Supplement for "Robust Factor Models with Covariates"

Jianqing Fan* Yuan Ke† Yuan Liao‡

September 24, 2016

Abstract

This document contains the remaining technical proofs, and an empirical study of testing proxy factors for S&P 500 returns. We also present a simulation for serially dependent data.

B Proofs for Section 3

Before formally present our proof, let us first provide a guideline of our proof's strategy. First of all, define, for $i = 1, ..., N$,

$$\mathbf{b}_i := \arg\min_{\mathbf{b}\in\mathbb{R}^J} E[y_{it} - \mathbf{b}'\Phi(\mathbf{x}_t)]^2, \quad \mathbf{b}_{i,\alpha} = \arg\min_{\mathbf{b}\in\mathbb{R}^J} E\alpha_T^2\rho\left(\frac{y_{it} - \Phi(\mathbf{x}_t)'\mathbf{b}}{\alpha_T}\right).$$

Recall that

$$\widehat{\mathbf{b}}_i = \arg\min_{\mathbf{b}\in\mathbb{R}^J} \frac{1}{T}\sum_{t=1}^{T} \alpha_T^2\rho\left(\frac{y_{it} - \Phi(\mathbf{x}_t)'\mathbf{b}}{\alpha_T}\right),$$

*Department of Operations Research and Financial Engineering, Bendheim Center for Finance, Princeton University

†Department of Operations Research and Financial Engineering, Princeton University

‡Department of Economics, Rutgers University

As $\alpha_T \to \infty$, $\mathbf{b}_{i,\alpha}$ is expected to converge to $\mathbf{b}_i$ uniformly in $i \leq N$. This is true given some moment conditions on

$$\mathbf{e}_t := \mathbf{y}_t - E(\mathbf{y}_t|\mathbf{x}_t).$$

We shall first prove this in Section B.1.

In addition, let $\widetilde{\mathbf{V}}$ be a $K \times K$ diagonal matrix, whose diagonal elements are the first K eigenvalues of $\widehat{\boldsymbol{\Sigma}}/N := \frac{1}{TN}\sum_{t=1}^{T}\widehat{E}(\mathbf{y}_t|\mathbf{x}_t)\widehat{E}(\mathbf{y}_t|\mathbf{x}_t)'$. By the definition of $\widehat{\boldsymbol{\Lambda}}$, $\frac{1}{N}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Lambda}} = \widehat{\boldsymbol{\Lambda}}\widetilde{\mathbf{V}}$. To effectively use this equality, we need to obtain the Bahadur representations of $\widehat{\mathbf{b}}_i$ and $\widehat{E}(\mathbf{y}_t|\mathbf{x}_t)$. These are achieved in Section B.2.

Furthermore, Section B.3 proves the rates of convergence for the estimated loadings. Sections B.4 and B.5 respectively present the proofs for the rates of convergence for $\mathbf{g}(\mathbf{x}_t)$ and $\boldsymbol{\gamma}_t$.

B.1 The approximation error of the robust estimator

Proposition B.1. *For any $4 < k < \zeta_2 + 2$,*

$$\max_{i \leq N} \|\mathbf{b}_{i,\alpha} - \mathbf{b}_i\| = O(\alpha_T^{-(k-1)}).$$

Proof. We first prove that for any $0 < k < \zeta_2 + 2$, $\max_{i\leq N}\sup_{\mathbf{x}} E(|e_{it}|^k|\mathbf{x}_t = \mathbf{x}) < \infty$. In fact, uniformly in $\mathbf{x}$ for $\mathbf{x}_t = \mathbf{x}$ and $i \leq N$, as long as $\zeta_2 + 2 > k$

$$\begin{aligned} E(|e_{it}|^k|\mathbf{x}_t) &= \int_0^\infty P(|e_{it}|^k > x|\mathbf{x}_t)dx \leq 1 + \int_1^\infty P(|e_{it}|^k > x|\mathbf{x}_t)dx \\ &\leq 1 + \int_1^\infty E(e_{it}^2 1\{|e_{it}| > x^{1/k}\}|\mathbf{x}_t)x^{-2/k}dx \leq 1 + \int_1^\infty Cx^{-(\zeta_2+2)/k}dx < \infty. \end{aligned}$$

Since $\zeta_2 > 2$ by assumption, there is $k > 4$ so that $\max_{i\leq N}\sup_{\mathbf{x}} E(|e_{it}|^k|\mathbf{x}_t = \mathbf{x}) < \infty$.

Now recall that $\mathbf{b}_i = \arg\min E(y_{it} - \mathbf{b}_i'\Phi(\mathbf{x}_t))^2$. Hence

$$E[(y_{it} - \mathbf{b}_{i,\alpha}'\Phi(\mathbf{x}_t))^2 - (y_{it} - \mathbf{b}_i'\Phi(\mathbf{x}_t))^2] = (\mathbf{b}_{i,\alpha}' - \mathbf{b}_i)'E\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'(\mathbf{b}_{i,\alpha} - \mathbf{b}_i) \geq \underline{c}\|\mathbf{b}_{i,\alpha} - \mathbf{b}_i\|^2$$

On the other hand, let $g_\alpha(z) := z^2 - \alpha_T^2 \rho(z/\alpha_T)$. Then for $C > 0$ as a generic constant,

$$
\begin{aligned}
& E[(y_{it} - \mathbf{b}_{i,\alpha}'\Phi(\mathbf{x}_t))^2 - (y_{it} - \mathbf{b}_i'\Phi(\mathbf{x}_t))^2] = Eg_\alpha(y_{it} - \mathbf{b}_{i,\alpha}'\Phi(\mathbf{x}_t)) - Eg_\alpha(y_{it} - \mathbf{b}_i'\Phi(\mathbf{x}_t)) \\
& +E[\alpha_T^2\rho(\alpha_T^{-1}(y_{it} - \mathbf{b}_{i,\alpha}'\Phi(\mathbf{x}_t))) - \alpha_T^2\rho(\alpha_T^{-1}(y_{it} - \mathbf{b}_i'\Phi(\mathbf{x}_t)))] \\
\leq_{(1)} \;& Eg_\alpha(y_{it} - \mathbf{b}_{i,\alpha}'\Phi(\mathbf{x}_t)) - Eg_\alpha(y_{it} - \mathbf{b}_i'\Phi(\mathbf{x}_t)) \leq_{(2)} E[|2\tilde{z} - \alpha_T\dot{\rho}(\alpha_T^{-1}\tilde{z})||\Phi(\mathbf{x}_t)'(\mathbf{b}_i - \mathbf{b}_{i,\alpha})|], \\
\leq_{(3)} \;& 2\alpha_T^{-(k-1)}E|\tilde{z}|^k|\Phi(\mathbf{x}_t)'(\mathbf{b}_i - \mathbf{b}_{i,\alpha})| \leq_{(4)} 2\alpha_T^{-(k-1)}E|z_{it} + e_{it} + (\mathbf{b}_i - \widetilde{\mathbf{b}}_i)'\Phi(\mathbf{x}_t)|^k|\Phi(\mathbf{x}_t)'(\mathbf{b}_i - \mathbf{b}_{i,\alpha})| \\
\leq \;& C\alpha_T^{-(k-1)}E(C + |(\mathbf{b}_i - \widetilde{\mathbf{b}}_i)'\Phi(\mathbf{x}_t)|^k)|\Phi(\mathbf{x}_t)'(\mathbf{b}_i - \mathbf{b}_{i,\alpha})|
\end{aligned}
$$

where (1) is due to the definition of $\mathbf{b}_{i,\alpha}$; (2) is by the mean value representation: $g_\alpha(z_1) - g_\alpha(z_2) = (2\tilde{z} - \alpha_T\dot{\rho}(\tilde{z}/\alpha_T))(z_1 - z_2)$, with $z_1 = y_{it} - \mathbf{b}_{i,\alpha}'\Phi(\mathbf{x}_t)$, $z_2 = y_{it} - \mathbf{b}_i'\Phi(\mathbf{x}_t)$, and $\tilde{z} = y_{it} - \widetilde{\mathbf{b}}_i'\Phi(\mathbf{x}_t)$ for some $\widetilde{\mathbf{b}}_i$ lying between $\mathbf{b}_i$ and $\mathbf{b}_{i,\alpha}$; (3) is due to

$$
|2\tilde{z} - \alpha_T\dot{\rho}(\alpha_T^{-1}\tilde{z})| \leq 2|\tilde{z}|1\{|\tilde{z}| > \alpha_T\} \leq 2|\tilde{z}|\frac{|\tilde{z}|^{k-1}}{\alpha_T^{k-1}}1\{|\tilde{z}| > \alpha_T\} \leq 2|\tilde{z}|^k/\alpha_T^{k-1}.
$$

(4) follows from $\tilde{z} = y_{it} - E(y_{it}|\mathbf{x}_t) + \mathbf{b}_i'\Phi(\mathbf{x}_t) + z_{it} - \widetilde{\mathbf{b}}_i'\Phi(\mathbf{x}_t)$, and that $e_{it} := y_{it} - E(y_{it}|\mathbf{x}_t)$.

Next, for ease of presentation, we introduce $M_{it} := C + |(\mathbf{b}_i - \widetilde{\mathbf{b}}_i)'\Phi(\mathbf{x}_t)|^k$ and $\mathbf{\Delta}_i := \mathbf{b}_i - \mathbf{b}_{i,\alpha}$. Then the above inequality can be further written as:

$$
\begin{aligned}
= \;& C\alpha_T^{-(k-1)}EM_{it}|\Phi(\mathbf{x}_t)'\mathbf{\Delta}_i| = C\alpha_T^{-(k-1)}E[M_{it}^2\mathbf{\Delta}_i'\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'\mathbf{\Delta}_i]^{1/2} \\
\leq_{(5)} \;& C\alpha_T^{-(k-1)}[\mathbf{\Delta}_i'EM_{it}^2\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'\mathbf{\Delta}_i]^{1/2} \leq C\alpha_T^{-(k-1)}\|EM_{it}^2\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'\|^{1/2}\|\mathbf{\Delta}_i\|.
\end{aligned}
$$

We now bound $\max_{i\leq N}\|EM_{it}^2\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'\| = \max_{i\leq N}\sup_{\|\boldsymbol{\nu}\|=1}EM_{it}^2(\Phi(\mathbf{x}_t)'\boldsymbol{\nu})^2 = O(1)$. By the Cauchy-Schwarz inequality, since $\Phi(\mathbf{x}_t)'\boldsymbol{\nu}$ is sub-Gaussian with the universal parameter,

$$
\begin{aligned}
& \sup_{\|\boldsymbol{\nu}\|=1}[EM_{it}^2(\Phi(\mathbf{x}_t)'\boldsymbol{\nu})^2]^2 \leq EM_{it}^4\sup_{\|\boldsymbol{\nu}\|=1}E(\Phi(\mathbf{x}_t)'\boldsymbol{\nu})^4 \leq CEM_{it}^4 \leq C(C + E|(\mathbf{b}_i - \widetilde{\mathbf{b}}_i)'\Phi(\mathbf{x}_t)|^{4k}) \\
\leq \;& C + CE\|\mathbf{b}_i - \widetilde{\mathbf{b}}_i\|^{4k}\left(\frac{(\mathbf{b}_i' - \widetilde{\mathbf{b}}_i)'}{\|\mathbf{b}_i - \widetilde{\mathbf{b}}_i\|}\Phi(\mathbf{x}_t)\right)^{4k} \leq C + C\|\mathbf{\Delta}_i\|^{4k}\sup_{\|\boldsymbol{\nu}\|=1}E(\boldsymbol{\nu}'\Phi(\mathbf{x}_t))^{4k} \leq C + C\|\mathbf{\Delta}_i\|^{4k}.
\end{aligned}
$$

Therefore, we have proved that uniformly in i,

$$E[(y_{it}-\mathbf{b}_{i,\alpha}'\Phi(\mathbf{x}_t))^2-(y_{it}-\mathbf{b}_i'\Phi(\mathbf{x}_t))^2] \leq C\alpha_T^{-(k-1)}(C+C\|\mathbf{\Delta}_i\|^{4k})^{1/4}\|\mathbf{\Delta}_i\| \leq C\alpha_T^{-(k-1)}(1+\|\mathbf{\Delta}_i\|^k)\|\mathbf{\Delta}_i\|$$

We have also proved that the left hand side is lower bounded by $\underline{c}\|\mathbf{\Delta}_i\|^2$. Uniformly in i,

$$\|\mathbf{\Delta}_i\| \leq C\alpha_T^{-(k-1)}(1+\|\mathbf{\Delta}_i\|^k).$$

If $\max_i \|\mathbf{\Delta}_i\| = O(1)$, then $\|\mathbf{\Delta}_i\| \leq C\alpha_T^{-(k-1)}$. Otherwise, $\max_i \|\mathbf{\Delta}_i\| \leq C\alpha_T^{-(k-1)}\max_i \|\mathbf{\Delta}_i\|^k$, which then implies $1 \leq C(\max_i \|\mathbf{\Delta}_i\|/\alpha_T)^{k-1}$. However, note that $\|\mathbf{\Delta}_i\| \leq \|\mathbf{b}_i\| + \|\mathbf{b}_{i,\alpha}\| \leq CJ^{1/2}$, and $J = o(\alpha_T^2)$, we have $\max_i \|\mathbf{\Delta}_i\|/\alpha_T = o(1)$, which is a contradiction. Therefore, $\max_i \|\mathbf{\Delta}_i\| \leq C\alpha_T^{-(k-1)}$. Q.E.D.

The following lemma shows the sieve approximation error is uniformly controlled.

Lemma B.1. *Under Assumption 3.2, there is $\eta \geq 1$, as $J \to \infty$,*

$$\max_{i\leq N}\sup_{\mathbf{x}} |E(y_{it}|\mathbf{x}_t = \mathbf{x}) - \mathbf{b}_i'\Phi(\mathbf{x})| = O(J^{-\eta}).$$

Proof. Recall that for $k \leq K$, $\mathbf{v}_k = \arg\min_{\mathbf{v}} E(f_{kt}-\mathbf{v}'\Phi(\mathbf{x}_t))^2 = (E\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)')^{-1}E\Phi(\mathbf{x}_t)f_{kt}$, and that $\mathbf{b}_i = \arg\min_{\mathbf{b}\in\mathbb{R}^J} E[y_{it} - \mathbf{b}'\Phi(\mathbf{x}_t)]^2 = (E\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)')^{-1}E\Phi(\mathbf{x}_t)y_{it}$. Also note that $y_{it} = \boldsymbol{\lambda}_i'\mathbf{f}_t + u_{it}$. We have $\mathbf{b}_i = \sum_{k=1}^K \mathbf{v}_k\lambda_{ik}$. Hence

$$\begin{aligned}
\max_{i\leq N}\sup_{\mathbf{x}} |E(y_{it}|\mathbf{x}_t = \mathbf{x}) - \mathbf{b}_i'\Phi(\mathbf{x})| &\leq \max_{i\leq N}\sup_{\mathbf{x}} |\sum_{k=1}^K \lambda_{ik}(E(f_{tk}|\mathbf{x}_t = \mathbf{x}) - \mathbf{v}_k'\Phi(\mathbf{x}))| \\
&\leq O(1)\max_k\sup_{\mathbf{x}} |E(f_{tk}|\mathbf{x}_t = \mathbf{x}) - \mathbf{v}_k'\Phi(\mathbf{x})| \\
&= O(J^{-\eta}).
\end{aligned}$$

B.2 Bahadur representation of the robust estimator

We now give the uniform convergence rate of $\widehat{\mathbf{b}}_i$ as well as its Bahadur representation.

Define

$$Q_i(\mathbf{b}) = \frac{1}{T}\sum_{t=1}^{T}\alpha_T^2\rho\left(\frac{y_{it} - \Phi(\mathbf{x}_t)'\mathbf{b}}{\alpha_T}\right).$$

Proposition B.2. *When $\alpha_T \leq C\sqrt{T/\log(NJ)}$ for any $C > 0$, and any $4 < k < \zeta_2 + 2$,*

$$\max_{i\leq N}\|\widehat{\mathbf{b}}_i - \mathbf{b}_i\| = O_P(\sqrt{\frac{J\log N}{T}} + \alpha_T^{-(k-1)}).$$

Proof. Let $m_T = \sqrt{\frac{J\log N}{T}}$. We aim to show, for any $\epsilon > 0$, there is $\delta > 0$, when for all large N, T,

$$P(\min_{i\leq N}\inf_{\|\boldsymbol{\nu}\|=\delta} Q_i(\mathbf{b}_{i,\alpha} + m_T\boldsymbol{\nu}) - Q_i(\mathbf{b}_{i,\alpha}) > 0) > 1 - \epsilon.$$

This then implies $\max_i \|\widehat{\mathbf{b}}_i - \mathbf{b}_{i,\alpha}\| = O_P(m_T)$. The result then follows from Proposition B.1.

By the definition of $\mathbf{b}_{i,\alpha}$,

$$E[\Phi(\mathbf{x}_t)\dot\rho(\alpha_T^{-1}e_{it,\alpha})] = 0, \qquad e_{it,\alpha} := y_{it} - \Phi(\mathbf{x}_t)'\mathbf{b}_{i,\alpha}.$$

In addition, we have $e_{it} = e_{it,\alpha} + \Delta_{it,\alpha}$, where $\Delta_{it,\alpha} := (\mathbf{b}_{i,\alpha} - \mathbf{b}_i)'\Phi(\mathbf{x}_t) - z_{it}$. Using the formula: $\rho(a+t)-\rho(a) = \dot\rho(a)t+\int_0^t(\dot\rho(a+x)-\dot\rho(a))dx$ for $a = \alpha_T^{-1}e_{it,\alpha}$ and $t = -m_T\alpha_T^{-1}\Phi(\mathbf{x}_t)'\boldsymbol{\nu}$,

$$\begin{aligned}
&Q_i(\mathbf{b}_{i,\alpha} + m_T\boldsymbol{\nu}) - Q_i(\mathbf{b}_{i,\alpha}) = -\frac{1}{T}\sum_{t=1}^{T} m_T\alpha_T\dot\rho(\alpha_T^{-1}e_{it,\alpha})\Phi(\mathbf{x}_t)'\boldsymbol{\nu} \\
&+\frac{1}{T}\sum_{t=1}^{T}1\{\Phi(\mathbf{x}_t)'\boldsymbol{\nu} < 0\}\alpha_T^2\int_0^{-m_T\alpha_T^{-1}\Phi(\mathbf{x}_t)'\boldsymbol{\nu}}\dot\rho(\alpha_T^{-1}e_{it,\alpha} + x) - \dot\rho(\alpha_T^{-1}e_{it,\alpha})dx \\
&-\frac{1}{T}\sum_{t=1}^{T}1\{\Phi(\mathbf{x}_t)'\boldsymbol{\nu} > 0\}\alpha_T^2\int_{-m_T\alpha_T^{-1}\Phi(\mathbf{x}_t)'\boldsymbol{\nu}}^{0}\dot\rho(\alpha_T^{-1}e_{it,\alpha} + x) - \dot\rho(\alpha_T^{-1}e_{it,\alpha})dx.
\end{aligned}$$

By the definition of $\dot{\rho}$, the integrant can be rewritten as:

$$
\begin{aligned}
& \dot{\rho}(\alpha_T^{-1}e_{it,\alpha}+x)-\dot{\rho}(\alpha_T^{-1}e_{it,\alpha}) = 2x1\{|\alpha_T^{-1}e_{it,\alpha}+x|<1, |\alpha_T^{-1}e_{it,\alpha}|<1\} \\
& +(\dot{\rho}(\alpha_T^{-1}e_{it,\alpha}+x)-\dot{\rho}(\alpha_T^{-1}e_{it,\alpha}))1\{|\alpha_T^{-1}e_{it,\alpha}+x|\geq 1, \text{ or } |\alpha_T^{-1}e_{it,\alpha}|\geq 1\} \\
= \ & 2x-(\dot{\rho}(\alpha_T^{-1}e_{it,\alpha}+x)-\dot{\rho}(\alpha_T^{-1}e_{it,\alpha})-2x)1\{|\alpha_T^{-1}e_{it,\alpha}+x|\geq 1, \text{ or } |\alpha_T^{-1}e_{it,\alpha}|\geq 1\}.
\end{aligned}
$$

In addition, note that

$$
|\dot{\rho}(x_1)-\dot{\rho}(x_1)|\leq 2|x_1-x_2|, \quad \forall x_1, x_2.
$$

Thus we can further write:

$$
\begin{aligned}
& Q_i(\mathbf{b}_{i,\alpha}+m_T\boldsymbol{\nu})-Q_i(\mathbf{b}_{i,\alpha}) = -\frac{1}{T}\sum_{t=1}^T m_T\alpha_T\dot{\rho}(\alpha_T^{-1}e_{it,\alpha})\Phi(\mathbf{x}_t)'\boldsymbol{\nu}+\frac{1}{T}\sum_{t=1}^T\alpha_T^2\int_0^{-m_T\alpha_T^{-1}\Phi(\mathbf{x}_t)'\boldsymbol{\nu}} 2xdx \\
& -\frac{1}{T}\sum_{t=1}^T 1\{\Phi(\mathbf{x}_t)'\boldsymbol{\nu}<0\}\alpha_T^2\int_0^{-m_T\alpha_T^{-1}\Phi(\mathbf{x}_t)'\boldsymbol{\nu}}(\dot{\rho}(\alpha_T^{-1}e_{it,\alpha}+x)-\dot{\rho}(\alpha_T^{-1}e_{it,\alpha})-2x) \\
& \times 1\{|\alpha_T^{-1}e_{it,\alpha}+x|\geq 1, \text{ or } |\alpha_T^{-1}e_{it,\alpha}|\geq 1\}dx \\
& +\frac{1}{T}\sum_{t=1}^T 1\{\Phi(\mathbf{x}_t)'\boldsymbol{\nu}>0\}\alpha_T^2\int_{-m_T\alpha_T^{-1}\Phi(\mathbf{x}_t)'\boldsymbol{\nu}}^0(\dot{\rho}(\alpha_T^{-1}e_{it,\alpha}+x)-\dot{\rho}(\alpha_T^{-1}e_{it,\alpha})-2x) \\
& \times 1\{|\alpha_T^{-1}e_{it,\alpha}+x|\geq 1, \text{ or } |\alpha_T^{-1}e_{it,\alpha}|\geq 1\}dx \\
\geq \ & \inf_{\|\boldsymbol{\nu}\|=\delta}\frac{1}{T}\sum_{t=1}^T\alpha_T^2(-m_T\alpha_T^{-1}\Phi(\mathbf{x}_t)'\boldsymbol{\nu})^2-\max_i\sup_{\|\boldsymbol{\nu}\|=\delta}|\frac{1}{T}\sum_{t=1}^T m_T\alpha_T\dot{\rho}(\alpha_T^{-1}e_{it,\alpha})\Phi(\mathbf{x}_t)'\boldsymbol{\nu}| \\
& -\max_i\sup_{\|\boldsymbol{\nu}\|=1}\frac{1}{T}\sum_{t=1}^T\alpha_T^2\int_0^{m_T\alpha_T^{-1}|\Phi(\mathbf{x}_t)'\boldsymbol{\nu}|} 4x1\{|\alpha_T^{-1}e_{it,\alpha}+x|\geq 1, \text{ or } |\alpha_T^{-1}e_{it,\alpha}|\geq 1\}dx \\
:= \ & A_1-A_2-A_3.
\end{aligned}
$$

We now lower bound A_1 and upper bound A_2, A_3.

