}\widehat{\sigma}_\alpha^2, \quad \widehat{\bar{\beta}}_l = \frac{1}{N}\sum_i \widehat{\beta}_{l,i};$$

for scenario (v), $s_0^2 = \widehat{\sigma}_\alpha^2$.

A.3.3 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. As we shall formally show below, when the dimension N is fixed, the two-pass regression method fails to consistently estimate *any* alpha, so it cannot be used in the FDR control or any multiple testing problems.

We shall focus on the case of balanced panel, and all factors are observable but not tradable.

Theorem A.7 (Inconsistent Estimation of α)**.** *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. Then we have for each $i \le N$, as long as $\beta_i \neq 0$, 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 $\eta_N = \frac{1}{N}S_\beta^{-1}\beta'\mathbb{M}_{1_N}\alpha$.

In many asset pricing contexts, the common goal is to test the null hypothesis: $\mathbb{H}_0$: all alphas are zero, see, e.g., Gibbons et al. (1989). The two-pass regression is consistent for alphas when N is fixed, because the null hypothesis is imposed. Under such null, $\mathrm{Var}(\alpha_i) = 0$ so $X_i = 0$ for all $i \le N$ in the above proposition. However, as long as there are at least one alpha that is nonzero, it holds that $\mathrm{Var}(\alpha_i) > 0$, then $\widehat{\alpha}_i$ would be inconsistent whenever $\beta_i \neq o(1)$ for that specific i.

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 with N or T.

B.1 Proof of Theorem A.1

Proof. By Proposition B.1, $\widehat{\alpha}_i - \alpha_i = \frac{1}{T}\sum_t u_{it}(1 - v_t'\Sigma_f^{-1}\lambda) - \frac{1}{N}\beta_i'S_\beta^{-1}\beta'\mathbb{M}_{1_N}\alpha + O_P(\frac{\log N}{T} + \frac{1}{N})$. 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 A.2

Proof. The proofs are similar throughout scenarios (i)-(iv). The case of scenario (iii) mixture of observable and latent factors, in the presence of missing data where we apply the matrix completion algorithm, is most challenging. Therefore, we mainly focus on the the proof of scenario (iii) below and briefly mention the proof of all other cases.

Scenarios (i)(v): observable factors only

When there are only observable factors, the matrix completion algorithm is not required. We apply fund-by-fund time series regressions. When factors are tradable, then the estimated intercept would then be the estimated alphas; when factors are not tradable, we additionally apply cross sectional regression to estimate the factor risk premia and alphas. The details are well known even for unbalanced data and are therefore omitted.

Scenario (ii): latent factors only

For any i, we have the following factor model: for $l_{it} = v_{l,t} - \bar{v}_{l,i}$ and $l_t = v_{l,t} - \bar{v}_l$,

$$r_{it} - \bar{r}_i = \beta_{l,i}' l_t + u_{it} - \bar{u}_i + \beta_{l,i}'(\bar{v}_l - \bar{v}_{l,i}).$$

To apply the matrix completion algorithm and Proposition B.2, we set $\kappa_{it} = 0$ and $g_{it} = \beta_{l,i}'(\bar{v}_{l,i} - \bar{v}_l)$. We need to verify $\max_i \frac{1}{T}\sum_t g_{it}^2 = o_P(\frac{1}{\sqrt{T \log N}})$ and $\frac{1}{T}\sum_t \|\frac{1}{N}\sum_j \omega_{jt} g_{jt} \beta_j\|^2 = o_P(\frac{1}{T \log N})$. The former is straightforward. Verifying the latter is very similar to that in scenario (iii) proved below.

Hence by Proposition B.2, there is H_l,

$$\widehat{\beta}_l - \beta_l H_l = J_l + o_P(\frac{1}{\sqrt{T \log N}}), \quad J_{l,i}' = \frac{1}{T_i}\sum_{t \in \mathcal{T}_i} u_{it} l_t' \bar{S}_l^{-1} H_l,$$

where $\bar{S}_l = \frac{1}{T}\sum_t l_t l_t'$, and $o_P(\frac{1}{\sqrt{T \log N}})$ is in the $\|.\|_\infty$ norm. The expansion for $\widehat{\lambda}$ would be the same as scenario (iii) below (and is much easier in the latent factor only case), leading to

$$\bar{r}_i - \widehat{\beta}_i' \widehat{\lambda} - \alpha_i \quad = \quad B_i + \bar{u}_i - (\Upsilon_i + \Xi_i)' H^{-1} \lambda - \beta_i' S_\beta^{-1} \frac{1}{N} \beta' \mathbb{M}_{1_N} \alpha + o_P(\frac{1}{\sqrt{T \log N}}),$$

where $B_i = \widehat{\beta}_i'[H^{-1}\bar{v}_i - (\widehat{\beta}' \mathbb{M}_{1_N} \widehat{\beta})^{-1} \widehat{\beta}' \mathbb{M}_{1_N} g]$. But the main difference here is that $\Upsilon_i = \Xi_i = 0$ in the absence of observable factors. Here $\widehat{\beta} := \widehat{\beta}_l$ and $\bar{v}_i := \bar{v}_{l,i}$ and $g_i = \beta_{l,i}' \bar{v}_{l,i}$. The bias correction effect $\widehat{B}_i - B_i$ would be the same. Unlike scenario (iii), there is no additional bias $H_{o,i} - H_o$ in the absence of observable factors. So the final estimator is just

$$\widehat{\alpha}_i = \bar{r}_i - \widehat{\beta}_{l,i}' \widehat{\lambda}_i - \widehat{B}_i,$$

which satisfies

$$\widehat{\alpha}_i - \alpha_i = \frac{1}{T_i}\sum_{t \in \mathcal{T}_i} u_{it}(1 - v_t' \Sigma_f^{-1} \lambda) - \beta_i' S_\beta^{-1} \frac{1}{N} \beta' \mathbb{M}_{1_N} \alpha + o_P(\frac{1}{\sqrt{T \log N}}).$$

Here $v_t := v_{l,t}$, $\Sigma_f = \mathrm{Cov}(v_{l,t})$ and $\lambda := \lambda_l$.

Scenario (iii): mixture of observable and latent factors

Step 1. estimate beta. Define $\bar{v}_{o,i} = \frac{1}{T_i}\sum_{t \in \mathcal{T}_i} v_{o,t}$, $\bar{v}_{l,i} = \frac{1}{T_i}\sum_{t \in \mathcal{T}_i} v_{l,t}$, and $S_{o,i} = \frac{1}{T_i}\sum_{t \in \mathcal{T}_i} v_{o,t} v_{o,t}'$. Then Lemma B.4 implies $\max_i \|\bar{v}_{o,i}\| = O_P(\sqrt{\frac{\log N}{T}}) = \max_i \|\bar{v}_{l,i}\|$ and $\max_i \|S_{o,i}^{-1}\| = O_P(1)$.

The first step OLS gives, for $H_{o,i} = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i} v_{l,t}v_{o,t}'S_{o,i}^{-1}$, and $H_o = \frac{1}{T}\sum_t v_{l,t}v_{o,t}'S_o^{-1}$,

$$\widehat{\beta}_{o,i} - \beta_{o,i} = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i} u_{it}S_{o,i}^{-1}v_{o,t} + H_{o,i}'\beta_{l,i} + \delta_i, \quad \max_i \|\delta_i\| = o_P(\frac{1}{\sqrt{T\log N}}). \tag{B.7}$$

Now let $l_{it} = (v_{l,t} - \bar{v}_{l,i}) - H_{o,i}(v_{o,t} - \bar{v}_{o,i})$, and $l_t = (v_{l,t} - \bar{v}_l) - H_o(v_{o,t} - \bar{v}_o)$. Then for any i, we have the following factor model: for $t \in \mathcal{T}_i$, and $z_{it} = r_{it} - \bar{r}_i - \widehat{\beta}_{o,i}(f_{o,t} - \bar{f}_{o,i})$,

$$z_{it} = \beta_{l,i}'l_t - \eta_{it} + u_{it} - \bar{u}_i + \beta_{l,i}'(l_{it} - l_t) - \delta_i'(v_{o,t} - \bar{v}_{o,i}), \quad \eta_{it} = \frac{1}{T_i}\sum_{s\in\mathcal{T}_i} u_{is}v_{o,s}'S_{o,i}^{-1}(v_{o,t} - \bar{v}_{o,i}).$$

To apply the matrix completion result in Proposition B.2, we set $\kappa_{it} = -\delta_i'(v_{o,t} - \bar{v}_{o,i})$ and $g_{it} = \beta_{l,i}'(l_{it} - l_t)$. We need to verify $\max_i \frac{1}{T}\sum_t g_{it}^2 = o_P(\frac{1}{\sqrt{T\log N}})$ and $\frac{1}{T}\sum_t \|\frac{1}{N}\sum_j \omega_{jt}g_{jt}\beta_j\|^2 = o_P(\frac{1}{T\log N})$. The former is straightforward. As for the latter,

$$\begin{aligned}
&\frac{1}{T}\sum_t \|\frac{1}{N}\sum_i \omega_{it}g_{it}\beta_i\|^2 \le \frac{1}{T}\sum_t \|\frac{1}{N}\sum_i \omega_{it}\beta_{l,i}'(\bar{v}_{l,i} - \bar{v}_l)\beta_i\|^2 + \frac{1}{T}\sum_t \|\frac{1}{N}\sum_i \omega_{it}\beta_{l,i}'(H_{o,i} - H_o)\beta_i\|^2 \\
&+\frac{1}{T}\sum_t \|\frac{1}{N}\sum_i \omega_{it}\beta_{l,i}'(H_{o,i} - H_o)\bar{v}_{o,i}\beta_i\|^2 + \frac{1}{T}\sum_t \|\frac{1}{N}\sum_i \omega_{it}\beta_{l,i}'H_o(\bar{v}_{o,i} - \bar{v}_o)\beta_i\|^2 \\
= \;& o_P(\frac{1}{T\log N}),
\end{aligned}$$

which follows from lemma B.4. Hence by Proposition B.2, there exists H_l, such that

$$\widehat{\beta}_l - \beta_l H_l = J_l + J_{2l} + o_P(\frac{1}{\sqrt{T\log N}}), \quad J_{l,i}' = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}(u_{it} - \eta_{it})l_t'\bar{S}_l^{-1}H_l, \quad J_{2l,i}' = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\beta_{l,i}'(l_{it} - l_t)l_t'\bar{S}_l^{-1}H_l,$$

where $\bar{S}_l = \frac{1}{T}\sum_t l_t l_t'$, and $o_P(\frac{1}{\sqrt{T\log N}})$ is in the $\|.\|_\infty$ norm. Hence

$$\begin{aligned}
\widehat{\beta} &= (\widehat{\beta}_o, \widehat{\beta}_l) = \beta H + \Upsilon + \Xi + o_P(\frac{1}{\sqrt{T\log N}}), \quad H = \begin{pmatrix} I & 0 \\ H_o & H_l \end{pmatrix}, \\
\Upsilon_i' &= (\frac{1}{T_i}\sum_{t\in\mathcal{T}_i} u_{it}v_{o,t}'S_{o,i}^{-1}, J_{l,i}'), \quad \Xi_i' = (\beta_{li}'(H_{o,i} - H_o), J_{2l,i}').
\end{aligned}$$

Also $\max_i \|\widehat{\beta}_i - H'\beta_i\| = o_P(1)$.

Step 2. estimate factor risk premium.

First, from Lemma B.4, $\|\frac{1}{N}\sum_i \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}(l_{it} - l_t)l_t'b_i\| = o_P(\frac{1}{\sqrt{T\log N}})$. Hence for any deterministic bounded sequence b_i, $\frac{1}{N}\sum_i b_i J_{2l,i} = o_P(\frac{1}{\sqrt{T\log N}}) = \frac{1}{N}\sum_i b_i\beta_{l,i}'(H_{o,i} - H_o)$. It is also straightforward to see $\frac{1}{N}\sum_i b_i J_{l,i} = o_P(\frac{1}{\sqrt{T\log N}})$. This implies

$$\frac{1}{N}\sum_i b_i(\widehat{\beta}_i - H'\beta_i) = o_P(\frac{1}{\sqrt{T\log N}}).$$

Thus $\widehat{\lambda} = (\widehat{\beta}'\mathbb{M}_{1_N}\widehat{\beta})^{-1}(\widehat{\beta}'\mathbb{M}_{1_N}\bar{r})$ implies, for g be the $N\times 1$ vector of $(\beta_i'\bar{v}_i)$,

$$\widehat{\lambda} - H^{-1}\lambda = (\widehat{\beta}'\mathbb{M}_{1_N}\widehat{\beta})^{-1}\widehat{\beta}'\mathbb{M}_{1_N}g + S_\beta^{-1}\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha + o_P(\frac{1}{\sqrt{T\log N}}).$$

So $\widehat{\beta}_i'\widehat{\lambda}-\beta_i'\lambda=(\Upsilon_i+\Xi_i)'H^{-1}\lambda+\beta_i'H(\widehat{\beta}'\mathbb{M}_{1_N}\widehat{\beta})^{-1}\widehat{\beta}'\mathbb{M}_{1_N}g+o_P(\frac{1}{\sqrt{T\log N}})$. Then

$$\bar{r}_i-\widehat{\beta}_i'\widehat{\lambda}-\alpha_i \;=\; B_i+\bar{u}_i-(\Upsilon_i+\Xi_i)'H^{-1}\lambda-\beta_i'S_\beta^{-1}\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha+o_P(\frac{1}{\sqrt{T\log N}}),$$

where $B_i=\widehat{\beta}_i'[H^{-1}\bar{v}_i-(\widehat{\beta}'\mathbb{M}_{1_N}\widehat{\beta})^{-1}\widehat{\beta}'\mathbb{M}_{1_N}g]$. In addition, λ_o can be consistently estimated due to:

$$H^{-1}=\begin{pmatrix} I & 0\\ -H_l^{-1}H_o & H_l^{-1}\end{pmatrix},\quad H^{-1}\lambda=\begin{pmatrix}\lambda_o\\ H_l^{-1}(\lambda_l-H_o\lambda_o)\end{pmatrix}.$$

We now work with $(\Upsilon_i+\Xi_i)'H^{-1}\lambda$. By the same argument as of (B.30), for $h=\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}(u_{it}-\eta_{it})l_t'\bar{S}_l^{-1}$, we have

$$\begin{aligned}\Upsilon_i'H^{-1}\lambda &= (\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}u_{it}v_{o,t}'S_{o,i}^{-1}-hH_o,h)\lambda+o_P(\frac{1}{\sqrt{T\log N}})\\ &= \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}u_{it}v_t'\Sigma_f^{-1}\lambda+o_P(\frac{1}{\sqrt{T\log N}}).\end{aligned}$$

For $\Xi_i'H^{-1}\lambda$, note that $\bar{S}_l=S_{ll}-S_{ol}'S_o^{-1}S_{ol}+O_P(T^{-1/2})$,

$$\begin{aligned}\Xi_i'H^{-1}\lambda &= \beta_{li}'(H_{o,i}-H_o)\lambda_o+\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\beta_{l,i}'(l_{it}-l_t)l_t'\bar{S}_l^{-1}(\lambda_l-H_o\lambda_o)\\ &= C_i+o_P(\frac{1}{\sqrt{T\log N}}),\quad \text{where}\\ C_i &= \beta_{li}'(H_{o,i}-H_o)\lambda_o.\end{aligned}$$

For the last equality, recall $l_{it}=(v_{l,t}-\bar{v}_{l,i})-H_{o,i}(v_{o,t}-\bar{v}_{o,i})$ and $l_t=(v_{l,t}-\bar{v}_l)-H_o(v_{o,t}-\bar{v}_o)$.

$$\begin{aligned}&\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\beta_{l,i}'(l_{it}-l_t)l_t'\bar{S}_l^{-1}(\lambda_l-H_o\lambda_o)\\ =\;& -\beta_{l,i}'(H_{o,i}-H_o)\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}v_{ot}l_t'\bar{S}_l^{-1}(\lambda_l-H_o\lambda_o)+o_P(\frac{1}{\sqrt{T\log N}})\\ =\;& \beta_{l,i}'(H_{o,i}-H_o)(S_oH_o'-S_{ol})\bar{S}_l^{-1}(\lambda_l-H_o\lambda_o)+o_P(\frac{1}{\sqrt{T\log N}})=o_P(\frac{1}{\sqrt{T\log N}}),\end{aligned}$$

where we used $H_o=S_{ol}'S_o^{-1}$. This implies

$$\bar{r}_i-\widehat{\beta}_i'\widehat{\lambda}-\alpha_i=\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}u_{it}(1-v_t'\Sigma_f^{-1}\lambda)-\beta_i'S_\beta^{-1}\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha-C_i+B_i+o_P(\frac{1}{\sqrt{T\log N}}).$$

Step 3. bias correction. We now respectively estimate C_i and B_i.

