INFERENCE FOR HETEROGENEOUS EFFECTS USING LOW-RANK ESTIMATIONS

VICTOR CHERNOZHUKOV, CHRISTIAN HANSEN, YUAN LIAO, AND YINCHU ZHU

Abstract. We study a panel data model with general heterogeneous effects, where slopes are allowed to be varying across both individuals and times. The key assumption for dimension reduction is that the heterogeneous slopes can be expressed as a factor structure so that the high-dimensional slope matrix is of low-rank, so can be estimated using low-rank regularized regression. Our paper makes an important theoretical contribution on the "post-SVD inference". Formally, we show that the post-SVD inference can be conducted via three steps: (1) apply the nuclear-norm penalized estimation; (2) extract eigenvectors from the estimated low-rank matrices, and (3) run least squares to iteratively estimate the individual and time effect components in the slope matrix. To properly control for the effect of the penalized low-rank estimation, we argue that this procedure should be embedded with "partial out the mean structure" and "sample splitting". The resulting estimators are asymptotically normal and admit valid inferences. In addition, we conduct global homogeneous tests, where under the null, the slopes are either common across individuals, time-invariant, or both. Empirically, we apply the proposed methods to estimate the county-level minimum wage effects on the employment.

Key words: nuclear norm penalization, post-SVD, sample splitting, interactive effects

Date: December 20, 2018.

1. Introduction

This paper studies the estimation and inference about the following panel data model:

$$y_{it} = x_{it}'\theta_{it} + \alpha_i' g_t + u_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T,$$

where x_{it} is a d-dimensional vector of observed covariates; α_i and g_t are unobserved fixed effects. Importantly, the model permits general heterogeneities, in the sense that not only the fixed effects appear in the model interactively, but also the slope θ_{it} is allowed to vary across both i and t. The main dimension reduction assumption employed in this paper is that θ_{it} can be expressed as:

$$\theta_{it} = \lambda_i' f_t,$$

that is, it can be represented by a factor structure, where λ_i is a matrix of loadings and f_t is a vector of factors. Here we allow f_t and g_t to have overlapping components.

Let Θ_r and X_r be the $N \times T$ matrices of the r th component $(\theta_{it,r})$ and $x_{it,r}$; let M, Y, U be the $N \times T$ matrices of $\alpha_i' g_t$, y_{it} and u_{it}. Also let $\odot$ denote the matrix element-wise product. We have the following matrix form:

$$Y = \sum_{r=1}^{d} X_r \odot \Theta_r + M + U.$$

With the factor structure in Θ_r and M, both the slope and fixed effect matrices are of low-rank, whose ranks are at most equal to their associated numbers of factors. This motivates employing the low-rank estimation:

$$\begin{aligned} \min_{\{\Theta_1,...,\Theta_d,M\}} \quad & \|Y - \sum_{r=1}^{d} X_r \odot \Theta_r - M\|_F^2 + P_0(\Theta_1, ..., \Theta_d, M) \\ \text{where} \quad & P_0(\Theta_1, ..., \Theta_d, M) = \sum_{r=1}^{d} \nu_r \|\Theta_r\|_n + \nu_0 \|M\|_n, \end{aligned} \tag{1.1}$$

for some tuning parameters $\nu_0, \nu_1, ..., \nu_d > 0$, where $\|.\|_F$ and $\|.\|_n$ respectively denote the matrix Frobenius norm and nuclear norm. In particular, let $\psi_1(\Theta) \geq ... \geq \psi_{\min\{T,N\}}(\Theta)$ be the sorted singular values of an $N \times T$ matrix Θ, then

$$\|\Theta\|_n := \sum_{i=1}^{\min\{T,N\}} \psi_i(\Theta),$$

which is a convex-relaxation of the rank of Θ, and can be casted using efficient algorithms such as the singular value decomposition (SVD).

While the penalized low-rank estimation (1.1) produces consistent estimators for the low-rank matrices, however, it is subjected to shrinkage biases. As such, this paper makes an important theoretical contribution on the inference for low-rank matrices, after employing the singular value decomposition type regularization, so-called "post-SVD inference". Formally, we show that the post-SVD inference can be conducted via three steps:

(1) apply the low-rank regularized estimation via SVD-based algorithms;
(2) extract eigenvectors from the estimated low-rank matrices, and
(3) run iterative least squares.

More specifically, we extract the eigenvectors of the estimated Θ_r obtained in (1.1), which form as preliminary estimators for λ_i. We then apply least squares to obtain $\widehat{f}_t$ as the estimated f_t, and re-estimate λ_i, still using least squares to obtain $\widehat{\lambda}_i$. This leads to our final estimator for θ_{it} as, for each i, t,

$$\widehat{\theta}_{it} = \widehat{\lambda}_i' \widehat{f}_t.$$

We shall show that the shrinkage bias can be offset from this iterative least squares procedure. The final estimator is asymptotically normally distributed and admits an asymptotically valid inference for θ_{it} at each fixed (i, t).

In addition, we assume that the covariates are generated from the following model:

$$x_{it} = \mu_{it} + e_{it},$$

where e_{it} is a zero-mean process that is serially independent and cross-sectionally weakly dependent. Here μ_{it} is the mean structure of x_{it}. We study two cases. The first case is: $\mu_{it} = \mu_i$, which assumes that x_{it} follows from a simple many-mean model. The second case is $\mu_{it} = l_i' w_t$ which also allows a factor-structure that captures the cross-sectional and serial dependences in x_{it}, with l_i, w_t respectively denoting the loading and factors in x_{it}. On the other hand, to properly control for the effect of the penalized low-rank estimation (1.1), the iterative least squares procedure should be embedded with "partial out the mean structure" and "sample splitting", so that in effect, we are using e_{it} as the regressors to apply the iterative least squares. Section 2.2 explains this in details.

We allow either f_t or λ_i or both to be constant across $i \leq N$ or $t \leq T$. So our model and estimation methods admit homogeneous slopes as special cases, and allow us to test for various homogeneity hypotheses. We consider three *global* homogeneity tests:

Time-invariant homogeneity:

$$H_0^1 : \theta_{it} = \theta_i, \quad \forall (i,t), \text{ for some } \theta_i.$$

Cross-sectionally invariant homogeneity:

$$H_0^2 : \theta_{it} = \theta_t, \quad \forall (i,t), \text{ for some } \theta_t.$$

Pure homogeneity:

$$H_0^3 : \theta_{it} = \theta, \quad \forall (i,t), \text{ for some } \theta.$$

The above are three popular homogeneity assumptions that have been commonly used in empirical studies. In addition, one can also test for *local* homogeneities, such as:

$$H_0^i : \theta_{it} = \theta_i, \quad \forall t, \text{ for some } \theta_i.$$

for a particular individual i; or $H_0^t : \theta_{it} = \theta_t$, $\forall i$, at a particular time.

Employing the nuclear norm penalization has been a popular technique in the statistical literature to achieve low-rank estimations, e.g., Negahban and Wainwright (2011); Recht et al. (2010); Sun and Zhang (2012); Candès and Tao (2010); Koltchinskii et al. (2011). However, there is little study on the element-wise inference for the components. The only work that we are aware of on the related issue is Cai et al. (2016), who proposed a debiased estimator to make inference about the low-rank matrix after applying the nuclear-norm penalization. We note that their debiased procedure does not apply in the model we are considering. The main issue is that under proper conditions, low-rank matrices deduced from a high-dimensional factor structure have "very spiked" eigenvalues, possessing leading eigenvalues that grow at rate $O(NT)^{1/2}$. The presence of these fast-growing eigenvalues calls for distinguished asymptotic behaviors in covariance estimations, eigen-spectral analysis, as well as matrix completions. The related literature can be found in Fan et al. (2013); Wang and Fan (2017).

There is also a growing literature on estimating panel data models with interactive fixed effects, but has been predominated by homogeneous models. Bai

(2009); Moon and Weidner (2015) studied the inference for a global homogeneous slope model where $\theta_{it} = \theta$ for some θ and all (i,t). In addition, Pesaran (2006); Ahn et al. (2013) studied time-invariant homogeneous models in which the slopes are allowed to vary across i. Su et al. (2015) considered nonlinear homogeneous effects where the effect of x_{it} is characterized by a nonparametric function that is time-invariant and common to all individuals. The homogeneity assumption, while simplifies the inference procedure and gains statistical efficiency if correctly specified, would lead to inconsistency and misleading inferences if fails to hold. Our standing point and recommended inference procedure is to start with a general heterogeneous model as we are proposing, and formally conduct homogeneous tests as formalized in this paper. Then conduct inference and estimations using the homogeneous method only if the homogeneous hypothesis is accepted.

The rest of the paper is organized as follows. Section 2 introduces the post-SVD algorithms that define our estimators. It also explains the rationale of using partial-out and sample splitting techniques. Section 3 provides asymptotic inferential theory. Section 4 studies tests of global homogeneity. Section 5 presents simulated results, and finally, Section 6 applies the proposed method to studying the county-level minimum wage effects. All proofs are presented in the appendix.

2. The model

We consider the following model

$$\begin{aligned} y_{it} &= x_{it}'\theta_{it} + \alpha_i' g_t + u_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T. \\ \theta_{it} &= \lambda_i' f_t. \end{aligned}$$

Here we observe (y_{it}, x_{it}). The goal is to make inference about θ_{it}, which has a latent factor structure. The model also allows a general form of fixed effect. As explained in Bai (2009), $\alpha_i' g_t$ can allow fixed effects that are either additive, or interactive, or both. We assume $\dim(\lambda_i) = K_1$, $\dim(\alpha_i) = K_2$, both fixed. For ease of presentation, we focus on the case that x_{it} is a univariate regressor, that is, $\dim(x_{it}) = 1$. It is straightforward to extend to the multivariate case, which is presented in Appendix A.

We allow arbitrary dependences among $\{\lambda_i, f_t : i \leq N, t \leq T\}$, and impose nearly no restrictions on the sequence for λ_i and f_t. In particular, this allows homogeneous models as special cases. For instance, by setting $\lambda_i = \lambda$ and $f_t = f$

for all (i, t) and dimension $\dim(\lambda_i) = \dim(f_t) = 1$, we can allow $\theta_{it} = \theta$ for all (i, t) and a common parameter θ. This then reduces to the homogeneous interactive effect model, studied by Bai (2009); Moon and Weidner (2015). In addition, f_t can also be thought of as a serially independent process, which leads to pure heterogeneities.

2.1. **Nuclear norm penalized estimation and an iterative algorithm.** Let Λ, A respectively be $N \times K_1$ and $N \times K_2$ matrices of (λ_i) and (α_i); let (F, G) be $T \times K_i$ matrices of (f_t, g_t); let (Y, X, U) be $N \times T$ matrices of (y_{it}, x_{it}, u_{it}). Then the matrix form is:

$$Y = AG' + X \odot (\Lambda F') + U.$$

where $\odot$ represents the element-wise product. Further let $\Theta := \Lambda F'$ and $M := AG'$. Note that both of them are $N \times T$ matrices, whose ranks are respectively (K_1, K_2). We let $N, T \to \infty$ but K_1, K_2 be fixed constants. Thus both are low-rank matrices. Motivated by this, we shall estimate (M, Θ) using the following penalized nuclear-norm optimization:

$$\begin{aligned} (\widetilde{\Theta}, \widetilde{M}) &= \min_{\Theta, M} F(\Theta, M), \\ F(\Theta, M) &:= \|Y - M - X \odot \Theta\|_F^2 + \nu_2 \|M\|_n + \nu_1 \|\Theta\|_n, \end{aligned} \tag{2.1}$$

for some tuning parameters $\nu_2, \nu_1 > 0$.

The computation of (2.1) can be carried out iteratively using *singular value thresholding estimations*. For a fixed matrix H, let $UDV' = H$ be its singular value decomposition (SVD). Define the singular value thresholding operator $S_\lambda(H) = UD_\lambda V'$, where D_λ is defined by replacing the diagonal entry D_{ii} of D by $\max\{D_{ii} - \lambda, 0\}$. Given Θ, solving for M in (2.1) leads to the following closed form solution: let $Z_\Theta = Y - X \odot \Theta$,

$$S_{\nu_2/2}(Z_\Theta) = \arg\min_M \|Z_\Theta - M\|_F^2 + \nu_2 \|M\|_n,$$

which applies the singular value thresholding operator on Z_Θ, with tuning $\nu_2/2$. On the other hand, given M, let $Z_M = Y - M$. Then the solution of Θ to (2.1) is given by:

$$\Theta_M := \arg\min_\Theta \|Z_M - X \odot \Theta\|_F^2 + \nu_1 \|\Theta\|_n,$$

which satisfies the following KKT condition, for any $\tau > 0$, (Ma et al., 2011)

$$\Theta_M = S_{\tau\nu_1/2}(\Theta_M - \tau X \odot (X \odot \Theta_M - Z_M)).$$

As such, we employ the following algorithm to iteratively solve for $\widetilde{M}$ and $\widetilde{\Theta}$ as the global solution to (2.1).

Algorithm 2.1. Compute the nuclear-norm penalized regression as follows.

Step 1: Fix the "step size" $\tau \in (0, 1/\max_{it} x_{it}^2)$. Initialize Θ_0, M_0 and set $k = 0$.

Step 2: Let

$$\begin{aligned}
\Theta_{k+1} &= S_{\tau\nu_1/2}(\Theta_k - \tau X \odot (X \odot \Theta_k - Y + M_k)),\\
M_{k+1} &= S_{\nu_2/2}(Y - X \odot \Theta_{k+1}).
\end{aligned}$$

Set k to $k+1$.

Step 3: Repeat step 2 until convergence.

As for the convergence property of the algorithm, note that given M, updating Θ_M is a standard gradient descent procedure (e.g., Beck and Teboulle (2009)), and therefore standard convergence analysis for the objective function $F(\Theta, M)$ applies. As we shall show below, the evaluated objective function $F(\Theta_{k+1}, M_{k+1})$ is monotonically decreasing, whose rate of convergence is $O(k^{-1})$.

Proposition 2.1. *Let $(\widetilde{\Theta}, \widetilde{M})$ be a global minimum for $F(\Theta, M)$. Then for any $\tau \in (0, 1/\max_{it} x_{it}^2)$, and any initial Θ_0, M_0, we have:*

$$F(\Theta_{k+1}, M_{k+1}) \leq F(\Theta_{k+1}, M_k) \leq F(\Theta_k, M_k),$$

for each $k \geq 0$. In addition, for all $k \geq 1$,

$$F(\Theta_{k+1}, M_{k+1}) - F(\widetilde{\Theta}, \widetilde{M}) \leq \frac{1}{k\tau}\|\Theta_1 - \widetilde{\Theta}\|_F^2. \tag{2.2}$$

Proposition 2.1 shows that the algorithm converges from an arbitrary initial value. The upper bound in (2.2) depends on the initial values through the accuracy of the first iteration $\Theta_1 - \widetilde{\Theta}$. Not surprisingly, the upper bound does not depend on $M_1 - \widetilde{M}$, because given Θ_1, minimizing with respect to M has a one-step closed form solution, whose accuracy is completely determined by Θ_1. In addition, the largest possible value for τ to ensure the convergence is $1/\max_{it} x_{it}^2$. In practice, we set $\tau = (1-\epsilon)/\max_{it} x_{it}^2$ for some small $\epsilon > 0$.

2.2. **Post-SVD inference.** The estimated Θ from the nuclear-norm penalized regression, however, cannot be used directly for inference because they are subjected to shrinkage biases. Our debiased method is based on iterative least squares, which iteratively estimates f_t and λ_i to control for the effect of shrinkage biases. We proceed in the following stages to obtain the final estimator of θ_{it}.

Stage 1: obtain estimated (M, Θ) from the nuclear norm optimization.

Stage 2: extract estimated loadings $(\widetilde{\alpha}_i, \widetilde{\lambda}_i)$'s from the singular vectors of the estimated (M, Θ).

Stage 3: iteratively estimate f_t and λ_i using least squares. More specifically:

(i) obtain unbiased estimators $\widehat{f}_t$ for f_t using the estimated loadings.

(ii) obtain unbiased estimators $\widehat{\lambda}_i$ for λ_i using $\widehat{f}_t$.

(iii) obtain the final estimator $\widehat{\theta}_{it} := \widehat{\lambda}_i' \widehat{f}_t$.

As Stage 1 employs a thresholding SVD method, we call our procedure as the "post-SVD inference". We shall show that $\widehat{\theta}_{it}$ is C_{NT}-consistent, where $C_{NT} = \sqrt{\min\{N, T\}}$, and is asymptotically normally distributed, centered around θ_{it}, for each fixed (i, t). Also note that due to the rotation discrepancy, λ_i and f_t are not separately identified, so they can be estimated without biases subject to a rotation transformation. On the other hand, the effect coefficient θ_{it} can be estimated well without rotation discrepancy.

To investigate the role of the dependence structure in x_{it}, we separately study two cases. In the first case, we assume x_{it} is serially independent, generated from the following mean process:

$$x_{it} = \mu_i + e_{it},$$

where e_{it} is a zero-mean process that is serially independent and cross-sectionally weakly dependent. In the second case, we relax this assumption and consider:

$$x_{it} = l_i' w_t + e_{it},$$

where w_t is a vector of (fixed dimensional) latent factors, with l_i as the loadings. Hence x_{it} also admits a factor structure, and w_t, or some of its components, are allowed to overlap with (f_t, g_t).

In both cases, Stage 3 is the essential stage to obtain well-behaved unbiased estimators. In this stage, $(\widetilde{\alpha}_i, \widetilde{\lambda}_i)$ are treated as the estimators for nuisance parameter, whose estimation error involves, among others, a term taking the form:

$$\Delta := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i (\widetilde{\alpha}_i - H\alpha_i)' x_{it},$$

where H is a square rotation matrix. It is however challenging to argue that Δ is negligible, due to two reasons. First, it can be shown that the nuclear-norm penalized estimator has a rate of convergence:

$$\frac{1}{N} \sum_{i=1}^{N} \|\widetilde{\alpha}_i - H\alpha_i\|^2 = O_P(C_{NT}^{-2}).$$

But this rate alone does not ensure $\Delta = o_P(1)$ in the case $N \geq T$. Secondly, when x_{it} admits a factor structure, correlations in both cross-sectional and serial directions introduce additional technical difficulties.

We resolve the above technical difficulties through two procedures: (i) partialling out the mean of x_{it} (or the factor structures in x_{it}) and (ii) sample splitting. When x_{it} admits a factor structure, note that the decomposition $x_{it} = l_i' w_t + e_{it}$ yields

$$\dot{y}_{it} = \alpha_i' g_t + e_{it} \lambda_i' f_t + u_{it}, \qquad \text{where } \dot{y}_{it} = y_{it} - l_i' w_t \theta_{it}. \tag{2.3}$$

We regress the estimated $\dot{y}_{it}$ onto $(\alpha_i, e_{it}\lambda_i)$ to estimate f_t. As such, the effect of estimating α_i becomes

$$\Delta' := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i (\widetilde{\alpha}_i - H\alpha_i)' e_{it},$$

and now e_{it} is a zero-mean process and independent across t. Therefore, by subtracting the estimated $l_i' w_t \theta_{it}$ from y_{it}, we have partialled out the strong dependent components in x_{it}, and essentially work with the following moment conditions for estimating f_t:

$$\mathsf{E} e_{it} (\dot{y}_{it} - \alpha_i' g_t - e_{it} \lambda_i' f_t) = 0, \quad \forall i = 1, ..., N. \tag{2.4}$$

Importantly, this moment condition enjoys the *Neyman's orthogonality*:

$$\frac{\partial}{\partial \alpha_i} \mathsf{E} e_{it} (\dot{y}_{it} - \alpha_i' g_t - e_{it} \lambda_i' f_t) = \mathsf{E} e_{it} g_t = 0.$$

This ensures that the effect of estimating α_i is indeed negligible. In the absence of factor structures in x_{it}, we shall simply partial out the sample mean $\frac{1}{T}\sum_t x_{it}$ from x_{it} to obtain a zero-mean process.

Next, we use the sample splitting technique to continue arguing that $\Delta' = o_P(1)$. For the fixed t, let $\{1, ..., T\} = I \cup I^c \cup \{t\}$ be a random disjoint partition. The cardinality of I is $[(T-1)/2]$ (the nearest integer to $(T-1)/2$). Let

$$D_I = \{(y_{is}, x_{is}) : i \leq N, s \in I\}.$$

We explain the rationale of the this splitting in more detail in Remark 2.1 later. We run the nuclear-norm optimization (2.1) using data D_I and estimate $\widetilde{\lambda}_i$. Because e_{it} is serially independent, and $t \notin I$, $\widetilde{\lambda}_i$ is independent of e_{it}. For example, in the case of cross-sectional independence, and $\dim(\lambda_i) = 1$, conditional on D_I, we have $\mathsf{E}(\Delta'|D_I) = 0$, whose variance is given by

$$\mathsf{Var}(\Delta'|D_I) = \frac{1}{N}\sum_{i=1}^{N} \mathsf{Var}(e_{it})(\widetilde{\alpha}_i - H\alpha_i)(\widetilde{\alpha}_i - H\alpha_i)'\lambda_i^2.$$

This allows us to argue that $\Delta' = o_P(1)$ so that the effect of estimating the nuisance parameters in Stage 3 is negligible.

2.3. **Formal Estimation Algorithm.** Let

$$x_{it} = \mu_{it} + e_{it}, \quad \mathsf{E}e_{it} = 0,$$

We present the estimation algorithm below, which works for both the cases of (i) many mean model, in which $\mu_{it} = \mu_i$, and (ii) factor model for the process of x_{it} in which $\mu_{it} = l_i' w_t$. When x_{it} admits a factor structure, we first apply the standard principal components method (PC) (e.g., Stock and Watson (2002)) to estimate $l_i' w_t$ and obtain $(\widehat{l_i' w_t}, \widehat{e}_{it})$. In both cases, we partial out the "common components" by subtracting the estimated μ_{it} from x_{it}, and essentially work with the model

$$\dot{y}_{it} = \alpha_i' g_t + e_{it}\theta_{it} + u_{it}, \quad \text{where } \dot{y}_{it} = y_{it} - \mu_{it} e_{it}. \tag{2.5}$$

Suppose we are interested in θ_{it} for some fixed (i, t). We propose the following steps to estimate θ_{it}.

Algorithm 2.2. Estimate θ_{it} as follows.

Step 1. Estimate the number of factors. Run nuclear-norm penalized regression:

$$(\widetilde{M}, \widetilde{\Theta}) := \arg\min_{M,\Theta} \|Y - M - X \odot \Theta\|_F^2 + \nu_2 \|M\|_n + \nu_1 \|\Theta\|_n. \tag{2.6}$$

Estimate K_1, K_2 by

$$\widehat{K}_1 = \sum_i 1\{\psi_i(\widetilde{\Theta}) \geq (\nu_2\|\widetilde{\Theta}\|)^{1/2}\}, \quad \widehat{K}_2 = \sum_i 1\{\psi_i(\widetilde{M}) \geq (\nu_1\|\widetilde{M}\|)^{1/2}\}$$

where $\psi_i(.)$ denotes the ith largest singular-value.

Step 2. Estimate the structure $x_{it} = \mu_{it} + e_{it}$.

In the many mean model, let $\widehat{e}_{it} = x_{it} - \frac{1}{T}\sum_{t=1}^T x_{it}$. In the factor model, use the PC estimator to obtain $(\widehat{l_i'w_t}, \widehat{e}_{it})$ for all $i = 1, ..., N, t = 1, .., T$.