First of all, there is $c>0$ independent of δ, with probability approaching one,

$$
A_1=\inf_{\|\boldsymbol{\nu}\|=\delta}\boldsymbol{\nu}'\frac{1}{T}\sum_{t=1}^T m_T^2\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'\boldsymbol{\nu}\geq\lambda_{\min}(\frac{1}{T}\sum_{t=1}^T\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)')m_T^2\delta^2\geq cm_T^2\delta^2.
$$

As for A_2, note that $|\alpha_T\dot\rho(\alpha_T^{-1}e_{it,\alpha})|\leq |e_{it,\alpha}|\leq |e_{it}|+|\Delta_{it,\alpha}|$. Uniformly in $i\leq N, j\leq J$, by Holder's inequality, with an arbitrarily small $v>0$, and $p=(1+v)^{-1}$,

$$\begin{aligned}
&E(\dot\rho(\alpha_T^{-1}e_{it,\alpha})\phi_j(\mathbf{x}_t))^2\leq \alpha_T^{-2}E(\alpha_T\dot\rho(\alpha_T^{-1}e_{it,\alpha})\phi_j(\mathbf{x}_t))^2\leq 2\alpha_T^{-2}E(e_{it}^2+\Delta_{it,\alpha}^2)\phi_j(\mathbf{x}_t)^2\\
\leq\ & 2\alpha_T^{-2}EE\{e_{it}^2|\mathbf{x}_t\}\phi_j(\mathbf{x}_t)^2+2\alpha_T^{-2}E\Delta_{it,\alpha}^2\phi_j(\mathbf{x}_t)^2\leq C\alpha_T^{-2}((E\{e_{it}^2|\mathbf{x}_t\}^{1+v})^{1/p}+C)\leq C\alpha_T^{-2}.
\end{aligned}$$

Note that $|\dot\rho|<2$ and $\{\phi_j(\mathbf{x}_t)\}$ is sub-Gaussian, thus by the Bernstein inequality, for $x=2\log(NJ)$,

$$P(|\frac{1}{T}\sum_{t=1}^T\dot\rho(\alpha_T^{-1}e_{it,\alpha})\phi_j(\mathbf{x}_t)|>\sqrt{\frac{2E(\dot\rho(\alpha_T^{-1}e_{it,\alpha})\phi_j(\mathbf{x}_t))^2x}{T}}+\frac{Cx}{T})\leq 2\exp(-x).$$

Note that when $\alpha_T\leq C\sqrt{T/\log(NJ)}$,

$$\sqrt{\frac{2E(\dot\rho(\alpha_T^{-1}e_{it,\alpha})\phi_j(\mathbf{x}_t))^2x}{T}}+\frac{Cx}{T}\leq\sqrt{\frac{C\log(NJ)}{\alpha_T^2T}}+\frac{C\log(NJ)}{T}\leq 2\sqrt{\frac{C\log(NJ)}{\alpha_T^2T}}.$$

Thus

$$P(\max_{ij}|\frac{1}{T}\sum_{t=1}^T\dot\rho(\alpha_T^{-1}e_{it,\alpha})\phi_j(\mathbf{x}_t)|>\sqrt{\frac{C\log(NJ)}{\alpha_T^2T}})\leq CNJ\exp(-2\log(NJ))=\frac{C}{NJ}.$$

Therefore, with probability approaching one,

$$\begin{aligned}
A_2\ &\leq\ m_T\alpha_T\delta\max_i\|\frac{1}{T}\sum_{t=1}^T\dot\rho(\alpha_T^{-1}e_{it,\alpha})\Phi(\mathbf{x}_t)'\|\leq m_T\alpha_T\sqrt{J}\delta\max_{i\leq N,j\leq J}|\frac{1}{T}\sum_{t=1}^T\dot\rho(\alpha_T^{-1}e_{it,\alpha})\phi_j(\mathbf{x}_t)|\\
&\leq\ \delta m_T\alpha_T\sqrt{\frac{CJ\log(N)}{\alpha_T^2T}}=\delta m_T\sqrt{\frac{CJ\log(N)}{T}}.
\end{aligned}$$

As for A_3, note that uniformly for $x\leq m_T\alpha_T^{-1}|\Phi(\mathbf{x}_t)'\boldsymbol{\nu}|$, and $e_{it}=e_{it,\alpha}+\Delta_{it,\alpha}$,

$$1\{|\alpha_T^{-1}e_{it,\alpha}+x|\geq 1,\text{ or }|\alpha_T^{-1}e_{it,\alpha}|\geq 1\}\leq 1\{|\alpha_T^{-1}e_{it,\alpha}+x|\geq 1\}+1\{|\alpha_T^{-1}e_{it,\alpha}|\geq 1\}$$

$$
\begin{aligned}
&\leq 2\times 1\{|e_{it,\alpha}|>3\alpha_T/4\}+1\{m_T|\Phi(\mathbf{x}_t)'\boldsymbol{\nu}|>\alpha_T/4\} \\
&\leq 2\times 1\{|e_{it}|>\alpha_T/2\}+1\{m_T|\Phi(\mathbf{x}_t)'\boldsymbol{\nu}|>\alpha_T/4\}+1\{|\Delta_{it,\alpha}|>\alpha_T/4\}.
\end{aligned}
$$

In addition, with probability at least $1-\epsilon/10$,

$$
\begin{aligned}
\max_i \frac{1}{T}\sum_{t=1}^T 1\{|e_{it}|>\alpha_T/2\} &\leq_{(1)} \max_i P(|e_{it}|>\alpha_T/2)+\sqrt{\frac{\log N}{T}}\max_i P(|e_{it}|>\alpha_T/2)^{1/2}, \\
\frac{1}{T}\sum_t 1\{m_T\|\Phi(\mathbf{x}_t)\|\delta>\alpha_T/4\} &\leq 10P(m_T\delta\|\Phi(\mathbf{x}_t)\|>\alpha_T/4)/\epsilon, \\
\max_i \frac{1}{T}\sum_{t=1}^T 1\{|\Delta_{it,\alpha}|>\alpha_T/4\} &\leq \max_i \frac{1}{T}\sum_{t=1}^T 1\{\|\Phi(\mathbf{x}_t)\|>C\alpha_T^k\}+1\{|z_{it}|>\alpha_T/4\} \\
&\leq 10P(\|\Phi(\mathbf{x}_t)\|>C\alpha_T^k)/\epsilon+CJ^{-2\eta}/\alpha_T^2+\sqrt{\frac{\log N}{T}}CJ^{-\eta}/\alpha_T,
\end{aligned}
$$

where (1) follows from the triangular inequality,

$$
\max_i \frac{1}{T}\sum_{t=1}^T 1\{|e_{it}|>\alpha_T/2\}\leq \max_i P(|e_{it}|>\alpha_T/2)+\max_i\Big|\frac{1}{T}\sum_{t=1}^T 1\{|e_{it}|>\alpha_T/2\}-P(|e_{it}|>\alpha_T/2)\Big|,
$$

and we used Bernstein inequality+union bound to bound the second term since the indicator function is bounded. Hence for an arbitrarily small $v>0$, by Holder's inequality, for some generic constant $C>0$, independent of δ,

$$
\begin{aligned}
A_3 &\leq \max_i \sup_{\|\boldsymbol{\nu}\|=\delta}\frac{1}{T}\sum_{t=1}^T 4(m_T|\Phi(\mathbf{x}_t)'\boldsymbol{\nu}|)^2[1\{|e_{it}|>\alpha_T/2\}+1\{m_T|\Phi(\mathbf{x}_t)'\boldsymbol{\nu}|>\alpha_T/4\}+1\{|\Delta_{it,\alpha}|>\alpha_T/4\}] \\
&\leq C\max_i(\frac{1}{T}\sum_{t=1}^T[1\{|e_{it}|>\alpha_T/2\}+1\{m_T\delta\|\Phi(\mathbf{x}_t)\|>\alpha_T/4\}+1\{|\Delta_{it,\alpha}|>\alpha_T/4\}])^{1-v} \\
&\quad\times(\frac{1}{T}\sum_{t=1}^T\|\Phi(\mathbf{x}_t)\|^{2/v})^v(m_T\delta)^2 \\
&\leq (m_T\delta)^2C\Bigg(\max_i P(|e_{it}|>\alpha_T/2)+\sqrt{\frac{\log N}{T}}\max_i P(|e_{it}|>\alpha_T/2)^{1/2}+10P(m_T\delta\|\Phi(\mathbf{x}_t)\|>\alpha_T/4)/\epsilon
\end{aligned}
$$

$$+10P(\|\Phi(\mathbf{x}_t)\| > C\alpha_T^k)/\epsilon + CJ^{-2\eta}/\alpha_T^2 + \sqrt{\frac{\log N}{T}}CJ^{-\eta}/\alpha_T\Big)^{1-v}(C + E\|\Phi(\mathbf{x}_t)\|^{2/v})^v.$$

We now upper bound $E\|\Phi(\mathbf{x}_t)\|^{2/v}$ and $P(\|\Phi(\mathbf{x}_t)\| > x)$ for any x. Since $\{\phi_j(w_t)\}_{j\leq J}$ is sub-Gaussian, by Lemma 14.12 of Bühlmann and van de Geer (2011),

$$\begin{aligned}
E\|\Phi(\mathbf{x}_t)\|^{2/v} &\leq J^{1/v}E(\max_{j\leq J}\phi_j(\mathbf{x}_t)^{2/v}) \leq J^{1/v}E(\max_{j\leq J}|\phi_j(\mathbf{x}_t)^{2/v} - E\phi_j(\mathbf{x}_t)^{2/v}|) \\
&\quad + J^{1/v}\max_j E\phi_j(\mathbf{x}_t)^{2/v} \leq J^{1/v}C\log(J). \\
P(\|\Phi(\mathbf{x}_t)\| > x) &\leq P(\max_j |\phi_j(\mathbf{x}_t)|^2 J > x^2) \leq J\max_j P(|\phi_j(\mathbf{x}_t)| > x/J^{1/2}) \leq J\exp(-Cx^2/J).
\end{aligned}$$

Therefore,

$$\begin{aligned}
A_3 &\leq (m_T\delta)^2C\Big(\max_i P(|e_{it}| > \alpha_T/2) + \sqrt{\frac{\log N}{T}}\max_i P(|e_{it}| > \alpha_T/2)^{1/2} + CJ\exp(-C\alpha_T^2/(Jm_T^2\delta^2))/\epsilon \\
&\quad + CJ\exp(-C\alpha_T^{2k}/J)/\epsilon + CJ^{-2\eta}/\alpha_T^2 + \sqrt{\frac{\log N}{T}}CJ^{-\eta}/\alpha_T\Big)^{1-v}J(\log J)^v := (m_T\delta)^2Cl_T.
\end{aligned}$$

Note that $l_T = o(1)$.

Consequently, for any $\epsilon > 0$, there are C, c, and c_ϵ independent of δ (may depend on ϵ), with probability at least $1-\epsilon$, uniformly in $i \leq N$ and $\|\boldsymbol{\nu}\| = \delta$, for $m_T = \sqrt{\frac{J\log N}{T}}$,

$$Q_i(\mathbf{b}_{i,\alpha} + m_T\boldsymbol{\nu}) - Q_i(\mathbf{b}_{i,\alpha}) \geq m_T^2\delta^2(c - c_\epsilon l_T) - \delta m_T C\sqrt{\frac{J\log N}{T}} \geq m_T\delta(m_T\delta c/2 - Cm_T) > 0$$

so long as $\delta c > 2C$. Thus $\max_i \|\widehat{\mathbf{b}}_i - \mathbf{b}_{i,\alpha}\| = O_P(m_T)$.

We now prove a simple lemma.

Lemma B.2. *There is $M > 0$ for all $x > M$,*

$$\begin{aligned}
\max_{i\leq N}\sup_{\mathbf{x}} P(|e_{it}| > x|\mathbf{x}_t = \mathbf{x}) &\leq Cx^{-\zeta_2-2} \\
\max_{i\leq N}\sup_{\mathbf{x}} E(|e_{it}|1\{|e_{it}| > x\}|\mathbf{x}_t = \mathbf{x}) &\leq Cx^{-\zeta_2-1}.
\end{aligned}$$

Proof. Uniformly in $\mathbf{x}=\mathbf{x}_t$ and $i\leq N$,

$$
\begin{aligned}
P(|e_{it}|>x|\mathbf{x}_t) &= E(1\{|e_{it}|>x\}|\mathbf{x}_t)\leq E(e_{it}^2 1\{|e_{it}|>x\}|\mathbf{x}_t)x^{-2}\leq Cx^{-\zeta_2-2}\\
E(|e_{it}|1\{|e_{it}|>x\}|\mathbf{x}_t) &\leq E(e_{it}^2 1\{|e_{it}|>x\}|\mathbf{x}_t)x^{-1}\leq Cx^{-\zeta_2-1}.
\end{aligned}
$$

Lemma B.3. *Uniformly for $i=1,...,N$,*

$$
\widehat{\mathbf{b}}_i-\mathbf{b}_{i,\alpha}=(2E\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)')^{-1}\frac{1}{T}\sum_{t=1}^T\alpha_T\dot{\rho}(\alpha_T^{-1}e_{it,\alpha})\Phi(\mathbf{x}_t)+\mathbf{R}_{i,b},
$$

where $\max_{i\leq N}\|\mathbf{R}_{i,b}\|=O_P(\alpha_T^{-(\zeta_1-1)}+\sqrt{\frac{\log J}{T}})J\sqrt{\frac{J\log N}{T}}$.

Proof. Note that $\nabla Q_i(\mathbf{b})=-\frac{1}{T}\sum_{t=1}^T\alpha_T\dot{\rho}(\alpha_T^{-1}(y_{it}-\Phi(\mathbf{x}_t)'\mathbf{b}))\Phi(\mathbf{x}_t)$. Define $\bar{Q}_i(\mathbf{b})=EQ_i(\mathbf{b})$,

$$
\boldsymbol{\mu}_i(\mathbf{b}):=\nabla Q_i(\mathbf{b})-\nabla\bar{Q}_i(\mathbf{b})=E\alpha_T\dot{\rho}(\alpha_T^{-1}(y_{it}-\Phi(\mathbf{x}_t)'\mathbf{b}))\Phi(\mathbf{x}_t)-\frac{1}{T}\sum_{t=1}^T\alpha_T\dot{\rho}(\alpha_T^{-1}(y_{it}-\Phi(\mathbf{x}_t)'\mathbf{b}))\Phi(\mathbf{x}_t).
$$

The first order condition gives $\nabla Q_i(\widehat{\mathbf{b}}_i)=0$. By the mean value expansion,

$$
\begin{aligned}
0 &= \nabla Q_i(\widehat{\mathbf{b}}_i)-\nabla\bar{Q}_i(\widehat{\mathbf{b}}_i)+\nabla\bar{Q}_i(\widehat{\mathbf{b}}_i)-\nabla\bar{Q}_i(\mathbf{b}_{i,\alpha})+\nabla\bar{Q}_i(\mathbf{b}_{i,\alpha})-\nabla Q_i(\mathbf{b}_{i,\alpha})+\nabla Q_i(\mathbf{b}_{i,\alpha})\\
&= \boldsymbol{\mu}_i(\widehat{\mathbf{b}}_i)+\nabla\bar{Q}_i(\widehat{\mathbf{b}}_i)-\nabla\bar{Q}_i(\mathbf{b}_{i,\alpha})-\boldsymbol{\mu}_i(\mathbf{b}_{i,\alpha})+\nabla Q_i(\mathbf{b}_{i,\alpha})\\
&= \nabla^2\bar{Q}_i(\widetilde{\mathbf{b}}_i)(\widehat{\mathbf{b}}_i-\mathbf{b}_{i,\alpha})+\nabla Q_i(\mathbf{b}_{i,\alpha})+\boldsymbol{\mu}_i(\widehat{\mathbf{b}}_i)-\boldsymbol{\mu}_i(\mathbf{b}_{i,\alpha}).
\end{aligned}
$$

for some $\widetilde{\mathbf{b}}_i$ in the segment joining $\widehat{\mathbf{b}}_i$ and $\mathbf{b}_{i,\alpha}$. We now proceed by: (i) upper bounding $\max_i\|\boldsymbol{\mu}_i(\widehat{\mathbf{b}}_i)-\boldsymbol{\mu}_i(\mathbf{b}_{i,\alpha})\|$, and (ii) finding the limit of $\nabla^2\bar{Q}_i(\widetilde{\mathbf{b}}_i)$ uniformly in i.

(i) Note that in the proof of Proposition B.2, we have proved that for any $\epsilon>0$, there is $\delta>0$, so that the following event holds with probability at least $1-\epsilon$:

$$
\max_i\|\widehat{\mathbf{b}}_i-\mathbf{b}_{i,\alpha}\|\leq\delta m_T,\quad m_T=\sqrt{\frac{J\log N}{T}}.
$$

We bound $E\max_i \sup_{\|\mathbf{b}-\mathbf{b}_{i,\alpha}\|\leq\delta m_T} \|\boldsymbol{\mu}_i(\mathbf{b})-\boldsymbol{\mu}_i(\mathbf{b}_{i,\alpha})\|$. Let $\mu_{ij}(\cdot)$ be the jth element of $\boldsymbol{\mu}_i$, $j\leq J$. Since $\{\mathbf{y}_t,\mathbf{x}_t\}_{t\leq T}$ are serially independent, there exists a Radamacher sequence $\{\varepsilon_t\}_{t\leq T}$ with $P(\varepsilon_t=1)=P(\varepsilon_t=-1)=1/2$, that is independent of $\{\mathbf{y}_t,\mathbf{x}_t\}$,

$$
\begin{aligned}
& E\max_{i\leq N, j\leq J}\sup_{\|\mathbf{b}-\mathbf{b}_{i,\alpha}\|\leq\delta m_T}|\mu_{ij}(\mathbf{b})-\mu_{ij}(\mathbf{b}_{i,\alpha})| \\
\leq_{(a)} \quad & 2E\max_{i\leq N, j\leq J}\sup_{\|\mathbf{b}-\mathbf{b}_{i,\alpha}\|\leq\delta m_T}|\frac{1}{T}\sum_{t=1}^T\varepsilon_t\alpha_T\left(\dot\rho(\alpha_T^{-1}(y_{it}-\Phi(\mathbf{x}_t)'\mathbf{b}))-\dot\rho(\alpha_T^{-1}(y_{it}-\Phi(\mathbf{x}_t)'\mathbf{b}_{i,\alpha}))\right)\phi_j(\mathbf{x}_t)| \\
\leq_{(b)} \quad & 4E\max_{i\leq N, j\leq J}\sup_{\|\mathbf{b}-\mathbf{b}_{i,\alpha}\|\leq\delta m_T}|\frac{1}{T}\sum_{t=1}^T\varepsilon_t\Phi(\mathbf{x}_t)'(\mathbf{b}_{i,\alpha}-\mathbf{b})\phi_j(\mathbf{x}_t)| \\
\leq \quad & 4\delta m_T E\max_{j\leq J}\|\frac{1}{T}\sum_{t=1}^T\varepsilon_t\phi_j(\mathbf{x}_t)\Phi(\mathbf{x}_t)'\| \leq 4\delta m_T\sqrt{J}E\max_{l,j\leq J}|\frac{1}{T}\sum_{t=1}^T\varepsilon_t\phi_j(\mathbf{x}_t)\phi_l(\mathbf{x}_t)| \\
\leq_{(c)} \quad & 4\delta m_T\sqrt{J}\frac{L}{T}\log E\exp\left(L^{-1}\max_{l,j\leq J}|\sum_{t=1}^T\varepsilon_t\phi_j(\mathbf{x}_t)\phi_l(\mathbf{x}_t)|\right) \\
\leq_{(d)} \quad & 4\delta m_T\sqrt{J}\frac{L}{T}\log\sum_{l,j\leq J}E\exp\left(L^{-1}|\sum_{t=1}^T\varepsilon_t\phi_j(\mathbf{x}_t)\phi_l(\mathbf{x}_t)|\right) \\
\leq_{(e)} \quad & 4\delta m_T\sqrt{J}\frac{L}{T}\log\sum_{l,j\leq J}\exp\left(\frac{T}{2(L^2-LK_0)}\right) = 4\delta m_T\sqrt{J}\frac{L}{T}\left(2\log J+\frac{T}{2(L^2-LK_0)}\right) \\
= \quad & 4\delta m_T\sqrt{J}\left(\frac{2L\log J}{T}+\sqrt{\frac{c_0\log J}{4T}}\right)\leq C\delta m_T\sqrt{\frac{J\log J}{T}}.
\end{aligned}
$$

Note that $|\dot\rho(\cdot)|\leq 2$ and $\{\phi_j(\cdot)\}$ is sub-Gaussian, hence (a) follows from the symmetrization theorem (see, e.g., Theorem 14.3 of Bühlmann and van de Geer (2011)); since $\dot\rho(\cdot)$ is Lipschitz continuous, (b) follows from the contraction theorem (e.g., Theorem 14.4 of Bühlmann and van de Geer (2011)). Let K_0 denote constant parameter of the sub-Gaussianity of $\{\phi_l(\mathbf{x}_t)\phi_j(\mathbf{x}_t)\}_{l,j\leq J}$; for some $c_0>0$, let

$$L = K_0+\sqrt{\frac{T}{c_0\log J}}.$$

Then (c) follows from the Jensen's inequality; (d) follows from the simple inequality that $\exp(\max) \leq \sum \exp$; (e) follows from an inequality of exponential moment of an average for sub-Gaussian random variables (Lemma 14.8 of Bühlmann and van de Geer (2011)).