For all $t\le T$, we have $\widehat{v}_t=(f_{o,t}-\bar{f}_o,\widehat{v}_{l,t})$, then it follows that

$$\widehat{v}_t-H^{-1}(v_t-\bar{v})=\widehat{v}_t-(f_{o,t}'-\bar{f}_o',(H_l^{-1}l_t)')'=(0',(\widehat{v}_{l,t}-H_l^{-1}l_t)')'.$$

It follows from Lemma B.4 that $\max_i \|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\widehat v_t - H^{-1}(v_t-\bar v)\| = o_P(\frac{1}{\sqrt{T\log N}})$. Recall that $\widehat g_i = \widehat\beta_i'\widehat{\bar v}_i$ and $g_i=\beta_i'\bar v_i$. So uniformly in i, for any bounded deterministic sequence $b=(b_1,...,b_N)'$,

$$\frac{1}{N}b'(\widehat g-(g-\beta\bar v)) = \frac{1}{N}\sum_i b_i(\widehat\beta_i - H'\beta_i)'\widehat{\bar v}_i + \frac{1}{N}\sum_i b_i\beta_i'H(\widehat{\bar v}_i - H^{-1}(\bar v_i-\bar v)) = o_P(\frac{1}{\sqrt{T\log N}}).$$

Now define $\widehat B_i := \widehat\beta_i'[\widehat{\bar v}_i - (\widehat\beta'\mathbb{M}_{1_N}\widehat\beta)^{-1}\widehat\beta'\mathbb{M}_{1_N}\widehat g] = \widehat\xi_i'\widehat g$. Uniformly in $i\le N$,

$$\begin{aligned}\widehat B_i - B_i &= \widehat\beta_i'[\widehat{\bar v}_i - H^{-1}\bar v_i - (\widehat\beta'\mathbb{M}_{1_N}\widehat\beta)^{-1}\widehat\beta'\mathbb{M}_{1_N}(\widehat g - g)]\\ &= \widehat\beta_i'[\widehat{\bar v}_i - H^{-1}(\bar v_i-\bar v) - (\widehat\beta'\mathbb{M}_{1_N}\widehat\beta)^{-1}\widehat\beta'\mathbb{M}_{1_N}(\widehat g-(g-\beta\bar v))\\ &\quad +(\widehat\beta'\mathbb{M}_{1_N}\widehat\beta)^{-1}\widehat\beta'\mathbb{M}_{1_N}(\beta H-\widehat\beta)H^{-1}\bar v] = o_P(\frac{1}{\sqrt{T\log N}}).\end{aligned}$$

As for $C_i := \beta_{li}'(H_{o,i}-H_o)\lambda_o$, we define,

$$\widehat H_{o,i} = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\widehat v_{l,t}(f_{o,t}-\bar f_{o,i})'\widehat S_{o,i}^{-1}, \quad \widehat H_o = \frac{1}{T}\sum_t \widehat v_{l,t}(f_{o,t}-\bar f_o)'\widehat S_o^{-1},$$

where $\widehat S_o$ and $\widehat S_{o,i}$ are respectively defined using $(f_{o,t}-\bar f_o)$ and $(f_{o,t}-\bar f_{o,i})$. The goal is to show $\max_i\|\widehat C_i - C_i\| = o_P(\frac{1}{\sqrt{T\log N}})$, where $\widehat C_i = \widehat\beta_{li}'(\widehat H_{o,i}-\widehat H_o)\widehat\lambda_o$. Note

$$\begin{aligned}\widehat C_i - C_i &= (\widehat\beta_{l,i}-H_l'\beta_{l,i})'(\widehat H_{o,i}-\widehat H_o)\widehat\lambda_o + \beta_{l,i}'H_l(\widehat H_{o,i}-\widehat H_o)(\widehat\lambda_o-\lambda_o)\\ &\quad +\beta_{l,i}'H_l[(\widehat H_{o,i}-\widehat H_o) - H_l^{-1}(H_{o,i}-H_o)]\lambda_o.\end{aligned}$$

It is straightforward to show the first two terms are $o_P(\frac{1}{\sqrt{T\log N}})$ uniformly in i. The most challenging term is the third one. To analyze it, we first show

$$H_{o,i}-H_o = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i} l_t v_{o,t}'S_{o,i}^{-1} - \frac{1}{T}\sum_{t=1}^T l_t v_{o,t}'S_o^{-1} + o_P(\frac{1}{\sqrt{T\log N}}),$$

where we recall $l_t = v_{l,t}-\bar v_l - H_o(v_{o,t}-\bar v_o)$. Note that $\frac{1}{T_i}\sum_{t\in\mathcal{T}_i} l_t v_{o,t}'S_{o,i}^{-1} = H_{o,i} - \bar v_l\bar v_{o,i}'S_{o,i}^{-1} - H_o + H_o\bar v_o\bar v_{o,i}'S_{o,i}^{-1}$ and $\frac{1}{T}\sum_t l_t v_{o,t}'S_o^{-1} = -\bar v_l\bar v_o'S_o^{-1} + H_o\bar v_o\bar v_o'S_o^{-1}$. Hence

$$\frac{1}{T_i}\sum_{t\in\mathcal{T}_i} l_t v_{o,t}'S_{o,i}^{-1} - \frac{1}{T}\sum_{t=1}^T l_t v_{o,t}'S_o^{-1} - [H_{o,i}-H_o] = -\bar v_l\bar v_{o,i}'(S_{o,i}^{-1}-S_o^{-1}) + H_o\bar v_o(\bar v_{o,i}'S_{o,i}^{-1} - \bar v_o'S_o^{-1}),$$

where the right hand side is $o_P(\frac{1}{\sqrt{T\log N}})$. So to show $\max_i\|\widehat C_i - C_i\| = o_P(\frac{1}{\sqrt{T\log N}})$, it suffices to show $\max_i\|(\widehat H_{o,i}-\widehat H_o) - H_l^{-1}[\frac{1}{T_i}\sum_{t\in\mathcal{T}_i} l_t v_{o,t}'S_{o,i}^{-1} - \frac{1}{T}\sum_{t=1}^T l_t v_{o,t}'S_o^{-1}]\| = o_P(\frac{1}{\sqrt{T\log N}})$, which is also sufficient to show

$$\max_i\|\widehat H_{o,i} - H_l^{-1}\frac{1}{T_i}\sum_{t\in\mathcal{T}_i} l_t v_{o,t}'S_{o,i}^{-1}\| = o_P(\frac{1}{\sqrt{T\log N}}).$$

This in fact immediately follows from Lemma B.4 that both $\max_i\|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}(\widehat v_{l,t}(f_{o,t}-\bar f_{o,i})' - H_l^{-1}l_t v_{o,t}')\|$ and $\max_i\|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}((f_{o,t}-\bar f_{o,i})(f_{o,t}-\bar f_{o,i})' - v_{o,t}v_{o,t}')\|$ are $o_P(\frac{1}{\sqrt{T\log N}})$.

Therefore, for $\widehat{\alpha}_i := \bar{r}_i - \widehat{\beta}_i'\widehat{\lambda} + \widehat{C}_i - \widehat{B}_i$,

$$\widehat{\alpha}_i - \alpha_i = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i} u_{it}(1 - v_t'\Sigma_f^{-1}\lambda) - \beta_i' S_\beta^{-1}\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha + o_P(\frac{1}{\sqrt{T\log N}}). \tag{B.8}$$

Scenario (iv): mixture of observable and latent factors and observable factors are tradable

The only difference from Scenario (iii) is that, $\lambda_o = \mathbb{E} f_{o,t}$, and λ_l is estimated by, for $\bar{A} = (\bar{r}_i - \widehat{\beta}_{o,i}'\bar{f}_{o,i})$ be an $N \times 1$ vector,

$$\widehat{\lambda}_l = (\widehat{\beta}_l'\mathbb{M}_{1_N}\widehat{\beta}_l)^{-1}\widehat{\beta}_l'\mathbb{M}_{1_N}\bar{A},$$

and

$$\widehat{\alpha}_i = \bar{r}_i - \widehat{\beta}_o'\bar{f}_{o,i} - \widehat{\beta}_l'\widehat{\lambda}_l - \text{estimated bias}.$$

Similar to the proof of Scenario (iii), we have

$$\begin{aligned}
\widehat{\beta} &= (\widehat{\beta}_o, \widehat{\beta}_l) = \beta H + \Upsilon + \Xi + o_P(\frac{1}{\sqrt{T\log N}}), \quad H = \begin{pmatrix} I & 0 \\ H_o & H_l \end{pmatrix}, \\
\Upsilon_i' &= \underbrace{(\frac{1}{T_i}\sum_{t\in\mathcal{T}_i} u_{it}v_{o,t}'S_{o,i}^{-1}}_{\Upsilon_{1i}'}, J_{l,i}'), \quad \Xi_i' = (\underbrace{\beta_{li}'(H_{o,i} - H_o)}_{\Xi_{1i}'}, J_{2l,i}'),
\end{aligned}$$

and for $C_i = \beta_{li}'(H_{o,i} - H_o)\lambda_o$, $(\Upsilon_i + \Xi_i)'H^{-1}\lambda = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i} u_{it}v_t'\Sigma_f^{-1}\lambda + C_i + o_P(\frac{1}{\sqrt{T\log N}})$. So

$$\widehat{\beta}_{o,i}' = \beta_{o,i}' + \beta_{l,i}'H_o + \Upsilon_{1i}' + \Xi_{1i}' + o_P(\frac{1}{\sqrt{T\log N}}),$$

and $\widehat{\beta}_{o,i}'\bar{f}_{o,i} = \beta_{o,i}'\bar{f}_{o,i} + \beta_{l,i}'H_o\bar{f}_{o,i} + (\Upsilon_{1i}' + \Xi_{1i}')\lambda_o + o_P(\frac{1}{\sqrt{T\log N}})$.

Also note that $\bar{r}_i = \alpha_i + \beta_{o,i}'\bar{f}_{o,i} + \beta_{l,i}'\lambda_l + \beta_{li}'\bar{v}_{l,i} + \bar{u}_i$, so for $g_l = (\beta_{li}'(\bar{v}_{l,i} - H_o\bar{v}_{o,i}) : i \le N)$,

$$\bar{A} = \alpha + \beta_l(\lambda_l - H_o\lambda_o) + g_l + \bar{u} - (\Upsilon_1 + \Xi_1)\lambda_o + o_P(\frac{1}{\sqrt{T\log N}}).$$

We also note from Lemma B.4 (iv) that for any deterministic and bounded sequence $b := (b_i : i \le N)$, $\|\frac{1}{N}\sum_{i=1}^N b_i(H_{o,i} - H_o)\| = o_P(\frac{1}{\sqrt{T\log N}})$. Hence $\frac{1}{N}b'(\Upsilon_1 + \Xi_1) = o_P(\frac{1}{\sqrt{T\log N}})$. So

$$\widehat{\lambda}_l = H_l^{-1}(\lambda_l - H_o\lambda_o) + (\widehat{\beta}_l'\mathbb{M}_{1_N}\widehat{\beta}_l)^{-1}\widehat{\beta}_l'\mathbb{M}_{1_N}g_l + (\widehat{\beta}_l'\mathbb{M}_{1_N}\widehat{\beta}_l)^{-1}\widehat{\beta}_l'\mathbb{M}_{1_N}\alpha + o_P(\frac{1}{\sqrt{T\log N}}).$$

This implies

$$\begin{pmatrix} \bar{f}_{o,i} \\ \widehat{\lambda}_l \end{pmatrix} - H^{-1}\lambda = \begin{pmatrix} \bar{v}_{o,i} \\ (\widehat{\beta}_l'\mathbb{M}_{1_N}\widehat{\beta}_l)^{-1}\widehat{\beta}_l'\mathbb{M}_{1_N}g_l + (\widehat{\beta}_l'\mathbb{M}_{1_N}\widehat{\beta}_l)^{-1}\widehat{\beta}'\mathbb{M}_{1_N}\alpha \end{pmatrix} + o_P(\frac{1}{\sqrt{T\log N}}).$$

Then for $B_i = \widehat{\beta}_{li}'[H_l^{-1}(\bar{v}_{l,i} - H_o\bar{v}_{o,i}) - (\widehat{\beta}_l'\mathbb{M}_{1_N}\widehat{\beta}_l)^{-1}\widehat{\beta}_l'\mathbb{M}_{1_N}g_l]$,

$$\bar{r}_i - \widehat{\beta}_{o,i}'\bar{f}_{o,i} - \widehat{\beta}_{l,i}'\widehat{\lambda}_l - \alpha_i = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i} u_{it}(1 - v_t'\Sigma_f^{-1}\lambda) - \beta_{l,i}'S_{\beta,l}^{-1}\frac{1}{N}\beta_l'\mathbb{M}_{1_N}\alpha - C_i + B_i + o_P(\frac{1}{\sqrt{T\log N}}).$$

Similar proof as before yields that (B_i, C_i) can be replaced with $(\widehat{B}, \widehat{C}_i)$ with negligible effects, where

$$\widehat{B}_i = \widehat{\beta}_{li}'[\widehat{\bar{v}}_i - (\widehat{\beta}_l'\mathbb{M}_{1_N}\widehat{\beta}_l)^{-1}\widehat{\beta}_l'\mathbb{M}_{1_N}\widehat{g}_l], \quad \widehat{g}_l = (\widehat{\beta}_{l,i}'\widehat{\bar{v}}_i : i \leq N).$$

□

B.3 Proof of Theorem A.3

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

$$\sqrt{T}\sigma_i^{-1}\alpha_i \geq 4\sqrt{\log N}, \quad \forall i \in \mathcal{H}. \tag{B.9}$$

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 $t^* = \Psi^{-1}(\tau|\mathcal{H}|/N)$,

$$\frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} 1\{t_i \geq x\} \leq \Psi(x)(1 + o_P(1)). \tag{B.10}$$

The remaining proof is divided into the following steps.

Step 1. We first show the inequality (B.10). 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 allow approximation errors Δ_i.

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 B.1. 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.

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

$$\frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} 1\{z_i \geq x\} \leq \Psi(x)(1+o_P(1)), \tag{B.11}$$

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 $0 < x \leq t^*$,

$$\begin{aligned} |\Psi(x) - \Psi(x - \|\Delta\|_\infty)| &\leq \phi(x+\eta_x)\|\Delta\|_\infty \leq \phi(x)\|\Delta\|_\infty \frac{\phi(x+\eta_x)}{\phi(x)} \\ &\leq C\phi(x)\|\Delta\|_\infty \exp(C\eta_x(\eta_x + t^*)) \\ &\leq Cx\Psi(x)\|\Delta\|_\infty(1+o(1)) \leq Ct^*\Psi(x)\|\Delta\|_\infty(1+o(1)) \\ &\leq o(1)\Psi(x), \end{aligned} \tag{B.12}$$

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 by (B.11), we have

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

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

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

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 $\mathcal{F} = \sum_{i\in\mathcal{H}_0} 1\{i \leq N : t_i \geq \widehat{t}\}$. To use inequality (B.10), we then aim to prove that $x = \widehat{t} \leq t^*$. To do so, note that

$$\Psi(\widehat{t}) = \tau\frac{1}{N}\max\{\sum_{i=1}^N 1\{t_i \geq \widehat{t}\}, 1\}, \tag{B.13}$$

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

$$\sum_{i=1}^N 1\{t_i \geq \widehat{t}\} \geq |\mathcal{H}|. \tag{B.14}$$

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

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

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

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

Next, $\sqrt{T}\max_i|\operatorname{se}(\widehat{\alpha})\sqrt{T} - \sigma_i| \leq O_P(\sqrt{\log N} + \sqrt{T/N})$ by (B.32). 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 B.2, 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.14).

Because $|\mathcal{H}| \to \infty$, $\Psi(x) \leq 0.5\exp(-x^2/2)$, we have $t^* = \Psi^{-1}(\tau|\mathcal{H}|/N) \leq \sqrt{2\log N}$. Then by step 3, $\Psi(t^*) = \frac{\tau|\mathcal{H}|}{N} \leq \frac{1}{N}\sum_{i=1}^{N} 1\{t_i \geq \sqrt{2\log N}\}\tau \leq \frac{1}{N}\sum_{i=1}^{N} 1\{t_i \geq t^*\}\tau$. 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

$$\Psi(\widehat{t}) = \tau\frac{1}{N}\sum_{i=1}^{N} 1\{t_i \geq \widehat{t}\} \geq \tau\frac{|\mathcal{H}|}{N}. \tag{B.15}$$

Step 5. We prove the FDR/FDP control.

In the proof of step 4, we have $\widehat{t} \leq t^*$ with probability converging to one, then by (B.10), $\mathcal{F} \leq \Psi(\widehat{t})|\mathcal{H}_0| + o_P(1)|\mathcal{H}_0|$. Also by (B.13),

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

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 $\widehat{t} \leq t^*$. Hence

$$\text{FDP} \leq \tau + o_P(1).$$

Together, for any $\epsilon > 0$,

$$\text{FDR} \leq \mathbb{E}(\tau\frac{|\mathcal{H}_0|}{N} + X|\widehat{t} \leq t^*) + \mathbb{P}(\widehat{t} > t^*)$$

$$\leq \quad \tau \frac{|\mathcal{H}_0|}{N} + \epsilon + \mathbb{P}(|X| \geq \epsilon | \mathcal{R} \geq 1) + o(1).$$

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

Step 6. We show FDR/FDP properties for the screening B-H.