Step 3: Sample splitting. Randomly split the sample into $\{1, ..., T\}/\{t\} = I \cup I^c$, so that $|I|_0 = [(T-1)/2]$. Denote by Y_I, X_I as the $N \times |I|_0$ matrices of (y_{is}, x_{is}) for observations at $s \in I$. Estimate the low-rank matrices Θ and M as in (2.6), with (Y, X) replaced with (Y_I, X_I), and obtain $(\widetilde{M}_I, \widetilde{\Theta}_I)$.

Let $\widetilde{\Lambda}_I = (\widetilde{\lambda}_1, ..., \widetilde{\lambda}_N)'$ be the $N \times \widehat{K}_1$ matrix, whose columns are defined as $\sqrt{N}$ times the first $\widehat{K}_1$ eigenvectors of $\widetilde{\Theta}_I\widetilde{\Theta}_I'$. Let $\widetilde{A}_I = (\widetilde{\alpha}_1, ..., \widetilde{\alpha}_N)'$ be the $N \times \widehat{K}_2$ matrix, whose columns are defined as $\sqrt{N}$ times the first $\widehat{K}_2$ eigenvectors of $\widetilde{M}_I\widetilde{M}_I'$.

Step 4. Estimate the "partial-out" components.

Substitute in $(\widetilde{\alpha}_i, \widetilde{\lambda}_i)$, and define

$$(\widetilde{f}_s, \widetilde{g}_s) := \arg\min_{f_s, g_s} \sum_{i=1}^N (y_{is} - \widetilde{\alpha}_i' g_s - x_{is}\widetilde{\lambda}_i' f_s)^2, \quad s \in I^c \cup \{t\}.$$

and

$$(\dot{\lambda}_i, \dot{\alpha}_i) = \arg\min_{\lambda_i, \alpha_i} \sum_{s \in I^c \cup \{t\}} (y_{is} - \alpha_i'\widetilde{g}_s - x_{is}\lambda_i'\widetilde{f}_s)^2, \quad i = 1, ..., N.$$

Step 5. Estimate (f_t, λ_i) for inferences.

Motivated by (2.5), for all $s \in I^c \cup \{t\}$, let

$$(\widehat{f}_{I,s}, \widehat{g}_{I,s}) := \arg\min_{f_s, g_s} \sum_{i=1}^N (\widehat{y}_{is} - \widetilde{\alpha}_i' g_s - \widehat{e}_{is}\widetilde{\lambda}_i' f_s)^2.$$

Fix $i \leq N$,

$$(\widehat{\lambda}_{I,i}, \widehat{\alpha}_{I,i}) = \arg\min_{\lambda_i, \alpha_i} \sum_{s \in I^c \cup \{t\}} (\widehat{y}_{is} - \alpha_i'\widehat{g}_{I,s} - \widehat{e}_{is}\lambda_i'\widehat{f}_{I,s})^2.$$

where $\widehat{y}_{is} = y_{is} - \bar{x}_i\dot{\lambda}_i'\widetilde{f}_s$ and $\widehat{e}_{is} = x_{is} - \bar{x}_i$ in the many mean model , and $\widehat{y}_{is} = y_{is} - \widehat{l_i'w_s}\dot{\lambda}_i'\widetilde{f}_s$, $\widehat{e}_{is} = x_{is} - \widehat{l_i'w_s}$ in the factor model.

Step 6. Estimate θ_{it}. Repeat steps 3-5 with I and I^c exchanged, and obtain $(\widehat{\lambda}_{I^c,i}, \widehat{f}_{I^c,s} : s \in I \cup \{t\}, i \le N)$. Define

$$\widehat{\theta}_{it} := \frac{1}{2}[\widehat{\lambda}'_{I,i}\widehat{f}_{I,t} + \widehat{\lambda}'_{I^c,i}\widehat{f}_{I^c,t}].$$

Remark 2.1. We split the sample to $\{1, ..., T\} = I \cup I^c \cup \{t\}$. As the partialled-out regressor e_{it} is serially independent, this ensures that at the fixed t, e_{it} is independent of both splitted sample I and I^c. In steps 3-5 we estimate θ_{is} for $s \in \{t\} \cup I^c$. In step 6 we switch the roles of I and I^c to estimate θ_{is} for $s \in \{t\} \cup I$. As such, the parameter has been estimated twice at $s = t$. The final estimator is taken as the average of the two, $\widehat{\lambda}'_{I,i}\widehat{f}_{I,t}$ and $\widehat{\lambda}'_{I^c,i}\widehat{f}_{I^c,t}$. This gains the asymptotic efficiency that would otherwise be lost due to the sample splitting. Finally, we have an expansion:

$$\widehat{\theta}_{it} - \theta_{it} = \lambda_i' M_1 \frac{1}{N}\sum_{j=1}^{N} \lambda_j e_{jt} u_{jt} + f_t' M_1 \frac{1}{T}\sum_{t=1}^{T} f_s e_{is} u_{is} + \text{negligible},$$

for M_1, M_2 to be defined later.

Note that step 4 is needed to obtain a consistent estimate for $\dot{y}_{it} = y_{it} - \mu_{it}\lambda_i' f_t$, in order to apply the "partialled out equation " (2.3). In particular, the consistency is required to hold for each fixed (i, t), but $(\widetilde{\lambda}_i, \widetilde{\alpha}_i)$ obtained in the low-rank estimation from step 1 does not satisfy this condition. This motives the need for the estimators $(\dot{\lambda}_i, \dot{\alpha}_i)$. On the other hand, the estimators in step 4, however, are not suitable for inference, because at this step we use the regressor x_{it}, which is however, not a zero-mean or serially independent process. This gives rise to the need for step 5.

2.4. **Choosing the tuning parameters.** The "scores" of the nuclear-norm penalized regression are given by $2\|U\|$ and $2\|X \odot U\|$, where $\|.\|$ denotes the matrix operator norm, and $U = \{u_{it}\}$, and $X \odot U = (x_{it}u_{it})$. The tuning parameters (ν_2, ν_1) are taken so that

$$2\|U\| < (1-c)\nu_2, \quad 2\|X \odot U\| < (1-c)\nu_1$$

for some $c > 0$.

To quantify the operator norm of these quantities, we shall assume that the columns of U and $X \odot U$, respectively $\{u_t\}$ and $\{x_t \odot u_t\}$, are sub-Gaussian vectors. In the absence of serial correlations, by the eigenvalue-concentration inequality for independent sub-Gaussian random vectors (Theorem 5.39 of Vershynin (2010)):

$$\|(X \odot U)(X \odot U)' - \mathsf{E}(X \odot U)(X \odot U)'\| = O_P(\sqrt{NT} + N),$$

which provides a sharp bound for $\|X \odot U\|$, and a similar upper bound holds for $\|UU' - \mathsf{E}UU'\|$. Hence ν_2 and ν_1 can be chosen to satisfy $\nu_2 \asymp \nu_1 \asymp \max\{\sqrt{N}, \sqrt{T}\}$.

In the presence of serial correlations, we assume the following representation:

$$X \odot U = \Omega_{NT} \Sigma_T^{1/2}$$

where Ω_{NT} is an $N \times T$ matrix with independent, zero-mean, sub-Gaussian columns. Then still by the eigenvalue-concentration inequality for sub-Gaussian random vectors,

$$\|\Omega_{NT}\Omega_{NT}' - \mathsf{E}\Omega_{NT}\Omega_{NT}'\| = O_P(\sqrt{NT} + N).$$

In addition, Σ_T is a $T \times T$ deterministic matrix, possibly non-diagonal, whose eigenvalues are bounded from both below and above by constants. Allowing Σ_T to be a non-diagonal matrix captures the serial-correlations in $\{x_t \odot u_t\}$. This also implies $\|X \odot U\| \leq O_P(\max\{\sqrt{N}, \sqrt{T}\})$. Therefore, the tuning parameters in general can be chosen to satisfy

$$\nu_2 \asymp \nu_1 \asymp \max\{\sqrt{N}, \sqrt{T}\}.$$

In the Gaussian case, they can be computed via simulations. Suppose u_{it} is independent across both (i, t) and $u_{it} \sim \mathcal{N}(0, \sigma_{ui}^2)$. Let Z be an $N \times T$ matrix whose element z_{it} is generated from $\mathcal{N}(0, \sigma_{ui}^2)$, independent across (i, t). Then $\|X \odot U\| =^d \|X \odot Z\|$ and $\|U\| =^d \|Z\|$ where $=^d$ means "identically distributed". For a fixed $\delta_{NT} = o(1)$, let

$$\nu_2 = 2(1 + c_1)\bar{Q}(\|Z\|; 1 - \delta_{NT}), \quad \nu_1 = 2(1 + c_1)\bar{Q}(\|X \odot Z\|; 1 - \delta_{NT})$$

respectively denote $(1+c_1)$ multiplied by the $1-\delta_{NT}$ quantiles of $\|Z\|$ and $\|X \odot Z\|$. Then

$$2\|U\| < (1 - \frac{c_1}{1 + c_1})\nu_2, \quad 2\|X \odot U\| < (1 - \frac{c_1}{1 + c_1})\nu_1$$

holds with probability $1 - \delta_{NT}$. In practice, we replace σ_{ui}^2 with a consistent estimator, and take $c_1 = 0.1$, $\delta_{NT} = 0.05$.

3. Asymptotic Results

3.1. **The effect of low-rank estimations on inference.** We first heuristically discuss the main technical arguments.

Note that one of the key ingredients is to establish the asymptotic normality of $\widehat{f}_t$ for a fixed t. By definition, the estimation of $\widehat{f}_t$ depends on the SVD estimators (the nuclear-norm penalized regression), $(\widetilde{\alpha}_i, \widetilde{\lambda}_i)$, obtained as the eigenvectors of the estimated low-rank matrices. The main asymptotic effect of the SVD estimators give rise to the following two components: $\frac{1}{\sqrt{N}}\sum_{i=1}^N \lambda_i(\widetilde{\alpha}_i - H_2'\alpha_i)'e_{it}$, and

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \lambda_i(\widetilde{\lambda}_i - H_1'\lambda_i)'e_{it}^2, \tag{3.1}$$

where H_2, H_1 are rotation matrices. As we have previously explained, the effect of $\widetilde{\alpha}_i - H_2'\alpha_i$ can be argued to be negligible using sample splitting. However, the effect of $\widetilde{\lambda}_i - H_1'\lambda_i$ in (3.1) is not necessarily $o_P(1)$, because the Neyman's orthogonality does not hold in the moment conditions:

$$\frac{\partial}{\partial \lambda_i}\mathsf{E}e_{it}(\dot{y}_{it} - \alpha_i'g_t - e_{it}\lambda_i'f_t) \neq 0.$$

Indeed, it can be proved that, there is a matrix $\mathcal{B}_{NT} \neq o_P(1)$, the following expansion holds:

$$\sqrt{N}(\widehat{f}_t - H_1^{-1}f_t) = H_1^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N \lambda_i e_{it}u_{it} + \mathcal{B}_{NT}f_t + o_P(1),$$

where the presence of $\mathcal{B}_{NT}$ is due to (3.1).

We shall argue that such an effect, however, does not affect our asymptotic inferential theory, by properly recentering $\widehat{f}_t$. An important observation is that $\mathcal{B}_{NT}$ is time-invariant and is associated with f_t, so the bias also belongs to the space of f_t. As such, we can define $H_f := H_1^{-1} + \mathcal{B}_{NT}N^{-1/2}$ and establish that

$$\sqrt{N}(\widehat{f}_t - H_f f_t) = H_1^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N \lambda_i e_{it}u_{it} + o_P(1).$$

Therefore, the effect of $\widetilde{\lambda}_i - H_1'\lambda_i$ is "absorbed" by the adjusted rotation matrix. This is sufficient for the inferential theory of $\widehat{\theta}_{it}$. Once f_t can be estimated without biases (up to rotation transformations), the least squares iteration using $\widehat{f}_t$ continues to producing unbiased estimators for λ_i.

3.2. **Estimating low rank matrices.** We first introduce a key assumption about the nuclear-norm SVD procedure. We require some "invertibility" condition for the operator:

$$(\Delta_1, \Delta_2) :\rightarrow \Delta_1 + \Delta_2 \odot X$$

when (Δ_1, Δ_2) is restricted on a "cone", consisting of low-rank matrices, which we call *restricted low-rank set*, and was previously introduced and studied by Negahban and Wainwright (2011). To describe this cone, we first introduce some notation. Define

$$\Theta_I^0 = (\lambda_i' f_t : i \leq N, t \in I), \quad M_I^0 = (\alpha_i' g_t : i \leq N, t \in I).$$

Define $U_1 D_1 V_1' = \Theta_I^0$ and $U_2 D_2 V_2' = M_I^0$ as the SVD's of M_I^0 and Θ_I^0, where the superscript 0 represents the "true" parameter values. Further decompose, for $j = 1, 2$,

$$U_j = (U_{j,r}, U_{j,c}), \quad V_j = (V_{j,r}, V_{j,c})$$

Here $(U_{j,r}, V_{j,r})$ corresponds to the nonzero singular values, while $(U_{j,c}, V_{j,c})$ corresponds to the zero singular values. In addition, for any $N \times T/2$ matrix Δ, let

$$\mathcal{P}_j(\Delta) = U_{j,c} U_{j,c}' \Delta V_{j,c} V_{j,c}', \quad \mathcal{M}_j(\Delta) = \Delta - \mathcal{P}_j(\Delta).$$

Here $U_{j,c} U_{j,c}'$ and $V_{j,c} V_{j,c}'$ respectively are the projection matrices onto the columns of $U_{j,c}$ and $V_{j,c}$. Therefore, $\mathcal{M}_1(\cdot)$ and $\mathcal{M}_2(\cdot)$ can be considered as the projection matrices onto the "low-rank spaces" of Θ_I^0 and M_I^0 respectively, and $\mathcal{P}_1(\cdot)$ and $\mathcal{P}_2(\cdot)$ ar projections onto their orthogonal spaces.

Assumption 3.1 (Restricted strong convexity)**.** *Define restricted low-rank set to be: for some $c > 0$,*

$$\mathcal{C}(c) = \{(\Delta_1, \Delta_2) : \|\mathcal{P}_1(\Delta_1)\|_1 + \|\mathcal{P}_2(\Delta_2)\|_1 \leq c\|\mathcal{M}_1(\Delta_1)\|_1 + c\|\mathcal{M}_2(\Delta_2)\|_1\}.$$

If $(\Delta_1, \Delta_2) \in \mathcal{C}(c)$ for some $c > 0$, then there is a constant $\kappa_c > 0$,

$$\|\Delta_1 + \Delta_2 \odot X\|_F^2 \geq \kappa_c \|\Delta_1\|_F^2 + \kappa_c \|\Delta_2\|_F^2.$$

The same condition holds when (M_I^0, Θ_I^0) are replaced with $(M_{I^c}^0, \Theta_{I^c}^0)$, or the full-sample (M^0, Θ^0).

Our next assumption allows non-stationary or arbitrary dependence in the time-series for (f_t, g_t). In particular, they hold for serially weakly dependent sequences, and even allow for perfect dependent sequence: $(f_t, g_t) = (f, g)$ for some time-invariant (f, g) by setting $\dim(f_t) = \dim(g_t) = 1$.

Assumption 3.2. *As $T \to \infty$, the sub-samples (I, I^c) satisfy:*

$$\frac{1}{|I|_0}\sum_{t\in I} f_t f_t' = \frac{1}{T}\sum_{t=1}^T f_t f_t' + O_P(T^{-1/2}) = \frac{1}{|I^c|_0}\sum_{t\in I^c} f_t f_t',$$
$$\frac{1}{|I|_0}\sum_{t\in I} g_t g_t' = \frac{1}{T}\sum_{t=1}^T g_t g_t' + O_P(T^{-1/2}) = \frac{1}{|I^c|_0}\sum_{t\in I^c} g_t g_t'.$$

In addition, there is $c > 0$, all the eigenvalues of $\frac{1}{T}\sum_{t=1}^T f_t f_t'$ and $\frac{1}{T}\sum_{t=1}^T g_t g_t'$ are bounded from below by c almost surely.

Assumption 3.3 (Valid factor structures with strong factors)**.** *There are constants $c_1 > ... > c_{K_1} > 0$, and $c_1' > ... > c_{K_2}' > 0$, so that, up to a term $o_P(1)$,*

(i) c_j' equals the j th largest eigenvalue of $(\frac{1}{T}\sum_t g_t g_t')^{1/2}\frac{1}{N}\sum_{i=1}^N \alpha_i\alpha_i'(\frac{1}{T}\sum_t g_t g_t')^{1/2}$ for all $j = 1, ..., K_1$, and

(ii) c_j equals the j th largest eigenvalue of $(\frac{1}{T}\sum_t f_t f_t')^{1/2}\frac{1}{N}\sum_{i=1}^N \lambda_i\lambda_i'(\frac{1}{T}\sum_t f_t f_t')^{1/2}$ for all $j = 1, ..., K_2$.

The above assumption requires that the factors be strong. In addition, we require distinct eigenvalues in order to identify their corresponding eigenvectors, and therefore, (λ_i, α_i).

Recall that $\widetilde{M}_S$ and $\widetilde{\Theta}_S$ respectively are the estimated low-rank matrices obtained by the nuclear-norm penalized estimations on sample $S \in \{I, I^c\}$. Given the above assumption, we have the consistency-in-frobenius-norm for the estimated low-rank matrices,

Proposition 3.1. *Suppose $2\|X \odot U\| < (1-c)\nu_1$, $2\|U\| < (1-c)\nu_2$ and $\nu_2 \asymp \nu_1$. Then under Assumption 3.1, for $S \in \{I, I^c, I\cup I^c\}$, (i)*

$$\frac{1}{NT}\|\widetilde{M}_S - M_S\|_F^2 = O_P(\frac{\nu_2^2+\nu_1^2}{NT}) = \frac{1}{NT}\|\widetilde{\Theta}_S - \Theta_S\|_F^2.$$

(ii) Additionally with Assumption 3.3, there are square matrices H_{S1}, H_{S2}, so that

$$\frac{1}{N}\|\widetilde{A}_S - AH_{S1}\|_F^2 = O_P(\frac{\nu_2^2+\nu_1^2}{NT}), \quad \frac{1}{N}\|\widetilde{\Lambda}_S - \Lambda H_{S2}\|_F^2 = O_P(\frac{\nu_2^2+\nu_1^2}{NT}).$$

(iii) Furthermore,

$$P(\widehat{K}_1 = K_1, \quad \widehat{K}_2 = K_2) \to 1.$$

3.3. **Asymptotic analysis when x_{it} admits a many mean model.** We now focus on the case $x_{it} = \mu_i + e_{it}$.

Assumption 3.4. *Let $\omega_t = (x_{1t}u_{1t}, ..., x_{Nt}u_{Nt}, u_{1t}, ..., u_{Nt})'$. Suppose $\{\omega_t\}_{t\le T}$ are independent sub-gaussian random vectors: there is $C > 0$,*

$$\max_{t\le T} \sup_{\|x\|=1} \mathsf{E} \exp(s\omega_t' x) \le \exp(s^2 C), \quad \forall s \in \mathbb{R}.$$

The sub-Gaussian condition allows us to apply the eigenvalue-concentration inequality for independent random vectors (Theorem 5.39 of Vershynin (2010)) to bound the scores $\|X \odot U\|$ and $\|U\|$. Under Assumption 3.4, we can take $\nu_2, \nu_1 = O_P(\sqrt{N+T})$. Therefore

$$\frac{1}{N}\|\widetilde{A}_S - AH_{S1}\|_F^2 = O_P(\frac{1}{N} + \frac{1}{T}) = \frac{1}{N}\|\widetilde{\Lambda}_S - \Lambda H_{S2}\|_F^2.$$

The assumptions below are also needed to achieve the inferential theory for the estimated treatment effect.

Assumption 3.5 (Dependence). *(i) $\{e_{it}\}$ is serially stationary, and*

$$\mathsf{E}(e_{it}|u_{it}, g_t, f_t) = 0, \quad \mathsf{E}(u_{it}|e_{it}, g_t, f_t) = 0.$$

(ii) $\{e_{it}, u_{it}\}$ are independent across t; $\{e_{it}\}$ are also conditionally independent across t, given $\{f_t, g_t, u_t\}$; $\{u_{it}\}$ are also conditionally independent across t, given $\{f_t, g_t, e_t\}$;

(iii) Weak conditional cross-sectional dependence: Let $\mathcal{W} = (F, G)$ and (E, U) be the $N \times T$ matrices of (e_{it}, u_{it}). Let $\omega_{it} = u_{it}e_{it}$, and let c_i be a bounded nonrandom sequence. Almost surely,

$$\max_t \frac{1}{N^3} \sum_{ijkl} |\mathsf{Cov}(e_{it}e_{jt}, e_{kt}e_{lt}|\mathcal{W}, U)| < C$$
$$\max_{i,t} \sum_j |\mathsf{Cov}(e_{it}^m, e_{jt}^r|\mathcal{W}, U)| < C, \quad m, r \in \{1, 2\}$$
$$\max_t \|\mathsf{E}(u_t u_t'|\mathcal{W}, E)\| < C$$
$$\max_t \mathsf{E}(|\frac{1}{\sqrt{N}} \sum_i c_i \omega_{it}|^4 |\mathcal{W}) < C$$
$$\max_t \mathsf{E}(|\frac{1}{\sqrt{N}} \sum_i c_i e_{it}|^4 |\mathcal{W}, U)$$
$$\max_t \mathsf{E}(|\frac{1}{\sqrt{N}} \sum_i c_i u_{it}|^4 |\mathcal{W}, E) < C$$

In addition, for each fixed $i \leq N$,

$$\max_{s\leq T}\frac{1}{N}\sum_{kj}|\mathsf{E}(e_{ks}e_{is}e_{js}|\mathcal{W},U)| < C$$
$$\max_{t}\frac{1}{N}\sum_{kj}|\mathsf{Cov}(\omega_{it}\omega_{jt},\omega_{it}\omega_{kt}|\mathcal{W})| < C$$
$$\max_{t}\sum_{j}|\mathsf{Cov}(\omega_{it},\omega_{jt})|\mathcal{W})| < C.$$

Let $\mathsf{diag}(X_s)$ be a diagonal matrix of x_{is} for a fixed $s \leq T$. Let $M_\alpha = \mathsf{I}_N - A(A'A)^{-1}A$, $M_g = \mathsf{I}_N - G(G'G)^{-1}G$, and

$$\begin{aligned} D_{fs} &= \frac{1}{N}\Lambda'\mathsf{diag}(X_s)M_\alpha\mathsf{diag}(X_s)\Lambda, \quad D_{\lambda i} = \frac{1}{T}F'\mathsf{diag}(X_i)M_g\mathsf{diag}(X_i)F \\ D_f &= \frac{1}{N}\Lambda'\mathsf{E}(\mathsf{diag}(X_s)M_\alpha\mathsf{diag}(X_s))\Lambda, \quad \bar{D}_{\lambda i} = \frac{1}{T}F'\mathsf{E}(\mathsf{diag}(X_i)M_g\mathsf{diag}(X_i))F. \end{aligned}$$

Assumption 3.6 (Moment bounds)**.** *(i)* $\max_i(\|\lambda_i\| + \|\alpha_i\| + \max_i|\mu_i|) < C$.

(ii) $\max_{it}\bar{e}_i^2\max_{it}e_{it}^2 = O_P(\min\{N,T\})$, $\max_i|\bar{e}_i| = O_P(1)$,
$C_{NT}^{-1}\max_{it}|e_{it}|^2 + \max_t\|\frac{1}{N}\sum_i e_{it}\alpha_i\lambda_i'\|_F + \max_t\|\frac{1}{N}\sum_i\lambda_i\lambda_i'\bar{e}_ie_{it}\|_F = o_P(1)$,

(iii) There is $c > 0$, so that almost surely, for all $s \leq T$ and $i \leq N$, $\min_{j\leq K_2}\psi_j(D_{\lambda i}) > c$, $\min_{j\leq K_2}\psi_j(D_{fs}) > c$, $\min_{j\leq K_2}\psi_j(\bar{D}_{\lambda i}) > c$ *and* $\min_{j\leq K_2}\psi_j(D_f) > c$.