Therefore,

$$E \max_i \sup_{\|\mathbf{b}-\mathbf{b}_{i,\alpha}\| \leq \delta m_T} \|\boldsymbol{\mu}_i(\mathbf{b}) - \boldsymbol{\mu}_i(\mathbf{b}_{i,\alpha})\| \leq CJm_T \sqrt{\frac{\log J}{T}} = \frac{CJ^{3/2}(\log N \log J)^{1/2}}{T}.$$

Hence

$$\max_i \|\boldsymbol{\mu}_i(\widehat{\mathbf{b}}_i) - \boldsymbol{\mu}_i(\mathbf{b}_{i,\alpha})\| = O_P(J^{3/2}(\log N \log J)^{1/2}/T).$$

(ii) Note that

$$\nabla \bar{Q}_i(\mathbf{b}) = -E\Phi(\mathbf{x}_t)\alpha_T \dot{\rho}(\alpha_T^{-1}(e_{it} + z_{it}) + \alpha_T^{-1}\Phi(\mathbf{x}_t)'(\mathbf{b}_i - \mathbf{b})) = -E\Phi(\mathbf{x}_t)A_{it}(\mathbf{b})$$

where $A_{it}(\mathbf{b}) = E[\alpha_T \dot{\rho}(\alpha_T^{-1}(e_{it} + z_{it}) + \alpha_T^{-1}\Phi(\mathbf{x}_t)'(\mathbf{b}_i - \mathbf{b}))|\mathbf{x}_t]$. Let $g_{e,i}$ denote the density of e_{it}, and let P_e denote the conditional probability measure conditioning on $\mathbf{x}_t$. Then careful calculations yield: $\nabla A_{it}(\mathbf{b}) = -2\Phi(\mathbf{x}_t)' + \sum_{j=1}^{8} B_{it,j}(\mathbf{b})\Phi(\mathbf{x}_t)'$, where

$$\begin{aligned}
B_{it,1}(\mathbf{b}) &= -2\alpha_T g_{e,i}(\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{x}_t) - z_{it}), \\
B_{it,2}(\mathbf{b}) &= -2\alpha_T g_{e,i}(-\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{x}_t) - z_{it}), \\
B_{it,3}(\mathbf{b}) &= -2P_e((\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{x}_t) + z_{it} + e_{it} > \alpha_T), \\
B_{it,4}(\mathbf{b}) &= 2((\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{x}_t) + z_{it})g_{e,i}(\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{x}_t) - z_{it}), \\
B_{it,5}(\mathbf{b}) &= 2P_e(e_{it} < -\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{x}_t) - z_{it}), \\
B_{it,6}(\mathbf{b}) &= -2((\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{x}_t) + z_{it})g_{e,i}(-\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{x}_t) - z_{it}), \\
B_{it,7}(\mathbf{b}) &= 2[\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{x}_t) - z_{it}]g_{e,i}(\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{x}_t) - z_{it}), \\
B_{it,8}(\mathbf{b}) &= -2(-\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{x}_t) - z_{it})g_{e,i}(-\alpha_T - (\mathbf{b}_i - \mathbf{b})'\Phi(\mathbf{x}_t) - z_{it}).
\end{aligned}$$

Since $\max_i \|\widehat{\mathbf{b}}_i - \mathbf{b}_i\| = o(m_T)$, $\max_{it} |z_{it}| = o_P(\alpha_T)$, $\Phi(\mathbf{x}_t)$ is sub-Gaussian and $J \log N \sqrt{\log T} =$

$o(T)$, we have: with probability approaching one, for any $\epsilon > 0$,

$$\max_{i,t} |(\mathbf{b}_i - \widetilde{\mathbf{b}}_i)'\Phi(\mathbf{x}_t)| + \max_{it} |z_{it}| < \epsilon \alpha_T.$$

Hence with probability approaching one,

$$\begin{aligned}
\max_i |\sum_{j\neq 3,5} B_{it,j}(\widetilde{\mathbf{b}}_i)| &\leq C\alpha_T \max_i \sup_{|x|<\epsilon\alpha_T} g_{e,i}(\pm\alpha_T + x) \leq C\alpha_T^{-(\zeta_1-1)}, \\
\max_i |B_{it,3}(\widetilde{\mathbf{b}}_i) + B_{it,5}(\widetilde{\mathbf{b}}_i)| &\leq C \max_i P(|e_{it}| > (1-\epsilon)\alpha_T) \leq C\alpha_T^{-(\zeta_2+2)}.
\end{aligned}$$

Hence

$$\|\nabla^2 \bar{Q}_i(\widetilde{\mathbf{b}}_i) - 2E\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'\| = \|\sum_{j=1}^{8} E\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'B_{it,j}(\widetilde{\mathbf{b}}_i)\| = O(J\alpha_T^{-(\zeta_1-1)} + J\alpha_T^{-(\zeta_2+2)}).$$

Consequently, $\widehat{\mathbf{b}}_i - \mathbf{b}_{i,\alpha} = -(2E\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)')^{-1}\nabla Q_i(\mathbf{b}_{i,\alpha}) + \mathbf{R}_{i,b}$, where

$$\begin{aligned}
\max_{i\leq N} \|\mathbf{R}_{i,b}\| &\leq \|(2E\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)')^{-1}\|(\|\nabla^2 \bar{Q}_i(\widetilde{\mathbf{b}}_i) - 2E\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'\|\|\widehat{\mathbf{b}}_i - \mathbf{b}_{i,\alpha}\| + \max_i \|\boldsymbol{\mu}_i(\widehat{\mathbf{b}}_i) - \boldsymbol{\mu}_i(\mathbf{b}_{i,\alpha})\|) \\
&= O_P(\alpha_T^{-(\zeta_1-1)} + \alpha_T^{-(\zeta_2+2)} + \sqrt{\frac{\log J}{T}})Jm_T
\end{aligned}$$

Proposition B.3. *Let $\widehat{E}(y_{it}|\mathbf{x}_t) = \widehat{\mathbf{b}}_i'\Phi(\mathbf{x}_t)$. Then for $\mathbf{A} = (2E\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)')^{-1}$,*

$$\widehat{E}(y_{it}|\mathbf{x}_t) = E(y_{it}|\mathbf{x}_t) + \Phi(\mathbf{x}_t)'\mathbf{A}\frac{1}{T}\sum_{s=1}^{T} \alpha_T \dot{\rho}(\alpha_T^{-1}e_{is})\Phi(\mathbf{x}_s) + R_{1,it} + R_{2,it} + R_{3,it},$$

where

$$\begin{aligned}
R_{1,it} &:= \Phi(\mathbf{x}_t)'\mathbf{A}\frac{1}{T}\sum_{s=1}^{T} \alpha_T[\dot{\rho}(\alpha_T^{-1}e_{is,\alpha}) - \dot{\rho}(\alpha_T^{-1}e_{is})]\Phi(\mathbf{x}_s) \\
R_{2,it} &:= \Phi(\mathbf{x}_t)'(\mathbf{R}_{i,b} + \mathbf{b}_{i,\alpha} - \mathbf{b}_i), \qquad R_{3,it} := -z_{it}.
\end{aligned}$$

Write $R_{it} := R_{1,it} + R_{2,it} + R_{3,it}$, *then*

$$\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2 = O_P(J^{1-2\eta} + \alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{T} + \frac{J^3\log N\log J}{T^2}),$$
$$\max_i \frac{1}{T}\sum_{t=1}^T |\widehat{E}(y_{it}|\mathbf{x}_t) - E(y_{it}|\mathbf{x}_t)|^2 = O_P(\frac{J\log N}{T} + J^{-2\eta}).$$

Proof. By Lemma B.3 and Proposition B.3,

$$\begin{aligned}
&\widehat{E}(y_{it}|\mathbf{x}_t) = E(y_{it}|\mathbf{x}_t) + \Phi(\mathbf{x}_t)'(\widehat{\mathbf{b}}_i - \mathbf{b}_{i,\alpha}) + \Phi(\mathbf{x}_t)'(\mathbf{b}_{i,\alpha} - \mathbf{b}_i) - z_{it} \\
= \;& E(y_{it}|\mathbf{x}_t) + \Phi(\mathbf{x}_t)'\mathbf{A}\frac{1}{T}\sum_{s=1}^T \alpha_T\dot{\rho}(\alpha_T^{-1}e_{is,\alpha})\Phi(\mathbf{x}_s) + \Phi(\mathbf{x}_t)'(\mathbf{R}_{i,b} + \mathbf{b}_{i,\alpha} - \mathbf{b}_i) - z_{it} \\
= \;& E(y_{it}|\mathbf{x}_t) + \Phi(\mathbf{x}_t)'\mathbf{A}\frac{1}{T}\sum_{s=1}^T \alpha_T\dot{\rho}(\alpha_T^{-1}e_{is})\Phi(\mathbf{x}_s) + R_{it}.
\end{aligned}$$

On the other hand, uniformly in i, for $a = \lambda_{\max}(\frac{1}{T}\sum_{t=1}^T \Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)')$,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^T R_{it}^2 \le aC\|\mathbf{A}\|^2\|\frac{1}{T}\sum_{s=1}^T \alpha_T[\dot{\rho}(\alpha_T^{-1}e_{is,\alpha}) - \dot{\rho}(\alpha_T^{-1}e_{is})]\Phi(\mathbf{x}_s)\|^2 \\
&+aC\|\mathbf{R}_{i,b} + \mathbf{b}_{i,\alpha} - \mathbf{b}_i\|^2 + C\frac{1}{T}\sum_t z_{it}^2 \\
\le \;& C(\frac{1}{T}\sum_s |e_{is,\alpha} - e_{is}|\|\Phi(\mathbf{x}_s)\|)^2 + C\|\mathbf{R}_{i,b}\|^2 + C\|\mathbf{b}_{i,\alpha} - \mathbf{b}\|^2 + C\frac{1}{T}\sum_t z_{it}^2 \\
\le \;& C\frac{1}{T}\sum_s(|z_{it}|^2 + \|\mathbf{b}_{i,\alpha} - \mathbf{b}_i\|^2\|\Phi(\mathbf{x}_t)\|^2)\frac{1}{T}\sum_t \|\Phi(\mathbf{x}_t)\|^2 + C\|\mathbf{R}_{i,b}\|^2 + C\|\mathbf{b}_{i,\alpha} - \mathbf{b}\|^2 + O_P(J^{-2\eta}) \\
= \;& O_P(J)(J^{-2\eta} + J\alpha_T^{-2(k-1)}) + O_P(\alpha_T^{-2(\zeta_1-1)} + \alpha_T^{-2(\zeta_2+2)} + \frac{\log J}{T})J^2 m_T^2.
\end{aligned}$$

Also note that $\alpha_T^{-2\zeta_2-4} = O(\log N/T)$. Finally,

$$\max_i \frac{1}{T}\sum_{t=1}^T |\widehat{E}(y_{it}|\mathbf{x}_t) - E(y_{it}|\mathbf{x}_t)|^2 \le \max_i \frac{1}{T}\sum_{t=1}^T |\Phi(\mathbf{x}_t)'(\widehat{\mathbf{b}}_i - \mathbf{b}_i)|^2 + \max_i \frac{1}{T}\sum_{t=1}^T z_{it}^2$$

$$\leq \quad a\|\widehat{\mathbf{b}}_i - \mathbf{b}_i\|^2 + \max_i \frac{1}{T}\sum_{t=1}^T z_{it}^2 = O_P(\frac{J\log N}{T} + \alpha_T^{-2(k-1)} + J^{-2\eta}).$$

The term involving $\alpha_T^{-2(k-1)}$ is negligible since it is smaller than $(\log N/T)^3$.

B.3 Estimating the loadings

Write $\mathbf{M}_\alpha$ be an $N \times J$ matrix, whose ith row is given by

$$\mathbf{M}_{i,\alpha}' := \frac{1}{T}\sum_{s=1}^T \alpha_T \dot{\rho}(\alpha_T^{-1} e_{is})\Phi(\mathbf{x}_s)'.$$

Write $\mathbf{R}_t = (R_{1t}, ..., R_{Nt})'$. Then the Bahadur representation in Proposition B.3 can be written in the vector form: $\mathbf{A} = (2E\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)')^{-1}$,

$$\widehat{E}(\mathbf{y}_t|\mathbf{x}_t) = E(\mathbf{y}_t|\mathbf{x}_t) + \mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{x}_t) + \mathbf{R}_t = \mathbf{\Lambda}E(\mathbf{f}_t|\mathbf{x}_t) + \mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{x}_t) + \mathbf{R}_t. \tag{B.1}$$

Let $\widetilde{\mathbf{V}}$ be a $K \times K$ diagonal matrix, whose diagonal elements are the first K eigenvalues of $\widehat{\mathbf{\Sigma}}/N := \frac{1}{TN}\sum_{t=1}^T \widehat{E}(\mathbf{y}_t|\mathbf{x}_t)\widehat{E}(\mathbf{y}_t|\mathbf{x}_t)'$. By step 2 of the proof of Theorem 2.1, all the eigenvalues of $\widetilde{\mathbf{V}}$ are bounded away from both zero and infinity with probability approaching one. By the definition of $\widehat{\mathbf{\Lambda}}$, $\frac{1}{N}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{\Lambda}} = \widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}$. Plugging in (B.1), we have,

$$\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H} = \sum_{i=1}^8 \mathbf{B}_i, \qquad \mathbf{H} = \frac{1}{TN}\sum_{t=1}^T E(\mathbf{f}_t|\mathbf{x}_t)E(\mathbf{f}_t|\mathbf{x}_t)'\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1} \tag{B.2}$$

where

$$\mathbf{B}_1 = \frac{1}{TN}\sum_{t=1}^T \mathbf{\Lambda}E(\mathbf{f}_t|\mathbf{x}_t)\Phi(\mathbf{x}_t)'\mathbf{A}\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}, \qquad \mathbf{B}_2 = \frac{1}{TN}\sum_{t=1}^T \mathbf{\Lambda}E(\mathbf{f}_t|\mathbf{x}_t)\mathbf{R}_t'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1},$$

$$\mathbf{B}_3 = \frac{1}{TN}\sum_{t=1}^T \mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{x}_t)E(\mathbf{f}_t|\mathbf{x}_t)'\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1} \qquad \mathbf{B}_4 = \frac{1}{TN}\sum_{t=1}^T \mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'\mathbf{A}\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1},$$

$$\mathbf{B}_5 = \frac{1}{TN}\sum_{t=1}^{T}\mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{x}_t)\mathbf{R}_t'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}, \qquad \mathbf{B}_6 = \frac{1}{TN}\sum_{t=1}^{T}\mathbf{R}_t E(\mathbf{f}_t|\mathbf{x}_t)'\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1},$$
$$\mathbf{B}_7 = \frac{1}{TN}\sum_{t=1}^{T}\mathbf{R}_t\Phi(\mathbf{x}_t)'\mathbf{A}\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}, \qquad \mathbf{B}_8 = \frac{1}{TN}\sum_{t=1}^{T}\mathbf{R}_t\mathbf{R}_t'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}.$$

Lemma B.4. *Recall that $\mathbf{V}$ is a $K \times K$ diagonal matrix, whose diagonal elements are the eigenvalues of $\mathbf{\Sigma}_\Lambda^{1/2} E\{E(\mathbf{f}_t|\mathbf{x}_t)E(\mathbf{f}_t|\mathbf{x}_t)'\}\mathbf{\Sigma}_\Lambda^{1/2}$. Then $\widetilde{\mathbf{V}} \to^P \mathbf{V}$ and $\|\widehat{\mathbf{V}} - \mathbf{V}_N\| = O_P(J^{-\eta} + \sqrt{\frac{\log N}{T}})$.*

Proof. Let $\mathbf{V}_N$ be a $K \times K$ diagonal matrix, whose diagonal elements are the first K eigenvalues of $\mathbf{\Sigma}/N$. From step 4 of the proof of Theorem 2.1, the diagonal entries are also the eigenvalues of $\{E\{E(\mathbf{f}_t|\mathbf{x}_t)E(\mathbf{f}_t|\mathbf{x}_t)'\}\}^{1/2}\mathbf{\Sigma}_{\Lambda,N}\{E\{E(\mathbf{f}_t|\mathbf{x}_t)E(\mathbf{f}_t|\mathbf{x}_t)'\}\}^{1/2}$. Since $\mathbf{\Sigma}_{\Lambda,N} \to \mathbf{\Sigma}_\Lambda$, by Weyl's theorem

$$\|\mathbf{V} - \mathbf{V}_N\| \leq \|\{E\{E(\mathbf{f}_t|\mathbf{x}_t)E(\mathbf{f}_t|\mathbf{x}_t)'\}\}^{1/2}(\mathbf{\Sigma}_{\Lambda,N} - \mathbf{\Sigma}_\Lambda)\{E\{E(\mathbf{f}_t|\mathbf{x}_t)E(\mathbf{f}_t|\mathbf{x}_t)'\}\}^{1/2}\| = o(1).$$

On the other hand, by Proposition B.3,

$$\begin{aligned}
\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_\infty &\leq \max_{ij}\frac{1}{T}\sum_t |\widehat{E}(y_{it}|\mathbf{x}_t)\widehat{E}(x_{jt}|\mathbf{x}_t) - E(y_{it}|\mathbf{x}_t)E(x_{jt}|\mathbf{x}_t)| \\
&\quad + \max_{ij}|\frac{1}{T}\sum_t E(y_{it}|\mathbf{x}_t)E(x_{jt}|\mathbf{x}_t) - E\{E(y_{it}|\mathbf{x}_t)E(x_{jt}|\mathbf{x}_t)\}| = o_P(1) \\
&\leq \max_i \|\widehat{\mathbf{b}}_i - \mathbf{b}_i\| O_P(\|\frac{1}{T}\sum_t \Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'\|) + O_P(J^{-\eta} + \sqrt{\frac{\log N}{T}}) \\
&= O_P(J^{-\eta} + \sqrt{\frac{\log N}{T}}).
\end{aligned}$$

By Weyls theorem and Davis-Kahan theorem , $\|\widetilde{\mathbf{V}} - \mathbf{V}_N\| \leq \frac{1}{N}\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\| \leq \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_\infty = O_P(J^{-\eta} + \sqrt{\frac{\log N}{T}}) = o_P(1)$. The result then follows from the triangular inequality.

B.3.1 Proof of Theorem 3.1: $\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}$

Result (3.2) follows from the following proposition.

Proposition B.4.

$$\frac{1}{N}\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}\|_F^2 = O_P(J/T + J^{1-2\eta} + \alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{T}).$$

Proof. Under the assumptions, $J\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2 = O_P(1)$, and $\|\mathbf{M}_\alpha\|^2/N = O_P(1)$. Hence from Lemma B.5 and Proposition B.3, and $\alpha_T = C\sqrt{T/\log(NJ)}$

$$\begin{aligned}
\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}\|_F^2 &= O_P(\sum_{i=1}^8 \|\mathbf{B}_i\|_F^2) = O_P(\|\mathbf{M}_\alpha\|^2 J\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2) \\
&\quad + O_P(\|\mathbf{M}_\alpha\|^2 + \|\mathbf{M}_\alpha\|_F^4/N) + O_P(N\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2 + N(\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)^2) \\
&= O_P(\|\mathbf{M}_\alpha\|^2) + O_P(N\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2).
\end{aligned}$$

The result then follows from Lemma B.5 and Proposition B.3. Note that the term that reflects the effect of $\alpha_T^{-(k-1)}$ is removed, since it is negligible when $k > 4$, which is guaranteed by our assumption that $\zeta_2 > 2$. Furthermore, by assumptions $J^2\log^3 N = O(T)$ and $\zeta_1 > 2$,

$$\alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{T} = (\frac{\log(JN)}{T})^{\zeta_1} J^3 = O(\frac{J}{T}).$$

Q.E.D.

B.3.2 Proof of Theorem 3.1: $\max_{i\le N}\|\boldsymbol{\lambda}_i - \mathbf{H}'\boldsymbol{\lambda}_i\|$

Note in Lemma B.5 below that terms $\mathbf{B}_1, \mathbf{B}_4$ and $\mathbf{B}_7$ have two upper bounds, where the second bound uses a simple inequality $\|\mathbf{M}_\alpha\widehat{\mathbf{\Lambda}}\|^2 \le \|\mathbf{M}_\alpha\|^2\|\widehat{\mathbf{\Lambda}}\|^2$. Such a simple inequality is crude, but is sufficient to prove Proposition B.4. On the other hand, given Proposition

B.4, a sharper rate for $\|\mathbf{M}_\alpha\widehat{\mathbf{\Lambda}}\|^2$ can be found. As a result, the first bounds for $\mathbf{B}_1, \mathbf{B}_4$ and $\mathbf{B}_7$ are used later to achieve sharp rates for $\widehat{\mathbf{g}}(\mathbf{x}_t)-\mathbf{g}(\mathbf{x}_t)$.

Lemma B.5. *(i)* $\|\mathbf{M}_\alpha\|^2=O_P(NJ/T+NJ^{1-2\eta})$,

(ii) $\|\mathbf{B}_1\|_F^2=O_P(\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2/N)=O_P(\|\mathbf{M}_\alpha\|^2)$, $\quad \|\mathbf{B}_3\|_F^2=O_P(\|\mathbf{M}_\alpha\|^2)$.