Recall that $\widehat{\mathcal{I}} = \{t_i > -\log(\log T)\sqrt{\log N}\}$. Proposition B.1 and Lemma B.1 imply

$$\max_{i \leq N} \frac{|\widehat{\alpha}_i - \alpha_i|}{\mathrm{se}(\widehat{\alpha}_i)} \leq o_P(\frac{1}{\sqrt{\log N}}) + O_P(\sqrt{\log N}) = O_P(\sqrt{\log N}). \tag{B.16}$$

Hence with probability approaching one,

$$A_1 \subseteq \widehat{\mathcal{I}} \subseteq A_2,$$

where

$$A_1 = \{\alpha_i > -\log\log T\sqrt{\frac{\log N}{T}}\sigma_i(1-\epsilon)\}, \quad A_2 = \{\alpha_i > -\log\log T\sqrt{\frac{\log N}{T}}\sigma_i(1+\epsilon)\}.$$

Let $\mathcal{H}_{0,2} = \mathcal{H}_0 \cap A_2 = \{-\log\log T\sqrt{\frac{\log N}{T}}\sigma_i(1+\epsilon) < \alpha_i \leq 0\}$. Thus $\mathcal{H}_{0,2} \subset \mathcal{H}_0$. Then the same proof as step 1 leads to, uniformly for $x \leq t^*$,

$$\frac{1}{|\mathcal{H}_{0,2}|} \sum_{i \in \mathcal{H}_{0,2}} 1\{t_i \geq x\} \leq \Psi(x)(1 + o_P(1)).$$

Let $\widehat{t}_1$, $\mathcal{F}_1$ and $\mathcal{R}_1$ be as $\widehat{t}$, $\mathcal{F}$ and $\mathcal{R}$ but defined on $\widehat{\mathcal{I}}$. Then

$$\Psi(\widehat{t}_1) = \frac{\tau}{|\widehat{\mathcal{I}}|} \max\{\sum_{i \in \widehat{\mathcal{I}}} 1\{t_i \geq \widehat{t}_1\}, 1\}, \quad \mathcal{R}_1 = \max\{\sum_{i \in \widehat{\mathcal{I}}} 1\{t_i \geq \widehat{t}_1\}, 1\} = \Psi(\widehat{t}_1)|\widehat{\mathcal{I}}|/\tau. \tag{B.17}$$

Suppose $\widehat{t}_1^* \leq t^*$, a claim to be proved later, then with probability approaching one,

$$\frac{\mathcal{F}_1}{|\mathcal{H}_{0,2}|} = \frac{\sum_{i \in \widehat{\mathcal{I}} \cap \mathcal{H}_0} 1\{t_i \geq \widehat{t}_1\}}{|\mathcal{H}_{0,2}|} \leq \frac{\sum_{i \in A_2 \cap \mathcal{H}_0} 1\{t_i \geq \widehat{t}_1\}}{|\mathcal{H}_{0,2}|} \leq \Psi(\widehat{t}_1)(1 + o_P(1)).$$

So with the assumption $|\mathcal{H}_{0,2}| \leq |A_1|$,

$$\mathrm{FDP}_{\mathrm{screening}} = \frac{\mathcal{F}_1}{\mathcal{R}_1} \leq \frac{\Psi(\widehat{t}_1)(1 + o_P(1))|\mathcal{H}_{0,2}|}{\Psi(\widehat{t}_1)|\widehat{\mathcal{I}}|/\tau} = \frac{\tau|\mathcal{H}_{0,2}|}{|\widehat{\mathcal{I}}|}(1+o_P(1)) \leq \frac{\tau|\mathcal{H}_{0,2}|}{|A_1|}(1+o_P(1)) \leq \tau(1+o_P(1)).$$

Then with the same proof as step 5,

$$\mathrm{FDR}_{\mathrm{screening}} = \mathbb{E}\,\mathrm{FDP}_{\mathrm{screening}} \leq \tau + o(1).$$

It remains to prove $\widehat{t}_1^* \leq t^*$. For any $i \in \mathcal{H}$, we note $\alpha_i \geq L_{NT}\sqrt{\frac{\log N}{T}}$. So (B.16) implies, for $L_{NT} \to \infty$ slowly,

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

Hence $i \in \widehat{\mathcal{I}}$. Hence $\mathcal{H} \subset \widehat{\mathcal{I}}$. This combined with step 3 imply with probability approaching one,

$$\sum_{i \in \widehat{\mathcal{I}}} 1\{t_i \geq \sqrt{2\log N}\} \geq |\mathcal{H}|.$$

Now let $t_1^* = \Psi^{-1}(\tau|\mathcal{H}|/|\widehat{\mathcal{I}}|)$. Since $|\mathcal{H}| \to \infty$, $\Psi(\sqrt{2\log N}) \leq 0.5\exp(-\log N) = \frac{1}{2N} \leq \frac{1}{|\widehat{\mathcal{I}}|} \leq \frac{\tau|\mathcal{H}|}{|\widehat{\mathcal{I}}|}$. So $t_1^* \leq \sqrt{2\log N}$, thus

$$\Psi(t_1^*) = \frac{\tau|\mathcal{H}|}{|\widehat{\mathcal{I}}|} \leq \frac{\tau}{|\widehat{\mathcal{I}}|}\sum_{i \in \widehat{\mathcal{I}}} 1\{t_i \geq \sqrt{2\log N}\} \leq \frac{1}{|\widehat{\mathcal{I}}|}\sum_{i \in \widehat{\mathcal{I}}} 1\{t_i \geq t_1^*\}\tau.$$

So by the definition of $\widehat{t}_1$, $\widehat{t}_1 \leq t_1^*$. Finally, $|\widehat{\mathcal{I}}| \leq N$ implies $t_1^* \leq t^*$ almost surely.

(b) The power property.

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

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

(c) To prove the power property, 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, $\widehat{k}_{\text{B-H}} \leq \widehat{k}_{\text{screening B-H}}$ and $p_{(\widehat{k}_{\text{B-H}})} \leq p_{(\widehat{k}_{\text{screening B-H}})}$. So

$$\begin{aligned}
\mathcal{G}_{\text{B-H}} &= \sum_{\mathbb{H}_0^i \text{ is false}} 1\{\mathbb{H}_0^i \text{ is rejected by B-H}\} = \sum_{\mathbb{H}_0^i \text{ is false}} 1\{p_i \leq p_{(\widehat{k}_{\text{B-H}})}\} \\
&= \sum_{\mathbb{H}_0^i \text{ is false}, i \notin \widehat{\mathcal{I}}} 1\{p_i \leq p_{(\widehat{k}_{\text{B-H}})}\} + \sum_{\mathbb{H}_0^i \text{ is false}, i \in \widehat{\mathcal{I}}} 1\{p_i \leq p_{(\widehat{k}_{\text{B-H}})}\} \\
&\leq \sum_{\mathbb{H}_0^i \text{ is false}, i \notin \widehat{\mathcal{I}}} 1\{p_i \leq 1/2\} + \sum_{\mathbb{H}_0^i \text{ is false}, i \in \widehat{\mathcal{I}}} 1\{p_i \leq p_{(\widehat{k}_{\text{screening B-H}})}\} \\
&= \sum_{\mathbb{H}_0^i \text{ is false}, i \notin \widehat{\mathcal{I}}} 1\{t_i \geq 0\} + \sum_{\mathbb{H}_0^i \text{ is false}, i \in \widehat{\mathcal{I}}} 1\{\mathbb{H}_0^i \text{ is rejected by screening B-H}\} \\
&= \mathcal{G}_{\text{screening B-H}}
\end{aligned}$$

where we used, $p_{(\widehat{k}_{\text{B-H}})} \leq \frac{\tau\widehat{k}_{\text{B-H}}}{N} \leq \tau < 1/2$, and if $i \in \widehat{\mathcal{I}}$ and $p_i \leq p_{(\widehat{k}_{\text{screening B-H}})}$, then $\mathbb{H}_0^i$ is rejected by screening B-H. Hence $\mathbb{E}\mathcal{G}_{\text{B-H}} \leq \mathbb{E}\mathcal{G}_{\text{screening B-H}}$.

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 $\widehat{\mathcal{I}}^c \subset \mathcal{H}_0$ asymptotically (to be proved in (iv) below), 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. Thus indeed $\mathbb{P}(A_{\text{screening B-H}}) \geq \mathbb{P}(A_{\text{B-H}})$.

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, $t_j > 0 > -\log(\log T)\sqrt{\log N}$, so indeed $j \in \widehat{\mathcal{I}}$.

(d) 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 $\widehat{\alpha}_i / \operatorname{se}(\widehat{\alpha}_i) \leq -\log(\log T)\sqrt{\log N}$. Thus (B.16) shows

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

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

□

B.4 Proof of Theorem A.4

Proof. (a) The main body of the proof is a standard argument of wild bootstrap. For brevity, in part (i) we focus on the case when latent factors are present. The case of observed-factors-only is well known and straightforward. Recall that

$$r_{it}^* = \widehat{\beta}_i' \widehat{\lambda} + \widehat{\beta}_{o,i}' \widehat{v}_{o,t} + \widehat{\beta}_{l,i}' \widehat{v}_{l,t} + \widehat{u}_{it}^*, \quad t \in \mathcal{T}_i.$$

For any $\zeta_{it} \in \{\widehat{u}_{it}, \widehat{u}_{it}\widehat{v}_t\}$, $\frac{1}{T_i}\sum_{t\in\mathcal{T}_i} \zeta_{it}^* = \frac{T}{T_i}\frac{1}{T}\sum_t \omega_{it}\zeta_{it} w_{it}^*$. We have $\max_i \|\frac{T}{T_i}\| = O_P(1)$, and $\mathbb{E}^* \frac{1}{T}\sum_t \omega_{it}\zeta_{it} w_{it}^* = 0$. Also, $P(|w_{it}| > x) \leq Ce^{1-Cx^c}$. Hence by Bernstein inequality and the union bound, we have $\max_i \|\frac{1}{T}\sum_t \omega_{it}\zeta_{it} w_{it}^*\| = O_{P^*}(\sqrt{\frac{\log N}{T}})$. This implies $\max_i \|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\zeta_{it}^*\| = O_{P^*}(\sqrt{\frac{\log N}{T}})$. Also we have

$$\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}(\widehat{v}_t - H^{-1}v_t)\widehat{u}_{it}^* = O_{P^*}(\sqrt{\frac{\log N}{T}})\sqrt{\max_i \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\|\widehat{v}_t - H^{-1}v_t\|^2} = o_{P^*}(\frac{1}{\sqrt{T\log N}}).$$

By Lemma B.4 (vii), we have

$$\widehat{\beta}_i^* - \widehat{\beta}_i = (\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}(\widehat{v}_t - \widehat{\bar{v}}_i)(\widehat{v}_t - \widehat{\bar{v}}_i)')^{-1}\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}(\widehat{v}_t - \widehat{\bar{v}}_i)\widehat{u}_{it}^* = H'\Sigma_f^{-1}\frac{1}{T_i}\sum_{t\in\mathcal{T}_i} v_t\widehat{u}_{it}^* + o_{P^*}(\frac{1}{\sqrt{T\log N}}).$$

Then $\frac{1}{N}\sum_i m_i(\widehat{\beta}_i^* - \widehat{\beta}_i)' = o_{P^*}(\frac{1}{\sqrt{T\log N}}) = \frac{1}{N}\sum_i \frac{1}{T_i}\sum_{t\in\mathcal{T}_i} m_i\widehat{u}_{it}^*$ for all $m_i \in \{\widehat{\beta}_i, 1\}$.

As a result, for $g^* = (g_i^* : i \leq N)$ with $g_i^* = \widehat{\beta}_i^{*\prime}\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\widehat{v}_t$, we have

$$\widehat{\lambda}^* - \widehat{\lambda} = (\widehat{\beta}^*\mathbb{M}_{1_N}\widehat{\beta}^*)^{-1}\widehat{\beta}^*\mathbb{M}_{1_N} g^* + o_{P^*}(\frac{1}{\sqrt{T\log N}}).$$

Thus it follows that

$$\begin{aligned}\bar{r}_i^* - \widehat{\beta}_i^{*\prime}\widehat{\lambda}^* &= \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\widehat{u}_{it}^* - (\widehat{\beta}_i^{*\prime} - \widehat{\beta}_i')\widehat{\lambda} + g_i^* - \widehat{\beta}_i'(\widehat{\lambda}^* - \widehat{\lambda}) + o_{P^*}(\frac{1}{\sqrt{T\log N}}) \\ &= \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\widehat{u}_{it}^*(1 - v_t'\Sigma_f^{-1}H\widehat{\lambda}) + g_i^* - \widehat{\beta}_i'(\widehat{\beta}^*\mathbb{M}_{1_N}\widehat{\beta}^*)^{-1}\widehat{\beta}^*\mathbb{M}_{1_N}g^* + o_{P^*}(\frac{1}{\sqrt{T\log N}}),\end{aligned}$$

which leads to the desired result.

(b) First, according to the Glivenko-Cantelli theorem, we can in spirit define $p_i^* = P^*(\widehat{\alpha}_i^* > \widehat{\alpha}_i)$, that is, replacing the bootstrap empirical measure $\frac{1}{B}\sum_{b=1}^B$ with the bootstrap measure P^*. We have the following decomposition: for $\max_i \|\Delta_i\| = o_{P^*}(\frac{1}{\sqrt{T\log N}})$ and for $\omega_{it} = 1\{r_{it}$ is observed$\}$,

$$\widehat{\alpha}_i^* = \frac{1}{T}\sum_{t=1}^T y_{it}^* + \Delta_i, \quad y_{it}^* = \widehat{u}_{it}^*\mu_{it}, \tag{B.18}$$

where

$$\mu_{it} := \frac{\omega_{it}}{\mathbb{E}\omega_{it}}(1 - v_t'\Sigma_f^{-1}\lambda).$$

Let $s_i^* = \sqrt{\frac{1}{T}\sum_t y_{it}^{*2}}$ and $s_i^2 = \frac{1}{T}\sum_t \widehat{u}_{it}^2\mu_{it}^2$. Then we have

$$\max_i |s_i^{*2} - T\,\mathrm{se}(\widehat{\alpha}_i^*)^2| \le \max_i |s_i^{*2} - s_i^2| + \max_i |s_i^2 - T\,\mathrm{se}(\widehat{\alpha}_i^*)^2| = o_P(1).$$

Also, recall that $\Psi(x) := 1 - \Phi(x)$. By definition $p_i = \Psi(\frac{\widehat{\alpha}_i^*}{\mathrm{se}(\widehat{\alpha}_i^*)})$, and

$$p_i^* = P^*(\widehat{\alpha}_i^* > \widehat{\alpha}_i^*) = P^*(\frac{\sqrt{T}\widehat{\alpha}_i^*}{s_i^*} > \frac{\widehat{\alpha}_i^*}{\mathrm{se}(\widehat{\alpha}_i^*)}\frac{\sqrt{T}\,\mathrm{se}(\widehat{\alpha}_i^*)}{s_i^*}) = P^*(\frac{\frac{1}{\sqrt{T}}\sum_t y_{it}^*}{s_i^*} > y_{1i}),$$

where $y_{1i} = \frac{\widehat{\alpha}_i^*}{\mathrm{se}(\widehat{\alpha}_i^*)}\frac{\sqrt{T}\,\mathrm{se}(\widehat{\alpha}_i^*)}{s_i^*} - \frac{\sqrt{T}\Delta_i}{s_i^*}$.