(iv) $\max_{is}\mathsf{E}(e_{is}^6|U,F) < C$, *and* $\max_{is}\mathsf{E}\|g_s\|^2\|f_s\|^2e_{is}^4 < \infty$, $\mathsf{E}u_{ks}^2\|f_s\|^2e_{is}^2e_{js}^2 < C$. $\mathsf{E}\|g_s\|^4 + \mathsf{E}\|f_s\|^4 < \infty$.

Theorem 3.1. *Under assumptions 3.1- 3.6, for any fixed $i \leq N$ and $t \leq T$, as $N,T \to \infty$,*

$$\frac{\widehat{\theta}_{it} - \theta_{it}}{(\frac{1}{N}\lambda_i'V_\lambda\lambda_i + \frac{1}{T}f_t'V_ff_t)^{1/2}} \to^d \mathcal{N}(0,1),$$

where $V_f = V_{f1}^{-1}V_{f2}V_{f1}^{-1}$ and $V_\lambda = V_{\lambda1}^{-1}V_{\lambda2}V_{\lambda1}^{-1}$, and the related quantities are defined as:

$V_{f1} = \frac{1}{T}\sum_s f_sf_s'\mathsf{E}e_{is}^2$, $V_{f2} = \frac{1}{T}\sum_s f_sf_s'\mathsf{E}(e_{is}^2u_{is}^2|F)$, $V_{\lambda1} = \frac{1}{N}\sum_j\lambda_j\lambda_j'\mathsf{E}e_{jt}^2$, *and* $V_{\lambda2} = \mathsf{Var}(\frac{1}{\sqrt{N}}\sum_j\lambda_je_{jt}u_{jt})$.

While factors and loadings are estimated up to a rotation matrix, both θ_{it} and its asymptotic variance are rotation free. On the other hand, the rotation matrices are different on the two splitted sample. Therefore, to preserve the rotation-free property of the asymptotic variance, the above quantities should be separately estimated on the two splitted sample, and the final asymptotic variance estimator should be taken as the average.

In addition, to estimate $V_{\lambda 2}$ we shall assume u_{jt} to be cross-sectionally independent given $\{e_{jt}\}$ for simplicity. More specifically, given a splitted sample $S \in \{I, I^c\}$, we respectively estimate the above quantities on S by $\widehat{V}_{f,1}^S = \frac{1}{|S|_0}\sum_{s\in S}\widehat{f}_s\widehat{f}_s'\widehat{e}_{is}^2$, $\widehat{V}_{f,2}^S = \frac{1}{|S|_0}\sum_{s\in S}\widehat{f}_s\widehat{f}_s'\widehat{e}_{is}^2\widehat{u}_{is}^2$, $\widehat{V}_{\lambda,1}^S = \frac{1}{N}\sum_j\widehat{\lambda}_j\widehat{\lambda}_j'\widehat{e}_{jt}^2$, and $\widehat{V}_{\lambda,2}^S = \frac{1}{N}\sum_j\widehat{\lambda}_j\widehat{\lambda}_j'\widehat{e}_{jt}^2\widehat{u}_{jt}^2$. We have

$$\begin{aligned}\widehat{v}_\lambda &= \frac{1}{2N}(\widehat{\lambda}_{I,i}'\widehat{V}_{\lambda,1}^{I-1}\widehat{V}_{\lambda,2}^{I}\widehat{V}_{\lambda,1}^{I-1}\widehat{\lambda}_{I,i} + \widehat{\lambda}_{I^c,i}'\widehat{V}_{\lambda,1}^{I^c-1}\widehat{V}_{\lambda,2}^{I^c}\widehat{V}_{\lambda,1}^{I^c-1}\widehat{\lambda}_{I^c,i})\\ \widehat{v}_f &= \frac{1}{2T}(\widehat{f}_{I,t}'\widehat{V}_{f,1}^{I-1}\widehat{V}_{f,2}^{I}\widehat{V}_{f,1}^{I-1}\widehat{f}_{I,t} + \widehat{f}_{I^c,t}'\widehat{V}_{f,1}^{I^c-1}\widehat{V}_{f,2}^{I^c}\widehat{V}_{f,1}^{I^c-1}\widehat{f}_{I^c,t})\end{aligned}$$

and the estimated asymptotic variance of $\widehat{\theta}_{it}$ is $\widehat{v}_\lambda + \widehat{v}_f$.

Corollary 3.1. *In addition to the assumptions of Theorem 3.1, assume u_{it} to be cross-sectionally independent given $\{e_{it}\}$. For any fixed $i \le N$ and $t \le T$,*

$$\frac{\widehat{\theta}_{it} - \theta_{it}}{(\widehat{v}_\lambda + \widehat{v}_f)^{1/2}} \to^d \mathcal{N}(0,1).$$

3.4. **Asymptotic analysis when x_{it} admits a factor model.** In the presence of cross-sectional and serial dependence in x_{it}, we assume $x_{it} = l_i'w_t + e_{it}$.

Assumption 3.7 (Dependence). *(ii) $\{e_{it}, u_{it}\}$ are independent across t; $\{e_{it}\}$ are also conditionally independent across t, given $\{f_t, g_t, w_t, u_t\}$; $\{u_{it}\}$ are also conditionally independent across t, given $\{f_t, g_t, w_t, e_t\}$;*

(ii)

$$\mathsf{E}(e_{it}|u_{it}, w_t, g_t, f_t) = 0, \quad \mathsf{E}(u_{it}|e_{it}, w_t, g_t, f_t) = 0.$$

(iii) The $N \times T$ matrix $X \odot U := (u_{it}x_{it})$ has the following decomposition:

$$X \odot U = \Omega_{NT}\Sigma_T^{1/2}$$

where : (a) $\Omega_{NT} := (\omega_1, ..., \omega_T)$ is an $N \times T$ matrix, whose columns $\{\omega_t\}_{t\le T}$ are independent sub-gaussian random vectors, with $\mathsf{E}\omega_t = 0$, more specifically, there is $C > 0$,

$$\max_{t\le T}\sup_{\|x\|=1}\mathsf{E}\exp(s\omega_t'x) \le \exp(s^2C), \quad \forall s \in \mathbb{R}.$$

(b) Σ_T is a $T \times T$ deterministic matrix, whose eigenvalues are bounded from both below and above by constants.

(iv) Cross-sectional weak dependence: Assumption 3.5(iii) holds with $\mathcal{W} = (F, G, W)$.

Condition (iii) allows x_{it} to be serially weakly dependent, where the serial correlations are captured by Σ_T. The required conditions allow us to apply the eigenvalue-concentration inequality for independent sub-Gaussian random vectors on Ω_{NT}. In addition, we allow arbitrary dependence among rows of Ω_{NT}, and thus the strong cross-sectional dependence among x_{it} are allowed, which is desirable given the factor structure.

Next, define

$$\begin{aligned}
b_{NT,1} &= \max_t \|\frac{1}{NT}\sum_{is} w_s(e_{is}e_{it} - \mathsf{E}e_{is}e_{it})\| \\
b_{NT,2} &= (\max_t \frac{1}{T}\sum_s(\frac{1}{N}\sum_i e_{is}e_{it} - \mathsf{E}e_{is}e_{it})^2)^{1/2} \\
b_{NT,3} &= \max_t \|\frac{1}{N}\sum_i l_i e_{it}\| \\
b_{NT,4} &= \max_i \|\frac{1}{T}\sum_s e_{is}w_s\| \\
b_{NT,5} &= \max_i \|\frac{1}{NT}\sum_{js} l_j(e_{js}e_{is} - \mathsf{E}e_{js}e_{is})\|
\end{aligned}$$

In addition,

$$\begin{aligned}
D_{fs} &= \frac{1}{N}\Lambda'(\mathsf{diag}(X_s)M_\alpha \mathsf{diag}(X_s)\Lambda \\
\bar{D}_{fs} &= \frac{1}{N}\Lambda'\mathsf{E}((\mathsf{diag}(e_s)M_\alpha \mathsf{diag}(e_s))\Lambda + \frac{1}{N}\Lambda'(\mathsf{diag}(Lw_s)M_\alpha \mathsf{diag}(Lw_s)\Lambda \\
D_{\lambda i} &= \frac{1}{T}F'(\mathsf{diag}(X_i)M_g \mathsf{diag}(X_i))F \\
\bar{D}_{\lambda i} &= \frac{1}{T}F'\mathsf{E}(\mathsf{diag}(E_i)M_g \mathsf{diag}(E_i))F + \frac{1}{T}F'(\mathsf{diag}(Wl_i)M_g \mathsf{diag}(Wl_i))F.
\end{aligned}$$

Assumption 3.8 (Moment bounds)**.** *(i)* $\max_i(\|\lambda_i\| + \|\alpha_i\| + \|l_i\|) < C$.

(ii) Let $\delta_{NT} := (C_{NT}^{-1} + b_{NT,4} + b_{NT,5})\max_t \|w_t\| + b_{NT,1} + b_{NT,3} + C_{NT}^{-1}b_{NT,2} + C_{NT}^{-1/2}$. *Then* $\delta_{NT}\max_{it}|e_{it}| = o_P(1)$, $\max_t \|\frac{1}{N}\sum_i e_{it}\alpha_i\lambda_i'\|_F = o_P(1)$,

(iii) There is $c > 0$*, so that almost surely, for all* $s \leq T$ *and* $i \leq N$, $\min_{j\leq K_2}\psi_j(D_{\lambda i}) > c$, $\min_{j\leq K_2}\psi_j(D_{fs}) > c$, $\min_{j\leq K_2}\psi_j(\bar{D}_{\lambda i}) > c$ *and* $\min_{j\leq K_2}\psi_j(\bar{D}_{fs}) > c$. *In addition,*

$$c < \min_j \psi_j(\frac{1}{N}\sum_i l_i l_i') \leq \max_j \psi_j(\frac{1}{N}\sum_i l_i l_i') < C$$

$$c < \min_j \psi_j(\frac{1}{T}\sum_t w_t w_t') \leq \max_j \psi_j(\frac{1}{T}\sum_t w_t w_t') < C$$

(iv) $\max_{is} \mathsf{E}(e_{is}^8|U,F) < C$, *and* $\mathsf{E}\|w_s\|^4 + \|g_s\|^4 + \mathsf{E}\|f_s\|^4 < C$, *and* $\mathsf{E}\|g_s\|^4\|f_s\|^4 + \mathsf{E}u_{ks}^4\|f_s\|^4 + \mathsf{E}e_{jt}^4\|f_t\|^8 + \mathsf{E}e_{jt}^4\|f_t\|^4\|g_t\|^4 + \mathsf{E}\|w_t\|^4\|g_t\|^4 < C$.

Theorem 3.2. *For any fixed $i \leq N$ and $t \leq T$, Under Assumptions 3.1-3.3, 4.1 and 3.8, as $N, T \to \infty$,*

$$\frac{\widehat{\theta}_{it} - \theta_{it}}{V^{1/2}} \to^d \mathcal{N}(0,1), \quad \frac{\widehat{\theta}_{it} - \theta_{it}}{\widehat{V}^{1/2}} \to^d \mathcal{N}(0,1),$$

where, for some invertible matrix H_f.

$$V = \frac{1}{N}\lambda_i' H_f^{-1} V_f H^{-1'} \lambda_i + \frac{1}{T} f_t' H_f' V_\lambda H_f f_t.$$

In addition, the conclusion of Corollary 3.1 continues to hold.

4. Testing about Homogeneous effects

In most empirical studies it is commonly assumed that the coefficients of the observed regressors are *globally* homogeneous. We can divide the null hypotheses into three classes:

Time-invariant homogeneity:

$$H_0^1 : \theta_{it} = \theta_i, \quad \forall (i,t).$$

Cross-sectionally invariant homogeneity:

$$H_0^2 : \theta_{it} = \theta_t, \quad \forall (i,t).$$

Pure homogeneity:

$$H_0^3 : \theta_{it} = \theta, \quad \text{for some } \theta \in \mathbb{R}, \quad \forall (i,t).$$

All the models regulated by these hypotheses are admitted as special cases as our general model. But note that in homogeneous models, we should require $\dim(\lambda_i) = \dim(f_t) = 1$ to avoid perfect multicolinearity. In particular, this is required by Assumption 3.3.

This section proposes tests for all three types of homogeneity. Our testing procedure is motivated by a simple argument as follows. Respectively let $\bar{\theta}_i = \frac{1}{T}\sum_t \theta_{it}$, $\bar{\theta}_t = \frac{1}{N}\sum_i \theta_{it}$, and $\bar{\theta} = \frac{1}{NT}\sum_{it}\theta_{it}$. Then the null hypotheses are respectively equivalent to

$$H_0^1 : \frac{1}{NT}\sum_{it}(\theta_{it} - \bar{\theta}_i)^2 = 0,$$

$$H_0^2 : \frac{1}{NT}\sum_{it}(\theta_{it} - \bar{\theta}_t)^2 = 0,$$

and

$$H_0^3 : \frac{1}{NT}\sum_{it}(\theta_{it} - \bar{\theta})^2 = 0.$$

Let $\bar{\widehat{\theta}}_i = \frac{1}{T}\sum_t \widehat{\theta}_{it}$, $\bar{\widehat{\theta}}_t = \frac{1}{N}\sum_i \widehat{\theta}_{it}$, and $\bar{\widehat{\theta}} = \frac{1}{NT}\sum_{it} \widehat{\theta}_{it}$. Therefore our test is based on the "sample variance" of θ_{it}:

$$\begin{aligned} S_{NT,1} &= \frac{1}{NT}\sum_{it}(\widehat{\theta}_{it} - \bar{\widehat{\theta}}_i)^2, \quad \textit{for } H_0^1, \\ S_{NT,2} &= \frac{1}{NT}\sum_{it}(\widehat{\theta}_{it} - \bar{\widehat{\theta}}_t)^2, \quad \textit{for } H_0^2, \\ S_{NT,3} &= \frac{1}{NT}\sum_{it}(\widehat{\theta}_{it} - \bar{\widehat{\theta}})^2, \quad \textit{for } H_0^3. \end{aligned} \tag{4.1}$$

We reject the null hypothesis for small values of $S_{NT,d}$ for each hypothesis.

Alternatively, one can construct test statistics in the spirit of Hausman's test. For instance, under the null hypothesis of pure homogeneous models:

$$y_{it} = \alpha_i' g_t + x_{it}\theta + e_{it},$$

one can estimate θ using the restricted estimator (e.g., Bai (2009)), and compare the restricted and unrestricted estimators. Our proposed tests, in contrast, is much simpler, as it only requires simply calculating the sample mean and variance of $\widehat{\theta}_{it}$, but no additional estimation steps.

4.1. **Computing the test statistics.** For each fixed (i, t), Algorithm 2.2 computes $\widehat{\theta}_{it}$ by splitting the sample into $\{1, ..., T\} = \{t\} \cup I \cup I^c$, taking out the t th observation. This ensures the partialled-out regressor e_{it} to be independent of both splitted sample I and I^c and gains the asymptotic efficiency for estimating θ_{it} for the fixed t. However, as the test statistics depend on $\{\widehat{\theta}_{it} : i \leq N, t \leq T\}$, it would require new sample splitting and nuclear-norm penalized regressions for each $t = 1, ..., T$, which could be very computationally demanding.

As such, we propose a revised algorithm to compute the estimated θ_{it}, which splits the sample into $\{1, ..., T\} = I \cup I^c$. In fact, it only requires sample splitting once, and computing the nuclear-norm penalized regression twice in total (respectively on I and I^c).

Algorithm 4.1. Compute $S_{NT,d}$ as follows.

Step 1. Estimate the structure $x_{it} = \mu_{it} + e_{it}$ as in step 2 of Algorithm 2.2.

Step 2: Sample splitting. Randomly split the sample into $\{1, ..., T\} = I \cup I^c$, so that $|I|_0 = [T/2]$ (the nearest integer to $T/2$).

Step 3. Estimate (f_t, λ_i) *for* $t \in I^c$. Run steps 3-5 of Algorithm 2.2 and obtain

$$\{\widehat{f}_t : t \in I^c\}, \quad \{\widehat{\lambda}_{I,i}, i \leq N\}.$$

Step 4. Estimate (f_t, λ_i) *for* $t \in I$. Run steps 3-5 of Algorithm 2.2 with I and I^c exchanged, and obtain

$$\{\widehat{f}_t : t \in I\}, \quad \{\widehat{\lambda}_{I^c,i}, i \leq N\}.$$

Step 5. For $i = 1, ..., N$, and $t = 1, ..., T$, let

$$\widetilde{\theta}_{it} = \frac{1}{2}(\widehat{\lambda}_{I,i} + \widehat{\lambda}_{I^c,i})' \widehat{f}_t, \tag{4.2}$$

and compute $S_{NT,d}$ as in (4.1) with $\widetilde{\theta}_{it}$ in place of $\widehat{\theta}_{it}$.

Remark 4.1. The main difference in estimating θ_{it} between Algorithms 2.2 and 4.1 is that in Algorithm 2.2, the sample splitting is specific to a fixed $t : \{1, ..., T\} = \{t\} \cup I \cup I^c$, and θ_{it} is estimated twice at time t. The final estimator is then constructed as the average of the two. In contrast, Algorithm 4.1 uses a "universe" sample splitting $\{1, ..., T\} = I \cup I^c$, and θ_{it} is estimated only once at each time t. In (4.2), the estimated λ_i is averaged to gain the asymptotic efficiency for the effect of estimating λ_i, but $\widehat{f}_t$ is not.

We note that Algorithm 4.1 is computationally efficient. While the computed $\widetilde{\theta}_{it}$ is *not* asymptotically equivalent to $\widehat{\theta}_{it}$, it cannot be used as inference for a fixed θ_{it}. [1] But they can be used for the homogeneous tests. Specifically, Algorithm 4.1 produces $S_{NT,d}$ whose asymptotic null hypothesis is equivalent to that of the statistic constructed based on $\widehat{\theta}_{it}$. To understand this, note that the asymptotic

[1]Note that λ_i and f_t are estimated consistently up to rotations, such rotations are canceled out in the construction of $\widehat{\theta}_{it}$, but not anymore for $\widetilde{\theta}_{it}$. For instance, when $t \in I^c$, then $(\widehat{f}_t, \widehat{\lambda}_{I,i}, \widehat{\lambda}_{I^c,i})$ respectively consistently estimate $(H_I^{-1'} f_t, H_I \lambda_i, H_{I^c} \lambda_i)$ for rotation matrices $H_I \neq H_{I^c}$. As such, $\widehat{\lambda}'_{I,i} \widehat{f}_t$ consistently estimates $\lambda'_i f_t$, but $\widehat{\lambda}'_{I^c,i} \widehat{f}_t$ consistently estimates $\lambda'_i H'_{I^c} H_I^{-1'} f_t \neq \lambda'_i f_t$. As such, unlike $\widehat{\theta}_{it}$, here $\widetilde{\theta}_{it}$ cannot be used for inference about θ_{it}, whose asymptotic expansion depends on an additional term $\mathcal{A}_{NT}$ that is not asymptotically negligible.

expansion of $\widetilde{\theta}_{it}$ is, for $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$

$$\widetilde{\theta}_{it} = \theta_{it} + \lambda_i' \mathcal{A}_{NT} f_t + \lambda_i' V_{\lambda 1}^{-1} \frac{1}{N} \sum_j \lambda_j e_{jt} u_{jt} + f_t' V_{f1}^{-1} \frac{1}{T} \sum_{s=1}^{T} f_s e_{is} u_{is} + O_P(C_{NT}^{-2}) \quad (4.3)$$

where the additional term $\lambda_i' \mathcal{A}_{NT} f_t = O_P(C_{NT}^{-1})$ arises from the rotation discrepancy, and is not asymptotically negligible. So $\widetilde{\theta}_{it}$ is biased for θ_{it}. In contrast, in the expansion of $\widehat{\theta}$, there is no such term $\lambda_i' \mathcal{A}_{NT} f_t$, which is the reason why $\widehat{\theta}_{it}$ can be used for inference at a fixed (i,t) while $\widetilde{\theta}_{it}$ cannot. On the other hand, for testing purposes, they are equivalent. Consider the null hypothesis $\theta_{it} = \theta$ for all (i,t). Take the average over (i,t), we have $\frac{1}{NT}\sum_{it} \widetilde{\theta}_{it} = \theta + \bar{\lambda} \mathcal{A}_{NT} \bar{f} + O_P(C_{NT}^{-2})$. Subtract from (4.3), and note that under the null $(\lambda_i, f_t) = (\bar{\lambda}, \bar{f})$ for all (i,t),

$$\widetilde{\theta}_{it} - \frac{1}{NT} \sum_{it} \widetilde{\theta}_{it} = \lambda_i' V_{\lambda 1}^{-1} \frac{1}{N} \sum_j \lambda_j e_{jt} u_{jt} + f_t' V_{f1}^{-1} \frac{1}{T} \sum_{s=1}^{T} f_s e_{is} u_{is} + O_P(C_{NT}^{-2}).$$

We thus reach the same asymptotic expansion as that of $\widehat{\theta}_{it} - \bar{\widehat{\theta}}$, leading to the asymptotic null distribution of $S_{NT,3}$.

4.2. **Asymptotical null distribution.** From (4.3), we note that the asymptotic null distributions depend on $\lambda_i' V_{\lambda 1}^{-1} \frac{1}{N} \sum_j \lambda_j e_{jt} u_{jt}$ and $f_t' V_{f1}^{-1} \frac{1}{T} \sum_{s=1}^{T} f_s e_{is} u_{is}$. As such, let

$$s_{2,t} = \frac{1}{N} \sum_i (\lambda_i' V_{\lambda 1}^{-1} \frac{1}{\sqrt{N}} \sum_j \lambda_j e_{jt} u_{jt})^2,$$

and

$$s_{1,i} = \frac{1}{T} \sum_t (f_t' V_{f1}^{-1} \frac{1}{\sqrt{T}} \sum_s f_s e_{is} u_{is})^2.$$

Then under H_0^1, we have $S_{NT,1} \approx \frac{1}{NT} \sum_i s_{1,i}$. Similarly, under H_0^2, $S_{NT,2} \approx \frac{1}{NT} \sum_t s_{2,t}$. Next, under H_0^3,

$$S_{NT,3} \approx \frac{1}{NT} \sum_i s_{1,i} + \frac{1}{NT} \sum_t s_{2,t}. \quad (4.4)$$

Therefore, properly scaled test statistics are asymptotically standard normal. For instance:

$$\frac{\sqrt{NT}(S_{NT,3} - a_{NT})}{\left(\frac{1}{N} V_1 + \frac{1}{T} V_2\right)^{1/2}} \to^d \mathcal{N}(0,1),$$

for properly defined asymptotic mean and variance.