(iii) $\|\mathbf{B}_2\|_F^2=O_P(N\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2)=\|\mathbf{B}_6\|_F^2$,

(iv) $\|\mathbf{B}_4\|_F^2=O_P(\|\mathbf{M}_\alpha\|^2\|\mathbf{M}_\alpha\widehat{\mathbf{\Lambda}}\|^2/N^2)=O_P(\|\mathbf{M}_\alpha\|_F^4/N)$, $\qquad \|\mathbf{B}_8\|_F^2=O_P(N(\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2)^2)$,

(v) $\|\mathbf{B}_5\|_F^2=O_P(\|\mathbf{M}_\alpha\|^2J\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2)$.

$\|\mathbf{B}_7\|_F^2=O_P(\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2J\|\mathbf{M}_\alpha\widehat{\mathbf{\Lambda}}\|^2/N)=O_P(\|\mathbf{M}_\alpha\|^2J\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2)$.

Proof. (i) Recall that $e_{it}=e_{it,\alpha}+\Delta_{it,\alpha}$, where $\Delta_{it,\alpha}:=(\mathbf{b}_{i,\alpha}-\mathbf{b}_i)'\Phi(\mathbf{x}_t)-z_{it}$.

$$
\begin{aligned}
&E\|\mathbf{M}_\alpha\|_F^2=E\sum_{i=1}^N\|\mathbf{M}_{i,\alpha}\|^2=\sum_{i=1}^N\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\alpha_T\dot\rho(\alpha_T^{-1}e_{is})\phi_j(\mathbf{x}_s))^2\\
\leq\;& 2\sum_{i=1}^N\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\alpha_T\dot\rho(\alpha_T^{-1}e_{is,\alpha})\phi_j(\mathbf{x}_s))^2+2\sum_{i=1}^N\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T 2|e_{is}-e_{is,\alpha}||\phi_j(\mathbf{x}_s)|)^2\\
\leq\;& 2\sum_{i=1}^N\sum_{j=1}^J\frac{1}{T}\operatorname{var}(\alpha_T\dot\rho(\alpha_T^{-1}e_{is,\alpha})\phi_j(\mathbf{x}_s))\\
&+C\sum_{i=1}^N\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T|(\mathbf{b}_{i,\alpha}-\mathbf{b}_i)'\Phi(\mathbf{x}_s)\phi_j(\mathbf{x}_s)|)^2+C\sum_{i=1}^N\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T|z_{is}\phi_j(\mathbf{x}_s)|)^2\\
\leq\;& O(NJ/T+NJ^2\alpha_T^{-2(k-1)}+NJ^{1-2\eta}),
\end{aligned}
$$

where the first inequality is due to the triangular inequality and $|\dot\rho(t_1)-\dot\rho(t_2)|\leq 2|t_1-t_2|$; the second inequality is due to $E\dot\rho(\alpha_T^{-1}e_{is,\alpha})\Phi(\mathbf{x}_s)=0$ and that $e_{is}-e_{is,\alpha}=(\mathbf{b}_{i,\alpha}-\mathbf{b}_i)'\Phi(\mathbf{x}_s)-z_{is}$.

(ii) Since $\frac{1}{T}\sum_t\Phi(\mathbf{x}_t)E(\mathbf{f}_t|\mathbf{x}_t)'E(\mathbf{f}_t|\mathbf{x}_t)\Phi(\mathbf{x}_t)'-(\|\frac{1}{T}\sum_t E(\mathbf{f}_t|\mathbf{x}_t)\Phi(\mathbf{x}_t)')'\frac{1}{T}\sum_t E(\mathbf{f}_t|\mathbf{x}_t)\Phi(\mathbf{x}_t)'$ is semipositive definite, we have

$$
\|\frac{1}{T}\sum_t E(\mathbf{f}_t|\mathbf{x}_t)\Phi(\mathbf{x}_t)'\|^2\leq\|\frac{1}{T}\sum_t\Phi(\mathbf{x}_t)E(\mathbf{f}_t|\mathbf{x}_t)'E(\mathbf{f}_t|\mathbf{x}_t)\Phi(\mathbf{x}_t)'\|=O(\|E\Phi(\mathbf{x}_t)\mathbf{f}_t'\mathbf{f}_t\Phi(\mathbf{x}_t)\|).
$$

$\|\mathbf{B}_1\|_F^2\leq\frac{1}{N^2}\|\mathbf{\Lambda}\|^2\|\frac{1}{T}\sum_{t=1}^T E(\mathbf{f}_t|\mathbf{x}_t)\Phi(\mathbf{x}_t)'\mathbf{A}\|^2\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2\|\widetilde{\mathbf{V}}^{-1}\|^2=O_P(\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2/N)=O_P(\|\mathbf{M}_\alpha\|^2)$.

The bound for $\|\mathbf{B}_3\|_F^2$ is similar.

(iii) By Proposition B.3, $\frac{1}{T}\sum_t \|\mathbf{R}_t\|^2 \leq N \max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2$. Hence

$$\|\mathbf{B}_2\|_F^2 \leq \|\mathbf{\Lambda}\|_F^2 \|\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|_F^2 \frac{1}{N^2}\frac{1}{T}\sum_t \|E(\mathbf{f}_t|\mathbf{x}_t)\|^2 \frac{1}{T}\sum_t \|\mathbf{R}_t\|^2 = O_P(N \max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2).$$

The bound for $\|\mathbf{B}_6\|_F^2$ is similar.

(iv) $\|\mathbf{B}_4\|_F^2$ is upper bounded by

$$\frac{1}{N^2}\|\mathbf{M}_\alpha \mathbf{A}\|^2 \|\frac{1}{T}\sum_{t=1}^T \Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'\mathbf{A}\|^2 \|\mathbf{M}_\alpha' \widehat{\mathbf{\Lambda}}\|^2 \|\widetilde{\mathbf{V}}^{-1}\|^2 = O_P(\|\mathbf{M}_\alpha\|^2 \|\mathbf{M}_\alpha \widehat{\mathbf{\Lambda}}\|^2/N^2) = O_P(\|\mathbf{M}_\alpha\|_F^4/N).$$

Also, $\|\mathbf{B}_8\|_F^2 \leq \frac{1}{N^2}(\frac{1}{T}\sum_{t=1}^T \|\mathbf{R}_t\|^2)^2 \|\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2 = O_P(N(\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)^2)$.

(v) $\mathbf{B}_5$ and $\mathbf{B}_7$ are bounded similarly. We have,

$$\begin{aligned}
\|\mathbf{B}_7\|_F^2 &\leq \frac{1}{N^2}\|\frac{1}{T}\sum_{t=1}^T \mathbf{R}_t \Phi(\mathbf{x}_t)'\mathbf{A}\|^2 \|\mathbf{M}_\alpha' \widehat{\mathbf{\Lambda}}\|^2 \|\widetilde{\mathbf{V}}^{-1}\|^2 = O_P(\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2 J \|\mathbf{M}_\alpha \widehat{\mathbf{\Lambda}}\|^2/N) \\
&= O_P(\|\mathbf{M}_\alpha\|^2 J \max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)
\end{aligned}$$

.

Q.E.D.

Given Proposition B.4, due to

$$\|\mathbf{M}_\alpha' \widehat{\mathbf{\Lambda}}\|_F^2 \leq 2\|\mathbf{M}_\alpha\|^2 \|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}\|_F^2 + 2\|\mathbf{M}_\alpha' \mathbf{\Lambda}\|^2 \|\mathbf{H}\|_F^2,$$

the rate of convergence for $\|\mathbf{M}_\alpha' \widehat{\mathbf{\Lambda}}\|_F^2$ can be improved, reaching a sharper bound than $\|\mathbf{M}_\alpha\|^2 \|\widehat{\mathbf{\Lambda}}\|_F^2$. This is given in Lemma B.6 below. As a result, rates for $\mathbf{B}_1, \mathbf{B}_4, \mathbf{B}_7$ can be improved as well.

Lemma B.6. *Given Proposition B.4, we have*

$$
\begin{aligned}
\|\mathbf{M}_\alpha'\boldsymbol{\Lambda}\|_F^2 &= O_P(JN^2\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|/T + JN/T + N^2J\alpha_T^{-\zeta_2}/T),\\
\|\mathbf{M}_\alpha'\widehat{\boldsymbol{\Lambda}}\|_F^2 &= O_P(JN^2\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|/T + JN/T + N^2J\alpha_T^{-\zeta_2}/T + NJ^{1-2\eta}(J^{1-2\eta} + \alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{T})).
\end{aligned}
$$

Proof. The proof is a straightforward calculation as follows:

$$
\begin{aligned}
& E\|\mathbf{M}_\alpha'\boldsymbol{\Lambda}\|_F^2 = E\|\sum_{i=1}^N \boldsymbol{\lambda}_i\mathbf{M}_{i,\alpha}'\|_F^2 = E\|\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N \boldsymbol{\lambda}_i\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is})\Phi(\mathbf{x}_s)'\|_F^2\\
= \;& \sum_{k=1}^K\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N \lambda_{ik}\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is})\phi_j(\mathbf{x}_s))^2 = \sum_{k=1}^K\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N 2\lambda_{ik}e_{is}1\{|e_{is}|<\alpha_T\}\phi_j(\mathbf{x}_s))^2\\
& + \sum_{k=1}^K\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N \lambda_{ik}\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is})1\{|e_{is}|\geq\alpha_T\}\phi_j(\mathbf{x}_s))^2\\
\leq \;& 8\sum_{k=1}^K\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N \lambda_{ik}e_{is}\phi_j(\mathbf{x}_s))^2 + 8\sum_{k=1}^K\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N \lambda_{ik}e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{x}_s))^2\\
& + \sum_{k=1}^K\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N 2|\lambda_{ik}|\alpha_T 1\{|e_{is}|\geq\alpha_T\}\phi_j(\mathbf{x}_s))^2\\
\leq \;& 8\sum_{k=1}^K\sum_{j=1}^J \operatorname{var}(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N \lambda_{ik}e_{is}\phi_j(\mathbf{x}_s)) + 12\sum_{k=1}^K\sum_{j=1}^J E(\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N |\lambda_{ik}e_{is}|1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{x}_s))^2.
\end{aligned}
$$

To bound the first term, Let E_w be the conditional expectation given $\mathbf{x}_s$. We need to bound $\sum_{i,l\leq N}|E_w e_{is}e_{ls}|$. Note that $e_{is} = x_{is} - E(x_{is}|\mathbf{x}_s) = \boldsymbol{\lambda}_i'\boldsymbol{\gamma}_s + u_{is}$. Since $E(\mathbf{u}_s|\mathbf{f}_s,\mathbf{x}_t) = 0$, we have

$$
E(\boldsymbol{\gamma}_s\mathbf{u}_s'|\mathbf{x}_s) = E(\mathbf{f}_s\mathbf{u}_s'|\mathbf{x}_s) - (E\mathbf{f}_s|\mathbf{x}_s)E(\mathbf{u}_s'|\mathbf{x}_s) = E(\mathbf{f}_s\mathbf{u}_s'|\mathbf{x}_s) = E(\mathbf{f}_sE(\mathbf{u}_s'|\mathbf{x}_s,\mathbf{f}_s)|\mathbf{x}_s) = 0.
$$

Hence $E_w(e_{is}e_{ls}) = E_w(\boldsymbol{\lambda}_i'\boldsymbol{\gamma}_s + u_{is})(\boldsymbol{\lambda}_l'\boldsymbol{\gamma}_s + u_{ls}) = \boldsymbol{\lambda}_i'\operatorname{cov}(\boldsymbol{\gamma}_s)\boldsymbol{\lambda}_l + E_w(u_{is}u_{ls})$. Therefore,

$$
\begin{aligned}
& 8\sum_{k=1}^{K}\sum_{j=1}^{J}\operatorname{var}(\frac{1}{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\lambda_{ik}e_{is}\phi_j(\mathbf{x}_s)) = 8\sum_{k=1}^{K}\sum_{j=1}^{J}\frac{1}{T}\operatorname{var}(\sum_{i=1}^{N}\lambda_{ik}e_{is}\phi_j(\mathbf{x}_s)) \\
= \;& 8\sum_{k=1}^{K}\sum_{j=1}^{J}\frac{1}{T}\sum_{i=1}^{N}\sum_{l=1}^{N}\lambda_{ik}\lambda_{lk}E\{E_w(e_{is}e_{ls})\phi_j(\mathbf{x}_s)^2\} \leq C\sum_{j=1}^{J}\frac{1}{T}E\phi_j(\mathbf{x}_s)^2\sup_{\mathbf{x}}\sum_{i=1}^{N}\sum_{l=1}^{N}|E_w(e_{is}e_{ls})| \\
\leq \;& \frac{CJ}{T}\sum_{i=1}^{N}\sum_{l=1}^{N}|\boldsymbol{\lambda}_i'\operatorname{cov}(\boldsymbol{\gamma}_s)\boldsymbol{\lambda}_l| + \frac{CJ}{T}\sup_{\mathbf{x}}\sum_{i=1}^{N}\sum_{l=1}^{N}|E_w(u_{is}u_{ls})| \\
\leq \;& \frac{CJ}{T}N^2\|\operatorname{cov}(\boldsymbol{\gamma}_s)\| + \frac{CJN}{T}\sup_{\mathbf{x}}\max_{i\leq N}\sum_{l=1}^{N}|E_w(u_{is}u_{ls})| = O(JN^2\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|/T + JN/T).
\end{aligned}
$$

Note that the second term is bounded by

$$
\begin{aligned}
\leq \;& C\sum_{k=1}^{K}\sum_{j=1}^{J}\frac{1}{T}\sum_{i=1}^{N}\sum_{l=1}^{N}E|e_{is}|1\{|e_{is}|>\alpha_T\}|e_{ls}|1\{|e_{ls}|>\alpha_T\}\phi_j(\mathbf{x}_s)^2 \\
& +C\sum_{k=1}^{K}\sum_{j=1}^{J}\frac{1}{T^2}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{t\neq s}^{T}\sum_{l=1}^{N}E|e_{is}|1\{|e_{is}|>\alpha_T\}|\phi_j(\mathbf{x}_s)|E|e_{lt}|1\{|e_{lt}|>\alpha_T\}|\phi_j(\mathbf{x}_t)| \\
\leq \;& C\sum_{k=1}^{K}\sum_{j=1}^{J}\frac{1}{T}\sum_{i=1}^{N}\sum_{l=1}^{N}\sup_{\mathbf{x}}E_w|e_{is}|1\{|e_{is}|>\alpha_T\}|e_{ls}|1\{|e_{ls}|>\alpha_T\} \\
& +C\sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{N}\sum_{l=1}^{N}(\sup_{\mathbf{x}}E|e_{is}|1\{|e_{is}|>\alpha_T\})^2 \\
\leq \;& C\frac{KJ}{T}N^2\max_i\sup_{\mathbf{x}}E_w e_{is}^2 1\{|e_{is}|>\alpha_T\} + CKJN^2(\max_i\sup_{\mathbf{x}}E|e_{is}|1\{|e_{is}|>\alpha_T\})^2 \\
= \;& O(N^2J\alpha_T^{-\zeta_2}/T + N^2J\alpha_T^{-2(\zeta_2+1)}).
\end{aligned}
$$

Hence $E\|\mathbf{M}_\alpha'\boldsymbol{\Lambda}\|_F^2 = O(JN^2\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|/T + JN/T + N^2J\alpha_T^{-\zeta_2}/T + N^2J\alpha_T^{-2(\zeta_2+1)})$.
The rate for $\|\mathbf{M}_\alpha'\widehat{\boldsymbol{\Lambda}}\|_F^2$ comes from

$$
\|\mathbf{M}_\alpha'\widehat{\boldsymbol{\Lambda}}\|_F^2 \leq 2\|\mathbf{M}_\alpha\|^2\|\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\mathbf{H}\|_F^2 + 2\|\mathbf{M}_\alpha'\boldsymbol{\Lambda}\|^2\|\mathbf{H}\|_F^2.
$$

Lemma B.7. $\max_{i\leq N}\|\mathbf{M}_{i,\alpha}\| = O_P(J^{-\eta}\sqrt{J} + \sqrt{J(\log N)/T})$

Proof. First, it follows from the proof of Proposition B.2 that

$$\max_i \|\frac{1}{T}\sum_{s=1}^{T}\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is,\alpha})\Phi(\mathbf{x}_s)\| = O_P(\sqrt{\frac{J\log N}{T}}).$$

Secondly, since $|\dot{\rho}(t_1)-\dot{\rho}(t_2)|\leq 2|t_1-t_2|$,

$$\begin{aligned}
&\max_i \|\frac{1}{T}\sum_{s=1}^{T}\alpha_T(\dot{\rho}(\alpha_T^{-1}e_{is})-\dot{\rho}(\alpha_T^{-1}e_{is,\alpha}))\Phi(\mathbf{x}_s)\| \leq \max_i \|\frac{1}{T}\sum_{s=1}^{T}2|e_{is}-e_{is,\alpha}|\Phi(\mathbf{x}_s)\| \\
\leq\ & \max_i \|\frac{1}{T}\sum_{s=1}^{T}2|(\mathbf{b}_{i,\alpha}-\mathbf{b}_i)'\Phi(\mathbf{x}_t)-z_{it}|\Phi(\mathbf{x}_s)\| \leq 2\max_i\|\mathbf{b}_{i,\alpha}-\mathbf{b}_i\|O_P(J)+O_P(J^{-\eta}\sqrt{J}) \\
=\ & O_P(J^{-\eta}\sqrt{J}+J\alpha_T^{-(k-1)}).
\end{aligned}$$

The result then follows from the triangular inequality.

Q.E.D.

Proof of Theorem 3.1 result (3.3): $\max_{i\leq N}\|\boldsymbol{\lambda}_i-\mathbf{H}'\boldsymbol{\lambda}_i\|$

Proof. Let $\mathbf{B}_{i1},...,\mathbf{B}_{i8}$ respectively denote the ith row of $\mathbf{B}_1,...,\mathbf{B}_8$. We have

$$\begin{aligned}
\max_i\|\mathbf{B}_{i1}\| &\leq O_P(\|\mathbf{M}_\alpha\widehat{\mathbf{\Lambda}}\|/N)\leq O_P(\max_i\|\mathbf{M}_{i,\alpha}\|) \\
\max_i\|\mathbf{B}_{i2}\| &\leq O_P(\max_i\frac{1}{T}\sum_{t=1}^{T}R_{it}^2)^{1/2}) \\
\max_i\|\mathbf{B}_{i3}\| &\leq O_P(\max_i\|\mathbf{M}_{i,\alpha}\|)=O_P(J^{-\eta}\sqrt{J}+J\alpha_T^{-(k-1)}+\sqrt{J(\log N)/T}) \\
\max_i\|\mathbf{B}_{i4}\| &\leq O_P(\max_i\|\mathbf{M}_{i,\alpha}\|)O_P(\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|/N) \\
\max_i\|\mathbf{B}_{i5}\| &\leq O_P(\max_i\|\mathbf{M}_{i,\alpha}\|)O_P(\sqrt{J}\max_i\frac{1}{T}\sum_t R_{it}^2)^{1/2} \\
\max_i\|\mathbf{B}_{i6}\| &\leq O_P(\max_i\frac{1}{T}\sum_t R_{it}^2)^{1/2})
\end{aligned}$$

$$
\begin{aligned}
\max_i \|\mathbf{B}_{i7}\| &\leq O_P(\max_i \frac{1}{T}\sum_t R_{it}^2)^{1/2}\sqrt{J})O_P(\|\mathbf{M}_\alpha\widehat{\mathbf{\Lambda}}\|/N) \\
\max_i \|\mathbf{B}_{i8}\| &\leq O_P(\max_i \frac{1}{T}\sum_t R_{it}^2).
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\max_{i\leq N} \|\boldsymbol{\lambda}_i - \mathbf{H}'\boldsymbol{\lambda}_i\| \leq O_P(\max_i \|\mathbf{B}_{i2}\| + \max_i \|\mathbf{B}_{i3}\|) \\
= \;& O_P(J^{-\eta}\sqrt{J} + \sqrt{J(\log N)/T} + \alpha_T^{-(\zeta_1-1)}\sqrt{\frac{J^3\log N}{T}}) = O_P(J^{-\eta}\sqrt{J} + \sqrt{J(\log N)/T}),
\end{aligned}
$$

where the last equality follows from

$$
\alpha_T^{-(\zeta_1-1)}\sqrt{\frac{J^3\log N}{T}} = (\frac{\log(NJ)}{T})^{\zeta_1/2}J^{3/2} = O(\sqrt{\frac{J\log N}{T}})
$$

under assumptions $(\log N)^3J^2 = O(T)$ and $\zeta_1 > 2$.

B.4 Proof of Theorem 3.2: $\frac{1}{T}\sum_{t=1}^T \|\widehat{\mathbf{g}}(\mathbf{x}_t) - \mathbf{H}^{-1}\mathbf{g}(\mathbf{x}_t)\|^2$

Recall that $\widehat{\mathbf{g}}(\mathbf{x}_t) = \frac{1}{N}\widehat{\mathbf{\Lambda}}'\widehat{E}(\mathbf{y}_t|\mathbf{x}_t)$. By (B.1), $\widehat{\mathbf{g}}(\mathbf{x}_t) - \mathbf{H}^{-1}\mathbf{g}(\mathbf{x}_t) = \sum_{i=1}^4 \mathbf{C}_{ti}$, where

$$
\begin{aligned}
\mathbf{C}_{t1} &= \frac{1}{N}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})'(\mathbf{\Lambda}\mathbf{H} - \widehat{\mathbf{\Lambda}})\mathbf{H}^{-1}E(\mathbf{f}_t|\mathbf{x}_t), \qquad \mathbf{C}_{t2} = -\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\mathbf{H}^{-1}E(\mathbf{f}_t|\mathbf{x}_t) \\
\mathbf{C}_{t3} &= \frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{x}_t), \qquad \mathbf{C}_{t4} = \frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{R}_t.
\end{aligned}
$$

Lemma B.8. *Assume $J = O(N)$,*

$$
\|\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\|_F = O_P(\sqrt{\frac{J}{TN}} + \frac{\sqrt{J^3\log N\log J}}{T} + \sqrt{\frac{J\|\operatorname{cov}(\boldsymbol{\gamma}_t)\|}{T}} + J^{1/2-\eta} + \alpha_T^{-(\zeta_1-1)}\sqrt{\frac{J^3\log N}{T}}),
$$

and $\|\frac{1}{N}\widehat{\mathbf{\Lambda}}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\|_F$ has the same rate of convergence.