The main technical tool is the moderate deviations for self-normalized sums (Peña et al., 2008) (also see Lemma 5 of Belloni et al. (2012)), which approximates $P^*(\frac{\frac{1}{\sqrt{T}}\sum_t y_{it}^*}{s_i^*} > y)$ using the standard normal distribution uniformly over $y \le T^{1/6}/l_T M$ for some $l_T \to \infty$ slowly and constant M. We thus consider two sets:

Set 1: $S_1 = \{i : \frac{\widehat{\alpha}_i^*}{\mathrm{se}(\widehat{\alpha}_i^*)} > T^{1/5}\}$. Then $\max_{i\in S_1} p_i \le \Psi(T^{1/5}) \to 0$. Also, $\min_{i\in S_1} s_i^* y_{1i} \ge cT^{1/5}$ for some $c > 0$ with probability P^* goes to one. Hence by the Chebyshev inequality,

$$\max_{i\in S_1} p_i^* \;\le\; \max_{i\in S_1} P^*(\frac{1}{\sqrt{T}}\sum_t y_{it}^* > cT^{1/5}) + o_P(1) \le \frac{C\max_{i\in S_1}\frac{1}{T}\sum_t \widehat{u}_{it}^2\mu_{it}^2}{T^{2/5}} + o_P(1) = o_P(1).$$

Set 2: $S_2 = \{i : \frac{\widehat{\alpha}_i^*}{\mathrm{se}(\widehat{\alpha}_i^*)} \le T^{1/5}\}$. Then $\max_{i\in S_2} y_{1i} \le 2T^{1/5}$ with P^* approaching one. Also, $|y_{1i} - \frac{\widehat{\alpha}_i^*}{\mathrm{se}(\widehat{\alpha}_i^*)}| < \frac{1}{\sqrt{\log N}}$. Hence $y_{2i} := \frac{\widehat{\alpha}_i^*}{\mathrm{se}(\widehat{\alpha}_i^*)} - \frac{1}{\sqrt{\log N}} < y_{1i} < \frac{\widehat{\alpha}_i^*}{\mathrm{se}(\widehat{\alpha}_i^*)} + \frac{1}{\sqrt{\log N}} := y_{3i}$. Then $P^*(\frac{\frac{1}{\sqrt{T}}\sum_t y_{it}^*}{s_i^*} >$

$y_{3i}) < p_i^* < P^*(\frac{\frac{1}{\sqrt{T}}\sum_t y_{it}^*}{s_i^*} > y_{2i})$. Hence

$$\max_{i\in S_2}|p_i^*-p_i| \leq \max_{i\in S_2}|P^*(\frac{\frac{1}{\sqrt{T}}\sum_t y_{it}^*}{s_i^*} > y_{2i}) - p_i| + \max_{i\in S_2}|P^*(\frac{\frac{1}{\sqrt{T}}\sum_t y_{it}^*}{s_i^*} > y_{3i}) - p_i|.$$

We now bound the first term on the right. The second one follows similarly.

$$\begin{aligned}\max_{i\in S_2}|P^*(\frac{\frac{1}{\sqrt{T}}\sum_t y_{it}^*}{s_i^*} > y_{2i}) - p_i| &\leq \max_{i\in S_2}|P^*(\frac{\frac{1}{\sqrt{T}}\sum_t y_{it}^*}{s_i^*} > y_{2i}) - \Psi(y_{2i})| + \max_{i\in S_2}|p_i - \Psi(y_{2i})| \\ &\leq \max_{i\in S_2}\sup_{y<T^{1/5}+\frac{1}{\sqrt{\log N}}}|P^*(\frac{\frac{1}{\sqrt{T}}\sum_t y_{it}^*}{s_i^*} > y) - \Psi(y)| + \frac{C}{\sqrt{\log N}} \\ &= o_P(1).\end{aligned}$$

Together, $\max_i |p_i - p_i^*| = o_P(1)$.

(c) Given the expansion (B.18), the proof of the FDR control is very standard (e.g., Proposition 2.3 of Liu and Shao (2014)), hence we only sketch it here. First, define

$$G_i^*(t) = P^*(\frac{\frac{1}{\sqrt{T}}\sum_{t=1}^T y_{it}^*}{s_i} > t).$$

Then $p_i^* = G_i^*(t_i - \delta_i)$ where $t_i = \sqrt{T}\widehat{\alpha}_i/s_i$ and $\max_i\|\delta_i\| = \max_i\|\sqrt{T}\Delta_i/s_i\| = o_P(\frac{1}{\sqrt{\log N}})$. Also, for the same μ_{it}, by Theorem A.2,

$$t_i - \sqrt{T}\alpha_i/s_i = z_i + \delta_{i2}, \quad z_i = \frac{1}{\sqrt{T}s_i}\sum_t u_{it}\mu_{it}, \quad \max_i\|\delta_{i2}\| = o_P(\frac{1}{\sqrt{\log N}}).$$

Next, by the same argument as in the proof of their Proposition 2.3, there exists $G_{\kappa,i}(t)$ such that

$$G_i^*(t) = G_{\kappa,i}(t)(1+o(1)), \quad \frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} 1\{z_i \geq t\} = \frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} G_{\kappa,i}(t)(1+o(1))$$

uniformly for $i \leq N$, $t \in A := [-C\sqrt{\log N}, C\sqrt{\log N}]$ for some $C > 0$ (for the left statement) and uniformly $t \in B := [0, G_{\kappa,i}^{-1}(b_N/N)]$ for any $b_N \to \infty$ (for the right statement). Next, the B-H rejects $\mathbb{H}_0^i$ if and only if $p_i^* \leq \hat{x}$, where $\hat{x} = \sup\{0 \leq x \leq 1 : Nx \leq \tau\max\{1, \sum_i 1\{p_i^* \leq x\}\}\}$ satisfies

$$\hat{x} = \frac{\tau\max\{1, \sum_i 1\{p_i^* \leq \hat{x}\}\}}{N} = \frac{\tau\max\{1, \mathcal{R}\}}{N}.$$

With probability approaching one, $t_i - \sqrt{T}\alpha_i/s_i - \delta_i \in A$ for all $i \in \mathcal{H}_0$. Also, because $|\mathcal{H}| \to \infty$, we have $x_0 := \frac{\tau|\mathcal{H}|}{N} \geq G_{\kappa,i}(C\sqrt{\log N})$,

$$\tau\sum_i 1\{p_i^* \leq x_0\} \geq \tau\sum_i 1\{p_i^* \leq G_{\kappa,i}(C\sqrt{\log N})\} \geq \frac{\tau|\mathcal{H}|}{N}N = x_0 N,$$

where the last inequality is from the similar argument of step 3 of the proof of Theorem A.3(i). Then the definition of $\hat{x}$ yields $x_0 \leq \hat{x}$. Thus $G_{\kappa,i}^{-1}(\hat{x}(1+o(1))) + \delta_{i3} \in B$ for all $i \in \mathcal{H}_0$. For $\delta_{i3} = \delta_i - \delta_{i2}$,

$$\frac{1}{|\mathcal{H}_0|}\mathcal{F} \quad = \quad \frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} 1\{p_i^* \leq \hat{x}\} = \frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} 1\{G_i^*(t_i - \delta_i) \leq \hat{x}\}$$

$$
\begin{aligned}
&\leq \frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} 1\{G_i^*(t_i-\sqrt{T}\alpha_i/s_i-\delta_i)\leq \hat{x}\}\\
&= \frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} 1\{G_{\kappa,i}(t_i-\sqrt{T}\alpha_i/s_i-\delta_i)\leq \hat{x}(1+o(1))\}\\
&= \frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} 1\{z_i\geq G_{\kappa,i}^{-1}(\hat{x}(1+o(1)))+\delta_{i3}\}\\
&= \frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} G_{\kappa,i}(m_i+\delta_{i3})(1+o(1)) = \frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} G_{\kappa,i}(m_i)(1+o(1))+\delta_4\\
&\overset{(1)}{=} \frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} G_{\kappa,i}(m_i)(1+o(1))\\
&= \hat{x}(1+o(1)) = \frac{\tau\max\{1,\mathcal{R}\}}{N}(1+o(1)),
\end{aligned}
$$

where the first inequality is due to $\alpha_i \leq 0$ for $i\in\mathcal{H}_0$ and that G_i^* is nonincreasing; $m_i = G_{\kappa,i}^{-1}(\hat{x}(1+o(1)))$; (1) is due to, following the same proof of (B.12),

$$
\delta_4 = \frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0}[G_{\kappa,i}(m_i+\delta_{i3})-G_{\kappa,i}(m_i)](1+o(1)) = \frac{1}{|\mathcal{H}_0|}\sum_{i\in\mathcal{H}_0} G_{\kappa,i}(m_i)o(1).
$$

Hence with probability approaching one,

$$
\frac{\mathcal{F}}{\max\{1,\mathcal{R}\}} \leq \frac{|\mathcal{H}_0|\tau}{N}(1+o(1)).
$$

From here, the remaining proof is the same as in Theorem A.3(i).

□

B.5 Proof of Theorem A.5

Proof. First of all, let $w = \mathbb{E}[(f_{l.t}-\mathbb{E}f_{l,t})f_{o,t}']\operatorname{Cov}(f_{o,t})^{-1}$. Then it is straightforward to check that

$$
\Gamma = \beta_l w + \beta_o. \tag{B.19}
$$

(Note that $\widehat{\beta}_0$ converges in probability to Γ, therefore $\widehat{\beta}_0$ is biased for β_0 unless $f_{o,t}$ and $f_{l,t}$ are uncorrelated, which is the omitted variable bias.) Next, define

$$
h_t = f_{l.t} - \mathbb{E}f_{l,t} - w(f_{o,t}-\mathbb{E}f_{o,t}).
$$

Then it is also straightforward to check that $Z_t = \beta_l h_t + u_t$. This proves the first equation.

Next, given the invertible matrix Q (whose existence is proved in the high-dimensional factor model literature, e.g., Fan et al. (2016)), we show that there is an invertible H so that $\beta H = (\Gamma, \beta_l Q)$. In fact, from (B.19),

$$
(\Gamma, \beta_l Q) = \underbrace{(\beta_o, \beta_l)}_{\beta}\underbrace{\begin{pmatrix} I & 0 \\ w & Q \end{pmatrix}}_{H},
$$

where $\det(H) = \det(Q) \neq 0$. This proves the second equation. (Also, $\widehat{\beta}_l$ converges in probability to $\beta_l Q$. Therefore $\widehat{\beta} = (\widehat{\beta}_o, \widehat{\beta}_l)$ converges to $(\Gamma, \beta_l Q) = \beta H$.)

Next, multiply $\mathcal{T}(\beta)\mathbb{M}_{1_N}$ to both sides of $\mathbb{E}r_t = \alpha + \beta\lambda$:

$$\beta\lambda = \underbrace{\beta(\beta'\mathbb{M}_{1_N}\beta)^{-1}\beta'}_{\mathcal{T}(\beta)}\mathbb{M}_{1_N}\beta\lambda = \mathcal{T}(\beta)\mathbb{M}_{1_N}\mathbb{E}r_t - \mathcal{T}(\beta)\mathbb{M}_{1_N}\alpha.$$

This proves the third equation. Finally, $\alpha = \mathbb{E}r_t - \beta\lambda$ follows immediately. $\square$

B.6 Proof of Theorem A.6

Proof. In cases (i)-(iii), 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.8),

$$\widehat{\alpha}_i - \alpha_i = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i} u_{it}(1 - v_t'\Sigma_f^{-1}\lambda) - \beta_i' S_\beta^{-1}\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha + \Delta_i,$$

where $\|\Delta\|_\infty = O_P(\frac{\log N}{T} + \frac{1}{N})$. In (B.31), we showed $\|\Delta\|_\infty = o_P(\frac{1}{\sqrt{T\log N}})$. In fact, 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. Thus

$$\widehat{\alpha}_0 - \bar{\alpha} = -\bar{\beta}' S_\beta^{-1}\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha + O_P(\frac{\log N}{T} + \frac{1}{N}).$$

For $1_N = (1, ..., 1)'$, $P_\beta = \beta(\beta'\beta)^{-1}\beta'$, $M_\beta = I - P_\beta$,

$$\begin{aligned}
\widehat{\alpha}_0 - \alpha_0 + O_P(\frac{\log N}{T} + \frac{1}{N}) &= \frac{1}{N}1_N'\alpha - \frac{1}{N}1_N'\beta(\frac{1}{N}\beta'\mathbb{M}_{1_N}\beta)^{-1}\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha - \alpha_0 \\
&= \frac{1}{N}1_N'\alpha - (1_N'M_\beta 1_N)^{-1}1_N'P_\beta\mathbb{M}_{1_N}\alpha - \alpha_0 \\
&= \frac{1}{N}1_N'\alpha - (1_N'M_\beta 1_N)^{-1}1_N'P_\beta\alpha + (1_N'M_\beta 1_N)^{-1}1_N'P_\beta 1_N 1_N'\frac{1}{N}\alpha - \alpha_0 \\
&= (1_N'M_\beta 1_N)^{-1}1_N'M_\beta(\alpha - 1_N\alpha_0) \\
&= (1_N'M_\beta 1_N)^{-1}\sum_i(\alpha_i - \alpha_0)(1 - \bar{\beta}(\frac{1}{N}\beta'\beta)^{-1}\beta_i).
\end{aligned}$$

The second equality uses the Woodbury matrix identity for $(\frac{1}{N}\beta'\mathbb{M}_{1_N}\beta)^{-1}$. It is easy to check that the triangular array Lindeberg 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 $\bar{\sigma}^2 > 0$.

In case (iv) that observable factors are tradable and there are also latent factors, the result is similar except that $\bar{\beta}$ and M_β should be replaced with $\bar{\beta}_l$ and $M_{\beta,l}$.

In case (v) that observable factors are tradable and there are no latent factors, we have

$$\widehat{\alpha}_i = \alpha_i - \bar{f}_i'(F_i'\mathbb{M}_{T_i}F_i)^{-1}F_i'\mathbb{M}_{T_i}u_i + \bar{u}_i.$$

Hence $\widehat{\alpha}_0 - \alpha_0 = \bar{\alpha} - \alpha_0 - \frac{1}{N}\sum_i \bar{f}_i'(F_i'\mathbb{M}_{T_i}F_i)^{-1}F_i'\mathbb{M}_{T_i}u_i + \frac{1}{N}\sum_i \bar{u}_i = \bar{\alpha} - \alpha_0 + O_P(\frac{1}{\sqrt{NT}})$. Then $\sqrt{N}\frac{\widehat{\alpha}-\alpha_0}{\widehat{\sigma}_\alpha} \xrightarrow{d} \mathcal{N}(0,1)$.

□

B.7 Proof of Theorem A.7

Proof. When N is bounded, (B.22) 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 $\beta_i \neq 0$. □

B.8 Technical Lemmas

The following proposition gives the asymptotic expansion for the estimated alphas. It applies to estimators that are obtained in any of the five factor scenarios: (i) observable factors only (Algorithm 3), (ii) latent factors only (Algorithm 4), (iii) the general case (mixed of observable and latent factors, Algorithm 3), (iv) mixed of observable and latent factors with additional condition that observable factors are tradable (Algorithm A.1), and (v) all factors are observable and tradable.

Proposition B.1. *Under the conditions of Theorem A.1,*

(a) Let $\|\Delta\|_\infty = O_P(\frac{\log N}{T} + \frac{1}{N})$. We have

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

where $\zeta = \beta S_\beta^{-1}\beta'$ for scenarios (i)-(iii), $\zeta = \beta_l S_{\beta,l}^{-1}\beta_l'$ for scenario (iv), and $\zeta = 0$ for scenario (v).

(b) Uniformly in $i \leq N$, when $T\log N = o(N)$,

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

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

Proof. Without loss of generality, we shall assume $\dim(f_t) = 1$ in order to simplify the notation. We use $C > 0$ to denote a generic constant.

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

$$\widehat{\beta} - \beta = (\frac{1}{T}\sum_t u_t v_t' - \bar{u}\bar{v}')S_f^{-1}, \tag{B.20}$$

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. Conditioning on α,

$$\begin{aligned}
\mathbb{E}[\|\alpha'\frac{1}{T}\sum_t u_t f_t'\|^2] &= \frac{1}{T^2}\sum_{k=1}^{K}\sum_t\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\mathbb{E}\mathbb{E}[u_{it}u_{jt}|f_t]f_{t,k}^2 \\
&\leq \frac{N}{T}\mathbb{E}\|f_t\|^2\max_{i,t}\sum_{j=1}^{N}|\mathbb{E}[u_{it}u_{jt}|f_t]| \leq \frac{CN}{T}.
\end{aligned}$$

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,

$$\widehat{\lambda} - \lambda = \bar{v} + \frac{1}{N}S_\beta^{-1}\beta'\mathbb{M}_{1_N}\alpha + O_P(\frac{1}{\sqrt{NT}}). \tag{B.21}$$

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.20) (B.21),

$$\widehat{\alpha} - \alpha = \bar{u} - \frac{1}{T}\sum_t u_t v_t' S_f^{-1}\widehat{\lambda} + \bar{u}\bar{v}'S_f^{-1}\widehat{\lambda} - \beta\frac{1}{N}S_\beta^{-1}\beta'\mathbb{M}_{1_N}\alpha - \beta\sum_{d=1}^{7}A_d. \tag{B.22}$$

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

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

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, and that

$$H = \frac{1}{NT}\sum_t (v_t-\bar{v})(v_t-\bar{v})'\beta'\widehat{\beta}D^{-1} + \frac{1}{NT}\sum_t (v_t-\bar{v})(u_t-\bar{u})'\widehat{\beta}D^{-1}.$$

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 variables" 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

$$\widehat{\beta}-\beta H = \frac{1}{NT}\sum_t u_t v_t'\beta'\widehat{\beta}D^{-1} + \frac{1}{NT}\sum_t (u_t u_t' - \mathbb{E}u_t u_t')\widehat{\beta}D^{-1} + \frac{1}{N}(\mathbb{E}u_t u_t')\widehat{\beta}D^{-1} - G, \quad \text{(B.23)}$$

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 eigenvectors, 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 $\max_{i,j\le N}\sum_{k=1}^N|\operatorname{Cov}(u_{it}u_{kt},u_{jt}u_{kt})|<C$ implies $\max_j\psi_1(\operatorname{Var}(u_tu_{jt}))<C$, so

$$\begin{aligned}
\mathbb{E}\|\frac{1}{\sqrt{N}}\beta'\frac{1}{NT}\sum_t(u_tu_t'-\mathbb{E}u_tu_t')\|^2 &= \frac{1}{N}\sum_{j=1}^N\frac{1}{N^2T}\beta'\operatorname{Var}(u_tu_{jt})\beta\\
&\le \max_j\psi_1(\operatorname{Var}(u_tu_{jt}))\frac{1}{N^2T}\|\beta\|^2\le\frac{C}{NT},\\
\mathbb{E}\|\frac{1}{\sqrt{N}N}\beta'1_N1_N'\frac{1}{NT}\sum_t(u_tu_t'-\mathbb{E}u_tu_t')\|^2 &\le \frac{C}{NT}.
\end{aligned}\tag{B.24}$$