Let

$$\widehat{a}_{NT,2} = \frac{1}{NT}\sum_t \frac{1}{N^2}\sum_{ij} \widehat{w}_{jt}^2(\widehat{\lambda}_i'\widetilde{M}_1\widehat{\lambda}_j)^2\widehat{u}_{jt}^2, \quad \widehat{a}_{NT,1} = \frac{1}{NT}\sum_i \frac{1}{T^2}\sum_{ts} \widehat{w}_{is}^2(\widehat{f}_t'\widetilde{M}_2\widehat{f}_s)^2\widehat{u}_{is}^2.$$

Let $\widehat{V}_2 = 2\frac{1}{T}\sum_t \frac{1}{N^2}\sum_{ij}(\widehat{\lambda}_i'\widetilde{M}_1'\frac{1}{N}\widehat{\Lambda}'\widehat{\Lambda}\widehat{M}_1\widehat{\lambda}_j)^2\widehat{w}_{it}^2\widehat{w}_{jt}^2\widehat{u}_{it}^2\widehat{u}_{jt}^2$, and
$\widehat{V}_1 = 2\frac{1}{N}\sum_i \frac{1}{T^2}\sum_{ts}(\widehat{f}_t'\widetilde{M}_2'\frac{1}{T}\sum_k \widehat{f}_k'\widehat{f}_k'\widehat{M}_2\widehat{f}_s)^2\widehat{u}_{it}^2\widehat{u}_{is}^2$.

Assumption 4.1. *u_{it} is cross-sectionally and serially independent, given $\{x_{it}, f_t\}$.*

Theorem 4.1. *Suppose Assumptions 3.1-4.1 hold. Then*

$$\text{under } H_0^1: \quad \frac{\sqrt{NT}(S_{NT,1} - \widehat{a}_{NT,1})}{\left(\frac{1}{N}\widehat{V}_1\right)^{1/2}} \to^d \mathcal{N}(0,1),$$

$$\text{under } H_0^2: \quad \frac{\sqrt{NT}(S_{NT,2} - \widehat{a}_{NT,2})}{\left(\frac{1}{T}\widehat{V}_2\right)^{1/2}} \to^d \mathcal{N}(0,1),$$

and

$$\text{under } H_0^3: \quad \frac{\sqrt{NT}(S_{NT,3} - \widehat{a}_{NT,1} - \widehat{a}_{NT,2})}{\left(\frac{1}{N}\widehat{V}_1 + \frac{1}{T}\widehat{V}_2\right)^{1/2}} \to^d \mathcal{N}(0,1).$$

5. Monte Carlo Simulations

5.1. **Static models.** We use simulations to assess the adequacy of the asymptotic distributions. We set

$$y_{it} = \alpha_i g_t + x_{it,1}\theta_{it} + x_{it,2}\beta_{it} + u_{it}$$

where $\theta_{it} = \lambda_{i,1}'f_{t,1}$, and $\beta_{it} = \lambda_{i,2}'f_{t,2}$. In addition, we have two processes for the regressors: either a many mean model

$$x_{it,r} = \mu_{i,r} + e_{it,r}, \quad r = 1, 2,$$

or a factor model:

$$x_{it,r} = l_{i,r}w_{t,r} + e_{it,r}, \quad r = 1, 2.$$

Here $e_{it,r}$ are generated independently from the standard normal distribution across (i,t). In addition, all the number of factors are set as one, and all the means, loadings and factors are independently generated from the standard normal distribution.

We define standardized estimates as:

$$\frac{\widehat{\theta}_{it}-\theta_{it}}{(\widehat{V}_{\lambda}+\widehat{V}_{f})^{1/2}}.$$

So the standardization is based on the theoretical mean and theoretical variance rather than the sample mean and sample variance from Monte Carlo repetitions. We next compute the sample mean and sample standard deviation (std) from 1,000 Monte Carlo repetitions of standardized estimates. If the asymptotic theory is adequate, the standardized estimates should be approximately $\mathcal{N}(0,1)$ and the std should be approximately one.

We only report the result for $t=i=r=1$; other values of (t,i,r) give similar results. In both processes that generated x_{it}, the sample means and sample standard deviations are close to zero and one, as is shown in Table 5.1. Figure 5.1 also plots the histogram of the standardized estimates, superimposed with the standard normal density. The histogram is scaled to be a density function. It appears that asymptotic theory provides a very good approximation to the finite sample distributions.

Table 5.1. Sample mean and standard deviations of the standardized estimate.

$N=T$	mean		standard deviation	
	many-mean	factor model	many-mean	factor model
50	0.110	-0.004	1.065	1.273
100	-0.010	0.002	0.944	1.329
200	0.002	-0.017	1.056	1.041

5.2. **Dynamic models.** Next we add lagged variables to allow for dynamics. We consider a model

$$y_{it}=\alpha_i' g_t+x_{it}\lambda_{i,1}' f_{t,1}+y_{i,t-1}\lambda_{i,2}' f_{t,2}+u_{it},$$

where $x_{it}=l_i' w_t+e_{it}$. Strictly speaking, our asymptotic theory does not allow the lagged variable $y_{i,t-1}$ because it does not analytically admit a static factor model structure. But it can be approximated by using a static factor model so long as $|\lambda_{i,2}' f_{t,2}|$ is bounded away from the unit root. So we investigate the finite sample performance of our method in this case.

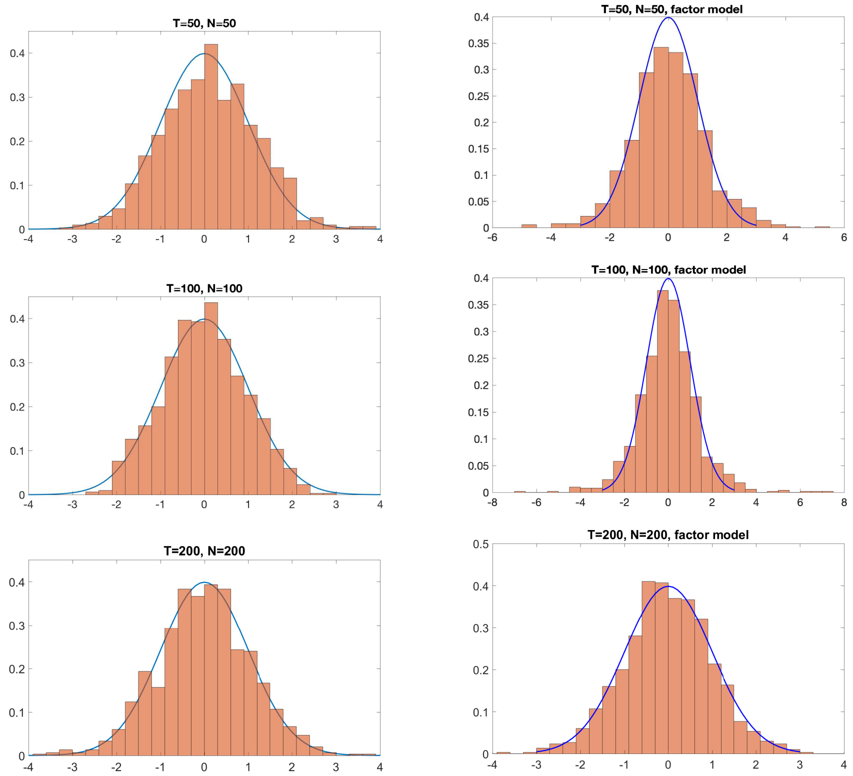

FIGURE 5.1. Histograms of standardized estimates in static models ($\widehat{\theta}_{11} - \theta_{11}$ divided by the estimated asymptotic standard deviation). The left three plots are from the many-mean model; the right three plots are from the factor model. The standard normal density function is superimposed on the histograms.

We calibrate the parameters of the data generating process from an empirical application of the democracy-income model as in Acemoglu et al. (2008). Here y_{it} is the democracy score of country i in period t, and x_{it} is the GDP per capita over the same period. We estimate their model and obtain the parameters for the simulated design. We assume all the factors and loadings are generated from normal distributions, whose mean vectors and covariance matrices are respectively calculated as the sample means and sample covariances of the estimators using the real democracy-income data. The number of factors in $\alpha_i' g_t$ is estimated to be one, while all the other numbers of factors are estimated to be two. The following tables show the calibrated sample means and covariances. The error term u_{it} is generated independently from $\mathcal{N}(0, \sigma^2)$ with $\sigma = 0.1287$ calibrated from the real data.

TABLE 5.2. Calibrated means for the dynamic model

λ_1	λ_2	f_1	f_2	l	w	α	g
-0.893	-0.863	-0.054	-0.056	-0.893	-0.054	0.298	0.005
0.075	0.045	-0.001	-0.001	0.075	-0.001		

TABLE 5.3. Calibrated covariances for the dynamic model

λ_1		λ_2		f_1		f_2		l		w		α	g
0.20	0.06	0.25	0.04	0.33	-0.21	0.45	-0.22	0.20	0.06	0.33	-0.21	0.92	0.01
0.06	1.01	0.04	1.01	-0.21	0.21	-0.22	0.22	0.06	1.01	-0.21	0.21		

We generate y_{i1} independently from $0.3\mathcal{N}(0,1) + 0.497$, whose parameters are calibrated from the real data at time $t = 1$, and then generate y_{it} iteratively. Let $z_{it} := y_{i,t-1}$. Figure 5.2 plots the first twenty eigenvalues of the $N \times N$ sample covariance of z_{it} when $N = T = 100$, and demonstrates one very spiked eigenvalue. Therefore we estimate a one-factor model on z_{it} in our estimation procedure.

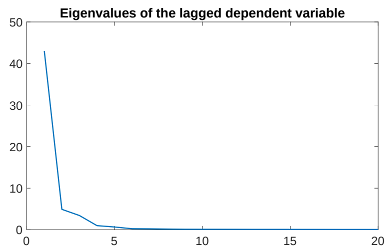

FIGURE 5.2. Sorted eigenvalues of the sample covariance of the lagged dependent variable

We calculate the t-statistic of the estimated $\lambda_{i,1}' f_{t,1}$ and $\lambda_{i,2}' f_{t,2}$ for $t = i = 1$, defined as

$$\frac{\widehat{\theta}_{it} - \theta_{it}}{(\widehat{V}_\lambda + \widehat{V}_f)^{1/2}}.$$

Table 5.4 reports the sample means and standard deviations of the t-statistics, over 1,000 repetitions. Figure 5.2 plots their histograms. Above all, the results are satisfactory. The estimated effect for x_{it} is basically unbiased and distributed as standard normality. On the other hand, the estimated lagged effect for $y_{i,t-1}$ is noticeably biased. As T, N become larger, the bias slowly decreases.

TABLE 5.4. Sample mean and standard deviations of the standardized estimated coefficients.

$N = T$	mean		standard deviation	
	x_{it}	$y_{i,t-1}$	x_{it}	$y_{i,t-1}$
50	0.019	-0.345	1.342	1.140
100	-0.096	-0.267	1.178	1.015
150	-0.010	-0.238	1.117	1.045

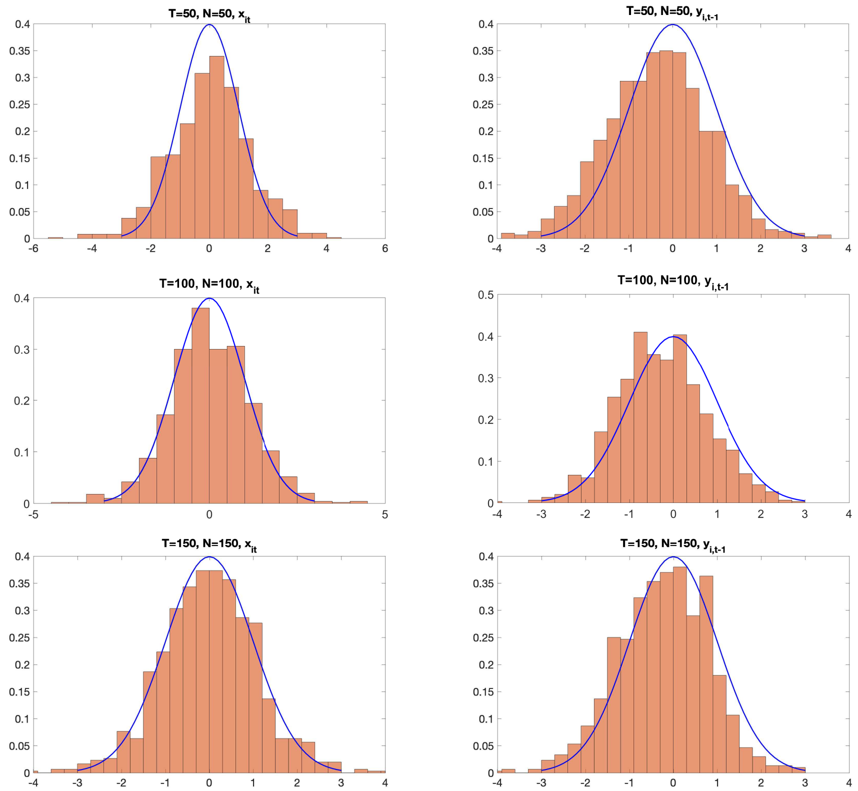

FIGURE 5.3. Histograms of standardized estimates in dynamic models ($\widehat{\theta}_{11} - \theta_{11}$ divided by the estimated asymptotic standard deviation). The left three plots are the effect of $x_{i,t}$; the right three plots are for the effect of $y_{i,t-1}$. The standard normal density function is superimposed on the histograms.

6. AN EMPIRICAL APPLICATION

6.1. **The background.** We illustrate our method in the application of studying the effect on employment of the minimum wage. Previous studies in the literature reach mixed conclusions. For instance, Card and Krueger (1994) find that the minimum wage has a positive effect on the employment using data of restaurants in New Jersey and Eastern Pennsylvania. Dube et al. (2010) concluded negative effects when county-level populations are controlled but also found no adverse employment effects from minimum wage increases when contiguous county-pairs are used for identifications. More recently, Wang et al. (2018) consider slope heterogeneity at the county level:

$$y_{it} = x_{it,1}\theta_i + x_{it,2}\beta_i + \alpha_i + g_t + u_{it} \tag{6.1}$$

where y_{it} is the log employment of county i at period t; $x_{it,1}$ is the log minimum wage, and $x_{it,2}$ is the log population; α_i and g_t are respectively the additive county and time fixed effects. They consider a grouping structure by assuming that θ_i and β_i are clustered by a small number of unknown groups, and reach mixed effects of the minimum wage across different groups.

While these results allow heterogeneity slopes across counties, the minimum wage effects have been treated as time invariant. However, these effects are decided by the equilibrium of both the supply and demand sides, which should vary over time. To illustrate this, we take the Baker County in Oregon for example, which has the highest average minimum wage during 1990-2006. Figure 6.1 plots the log minimum wage and the employment during this period. It is clear that while the minimum wage increases throughout years, the employment rate is very volatile, indicating a possible time-varying effect. In addition, using additive fixed effects in (6.1) may fail to control for complex unobserved heterogeneity correlated with minimum wages. To control for unobservable factors and heterogeneity that may change over time and counties interactively, we allow the interactive effect and estimate the following model:

$$\begin{aligned} y_{it} &= x_{it,1}\theta_{it} + x_{it,2}\beta_{it} + \alpha_i' g_t + u_{it}, \\ \theta_{it} &= \lambda_{i,1}' f_{t,1}, \quad \beta_{it} = \lambda_{i,2}' f_{t,2}. \end{aligned}$$

We use the same county-level data as in Dube et al. (2010) and Wang et al. (2018). [2] The balanced panel contains data of 1378 counties in the US that ranges from the first quarter of 1990 to the second quarter of 2006, so $T = 66$ and $N = 1378$. The preliminary analysis on the eigenvalues of the regressors $x_{it,1}$ and $x_{it,2}$ clearly demonstrates the presence of a single very spiked eigenvalue for each regressor, so we apply the "partial-out" approach by extracting one estimated

[2]We are grateful to Wuyi Wang for sharing the data with us.

factor from each regressor. The tuning parameters for the nuclear-norm regularization are chosen as in Section 2.4, where we estimate the noise variance $\mathsf{Var}(u_{it})$ by first setting it to $\mathsf{Var}(y_{it})$ and updating and iterating it in the tuning parameters. In addition, the numbers of factors in θ_{it} and β_{it} are both selected as one by the algorithm.

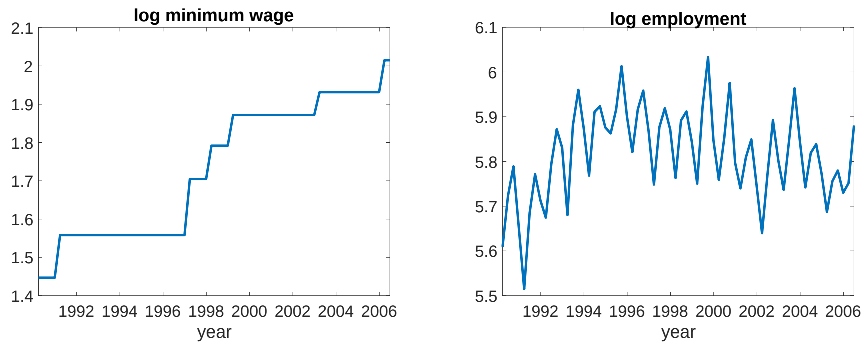

FIGURE 6.1. The minimum wage and employment in Baker County, Oregon. The left panel plots the log minimum wage, and the right panel lots the log employment, starting from the first quarter of 1990 and ending at the second quarter of 2006. Baker County in Oregon is one of the counties with the highest average minimum wage in this period.

6.2. **The results.** For each county, we estimate the minimum wage effect at each period. To present the results, we group the counties according to their averaged minimum wage throughout the time, and obtain three groups: high, medium and low minimum wage. Figure 6.2 plots the averaged effects within each group across time. The estimated effects of counties of either high or medium minimum wages are much more stable over time than counties of low minimum wages; the latter's effects are mostly negative, between -0.67 and -0.26 during 1992 through 1998, but then jump to between 0.64 and 1.78 during 2000 through 2004. In contrast, the estimated effects for counties of high minimum wages are between -0.05 and -0.02 during 1992 through 1998, and between 0.03 and 0.14 during 2000 through 2004. Table 6.1 summarizes the averaged effects (by counties) of the three groups during these two periods.

Our method also allows to study the dynamics of the county specific minimum wage effects. To illustrate this, we now focus on the Baker County in Oregon, which has the highest averaged (over time) minimum wage, and Uinta County in Wyoming, which has the lowest averaged (over time) minimum wage. Figure 6.2

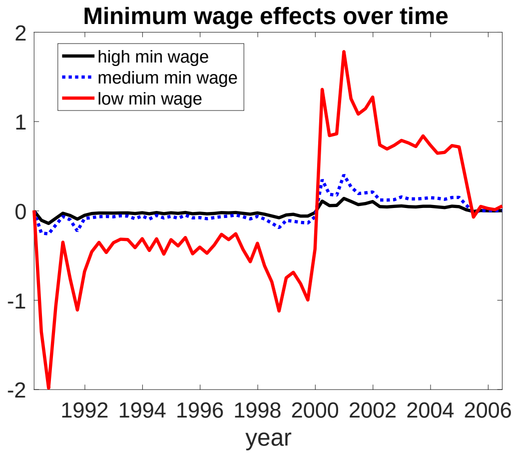

FIGURE 6.2. Averaged (by counties) minimum wage effects in each group

TABLE 6.1. Averaged (by counties) minimum wage effects of three groups

	1992 through 1998			2000 through 2004		
	mean	min	max	mean	min	max
high min wage	-0.026	-0.049	-0.018	0.065	0.037	0.139
medium min wage	-0.069	-0.092	-0.045	0.182	0.121	0.404
low min wage	-0.403	-0.677	-0.256	0.918	0.645	1.782

plots the scatter plot and the least squares regression line between the estimated effects and the minimum wages for each county. In both counties, the effect is generally increasing with respect to the level of the minimum wage. This means, during periods when the minimum wage is higher, the effect is also expected to become larger. We interpret the positive relationship between the minimum wage and its effect as the consequence of the increasing trend of the two throughout the sampling period in both counties.

7. CONCLUSION

We study a panel data model with general heterogeneous effects, in the sense that the slopes are allowed to be varying across both individuals and times, and the interactive fixed effects are allowed. The key assumption for dimension reduction is that the heterogeneous slopes can be expressed as a factor structure so that the

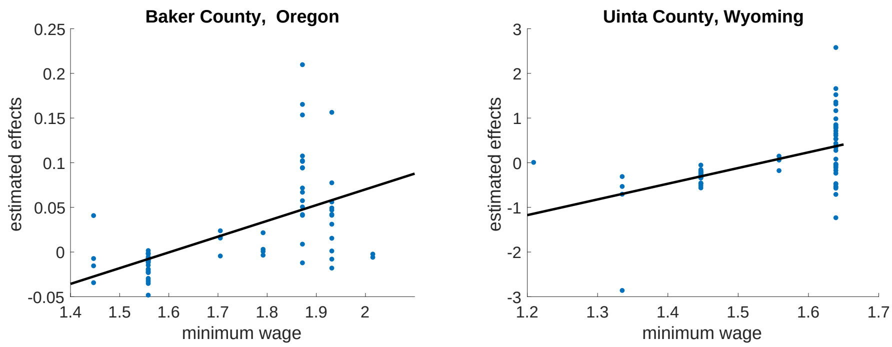

FIGURE 6.3. Scatter plots of estimated effects with respect to the minimum wage. Here dots represents different times. The left panel is for the Baker County (the highest averaged minimum wage), and the right panel is for the Uinta County (the lowest averaged minimum wage). Each plot is fitted using a least squares line.

high-dimensional slope matrix is of low-rank, so can be estimated using nuclear-norm based SVD regression. We show that the inference can be conducted via three steps:

(1) apply the low-rank SVD estimation;

(2) extract eigenvectors from the estimated low-rank matrices, and

(3) run least squares to iteratively estimate the individual and time effect components in the slope matrix.

To properly control for the effect of the penalized low-rank estimation, we argue that this procedure should be embedded with "partial out the mean structure" and "sample splitting". The resulting estimators are asymptotically normal and admit valid asymptotic inferences. In addition, we conduct global homogeneous tests, where under the null, the slopes are either common across individuals, time-invariant, or both.

References

Acemoglu, D., Johnson, S., Robinson, J. A. and Yared, P. (2008). Income and democracy. *American Economic Review* **98** 808–42.

Ahn, S. C., Lee, Y. H. and Schmidt, P. (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics* **174** 1–14.

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

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* **77** 1229–1279.

Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences* **2** 183–202.

Cai, T. T., Liang, T., Rakhlin, A. et al. (2016). Geometric inference for general high-dimensional linear inverse problems. *The Annals of Statistics* **44** 1536–1563.

Candès, E. J. and Tao, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory* **56** 2053–2080.

Card, D. and Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania. *American Economic Review* **84** 772–793.

Dube, A., Lester, T. W. and Reich, M. (2010). Minimum wage effects across state borders: Estimates using contiguous counties. *The Review of Economics and Statistics* **92** 945–964.

Fan, J., Liao, Y. and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements (with discussion). *Journal of the Royal Statistical Society, Series B* **75** 603–680.

Koltchinskii, V., Lounici, K., Tsybakov, A. B. et al. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics* **39** 2302–2329.

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

Moon, R. and Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* **83** 1543–1579.

Negahban, S. and Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics* **39** 1069–1097.

Pesaran, H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* **74** 967–1012.

Recht, B., Fazel, M. and Parrilo, P. A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review* **52** 471–501.

Stock, J. and Watson, M. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* **97** 1167–1179.

Su, L., Jin, S. and Zhang, Y. (2015). Specification test for panel data models with interactive fixed effects. *Journal of Econometrics* **186** 222–244.

Sun, T. and Zhang, C.-H. (2012). Calibrated elastic regularization in matrix completion. In *Advances in Neural Information Processing Systems*.

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027* .

Wang, W. and Fan, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Annals of statistics* **45** 1342.

Wang, W., Su, L. and Phillips, P. C. B. (2018). The heterogeneous effects of the minimum wage on employment across states. *Economics Letters* .