Proof. $\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}) = \sum_{i=1}^8 \mathbf{\Lambda}'\mathbf{B}_i$. Keep in mind that $\|\mathbf{\Lambda}'\mathbf{M}_\alpha\|$ and $\|\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\|$ have sharper bounds than $\|\mathbf{\Lambda}\|\|\mathbf{M}_\alpha\|$, $\|\widehat{\mathbf{\Lambda}}\|\|\mathbf{M}_\alpha\|$, given in Lemma B.6.

For $i \neq 3, 4, 5$, we simply use $\|\mathbf{\Lambda}'\mathbf{B}_i\| \leq \|\mathbf{\Lambda}\|\|\mathbf{B}_i\| = O(\sqrt{N})\|\mathbf{B}_i\|$ and Lemma B.5. But note that for $\mathbf{B}_1, \mathbf{B}_7$, the first upper bound in the lemma is used.

$$
\begin{aligned}
\|\mathbf{\Lambda}'\mathbf{B}_1\| &= O_P(\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|) \qquad \|\mathbf{\Lambda}'\mathbf{B}_2\| = O_P((\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)^{1/2}N) = \|\mathbf{\Lambda}'\mathbf{B}_6\| \\
\|\mathbf{\Lambda}'\mathbf{B}_7\| &= O_P((\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)^{1/2}J^{1/2}\|\mathbf{M}_\alpha\widehat{\mathbf{\Lambda}}\|), \qquad \|\mathbf{\Lambda}'\mathbf{B}_8\| = O_P(\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2 N).
\end{aligned}
$$

As for $\mathbf{B}_3, \mathbf{B}_4, \mathbf{B}_5$, we have

$$
\begin{aligned}
\|\mathbf{\Lambda}'\mathbf{B}_3\| &\leq O_P(1)\|\mathbf{\Lambda}'\mathbf{M}_\alpha\|\|\frac{1}{T}\sum_{t=1}^T \Phi(\mathbf{x}_t)E(\mathbf{f}_t|\mathbf{x}_t)'\| = O_P(\|\mathbf{\Lambda}'\mathbf{M}_\alpha\|). \\
\|\mathbf{\Lambda}'\mathbf{B}_4\| &\leq O_P(1)\|\mathbf{\Lambda}'\mathbf{M}_\alpha\|\|\frac{1}{TN}\sum_{t=1}^T \Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'\|\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\| = O_P(\|\mathbf{\Lambda}'\mathbf{M}_\alpha\|\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|/N) \\
\|\mathbf{\Lambda}'\mathbf{B}_5\| &\leq O_P(1)\|\mathbf{\Lambda}'\mathbf{M}_\alpha\|\|\frac{1}{T}\sum_{t=1}^T \Phi(\mathbf{x}_t)\mathbf{R}_t'\|/\sqrt{N} = O_P(\|\mathbf{\Lambda}'\mathbf{M}_\alpha\|(J\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)^{1/2}).
\end{aligned}
$$

Note that $\sqrt{J}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|_F \leq N$. Hence

$$
\begin{aligned}
\|\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\|_F &\leq \sqrt{K}\|\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\| = \frac{1}{N}O_P(\|\mathbf{\Lambda}'\mathbf{B}_1\| + \|\mathbf{\Lambda}'\mathbf{B}_2\| + \|\mathbf{\Lambda}'\mathbf{B}_3\| + \|\mathbf{\Lambda}'\mathbf{B}_6\|) \\
&= O_P(\frac{1}{N}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\| + (\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)^{1/2}).
\end{aligned}
$$

In addition, we have

$$
\begin{aligned}
\|\frac{1}{N}\widehat{\mathbf{\Lambda}}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\|_F &\leq \|\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\|_F + \frac{1}{N}\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}\|_F^2 \\
&= O_P(\frac{1}{N}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\| + \frac{1}{N}\|\mathbf{M}_\alpha\|^2 + (\max_i \frac{1}{T}\sum_{t=1}^T R_{it}^2)^{1/2}),
\end{aligned}
$$

which is straightforward to be verified to have the same rate as $\|\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})\|_F$.

Q.E.D.

Proof of Theorem 3.2: first result

The convergence of $\frac{1}{T}\sum_{t=1}^T\|\widehat{\mathbf{g}}(\mathbf{x}_t)-\mathbf{H}^{-1}\mathbf{g}(\mathbf{x}_t)\|^2$ in this theorem is proved in the following proposition.

Proposition B.5.

$$\frac{1}{T}\sum_{t=1}^T\|\widehat{\mathbf{g}}(\mathbf{x}_t)-\mathbf{H}^{-1}\mathbf{g}(\mathbf{x}_t)\|^2=O_P(\frac{J}{TN}+\frac{J^3\log N\log J}{T^2}+\frac{J\|\operatorname{cov}(\boldsymbol{\gamma}_t)\|}{T}+J^{1-2\eta}+\alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{T}).$$

Proof. By the proof of Proposition B.4 and Lemma B.8, $\frac{1}{N}\|\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H}\|_F^2=O_P(\frac{1}{N}\|\mathbf{M}_\alpha\|^2)+O_P(\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2)$

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^T\|\mathbf{C}_{t1}\|^2 &= O_P(\frac{1}{N^2}\|\mathbf{M}_\alpha\|^4+(\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2)^2)\\
\frac{1}{T}\sum_{t=1}^T\|\mathbf{C}_{t2}\|^2 &= O_P(\|\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})\|^2)=O_P(\frac{1}{N^2}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2+\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2)\\
\frac{1}{T}\sum_{t=1}^T\|\mathbf{C}_{t4}\|^2 &= O_P(\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2).
\end{aligned}$$

Finally, let $\boldsymbol{\beta}_i$ denote the ith row of $\frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\mathbf{A}$, $i\leq K$. Then

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^T\|\mathbf{C}_{t3}\|^2=\frac{1}{T}\sum_{t=1}^T\|\frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{x}_t)\|^2=\sum_{i=1}^K\frac{1}{T}\sum_{t=1}^T(\boldsymbol{\beta}_i'\Phi(\mathbf{x}_t))^2\\
\leq&\sum_{i=1}^K\|\boldsymbol{\beta}_i\|^2\|\frac{1}{T}\sum_{t=1}^T\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'\|=O_P(1)\|\frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\mathbf{A}\|_F^2=O_P(\frac{1}{N^2}\|\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\|^2).
\end{aligned}$$

Hence

$$\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{g}}(\mathbf{x}_t)-\mathbf{g}(\mathbf{x}_t)\|^2 = O_P(\frac{1}{N^2}\|\mathbf{M}_\alpha\|^4+\frac{1}{N^2}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2+\max_i\frac{1}{T}\sum_{t=1}^{T}R_{it}^2).$$

$$= O_P(\frac{J}{TN}+\frac{J^3\log N\log J}{T^2}+\frac{J\|\operatorname{cov}(\boldsymbol{\gamma}_t)\|}{T}+J^{1-2\eta}+\alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{T}).$$

Finally, due to $\zeta_1\geq 2.5$ and $\log N^2=O(T)$,

$$\alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{T}=O(\frac{J^3\log N\log J}{T^2}).$$

B.5 Proof of Theorem 3.2: $\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}^{-1}\boldsymbol{\gamma}_t\|^2$

Note that $\mathbf{y}_t-E(\mathbf{y}_t|\mathbf{x}_t)=\mathbf{\Lambda}\boldsymbol{\gamma}_t+\mathbf{u}_t$. and $\widehat{\boldsymbol{\gamma}}_t=\frac{1}{N}\widehat{\mathbf{\Lambda}}'(\mathbf{y}_t-\widehat{E}(\mathbf{y}_t|\mathbf{x}_t))$. Hence from (B.1)

$$\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}^{-1}\boldsymbol{\gamma}_t=\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t+\sum_{i=1}^{d}\mathbf{D}_{ti}, \tag{B.3}$$

where for $\mathbf{C}_{t3},\mathbf{C}_{t4}$ defined earlier,

$$\begin{aligned}\mathbf{D}_{t1} &= \frac{1}{N}\widehat{\mathbf{\Lambda}}'(\mathbf{\Lambda}\mathbf{H}-\widehat{\mathbf{\Lambda}})\mathbf{H}^{-1}\boldsymbol{\gamma}_t, \qquad \mathbf{D}_{t2}=\frac{1}{N}(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})'\mathbf{u}_t\\ \mathbf{D}_{t3} &= \frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{x}_t)=\mathbf{C}_{t3}, \qquad \mathbf{D}_{t4}=\frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{R}_t=\mathbf{C}_{t4}.\end{aligned}$$

Hence for a constant $C>0$, $\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}^{-1}\boldsymbol{\gamma}_t\|^2\leq C\sum_{i=1}^{4}\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{D}_{ti}\|^2$. We look at terms on the right hand side one by one. First of all,

$$\begin{aligned}E\|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t'-\operatorname{cov}(\boldsymbol{\gamma}_t)\|_F^2 &= \sum_{i=1}^{K}\sum_{j=1}^{K}\operatorname{var}(\frac{1}{T}\sum_{t=1}^{T}\gamma_{it}\gamma_{jt})=\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{1}{T}\operatorname{var}(\gamma_{it}\gamma_{jt})\\ &= O(T^{-1})\max_{i,j\leq K}\operatorname{var}(\gamma_{it}\gamma_{jt}).\end{aligned}$$

As for $\mathbf{D}_{t1}$, let $\mathbf{G} = \frac{1}{N}\widehat{\mathbf{\Lambda}}'(\mathbf{\Lambda}\mathbf{H} - \widehat{\mathbf{\Lambda}})\mathbf{H}^{-1}$ and let $\mathbf{G}_i'$ denote its ith row, $i \leq K$. Then

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{D}_{t1}\|^2 &= \sum_{i=1}^{K}\frac{1}{T}\sum_{t=1}^{T}(\mathbf{G}_i'\boldsymbol{\gamma}_t)^2 = \sum_{i=1}^{K}\mathbf{G}_i'\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t'\mathbf{G}_i \leq \|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t'\|\|\mathbf{G}\|_F^2 \\
&= \|\mathbf{G}\|_F^2(\|\operatorname{cov}(\boldsymbol{\gamma}_t)\| + O(\frac{1}{T})\max_{i,j\leq K}\operatorname{var}(\gamma_{it}\gamma_{jt})) \\
&= O_P(\|\operatorname{cov}(\boldsymbol{\gamma}_t)\| + \frac{1}{T}\max_{i,j\leq K}\operatorname{var}(\gamma_{it}\gamma_{jt}))O_P(\frac{1}{N^2}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2 + \frac{1}{N^2}\|\mathbf{M}_\alpha\|^4 + \max_i\frac{1}{T}\sum_{t=1}^{T}R_{it}^2)
\end{aligned}$$

Terms $\mathbf{D}_{t3}$ and $\mathbf{D}_{t4}$ were bounded earlier. Term $\mathbf{D}_{t2}$ is given in Lemma B.11 below. Also note that the rate for $\frac{1}{N^2}\|\mathbf{M}_\alpha\|^4$ is dominated by the rate of $\frac{1}{N^2}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2 + \max_i\frac{1}{T}\sum_{t=1}^{T}R_{it}^2$. Thus by Proposition B.3, Lemmas B.6, B.9, B.10,

$$\begin{aligned}
\sum_{i=1}^{4}\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{D}_{ti}\|^2 &= O_P(\frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2 + \frac{1}{N^2T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}|\mathbf{u}_s'\mathbf{R}_t|^2 + \frac{1}{N^2}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2 + \max_i\frac{1}{T}\sum_{t=1}^{T}R_{it}^2) \\
&= O_P(\frac{J}{TN} + \frac{J^4\log N\log J}{T^2} + \frac{J\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|}{T} + J^{1-2\eta} + \alpha_T^{-2(\zeta_1-1)}\frac{J^4\log N}{T}).
\end{aligned}$$

Finally, $\frac{1}{T}\sum_{t=1}^{T}\|\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t\|^2 = O_P(\frac{1}{TN^2}\sum_{t=1}^{T}E\|\mathbf{\Lambda}'\mathbf{u}_t\|^2) = O_P(\frac{1}{N})$. Hence

$$\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{\gamma}}_t - \mathbf{H}^{-1}\boldsymbol{\gamma}_t\|^2 = O_P(\frac{1}{N} + \frac{J^4\log N\log J}{T^2} + \frac{J\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|}{T} + J^{1-2\eta} + \alpha_T^{-2(\zeta_1-1)}\frac{J^4\log N}{T}).$$

Lemma B.9. $\sum_{t=1}^{T}\|\mathbf{u}_t'\mathbf{M}_\alpha\|^2 = O_P(JN\|\operatorname{cov}(\boldsymbol{\gamma}_s)\| + JN^2/T + J + JN^2\alpha_T^{-\zeta_2})$.

Proof. Note that $E\sum_{t=1}^{T}\|\mathbf{u}_t'\mathbf{M}_\alpha\|^2 = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{J}E(\sum_{i=1}^{N}\sum_{s=1}^{T}u_{it}\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is})\phi_j(\mathbf{x}_s))^2$. We now bound the right hand side. In fact, since $e_{is} = \boldsymbol{\lambda}_i'\boldsymbol{\gamma}_s + u_{is}$,

$$\begin{aligned}
&E(\sum_{i=1}^{N}\sum_{s=1}^{T}u_{it}\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is})\phi_j(\mathbf{x}_s))^2 \\
\leq\ & 8E(\sum_{i=1}^{N}\sum_{s=1}^{T}u_{it}e_{is}1\{|e_{is}| < \alpha_T\}\phi_j(\mathbf{x}_s))^2 + 2E(\sum_{i=1}^{N}\sum_{s=1}^{T}u_{it}\alpha_T\dot{\rho}(\alpha_T^{-1}e_{is})1\{|e_{is}| \geq \alpha_T\}\phi_j(\mathbf{x}_s))^2
\end{aligned}$$

$$
\begin{aligned}
&\leq CE(\sum_{i=1}^{N}\sum_{s=1}^{T}u_{it}e_{is}\phi_j(\mathbf{x}_s))^2+CE(\sum_{i=1}^{N}\sum_{s=1}^{T}|u_{it}e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{x}_s)|)^2\\
&\leq CE(\sum_{i=1}^{N}\sum_{s=1}^{T}u_{it}\boldsymbol{\lambda}_i'\boldsymbol{\gamma}_s\phi_j(\mathbf{x}_s))^2+CE(\sum_{i=1}^{N}\sum_{s=1}^{T}(u_{it}u_{is}-E(u_{it}u_{is}))\phi_j(\mathbf{x}_s))^2+\\
&\quad +CE(\sum_{i=1}^{N}\sum_{s=1}^{T}(Eu_{it}u_{is})\phi_j(\mathbf{x}_s))^2+CE(\sum_{i=1}^{N}\sum_{s=1}^{T}|u_{it}e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{x}_s)|)^2.
\end{aligned}
$$

The first term is bounded as: uniformly in t,

$$
\begin{aligned}
&E(\sum_{i=1}^{N}\sum_{s=1}^{T}u_{it}\boldsymbol{\lambda}_i'\boldsymbol{\gamma}_s\phi_j(\mathbf{x}_s))^2=\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{l=1}^{N}\sum_{k=1}^{T}\boldsymbol{\lambda}_i'E\boldsymbol{\gamma}_s\phi_j(\mathbf{x}_s)u_{lt}u_{it}\phi_j(\mathbf{x}_k)\boldsymbol{\gamma}_k'\boldsymbol{\lambda}_l\\
&=\sum_{i=1}^{N}\sum_{l=1}^{N}\boldsymbol{\lambda}_i'E\boldsymbol{\gamma}_t u_{lt}u_{it}\phi_j(\mathbf{x}_t)^2\boldsymbol{\gamma}_t'\boldsymbol{\lambda}_l+\sum_{i=1}^{N}\sum_{s\neq t}\sum_{l=1}^{N}\boldsymbol{\lambda}_i'\operatorname{cov}(\boldsymbol{\gamma}_s)E\phi_j(\mathbf{x}_s)^2\boldsymbol{\lambda}_l Eu_{lt}u_{it}\\
&\leq\sum_{i=1}^{N}\sum_{l=1}^{N}E[|(Eu_{lt}u_{it}|\mathbf{x}_t,\mathbf{f}_t)|\phi_j(\mathbf{x}_t)^2\|\boldsymbol{\gamma}_t\|^2]\max_i\|\boldsymbol{\lambda}_i\|^2+T\sum_{i=1}^{N}\sum_{l=1}^{N}\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|E\phi_j(\mathbf{x}_s)^2|Eu_{lt}u_{it}|\max_i\|\boldsymbol{\lambda}_i\|^2\\
&\leq NC\sup_{\mathbf{x},\mathbf{f}}\max_i\sum_{l=1}^{N}|(Eu_{lt}u_{it}|\mathbf{x}_t,\mathbf{f}_t)|\sup_{\mathbf{x}}E(\|\boldsymbol{\gamma}_t\|^2|\mathbf{x}_t=\mathbf{x})E\phi_j(\mathbf{x}_t)^2+\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|TNC\max_i\sum_{l=1}^{N}|Eu_{lt}u_{it}|\\
&\leq NC\sup_{\mathbf{x},\mathbf{f}}\max_i\sum_{l=1}^{N}|(Eu_{lt}u_{it}|\mathbf{x}_t,\mathbf{f}_t)|\|\operatorname{cov}(\boldsymbol{\gamma}_t)\|+\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|TNC\max_i\sum_{l=1}^{N}|Eu_{lt}u_{it}|\\
&=O(TN\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|).
\end{aligned}
$$

The second term: note that for some $v>1$, $E\{Eu_{it}^4|\mathbf{x}_t\}^v<\infty$, uniformly in t,

$$
\begin{aligned}
&E(\sum_{i=1}^{N}\sum_{s=1}^{T}(u_{it}u_{is}-E(u_{it}u_{is}))\phi_j(\mathbf{x}_s))^2\\
&=\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{l=1}^{N}\sum_{k=1}^{T}E(u_{it}u_{is}-E(u_{it}u_{is}))(u_{lt}u_{lk}-E(u_{lt}u_{lk}))\phi_j(\mathbf{x}_k)\phi_j(\mathbf{x}_s)\\
&=\sum_{i=1}^{N}\sum_{l=1}^{N}E(u_{it}^2-Eu_{it}^2)(u_{lt}^2-Eu_{lt}^2)\phi_j(\mathbf{x}_t)^2+\sum_{i=1}^{N}\sum_{s\neq t}\sum_{l=1}^{N}Eu_{it}(u_{lt}^2-Eu_{lt}^2)\phi_j(\mathbf{x}_t)Eu_{is}\phi_j(\mathbf{x}_s)
\end{aligned}
$$

$$
\begin{aligned}
& +\sum_{i=1}^N\sum_{l=1}^N\sum_{k\neq t}E(u_{it}^2-Eu_{it}^2)u_{lt}\phi_j(\mathbf{x}_t)E\phi_j(\mathbf{x}_k)u_{lk}+\sum_{i=1}^N\sum_{s\neq t}\sum_{l=1}^N Eu_{it}u_{lt}Eu_{ls}u_{is}\phi_j(\mathbf{x}_s)^2 \\
& +\sum_{i=1}^N\sum_{s\neq t}\sum_{l=1}^N\sum_{k\neq t,s}Eu_{it}u_{lt}Eu_{is}\phi_j(\mathbf{x}_s)E\phi_j(\mathbf{x}_k)u_{lk} \\
= & \sum_{i=1}^N\sum_{l=1}^N E(u_{it}^2-Eu_{it}^2)(u_{lt}^2-Eu_{lt}^2)\phi_j(\mathbf{x}_t)^2+\sum_{i=1}^N\sum_{s\neq t}\sum_{l=1}^N Eu_{it}u_{lt}Eu_{ls}u_{is}\phi_j(\mathbf{x}_s)^2 \\
\leq & \sum_{i=1}^N\sum_{l=1}^N E(u_{it}^2-Eu_{it}^2)(u_{lt}^2-Eu_{lt}^2)\phi_j(\mathbf{x}_t)^2+CT(\max_i\sum_{l=1}^N|Eu_{it}u_{lt}|)(\sup_{\mathbf{x}}\max_l\sum_{i=1}^N|Eu_{ls}u_{is}|\mathbf{x})|E\phi_j(\mathbf{x}_s)^2) \\
= & O(N^2+T).
\end{aligned}
$$

The third term is bounded as: uniformly in t,

$$
E(\sum_{i=1}^N\sum_{s=1}^T(Eu_{it}u_{is})\phi_j(\mathbf{x}_s))^2=E(\sum_{i=1}^N(Eu_{it}^2)\phi_j(\mathbf{x}_t))^2=O(N^2).
$$

Finally, the fourth term is :

$$
\begin{aligned}
& E(\sum_{i=1}^N\sum_{s=1}^T|u_{it}e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{x}_s)|)^2 \\
= & E\sum_{i=1}^N\sum_{s=1}^T\sum_{l=1}^N\sum_{k=1}^T|u_{it}e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{x}_s)||u_{lt}e_{lk}1\{|e_{lk}|>\alpha_T\}\phi_j(\mathbf{x}_k)| \\
= & \sum_{i=1}^N\sum_{s\neq t}\sum_{l=1}^N E|u_{it}u_{lt}e_{lt}1\{|e_{lt}|>\alpha_T\}\phi_j(\mathbf{x}_t)|E|e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{x}_s)| \\
& +\sum_{i=1}^N\sum_{l=1}^N E|u_{it}e_{it}1\{|e_{it}|>\alpha_T\}u_{lt}e_{lt}1\{|e_{lt}|>\alpha_T\}|\phi_j(\mathbf{x}_t)^2 \\
& +E\sum_{i=1}^N\sum_{l=1}^N\sum_{k\neq t}|u_{it}u_{lt}e_{it}1\{|e_{it}|>\alpha_T\}\phi_j(\mathbf{x}_t)|E|e_{lk}1\{|e_{lk}|>\alpha_T\}\phi_j(\mathbf{x}_k)| \\
& +\sum_{i=1}^N\sum_{s\neq t}\sum_{l=1}^N E|u_{it}u_{lt}|E|e_{ls}1\{|e_{ls}|>\alpha_T\}e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{x}_s)^2|
\end{aligned}
$$

$$
\begin{aligned}
& +\sum_{i=1}^{N}\sum_{s\neq t}\sum_{l=1}^{N}\sum_{k\neq s,t} E|u_{it}u_{lt}|E|e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{x}_s)|E|e_{lk}1\{|e_{lk}|>\alpha_T\}\phi_j(\mathbf{x}_k)| \\
:= & \sum_{i=1}^{5} a_i.
\end{aligned}
$$