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, we have

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

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,

$$\widehat{\lambda}-H^{-1}\lambda = H^{-1}\bar{v}+\widehat{S}_\beta^{-1}\frac{1}{N}H'\beta'\mathbb{M}_{1_N}\alpha+O_P(\frac{1}{N}+\frac{1}{T}).\tag{B.25}$$

Step 3. Expand $\widehat{\alpha}-\alpha$. Substitute in the expansions (B.23) and (B.25) in steps 2, 3,

$$\widehat{\alpha}-\alpha = \beta\bar{v}+\bar{u}-\beta H(\widehat{\lambda}-H^{-1}\lambda)+(\beta H-\widehat{\beta})\widehat{\lambda}$$

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

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

$$
\begin{aligned}
\frac{1}{N} \beta' \widehat{\beta} D^{-1} \widehat{\lambda} - \Sigma_f^{-1} \lambda &= \frac{1}{N} H'^{-1} (H' \beta' - \widehat{\beta}') \widehat{\beta} D^{-1} \widehat{\lambda} + H'^{-1} D^{-1} (\widehat{\lambda} - H^{-1} \lambda) + (HDH')^{-1} \lambda - \Sigma_f^{-1} \lambda \\
&= O_P(\frac{1}{\sqrt{T}} + \frac{1}{N}) + [(HDH')^{-1} - \Sigma_f^{-1}] \lambda.
\end{aligned}
$$

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

$$
\begin{aligned}
\|\frac{1}{NT} \sum_t (u_t u_t' - \mathbb{E} u_t u_t') \widehat{\beta} D^{-1}\|_\infty &\leq \max_i \|\frac{1}{\sqrt{NT}} \sum_t (u_{it} u_t' - \mathbb{E} u_{it} u_t')\| \frac{1}{\sqrt{N}} \|\widehat{\beta} - \beta H\| \|D^{-1}\| \\
&\quad + \|\frac{1}{NT} \sum_t (u_t u_t' - \mathbb{E} u_t u_t') \beta H D^{-1}\|_\infty \\
&= O_P(\sqrt{\frac{\log N}{NT}} + \sqrt{\frac{\log N}{T^2}}). \qquad \text{(B.26)}
\end{aligned}
$$

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

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

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.26) and Lemma B.1,

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

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

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

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

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

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 $K_o = \dim(f_{o,t})$. 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 exists a matrix

$$A = \begin{pmatrix} I_{K_o} \\ \frac{1}{T}\sum_t (f_{l,t} - \bar{f}_l)v_{o,t}'S_o^{-1} \end{pmatrix} = \begin{pmatrix} I_{K_o} \\ a \end{pmatrix},$$

such that

$$\widehat{\beta}_o - \beta A = \xi_1, \quad \xi_1 = (\frac{1}{T}\sum_t u_t v_{o,t}' - \bar{u}\bar{v}_o')S_o^{-1}. \tag{B.27}$$

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 $\beta_l \frac{1}{T}\sum_t (f_{l,t} - \bar{f}_l)v_{o,t}'S_o^{-1}$, which 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. We also have $Z_t = \beta_l l_{1t} + \mu_t$, where

$$\mu_t = u_t - \bar{u} - \xi_1(f_{o,t} - \bar{f}_o), \quad l_{1t} := f_{l,t} - \bar{f}_l - a(f_{o,t} - \bar{f}_o).$$

Therefore, we let D 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

$$H_1 = \begin{pmatrix} 0 \\ b \end{pmatrix}. \quad H = (A, H_1) = \begin{pmatrix} I_{K_o} & 0 \\ a & b \end{pmatrix}, \quad b = \frac{1}{TN}\sum_t l_{1t}(l_{1t}'\beta_l' + \mu_t')\widehat{\beta}_l D^{-1}.$$

Then $\widehat{\beta}_l - \beta_l b = \xi_2$, where

$$\xi_2 = \frac{1}{TN}\sum_t u_t(l_{1t}'\beta_l' + \mu_t')\widehat{\beta}_l D^{-1} - \xi_1\frac{1}{TN}\sum_t f_{o,t}(l_{1t}'\beta_l' + \mu_t')\widehat{\beta}_l D^{-1}.$$

Let $\widehat{\beta} = (\widehat{\beta}_o, \widehat{\beta}_l)$, and $\xi_3 = (\xi_1, \xi_2)$. So

$$\widehat{\beta} = \beta H + \xi_3. \tag{B.28}$$

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 $\mathbb{M}_{1_N}=I-1_N1_N'/N$, 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.24) 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.25) still carries over. So we can avoid repeating ourselves but directly conclude that

$$\widehat{\lambda}-H^{-1}\lambda=H^{-1}\bar{v}+\widehat{S}_\beta^{-1}\frac{1}{N}H'\beta'\mathbb{M}_{1_N}\alpha+O_P(\frac{1}{N}+\frac{1}{T}). \tag{B.29}$$

Also, if we further write

$$H^{-1}=\begin{pmatrix} I_{K_o} & 0\\ -b^{-1}a & b^{-1}\end{pmatrix},$$

then it is easy to see that the first K_o components of $H^{-1}\lambda$ is λ_o. That is, the risk premia of observed factors are consistently estimated: $\widehat{\lambda}_o\to^P\lambda_o$, which is rotation-free, while the latent factor premia are still estimated up to a rotation.

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

$$\begin{aligned}
\widehat{\alpha}-\alpha &= \beta\bar{v}+\bar{u}-\beta H(\widehat{\lambda}-H^{-1}\lambda)+(\beta H-\widehat{\beta})\widehat{\lambda}\\
&= \bar{u}-\beta H\widehat{S}_\beta^{-1}\frac{1}{N}H'\beta'\mathbb{M}_{1_N}\alpha-\xi_3H^{-1}\lambda+\Delta
\end{aligned}$$

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 ξ_3H^{-1}, which we now focus on.

Note that by definition, for $m_2:=\frac{1}{N}\beta_l'\widehat{\beta}_lD^{-1}$, $L_1=(l_{1t}:t\le T)$ be $K_l\times T$, $v_o=(v_{o,t}:t\le T)$ be $K_o\times T$, $v_l=(v_{l,t}:t\le T)$ be $K_l\times T$ matrix, and U be $N\times T$ matrix of u_{it}, we can write in a matrix form

$$\xi_3 = \left(\frac{1}{T}Uv_o'S_o^{-1},\quad (\frac{1}{T}U-\frac{1}{T}Uv_o'S_o^{-1}\frac{1}{T}v_o)L_1'm_2\right)+\Delta.$$

Write $J = (\frac{1}{T}U - \frac{1}{T}Uv_o'S_o^{-1}\frac{1}{T}v_o)L_1'(\frac{1}{T}L_1L_1')^{-1}$. It can be verified that, for $\|\Delta\|_\infty = O_P(\frac{\log N}{T} + \frac{1}{N})$,

$$\begin{aligned} \xi_3 H^{-1} &\overset{(1)}{=} \left(\frac{1}{T}\sum_t u_t v_{o,t}' S_o^{-1} - Ja, J\right) + \Delta \\ &\overset{(2)}{=} \frac{1}{T}\sum_t u_t v_t' \Sigma_f^{-1} + \Delta. \end{aligned} \tag{B.30}$$

We now prove both equalities.

As for (1), note that

$$H_1 = \begin{pmatrix} 0 \\ b \end{pmatrix}, \quad b = \frac{1}{TN}\sum_t l_{1t}l_{1t}'\beta_l'\widehat{\beta}_l D^{-1} + O_P(\frac{1}{T} + \frac{1}{N}) = \frac{1}{T}L_1L_1'm_2 + O_P(\frac{1}{T} + \frac{1}{N}).$$

Also note that $H = (A, H_1)$ so

$$\begin{aligned} \left(\frac{1}{T}Uv_o'S_o^{-1} - Ja, \quad J\right) A &= \frac{1}{T}Uv_o'S_o^{-1}, \\ \left(\frac{1}{T}Uv_o'S_o^{-1} - Ja, \quad J\right) H_1 &= J\frac{1}{T}L_1L_1'm_2 = (\frac{1}{T}U - \frac{1}{T}Uv_o'S_o^{-1}\frac{1}{T}v_o)L_1'm_2 + O_P(\frac{1}{T} + \frac{1}{N}). \end{aligned}$$

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

As for (2),

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

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

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

Note that $a = S_{ol}'S_o^{-1} + O_P(T^{-1/2})$, so $(\frac{1}{T}L_1L_1') = \frac{1}{T}(v_l' - av_o')(v_l - v_o a') = W + O_P(T^{-1/2})$. So it holds that

$$J = \left[\frac{1}{T}Uv_l' - \frac{1}{T}Uv_o'S_o^{-1}S_{ol}\right] W^{-1} + \Delta = a_2 + \Delta,$$

and

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

Then we have

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

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

$$\widehat{\alpha}-\alpha=\frac{1}{T}\sum_t u_t(1-v_t'\Sigma_f^{-1}\lambda)-\beta S_\beta^{-1}\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha+\Delta, \tag{B.31}$$

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

Scenario (iv) & (v). In the mixed factor case with tradable observable factors, the proof is very similar to scenario (iii). In the case of observed factors only and they are all tradable, the problem becomes the regular fund-by-fund time series regression. We omit details to avoid repetitions.

(b) We only provide proof for the latent factor case, since the other scenarios are very similar.

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

$$\frac{\widehat{\alpha}_i-\alpha_i}{\mathrm{se}(\widehat{\alpha}_i)}=\frac{m_i+\Delta_i\sqrt{T}}{\sqrt{T}\,\mathrm{se}(\widehat{\alpha}_i)}-\frac{\beta_i'S_\beta^{-1}\frac{1}{N}\beta'\mathbb{M}_{1_N}\alpha}{\mathrm{se}(\widehat{\alpha}_i)}.$$

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

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

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

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

The second term on the right is $O_P(\sqrt{\log N/T})$ by Lemma B.1. We now focus on the first term. The first term is bounded by $Q_1+Q_2+Q_3$, where

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

(1) 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

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

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

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

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

(2) 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

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

(3) 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

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

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 $w_t = \frac{1}{\sqrt{N}}\beta' u_t$. Next, by direct calculations

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

Then we have

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

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

$$\max_i |\sigma_i^2 - T\,\mathrm{se}(\widehat{\alpha}_i)^2| \leq O_P(\sqrt{\frac{\log N}{T}} + \frac{1}{\sqrt{N}}). \tag{B.32}$$

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

$\square$

Proposition B.2. *Consider the latent factor model*

$$Z = X + \mathcal{E} + \mathcal{K} + \mathcal{G}, \quad X = \beta_l F$$

where F denotes the $K \times T$ matrix of latent factors, and Z is subject to missing data satisfying Assumptions A.1-A.5; $\mathcal{E}$ is the idiosyncratic noise. Suppose

(i) with probability approaching one, $1.1\|\mathcal{E} \circ X\| < \lambda_{NT}$;

(ii) $\mathcal{K}$ is an approximation error so that $\mathcal{K} = (\kappa_{it})$, $\kappa_{it} = a_i' b_t$ where a_i, b_t are such that $\max_i \|a_i\| = o_P(\frac{1}{\sqrt{T \log N}})$ and $\frac{1}{T}\sum_t \|b_t\|^4 = O_P(1)$.

(iii) Also $\max_{it} \frac{1}{N}\sum_{j,k} |\operatorname{Cov}(\varepsilon_{jt}\varepsilon_{it}, \varepsilon_{kt}\varepsilon_{it})| < C$.

(iv) $\mathcal{G} = (g_{it})$ is that $\max_i \frac{1}{T}\sum_t g_{it}^2 = o_P(\frac{1}{\sqrt{T \log N}})$ and $\frac{1}{T}\sum_t (\frac{1}{N}\sum_j \omega_{jt} g_{jt} \beta_j)^2 = o_P(\frac{1}{T \log N})$.

Then there is a rotation matrix H so that

$$\widehat{\beta}_{l,i} - H'\beta_{l,i} = H'(\frac{1}{T}\sum_{t=1}^{T} F_t F_t')^{-1} \frac{1}{T_i} \sum_{t \in \mathcal{T}_i} F_t \varepsilon_{it} + o_P(\frac{1}{\sqrt{T \log N}}) \tag{B.33}$$

where the o_P term is in $\|.\|_\infty$.

Proof. First of all, a standard argument, based on the restricted strong convexity condition, yields that the nuclear-norm penalized regression yields:

$$\frac{1}{NT}\|\widetilde{M} - M\|_F^2 = O_P(C_{NT}^2), \quad C_{NT} = \frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}.$$

The novelty of the proof is to show the asymptotic normality of the debiased estimators for low-rank inference, where we propose a new inference algorithm that is different from the existing literature (e.g., Chen et al. (2019)), and is more suitable in the context of asset pricing.

Let $\widetilde{\beta}$ be $N \times K$ whose columns are $\sqrt{N}$ times the top K left singular vectors of $\widetilde{M}$. Then by the sine-theta inequality, there is matrix H_1 such that $\frac{1}{\sqrt{N}}\|\widetilde{\beta} - \beta_l H_1\|_F = O_P(C_{NT})$. Next, by definition, $\widehat{F}_t = (\sum_{i=1}^{N} \omega_{it}\widetilde{\beta}_i\widetilde{\beta}_i')^{-1}\sum_{i=1}^{N}\omega_{it}\widetilde{\beta}_i z_{it}$, where $\omega_{it} = 1\{r_{it}$ is not missing$\}$. Let $\widetilde{B}_t = \frac{1}{N}\sum_{i=1}^{N}\omega_{it}\widetilde{\beta}_i\widetilde{\beta}_i'$ and $B = H_1'\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(\omega_{it})\widetilde{\beta}_i\widetilde{\beta}_i' H_1$.

Then let $H_2 := H_1^{-1} + B^{-1}H'\frac{1}{N}\sum_{i=1}^{N}(\mathbb{E}\omega_{it})\beta_{l,i}(\beta_{l,i}'H_1 - \widetilde{\beta}_i)H_1^{-1}$. Both B and H_2 are independent of (i, t). Basic algebras show the following identity:

$$\begin{aligned}
\widehat{F}_t - H_2 F_t &= \widetilde{B}_t^{-1} H_1' \frac{1}{N}\sum_{j=1}^{N}\omega_{jt}\beta_j(\varepsilon_{jt} + \kappa_{jt} + g_{jt}) + \sum_{d=1}^{4}\Delta_{t,d}, \\
\Delta_{t,1} &= -B^{-1}H_1'\frac{1}{N}\sum_{j=1}^{N}(\omega_{jt} - \mathbb{E}\omega_{jt})\beta_{l,j}F_t' H_1^{'-1}(\widetilde{\beta}_j - H_1'\beta_{l,j}) + \widetilde{B}_t^{-1}\frac{1}{N}\sum_{j=1}^{N}\omega_{jt}\varepsilon_{jt}(\widetilde{\beta}_j - H_1'\beta_{l,j}),
\end{aligned}$$

$$
\begin{aligned}
\Delta_{t,2} &= (\widetilde{B}_t^{-1}-B^{-1})\frac{1}{N}\sum_{j=1}^{N}\omega_{jt}\widetilde{\beta}_j(\beta_{l,j}'H_1-\widetilde{\beta}_j')H_1^{-1}F_t,\\
\Delta_{t,3} &= B^{-1}\frac{1}{N}\sum_{j=1}^{N}\omega_{jt}(\widetilde{\beta}_j-H_1'\beta_{l,j})(\beta_{l,j}'H_1-\widetilde{\beta}_j')H_1^{-1}F_t,\\
\Delta_{t,4} &= \widetilde{B}_t^{-1}\frac{1}{N}\sum_{j=1}^{N}\omega_{jt}(\kappa_{jt}+g_{jt})(\widetilde{\beta}_j-H_1'\beta_{l,j}).
\end{aligned}
\tag{B.34}
$$

Next, let $\widehat{A}_i=\frac{1}{T}\sum_{t=1}^{T}\omega_{it}\widehat{F}_t\widehat{F}_t'$. By definition, $\widehat{\beta}_{l,i}=(\sum_{t=1}^{T}\omega_{it}\widehat{F}_t\widehat{F}_t')^{-1}\sum_{t=1}^{T}\omega_{it}\widehat{F}_tz_{it}$. Substituting (B.34), basic algebras show the following identity:

$$
\begin{aligned}
\widehat{\beta}_{l,i}-H_2'^{-1}\beta_{l,i} &= H_2'^{-1}(\frac{1}{T}\sum_{t=1}^{T}\omega_{it}F_tF_t')^{-1}\frac{1}{T}\sum_{t=1}^{T}\omega_{it}F_t(\varepsilon_{it}+g_{it})+\sum_{d=1}^{6}\delta_{i,d}\\
\delta_{i,1} &= \widehat{A}_i^{-1}\frac{1}{T}\sum_{t=1}^{T}\omega_{it}\widetilde{B}_t^{-1}H_1'\frac{1}{N}\sum_{j=1}^{N}\omega_{jt}\beta_j(\varepsilon_{jt}+\kappa_{jt}+g_{jt})(\varepsilon_{it}+\kappa_{it}+g_{it})\\
\delta_{i,2} &= \widehat{A}_i^{-1}\frac{1}{T}\sum_{t=1}^{T}\omega_{it}\widehat{F}_t\frac{1}{N}\sum_{j}\omega_{jt}(\varepsilon_{jt}+\kappa_{jt}+g_{jt})\beta_j'H_1\widetilde{B}_t^{-1}H_2'^{-1}\beta_{l,i}\\
\delta_{i,3} &= \sum_{d=1}^{4}\widehat{A}_i^{-1}\frac{1}{T}\sum_{t=1}^{T}\omega_{it}\widehat{F}_t\Delta_{t,d}'H_2'^{-1}\beta_{l,i}\\
\delta_{i,4} &= \sum_{d=1}^{4}\widehat{A}_i^{-1}\frac{1}{T}\sum_{t=1}^{T}\omega_{it}\Delta_{t,d}(\varepsilon_{jt}+\kappa_{jt}+g_{jt})\\
\delta_{i,5} &= (\widehat{A}_i^{-1}-A_i^{-1})H_2\frac{1}{T}\sum_{t=1}^{T}\omega_{it}F_t(\varepsilon_{it}+\kappa_{it}+g_{it})\\
\delta_{i,6} &= H_2'^{-1}(\frac{1}{T}\sum_{t=1}^{T}\omega_{it}F_tF_t')^{-1}\frac{1}{T}\sum_{t=1}^{T}\omega_{it}F_t\kappa_{it}
\end{aligned}
\tag{B.35}
$$

where $A_i=H_2\frac{1}{T}\sum_{t=1}^{T}\omega_{it}F_tF_t'H_2'$ and we note $A_i^{-1}H_2=H_2'^{-1}(\frac{1}{T}\sum_{t=1}^{T}\omega_{it}F_tF_t')^{-1}$.