Appendix A. Estimation algorithm for the multivariate case

Consider

$$\begin{aligned} y_{it} &= \sum_{r=1}^{R} x_{it,r}\theta_{it,r} + \alpha_i' g_t + u_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T. \\ \theta_{it,r} &= \lambda_{i,r}' f_{t,r}. \end{aligned}$$

where $(x_{it,1}, ..., x_{it,R})'$ is an R-dimensional vector of covariate. Each coefficient $\theta_{it,r}$ admits a factor structure with $\lambda_{i,r}$ and $f_{t,r}$ as the "loadings" and "factors"; the factors and loadings are $\theta_{it,r}$ specific, but are allowed to have overlap. Here $(R, \dim(\lambda_{i,1}), ..., \dim(\lambda_{i,R}))$ are all assumed fixed.

Suppose $x_{it,r} = \mu_{it,r} + e_{it,r}$. For instance, $\mu_{it,r} = l_{i,r}' w_{t,r}$ also admits a factor structure. Then after partialing out $\mu_{it,r}$, the model can also written as:

$$\dot{y}_{it} = \sum_{r=1}^{R} e_{it,r}\lambda_{i,r}' f_{t,r} + \alpha_i' g_t + u_{it}, \quad \dot{y}_{it} = y_{it} - \sum_{r=1}^{R} \mu_{it,r}\theta_{it,r}$$

As such, it is straightforward to extend the estimation algorithm to the multivariate case. Let X_r be the $N \times T$ matrix of $x_{it,r}$, Let Θ_r be the $N \times T$ matrix of $\theta_{it,r}$. We first estimate these low rank matrices using penalized nuclear-norm regression. We then apply sample splitting, and employ steps 2-4 to iteratively estimate $(f_{t,r}, \lambda_{i,r})$. The formal algorithm is stated as follows.

Algorithm A.1. Estimate $\theta_{it,r}$ as follows.

Step 1. Estimate the number of factors. Run nuclear-norm penalized regression:

$$(\widetilde{M}, \widetilde{\Theta}_r) := \arg\min_{M, \Theta_r} \|Y - M - \sum_{r=1}^{R} X_r \odot \Theta_r\|_F^2 + \nu_0 \|M\|_n + \sum_{r=1}^{R} \nu_r \|\Theta\|_n.$$

Estimate $K_r = \dim(\lambda_{i,r}), K_0 = \dim(\alpha_i)$ by

$$\widehat{K}_r = \sum_i 1\{\psi_i(\widetilde{\Theta}_r) \geq (\nu_r \|\widetilde{\Theta}_r\|)^{1/2}\}, \quad \widehat{K}_0 = \sum_i 1\{\psi_i(\widetilde{M}) \geq (\nu_0 \|\widetilde{M}\|)^{1/2}\}.$$

Step 2. Estimate the structure $x_{it,r} = \mu_{it,r} + e_{it,r}$.

In the many mean model, let $\widehat{e}_{it,r} = x_{it,r} - \frac{1}{T}\sum_{t=1}^{T} x_{it,r}$. In the factor model, use the PC estimator to obtain $(\widehat{l_{i,r}' w_{t,r}}, \widehat{e}_{it,r})$ for all $i = 1, ..., N, t = 1, .., T$ and $r = 1, ..., R$.

Step 3: Sample splitting. Randomly split the sample into $\{1, ..., T\}/\{t\} = I \cup I^c$, so that $|I|_0 = [(T-1)/2]$. Denote by Y_I, $X_{I,r}$ as the $N \times |I|_0$ matrices of $(y_{is}, x_{is,r})$

for observations at $s \in I$. Estimate the low-rank matrices Θ and M as in step 1, with (Y, X_r) replaced with $(Y_I, X_{I,r})$, and obtain $(\widetilde{M}_I, \widetilde{\Theta}_{I,r})$.

Let $\widetilde{\Lambda}_{I,r} = (\widetilde{\lambda}_{1,r}, ..., \widetilde{\lambda}_{N,r})'$ be the $N \times \widehat{K}_r$ matrix, whose columns are defined as $\sqrt{N}$ times the first $\widehat{K}_r$ eigenvectors of $\widetilde{\Theta}_{I,r}\widetilde{\Theta}'_{I,r}$. Let $\widetilde{A}_I = (\widetilde{\alpha}_1, ..., \widetilde{\alpha}_N)'$ be the $N \times \widehat{K}_0$ matrix, whose columns are defined as $\sqrt{N}$ times the first $\widehat{K}_0$ eigenvectors of $\widetilde{M}_I \widetilde{M}'_I$.

Step 4. Estimate the "partial-out" components.

Substitute in $(\widetilde{\alpha}_i, \widetilde{\lambda}_{i,r})$, and define

$$(\widetilde{f}_{s,r}, \widetilde{g}_s) := \arg\min_{f_{s,r}, g_s} \sum_{i=1}^{N} (y_{is} - \widetilde{\alpha}'_i g_s - \sum_{r=1}^{R} x_{is,r} \widetilde{\lambda}'_{i,r} f_{s,r})^2, \quad s \in I^c \cup \{t\}.$$

and

$$(\dot{\lambda}_{i,r}, \dot{\alpha}_i) = \arg\min_{\lambda_{i,r}, \alpha_i} \sum_{s \in I^c \cup \{t\}} (y_{is} - \alpha'_i \widetilde{g}_s - \sum_{r=1}^{R} x_{is,r} \lambda'_{i,r} \widetilde{f}_{s,r})^2, \quad i = 1, ..., N.$$

Step 5. Estimate $(f_{t,r}, \lambda_{i,r})$ for inferences.

Motivated by (2.5), for all $s \in I^c \cup \{t\}$, let

$$(\widehat{f}_{I,s,r}, \widehat{g}_{I,s}) := \arg\min_{f_{s,r}, g_s} \sum_{i=1}^{N} (\widehat{y}_{is} - \widetilde{\alpha}'_i g_s - \sum_{r=1}^{R} \widehat{e}_{is,r} \widetilde{\lambda}'_{i,r} f_{s,r})^2.$$

Fix $i \leq N$,

$$(\widehat{\lambda}_{I,i,r}, \widehat{\alpha}_{I,i}) = \arg\min_{\lambda_{i,r}, \alpha_i} \sum_{s \in I^c \cup \{t\}} (\widehat{y}_{is} - \alpha'_i \widehat{g}_{I,s} - \sum_{r=1}^{R} \widehat{e}_{is,r} \lambda'_{i,r} \widehat{f}_{I,s,r})^2.$$

where $\widehat{y}_{is} = y_{is} - \bar{x}_{i,r} \dot{\lambda}'_i \widetilde{f}_s$ and $\widehat{e}_{is,r} = x_{is,r} - \bar{x}_{i,r}$ in the many mean model, and $\widehat{y}_{is} = y_{is} - \widehat{l'_{i,r} w_{s,r}} \dot{\lambda}'_{i,r} \widetilde{f}_{s,r}$, $\widehat{e}_{is,r} = x_{is,r} - \widehat{l'_{i,r} w_{s,r}}$ in the factor model.

Step 6. Estimate $\theta_{it,r}$. Repeat steps 3-5 with I and I^c exchanged, and obtain $(\widehat{\lambda}_{I^c,i,r}, \widehat{f}_{I^c,s,r} : s \in I \cup \{t\}, i \leq N, r \leq R)$. Define

$$\widehat{\theta}_{it,r} := \frac{1}{2}[\widehat{\lambda}'_{I,i,r} \widehat{f}_{I,t,r} + \widehat{\lambda}'_{I^c,i,r} \widehat{f}_{I^c,t,r}].$$

The asymptotic variance can be estimated by $\widehat{v}_{\lambda,r} + \widehat{v}_{f,r}$, where

$$\begin{aligned}
\widehat{v}_{\lambda,r} &= \frac{1}{2N}(\widehat{\lambda}'_{I,i,r} \widehat{V}_{\lambda,1}^{I-1} \widehat{V}_{\lambda,2}^{I} \widehat{V}_{\lambda,1}^{I-1} \widehat{\lambda}_{I,i,r} + \widehat{\lambda}'_{I^c,i,r} \widehat{V}_{\lambda,1}^{I^c-1} \widehat{V}_{\lambda,2}^{I^c} \widehat{V}_{\lambda,1}^{I^c-1} \widehat{\lambda}_{I^c,i,r}) \\
\widehat{v}_{f,r} &= \frac{1}{2T}(\widehat{f}'_{I,t,r} \widehat{V}_{f,1}^{I-1} \widehat{V}_{f,2}^{I} \widehat{V}_{f,1}^{I-1} \widehat{f}_{I,t,r} + \widehat{f}'_{I^c,t,r} \widehat{V}_{f,1}^{I^c-1} \widehat{V}_{f,2}^{I^c} \widehat{V}_{f,1}^{I^c-1} \widehat{f}_{I^c,t,r})
\end{aligned}$$

with $\widehat{V}_{f,1}^S = \frac{1}{|S|_0}\sum_{s\in S}\widehat{f}_{s,r}\widehat{f}_{s,r}'\widehat{e}_{is,r}^2$, $\widehat{V}_{f,2}^S = \frac{1}{|S|_0}\sum_{s\in S}\widehat{f}_{s,r}\widehat{f}_{s,r}'\widehat{e}_{is}^2\widehat{u}_{is,r}^2$, $\widehat{V}_{\lambda,1}^S = \frac{1}{N}\sum_j\widehat{\lambda}_{j,r}\widehat{\lambda}_{j,r}'\widehat{e}_{jt,r}^2$, and $\widehat{V}_{\lambda,2}^S = \frac{1}{N}\sum_j\widehat{\lambda}_{j,r}\widehat{\lambda}_{j,r}'\widehat{e}_{jt,r}^2\widehat{u}_{jt,r}^2$.

It is also straightforward to extend the univariate asymptotic analysis to the multivariate case, and establish the asymptotic normality for $\widehat{\theta}_{it,r}$. The proof techniques are the same, subjected to more complicated notation. Therefore our proofs below focus on the univariate case.

Appendix B. Proof of Proposition 2.1

Recall that $\Theta_{k+1} = S_{\tau\nu_1/2}(\Theta_k - \tau A_k)$, where

$$A_k = X \odot (X \odot \Theta_k - Y + M_k).$$

By Lemma B.2, set $\Theta = \widetilde{\Theta}$ and $M = \widetilde{M}$, and replace k with subscript m,

$$F(\widehat{\Theta}, \widehat{M}) - F(\Theta_{m+1}, M_{m+1}) \geq \frac{1}{\tau}\left(\|\Theta_{m+1} - \widehat{\Theta}\|_F^2 - \|\Theta_m - \widehat{\Theta}\|_F^2\right).$$

Let $m = 1, ..., k$, and sum these inequalities up, since $F(\Theta_{m+1}, M_{m+1}) \geq F(\Theta_{k+1}, M_{k+1})$ by Lemma B.1,

$$\begin{aligned} kF(\widehat{\Theta}, \widehat{M}) - kF(\Theta_{k+1}, M_{k+1}) &\geq kF(\widehat{\Theta}, \widehat{M}) - \sum_{m=1}^{k} F(\Theta_{m+1}, M_{m+1}) \\ \geq \frac{1}{\tau}\left(\|\Theta_{k+1} - \widehat{\Theta}\|_F^2 - \|\Theta_1 - \widehat{\Theta}\|_F^2\right) &\geq -\frac{1}{\tau}\|\Theta_1 - \widehat{\Theta}\|_F^2. \end{aligned}$$

Q.E.D.

The above proof depends on the following lemmas.

Lemma B.1. *We have: (i)*

$$\begin{aligned} \Theta_{k+1} &= \arg\min_{\Theta} p(\Theta, \Theta_k, M_k) + \nu_1\|\Theta\|_n, \\ p(\Theta, \Theta_k, M_k) &:= \tau^{-1}\|\Theta_k - \Theta\|_F^2 - 2\mathsf{tr}\left((\Theta_k - \Theta)' A_k\right). \end{aligned}$$

(ii) For any $\tau \in (0, 1/\max x_{it}^2)$,

$$F(\Theta, M_k) \leq p(\Theta, \Theta_k, M_k) + \nu_1\|\Theta\|_n + \nu_2\|M_k\|_n + \|Y - M_k - X \odot \Theta_k\|_F^2.$$

(iii) $F(\Theta_{k+1}, M_{k+1}) \leq F(\Theta_{k+1}, M_k) \leq F(\Theta_k, M_k)$.

Proof. (i) We have $\|\Theta_k - \tau A_k - \Theta\|_F^2 = \|\Theta_k - \Theta\|_F^2 + \tau^2\|A_k\|_F^2 - 2\mathsf{tr}[(\Theta_k - \Theta)'A_k]\tau$. So

$$\begin{aligned} & \arg\min_{\Theta} \|\Theta_k - \tau A_k - \Theta\|_F^2 + \tau\nu_1\|\Theta\|_n \\ = \ & \arg\min_{\Theta} \|\Theta_k - \Theta\|_F^2 - 2\mathsf{tr}[(\Theta_k - \Theta)'A_k]\tau + \tau\nu_1\|\Theta\|_n \\ = \ & \arg\min_{\Theta} \tau \cdot p(\Theta, \Theta_k, M_k) + \tau\nu_1\|\Theta\|_n. \end{aligned}$$

On the other hand, it is well known that $\Theta_{k+1} = S_{\tau\nu_1/2}(\Theta_k - \tau A_k)$ is the solution to the first problem in the above equalities (Ma et al., 2011). This proves (i).

(ii) Note that for $\Theta_k = (\theta_{k,it})$ and $\Theta = (\theta_{it})$, and any $\tau^{-1} > \max_{it} x_{it}^2$, we have

$$\|X \odot (\Theta_k - \Theta)\|_F^2 = \sum_{it} x_{it}^2(\theta_{k,it} - \theta_{it})^2 < \tau^{-1}\|\Theta_k - \Theta\|_F^2.$$

So

$$\begin{aligned} F(\Theta, M_k) &= \|Y - M_k - X \odot \Theta\|_F^2 + \nu_2\|M_k\|_n + \nu_1\|\Theta\|_n \\ &= \|Y - M_k - X \odot \Theta_k\|_F^2 + \|X \odot (\Theta_k - \Theta)\|_F^2 - 2\mathsf{tr}[A_k'(\Theta_k - \Theta)] \\ &\quad + \nu_2\|M_k\|_n + \nu_1\|\Theta\|_n \\ &\leq \|Y - M_k - X \odot \Theta_k\|_F^2 + \tau^{-1}\|\Theta_k - \Theta\|_F^2 - 2\mathsf{tr}[A_k'(\Theta_k - \Theta)] \\ &\quad + \nu_2\|M_k\|_n + \nu_1\|\Theta\|_n \\ &= p(\Theta, \Theta_k, M_k) + \|Y - M_k - X \odot \Theta_k\|_F^2 + \nu_2\|M_k\|_n + \nu_1\|\Theta\|_n. \end{aligned}$$

(iii) By definition, $p(\Theta_k, \Theta_k, M_k) = 0$. So

$$\begin{aligned} F(\Theta_{k+1}, M_{k+1}) &= \|Y - X \odot \Theta_{k+1} - M_{k+1}\|_F^2 + \nu_1\|\Theta_{k+1}\|_n + \nu_2\|M_{k+1}\|_n \\ &\leq^{(a)} \|Y - X \odot \Theta_{k+1} - M_k\|_F^2 + \nu_1\|\Theta_{k+1}\|_n + \nu_2\|M_k\|_n \\ &= F(\Theta_{k+1}, M_k) \\ &\leq^{(b)} p(\Theta_{k+1}, \Theta_k, M_k) + \nu_1\|\Theta_{k+1}\|_n + \nu_2\|M_k\|_n + \|Y - M_k - X \odot \Theta_k\|_F^2 \\ &\leq^{(c)} p(\Theta_k, \Theta_k, M_k) + \nu_1\|\Theta_k\|_n + \nu_2\|M_k\|_n + \|Y - M_k - X \odot \Theta_k\|_F^2 \\ &= F(\Theta_k, M_k). \end{aligned}$$

(a) is due to the definition of M_{k+1}; (b) is due to (ii); (c) is due to (i). Q.E.D.

Lemma B.2. *For any $\tau \in (0, 1/\max x_{it}^2)$, any (Θ, M) and any $k \geq 1$,*

$$F(\Theta, M) - F(\Theta_{k+1}, M_{k+1}) \geq \frac{1}{\tau}\left(\|\Theta_{k+1} - \Theta\|_F^2 - \|\Theta_k - \Theta\|_F^2\right).$$

Proof. The proof is similar to that of Lemma 2.3 of Beck and Teboulle (2009), with the extension that M_{k+1} is updated after Θ_{k+1}. The key difference here is that, while an update to M_{k+1} is added to the iteration, we show that the lower bound does not depend on M_{k+1} or M_k. Therefore, the convergence property of the algorithm depends mainly on the step of updating Θ.

Let $\partial\|A\|_n$ be an element that belongs to the subgradient of $\|A\|_n$. Note that $\partial\|A\|_n$ is convex in A. Also, $\|Y - X \odot \Theta - M\|_F^2$ is convex in (Θ, M), so for any Θ, M, we have the following three inequalities:

$$\begin{aligned}
\|Y - X \odot \Theta - M\|_F^2 &\geq \|Y - X \odot \Theta_k - M_k\|_F^2 - 2\mathsf{tr}[(\Theta - \Theta_k)'(X \odot (Y - X \odot \Theta_k - M_k))] \\
&\quad -2\mathsf{tr}[(M - M_k)'(Y - X \odot \Theta_k - M_k)] \\
\nu_1\|\Theta\|_n &\geq \nu_1\|\Theta_{k+1}\|_n + \nu_1\mathsf{tr}[(\Theta - \Theta_{k+1})'\partial\|\Theta_{k+1}\|_n] \\
\nu_2\|M\|_n &\geq \nu_2\|M_k\|_n + \nu_2\mathsf{tr}[(M - M_k)'\partial\|M_k\|_n].
\end{aligned}$$

In addition,

$$\begin{aligned}
-F(\Theta_{k+1}, M_{k+1}) &\geq -F(\Theta_{k+1}, M_k) \\
&\geq -p(\Theta_{k+1}, \Theta_k, M_k) - \nu_1\|\Theta_{k+1}\|_n - \nu_2\|M_k\|_n - \|Y - M_k - X \odot \Theta_k\|_F^2.
\end{aligned}$$

where the two inequalities are due to Lemma B.1. Sum up the above inequalities,

$$\begin{aligned}
& F(\Theta, M) - F(\Theta_{k+1}, M_{k+1}) \geq (A) \\
(A) \quad := \quad & -2\mathsf{tr}[(\Theta - \Theta_k)'(X \odot (Y - X \odot \Theta_k - M_k))] - 2\mathsf{tr}[(M - M_k)'(Y - X \odot \Theta_k - M_k)] \\
& +\nu_1\mathsf{tr}[(\Theta - \Theta_{k+1})'\partial\|\Theta_{k+1}\|_n] + \nu_2\mathsf{tr}[(M - M_k)'\partial\|M_k\|_n] - p(\Theta_{k+1}, \Theta_k, M_k).
\end{aligned}$$

We now simplify (A). Since $k \geq 1$, both M_k and Θ_{k+1} should satisfy the KKT condition. By Lemma B.1, they are:

$$\begin{aligned}
0 &= \nu_1\partial\|\Theta_{k+1}\|_n - \tau^{-1}2(\Theta_k - \Theta_{k+1}) + 2A_k \\
0 &= \nu_2\partial\|M_k\|_n - 2(Y - X \odot \Theta_k - M_k).
\end{aligned}$$

Plug in, we have

$$\begin{aligned}
(A) &= \tau^{-1}2\mathsf{tr}[(\Theta - \Theta_{k+1})'(\Theta_k - \Theta_{k+1})] - \tau^{-1}\left\|\Theta_k - \Theta_{k+1}\right\|_F^2 \\
&= \frac{1}{\tau}\left(\|\Theta_{k+1} - \Theta\|_F^2 - \|\Theta_k - \Theta\|_F^2\right).
\end{aligned}$$

Q.E.D.

Appendix C. Proof of Proposition 3.1

C.1. **Level of the Score.**

Lemma C.1. *In the presence of serial correlations in x_{it} (Assumption 4.1), $\|X \odot U\| = O_P(\sqrt{N+T}) = \|U\|$. In addition, the chosen ν_2, ν_1 in Section 2.4 are $O_P(\sqrt{N+T})$.*

Proof. The assumption that Ω_{NT} contains independent sub-Gaussian columns ensures that, by the eigenvalue-concentration inequality for sub-Gaussian random vectors (Theorem 5.39 of Vershynin (2010)):

$$\|\Omega_{NT}\Omega_{NT}' - \mathsf{E}\Omega_{NT}\Omega_{NT}'\| = O_P(\sqrt{NT} + N).$$

In addition, let w_i be the $T \times 1$ vector of $\{x_{it}u_{it} : t \leq T\}$. We have, for each (i, j, t, s),

$$\mathsf{E}(w_i w_j')_{s,t} = \mathsf{E}x_{it}x_{js}u_{it}u_{js} = \begin{cases} \mathsf{E}x_{it}^2 u_{it}^2, & i = j, t = s \\ 0, & \text{otherwise} \end{cases}$$

due to the conditional cross-sectional and serial independence in u_{it}. Then for the (i, j)'th entry of $\mathsf{E}\Omega_{NT}\Omega_{NT}'$,

$$\begin{aligned} (\mathsf{E}\Omega_{NT}\Omega_{NT}')_{i,j} &= (\mathsf{E}(X \odot U)\Sigma_T^{-1}(X \odot U)')_{i,j} \\ &= \mathsf{E}w_i'\Sigma_T^{-1}w_j = tr(\Sigma_T^{-1}\mathsf{E}w_j w_i') = \begin{cases} \sum_{t=1}^T (\Sigma_T^{-1})_{tt}\mathsf{E}x_{it}^2 u_{it}^2, & i = j \\ 0, & i \neq j. \end{cases} \end{aligned}$$

Hence $\|\mathsf{E}\Omega_{NT}\Omega_{NT}'\| \leq O(T)$. This implies $\|\Omega_{NT}\Omega_{NT}'\| \leq O(T + N)$. Hence $\|X \odot U\| \leq \|\Omega_{NT}\|\|\Sigma_T^{1/2}\| \leq O_P(\max\{\sqrt{N}, \sqrt{T}\})$. The rate for $\|U\|$ follows from the same argument. The second claim that ν_2, ν_1 satisfy the same rate constraint follows from the same argument, by replacing U with Z, and Assumption 4.1 is still satisfied by Z and $X \odot Z$. Q.E.D.

C.2. **Useful Claims.** The proof of Proposition 3.1 uses some claims that are proved in the following lemma. Let us first recall the notations. Define $U_2D_2V_2' = \Theta_I^0$ and $U_1D_1V_1' = M_I^0$ as the singular value decompositions of the true values Θ_I^0 and M_I^0. Further decompose, for $j = 1, 2$,

$$U_j = (U_{j,r}, U_{j,c}), \quad V_j = (V_{j,r}, V_{j,c})$$

Here $(U_{j,r}, V_{j,r})$ corresponds to the nonzero singular values, while $(U_{j,c}, V_{j,c})$ corresponds to the zero singular values. In addition, for any $N \times T/2$ matrix Δ, let

$$\mathcal{P}_j(\Delta) = U_{j,c}U_{j,c}'\Delta V_{j,c}V_{j,c}', \quad \mathcal{M}_j(\Delta) = \Delta - \mathcal{P}_j(\Delta).$$

Here $U_{j,c}U_{j,c}'$ and $V_{j,c}V_{j,c}'$ respectively are the projection matrices onto the columns of $U_{j,c}$ and $V_{j,c}$. Therefore, $\mathcal{M}_1(\cdot)$ and $\mathcal{M}_2(\cdot)$ can be considered as the projection matrices onto the "low-rank" spaces of Θ_I^0 and M_I^0 respectively, and $\mathcal{P}_1(\cdot)$ and $\mathcal{P}_2(\cdot)$ ar projections onto their orthogonal spaces.