We look at $a_i, i = 1, ..., 5$ respectively. By Holder's inequality, and the assumption that $E\{E(u_{it}^4|\mathbf{x})\}^v < \infty$, and by repeatedly using Cauchy-Schwarz inequality, uniformly in t,

$$
\begin{aligned}
a_1 &= \sum_{i=1}^{N}\sum_{s\neq t}\sum_{l=1}^{N} E|u_{it}u_{lt}e_{lt}1\{|e_{lt}|>\alpha_T\}\phi_j(\mathbf{x}_t)|E|e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{x}_s)| \\
&\leq \sum_{i=1}^{N}\sum_{s\neq t}\sum_{l=1}^{N}(Ee_{lt}^2 1\{|e_{lt}|>\alpha_T\})^{1/2}(Eu_{it}^2u_{lt}^2\phi_j(\mathbf{x}_t)^2)^{1/2}\sup_{\mathbf{x}} E(|e_{is}|1\{|e_{is}|>\alpha_T\}|\mathbf{x})E|\phi_j(\mathbf{x}_s)| \\
&\leq CTN^2\max_i\{E[Eu_{it}^4|\mathbf{x}_t]^v\}^{1/(2v)}\alpha_T^{-(\zeta_2+1)-\zeta_2/2} = O(TN^2\alpha_T^{-(\zeta_2+1)-\zeta_2/2}) \\
a_2 &= \sum_{i=1}^{N}\sum_{l=1}^{N} E|u_{it}e_{it}1\{|e_{it}|>\alpha_T\}u_{lt}e_{lt}1\{|e_{lt}|>\alpha_T\}|\phi_j(\mathbf{x}_t)^2 \\
&\leq \sum_{i=1}^{N}\sum_{l=1}^{N} E|u_{it}\boldsymbol{\lambda}_i'\boldsymbol{\gamma}_t 1\{|e_{it}|>\alpha_T\}u_{lt}\boldsymbol{\lambda}_l'\boldsymbol{\gamma}_t 1\{|e_{lt}|>\alpha_T\}|\phi_j(\mathbf{x}_t)^2 \\
&\quad+\sum_{i=1}^{N}\sum_{l=1}^{N} E|u_{it}\boldsymbol{\lambda}_i'\boldsymbol{\gamma}_t 1\{|e_{it}|>\alpha_T\}u_{lt}^2 1\{|e_{lt}|>\alpha_T\}|\phi_j(\mathbf{x}_t)^2 \\
&\quad+\sum_{i=1}^{N}\sum_{l=1}^{N} E|u_{it}^2 1\{|e_{it}|>\alpha_T\}u_{lt}\boldsymbol{\lambda}_l'\boldsymbol{\gamma}_t 1\{|e_{lt}|>\alpha_T\}|\phi_j(\mathbf{x}_t)^2 \\
&\quad+\sum_{i=1}^{N}\sum_{l=1}^{N} E|u_{it}^2 1\{|e_{it}|>\alpha_T\}u_{lt}^2 1\{|e_{lt}|>\alpha_T\}|\phi_j(\mathbf{x}_t)^2 \\
&\leq C\sum_{i=1}^{N}\sum_{l=1}^{N}\max_i(Eu_{it}^4)^{1/2}(E\{E\|\boldsymbol{\gamma}_t\|^4|\mathbf{x}_t\}^v)^{1/(2v)} + C\sum_{i=1}^{N}\sum_{l=1}^{N}[E(u_{it}u_{lt}^2)^{4/3}]^{3/4}(E\|\boldsymbol{\gamma}_t\|^4\phi_j(\mathbf{x}_t)^8)^{1/4} \\
&\quad+C\sum_{i=1}^{N}\sum_{l=1}^{N}\{E[E(u_{it}^2u_{lt}^2|\mathbf{x}_t)]^v\}^{1/v} = O(N^2)
\end{aligned}
$$

$$
\begin{aligned}
a_3 &= E\sum_{i=1}^N\sum_{l=1}^N\sum_{k\neq t}|u_{it}u_{lt}e_{it}1\{|e_{it}|>\alpha_T\}\phi_j(\mathbf{x}_t)|E|e_{lk}1\{|e_{lk}|>\alpha_T\}\phi_j(\mathbf{x}_k)| \\
&\leq \sum_{i=1}^N\sum_{l=1}^N\sum_{k\neq t}(E|u_{it}u_{lt}\phi_j(\mathbf{x}_t)|^2)^{1/2}(Ee_{it}^2 1\{|e_{it}|>\alpha_T\})^{1/2}E|e_{lk}1\{|e_{lk}|>\alpha_T\}\phi_j(\mathbf{x}_k)| \\
&\leq TC\sum_{i=1}^N\sum_{l=1}^N\{E[E(u_{it}^2u_{lt}^2|\mathbf{x}_t)]^v\}^{1/2v}\alpha_T^{-\zeta_2/2-(\zeta_2+1)} = O(N^2T\alpha_T^{-\zeta_2/2-(\zeta_2+1)}) \\
a_4 &= \sum_{i=1}^N\sum_{s\neq t}\sum_{l=1}^N E|u_{it}u_{lt}|E|e_{ls}1\{|e_{ls}|>\alpha_T\}e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{x}_s)^2| = O(TN^2\alpha_T^{-\zeta_2}) \\
a_5 &= \sum_{i=1}^N\sum_{s\neq t}\sum_{l=1}^N\sum_{k\neq s,t} E|u_{it}u_{lt}|E|e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{x}_s)|E|e_{lk}1\{|e_{lk}|>\alpha_T\}\phi_j(\mathbf{x}_k)| \\
&= O(N^2T^2\alpha_T^{-2(\zeta_2+1)}).
\end{aligned}
$$

Therefore, uniformly in $t\leq T$,

$$
E(\sum_{i=1}^N\sum_{s=1}^T|u_{it}e_{is}1\{|e_{is}|>\alpha_T\}\phi_j(\mathbf{x}_s)|)^2 = O(TN^2\alpha_T^{-(\zeta_2+1)-\zeta_2/2}+N^2+TN^2\alpha_T^{-\zeta_2}+N^2T^2\alpha_T^{-2(\zeta_2+1)}).
$$

Consequently, (note that $JN^2\alpha_T^{-\zeta_2}+JN^2T\alpha_T^{-2(\zeta_2+1)}\geq JN^2\sqrt{T}\alpha_T^{-(\zeta_2+1)-\zeta_2/2}$)

$$
E\sum_{t=1}^T\|\mathbf{u}_t'\mathbf{M}_\alpha\|^2 = O(JN\|\operatorname{cov}(\boldsymbol{\gamma}_s)\|+JN^2/T+J+JN^2\alpha_T^{-\zeta_2}+JN^2T\alpha_T^{-2(\zeta_2+1)}).
$$

Lemma B.10.

$$
\sum_{s=1}^T\sum_{t=1}^T|\mathbf{u}_s'\mathbf{R}_t|^2 \quad = \quad O_P(N^2J^4\log N\log J+J^{2-2\eta}T^2N+T\alpha_T^{-2(\zeta_1-1)}N^2J^4\log N).
$$

Proof. Recall that $R_{it} = R_{1,it} + R_{2,it} + R_{3,it}$, where

$$\begin{aligned} R_{1,it} &:= \frac{1}{T}\sum_{k=1}^{T}\alpha_T[\dot{\rho}(\alpha_T^{-1}e_{ik,\alpha}) - \dot{\rho}(\alpha_T^{-1}e_{ik})]\Phi(\mathbf{x}_k)'\mathbf{A}\Phi(\mathbf{x}_t) \\ R_{2,it} &:= \Phi(\mathbf{x}_t)'(\mathbf{R}_{i,b} + \mathbf{b}_{i,\alpha} - \mathbf{b}_i), \qquad R_{3,it} := -z_{it}. \end{aligned}$$

In addition, recall $e_{it} = e_{it,\alpha} + \Delta_{it,\alpha}$, where $\Delta_{it,\alpha} = (\mathbf{b}_{i,\alpha} - \mathbf{b}_i)'\Phi(\mathbf{x}_t) - z_{it}$. For notational simplicity, we also write $H_{kt} := \Phi(\mathbf{x}_k)'\mathbf{A}\Phi(\mathbf{x}_t)$.

$$\sum_{s=1}^{T}\sum_{t=1}^{T}|\mathbf{u}_s'\mathbf{R}_t|^2 \leq C\sum_{s=1}^{T}\sum_{t=1}^{T}(\sum_{i=1}^{N}u_{is}R_{1,it})^2 + C\sum_{s=1}^{T}\sum_{t=1}^{T}(\sum_{i=1}^{N}u_{is}R_{2,it})^2 + C\sum_{s=1}^{T}\sum_{t=1}^{T}(\sum_{i=1}^{N}u_{is}R_{3,it})^2.$$

We look at these terms respectively.

bounding the first term

$$\begin{aligned} & \sum_{s=1}^{T}\sum_{t=1}^{T}E(\sum_{i=1}^{N}u_{is}R_{1,it})^2 = \sum_{s=1}^{T}\sum_{t=1}^{T}E\{\sum_{i=1}^{N}u_{is}\frac{1}{T}\sum_{k=1}^{T}\alpha_T[\dot{\rho}(\alpha_T^{-1}e_{ik,\alpha}) - \dot{\rho}(\alpha_T^{-1}e_{ik})]\Phi(\mathbf{x}_k)'\mathbf{A}\Phi(\mathbf{x}_t)\}^2 \\ \leq\ & C\sum_{s=1}^{T}\sum_{t=1}^{T}E\{\sum_{i=1}^{N}u_{is}\frac{1}{T}\sum_{k=1}^{T}\Delta_{ik,\alpha}H_{kt}\}^2 \\ & + C\sum_{s=1}^{T}\sum_{t=1}^{T}E\{\sum_{i=1}^{N}|u_{is}|\frac{1}{T}\sum_{k=1}^{T}|\Delta_{ik,\alpha}|1\{|e_{ik}| > \alpha_T \text{ or } |e_{it,\alpha}| > \alpha_T\}|H_{kt}|\}^2 \\ :=\ & Ca_1 + Ca_2. \end{aligned}$$

For notational simplicity, let $I_{i,kt} := 1\{|e_{ik}| > \alpha_T \text{ or } |e_{it,\alpha}| > \alpha_T\}$.

$$\begin{aligned} a_1 &= \sum_{s=1}^{T}\sum_{t=1}^{T}E\{\sum_{i=1}^{N}u_{is}\frac{1}{T}\sum_{k=1}^{T}\Delta_{ik,\alpha}H_{kt}\}^2 \\ &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{T}\sum_{k=1}^{T}E(Eu_{is}u_{js}|\{\mathbf{x}_l\}_{l\leq T})\Delta_{ik,\alpha}H_{kt}\Delta_{jm,\alpha}H_{mt} \end{aligned}$$

$$
\begin{aligned}
&\leq \sup_{\mathbf{x}} \sum_{i=1}^{N} |E(u_{is}u_{js}|\mathbf{x}_s)| \frac{1}{T^2} \sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{m=1}^{T}\sum_{k=1}^{T} E \max_i |\Delta_{ik,\alpha}||H_{kt}\Delta_{jm,\alpha}H_{mt}| \\
&\leq CT^2 N(\alpha_T^{-(k-1)}\sqrt{J} + J^{-\eta})^2 J^2. \\
a_2 &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T} E\{\sum_{i=1}^{N}\sum_{k=1}^{T} |u_{is}\Delta_{ik,\alpha}I_{i,kt}H_{kt}|\}^2 = \frac{1}{T^2}\sum_{s,t,k,l\leq T}\sum_{i=1}^{N}\sum_{j=1}^{N} E|u_{is}u_{js}\Delta_{ik,\alpha}H_{kt}\Delta_{jl,\alpha}I_{i,kt}I_{j,lt}H_{lt}| \\
&\leq \frac{1}{T^2}\sum_{t,k,l\leq T}\sum_{s=t \text{ or } k \text{ or } l}\sum_{i=1}^{N}\sum_{j=1}^{N}(E(u_{is}u_{js}\Delta_{ik,\alpha}H_{kt}\Delta_{jl,\alpha}H_{lt})^2)^{1/2}(EI_{i,kt}I_{j,lt})^{1/2} \\
&\quad + \frac{1}{T^2}\sum_{t,k,l\leq T}\sum_{s\neq t,k,l}\sum_{i=1}^{N}\sum_{j=1}^{N} E|u_{is}u_{js}|(E(\Delta_{ik,\alpha}H_{kt}H_{lt}\Delta_{jl,\alpha})^2)^{1/2}(EI_{i,kt}I_{j,lt})^{1/2} \\
&\leq \frac{CN(N+T)}{T^2}(\alpha_T^{-(k-1)}\sqrt{J} + J^{-\eta})^2 J^2 \sum_{t,k,l\leq T}(EI_{i,kt}I_{j,lt})^{1/2} \\
&\leq CJ^2 NT(N+T)(\alpha_T^{-(k-1)}\sqrt{J} + J^{-\eta})^2 \alpha_T^{-(\zeta_2+2)/2}
\end{aligned}
$$

where the last inequality is due to, uniformly in i, j,

$$
\begin{aligned}
P(|e_{it,\alpha}| > \alpha_T) &\leq P(|e_{it}| > 3\alpha_T/4) + P(\|\Phi(\mathbf{x}_t)\| > C\alpha_T^k) \leq C\alpha_T^{-(\zeta_2+2)}, \\
\sum_{t,k,l\leq T}(EI_{i,kt}I_{j,lt})^{1/2} &\leq CT^3\alpha_T^{-(\zeta_2+2)/2}.
\end{aligned}
$$

Therefore, $\sum_{s=1}^{T}\sum_{t=1}^{T}(\sum_{i=1}^{N} u_{is}R_{1,it})^2 = O_P((\alpha_T^{-(k-1)}\sqrt{J} + J^{-\eta})^2 J^2 TN(T + N\alpha_T^{-(\zeta_2+2)/2}))$.

bounding the second term

By Lemma B.3, $\max_{i\leq N}\|\mathbf{R}_{i,b}\|^2 = O_P(\alpha_T^{-2(\zeta_1-1)} + \alpha_T^{-2(\zeta_2+2)} + \frac{\log J}{T})\frac{J^3\log N}{T}$. Hence

$$
\begin{aligned}
&\sum_{s=1}^{T}\sum_{t=1}^{T}(\sum_{i=1}^{N} u_{is}R_{2,it})^2 = \sum_{s=1}^{T}\sum_{t=1}^{T}(\sum_{i=1}^{N} u_{is}\Phi(\mathbf{x}_t)'(\mathbf{R}_{i,b} + \mathbf{b}_{i,\alpha} - \mathbf{b}_i))^2 \\
&\leq 2\sum_{s=1}^{T}\sum_{t=1}^{T}(\sum_{i=1}^{N} u_{is}\Phi(\mathbf{x}_t)'\mathbf{R}_{i,b})^2 + 2\sum_{s=1}^{T}\sum_{t=1}^{T}(\sum_{i=1}^{N} u_{is}\Phi(\mathbf{x}_t)'(\mathbf{b}_{i,\alpha} - \mathbf{b}_i))^2 := a_1 + a_2, \text{ say} \\
a_1 &\leq 2\sum_{s=1}^{T}\sum_{t=1}^{T}(\sum_{i=1}^{N} \|u_{is}\Phi(\mathbf{x}_t)\|)^2 \max_i \|\mathbf{R}_{i,b}\|^2 = O_P(T^2N^2J)\max_i\|\mathbf{R}_{i,b}\|^2
\end{aligned}
$$

$$
\begin{aligned}
&= O_P((T\alpha_T^{-2(\zeta_1-1)} + T\alpha_T^{-2(\zeta_2+2)} + \log J)N^2J^4\log N). \\
E|a_2| &= 2\sum_{s=1}^T\sum_{i=1}^N\sum_{j=1}^N(\mathbf{b}_{i,\alpha}-\mathbf{b}_i)'Eu_{is}u_{js}\sum_{t=1}^T\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'(\mathbf{b}_{j,\alpha}-\mathbf{b}_j) \\
&\leq 2\sup_{\mathbf{x}}\max_i\sum_{j=1}^N|(Eu_{is}u_{js}|\mathbf{x}_s)|\max_i\|\mathbf{b}_{i,\alpha}-\mathbf{b}_i\|^2\sum_{s=1}^T\sum_{i=1}^N E\|\sum_{t=1}^T\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)\| \\
&\leq O(T^2\max_i\|\mathbf{b}_{i,\alpha}-\mathbf{b}_i\|^2N) = O(T^2N\alpha_T^{-2(k-1)}).
\end{aligned}
$$

Therefore, $\sum_{s=1}^T\sum_{t=1}^T(\sum_{i=1}^N u_{is}R_{2,it})^2 = O_P((T\alpha_T^{-2(\zeta_1-1)} + T\alpha_T^{-2(\zeta_2+2)} + \log J)N^2J^4\log N + T^2N\alpha_T^{-2(k-1)})$.

bounding the third term

$$
E\sum_{s=1}^T\sum_{t=1}^T(\sum_{i=1}^N u_{is}R_{3,it})^2 = \sum_{s=1}^T\sum_{t=1}^T(\sum_{i=1}^N u_{is}z_{it})^2 = \sum_{s=1}^T\sum_{t=1}^T\sum_{i=1}^N\sum_{j=1}^N Eu_{is}u_{js}z_{it}z_{jt} = O(NT^2J^{-2\eta}).
$$

Hence the result follows.

Lemma B.11.

$$
\frac{1}{T}\sum_{t=1}^T\|\mathbf{D}_{t2}\|^2 = O_P(\frac{1}{N^3}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2 + \frac{1}{N}\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2 + \frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2 + \frac{1}{N^2T^2}\sum_{s=1}^T\sum_{t=1}^T|\mathbf{u}_s'\mathbf{R}_t|^2).
$$

Proof. First of all, note that $\max_i\sum_j|Eu_{is}u_{js}| < \infty$, hence

$$
E\frac{1}{T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{\Lambda}\|^2 = \sum_{j=1}^K E(\mathbf{u}_s'\boldsymbol{\lambda}_j)^2 = O(N).
$$

In addition, $\frac{1}{T}\sum_{t=1}^T\|\mathbf{D}_{t2}\|^2 = \frac{1}{T}\sum_{t=1}^T\|\frac{1}{N}(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})'\mathbf{u}_t\|^2 \leq C\sum_{i=1}^8\frac{1}{N^2T}\sum_{t=1}^T\|\mathbf{u}_t'\mathbf{B}_i\|^2$.

$$
\frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{B}_1\|^2 = \frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{\Lambda}\frac{1}{TN}\sum_{t=1}^T E(\mathbf{f}_t|\mathbf{x}_t)\Phi(\mathbf{x}_t)'\mathbf{A}\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2
$$

$$
\begin{aligned}
&\leq \frac{1}{N^4T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{\Lambda}\|^2\|\frac{1}{T}\sum_{t=1}^T E(\mathbf{f}_t|\mathbf{x}_t)\Phi(\mathbf{x}_t)'\|^2\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2 O_P(1) \\
&= O_P(1)\frac{1}{N^4T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{\Lambda}\|^2\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2 = O_P(\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2/N^3),
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{B}_2\|^2 &= \frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{\Lambda}\frac{1}{TN}\sum_{t=1}^T E(\mathbf{f}_t|\mathbf{x}_t)\mathbf{R}_t'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2 \\
&\leq \frac{1}{N^3T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{\Lambda}\|^2\|\frac{1}{T}\sum_{t=1}^T E(\mathbf{f}_t|\mathbf{x}_t)\mathbf{R}_t'\|^2 O_P(1) = O_P(\frac{1}{N}\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2), \\
\frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{B}_3\|^2 &= \frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{M}_\alpha\mathbf{A}\frac{1}{TN}\sum_{t=1}^T\Phi(\mathbf{x}_t)E(\mathbf{f}_t|\mathbf{x}_t)'\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2 \\
&\leq O_P(1)\frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2\|\frac{1}{T}\sum_{t=1}^T\Phi(\mathbf{x}_t)E(\mathbf{f}_t|\mathbf{x}_t)'\|^2 = \frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2 O_P(1), \\
\frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{B}_4\|^2 &= \frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{M}_\alpha\mathbf{A}\frac{1}{TN}\sum_{t=1}^T\Phi(\mathbf{x}_t)\Phi(\mathbf{x}_t)'\mathbf{A}\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2 \\
&\leq \frac{1}{N^4T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2 O_P(1) \\
\frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{B}_5\|^2 &= \frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{M}_\alpha\mathbf{A}\frac{1}{TN}\sum_{t=1}^T\Phi(\mathbf{x}_t)\mathbf{R}_t'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2 \\
&\leq O_P(\frac{1}{N^3T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2\|\frac{1}{T}\sum_{t=1}^T\Phi(\mathbf{x}_t)\mathbf{R}_t'\|^2) = O_P(\frac{J}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2) \\
\frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{B}_6\|^2 &= \frac{1}{N^2T}\sum_{s=1}^T\|\frac{1}{TN}\sum_{t=1}^T\mathbf{u}_s'\mathbf{R}_t E(\mathbf{f}_t|\mathbf{x}_t)'\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2 \\
&\leq \leq O_P(\frac{1}{N^2T^3}\sum_{s=1}^T\|\sum_{t=1}^T\mathbf{u}_s'\mathbf{R}_t E(\mathbf{f}_t|\mathbf{x}_t)'\|^2) \leq O_P(\frac{1}{N^2T^2}\sum_{s=1}^T\sum_{t=1}^T|\mathbf{u}_s'\mathbf{R}_t|^2) \\
\frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{B}_7\|^2 &= \frac{1}{N^2T}\sum_{s=1}^T\|\frac{1}{TN}\sum_{t=1}^T\mathbf{u}_s'\mathbf{R}_t\Phi(\mathbf{x}_t)'\mathbf{A}\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2 \leq O_P(\frac{J}{N^4T^2}\sum_{s=1}^T\sum_{t=1}^T|\mathbf{u}_s'\mathbf{R}_t|^2\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2) \\
\frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\mathbf{B}_8\|^2 &= \frac{1}{N^2T}\sum_{s=1}^T\|\mathbf{u}_s'\frac{1}{TN}\sum_{t=1}^T\mathbf{R}_t\mathbf{R}_t'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2 \leq O_P(1)\frac{1}{N^2T^2}\sum_{s=1}^T\sum_{t=1}^T|\mathbf{u}_s'\mathbf{R}_t|^2\max_i\frac{1}{T}\sum_{t=1}^T R_{it}^2.
\end{aligned}
$$

Summarizing, we have

$$\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{D}_{t2}\|^2 = O_P(\frac{1}{N^3}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2+\frac{1}{N}\max_i\frac{1}{T}\sum_{t=1}^{T}R_{it}^2+\frac{1}{N^2T}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2+\frac{1}{N^2T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}|\mathbf{u}_s'\mathbf{R}_t|^2).$$

C Proof of Theorem 4.1

The proof of the limiting distribution of S under the null is divided into two major steps.

step 1: Asymptotic expansion: under H_0,

$$S=\frac{1}{TN}\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}\mathbf{H}\widehat{\mathbf{W}}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t+o_P(T^{-1/2}).$$

step 2: The effect of estimating $\mathbf{\Sigma}_u$ is first-order negligible:

$$\frac{1}{TN}\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}\mathbf{H}\widehat{\mathbf{W}}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t=\frac{1}{TN}\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}(\frac{1}{N}\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{u}_t+o_P(T^{-1/2}).$$

The result then follows from the asymptotic normality of the first term on the right hand side. We shall prove this using Lindeberg's central limit theorem.