We shall bound $\max_i\|\delta_{i,1}+\delta_{i,2}\|$ in Lemma B.3. Note that $\max_i\|\widehat{A}_i-A_i\|=o_P(1)$. Also, $\min_i\psi_{\min}(A_i)>c_0$. So $\max_i\|\widehat{A}_i^{-1}\|=O_P(1)$. Similarly, $\max_t\|\widetilde{B}_t^{-1}\|=O_P(1)$. So

$$
\max_i\|A_i^{-1}-\widehat{A}_i^{-1}\|\leq O_P(1)\max_i\|A_i-\widehat{A}_i\|=O_P(1)\max_i\|\frac{1}{T}\sum_{t=1}^{T}\omega_{it}(\widehat{F}_t\widehat{F}_t'-H_2F_tF_t'H_2')\|=O_P(C_{NT}).
$$

This implies $\max_i\|\delta_{i,5}\|=o_P(\frac{1}{\sqrt{T}\log N})$. Also, it is easy to see $\max_i\|\delta_{i,6}\|=o_P(\frac{1}{\sqrt{T}\log N})$.

As for $\max_i\|\delta_{i,3}+\delta_{i,4}\|$, we note $\max_i\frac{1}{T}\sum_t\varepsilon_{it}^2\leq\max_i|\frac{1}{T}\sum_t\varepsilon_{it}^2-\mathbb{E}\varepsilon_{it}^2|+O(1)=O_P(1)$. Hence $\max_i\|\delta_{i,3}+\delta_{i,4}\|=O_P(1)\sum_{d=1}^{4}\sqrt{\frac{1}{T}\sum_t\|\Delta_{t,d}\|^2}$.

Bounding $\frac{1}{T}\sum_t\|\Delta_{t,1}\|^2$ is more technically involved because it is challenging to directly obtain an expansion for $\widetilde{\beta}_j-H_1'\beta_{l,j}$. Meanwhile, the proof of $\max_i\|\delta_{1,i}+\delta_{2,i}\|+\sum_{d\neq1}\sqrt{\frac{1}{T}\sum_t\|\Delta_{t,d}\|^2}=$

$o_P(\frac{1}{\sqrt{T\log N}})$ is given in Lemma B.3. Then uniformly in i,

$$\begin{aligned}\widehat{\beta}_{l,i} - H_2^{'-1}\beta_{l,i} &= H_2^{'-1}(\frac{1}{T}\sum_{t=1}^{T}\omega_{it}F_tF_t')^{-1}\frac{1}{T}\sum_{t=1}^{T}\omega_{it}F_t\varepsilon_{it} + o_P(\frac{1}{\sqrt{T\log N}}) \\ &= H_2^{'-1}(\frac{1}{T}\sum_{t=1}^{T}F_tF_t')^{-1}\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}F_t\varepsilon_{it} + o_P(\frac{1}{\sqrt{T\log N}}),\end{aligned} \tag{B.36}$$

where the last equality follows from the fact that $\max_i \|\frac{1}{T}\sum_{t=1}^{T}(\omega_{it}-\mathbb{E}\omega_{it})F_tF_t'\|$, $\max_i \|\frac{1}{T}\sum_{t=1}^{T}(\omega_{it}-\mathbb{E}\omega_{it})\|$ and $\max_i \|\frac{1}{T}\sum_{t=1}^{T}\omega_{it}F_t\varepsilon_{it}\|$ are all $O_P(\sqrt{\frac{\log N}{T}})$.

It remains to prove $\frac{1}{T}\sum_t \|\Delta_{t,1}\|^2 = o_P(\frac{1}{T\log N})$.

We focus on bounding $\xi := \frac{1}{T}\sum_t \|\frac{1}{N}\sum_{j=1}^{N}(\omega_{jt}-\mathbb{E}\omega_{jt})\beta_{l,j}F_t'H_1^{'-1}(\widetilde{\beta}_j - H_1'\beta_{l,j})\|^2$. To achieve a sharp bound, we apply a computation result and the auxiliary leave-one-out argument in Chen et al. (2019). Define for each $t \le T$,

$$\begin{aligned}(\widetilde{W},\widetilde{Y}) &= \arg\min_{W\in\mathbb{R}^{N\times K}}\min_{Y\in\mathbb{R}^{T\times K}} \|\Omega\circ(Z-WY')\|_F^2 + \lambda_{NT}\|W\|_F^2 + \lambda_{NT}\|Y\|_F^2 \\ (\widetilde{W}^{(-t)},\widetilde{Y}^{(-t)}) &= \arg\min_{W,Y\in\mathbb{R}^{N\times K}} \|\Omega^{(-t)}\circ(Z-WY')\|_F^2 + \|E^{(t)}\circ(M-WY')\|_F^2 + \lambda_{NT}\|W\|_F^2 + \lambda_{NT}\|Y\|_F^2\end{aligned}$$

where we recall M is the true value of $\beta_l F$,

$$\Omega^{(-t)} = (\omega_{is}1\{s\neq t\})_{N\times T}, \quad E^{(t)} = (\sqrt{\mathbb{E}\omega_{is}}1\{s=t\})_{N\times T}.$$

So $(\widetilde{W}^{(-t)},\widetilde{Y}^{(-t)})$ are the "leave-one-out" versions of $(\widetilde{W},\widetilde{Y})$. Importantly, $(\widetilde{W}^{(-t)},\widetilde{Y}^{(-t)})$ are independent of $Z_t, (\omega_{it} : i\le N)$. We now apply three results from Chen et al. (2019), which are their Lemma 12, Lemma 18(3), Lemma 5, and Lemma 2:

(a) There is a $K\times K$ orthonormal matrix $H^{(-t)}$ that only depends on $\widetilde{W}^{(-t)}$ and $\beta_l F$, and another orthonormal matrix H_3

$$\max_t \|\widetilde{W}^{(-t)}H^{(-t)} - \widetilde{W}H_3\|_F = o_P((\frac{N}{T})^{1/4}\frac{1}{\sqrt{\log N}}).$$

(b) Let $U_M D_M V_M'$ be the SVD of $M = \beta_l F$, and $W = U_M D_M^{1/2}$. Then

$$\|\widetilde{W}H_3 - W\|_F = O_P(C_{NT}(NT)^{1/4}).$$

(c) $\|\widetilde{W}\widetilde{Y}' - \widetilde{M}\|_F = o_P(\sqrt{N/\log N})$.

Strictly speaking, Chen et al. (2019) considered the case $N = T$. However, by carefully examining their proofs, the proof of (A.12) (A.9a) and (A.14b) still carries over to the case when $N \neq T$.

We are now ready to bound ξ. We shall use a generic notation μ throughout the following steps without causing confusions.

Step 1. There is a rotation $H_4 = O_P(1)$ so that $\frac{1}{(NT)^{1/4}}\widetilde{W}H_4$ equals the left singular-vectors of $\widetilde{W}\widetilde{Y}'$. So let $\Omega_t = (\omega_{1t}, ..., \omega_{Nt})'$,

$$
\begin{aligned}
\xi &= \frac{1}{T}\sum_t \|\frac{1}{N}\beta_l'\mathrm{diag}(\Omega_t - \mathbb{E}\Omega_t)(\widetilde{\beta} - \beta_l H_1)H_1^{-1}F_t\|^2 \leq 4\xi_1 + 4\mu \\
\xi_1 &= \frac{1}{T}\sum_t \|\frac{1}{N}\beta_l'\mathrm{diag}(\Omega_t - \mathbb{E}\Omega_t)(\frac{\sqrt{N}}{(NT)^{1/4}}\widetilde{W}H_4 - \beta_l H_1)H_1^{-1}F_t\|^2 \\
\mu &= \frac{1}{T}\sum_t \|\frac{1}{N}\beta_l'\mathrm{diag}(\Omega_t - \mathbb{E}\Omega_t)(\widetilde{\beta} - \frac{\sqrt{N}}{(NT)^{1/4}}\widetilde{W}H_4)H_1^{-1}F_t\|^2.
\end{aligned}
$$

By (c) and sine-theta inequality, $\mu \leq O_P(\frac{1}{N})\|\widetilde{\beta} - \frac{\sqrt{N}}{(NT)^{1/4}}\widetilde{W}H_4\|^2 \leq O_P(\frac{1}{NT})\|\widetilde{M} - \widetilde{W}\widetilde{Y}'\|_F^2 = o_P(\frac{1}{T\log N})$. The problem then becomes bounding ξ_1.

Step 2. Note that

$$
\begin{aligned}
\xi_1 &\leq 4\xi_2 + 4\mu \\
\xi_2 &= \frac{1}{T}\sum_t \|\frac{1}{N}\beta_l'\mathrm{diag}(\Omega_t - \mathbb{E}\Omega_t)(\sqrt{N}\widetilde{W}H_3 D_M^{-1/2} - \beta_l H_1)H_1^{-1}F_t\|^2 \\
\mu &= \frac{1}{T}\sum_t \|\frac{1}{N}\beta_l'\mathrm{diag}(\Omega_t - \mathbb{E}\Omega_t)\frac{\sqrt{N}}{(NT)^{1/4}}\widetilde{W}(H_4 - H_3 D_M^{-1/2}(NT)^{1/4})H_1^{-1}F_t\|^2.
\end{aligned}
$$

We now bound μ. For notational simplicity, we shall assume $\dim(F_t) = 1$ as one can apply an element-by-element analysis for the multivariate case without changing the result, given that the number of factors is fixed. We first have, by (b)

$$
\begin{aligned}
\|H_3 D_M^{-1/2}(NT)^{1/4} - H_4\|_F &\leq (NT)^{1/4}\|(\widetilde{W}'\widetilde{W})^{-1}\widetilde{W}'\|(NT)^{-1/4}\|\widetilde{W}H_3 D_M^{-1/2}(NT)^{1/4} - \widetilde{W}H_4\|_F \\
&\leq O_P(1)\|\widetilde{W}H_3 - W\|_F\|D_M^{-1/2}\| + (NT)^{-1/4}\|WD_M^{-1/2}(NT)^{1/4} - \widetilde{W}H_4\|_F \\
&\leq O_P(C_{NT}) + \|U_M - (NT)^{-1/4}\widetilde{W}H_4\|_F \\
&\leq O_P(C_{NT}) + O_P((NT)^{-1/2})\|\widetilde{W}\widetilde{Y}' - M\|_F = O_P(C_{NT}).
\end{aligned}
$$

Let $d := \frac{\sqrt{N}}{(NT)^{1/4}}W$. Then $\|d\|_\infty < C$. Hence by (b) and

$$
\begin{aligned}
\mu &= O_P(1)\frac{1}{T}\sum_t \|F_t\frac{1}{N}\beta_l'\mathrm{diag}(\Omega_t - \mathbb{E}\Omega_t)\frac{\sqrt{N}}{(NT)^{1/4}}\widetilde{W}\|^2\|H_4 - H_3 D_M^{-1/2}(NT)^{1/4}\|^2 \\
&\leq O_P(C_{NT}^2)\frac{1}{T}\sum_t \|F_t\frac{1}{N}\beta_l'\mathrm{diag}(\Omega_t - \mathbb{E}\Omega_t)\|^2\|\frac{\sqrt{N}}{(NT)^{1/4}}(\widetilde{W}H_3 - W)\|^2 \\
&\quad + O_P(C_{NT}^2)\frac{1}{T}\sum_t \|F_t\frac{1}{N}\beta_l'\mathrm{diag}(\Omega_t - \mathbb{E}\Omega_t)\frac{\sqrt{N}}{(NT)^{1/4}}W\|^2 \\
&\leq O_P(C_{NT}^4) + O_P(C_{NT}^2)\frac{1}{T}\sum_t (F_t\frac{1}{N}\sum_j \beta_{lj}d_j(\omega_{jt} - \mathbb{E}\omega_{jt}))^2 = O_P(C_{NT}^4) = o_P(\frac{1}{T\log N})
\end{aligned}
$$

given that $\frac{1}{N}\sum_{ij}|\operatorname{Cov}(\omega_{jt},\omega_{it})|<C$. The problem then becomes bounding ξ_2.

Step 3. Note that

$$\begin{aligned}
\xi_2 &\leq 4\xi_3+4\mu \\
\xi_3 &= \frac{1}{T}\sum_t \|\frac{1}{N}\beta_l'\operatorname{diag}(\Omega_t-\mathbb{E}\Omega_t)(\sqrt{N}\widetilde{W}^{(-t)}H^{(-t)}D_M^{-1/2}-\beta_l H_1)H_1^{-1}F_t\|^2 \\
\mu &= N\frac{1}{T}\sum_t \|\frac{1}{N}\beta_l'\operatorname{diag}(\Omega_t-\mathbb{E}\Omega_t)(\widetilde{W}H_3-\widetilde{W}^{(-t)}H^{(-t)})D_M^{-1/2}H_1^{-1}F_t\|^2.
\end{aligned}$$

We now bound μ. By (a) $\mu\leq O_P((NT)^{-1/2})\max_t\|\widetilde{W}H_3-\widetilde{W}^{(-t)}H^{(-t)}\|_F^2=o_P(\frac{1}{T\log N})$. The problem then becomes bounding ξ_3.

Step 4. To bound ξ_3, we still consider the case $\dim(F_t)=1$ without loss of generality. Note that $\frac{1}{\sqrt{N}}\beta_l H_1$ equals the top left eigenvectors of M. Let $p_t:=(\sqrt{N}\widetilde{W}^{(-t)}H^{(-t)}D_M^{-1/2}-\beta_l H_1)$. Note that p_t only depends on M and Z excluding the t th column of Z. As such, for $P:=(p_t:t\leq T)$ we have $\mathbb{E}(\beta_l'\operatorname{diag}(\Omega_t-\mathbb{E}\Omega_t)p_tH_1^{-1}F_t|\beta_l,F,P)=0$, therefore

$$\begin{aligned}
\mathbb{E}(\xi_3|\beta_l,F,P) &= O_P(\frac{1}{N^2})\frac{1}{T}\sum_t F_t^2p_t'Sp_t\leq O_P(\frac{1}{N^2})\max_t\|p_t\|^2 \\
&\leq O_P(\frac{1}{N^2})\max_t\|\sqrt{N}U_M-\beta_lH_1\|_F^2+O_P(\frac{1}{N\sqrt{NT}})\max_t\|(\widetilde{W}H_3-W)\|_F^2 \\
&\quad +O_P(\frac{1}{N^2\sqrt{T}})\max_t\|(\widetilde{W}^{(-t)}H^{(-t)}-\widetilde{W}H_3)\|_F^2=o_P(\frac{1}{T\log N}),
\end{aligned}$$

where $S=\operatorname{Var}(\beta_l'\operatorname{diag}(\Omega_t-\mathbb{E}\Omega_t)|\beta_l,F,P)$ and almost surely,

$$\|S\|_1=\max_{i\leq N}\sum_j|S_{ij}|\leq\max_{i\leq N}\sum_j|\beta_i\beta_j||\operatorname{Cov}(\omega_{it},\omega_{jt})|<C.$$

Hence $\xi_3=o_P(\frac{1}{T\log N})$. This proves $\xi=o_P(\frac{1}{T\log N})$.