Lemma C.2 (claims)**.** *Same results below also apply to $\mathcal{P}_2$ Θ_I^0, and samples on I^c. For any matrix Δ,*

(i) $\|\mathcal{P}_1(\Delta) + M_I^0\|_n = \|\mathcal{P}_1(\Delta)\|_n + \|M_I^0\|_n$.

(ii) $\|\Delta\|_F^2 = \|\mathcal{M}_1(\Delta)\|_F^2 + \|\mathcal{P}_1(\Delta)\|_F^2$

(iii) $\mathsf{rank}(\mathcal{M}_1(\Delta)) \leq 2K_1$*, where* $K_1 = \mathsf{rank}(M_I^0)$.

(iv) $\|\Delta\|_F^2 = \sum_j \sigma_j^2$ *and* $\|\Delta\|_n^2 \leq \|\Delta\|_F^2 \mathsf{rank}(\Delta)$*, with σ_j as the singular values of Δ.*

(v) For any Δ_1, Δ_2, $|\mathsf{tr}(\Delta_1\Delta_2)| \leq \|\Delta_1\|_n\|\Delta_2\|$*, Here $\|.\|$ denotes the operator norm.*

Proof. (i) Note that $M_I^0 = U_{1,r}D_{1,r}V_{1,r}'$ where $D_{1,r}$ are the subdiagonal matrix of nonzero singular values. The claim follows from Lemma 2.3 of Recht et al. (2010).

(ii) Write

$$U_1'\Delta V_1 = \begin{pmatrix} A & B \\ C & U_{1,c}'\Delta V_{1,c} \end{pmatrix} = \begin{pmatrix} A & B \\ C & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & U_{1,c}'\Delta V_{1,c} \end{pmatrix} := H_2 + H_1.$$

Then $\mathcal{P}_1(\Delta) = U_1H_1V_1'$ and $\mathcal{M}_1(\Delta) = U_1H_2V_1'$. So

$$\|\mathcal{P}_1(\Delta)\|_F^2 = \mathsf{tr}(U_1H_1V_1'V_1H_1'U_1') = \mathsf{tr}(H_1H_1') = \|H_1\|_F^2.$$

Similarly, $\|\mathcal{M}_1(\Delta)\|_F^2 = \|H_2\|_F^2$. So

$$\|H_2\|_F^2 + \|H_1\|_F^2 = \|U_1'\Delta V_1\|_F^2 = \|\Delta\|_F^2.$$

(iii) This is Lemma 1 of Negahban and Wainwright (2011).

(iv) The first is a basic equality, and the second follows from the Cauchy-Schwarz inequality.

(v) Let $UDV' = \Delta_1$ be the SVD of Δ_1, then

$$|\mathsf{tr}(\Delta_1\Delta_2)| = |\sum_i D_{ii}(V'\Delta_2 U)_{ii}| \leq \max_i |(V'\Delta_2 U)_{ii}| \sum_i D_{ii} \leq \|\Delta_1\|_n\|\Delta_2\|.$$

C.3. **Proof of Proposition 3.1: convergence of $\widetilde{\Theta}_S, \widetilde{M}_S$.** In the proof below, we set $S = I$, that is, on the splitted sample $t \in I$, and set $T_0 = |I|_0$. The proof carries over to $S = I^c$ or $S = I \cup I^c$. We surpass the subscript S for notational simplicity. Let $\Delta_1 = \widetilde{M} - M$ and $\Delta_2 = \widetilde{\Theta} - \Theta$. Then

$$\|Y - \widetilde{M} - X \odot \widetilde{\Theta}\|_F^2 = \|\Delta_1 + X \odot \Delta_2\|_F^2 + \|U\|_F^2 - 2\mathsf{tr}[U'(\Delta_1 + X \odot \Delta_2)].$$

Note that $\mathsf{tr}(U'(X \odot \Delta_2)) = \mathsf{tr}(\Delta_2'(X \odot U))$. Thus by claim (v),

$$\begin{aligned} |2\mathsf{tr}[U'(\Delta_1 + X \odot \Delta_2)]| &\leq 2\|U\|\|\Delta_1\|_n + 2\|X \odot U\|\|\Delta_2\|_n \\ &\leq (1-c)\nu_2\|\Delta_1\|_n + (1-c)\nu_1\|\Delta_2\|_n. \end{aligned}$$

Thus $\|Y - \widetilde{M} - X \odot \widetilde{\Theta}\|_F^2 \leq \|Y - M - X \odot \Theta\|_F^2$ (evalued at the true parameters) implies

$$\begin{aligned} \|\Delta_1 + X \odot \Delta_2\|_F^2 + \nu_2\|\widetilde{M}\|_n + \nu_1\|\widetilde{\Theta}\|_n &\leq (1-c)\nu_2\|\Delta_1\|_n + (1-c)\nu_1\|\Delta_2\|_n \\ &\quad + \nu_2\|M\|_n + \nu_1\|\Theta\|_n. \end{aligned}$$

Now

$$\begin{aligned} \|\widetilde{M}\|_n &= \|\Delta_1 + M\|_n = \|M + \mathcal{P}_1(\Delta_1) + \mathcal{M}_1(\Delta_1)\|_n \\ &\geq \|M + \mathcal{P}_1(\Delta_1)\|_n - \|\mathcal{M}_1(\Delta_1)\|_n \\ &= \|M\|_n + \|\mathcal{P}_1(\Delta_1)\|_n - \|\mathcal{M}_1(\Delta_1)\|_n, \end{aligned}$$

where the last equality follows from claim (i). Similar lower bound applies to $\|\widetilde{\Theta}\|_n$. Therefore,

$$\begin{aligned} &\|\Delta_1 + X \odot \Delta_2\|_F^2 + c\nu_2\|\mathcal{P}_1(\Delta_1)\|_n + c\nu_1\|\mathcal{P}_2(\Delta_2)\|_n \\ \leq\ &(2-c)\nu_2\|\mathcal{M}_1(\Delta_1)\|_n + (2-c)\nu_1\|\mathcal{M}_2(\Delta_2)\|_n. \end{aligned} \tag{C.1}$$

In the case U is Gaussian, $\|U\|$ and $\|X \odot U\| \asymp \max\{\sqrt{N}, \sqrt{T}\}$, while in the more general case, set $\nu_2 \asymp \nu_1 \asymp \max\{\sqrt{N}, \sqrt{T}\}$. Thus the above inequality implies $(\Delta_1, \Delta_2) \in \mathcal{C}(a)$ for some $a > 0$. Thus apply Assumption 3.1 and claims to (C.1) , for a generic $C > 0$,

$$\begin{aligned} \|\Delta_1\|_F^2 + \|\Delta_2\|_F^2 &\leq C\nu_2\|\mathcal{M}_1(\Delta_1)\|_n + C\nu_1\|\mathcal{M}_2(\Delta_2)\|_n \\ &\leq^{\text{claim (iv)}} C\nu_2\|\mathcal{M}_1(\Delta_1)\|_F\sqrt{\mathsf{rank}(\mathcal{M}_2(\Delta_1))} \\ &\quad + C\nu_1\|\mathcal{M}_2(\Delta_2)\|_F\sqrt{\mathsf{rank}(\mathcal{M}_2(\Delta_2))} \\ &\leq^{\text{claim (iii)}} C\nu_2\|\mathcal{M}_1(\Delta_1)\|_F\sqrt{2K_1} + C\nu_1\|\mathcal{M}_2(\Delta_2)\|_F\sqrt{2K_2} \\ &\leq^{\text{claim (ii)}} C\nu_2\|\Delta_1\|_F + C\nu_1\|\Delta_2\|_F \\ &\leq C\max\{\nu_2, \nu_1\}\sqrt{\|\Delta_1\|_F^2 + \|\Delta_2\|_F^2}. \end{aligned}$$

Thus $\|\Delta_1\|_F^2 + \|\Delta_2\|_F^2 \leq C(\nu_2^2 + \nu_1^2)$.

C.4. **Proof of Proposition 3.1: convergence of $\widetilde{\Lambda}_S, \widetilde{A}_S$.** We proceed the proof in the following steps.

step 1: bound the eigenvalues

Replace $\nu_2^2 + \nu_1^2$ with $O_P(N+T)$, then

$$\|\widetilde{\Theta}_S - \Theta_S\|_F^2 = O_P(N+T).$$

Let $S_f = \frac{1}{T_0}\sum_{t\in I} f_t f_t'$, $\Sigma_f = \frac{1}{T}\sum_{t=1}^T f_t f_t'$ and $S_\Lambda = \frac{1}{N}\Lambda'\Lambda$. Let $\psi_{I,1}^2 \geq ... \geq \psi_{I,K_1}^2$ be the K_1 nonzero eigenvalues of $\frac{1}{NT_0}\Theta_I\Theta_I' = \frac{1}{N}\Lambda S_f \Lambda'$. Let $\widetilde{\psi}_1^2 \geq ... \geq \widetilde{\psi}_{K_2}^2$ be the first K_2 nonzero singular values of $\frac{1}{NT_0}\widetilde{\Theta}_I\widetilde{\Theta}_I'$. Also, let ψ_j^2 be the j th largest eigenvalue of $\frac{1}{N}\Lambda\Sigma_f\Lambda'$. Note that $\psi_1^2 ... \psi_{K_1}^2$ are the same as the eigenvalues of $\Sigma_f^{1/2} S_\Lambda \Sigma_f^{1/2}$. Hence by Assumption 3.3, there are constants $c_1, ..., c_{K_1} > 0$, so that

$$\psi_j^2 = c_j, \quad j = 1, ..., K_1.$$

Then by Weyl's theorem, for $j = 1, ..., \min\{T_0, N\}$, with the assumption that $\|S_f - \Sigma_f\| = O_P(\frac{1}{\sqrt{T}})$, $|\psi_{I,j}^2 - \psi_j^2| \leq \frac{1}{N}\|\Lambda(S_f - \Sigma_f)\Lambda'\| \leq O(1)\|S_f - \Sigma_f\| = O_P(\frac{1}{\sqrt{T}})$. This also implies $\|\Theta_I\| = \psi_{I,1}\sqrt{NT_0} = \sqrt{(c_1 + o_P(1))T_0 N}$.

Still by Weyl's theorem, for $j = 1, ..., \min\{T_0, N\}$,

$$\begin{aligned} |\widetilde{\psi}_j^2 - \psi_{I,j}^2| &\leq \frac{1}{NT_0}\|\widetilde{\Theta}_I\widetilde{\Theta}_I' - \Theta_I\Theta_I'\| \\ &\leq \frac{2}{NT_0}\|\Theta_I\|\|\widetilde{\Theta}_I - \Theta_I\| + \frac{1}{NT_0}\|\widetilde{\Theta}_I - \Theta_I\|^2 = O_P(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}). \end{aligned}$$

implying

$$|\widetilde{\psi}_j^2 - \psi_j^2| = O_P(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}).$$

Then for all $j \leq K_1$, with probability approaching one,

$$\begin{aligned} |\psi_{j-1}^2 - \widetilde{\psi}_j^2| &\geq |\psi_{j-1}^2 - \psi_j^2| - |\psi_j^2 - \widetilde{\psi}_j^2| \geq (c_{j-1} - c_j)/2 \\ |\widetilde{\psi}_j^2 - \psi_{j+1}^2| &\geq |\psi_j^2 - \psi_{j+1}^2| - |\widetilde{\psi}_j^2 - \psi_j^2| \geq (c_j - c_{j+1})/2 \end{aligned} \tag{C.2}$$

with $\psi_{K_1+1}^2 = c_{K_1+1} = 0$ because $\Theta_I\Theta_I'$ has at most K_1 nonzero eigenvalues.

step 2: characterize the eigenvectors

Next, we show that there is a $K_1 \times K_1$ matrix H_1, so that the columns of $\frac{1}{\sqrt{N}}\Lambda H_1$ are the first K_1 eigenvectors of $\Lambda\Sigma_f\Lambda'$. Let $L = S_\Lambda^{1/2}\Sigma_f S_\Lambda^{1/2}$. Let R be a $K_1 \times K_1$ matrix whose columns are the eigenvectors of L. Then $D = R'LR$ is a diagonal matrix of the eigenvalues of L that are distinct nonzeros according to Assumption

3.3. Let $H_1 = S_\Lambda^{-1/2} R$. Then

$$\begin{aligned}\frac{1}{N}\Lambda\Sigma_f\Lambda'\Lambda H_1 &= \Lambda S_\Lambda^{-1/2} S_\Lambda^{1/2}\Sigma_f S_\Lambda^{1/2} S_\Lambda^{1/2} H_1 = \Lambda S_\Lambda^{-1/2} RR'LR \\ &= \Lambda H_1 D.\end{aligned}$$

Now $\frac{1}{N}(\Lambda H_1)'\Lambda H_1 = H_1' S_\Lambda H_1 = R'R = I$. So the columns of $\Lambda H_1/\sqrt{N}$ are the eigenvectors of $\Lambda\Sigma_f\Lambda'$, corresponding to the eigenvalues in D.

Importantly, the rotation matrix H_1, by definition, depends only on S_Λ, Σ_f, which is time-invariant, and does not depend on the splitted sample.

step 3: prove the convergence

We first assume $\widehat{K}_1 = K_1$. The proof of the consistency is given in step 4 below. Once this is true, then the following argument can be carried out conditional on the event $\widehat{K}_1 = K_1$. Apply Davis-Kahan sin-theta inequality, and by (C.2),

$$\begin{aligned}\|\frac{1}{\sqrt{N}}\widetilde{\Lambda}_I - \frac{1}{\sqrt{N}}\Lambda H_1\|_F &\leq \frac{\frac{1}{N}\|\Lambda\Sigma_f\Lambda' - \frac{1}{T_0}\widetilde{\Theta}_I\widetilde{\Theta}_I'\|}{\min_{j\leq K_2}\min\{|\psi_{j-1}^2 - \widetilde{\psi}_j^2|, |\widetilde{\psi}_j^2 - \psi_{j+1}^2|\}} \\ \leq\ & O_P(1)\frac{1}{N}\|\Lambda\Sigma_f\Lambda' - \frac{1}{T_0}\widetilde{\Theta}_I\widetilde{\Theta}_I'\| \\ \leq\ & O_P(1)\frac{1}{N}\|\Lambda(\Sigma_f - S_f)\Lambda'\| + \frac{1}{NT_0}\|\Theta_I\Theta_I' - \widetilde{\Theta}_I\widetilde{\Theta}_I'\| = O_P(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{N}}).\end{aligned}$$

step 4: prove $P(\widehat{K}_1 = K_1) = 1$.

Note that $\psi_j(\widetilde{\Theta}) = \widetilde{\psi}_j\sqrt{NT}$. By step 1, for all $j \leq K_1$, $\widetilde{\psi}_j^2 \geq c_j - o_P(1) \geq c_j/2$ with probability approaching one. Also, $\widetilde{\psi}_{K_1+1}^2 \leq O_P(T^{-1/2} + N^{-1/2})$, implying that

$$\min_{j\leq K_1}\psi_j(\widetilde{\Theta}) \geq c_{K_1}\sqrt{NT}/2, \quad \max_{j>K_1}\psi_j(\widetilde{\Theta}) \leq O_P(T^{-1/2} + N^{-1/2})\sqrt{NT}.$$

In addition, $(\nu_2\|\widetilde{\Theta}\|)^{1/2} \asymp (\sqrt{N+T}\psi_1(\widetilde{\Theta}))^{1/2} \asymp (\sqrt{N+T}\sqrt{NT})^{1/2}$. Thus

$$\min_{j\leq K_1}\psi_j(\widetilde{\Theta}) \geq (\nu_2\|\widetilde{\Theta}\|)^{1/2}, \quad \max_{j>K_1}\psi_j(\widetilde{\Theta}) \leq o_P(1)(\nu_2\|\widetilde{\Theta}\|)^{1/2}.$$

This proves the consistency of $\widehat{K}_1$.

Finally, the proof of the convergence for $\widetilde{A}_I$ and the consistency of $\widehat{K}_2$ follows from the same argument. Q.E.D.

Appendix D. Proof of Theorems 3.1 and 3.2

In the many mean model

$$x_{it} = \mu_i + e_{it},$$

we write $\widehat{e}_{it} = x_{it} - \bar{x}_i$, $\widehat{\mu}_{it} = \bar{x}_i$ and $\mu_{it} = \mu_i$. In the factor model

$$x_{it} = l_i' w_t + e_{it},$$

we write $\widehat{e}_{it} = x_{it} - \widehat{l_i' w_t}$, $\widehat{\mu}_{it} = \widehat{l_i' w_t}$ and $\mu_{it} = l_i' w_t$. The proof in this section works for both models.

Let

$$C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}.$$

First recall that $(\widetilde{f}_s, \dot{\lambda}_i)$ are computed as the preliminary estimators in step 3. The main technical requirement of these estimators is that their estimation effects are negligible, specifically, there is a rotation matrix H_1 that is independent of the sample splitting, for each fixed $s \in I^c \cup \{t\}$, and fixed $i \leq N$,

$$\begin{aligned} \frac{1}{N}\sum_j (H_1'\lambda_j - \dot{\lambda}_j)e_{js} &= O_P(C_{NT}^{-2}), \\ \frac{1}{T}\sum_{s \in I^c \cup \{t\}} f_s(H_1^{-1}f_s - \widetilde{f}_s)e_{is} &= O_P(C_{NT}^{-2}). \end{aligned}$$

These are given in Lemmas E.2 and E.4 below for the many mean model, and Lemmas F.4 and F.6 for the factor model.

D.1. Behavior of $\widehat{f}_t$.

Recall that for each $t \notin I$,

$$(\widehat{f}_{I,t}, \widehat{g}_{I,t}) := \arg\min_{f_t, g_t} \sum_{i=1}^N (\widehat{y}_{it} - \widetilde{\alpha}_i' g_t - \widehat{e}_{it}\widetilde{\lambda}_i' f_t)^2.$$

For notational simplicity, we simply write $\widehat{f}_t = \widehat{f}_{I,t}$ and $\widehat{g}_t = \widehat{g}_{I,t}$, but keep in mind that $\widetilde{\alpha}$ and $\widetilde{\lambda}$ are estimated through the low rank estimations on data I. Note that $\widetilde{\lambda}_i$ consistently estimates λ_i up to a rotation matrix H_1', so $\widehat{f}_t$ is consistent for $H_1^{-1}f_t$. However, as we shall explain below, it is difficult to establish the asymptotic normality for $\widehat{f}_t$ centered at $H_1^{-1}f_t$. Instead, we obtain a new centering quantity, and obtain an expansion for

$$\sqrt{N}(\widehat{f}_t - H_f f_t)$$

with a new rotation matrix H_f that is also independent of t. For the purpose of inference for θ_{it}, this is sufficient.

Let $\widehat{w}_{it} = (\widetilde{\lambda}_i'\widehat{e}_{it}, \widetilde{\alpha}_i')'$, and $\widehat{B}_t = \frac{1}{N}\sum_i \widehat{w}_{it}\widehat{w}_{it}'$. Define $w_{it} = (\lambda_i' e_{it}, \alpha_i')'$, and

$$\widehat{Q}_t = \frac{1}{N}\sum_i \widehat{w}_{it}(\mu_{it}\lambda_i' f_t - \widehat{\mu}_{it}\dot{\lambda}_i'\widetilde{f}_t + u_{it}).$$

We have

$$\begin{aligned}\begin{pmatrix}\widehat{f}_t\\ \widehat{g}_t\end{pmatrix} &= \widehat{B}_t^{-1}\frac{1}{N}\sum_i \widehat{w}_{it}(y_{it} - \widehat{\mu}_{it}\dot{\lambda}_i'\widetilde{f}_t)\\ &= \begin{pmatrix}H_1^{-1}f_t\\ H_2^{-1}g_t\end{pmatrix} + \widehat{B}_t^{-1}\widehat{S}_t\begin{pmatrix}H_1^{-1}f_t\\ H_2^{-1}g_t\end{pmatrix} + \widehat{B}_t^{-1}\widehat{Q}_t \qquad \text{(D.1)}\end{aligned}$$

where

$$\widehat{S}_t = \frac{1}{N}\sum_i \begin{pmatrix}\widetilde{\lambda}_i\widehat{e}_{it}(\lambda_i' H_1 e_{it} - \widetilde{\lambda}_i'\widehat{e}_{it}) & \widetilde{\lambda}_i\widehat{e}_{it}(\alpha_i' H_2 - \widetilde{\alpha}_i')\\ \widetilde{\alpha}_i(\lambda_i' H_1 e_{it} - \widetilde{\lambda}_i'\widehat{e}_{it}) & \widetilde{\alpha}_i(\alpha_i' H_2 - \widetilde{\alpha}_i')\end{pmatrix}.$$

Note that the "upper block" of $\widehat{B}_t^{-1}\widehat{S}_t$ is not first-order negligible. Essentially this is due to the fact that the moment condition

$$\frac{\partial}{\partial\lambda_i}\mathsf{E}e_{it}(\dot{y}_{it} - \alpha_i' g_t - e_{it}\lambda_i' f_t) \neq 0,$$

so is not "Neyman orthogonal" with respect to λ_i. On the other hand, we can get around such difficulty. In Lemma E.5 below, we show that $\widehat{B}_t$ and $\widehat{S}_t$ both converge in probability to block diagonal matrices that are independent of t. So g_t and f_t are "orthogonal", and

$$\widehat{B}_t^{-1}\widehat{S}_t = \begin{pmatrix}\bar{H}_3 & 0\\ 0 & \bar{H}_4\end{pmatrix} + o_P(N^{-1/2}).$$

Define $H_f := H_1^{-1} + \bar{H}_3 H_1^{-1}$. Then (D.1) implies that

$$\widehat{f}_t = H_f f_t + \text{upper block of } \widehat{B}_t^{-1}\widehat{Q}_t.$$

Therefore $\widehat{f}_t$ converges to f_t up to a new rotation matrix H_f, which equals H_1^{-1} up to an $o_P(1)$ term $\bar{H}_3 H_1^{-1}$. While the effect of $\bar{H}_3$ is not negligible, it is "absorbed" into the new rotation matrix. As such, we are able to establish the asymptotic normality for $\sqrt{N}(\widehat{f}_t - H_f f_t)$.

Proposition D.1. *For each fixed $t \notin I$, for both the (i) many mean model and (ii) factor model, we have*

$$\widehat{f}_t - H_f f_t = H_f(\frac{1}{N}\sum_i \lambda_i\lambda_i'\mathsf{E}e_{it}^2)^{-1}\frac{1}{N}\sum_i \lambda_i e_{it}u_{it} + O_P(C_{NT}^{-2}).$$

Proof. Define

$$B = \frac{1}{N}\sum_i \begin{pmatrix} H_1'\lambda_i\lambda_i'H_1\mathsf{E}e_{it}^2 & 0 \\ 0 & H_2'\alpha_i\alpha_i'H_2 \end{pmatrix}, \quad S = \frac{1}{N}\sum_i \begin{pmatrix} H_1'\lambda_i(\lambda_i'H_1 - \widetilde{\lambda}_i')\mathsf{E}e_{it}^2 & 0 \\ 0 & \widetilde{\alpha}_i(\alpha_i'H_2 - \widetilde{\alpha}_i') \end{pmatrix}.$$

Both B, S are independent of t due to the stationarity of e_{it}^2. But S depends on the sample splitting through $(\widetilde{\lambda}_i, \widetilde{\alpha}_i)$.