We achieve each step in the following subsections.

C.1 Step 1 asymptotic expansion of S

Proposition C.1. *Under H_0,*

$$S=\frac{1}{TN}\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}\mathbf{H}\widehat{\mathbf{W}}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t+o_P(T^{-1/2})$$

Proof. Since $\|\widehat{\mathbf{W}}\|\leq\max_i\widehat{\sigma}_{ii}=O_P(1)$, it follows from (B.3) that it suffices to prove under H_0, $\frac{N}{T}\sum_{t=1}^{T}\mathbf{D}_{ti}'\widehat{\mathbf{W}}\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t=o_P(T^{-1/2})$, and $\frac{N}{T}\sum_{t=1}^{T}\|\mathbf{D}_{ti}\|^2=o_P(T^{-1/2}), i=2,3,4$.

By the proof of Propositions B.5, B.3, Lemmas B.6, B.11 and that $\mathbf{D}_{t3}=\mathbf{C}_{t3}$,$\mathbf{D}_{t4}=\mathbf{C}_{t4}$,

$$\begin{aligned}\frac{N}{T}\sum_{t=1}^{T}\|\mathbf{D}_{t4}\|^2 &= O_P(\max_i\frac{N}{T}\sum_{t=1}^{T}R_{it}^2)=O_P(NJ^{1-2\eta}+\frac{NJ^3\log N}{\alpha_T^{2(\zeta_1-1)}T}+\frac{NJ^3\log N\log J}{T^2})=o_P(\frac{1}{\sqrt{T}})\\ \frac{N}{T}\sum_{t=1}^{T}\|\mathbf{D}_{t3}\|^2 &= O_P(\frac{1}{N}\|\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\|^2)=O_P(\frac{J}{T}+\frac{NJ\alpha_T^{-\zeta_2}}{T}+J^{2-4\eta}+\alpha_T^{-2(\zeta_1-1)}\frac{J^3\log N}{TJ^{2\eta-1}})=o_P(\frac{1}{\sqrt{T}})\end{aligned}$$

The last equality holds so long as $N\sqrt{T}=o(J^{2\eta-1})$, $NJ^4\log N\log J=o(T^{3/2})$, $\zeta_1>2$.

By Lemma B.10,

$$\frac{N}{T}\sum_{t=1}^{T}\|\mathbf{D}_{t2}\|^2=O_P(\frac{1}{N^2}\|\mathbf{M}_\alpha'\widehat{\mathbf{\Lambda}}\|^2+\max_i\frac{1}{T}\sum_{t=1}^{T}R_{it}^2+\frac{1}{NT}\sum_{s=1}^{T}\|\mathbf{u}_s'\mathbf{M}_\alpha\|^2+\frac{1}{NT^2}\sum_{s=1}^{T}\sum_{t=1}^{T}|\mathbf{u}_s'\mathbf{R}_t|^2)=o_P(\frac{1}{\sqrt{T}}).$$

The proof of $\frac{N}{T}\sum_{t=1}^{T}\mathbf{D}_{ti}'\widehat{\mathbf{W}}\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t=o_P(T^{-1/2})$ is given in Lemmas C.1 and C.2 below. It then leads to the desired result.

Lemma C.1. *Suppose* $(N+T)J^{1-2\eta}=o(1)$. *Then*

$$\mathrm{tr}(\frac{N}{T}\sum_{t=1}^{T}\mathbf{D}_{t2}'\widehat{\mathbf{W}}\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t)=o_P(T^{-1/2})$$

Proof. It suffics to prove $\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{D}_{t2}\mathbf{u}_t'\mathbf{\Lambda}\|^2=\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})\|^2=o_P(\frac{1}{T})$. To this end, we need to decompose $\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H}=\sum_{i=1}^{8}\mathbf{B}_i$ again as in (B.2). Every term can be bounded using established bounds except for the term involving $\mathbf{B}_3$. More specifically, for $i\neq 3$, we use $\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\mathbf{B}_i\|^2\leq\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\|_F^2\|\mathbf{B}_i\|^2$. On the other hand, $\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\|_F^2\leq 2\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\mathbf{\Lambda}'\mathbf{\Sigma}_u\|_F^2+2\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\mathbf{\Lambda}'(\mathbf{u}_t\mathbf{u}_t'-\mathbf{\Sigma}_u)\|_F^2$. The first term is $O_P(\frac{1}{N})$. As for the second tern,

$$E\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\mathbf{\Lambda}'(\mathbf{u}_t\mathbf{u}_t'-\mathbf{\Sigma}_u)\|_F^2=\sum_{k=1}^{K}\sum_{i=1}^{N}E(\frac{1}{TN}\sum_{t=1}^{T}\sum_{j=1}^{N}\lambda_{jk}(u_{jt}u_{it}-Eu_{jt}u_{it}))^2$$

$$
\begin{aligned}
&= \frac{1}{T^2N^2}\sum_{k=1}^K\sum_{i=1}^N\sum_{t=1}^T \mathrm{var}(\sum_{j=1}^N \lambda_{jk}(u_{jt}u_{it}-Eu_{jt}u_{it})) = \frac{1}{T^2N^2}\sum_{k=1}^K\sum_{i=1}^N\sum_{t=1}^T\sum_{j=1}^N\sum_{l=1}^N \lambda_{jk}\lambda_{lk}\,\mathrm{cov}(u_{jt}u_{it},u_{lt}u_{it})\\
&= O(\frac{1}{T}) + \frac{1}{T^2N^2}\sum_{k=1}^K\sum_{i=1}^N\sum_{t=1}^T\sum_{j=1}^N\sum_{l\neq i,t}\lambda_{jk}\lambda_{lk}E(u_{jt}u_{it}-\sigma_{ij})u_{it}u_{lt} = O(\frac{1}{T}).
\end{aligned}
$$

Hence $\|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\mathbf{B}_i\|^2 \le O_P(\frac{1}{T}+\frac{1}{N})\|\mathbf{B}_i\|^2 = o(\frac{1}{T})$, for $i\neq 3$, where the last equality holds by straightforward verifying $(\frac{T}{N}+1)\|\mathbf{B}_i\|^2 = o(1)$ using Lemma B.5, assuming $(N+T)J^{1-2\eta} = o(1)$.

To allow $N/T\to\infty$, the term involving $\mathbf{B}_3$ requires a different and sharper bound:

$$
\begin{aligned}
&\|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\mathbf{B}_3\|^2 = \|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\mathbf{M}_\alpha\frac{1}{TN}\sum_{s=1}^T\mathbf{A}\Phi(\mathbf{x}_s)E(\mathbf{f}_s|\mathbf{x}_s)'\mathbf{\Lambda}'\widehat{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^{-1}\|^2\\
\le\;& \|\frac{1}{TN}\sum_{t=1}^T\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\mathbf{M}_\alpha\|^2O_P(1) = O_P(1)\|\frac{1}{TN}\sum_t\mathbf{\Lambda}'\mathbf{u}_t\sum_{i=1}^N u_{it}\frac{1}{T}\sum_{s=1}^T\alpha_T\dot\rho(\alpha_T^{-1}e_{is})\Phi(\mathbf{x}_s)\|^2\\
\le\;& O_P(1)\|\frac{1}{TN}\sum_t\mathbf{\Lambda}'\mathbf{u}_t\sum_{i=1}^N u_{it}\frac{1}{T}\sum_{s=1}^T u_{is}1\{|u_{is}|<\alpha_T\}\Phi(\mathbf{x}_s)\|^2\\
&+O_P(1)\|\frac{1}{TN}\sum_t\mathbf{\Lambda}'\mathbf{u}_t\sum_{i=1}^N u_{it}\frac{1}{T}\sum_{s=1}^T\alpha_T\dot\rho(\alpha_T^{-1}u_{is})1\{|u_{is}|>\alpha_T\}\Phi(\mathbf{x}_s)\|^2\\
\le\;& O_P(1)\|\frac{1}{T^2N}\sum_t\sum_{s=1}^T\sum_{i=1}^N\mathbf{\Lambda}'\mathbf{u}_tu_{it}u_{is}\Phi(\mathbf{x}_s)\|^2\\
&+O_P(1)\left(\frac{1}{T^2N}\sum_t\sum_{s=1}^T\sum_{i=1}^N\|\mathbf{\Lambda}'\mathbf{u}_t\||u_{it}||u_{is}|1\{|u_{is}|>\alpha_T\}\|\Phi(\mathbf{x}_s)\|\right)^2, \qquad\text{(C.1)}
\end{aligned}
$$

where we used the fact that under H_0, $e_{is}=u_{is}$. We respectively bound the two terms on the right hand side.

First term in (C.1) Note that

$$
E\|\frac{1}{T^2N}\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^N\mathbf{\Lambda}'\mathbf{u}_tu_{it}u_{is}\Phi(\mathbf{x}_s)\|^2 = \frac{1}{T^4N^2}\sum_{l=1}^J\sum_{k=1}^K E(\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^N\sum_{j=1}^N\lambda_{jk}u_{jt}u_{it}u_{is}\phi_l(\mathbf{x}_s))^2.
$$

We then expand the term on the right hand side, which leads to many additive terms in the expansion. Using the assumption of serial independence to analyze each term, we conclude that $E\|\frac{1}{T^2N}\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^N \mathbf{\Lambda}'\mathbf{u}_t u_{it}u_{is}\Phi(\mathbf{x}_s)\|^2 = O_P(\frac{J}{TN}+\frac{J}{T^2}+\frac{JN}{T^3})$. We omit the lengthy details.

Second term in (C.1) As for the second term, first note that under H_0, $u_{it}=e_{it}$. So Lemma B.2 implies $(E|u_{is}|1\{|u_{is}|>\alpha_T\}|\mathbf{x}_t=\mathbf{x})\leq C\alpha_T^{-\zeta_2-1}$. On the other hand, by assumption, for some $C>0$, $\sup_{\mathbf{x}} E(u_{it}^4 1\{|u_{it}|>\alpha_T\}|\mathbf{x}_t=\mathbf{x})\leq \alpha_T^{-\zeta_5}C$, $E\|\mathbf{\Lambda}'\mathbf{u}_t\|^2=O(N)$. Hence

$$
\begin{aligned}
& \frac{1}{T^2N}\sum_t\sum_{s=1}^T\sum_{i=1}^N E\|\mathbf{\Lambda}'\mathbf{u}_t\||u_{it}||u_{is}|1\{|u_{is}|>\alpha_T\}\|\Phi(\mathbf{x}_s)\| \\
= & \frac{1}{T^2N}\sum_t\sum_{i=1}^N E\|\mathbf{\Lambda}'\mathbf{u}_t\|u_{it}^2 1\{|u_{it}|>\alpha_T\}\|\Phi(\mathbf{x}_t)\| \\
& +\frac{1}{T^2N}\sum_t\sum_{s\neq t}\sum_{i=1}^N E\|\mathbf{\Lambda}'\mathbf{u}_t\||u_{it}|E|u_{is}|1\{|u_{is}|>\alpha_T\}\|\Phi(\mathbf{x}_s)\| \\
\leq & \frac{1}{T^2N}\sum_t\sum_{i=1}^N (E\|\mathbf{\Lambda}'\mathbf{u}_t\|^2)^{1/2}(E\|\Phi(\mathbf{x}_t)\|^2)^{1/2}\sup_{\mathbf{x}}(Eu_{it}^4 1\{|u_{it}|>\alpha_T\}|\mathbf{x}_t=\mathbf{x})^{1/2} \\
& +\frac{1}{T^2N}\sum_t\sum_{s\neq t}\sum_{i=1}^N (E\|\mathbf{\Lambda}'\mathbf{u}_t\|^2)^{1/2}(Eu_{it}^2)^{1/2}E\|\Phi(\mathbf{x}_s)\|\sup_{\mathbf{x}}(E|u_{is}|1\{|u_{is}|>\alpha_T\}|\mathbf{x}_t=\mathbf{x}) \\
= & O_P(\frac{\sqrt{JN}}{T}\alpha_T^{-\zeta_5/2}+\sqrt{NJ}\alpha_T^{-\zeta_2-1}).
\end{aligned}
$$

It then implies the second term in (C.1) is $O_P(\frac{JN}{T^2}\alpha_T^{-\zeta_5}+NJ\alpha_T^{-2\zeta_2-2})$.

Thus, when $\zeta_5\geq 1$, $T=o(J^{2\eta-1})$

$$
\|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\mathbf{B}_3\|^2 = O_P(\frac{J}{TN}+\frac{J}{T^2}+\frac{JN}{T^3}+\frac{JN}{T^2}\alpha_T^{-\zeta_5}+NJ\alpha_T^{-2\zeta_2-2}) = o_P(\frac{1}{T}).
$$

As a result, $\|\frac{1}{T}\sum_{t=1}^T\mathbf{D}_{t2}\mathbf{u}_t'\mathbf{\Lambda}\|^2 = \|\frac{1}{T}\sum_{t=1}^T\frac{1}{N}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'(\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\mathbf{H})\|^2 = o_P(\frac{1}{T})$.

Lemma C.2. *For $i = 3, 4$,*

$$\mathrm{tr}(\frac{N}{T}\sum_{t=1}^{T}\mathbf{D}_{ti}'\widehat{\mathbf{W}}\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t) = o_P(T^{-1/2})$$

Proof. Again, it suffices to verify $\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{D}_{ti}\mathbf{u}_t'\mathbf{\Lambda}\|^2 = o_P(\frac{1}{T})$ for $i = 3, 4$. Note that $\|\frac{1}{T}\sum_{t=1}^{T}\Phi(\mathbf{x}_t)\mathbf{u}_t'\mathbf{\Lambda}\|^2 = O_P(\frac{NJ}{T})$. Then by definition,

$$\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{D}_{t3}\mathbf{u}_t'\mathbf{\Lambda}\|^2 = \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\mathbf{A}\Phi(\mathbf{x}_t)\mathbf{u}_t'\mathbf{\Lambda}\|^2 \leq O_P(\frac{1}{N^2})\|\widehat{\mathbf{\Lambda}}'\mathbf{M}_\alpha\|^2\|\frac{1}{T}\sum_{t=1}^{T}\Phi(\mathbf{x}_t)\mathbf{u}_t'\mathbf{\Lambda}\|^2 = o_P(\frac{1}{T}).$$

On the other hand, recall the definition $R_{it} := R_{1,it} + R_{2,it} + R_{3,it}$, where

$$\begin{aligned} R_{1,it} &:= \Phi(\mathbf{x}_t)'\mathbf{A}\frac{1}{T}\sum_{s=1}^{T}\alpha_T[\dot\rho(\alpha_T^{-1}e_{is,\alpha}) - \dot\rho(\alpha_T^{-1}e_{is})]\Phi(\mathbf{x}_s) \\ R_{2,it} &:= \Phi(\mathbf{x}_t)'(\mathbf{R}_{i,b} + \mathbf{b}_{i,\alpha} - \mathbf{b}_i), \qquad R_{3,it} := -z_{it}. \end{aligned}$$

Thus it can be verified similarly that

$$\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{D}_{t4}\mathbf{u}_t'\mathbf{\Lambda}\|^2 = \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\widehat{\mathbf{\Lambda}}'\mathbf{R}_t\mathbf{u}_t'\mathbf{\Lambda}\|^2 = O_P(\frac{1}{NT^2})\sum_{i=1}^{N}\|\sum_{t=1}^{T}R_{it}\mathbf{u}_t'\mathbf{\Lambda}\|^2 = o_P(\frac{1}{T}).$$

The verification is very similar as before, and is omitted here.

C.2 Step 2 Completion of the proof

We now aim to show $\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Sigma}}_u\widehat{\mathbf{\Lambda}}/N = \mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}/N + o_P(T^{-1/2})$. Once this is done, it then follows from the facts that $\mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}/N = O_P(1)$ and $(\mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}/N)^{-1} = O_P(1)$,

$$(\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Sigma}}_u\widehat{\mathbf{\Lambda}}/N)^{-1} = (\mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}/N)^{-1} + o_P(T^{-1/2}).$$

As a result, by Proposition C.1,

$$
\begin{aligned}
S &= \frac{1}{TN}\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}\mathbf{H}(\mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}/N)^{-1}\mathbf{H}'\mathbf{\Lambda}'\mathbf{u}_t + o_P(T^{-1/2}) \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{u}_t + o_P(T^{-1/2}).
\end{aligned}
$$

Hence

$$
\frac{TS - TK}{\sqrt{2TK}} = \frac{\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{u}_t - TK}{\sqrt{2TK}} + o_P(1) \to^d \mathcal{N}(0,1).
$$

To finish the proof, we now show two claims:

(1)

$$
\frac{\sum_{t=1}^{T}\mathbf{u}_t'\mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{u}_t - TK}{\sqrt{2TK}} \to^d \mathcal{N}(0,1).
$$

(2) $\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Sigma}}_u\widehat{\mathbf{\Lambda}}/N = \mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}/N + o_P(T^{-1/2})$.

Proof of (1) We define $X_t = \mathbf{u}_t'\mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{u}_t$ and $s_T^2 = \sum_{t=1}^{T}\mathrm{var}(X_t)$. Then $E(X_t) = \operatorname{tr} E((\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{u}_t\mathbf{u}_t'\mathbf{\Lambda}) = K$. Also by Assumption 4.1, $s_T^2/T \to 2K$, hence we have $E\frac{1}{T}\sum_{t=1}^{T}(X_t - K)^2 < \infty$ for all large N, T. For any $\epsilon > 0$, by the dominated convergence theorem, for all large N, T,

$$
\frac{1}{T}\sum_{t=1}^{T}E(X_t - K)^2 1\{|X_t - K| > \epsilon s_T\} \leq \frac{1}{T}\sum_{t=1}^{T}E(X_t - K)^2 1\{|X_t - K| > \epsilon\sqrt{KT}\} = o(1).
$$

This then implies the Lindeberg condition, $\frac{1}{s_T^2}\sum_{t=1}^{T}E(X_t - K)^2 1\{|X_t - K| > \epsilon s_T\} = o(1)$. Hence by the Lindeberg central limit theorem,

$$
\frac{\sum_t X_t - TK}{s_T} \to^d \mathcal{N}(0,1).
$$

The result then follows since $s_T^2/T \to 2K$.

Proof of (2) By the triangular inequality,

$$\|\frac{1}{N}\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Sigma}}_u\widehat{\mathbf{\Lambda}} - \frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}\| \leq \|\frac{1}{N}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\widehat{\mathbf{\Lambda}}\| + \|\frac{1}{N}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})'\mathbf{\Sigma}_u(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\|$$
$$+\|\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\| + \|\frac{1}{N}\mathbf{H}'\mathbf{\Lambda}'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\mathbf{\Lambda}\mathbf{H}\| + 2\|\frac{1}{N}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}\|.$$

Using the established bounds for $\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}\|$ in Theorem 3.1, it is straightforward to verify $\|\frac{1}{N}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})'\mathbf{\Sigma}_u(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})\| = o_P(T^{-1/2})$. Other terms require sharper bounds yet to be established. These are given in Proposition C.2 below. It then follows that $\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Sigma}}_u\widehat{\mathbf{\Lambda}}/N = \mathbf{H}'\mathbf{\Lambda}'\mathbf{\Sigma}_u\mathbf{\Lambda}\mathbf{H}/N + o_P(T^{-1/2})$. This completes the proof.

Proposition C.2. *(i)* $\frac{1}{N}\mathbf{\Lambda}'\mathbf{\Sigma}_u(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}) = o_P(T^{-1/2})$*;*
(ii) $\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\mathbf{\Lambda} = o_P(T^{-1/2})$*;*
(iii) $\|\frac{1}{N}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\mathbf{G}\| = o_P(T^{-1/2})$*, for either* $\mathbf{G} = \mathbf{\Lambda}$ *or* $\mathbf{G} = \widehat{\mathbf{\Lambda}}$*.*

Proof. Define $\widetilde{\mathbf{\Lambda}} = \mathbf{\Sigma}_u\mathbf{\Lambda}$. Note that we cannot simply bound these terms by $\frac{1}{N}\|\widetilde{\mathbf{\Lambda}}\|\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}\|$ or $\frac{1}{N}\|\mathbf{\Lambda}\|^2\|\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u\|$, as these bounds are too crude to achieve the desired rate of convergence when $N/T \to \infty$. More careful analysis is called for.