The bound for $\frac{1}{T}\sum_t\|\frac{1}{N}\sum_{j=1}^N\omega_{jt}\varepsilon_{jt}(\widetilde{\beta}_j-H_1'\beta_{l,j})\|^2$ is very similar. We just need to replace S in Step 4 with $\tilde{S}=\operatorname{Var}(\varepsilon_t'\operatorname{diag}(\Omega_t)|\beta_l,F,P)$, whose $\|.\|_1$ norm is also bounded by a constant. This completes the proof.

□

Next, we prove the following lemmas.

Lemma B.1. $\max_{i\leq N}\|\frac{1}{T}\sum_t u_{it}^m f_{kt}^n f_{qt}^v-\mathbb{E}u_{it}^m f_{kt}^n f_{qt}^v\|=O_P(\sqrt{\frac{\log N}{T}})$, *for* $m,n,v\in\{0,1,2\}$ *for any* $q,k\leq K$. *Also,* $\max_{ij}\|\frac{1}{T}\sum_t(u_{it}u_{jt}-\mathbb{E}u_{it}u_{jt})\|=O_P(\sqrt{\frac{\log N}{T}})$, $\max_{ijk}\|\frac{1}{T}\sum_t(u_{it}^2u_{jt}u_{kt}-\mathbb{E}u_{it}^2u_{jt}u_{kt})\|=O_P(\sqrt{\frac{\log N}{T}})$, *and for* $w_t=\frac{1}{\sqrt{N}}\beta'u_t$, $\max_{i\leq N}\|\frac{1}{T}\sum_t u_{it}^dw_t^d-\mathbb{E}u_{it}^dw_t^d\|=O_P(\sqrt{\frac{\log N}{T}})$ *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 Ω_t as $u_{it}^m f_{kt}^n f_{qt}^v$, $u_{it}u_{jt}$ and $u_{it}w_t$.

□

Lemma B.2. *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 A.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, $\mathbb{P}(\max_{i\leq N}|Y_i| > s) \leq 2N(1-\Phi(s)) \leq 2\exp(\log N - s^2/2) = o(1)$ 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 B.1 .

□

Lemma B.3. *Recall the definitions of $\delta_{1,i}$, $\delta_{2,i}$, $\Delta_{t,d}$ in (B.34) and (B.35). We have $\max_i\|\delta_{1,i}\| = o_P(\frac{1}{\sqrt{T}\log N}) = \max_i\|\delta_{2,i}\| = o_P(\frac{1}{\sqrt{T}\log N})$, and $\sum_{d=2}^4\sqrt{\frac{1}{T}\sum_t\|\Delta_{t,d}\|^2} = o_P(\frac{1}{\sqrt{T}\log N})$.*

Proof. First note that $\max_t\|\widetilde{B}_t - B\| = o_P(1)$ and thus $\max_t\|\widetilde{B}_t^{-1}\| = O_P(1)$, so

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^T\|\widetilde{B}_t^{-1} - B^{-1}\|^2 \quad &\leq \quad O_P(1)\frac{1}{T}\sum_{t=1}^T\|\widetilde{B}_t - B\|^2 \leq O_P(1)\frac{1}{T}\sum_{t=1}^T\|\frac{1}{N}\sum_{i=1}^N\omega_{it}(\widetilde{\beta}_i\widetilde{\beta}_i' - H_1'\beta_{l,i}\beta_{l,i}'H_1)\|^2 \\ &+O_P(1)\frac{1}{T}\sum_{t=1}^T\|\frac{1}{N}\sum_{i=1}^N(\omega_{it} - \mathbb{E}\omega_{it})\beta_{l,i}\beta_{l,i}'\|^2 = O_P(C_{NT}^2).\end{aligned}$$

Similarly, $\frac{1}{T}\sum_{t=1}^T \|F_t\|^2\|\widetilde{B}_t^{-1}-B^{-1}\|^2 = O_P(C_{NT}^2)$.

(i) Recall that $\delta_{i,1} = \widehat{A}_i^{-1}\frac{1}{T}\sum_{t=1}^T \omega_{it}\widetilde{B}_t^{-1}H_1'\frac{1}{N}\sum_{j=1}^N \omega_{jt}\beta_j(\widetilde{\varepsilon}_{jt}+g_{jt})(\widetilde{\varepsilon}_{it}+g_{it})$ where we write $\widetilde{\varepsilon}_{it} = \varepsilon_{it}+\kappa_{it}$. Then $\max_i \|\delta_{1,i}\| \leq I + II + III + o_P(\frac{1}{\sqrt{T\log N}})$ with

$$
\begin{aligned}
I &\leq O_P(1)\max_i \|(\frac{1}{T}\sum_{t=1}^T \|\widetilde{B}_t^{-1}-B^{-1}\|^2)^{1/2}(\frac{1}{T}\sum_{t=1}^T \|\frac{1}{N}\sum_{j=1}^N \omega_{jt}\beta_j\widetilde{\varepsilon}_{jt}\widetilde{\varepsilon}_{it}\|^2)^{1/2} \\
&\quad + O_P(1)\max_i \|\frac{1}{TN}\sum_{t=1}^T\sum_{j=1}^N \omega_{it}\omega_{jt}\beta_j\widetilde{\varepsilon}_{jt}\widetilde{\varepsilon}_{it}\| \\
&\leq O_P(C_{NT})\max_i(\frac{1}{T}\sum_t \widetilde{\varepsilon}_{it}^4)^{1/4}(\frac{1}{T}\sum_{t=1}^T\|\frac{1}{N}\sum_{j=1}^N \omega_{jt}\beta_j\widetilde{\varepsilon}_{jt}\|^4)^{1/4} \\
&\quad + O_P(1)\max_i \|\frac{1}{TN}\sum_{t=1}^T\sum_{j=1}^N \omega_{it}\omega_{jt}\beta_j\mathbb{E}\varepsilon_{jt}\varepsilon_{it}\| \\
&\quad + O_P(1)\max_i \|\frac{1}{TN}\sum_{t=1}^T\sum_{j=1}^N \omega_{it}\omega_{jt}\beta_j(\varepsilon_{jt}\varepsilon_{it}-\mathbb{E}\varepsilon_{jt}\varepsilon_{it})\| + o_P(\frac{1}{\sqrt{T\log N}}) \\
&= O_P(C_{NT}\sqrt{\log(NT)}N^{-1/2} + N^{-1} + \sqrt{\frac{\mathrm{Var}(\frac{1}{N}\sum_j \omega_{jt}\beta_j\varepsilon_{jt}\varepsilon_{it})\log N}{T}}) = o_P(\frac{1}{\sqrt{T\log N}}),
\end{aligned}
$$

given that $\max_{it}\frac{1}{N}\sum_{j,k}|\mathrm{Cov}(\varepsilon_{jt}\varepsilon_{it},\varepsilon_{kt}\varepsilon_{it})| < C$. Also, we have

$$
\begin{aligned}
II &\leq O_P(1)\max_i \|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\sum_{j=1}^N \omega_{it}\omega_{jt}\beta_j\widetilde{\varepsilon}_{jt}g_{it})\| = o_P(\frac{1}{\sqrt{T\log N}}), \\
III &\leq O_P(1)\max_i \|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\sum_{j=1}^N \omega_{it}\omega_{jt}\beta_j\widetilde{\varepsilon}_{it}g_{jt})\| = o_P(\frac{1}{\sqrt{T\log N}}).
\end{aligned}
$$

(ii) Note that $\frac{1}{T}\sum_{t=1}^T(\widehat{F}_t - H_2F_t)^2 = O_P(C_{NT}^2)$.

$$
\begin{aligned}
\max_i \|\delta_{2,i}\| &\leq O_P(1)(\frac{1}{T}\sum_{t=1}^T(\widehat{F}_t-H_2F_t)^2)^{1/2}(\frac{1}{T}\sum_{t=1}^T\|\frac{1}{N}\sum_j \omega_{jt}(\widetilde{\varepsilon}_{jt}+g_{jt})\beta_j'\|^2)^{1/2} \\
&\quad + O_P(1)(\frac{1}{T}\sum_{t=1}^T\|\frac{1}{N}\sum_j \omega_{jt}F_t(\widetilde{\varepsilon}_{jt}+g_{jt})\beta_j'\|^2)^{1/2}(\frac{1}{T}\sum_{t=1}^T(\widetilde{B}_t^{-1}-B^{-1})^2)^{1/2} \\
&\quad + O_P(1)\max_i \|\frac{1}{TN}\sum_{t=1}^T\sum_j \omega_{it}\omega_{jt}F_t\widetilde{\varepsilon}_{jt}\beta_j'\| + O_P(1)\max_i\|\frac{1}{TN}\sum_{t=1}^T\sum_j \omega_{it}\omega_{jt}F_tg_{jt}\beta_j'\| \\
&= O_P(\sqrt{\frac{\log N}{NT}} + C_{NT}^2) = o_P(\frac{1}{\sqrt{T\log N}}),
\end{aligned}
$$

where we used

$$
\max_i \|\frac{1}{TN}\sum_{t=1}^T\sum_j \omega_{it}\omega_{jt}F_tg_{jt}\beta_j'\| \leq O_P(1)(\frac{1}{T}\sum_t(\frac{1}{N}\sum_j \omega_{jt}g_{jt}\beta_j)^2)^{1/2} = o_P(\frac{1}{\sqrt{T\log N}}).
$$

(iii) By Cauchy-Schwarz, we have

$$\begin{aligned}
\frac{1}{T}\sum_t \|\Delta_{t,2}\|^2 &\leq O_P(1)\frac{1}{T}\sum_t \|(\widetilde{B}_t^{-1}-B^{-1})\|_F^2\|F_t\|^2\frac{1}{N}\sum_{j=1}^N\|\beta_{l,j}'H_1-\widetilde{\beta}_j'\|^2 = o_P(\frac{1}{T\log N}),\\
\frac{1}{T}\sum_t \|\Delta_{t,3}\|^2 &\leq O_P(1)\frac{1}{T}\sum_t \|F_t\|^2(\frac{1}{N}\sum_{j=1}^N\|\widetilde{\beta}_j-H_1'\beta_{l,j}\|^2)^2 = o_P(\frac{1}{T\log N}),\\
\frac{1}{T}\sum_t \|\Delta_{t,4}\|^2 &\leq O_P(C_{NT}^2)\max_j\frac{1}{T}\sum_t(\kappa_{jt}^2+g_{jt}^2) = o_P(\frac{1}{T\log N}).
\end{aligned}$$

□

Lemma B.4. *Suppose* $\min_i \mathbb{E}\omega_{it} > c$. *For any deterministic and bounded sequence* $\{b_i : i \leq N\}$,
(i) $\max_{i\leq N}\|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\zeta_{it}-\frac{1}{T}\sum_t\zeta_{it}\| = O_P(\sqrt{\frac{\log N}{T}})$ *for* $\zeta_{it}\in\{v_{o,t}v_{o,t}', v_{o,t}, v_{l,t}, v_{o,t}v_{l,t}, u_{it}\}$.
(ii) $\|\frac{1}{N}\sum_{i=1}^N b_i(\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\zeta_{it}-\frac{1}{T}\sum_t\zeta_{it})\| = o_P(\frac{1}{\sqrt{T\log N}})$.
(iii) $\frac{1}{T}\sum_t\|\frac{1}{N}\sum_i\omega_{it}b_i(\bar{\zeta}_i-\bar{\zeta})\eta_t\|^2 = o_P(\frac{1}{T\log N})$ *for* $\eta_t\in\{v_t, v_tv_t', 1\}$ *where* $\bar{\zeta}_i = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\zeta_t$ *for* $\zeta_t\in\{v_t, v_{l,t}v_{o,t}', v_{o,t}v_{o,t}'\}$.
(iv) $\|\frac{1}{N}\sum_{i=1}^N b_i(H_{o,i}-H_o)\| = o_P(\frac{1}{\sqrt{T\log N}}) = \|\frac{1}{N}\sum_i b_i(H_{o,i}-H_o)\frac{1}{T}\sum_t\zeta_t\omega_{it}\|$, *where* $\zeta_t\in\{v_t, v_{l,t}v_{o,t}', v_{o,t}v_{o,t}'\}$, $H_{o,i} = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}v_{l,t}v_{o,t}'S_{o,i}^{-1}$, *and* $H_o = \frac{1}{T}\sum_t v_{l,t}v_{o,t}'S_o^{-1}$. *Also,* $\frac{1}{T}\sum_t\|\frac{1}{N}\sum_i\omega_{it}b_i(H_{o,i}-H_o)\eta_t\|^2 = o_P(\frac{1}{T\log N})$ *for* $\eta_t\in\{v_t, v_tv_t', 1\}$.
(v) $\|\frac{1}{N}\sum_i\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}(l_{it}-l_t)l_t'b_i\| = o_P(\frac{1}{\sqrt{T\log N}})$ *and* $\frac{1}{T}\sum_{t=1}^T\|\frac{1}{N}\sum_{i=1}^N\omega_{it}b_i(l_{it}-l_t)m_t\|^2 = o(\frac{1}{T\log N})$ *where* $l_{it} = (v_{l,t}-\bar{v}_{l,i})-H_{o,i}(v_{o,t}-\bar{v}_{o,i})$, *and* $l_t = (v_{l,t}-\bar{v}_l)-H_o(v_{o,t}-\bar{v}_o)$, *for all* $m_t\in\{1, v_t\}$.
(vi) All these terms are $o_P(\frac{1}{\sqrt{T\log N}})$*:* $\max_i\|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\widehat{v}_{l,t}-H_l^{-1}l_t\|$, $\max_i\|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\widehat{v}_t-H^{-1}v_t\|$, $\max_i\|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}(\widehat{v}_{l,t}(f_{o,t}-\bar{f}_{o,i})'-H_l^{-1}l_tv_{o,t}')\|$ *and* $\max_i\|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}((f_{o,t}-\bar{f}_{o,i})(f_{o,t}-\bar{f}_{o,i})'-v_{o,t}v_{o,t}')\|$.
(vii) $\max_i\|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\widehat{v}_t\widehat{v}_t'-H^{-1}\frac{1}{T}\sum_t v_tv_t'H'^{-1}\| = O_P(\sqrt{\frac{\log N}{T}})$, *and* $\max_i\|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\widehat{v}_{l,t}\widehat{v}_{o,t}'\| = O_P(\sqrt{\frac{\log N}{T}})$.

Proof. (i) For $p_i := \mathbb{E}\omega_{it}$, conditioning on ζ_{it}, ω_{it} are independent across (i,t). Hence

$$\max_i\|\frac{1}{T}\sum_t\zeta_{it}(\omega_{it}-p_i)\| = O_P(\sqrt{\frac{\log N}{T}}) = \max_i\|\frac{1}{T}\sum_t(\omega_{it}-p_i)\|.$$

Also, $\min_i p_i > c$. This implies: $\max_i\|\frac{1}{Tp_i}\sum_t\zeta_{it}(\omega_{it}-p_i)\| = O_P(\sqrt{\frac{\log N}{T}})$, $\max_i\|\frac{1}{T}\sum_t\zeta_{it}\omega_{it}\| = O_P(1)$, and $\min_i\frac{1}{T}\sum_t\omega_{it} > c$. Therefore, the result follows from the following identity:

$$\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\zeta_{it}-\frac{1}{T}\sum_t\zeta_{it} = \frac{p_i-\frac{1}{T}\sum_t\omega_{it}}{p_i\frac{1}{T}\sum_t\omega_{it}}\frac{1}{T}\sum_t\zeta_{it}\omega_{it}+\frac{1}{Tp_i}\sum_t\zeta_{it}(\omega_{it}-p_i).$$

(ii) Without loss of generality, we assume ζ_{it} is a scalar as the analysis can be carried out elementwise. We have

$$\frac{1}{N}\sum_{i=1}^N b_i(\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\zeta_{it}-\frac{1}{T}\sum_t\zeta_{it}) = I+II+o_P(\frac{1}{\sqrt{T\log N}}),\quad\text{where}$$

$$I \;:=\; \frac{1}{T^2N}\sum_{i=1}^N \frac{b_i}{p_i^2}\sum_s\sum_t \zeta_{it}\omega_{it}(p_i-\omega_{is}), \quad II = \frac{1}{N}\sum_{i=1}^N \frac{b_i}{Tp_i}\sum_t \zeta_{it}(\omega_{it}-p_i).$$

It is straightforward to show $\mathbb{E}I^2 = O(\frac{1}{T^2})$ and $\mathbb{E}II^2 = O(\frac{1}{TN})$. Hence this leads to the desired result.

(iii) $\frac{1}{T}\sum_t \|\frac{1}{N}\sum_i \omega_{it}b_i(\bar\zeta_i-\bar\zeta)\eta_t\|^2 \le 4I + 4II + o_P(\frac{1}{T\log N})$ where

$$\begin{aligned} I &= \frac{1}{T}\sum_s \|\frac{1}{N}\sum_i \frac{\omega_{is}b_i}{Tp_i}\sum_t \zeta_{it}\eta_t(\omega_{it}-p_i)\|^2, \\ II &= \frac{1}{T}\sum_s \|\frac{1}{NT^2}\sum_i\sum_t\sum_k \frac{b_i}{p_i^2}\zeta_{it}\eta_t(\omega_{ik}-p_i)\omega_{it}\omega_{is}\|^2. \end{aligned}$$

It is straightforward to show $\mathbb{E}I = O(\frac{1}{TN}+\frac{1}{T^2}) = \mathbb{E}II = o_P(\frac{1}{T\log N})$.