From (D.1),

$$\begin{aligned}
\begin{pmatrix} \widehat{f}_t \\ \widehat{g}_t \end{pmatrix} &= (B^{-1}S + \mathsf{I})\begin{pmatrix} H_1^{-1}f_t \\ H_2^{-1}g_t \end{pmatrix} + B^{-1}\frac{1}{N}\sum_i \begin{pmatrix} H_1'\lambda_i e_{it} \\ H_2'\alpha_i \end{pmatrix} u_{it} \\
&\quad + \sum_{d=1}^5 A_{dt}, \text{ where} \\
A_{1t} &= (\widehat{B}_t^{-1}\widehat{S}_t - B^{-1}S)\begin{pmatrix} H_1^{-1}f_t \\ H_2^{-1}g_t \end{pmatrix} \\
A_{2t} &= (\widehat{B}_t^{-1} - B^{-1})\frac{1}{N}\sum_i \begin{pmatrix} H_1'\lambda_i e_{it} \\ H_2'\alpha_i \end{pmatrix} u_{it} \\
A_{3t} &= \widehat{B}_t^{-1}\frac{1}{N}\sum_i \begin{pmatrix} \widetilde{\lambda}_i\widehat{e}_{it} - H_1'\lambda_i e_{it} \\ \widetilde{\alpha}_i - H_2'\alpha_i \end{pmatrix} u_{it} \\
A_{4t} &= \widehat{B}_t^{-1}\frac{1}{N}\sum_i \begin{pmatrix} \widetilde{\lambda}_i\widehat{e}_{it} - H_1'\lambda_i e_{it} \\ \widetilde{\alpha}_i - H_2'\alpha_i \end{pmatrix} (\mu_{it}\lambda_i'f_t - \widehat{\mu}_{it}\dot{\lambda}_i'\widetilde{f}_t) \\
A_{5t} &= (\widehat{B}_t^{-1} - B^{-1})\frac{1}{N}\sum_i \begin{pmatrix} H_1'\lambda_i e_{it} \\ H_2'\alpha_i \end{pmatrix} (\mu_{it}\lambda_i'f_t - \widehat{\mu}_{it}\dot{\lambda}_i'\widetilde{f}_t) \\
A_{6t} &= B^{-1}\frac{1}{N}\sum_i \begin{pmatrix} H_1'\lambda_i e_{it} \\ H_2'\alpha_i \end{pmatrix} (\mu_{it}\lambda_i'f_t - \widehat{\mu}_{it}\dot{\lambda}_i'\widetilde{f}_t).
\end{aligned} \tag{D.2}$$

Note that $B^{-1}S$ is a block-diagonal matrix, with the upper block being

$$\bar{H}_3 := H_1^{-1}(\frac{1}{N}\sum_i \lambda_i\lambda_i'\mathsf{E}e_{it}^2)^{-1}\frac{1}{N}\sum_i \lambda_i(\lambda_i'H_1 - \widetilde{\lambda}_i')\mathsf{E}e_{it}^2.$$

Define

$$H_f := (\bar{H}_3 + \mathsf{I})H_1^{-1}.$$

Fixed $t \in I^c$, in the many mean model, in Lemma E.7 we show that $\sum_{d=1}^{5} A_{dt} = O_P(C_{NT}^{-2})$ and for the "upper block" of A_{6t}, $\frac{1}{N}\sum_i \lambda_i e_{it}(\mu_i \lambda_i' f_t - \bar{x}_i \dot{\lambda}_i' \widetilde{f}_t) = O_P(C_{NT}^{-2})$.

On the other hand, in the factor model, in Lemma F.9 we show that $\sum_{d=2}^{5} A_{dt} = O_P(C_{NT}^{-2})$, the "upper block" of A_{6t}, $\frac{1}{N}\sum_i \lambda_i e_{it}(l_i' w_t \lambda_i' f_t - \widehat{l_i' w_t} \dot{\lambda}_i' \widetilde{f}_t) = O_P(C_{NT}^{-2})$ and the upper block of A_{1t} is $O_P(C_{NT}^{-2})$. Therefore, in both models,

$$\widehat{f}_t = H_f f_t + H_1^{-1}(\frac{1}{N}\sum_i \lambda_i \lambda_i' \mathsf{E} e_{it}^2)^{-1} \frac{1}{N}\sum_i \lambda_i e_{it} u_{it} + O_P(C_{NT}^{-2}).$$

Given that $\bar{H}_3 = O_P(C_{NT}^{-1})$ we have $H_1^{-1} = H_f + O_P(C_{NT}^{-1})$. By $\frac{1}{N}\sum_i \lambda_i e_{it} u_{it} = O_P(N^{-1/2})$,

$$\widehat{f}_t = H_f f_t + H_f(\frac{1}{N}\sum_i \lambda_i \lambda_i' \mathsf{E} e_{it}^2)^{-1} \frac{1}{N}\sum_i \lambda_i e_{it} u_{it} + O_P(C_{NT}^{-2}). \tag{D.3}$$

Q.E.D.

D.2. **Behavior of $\widehat{\lambda}_i$.** Recall that fix $i \leq N$,

$$(\widehat{\lambda}_{I,i}, \widehat{\alpha}_{I,i}) = \arg\min_{\lambda_i, \alpha_i} \sum_{s \in I^c \cup \{t\}} (\widehat{y}_{is} - \alpha_i' \widehat{g}_{I,s} - \widehat{e}_{is} \lambda_i' \widehat{f}_{I,s})^2.$$

For notational simplicity, we simply write $\widehat{\lambda}_i = \widehat{\lambda}_{I,i}$ and $\widehat{\alpha}_i = \widehat{\alpha}_{I,i}$, but keep in mind that $\widetilde{\alpha}$ and $\widetilde{\lambda}$ are estimated through the low rank estimations on data I. Write

$$T_0 = |I^c|_0.$$

$$(B^{-1}S + \mathsf{I})\begin{pmatrix} H_1^{-1} & 0 \\ 0 & H_2^{-1} \end{pmatrix} := \begin{pmatrix} H_f & 0 \\ 0 & H_g \end{pmatrix}$$
$$\widehat{D}_i = \frac{1}{T_0}\sum_{s \in I^c \cup \{t\}} \begin{pmatrix} \widehat{f}_s \widehat{f}_s' \widehat{e}_{is}^2 & \widehat{f}_s \widehat{g}_s' \widehat{e}_{is} \\ \widehat{g}_s \widehat{f}_s' \widehat{e}_{is} & \widehat{g}_s \widehat{g}_s' \end{pmatrix}, \quad D_i = \frac{1}{T_0}\sum_{s \in I^c \cup \{t\}} \begin{pmatrix} H_f f_s f_s' H_f' \mathsf{E} e_{is}^2 & 0 \\ 0 & H_g g_s g_s' H_g' \end{pmatrix} \tag{D.4}$$

Proposition D.2. *For both the (i) many mean model and (ii) factor model,*

$$\widehat{\lambda}_i = H_f'^{-1}\lambda_i + H_f'^{-1}(\frac{1}{T_0}\sum_{s \in I^c \cup \{t\}} f_s f_s' \mathsf{E} e_{is}^2)^{-1} \frac{1}{T_0}\sum_{s \in I^c \cup \{t\}} f_s e_{is} u_{is} + O_P(C_{NT}^{-2}).$$

Proof. By definition and (D.2),

$$\begin{pmatrix} \widehat{\lambda}_i \\ \widehat{\alpha}_i \end{pmatrix} = \widehat{D}_i^{-1} \frac{1}{T_0}\sum_{s \in I^c \cup \{t\}} \begin{pmatrix} \widehat{f}_s \widehat{e}_{is} \\ \widehat{g}_s \end{pmatrix} (y_{is} - \widehat{\mu}_{is} \dot{\lambda}_i' \widetilde{f}_s)$$

$$
\begin{aligned}
&= \begin{pmatrix} H_f'^{-1}\lambda_i \\ H_g'^{-1}\alpha_i \end{pmatrix} + D_i^{-1}\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} \begin{pmatrix} H_f f_s e_{is} \\ H_g g_s \end{pmatrix} u_{is} + \sum_{d=1}^{6} R_{di}, \quad \text{where,} \\
R_{1i} &= (\widehat{D}_i^{-1} - D_i^{-1})\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} \begin{pmatrix} H_f f_s e_{is} \\ H_g g_s \end{pmatrix} u_{is} \\
R_{2i} &= \widehat{D}_i^{-1}\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} \begin{pmatrix} H_f f_s(\widehat{e}_{is} - e_{is})u_{is} \\ 0 \end{pmatrix} \\
R_{3i} &= (\widehat{D}_i^{-1} - D_i^{-1})\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} \begin{pmatrix} \widehat{f}_s\widehat{e}_{is} \\ \widehat{g}_s \end{pmatrix} (\mu_{is}\lambda_i' f_s - \widehat{\mu}_{is}\dot{\lambda}_i'\widetilde{f}_s) \\
R_{4i} &= D_i^{-1}\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} \begin{pmatrix} \widehat{f}_s\widehat{e}_{is} \\ \widehat{g}_s \end{pmatrix} (\mu_{is}\lambda_i' f_s - \widehat{\mu}_{is}\dot{\lambda}_i'\widetilde{f}_s) \\
R_{5i} &= \widehat{D}_i^{-1}\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} \begin{pmatrix} \widehat{e}_{is}\mathsf{I} & 0 \\ 0 & \mathsf{I} \end{pmatrix} \begin{pmatrix} \widehat{f}_s - H_f f_s \\ \widehat{g}_s - H_g g_s \end{pmatrix} u_{is} \\
R_{6i} &= \widehat{D}_i^{-1}\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} \begin{pmatrix} \widehat{f}_s\widehat{e}_{is} \\ \widehat{g}_s \end{pmatrix} (\lambda_i' H_f^{-1}, \alpha_i' H_g^{-1}) \begin{pmatrix} \widehat{e}_{is}\mathsf{I} & 0 \\ 0 & \mathsf{I} \end{pmatrix} \begin{pmatrix} \widehat{f}_s - H_f f_s \\ \widehat{g}_s - H_g g_s \end{pmatrix} \qquad (D.5)
\end{aligned}
$$

In addition, $\widehat{D}_i - D_i = o_P(1)$ and the upper blocks of R_{di} are all $O_P(C_{NT}^{-2})$ for $d = 1, ..., 6$. This is proved by Lemmas E.9, E.10 in the many mean model, and by Lemma F.13, F.15 in the factor model. Hence

$$
\widehat{\lambda}_i = H_f'^{-1}\lambda_i + H_f'^{-1}\left(\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} f_s f_s' \mathsf{E} e_{is}^2\right)^{-1}\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} f_s e_{is} u_{is} + O_P(C_{NT}^{-2}).
$$

Q.E.D.

D.3. **Proof of normality of $\widehat{\theta}_{it}$.** Suppose T is odd and $|I|_0 = |I^c|_0 = (T-1)/2$. By Propositions D.1 D.2, for fixed $t \notin I$,

$$
\begin{aligned}
\widehat{\lambda}_{I,i}'\widehat{f}_{I,t} - \lambda_i' f_t &= f_t'\left(\frac{1}{|I^c|_0}\sum_{s\in I^c\cup\{t\}} f_s f_s' \mathsf{E} e_{is}^2\right)^{-1}\frac{1}{|I^c|_0}\sum_{s\in I^c\cup\{t\}} f_s e_{is} u_{is} \\
&\quad + \lambda_i'\left(\frac{1}{N}\sum_i \lambda_i\lambda_i' \mathsf{E} e_{it}^2\right)^{-1}\frac{1}{N}\sum_i \lambda_i e_{it} u_{it} \\
&\quad + O_P(C_{NT}^{-2}).
\end{aligned}
$$

Exchanging I with I^c, we have, for $t \notin I^c$,

$$
\widehat{\lambda}_{I^c,i}'\widehat{f}_{I^c,t} - \lambda_i' f_t = f_t'\left(\frac{1}{|I|_0}\sum_{s\in I} f_s f_s' \mathsf{E} e_{is}^2\right)^{-1}\frac{1}{|I|_0}\sum_{s\in I} f_s e_{is} u_{is}
$$

$$+\lambda_i'(\frac{1}{N}\sum_i \lambda_i\lambda_i'\mathsf{E}e_{it}^2)^{-1}\frac{1}{N}\sum_i \lambda_i e_{it}u_{it}$$
$$+O_P(C_{NT}^{-2}).$$

Note that the fixed $t \notin I \cup I^c$, so take the average:

$$\begin{aligned}\widehat{\theta}_{it} - \lambda_i' f_t &= f_t'\left[(\frac{1}{|I|_0}\sum_{s\in I} f_s f_s'\mathsf{E}e_{is}^2)^{-1}\frac{1}{2|I|_0}\sum_{s\in I} f_s e_{is}u_{is} + (\frac{1}{|I^c|_0}\sum_{s\in I^c\cup\{t\}} f_s f_s'\mathsf{E}e_{is}^2)^{-1}\frac{1}{2|I^c|_0}\sum_{s\in I^c\cup\{t\}} f_s e_i\right. \\ &\quad +\lambda_i'(\frac{1}{N}\sum_i \lambda_i\lambda_i'\mathsf{E}e_{it}^2)^{-1}\frac{1}{N}\sum_i \lambda_i e_{it}u_{it} + O_P(C_{NT}^{-2}).\end{aligned}$$

Let $V_{f1} = \frac{1}{T}\sum_{s=1}^T f_s f_s'\mathsf{E}e_{is}^2$. By the assumption that

$$\frac{1}{|I|_0}\sum_{s\in I} f_s f_s'\mathsf{E}e_{is}^2 = V_{f1} + o_P(1) = \frac{1}{|I^c|_0}\sum_{s\in I^c\cup\{t\}} f_s f_s'\mathsf{E}e_{is}^2,$$

the first term in the above expansion is then $o_P(T^{-1/2})$ plus

$$f_t'V_{f1}^{-1}\left[\frac{1}{T}\sum_{s\in I} f_s e_{is}u_{is} + \frac{1}{T}\sum_{s\in I^c\cup\{t\}} f_s e_{is}u_{is}\right] = f_t'V_{f1}^{-1}\frac{1}{T}\sum_{s=1}^T f_s e_{is}u_{is} + O_P(T^{-1}).$$

In addition, let $\xi_{NT} = \lambda_i'V_{\lambda 1}^{-1}\frac{1}{\sqrt{N}}\sum_j \lambda_j e_{jt}u_{jt}$ and $\zeta_{NT} = f_t'V_{f1}^{-1}\frac{1}{\sqrt{T}}\sum_{s=1}^T f_s e_{is}u_{is}$. Then

$$\widehat{\theta}_{it} - \theta_{it} = \zeta_{NT}/\sqrt{T} + \xi_{NT}/\sqrt{N} + O_P(C_{NT}^{-2}) + o_P(T^{-1/2}).$$

Next, write $\Sigma_{NT}^{1/2} = (\frac{1}{T}f_t'V_f f_t + \frac{1}{N}\lambda_i V_\lambda \lambda_i)^{1/2}$. Then regardless of $T/N \in [0,\infty]$,

$$\begin{aligned}\left(\frac{O_P(C_{NT}^{-2}) + o_P(T^{-1/2})}{\Sigma_{NT}^{1/2}}\right)^2 &= O_P\left(\frac{1}{Tf_t'V_f f_t + \frac{T^2}{N}\lambda_i V_\lambda \lambda_i}\right) \\ &\quad + O_P\left(\frac{1}{\frac{N^2}{T}f_t'V_f f_t + N\lambda_i V_\lambda \lambda_i}\right) + o_P(1) = o_P(1)\end{aligned}$$

Next, $\mathsf{Cov}(\xi_{NT}, \zeta_{NT}) = \frac{1}{\sqrt{NT}}\sum_j \mathsf{E}\lambda_i'V_{\lambda 1}^{-1}\lambda_j f_t'V_{f1}^{-1}f_s e_{jt}e_{it}\mathsf{E}(u_{jt}u_{it}|E,F) \to 0$. So $(\zeta_{NT}, \xi_{NT}) \to^d (\zeta, \xi)$, where $(\zeta.\xi)$ is a bivariate Gaussian random vector, with mean zero, and covariance $\mathsf{diag}\{f_t'V_f f_t, \lambda_i'V_\lambda \lambda_i\}$. We now use the same argument as the proof of Theorem 3 in Bai (2003). There exist $(\zeta_{NT}^*, \xi_{NT}^*)$ and (ζ^*, ξ^*) with the same distribution as (ζ_{NT}, ξ_{NT}) and (ζ, ξ) such that $(\zeta_{NT}^*, \xi_{NT}^*) \to (\zeta^*, \xi^*)$ almost surely (almost sure representation). Then

$$
\begin{aligned}
\frac{\zeta_{NT}/\sqrt{T}+\xi_{NT}/\sqrt{N}}{\Sigma_{NT}^{1/2}} &=^d \frac{\zeta_{NT}^*/\sqrt{T}+\xi_{NT}^*/\sqrt{N}}{\Sigma_{NT}^{1/2}} \\
&= \underbrace{\frac{\zeta^*/\sqrt{T}+\xi^*/\sqrt{N}}{\Sigma_{NT}^{1/2}}}_{\mathcal{N}(0,1)} + \underbrace{\frac{(\zeta_{NT}^*-\zeta^*)}{(f_t'V_ff_t+\frac{T}{N}\lambda_iV_\lambda\lambda_i)^{1/2}}}_{a_1} + \underbrace{\frac{(\xi_{NT}^*-\xi^*)}{(\frac{N}{T}f_t'V_ff_t+\lambda_iV_\lambda\lambda_i)^{1/2}}}_{a_2} \\
&=^d \mathcal{N}(0,1)+o(1),
\end{aligned}
$$

where $a_1 \to 0$ and $a_2 \to 0$ almost surely regardless of $T/N \in [0,\infty]$. Therefore,

$$\frac{\widehat{\theta}_{it}-\theta_{it}}{\Sigma_{NT}^{1/2}} \to^d \mathcal{N}(0,1).$$

Q.E.D.

Appendix E. Technical lemmas in the many mean model

E.1. **Behavior of the preliminary in the many mean model.** Recall that

$$(\widetilde{f}_s,\widetilde{g}_s) := \arg\min_{f_s,g_s}\sum_{i=1}^N (y_{is}-\widetilde{\alpha}_i'g_s - x_{is}\widetilde{\lambda}_i'f_s)^2, \quad s \in I^c\cup\{t\}.$$

and

$$(\dot{\lambda}_i,\dot{\alpha}_i) = \arg\min_{\lambda_i,\alpha_i}\sum_{s\in I^c\cup\{t\}}(y_{is}-\alpha_i'\widetilde{g}_s - x_{is}\lambda_i'\widetilde{f}_s)^2, \quad i=1,...,N.$$

The goal of this section is to show that the effect of the preliminary estimation is negligible. Specifically, we aim to show, for each fixed $t\in I^c$, fixed $i \le N$,

$$
\begin{aligned}
\frac{1}{\sqrt{N}}\sum_j (H_1'\lambda_j-\dot{\lambda}_j)e_{jt} &= O_P(\sqrt{N}C_{NT}^{-2}), \\
\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s(H_1^{-1}f_s-\widetilde{f}_s)e_{is} &= O_P(\sqrt{T}C_{NT}^{-2}).
\end{aligned}
$$

Throughout the proof below, we treat $|I^c|=T$ instead of $T/2$ to avoid keeping the constant "2". In addition, for notational simplicity, we write $\widetilde{\Lambda}=\widetilde{\Lambda}_I$ and $\widetilde{A}=\widetilde{A}_I$ by suppressing the subscripts, but we should keep in mind that $\widetilde{\Lambda}$ and $\widetilde{A}$ are estimated on data D_I as defined in step 2. In addition, let E_I and Var_I be the conditional expectation and variance, given D_I. Recall that X_s be the vector of x_{is} fixing $s\le T$, and $M_{\widetilde{\alpha}} = I_N - \widetilde{A}(\widetilde{A}'\widetilde{A})^{-1}\widetilde{A}'$; X_i be the vector of x_{is} fixing

$i \leq N$, and $M_{\widetilde{g}} = I - \widetilde{G}(\widetilde{G}'\widetilde{G})^{-1}\widetilde{G}'$, for $\widetilde{G}$ as the $|I^c|_0 \times K_1$ matrix of $\widetilde{g}_s$. Define $\widetilde{F}$ similarly.

Also,

$$\begin{aligned}
\widetilde{D}_{fs} &= \frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)\widetilde{\Lambda} \\
D_{fs} &= \frac{1}{N}\Lambda'(\mathsf{diag}(X_s)M_{\alpha}\mathsf{diag}(X_s)\Lambda \\
D_{f} &= \frac{1}{N}\Lambda'\mathsf{E}((\mathsf{diag}(X_s)M_{\alpha}\mathsf{diag}(X_s))\Lambda \\
\widetilde{D}_{\lambda i} &= \frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}\mathsf{diag}(X_i)\widetilde{F} \\
D_{\lambda i} &= \frac{1}{T}F'(\mathsf{diag}(X_i)M_{g}\mathsf{diag}(X_i))F \\
\bar{D}_{\lambda i} &= \frac{1}{T}F'\mathsf{E}(\mathsf{diag}(X_i)M_{g}\mathsf{diag}(X_i))F
\end{aligned}$$

By the stationarity, D_f does not depend on s.

Lemma E.1. *Suppose* $\max_{is} x_{is}^2 = o_P(C_{NT})$. *Also, there is* $c > 0$*, so that* $\min_s \min_j \psi_j(D_{fs}) > c$. *Then*

(i) $\max_s \|\widetilde{D}_{fs}^{-1}\| = O_P(1)$.

(ii) $\frac{1}{T}\sum_{s \in I^c \cup \{t\}} \|\widetilde{D}_{fs}^{-1} - (H_1'D_fH_1)^{-1}\|^2 = O_P(C_{NT}^{-2})$.

Proof. (i) The eigenvalues of D_{fs} are bounded from zero uniformly in $s \leq T$. Also,

$$\begin{aligned}
\widetilde{D}_{fs} - H_1'D_{fs}H_1 &= \sum_l \delta_l, \qquad \text{where} \\
\delta_1 &= \frac{1}{N}(\widetilde{\Lambda} - \Lambda H_1)'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)\widetilde{\Lambda}, \\
\delta_2 &= \frac{1}{N}H_1'\Lambda'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)(\widetilde{\Lambda} - \Lambda H_1) \\
\delta_3 &= \frac{1}{N}H_1'\Lambda'\mathsf{diag}(X_s)(M_{\widetilde{\alpha}} - M_{\alpha})\mathsf{diag}(X_s)\Lambda H_1. \qquad \text{(E.1)}
\end{aligned}$$

We now bound each term uniformly in $s \leq T$. The first term is

$$\frac{1}{N}(\widetilde{\Lambda} - \Lambda H_1)'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)\widetilde{\Lambda} \leq O_P(1)\frac{1}{\sqrt{N}}\|\widetilde{\Lambda} - \Lambda H_1\|_F \max_{is} x_{is}^2 = o_P(1)$$

provided that $\max_{is} x_{is}^2 = o_P(C_{NT})$. The second term is bounded similarly. The third term is bounded by

$$O_P(1)\max_{is} x_{is}^2 \|M_{\widetilde{\alpha}} - M_{\alpha}\| = O_P(1)\frac{1}{\sqrt{N}}\|\widetilde{A} - AH_2\|_F \max_{is} x_{is}^2 = o_P(1).$$

This implies $\max_s \|\widetilde{D}_{fs} - H_1' D_{fs} H_1\| = o_P(1)$. In addition, because of the convergence of $\|\frac{1}{\sqrt{N}}(\widetilde{\Lambda} - \Lambda H_1)\|_F$, we have $\min_j \psi_j(H_1'H_1) \geq C \min_j \psi_j(\frac{1}{N}H_1'\Lambda'\Lambda H_1)$, bounded away from zero. Thus $\min_s \min_j \psi_j(H_1'D_{fs}H_1) \geq \min_s \min_j \psi_j(D_{fs})C$, bounded away from zero. This together with $\max_s \|\widetilde{D}_{fs} - H_1'D_{fs}H_1\| = o_P(1)$ imply $\min_s \|\widetilde{D}_{fs}^{-1}\| = O_P(1)$.