(i) Proving $\frac{1}{N}\widetilde{\mathbf{\Lambda}}'(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}) = o_P(T^{-1/2})$ is exactly the same as that of Lemma B.8. Note that replacing $\mathbf{\Lambda}$ with $\widetilde{\mathbf{\Lambda}}$ does not introduce any complications as $\mathbf{\Sigma}_u$ is a diagonal matrix. Hence the proof is omitted here to avoid repetitions.

(ii) For any $k, l \leq K$, the (k, l) element of $\frac{1}{N}\mathbf{\Lambda}'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\mathbf{\Lambda}$ is given by

$$\frac{1}{N}\sum_{i=1}^{N}\lambda_{ik}\lambda_{il}(\widehat{\sigma}_{ii} - \sigma_{ii}) = \frac{1}{N}\frac{1}{T}\sum_{t}\sum_{i=1}^{N}\lambda_{ik}\lambda_{il}(u_{it}^2 - Eu_{it}^2) + \frac{1}{N}\sum_{i=1}^{N}\lambda_{ik}\lambda_{il}\frac{1}{T}\sum_{t}(\widehat{u}_{it}^2 - u_{it}^2)$$

As for the first term,

$$E|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{ik}\lambda_{il}(u_{it}^2 - \sigma_{ii})| \leq \left[E\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{ik}\lambda_{il}(u_{it}^2 - \sigma_{ii})\right)^2\right]^{1/2}$$

$$
\begin{aligned}
&= \left[\frac{1}{N^2T^2}\sum_{i=1}^N\sum_{t=1}^T\sum_{j=1}^N\sum_{s=1}^T \lambda_{ik}\lambda_{il}\lambda_{jk}\lambda_{jl}\operatorname{cov}(u_{it}^2, u_{js}^2)\right]^{1/2} \\
&= \left[\frac{1}{N^2T^2}\sum_{i=1}^N\sum_{t=1}^T \lambda_{ik}^2\lambda_{il}^2\operatorname{var}(u_{it}^2)\right]^{1/2} = o(\frac{1}{\sqrt{T}}).
\end{aligned}
$$

As for the second term, we have

$$
\begin{aligned}
& |\frac{1}{TN}\sum_{i=1}^N\sum_{t=1}^T \lambda_{ik}\lambda_{il}(\widehat{u}_{it}^2 - u_{it}^2)| \leq 2|\frac{1}{TN}\sum_{i=1}^N\sum_{t=1}^T \lambda_{ik}\lambda_{il}(\widehat{u}_{it} - u_{it})u_{it}| + |\frac{1}{TN}\sum_{i=1}^N\sum_{t=1}^T \lambda_{ik}\lambda_{il}(\widehat{u}_{it} - u_{it})^2| \\
\leq\ & O_P(1)|\frac{1}{T}\sum_{t=1}^T(\widehat{\mathbf{f}}_t - \mathbf{f}_t)'\frac{1}{N}\sum_{i=1}^N \lambda_{ik}\lambda_{il}u_{it}\boldsymbol{\lambda}_i| + O_P(1)|\frac{1}{TN}\sum_{i=1}^N \lambda_{ik}\lambda_{il}(\widehat{\boldsymbol{\lambda}}_i - \mathbf{H}'\boldsymbol{\lambda}_i)'\sum_{t=1}^T u_{it}\mathbf{f}_t| \\
& + O_P(1)|\frac{1}{TN}\sum_{i=1}^N\sum_{t=1}^T \lambda_{ik}\lambda_{il}u_{it}(\widehat{\boldsymbol{\lambda}}_i - \mathbf{H}'\boldsymbol{\lambda}_i)'(\widehat{\mathbf{f}}_t - \mathbf{H}^{-1}\mathbf{f}_t)| + O(1)\max_i|\frac{1}{T}\sum_{t=1}^T(\widehat{u}_{it} - u_{it})^2| \\
\leq\ & O_P(1)(\frac{1}{T}\sum_t\|\widehat{\mathbf{f}}_t - \mathbf{H}^{-1}\mathbf{f}_t\|^2)^{1/2}(\frac{1}{T}\sum_t\|\frac{1}{N}\sum_{i=1}^N \lambda_{ik}\lambda_{il}u_{it}\boldsymbol{\lambda}_i\|^2)^{1/2} \\
& + O_P(1)(\frac{1}{N}\sum_i\|\lambda_{ik}\lambda_{il}(\widehat{\boldsymbol{\lambda}}_i - \mathbf{H}'\boldsymbol{\lambda}_i)\|^2)^{1/2}(\frac{1}{N}\sum_i\|\frac{1}{T}\sum_t u_{it}\mathbf{f}_t\|^2)^{1/2} + O(1)\max_i\frac{1}{T}\sum_{t=1}^T(\widehat{u}_{it} - u_{it})^2 \\
& + O_P(1)(\frac{1}{T}\sum_t\|\widehat{\mathbf{f}}_t - \mathbf{H}^{-1}\mathbf{f}_t\|^2)^{1/2}(\frac{1}{TN}\sum_{it} u_{it}^2)^{1/2}(\frac{1}{N}\sum_i\|\widehat{\boldsymbol{\lambda}}_i - \mathbf{H}'\boldsymbol{\lambda}_i\|^2)^{1/2}.
\end{aligned}
$$

Note that $\frac{1}{T}\sum_t\|\widehat{\mathbf{f}}_t - \mathbf{H}^{-1}\mathbf{f}_t\|^2 = O_P(\psi_{NT}^2)$, $\max_i\frac{1}{T}\sum_t(\widehat{u}_{it} - u_{it})^2 = O_P(\psi_{NT}^2 + \frac{J\log N}{T})$ by Lemma C.3. Also, $\frac{1}{N}\sum_{i=1}^N\|\widehat{\boldsymbol{\lambda}}_i - \mathbf{H}'\boldsymbol{\lambda}_i\|^2 = O_P\left(\frac{J}{T} + \frac{1}{J^{2\eta-1}} + (\frac{\log N}{T})^{\zeta_1}J^3\right)$ by Theorem 3.1. In addition,

$$
\begin{aligned}
& E\frac{1}{T}\sum_t\|\frac{1}{N}\sum_{i=1}^N \lambda_{ik}\lambda_{il}u_{it}\boldsymbol{\lambda}_i\|^2 = \sum_{m=1}^K E(\frac{1}{N}\sum_i \lambda_{ik}\lambda_{il}\lambda_{im}u_{it})^2 \\
=\ & \sum_{m=1}^K\frac{1}{N^2}\sum_i\sum_j \lambda_{ik}\lambda_{il}\lambda_{im}\lambda_{jk}\lambda_{jl}\lambda_{jm}Eu_{it}u_{jt} = \sum_{m=1}^K\frac{1}{N^2}\sum_i \lambda_{ik}^2\lambda_{il}^2\lambda_{im}^2Eu_{it}^2 = O(\frac{1}{N}),
\end{aligned}
$$

$$E\frac{1}{N}\sum_i \|\frac{1}{T}\sum_t u_{it}\mathbf{f}_t\|^2 = \frac{1}{N}\sum_i\sum_k E(\frac{1}{T}\sum_t u_{it}f_{kt})^2 = \frac{1}{N}\sum_i\sum_k \frac{1}{T^2}\sum_t Eu_{it}^2 f_{kt}^2 = O(\frac{1}{T}).$$

Hence it is straightforward to verify that $|\frac{1}{TN}\sum_{i=1}^N\sum_{t=1}^T \lambda_{ik}\lambda_{il}(\widehat{u}_{it}^2 - u_{it}^2)| = o_P(T^{-1/2})$ so long as $T = o(N^2)$, $T = o(J^{2\eta-1}N)$, $J^4\log N = o(NT)$.

(iii) Let G_{ik} denote the (i,k) element of $\mathbf{G}$, and let δ_{ik} denote the (i,k) element of $\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H}$. Since $\max_i \|\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i\| = o_P(1)$, we have $\max_{ik}|G_{ik}| = O_P(1)$, regardless of $\mathbf{G} = \mathbf{\Lambda}$ or $\mathbf{G} = \widehat{\mathbf{\Lambda}}$. Then the (l,k) element of the $K\times K$ matrix $\frac{1}{N}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\mathbf{H})'(\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u)\mathbf{G}$ is bounded by

$$|\frac{1}{N}\sum_{i=1}^N \delta_{il}G_{ik}\frac{1}{T}\sum_t(\widehat{u}_{it}^2 - \sigma_{ii})| \leq \max_{ilk}|\delta_{il}G_{ik}|\frac{1}{NT}\sum_{i=1}^N|\sum_{t=1}^T(\widehat{u}_{it}^2 - u_{it}^2) + (u_{it}^2 - \sigma_{ii})|.$$

On one hand, by Lemma C.3,

$$\max_{ilk}|\delta_{il}G_{ik}|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T|\widehat{u}_{it}^2 - u_{it}^2| = O_P(\psi_{NT} + \sqrt{\frac{J\log N}{T}})\max_i\|\widehat{\boldsymbol{\lambda}}_i - \mathbf{H}'\boldsymbol{\lambda}_i\| = o_P(\frac{1}{\sqrt{T}}).$$

On the other hand, $E|\sum_{t=1}^T(u_{it}^2 - \sigma_{ii})| \leq \mathrm{var}(\sum_{t=1}^T(u_{it}^2 - \sigma_{ii}))^{1/2} = (T\,\mathrm{var}(u_{it}^2))^{1/2} = O(T^{1/2})$. Hence

$$\max_{ilk}|\delta_{il}G_{ik}|\frac{1}{NT}\sum_{i=1}^N|\sum_{t=1}^T(u_{it}^2 - \sigma_{ii})| = O_P(\frac{1}{\sqrt{T}})\max_i\|\widehat{\boldsymbol{\lambda}}_i - \mathbf{H}'\boldsymbol{\lambda}_i\| = o_P(\frac{1}{\sqrt{T}}).$$

Lemma C.3. *Define*

$$\psi_{NT} = \frac{1}{J^{\eta-1/2}} + \frac{1}{\sqrt{N}} + \frac{J^2(\log N\log J)^{1/2}}{T} + (\frac{\log N}{T})^{\zeta_1/2}J^2.$$

Under H_0, when $N = O(T^2)$,
(i) $\frac{1}{T}\sum_{t=1}^T\|\widehat{\mathbf{f}}_t - \mathbf{H}^{-1}\mathbf{f}_t\|^2 = O_P\left(\psi_{NT}^2\right)$.
(ii) $\max_i\frac{1}{T}\sum_t(\widehat{u}_{it} - u_{it})^2 = O_P(\psi_{NT}^2 + \frac{J\log N}{T})$.

(iii) $\frac{1}{NT}\sum_i\sum_t|\widehat{u}_{it}^2-u_{it}^2|=O_P(\psi_{NT}+\sqrt{\frac{J\log N}{T}})$.

Proof. (i) By Theorem 3.2, under H_0,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{f}}_t-\mathbf{H}^{-1}\mathbf{f}_t\|^2 &\leq 2\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{g}}(\mathbf{x}_t)-\mathbf{H}^{-1}\mathbf{g}(\mathbf{x}_t)\|^2+2\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}^{-1}\boldsymbol{\gamma}_t\|^2\\ &= O_P\left(\frac{1}{N}+\frac{J^4\log N\log J}{T^2}+\frac{1}{J^{2\eta-1}}+(\frac{\log N}{T})^{\zeta_1}J^4\right). \qquad \text{(C.2)}\end{aligned}$$

(ii) Uniformly in i, by Theorem 3.1,

$$\frac{1}{T}\sum_t(\widehat{u}_{it}-u_{it})^2\leq C\frac{1}{T}\sum_t\|\widehat{\boldsymbol{\lambda}}_i-\boldsymbol{\lambda}_i\|^2\|\widehat{\mathbf{f}}_t\|^2+C\frac{1}{T}\sum_t\|\boldsymbol{\lambda}_i\|^2\|\widehat{\mathbf{f}}_t-\mathbf{f}_t\|^2=O_P(\psi_{NT}^2).$$

(iii) We have, using $|a^2-b^2|\leq|a-b||a+b|$ and the Cauchy-Schwarz inequality,

$$\begin{aligned}&(\frac{1}{NT}\sum_i\sum_t|\widehat{u}_{it}^2-u_{it}^2|)^2\leq\frac{1}{NT}\sum_{it}(\widehat{u}_{it}-u_{it})^2\frac{1}{NT}\sum_{it}(\widehat{u}_{it}+u_{it})^2\\ \leq\;& \max_i\frac{1}{T}\sum_t(\widehat{u}_{it}-u_{it})^2\frac{1}{NT}\sum_{it}[2(\widehat{u}_{it}-u_{it})^2+4u_{it}^2]\\ \leq\;& 2(\max_i\frac{1}{T}\sum_t(\widehat{u}_{it}-u_{it})^2)^2+4\max_i\frac{1}{T}\sum_t(\widehat{u}_{it}-u_{it})^2\frac{1}{NT}\sum_{it}u_{it}^2=O_P(\max_i\frac{1}{T}\sum_t(\widehat{u}_{it}-u_{it})^2)\\ =\;& O_P(\psi_{NT}^2+\frac{J\log N}{T}).\end{aligned}$$

D Choice of the tuning parameter

As mentioned earlier, throughout this paper, we take

$$\alpha_T=C\sqrt{\frac{T}{\log(NJ)}} \qquad \text{(D.1)}$$

for a constant C chosen by the cross-validation. The Huber-estimator is biased for estimating the mean coefficient, whose population counterpart is

$$\mathbf{b}_{i,\alpha} := \arg\min_{\mathbf{b}\in\mathbb{R}^J} E\rho\left(\frac{y_{it} - \Phi(\mathbf{x}_t)'\mathbf{b}}{\alpha_T}\right),$$

As α_T increases, the Huber loss behaves like a quadratic loss. In fact, we show (Proposition B.1) that for $\mathbf{b}_i := \arg\min_{\mathbf{b}\in\mathbb{R}^J} E[y_{it} - \mathbf{b}'\Phi(\mathbf{x}_t)]^2$,

$$\max_{i\leq N} \|\mathbf{b}_{i,\alpha} - \mathbf{b}_i\| = O(\alpha_T^{-(\zeta_2+1)+\epsilon})$$

for an arbitrarily smal $\epsilon > 0$, where ζ_2 is defined in Assumption 3.1. Hence the bias decreases as α_T grows as expected. On the other hand, we shall investigate the uniform convergence (in $i = 1, ..., N$) of

$$\max_{i\leq N} \|\frac{1}{T}\sum_{t=1}^{T} \dot{\rho}(\alpha_T^{-1} e_{it})\Phi(\mathbf{x}_t)\|, \tag{D.2}$$

which is the leading term in the Bahadur expansion of the Huber-estimator. It turns out that α_T cannot grow faster than $O(\sqrt{\frac{T}{\log(NJ)}})$ in order to guard for robustness and to have a sharp uniform convergence, where J is the number of sieve basis. Hence the choice (D.1) leads to the asymptotically least-biased robust estimation.

E Testing proxy factors for S&P 500 returns

In this section, we use the test in Section 4 to test the explanatory power of the observable proxies for the true factors using S&P 500 returns. For each given group of observable proxies, we set the number of common factors K equals the number of observable proxies. The sieve basis is chosen as the additive Fourier basis with $J = 5$. We set the tuning parameter $\alpha_T = C\sqrt{\frac{T}{\log(NJ)}}$ with constant C selected by the 5-fold cross validation.

We calculate the daily excess returns for the stocks in S&P 500 index that have complete daily closing prices from January 2005 to December 2013. The data, collected from CRSP, contains 393 stocks with a time span of 2265 trading days. As we know, two stylized features of S&P 500 daily returns are asymmetry and heavy tails. The proxy factors ($\mathbf{x}_t$) are chosen to be the Fama-French 3/5 factors and the sector SPDR ETF's, which are intended to track the 9 largest S&P sectors. The detailed descriptions of sector SPDR ETF's are listed in Table E.1. In this study, we consider three groups of proxy factors with increasing information: (1) Fama-French 3 factors (FF3); (2) Fama-French 5 factors (FF5); and (3) Fama-French 5 factors plus 9 sector SPDR ETF's (FF5+ETF9).

We apply moving window tests with the window size (T) equals three months or six months. The testing window moves one trading day forward per test. Within each testing window, we calculate the standardized test statistic S for three groups of proxy factors. The plots of S under various scenarios are reported in Figure E.1.

According to Figure E.1, under all scenarios, the null hypothesis ($H_0 : \text{cov}(\boldsymbol{\gamma}_t) = 0$) is rejected as S is always larger than the critical value 1.96. This suggests a strong evidence that the proxy factors can not fully explain the estimated common factors. Under all window sizes, a larger group of proxy factors tends to yield smaller statistics, demonstrating stronger explanatory power for estimated common factors. Also, we find the test statistics increase while the window size increases.

Table E.1: Sector SPDR ETF's (data available from Yahoo finance)

Code	Representative sector
XLE	Energy
XLB	Materials
XLI	Industrials
XLY	Consumer discretionary
XLP	Consumer staples
XLV	Health care
XLF	Financial
XLK	Information technology
XLU	Utilities

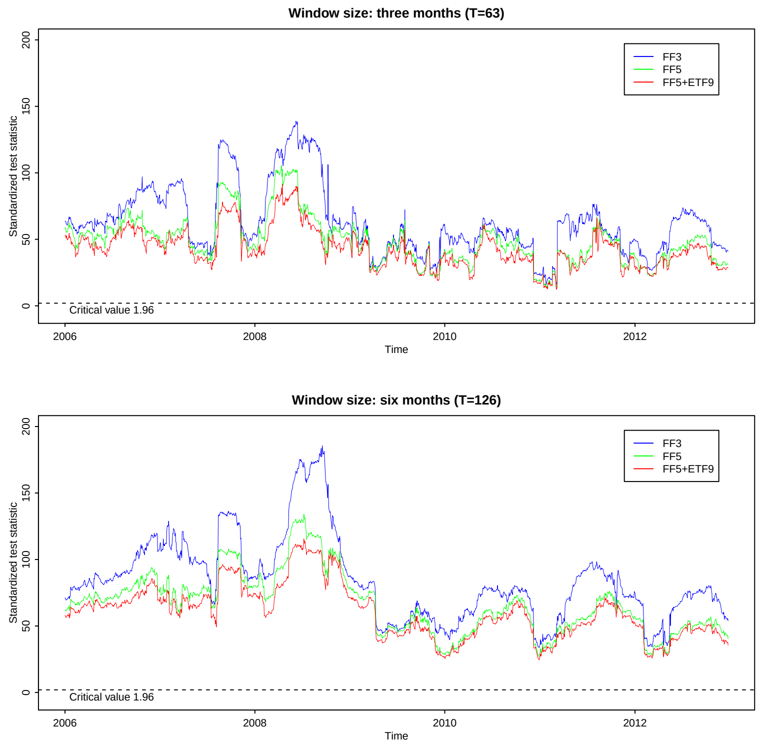

Figure E.1: S&P 500 daily returns: plots for standardized test statistic S for various window sizes. The dotted line is critical value 1.96.

F Additional simulation under serial dependence

In this section, we numerically compare RPR, Sieve-LS and PCA under serial dependences. We consider the similar simulation settings as in Section 5.1 except both $\mathbf{x}_t$ and $\boldsymbol{\gamma}_t$ are generated from a stationary VAR(1) model as follows

$$\mathbf{x}_t = \mathbf{\Pi}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\gamma}_t = \mathbf{\Pi}\boldsymbol{\gamma}_{t-1} + \eta_t, \quad t = 1, \cdots, T,$$

with $\mathbf{x}_0 = \mathbf{0}$ and $\boldsymbol{\gamma}_0 = \mathbf{0}$. The (i,j)th entry of $\mathbf{\Pi}$ is set to be 0.5 when $i = j$ and $0.4^{|i-j|}$ when $i \neq j$. In addition, $\boldsymbol{\varepsilon}_t$ and η_t are drawn form i.i.d. $N(\mathbf{0}, \mathbf{I})$.

As the factors and loading may be estimated up to a rotation matrix, the canonical correlations between the parameter and its estimator can be used to measure the estimation accuracy (Bai, 2003). The performance among RPR, Sieve-LS and PCA are presented in Table F.1 below. Our numerical findings for the independent data continue to hold for serially dependent data: both RPR and Sieve-LS outperform PCA when $\mathbf{x}_t$ and $\mathbf{f}_t$ are serially correlated. RPR gives the best performance when the error distributions are heavy-tailed.

Table F.1: **Dependent data**: Median of canonical correlations of the estimated loadings/factors and the true ones when $N = 50, T = 100$: the larger the better

			Model (I)			Model (II)		
	$\mathbf{u}_t$	σ	RPR	Sieve-LS	PCA	RPR	Sieve-LS	PCA
Loadings	$N(0,\ 8)$	0.01	0.92	0.92	0.86	0.90	0.90	0.86
		0.3	0.92	0.92	0.92	0.89	0.89	0.90
		1.0	0.90	0.90	0.98	0.84	0.85	0.96
	LogN	0.01	0.63	0.31	0.25	0.65	0.33	0.26
		0.3	0.62	0.30	0.26	0.62	0.30	0.27
		1.0	0.60	0.28	0.29	0.60	0.28	0.29
Factors	$N(0,\ 8)$	0.01	0.91	0.92	0.89	0.89	0.89	0.85
		0.3	0.90	0.91	0.92	0.88	0.88	0.88
		1.0	0.88	0.88	0.97	0.85	0.86	0.97
	LogN	0.01	0.65	0.41	0.30	0.63	0.36	0.25
		0.3	0.63	0.40	0.31	0.58	0.31	0.27
		1.0	0.61	0.39	0.33	0.56	0.30	0.31

References

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

Bühlmann, P. and van de Geer, S. (2011). *Statistics for high-dimensional data, methods, theory and applications.* The first edition ed. Springer, New York.