(iv) Let $\xi_t = v_{l,t}v_{o,t}'$. Let $\bar\xi = \frac{1}{T}\sum_t \xi_t$ and $\bar\xi_i = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\xi_t$. By parts (i)(ii), $\max_i \|S_{o,i}^{-1}-S_o^{-1}\| = O_P(\sqrt{\frac{\log N}{T}})$. Therefore,

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^N b_i(H_{o,i}-H_o) = \frac{1}{N}\sum_{i=1}^N b_i(\bar\xi_i-\bar\xi)S_{o,i}^{-1} + \frac{\bar\xi}{N}\sum_{i=1}^N b_i(S_{o,i}^{-1}-S_o^{-1}) \\ =\;& \frac{1}{N}\sum_{i=1}^N b_i(\bar\xi_i-\bar\xi)S_o^{-1} + \frac{\bar\xi}{N}S_o^{-1}\sum_{i=1}^N b_i(S_o-S_{o,i})S_o^{-1} + o_P(\frac{1}{\sqrt{T\log N}}) \\ =\;& o_P(\frac{1}{\sqrt{T\log N}}). \end{aligned}$$

In addition, we have

$$\begin{aligned} &\|\frac{1}{N}\sum_i b_i(H_{o,i}-H_o)\frac{1}{T}\sum_t \zeta_t\omega_{it}\| \\ \le\;& \|\frac{1}{N}\sum_i b_i(\bar\xi_i-\bar\xi)S_{o,i}^{-1}\frac{1}{T}\sum_t\zeta_t\omega_{it}\| + \|\frac{1}{N}\sum_i b_i\bar\xi(S_{o,i}^{-1}-S_o^{-1})\frac{1}{T}\sum_t\zeta_t\omega_{it}\| \\ \le\;& o_P(\frac{1}{\sqrt{T\log N}}) + \|\frac{1}{N}\sum_i b_i(\bar\xi_i-\bar\xi)S_o^{-1}\frac{1}{T}\sum_t\zeta_t\omega_{it}\| + \|\bar\xi S_o^{-1}\frac{1}{N}\sum_i b_i(S_{o,i}-S_o)S_o^{-1}\frac{1}{T}\sum_t\zeta_t\omega_{it}\|. \end{aligned}$$

Using the same argument as in part (ii), it can be shown that both terms are $o_P(\frac{1}{\sqrt{T\log N}})$.

Additionally, for $\zeta_t = v_{l,t}v_{o,t}'$,

$$\begin{aligned} &\frac{1}{T}\sum_t \|\frac{1}{N}\sum_i \omega_{it}b_i(H_{o,i}-H_o)\eta_t\|^2 \\ =\;& \frac{1}{T}\sum_t \|\frac{1}{N}\sum_i \omega_{it}b_i(\bar\zeta_i-\bar\zeta)S_o^{-1}\eta_t\|^2 + \frac{1}{T}\sum_t\|\frac{1}{N}\sum_i \omega_{it}b_i\bar\zeta S_o^{-1}(S_{o,i}-S_o)S_o^{-1}\eta_t\|^2 + o_P(\frac{1}{T\log N}) \\ =\;& o_P(\frac{1}{T\log N}). \end{aligned}$$

The last equality follows from (iii) that $\frac{1}{T}\sum_t \|\frac{1}{N}\sum_i \omega_{it}b_i(\bar\zeta_i-\bar\zeta)\eta_t\|^2 = o_P(\frac{1}{T\log N})$.

(v) First, note that

$$
\begin{aligned}
&\|\frac{1}{N}\sum_i \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}(l_{it}-l_t)l_t'b_i\| \le o_P(\frac{1}{T\log N}) + \|\frac{1}{N}\sum_i \frac{b_i}{p_i}(H_{o,i}-H_o)\frac{1}{T}\sum_t v_{o,t}l_t'\omega_{it}\| \\
&+\|\frac{1}{N}\sum_i \frac{b_i}{p_i}(\bar{v}_{l,i}-\bar{v}_l)\frac{1}{T}\sum_t l_t'\omega_{it}\| + \|\frac{1}{N}\sum_i \frac{b_i}{p_i}H_o(\bar{v}_{o,i}-\bar{v}_o)\frac{1}{T}\sum_t l_t'\omega_{it}\|.
\end{aligned}
$$

Since $\max_i \|\frac{1}{T}\sum_t l_t'\omega_{it}\| = O_P(\sqrt{\frac{\log N}{T}})$, the last two terms are $o_P(\frac{1}{\sqrt{T\log N}})$. The first term follows from part (iii). Next, it follows directly from (iii)(iv) that for $m_t \in \{1, v_t\}$,

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^T \|\frac{1}{N}\sum_{i=1}^N \omega_{it}b_i(l_{it}-l_t)m_t\|^2 \le \frac{1}{T}\sum_{t=1}^T \|\frac{1}{N}\sum_{i=1}^N \omega_{it}b_i(\bar{v}_{l,i}-\bar{v}_l)m_t\|^2 \\
&+\frac{1}{T}\sum_{t=1}^T \|\frac{1}{N}\sum_{i=1}^N \omega_{it}b_i(H_{o,i}-H_o)v_{o,t}m_t\|^2 + \frac{1}{T}\sum_{t=1}^T \|\frac{1}{N}\sum_{i=1}^N \omega_{it}b_i(H_{o,i}-H_o)\bar{v}_{o,i}m_t\|^2 \\
&+\frac{1}{T}\sum_{t=1}^T \|\frac{1}{N}\sum_{i=1}^N \omega_{it}b_iH_o(\bar{v}_{o,i}-\bar{v}_o)m_t\|^2 = o_P(\frac{1}{T\log N}).
\end{aligned}
$$

(vi) We have $\widehat{v}_t = (f_{o,t}-\bar{f}_o, \widehat{v}_{l,t})$. Then

$$
\widehat{v}_t - H^{-1}(v_t-\bar{v}) = \widehat{v}_t - (f_{o,t}'-\bar{f}_o', (H_l^{-1}l_t)')' = (0', (\widehat{v}_{l,t}-H_l^{-1}l_t)')'.
$$

We can apply (B.34) in the proof of Proposition B.2 for $H_2 := H_l^{-1}$, $\widehat{F}_t = \widehat{v}_{l,t}$ and $F_t := l_t$. Hence

$$
\widehat{v}_{l,t} - H_l^{-1}l_t = \widetilde{B}_t^{-1}H_1'\frac{1}{N}\sum_{j=1}^N \omega_{jt}\beta_{l,j}(-\eta_{jt}+u_{jt}-\bar{u}_j+\beta_{l,j}'(l_{jt}-l_t)) + \sum_{d=1}^4 \Delta_{t,d},
$$

where by the proof of Proposition B.2, $\frac{1}{T}\sum_t \|\Delta_{t,d}\|^2 = o_P(\frac{1}{T\log N})$ for $d = 1...4$ and $\max_t \|\widetilde{B}_t^{-1}\| = O_P(1)$. Therefore, $\max_i \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\|\Delta_{t,d}\|^2 = o_P(\frac{1}{T\log N})$ and thus for $m_t \in \{1, v_{o,t}\}$,

$$
\begin{aligned}
&\max_i \|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}(\widehat{v}_{l,t}-H_l^{-1}l_t)m_t'\| \\
\le\ &\max_i \|\frac{1}{T_i}\sum_{t=1}^T \frac{1}{N}\sum_{j=1}^N \omega_{jt}\omega_{it}\beta_{l,j}(-\eta_{jt}+u_{jt}-\bar{u}_j+\beta_{l,j}'(l_{jt}-l_t))m_t'\| + o_P(\frac{1}{\sqrt{T\log N}}) \\
\le\ &o_P(\frac{1}{\sqrt{T\log N}}) + O_P(1)(\frac{1}{T}\sum_{t=1}^T \|\frac{1}{N}\sum_{j=1}^N \omega_{jt}\beta_{l,j}\beta_{l,j}'(l_{jt}-l_t)m_t'\|^2)^{1/2} = o_P(\frac{1}{\sqrt{T\log N}}).
\end{aligned}
$$

The last equality follows from (v). For $\bar{v}_{o,i} = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i} v_{o,t}$,

$$
\max_i \|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}(\widehat{v}_{l,t}(f_{o,t}-\bar{f}_{o,i})'-H_l^{-1}l_tv_{o,t}')\| = \max_i \|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}(\widehat{v}_{l,t}-H_l^{-1}l_t)v_{o,t}' - \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\widehat{v}_{l,t}\bar{v}_{o,i}'\| = o_P(\frac{1}{\sqrt{T\log N}}).
$$

Finally, $\max_i \|\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}((f_{o,t}-\bar{f}_{o,i})(f_{o,t}-\bar{f}_{o,i})' - v_{o,t}v_{o,t}')\| = \max_i \|-\bar{v}_{o,i}\bar{v}_{o,i}'\| = o_P(\frac{1}{\sqrt{T\log N}})$.

(vii) By (vi), $\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\widehat{v}_{l,t}\widehat{v}_{o,t}' = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i} H_l^{-1} l_t v_{o,t}' + o_P(\frac{1}{\sqrt{T}\log N})$, and

$$\frac{1}{T_i}\sum_{t\in\mathcal{T}_i} H_l^{-1} l_t v_{o,t}' + o_P(\frac{1}{\sqrt{T}\log N}) = H_l^{-1}\frac{1}{T}\sum_t (v_{l,t} - H_o v_{o,t})v_{o,t}' + O_P(\sqrt{\frac{\log N}{T}})$$
$$= H_l^{-1}(S_{lo} - H_o S_o) + O_P(\sqrt{\frac{\log N}{T}}) = O_P(\sqrt{\frac{\log N}{T}}).$$

Finally for $p_i = \mathbb{E}\omega_{it}$, we have

$$\frac{1}{T_i}\sum_{t\in\mathcal{T}_i}\widehat{v}_t\widehat{v}_t' - H^{-1}\frac{1}{T}\sum_t v_t v_t' H'^{-1} = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i}(\widehat{v}_t - H^{-1}v_t)\widehat{v}_t' + \frac{1}{T_i}\sum_{t\in\mathcal{T}_i} H^{-1}v_t(\widehat{v}_t' - v_t' H'^{-1})$$
$$+\frac{T}{T_i p_i}(p_i - \frac{1}{T}\sum_t \omega_{it})\frac{1}{T}\sum_{t\in\mathcal{T}_i} H^{-1}v_t v_t' H'^{-1} + \frac{1}{T}\sum_{t=1}^{N} H^{-1}v_t v_t' H'^{-1}(\frac{\omega_{it}}{p_i} - 1) = O_P(\sqrt{\frac{\log N}{T}}).$$

$\square$

C Data Appendix

C.1 Lipper-TASS

We follow Getmansky et al. (2015) and Sinclair (2018) in cleaning the TASS data. We receive the TASS data in the form of roughly yearly snapshots,[1] which include both dead and alive funds. In order to adjust for backfill bias, we remove returns that were inputted prior to date when the fund was reportedly added to TASS; if that date is missing, we use the date of the first snapshot in which the fund appears.[2] We remove funds that do not report monthly or do not report net returns. For funds which report NAV, we recompute monthly returns to equal percent changes in NAV; otherwise, we keep their reported returns. For firms reporting in an international currency, we adjust their returns and AUMs into USD, whenever possible (this excludes returns before the start of the Euro for European funds, as the pre-Euro currency cannot be determined). Finally, we remove suspicious returns (returns more extreme than -100% or +200% returns in a month), suspicious funds (funds with an 100-fold increase in AUM followed by a 99% decrease in returns; funds where an extreme change in returns does not appear in the AUM: a return is lower than -50% in a month, but AUM does not drop by at least 10% or vice-versa on the positive side), and stale returns (returns equal for three consecutive months). 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 funds with more than 5% of

[1] Specifically: the first snapshot is for 2000. We then have a snapshot for 2002, and at least one snapshot per year from 2005 to 2018, except 2006, 2010, 2014, 2017.

[2] An alternative procedure, proposed by Jorion and Schwarz (2019), could be used in the absence of snapshots.

returns missing. As the literature has noted (e.g., Aggarwal and Jorion (2009)), hedge fund datasets sometimes report duplicate series (for example, multiple share classes or cases in which multiple feeder funds channel capital to one investing master fund, see Joenväärä et al. (2012) and Bali et al. (2014)). To prevent this, we screen for cases in which two funds have return correlations of 99% or more while overlapping for 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 do not exactly overlap).

We impose two further constraints on the funds, again following the existing literature. First, we require funds to have reported returns to the dataset for at least a certain amount of months. This also helps reduce the total number of missing values in our data, which, as the simulation shows, is important to be able to apply the FDR control. Based on the simulations, we choose 36 months and check for robustness increasing this number to 60 months. Second, we follow Kosowski et al. (2007) in only using funds above an AUM threshold; we require funds to have at least $10m of AUM, 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.

C.2 Evestment

The eVestment dataset has been used in a smaller literature. We base our cleaning procedure partly on the one employed for TASS, and partly by looking at the literature that has used eVestment, among which Li et al. (2015); Sebastian and Attaluri (2014); Mozes and Steffens (2016); Cookson et al. (2018); Jenkinson et al. (2016); Chava et al. (2019).

Specifically, we receive the eVestment data in a single large database containing both dead and alive funds, with fund performance broken down into separate investment vehicles. Similarly to the TASS data, we filter to only include returns which were reported monthly and on a net basis. We remove any non-hedge fund products from the sample. For international returns, we recompute returns and AUMs into their USD equivalents. We compute a fund's returns by taking an AUM-weighted average of its vehicle's returns when available; if vehicle AUM is not available, we take an equal-weighted average. We use the same criteria as for TASS to remove duplicate funds and suspicious returns. We remove funds with likely reporting errors in AUMs. In particular, AUMs are sometimes off by many orders of magnitude, sometimes for several months at a time. Given that we compute value-weighted returns, including funds with AUMs that are erroneously high by orders of magnitudes might bias our results. To be conservative, we exclude funds in which the AUM increases more than 10-fold in any a month or more than 100-fold in any 12-month period (and only among those funds that are always above the size cutoff). We also remove hedge funds whose AUM is always reported to be above half a trillion dollars throughout the life of the fund (these are most likely errors). We remove funds where AUM or returns are missing for more than 95% of the time.

As for TASS, we require funds to have reported returns for at least 36 months, and with a size of at least \$10m. Finally, we note that eVestment reports the date in which each entry was added to the database. Given that all the hedge fund data we observe was entered after 2009, we cannot eliminate returns prior to data entry to avoid backfilling bias, as we would not have enough data for the estimation. We do however keep the earliest observation whenever we see an entry being subsequently changed.

References

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

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

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

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.

Benjamini, Y. and Liu, W. (1999). A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. *Journal of Statistical Planning and Inference*, 82(1-2):163–170.

Chava, S., Kim, S., and Weagley, D. (2019). Revealed heuristics: Evidence from investment consultants' search behavior. *Georgia Tech Scheller College of Business Research Paper*, (18-44).

Chen, Y., Chi, Y., Fan, J., Ma, C., and Yan, Y. (2019). Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *arXiv preprint arXiv:1902.07698*.

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

Chernozhukov, V., Hansen, C., Liao, Y., and Zhu, Y. (2018). Inference for heterogeneous effects using low-rank estimations. *arXiv preprint arXiv:1812.08089*.

Cookson, G., Jenkinson, T., Jones, H., and Martinez, J. V. (2018). Investment consultants' claims about their own performance: What lies beneath?

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

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

Gibbons, M., Ross, S., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 57:1121–1152.

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

Jenkinson, T., Jones, H., and Martinez, J. V. (2016). Picking winners? investment consultants' recommendations of fund managers. *The Journal of Finance*, 71(5):2333–2370.

Joenväärä, J., Kosowski, R., and Tolonen, P. (2012). New 'stylized facts' about hedge funds and database selection bias. *Imperial College Business School*.

Jorion, P. and Schwarz, C. (2019). The fix is in: Properly backing out backfill bias. *The Review of Financial Studies*, 32(12):5048–5099.

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

Lee, M.-L. T. and Whitmore, G. A. (2002). Power and sample size for dna microarray studies. *Statistics in medicine*, 21(23):3543–3570.

Li, X., Yao, T., and Zhang, Z. (2015). On the flow-performance relations among delegated institutional portfolios.

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

Ma, S., Goldfarb, D., and Chen, L. (2011). Fixed point and bregman iterative methods for matrix rank minimization. *Mathematical Programming*, 128(1-2):321–353.

Mozes, H. A. and Steffens, J. L. (2016). Hedge fund illiquidity, age, and performance. *The Journal of Wealth Management*, 19(3):87–98.

Peña, V. H., Lai, T. L., and Shao, Q.-M. (2008). *Self-normalized processes: Limit theory and Statistical Applications*. Springer Science & Business Media.

Sebastian, M. and Attaluri, S. (2014). Conviction in equity investing. *The Journal of Portfolio Management*, 40(4):77–88.

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

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

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

Storey, J. D., Taylor, J. E., and Siegmund, D. (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.