(ii) By (E.1),

$$
\begin{aligned}
&\frac{1}{T}\sum_{s\in I^c\cup\{t\}} \|\widetilde{D}_{fs}^{-1} - (H_1'D_fH_1)^{-1}\|^2 \\
\leq\ & \frac{1}{T}\sum_s \|\widetilde{D}_{fs} - (H_1'D_fH_1)\|^2 \|(H_1'D_fH_1)^{-2}\| \max_s \|\widetilde{D}_{fs}^{-2}\| \\
\leq\ & O_P(1)\frac{1}{T}\sum_s \|\widetilde{D}_{fs} - (H_1'D_fH_1)\|^2 \\
=\ & O_P(1)\sum_{l=1}^{3}\frac{1}{T}\sum_s \|\delta_l\|^2 + O_P(1)\frac{1}{T}\sum_s \|D_{fs} - D_f\|^2.
\end{aligned}
$$

We bound each term below.

$$
\frac{1}{T}\sum_{s\in I^c\cup\{t\}} \|\delta_1\|^2 \leq \frac{1}{N^2}\sum_{ij} \|\widetilde{\lambda}_i - H_1'\lambda_i\|^2 \|\widetilde{\lambda}_j\|^2 \frac{1}{T}\sum_{s\in I^c\cup\{t\}} x_{is}^2 x_{js}^2.
$$

Note that $\{e_{is}\}$ is serially independent, so $x_{is}^2x_{js}^2$ is independent of $\|\widetilde{\lambda}_i - H_1'\lambda_i\|^2\|\widetilde{\lambda}_j - H_1'\lambda_j\|^2$ for $s \in I^c \cup \{t\}$. Take the conditional expectation E_I. Then $\mathsf{E}_I x_{is}^2 x_{js}^2$ equals the unconditional expectation, and is bounded uniformly over (s, i, j).

$$
\begin{aligned}
\frac{1}{T}\sum_{s\in I^c\cup\{t\}} \|\delta_1\|^2 &\leq O_P(1)\frac{1}{N^2}\sum_{ij} \|\widetilde{\lambda}_i - H_1'\lambda_i\|^2 \|\widetilde{\lambda}_j\|^2 \frac{1}{T}\sum_{s\in I^c\cup\{t\}} \mathsf{E}_I x_{is}^2 x_{js}^2 \\
&\leq O_P(1)\frac{1}{N^2}\sum_{ij} \|\widetilde{\lambda}_i - H_1'\lambda_i\|^2 \|\widetilde{\lambda}_j\|^2 = O_P(C_{NT}^{-2}).
\end{aligned}
$$

Term of $\frac{1}{T}\sum_{s\in I^c\cup\{t\}} \|\delta_2\|^2$ is bounded similarly.

$$
\frac{1}{T}\sum_{s\in I^c\cup\{t\}} \|\delta_3\|^2 \leq O_P(1)\|M_{\widetilde{\alpha}} - M_\alpha\| = O_P(C_{NT}^{-2}).
$$

Finally, let $M_{\alpha,ij}$ and $P_{\alpha,ij}$ be the (i,j) th component of M_α and $A(A'A)^{-1}A'$. Then $P_{\alpha,ij} = \frac{1}{N}\alpha_i'(\frac{1}{N}A'A)^{-1}\alpha_j$. Let $p_i = (\frac{1}{N}A'A)^{-1/2}\alpha_i$. We have,

$$
\frac{1}{T}\sum_s \|D_{fs} - D_f\|^2 \leq \frac{1}{T}\sum_s \|\frac{1}{N}\sum_{ij}\lambda_i\lambda_j' M_{\alpha,ij}(x_{is}x_{js} - \mathsf{E}x_{is}x_{js})\|_F^2
$$

$$= \frac{1}{T}\sum_s \|\frac{1}{N}\sum_i \lambda_i\lambda_i' M_{\alpha,ii}(x_{is}^2 - \mathsf{E}x_{is}^2) - \frac{1}{N^2}\sum_{i\neq j}\lambda_i\lambda_j' p_i' p_j(x_{is}x_{js} - \mathsf{E}x_{is}x_{js})\|_F^2.$$

As for the first term, it equals $\frac{1}{T}\sum_s \|\frac{1}{N}\sum_i \lambda_i\lambda_i' M_{\alpha,ii}(e_{is}^2 - \mathsf{E}e_{is}^2) + \frac{2}{N}\sum_i \lambda_i\lambda_i' M_{\alpha,ii}\mu_i e_{is}\|_F^2 = O_P(N^{-1})$ provided that $\frac{1}{N}\sum_{ij}|\mathsf{Cov}(e_{is}^2, e_{js}^2)| < \infty$ and $\|\mathsf{E}e_s e_s'\| < \infty$. As for the second term, we can consider each element of λ_i and p_i, as the dimensions of both vectors are finite. So without loss of generality we assume $\dim(\lambda_i) = \dim(p_i) = 1$. Then,

$$\begin{aligned}
& \mathsf{E}\frac{1}{T}\sum_s |\frac{1}{N^2}\sum_{i\neq j}\lambda_i\lambda_j p_i p_j(x_{is}x_{js} - \mathsf{E}x_{is}x_{js})|^2 \\
\leq\ & 2\frac{1}{T}\sum_s \mathsf{E}|\frac{1}{N^2}\sum_{ij}\lambda_i\lambda_j p_i p_j(x_{is}x_{js} - \mathsf{E}x_{is}x_{js})|^2 + O(N^{-2}) \\
\leq\ & 8\mathsf{Var}(\frac{1}{N}\sum_j \lambda_j p_j e_{js})(\frac{1}{N}\sum_i \lambda_i p_i \mu_i)^2 + \mathsf{Var}((\frac{1}{N}\sum_i \lambda_i p_i e_{is})^2) + O(N^{-2}) \\
=\ & O(N^{-1}).
\end{aligned}$$

provided that $\|\mathsf{E}e_s e_s'\| < \infty$ and $\frac{1}{N^3}\sum_{ijkl}\mathsf{Cov}(e_{is}e_{js}, e_{ks}e_{ls}) < C$. So $\frac{1}{T}\sum_s \|D_{fs} - D_f\|^2 = O_P(N^{-1})$. Put together, $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|\widetilde{D}_{fs}^{-1} - (H_1' D_f H_1)^{-1}\|^2 = O_P(C_{NT}^{-2})$.

Lemma E.2. *(i) For each fixed $t \in I^c$, $H_1^{-1}f_t - \widetilde{f}_t = O_P(C_{NT}^{-1})$.*

(ii) For each fixed $i \leq N$, $\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s(H_1^{-1}f_s - \widetilde{f}_s)e_{is} = O_P(\sqrt{T}C_{NT}^{-2})$.

Proof. Without loss of generality, we assume $\dim(g_s) = \dim(f_s) = 1$, as we can always work with their elements given that the number of factors is fixed. Then for $y_s = Ag_s + \mathsf{diag}(X_s)\Lambda f_s + u_s$,

$$\begin{aligned}
\widetilde{f}_s &= \widetilde{D}_{fs}^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}y_s \\
&= H_1^{-1}f_s + \widetilde{D}_{fs}^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}(AH_2 - \widetilde{A})H_2^{-1}g_s \\
&\quad + \widetilde{D}_{fs}^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)(\Lambda H_1 - \widetilde{\Lambda})H_1^{-1}f_s \\
&\quad + \widetilde{D}_{fs}^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}u_s.
\end{aligned} \tag{E.2}$$

Part (ii) that $H_1^{-1}f_t - \widetilde{f}_t = O_P(C_{NT}^{-1})$ follows from a straightforward calculation. So we omit the details. We now prove the much harder part (i) by plugging in each term in the above expansion to $\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s(H_1^{-1}f_s - \widetilde{f}_s)e_{is}$.

First term: note that $\frac{1}{NT}\sum_s \|g_s f_s e_{is}\|^2 \|\widetilde{\Lambda}'\mathsf{diag}(X_s)\|_F^2 < C$ provided that $\mathsf{E} g_s^2 f_s^2 e_{is}^2 x_{is}^2 < \infty$. So for a fixed $i \leq N$, by Lemma E.1.

$$
\begin{aligned}
& \frac{1}{\sqrt{T}}\sum_s f_s e_{is}(\widetilde{D}_{fs}^{-1} - (H_1' D_x H_1)^{-1})\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s) M_{\widetilde{\alpha}}(AH_2 - \widetilde{A})H_2^{-1} g_s \\
\leq\ & O_P(\sqrt{T} C_{NT}^{-1})(\frac{1}{T}\sum_s \|\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)\|^2 \|g_s\|^2 \|f_s e_{is}\|^2 \|AH_2 - \widetilde{A}\|^2)^{1/2} \\
\leq\ & O_P(\sqrt{T} C_{NT}^{-2}) = o_P(1). \\
& \frac{1}{\sqrt{T}}\sum_s f_s e_{is}(H_1' D_x H_1)^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s) M_{\widetilde{\alpha}}(AH_2 - \widetilde{A})H_2^{-1} g_s \\
\leq\ & O_P(C_{NT}^{-1})\|\frac{1}{\sqrt{TN}}\sum_s f_s g_s e_{is}(H_1' D_x H_1)^{-1}\widetilde{\Lambda}'\mathsf{diag}(X_s)\| \\
\leq\ & O_P(\sqrt{T} C_{NT}^{-1})[\frac{1}{N}\sum_j \widetilde{\lambda}_j^2(\frac{1}{T}\sum_s f_s g_s e_{is} x_{js})^2]^{1/2}
\end{aligned}
$$

To bound the last line, note that e_{is} is conditionally serially independent, given

$$
\begin{aligned}
& \mathsf{E}_I \frac{1}{N}\sum_j \widetilde{\lambda}_j^2(\frac{1}{T}\sum_s f_s g_s e_{is} x_{js})^2 = \frac{1}{N}\sum_j \widetilde{\lambda}_j^2 \mathsf{E}_I(\frac{1}{T}\sum_s f_s g_s e_{is} x_{js})^2 \\
\leq\ & \frac{1}{N}\sum_j (\widetilde{\lambda}_j - H_1'\lambda_j)^2 + O_P(1)\frac{1}{N}\sum_j \mathsf{E}_I(\frac{1}{T}\sum_s f_s g_s e_{is} x_{js})^2 \\
=\ & O_P(C_{NT}^{-2}) + O_P(1)\frac{1}{N}\sum_j \mathsf{E}_I(\frac{1}{T}\sum_s f_s g_s e_{is} e_{js})^2 + O_P(1)\mathsf{E}_I(\frac{1}{T}\sum_s f_s g_s e_{is})^2 \\
\leq\ & O_P(C_{NT}^{-2}) + O_P(1)\frac{1}{N}\sum_j (\frac{1}{T}\sum_s \mathsf{E}_I f_s g_s \mathsf{E}(e_{is} e_{js} | D_I, f_s, g_s))^2 \\
& + O_P(1)\mathsf{Var}_I(\frac{1}{T}\sum_s f_s g_s e_{is} e_{js}) + O_P(1)\mathsf{Var}_I(\frac{1}{T}\sum_s f_s g_s e_{is}) \\
\leq^*\ & O_P(C_{NT}^{-2}) + O_P(1)\frac{1}{N}\sum_j |\frac{1}{T}\sum_s \mathsf{E}_I f_s g_s \mathsf{E}(e_{is} e_{js} | D_I, f_s, g_s)| + O_P(T^{-1}) \\
\leq\ & O_P(C_{NT}^{-2}) + O_P(N^{-1})\frac{1}{T}\sum_s \mathsf{E}_I |f_s g_s| \max_{i,s,f_s,g_s}\sum_j |\mathsf{E}(e_{is} e_{js} | f_s, g_s)| + O_P(T^{-1}) \\
\leq\ & O_P(C_{NT}^{-2}),
\end{aligned}
$$

where $\leq^*$ is due to $|\mathsf{E}_I f_s g_s e_{is} e_{js}| < \infty$. Put together,

$$
\frac{1}{\sqrt{T}}\sum_s f_s e_{is}\widetilde{D}_{fs}^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s) M_{\widetilde{\alpha}}(AH_2 - \widetilde{A})H_2^{-1} g_s = O_P(\sqrt{T} C_{NT}^{-2}).
$$

Second term: recall that $M_{\widetilde{\alpha},ij}$ and $P_{\widetilde{\alpha},ij}$ are the (i,j) th component of $M_{\widetilde{\alpha}}$ and $\widetilde{A}(\widetilde{A}'\widetilde{A})^{-1}\widetilde{A}'$ and $P_{\widetilde{\alpha},ij} = \frac{1}{N}p_i'p_j$.

$$
\begin{aligned}
& \frac{1}{\sqrt{T}}\sum_s f_s e_{is}(\widetilde{D}_{fs}^{-1} - (H_1'D_xH_1)^{-1})\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)(\Lambda H_1 - \widetilde{\Lambda})H_1^{-1}f_s \\
\leq\ & O_P(\sqrt{TN}C_{NT}^{-2})(\frac{1}{T}\sum_s \|\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)\|^2\|f_s\|^2\|f_se_{is}\|^2)^{1/2} \\
\leq\ & O_P(\sqrt{TN}C_{NT}^{-2})(\frac{1}{T}\sum_s \frac{1}{N^2}\sum_j \widetilde{\lambda}_j^2 x_{js}^4 M_{\widetilde{\alpha},jj}^2\|f_s\|^2\|f_se_{is}\|^2)^{1/2} \\
& +O_P(\sqrt{TN}C_{NT}^{-2})(\frac{1}{T}\sum_s \frac{1}{N^2}\sum_j [\frac{1}{N}\sum_{k\neq j}\widetilde{\lambda}_k x_{ks}p_k'p_jx_{js}]^2\|f_s\|^2\|f_se_{is}\|^2)^{1/2} \\
\leq\ & O_P(\sqrt{T}C_{NT}^{-2})(\frac{1}{T}\sum_s \frac{1}{N}\sum_j \widetilde{\lambda}_j^2\mathsf{E}_I x_{js}^4\|f_s\|^2\|f_se_{is}\|^2)^{1/2} \\
& +O_P(\sqrt{T}C_{NT}^{-2})(\frac{1}{T}\sum_s \frac{1}{N^2}\sum_{k\neq j} x_{ks}^2(p_k'p_j)^2x_{js}^2\|f_s\|^2\|f_se_{is}\|^2)^{1/2} \\
=\ & O_P(\sqrt{T}C_{NT}^{-2}).
\end{aligned}
$$

It is also straightforward to prove that

$$
\begin{aligned}
& (\frac{1}{\sqrt{T}}\sum_s f_se_{is}(H_1'D_xH_1)^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)(M_{\widetilde{\alpha}} - M_\alpha)\mathsf{diag}(X_s)(\Lambda H_1 - \widetilde{\Lambda})H_1^{-1}f_s)^2 \\
\leq\ & O_P(T)\frac{1}{T}\sum_s \|f_s^2e_{is}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)\|^2\frac{1}{T}\sum_s \|\mathsf{diag}(X_s)(\Lambda H_1 - \widetilde{\Lambda})\|^2\|M_{\widetilde{\alpha}} - M_\alpha\|^2 \\
\leq\ & O_P(TC_{NT}^{-4}).
\end{aligned}
$$

Next, let $z_{js} = \sum_k \widetilde{\lambda}_k M_{\alpha,kj}x_{ks}$,

$$
\begin{aligned}
& \frac{1}{\sqrt{T}}\sum_s f_se_{is}(H_1'D_xH_1)^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_\alpha\mathsf{diag}(X_s)(\Lambda H_1 - \widetilde{\Lambda})H_1^{-1}f_s \\
\leq\ & O_P(\sqrt{T}C_{NT}^{-1})\|\frac{1}{T}\sum_s f_s^2e_{is}(H_1'D_xH_1)^{-1}\frac{1}{\sqrt{N}}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_\alpha\mathsf{diag}(X_s)\| \\
\leq\ & O_P(\sqrt{T}C_{NT}^{-1})[\frac{1}{N}\sum_j(\frac{1}{T}\sum_s f_s^2e_{is}z_{js}x_{js})^2]^{1/2}
\end{aligned}
$$

To bound the last line, note that

$$
\frac{1}{N}\sum_j \mathsf{E}_I(\frac{1}{T}\sum_s f_s^2e_{is}z_{js}x_{js})^2 = \frac{1}{N}\sum_j(\frac{1}{T}\sum_s \mathsf{E}_I f_s^2z_{js}e_{is}x_{js})^2 + \frac{1}{N}\sum_j \mathsf{Var}_I(\frac{1}{T}\sum_s f_s^2e_{is}z_{js}x_{js})
$$

$$
\begin{aligned}
&\leq \frac{1}{N}\sum_j(\frac{1}{T}\sum_s \mathsf{E}_I f_s^2 e_{is}x_{js}^2)^2(\widetilde{\lambda}_j M_{\alpha,jj})^2 + O_P(1)\frac{1}{N}\sum_j \frac{1}{N}\sum_{k\neq j}(\frac{1}{T}\sum_s \mathsf{E}_I f_s^2 x_{ks}e_{is}x_{js})^2 \\
&\quad +\frac{1}{N}\sum_j \mathsf{Var}_I(\frac{1}{T}\sum_s f_s^2 e_{is}z_{js}x_{js}) \\
&\leq \frac{1}{N}\sum_j(\frac{1}{T}\sum_s \mathsf{E}_I f_s^2 e_{is}x_{js}^2)^2(\widetilde{\lambda}_j - H_1'\lambda_j)^2 + O_P(1)\frac{1}{N}\sum_j(\frac{1}{T}\sum_s \mathsf{E}_I f_s^2 e_{is}x_{js}^2)^2 \\
&\quad +O_P(1)\frac{1}{N}\sum_j \frac{1}{N}\sum_{k\neq j}(\frac{1}{T}\sum_s \mathsf{E}_I f_s^2 x_{ks}e_{is}x_{js})^2 + O_P(T^{-1}) \\
&\leq O_P(1)\frac{1}{N}\frac{1}{T}\sum_s \mathsf{E}_I f_s^2\sum_j |\mathsf{E}_I(e_{is}e_{js}^2|f_s)| + O_P(1)\frac{1}{N}\frac{1}{T}\sum_s \mathsf{E}_I f_s^2 \sum_j |\mathsf{E}_I(e_{is}e_{js}|f_s)| \\
&\quad +O_P(1)\frac{1}{N}\frac{1}{N}\sum_{k\neq j}|\frac{1}{T}\sum_s \mathsf{E}_I f_s^2 e_{ks}e_{is}e_{js}| + O_P(C_{NT}^{-2}) = O_P(C_{NT}^{-2})
\end{aligned}
$$

given that $\sum_j |\mathsf{E}_I(e_{is}e_{js}^2|f_s)| + \sum_j |\mathsf{E}_I(e_{is}e_{js}|f_s)| + \frac{1}{N}\sum_{k\neq j}|\mathsf{E}_I(e_{ks}e_{is}e_{js}|f_s)| < \infty$. Put together,

$$
\frac{1}{\sqrt{T}}\sum_s f_s e_{is}\widetilde{D}_{fs}^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)(\Lambda H_1 - \widetilde{\Lambda})H_1^{-1}f_s = O_P(\sqrt{T}C_{NT}^{-2}).
$$

Third term: note that $\mathsf{E}_I(u_s u_s'|X_s, f_s) < C$.

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s e_{is}(\widetilde{D}_{fs}^{-1} - (H_1'D_xH_1)^{-1})\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}u_s \\
&\leq O_P(\sqrt{T}C_{NT}^{-1})(\frac{1}{T}\sum_{s\in I^c\cup\{t\}} |f_s e_{is}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}u_s|^2)^{1/2} \\
&= O_P(\sqrt{T}C_{NT}^{-1})(\frac{1}{T}\sum_{s\in I^c\cup\{t\}} \mathsf{E}_I|f_s e_{is}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}u_s|^2)^{1/2} \\
&= O_P(\sqrt{T}C_{NT}^{-1})(\frac{1}{T}\sum_{s\in I^c\cup\{t\}} \mathsf{E}_I f_s^2 e_{is}^2\frac{1}{N^2}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{E}_I(u_s u_s'|X_s, f_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)\widetilde{\Lambda})^{1/2} \\
&\leq O_P(\sqrt{T}C_{NT}^{-1})\frac{1}{N}(\frac{1}{T}\sum_{s\in I^c\cup\{t\}} \mathsf{E}_I f_s^2 e_{is}^2\|\widetilde{\Lambda}'\mathsf{diag}(X_s)\|^2)^{1/2} \\
&\leq O_P(\sqrt{T}C_{NT}^{-1})\frac{1}{N}(\frac{1}{T}\sum_{s\in I^c\cup\{t\}} \mathsf{E}_I f_s^2 e_{is}^2\sum_j \widetilde{\lambda}_j^2 x_{js}^2)^{1/2} = O_P(\sqrt{T}C_{NT}^{-2}).
\end{aligned}
$$

Next, due to $\mathsf{E}(u_s|f_s, D_I, e_s) = 0$, and e_s is conditionally serially independent given (f_s, u_s),

$$
\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s e_{is}(H_1'D_xH_1)^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)(M_{\widetilde{\alpha}} - M_\alpha)u_s
$$

$$
\begin{aligned}
&= \mathsf{tr}\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} u_s f_s e_{is}(H_1'D_xH_1)^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)(M_{\widetilde{\alpha}}-M_\alpha)\\
&\leq O_P(\sqrt{T}C_{NT}^{-1})\|\frac{1}{T}\sum_{s\in I^c\cup\{t\}} u_s f_s e_{is}(H_1'D_xH_1)^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)\|_F\\
&\leq O_P(\sqrt{T}C_{NT}^{-1})(\frac{1}{N}\sum_j\|\frac{1}{T}\sum_{s\in I^c\cup\{t\}} u_s f_s e_{is}\frac{1}{\sqrt{N}}x_{js}\|_F^2\widetilde{\lambda}_j^2)^{1/2}\\
&\leq O_P(\sqrt{T}C_{NT}^{-1})(\max_j \mathsf{E}_I\|\frac{1}{T}\sum_{s\in I^c\cup\{t\}} u_s f_s e_{is}\frac{1}{\sqrt{N}}x_{js}\|_F^2)^{1/2}\\
&\leq O_P(C_{NT}^{-1})(\max_{jk}\frac{1}{T}\mathsf{Var}_I(\sum_{s\in I^c\cup\{t\}} u_{ks} f_s e_{is}x_{js}))^{1/2}\\
&\leq O_P(C_{NT}^{-1})(\max_{jks}\mathsf{Var}_I(u_{ks}f_s e_{is}x_{js}))^{1/2} = O_P(C_{NT}^{-1}) = O_P(\sqrt{T}C_{NT}^{-2}).
\end{aligned}
$$

Next,

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s e_{is}(H_1'D_xH_1)^{-1}\frac{1}{N}(\widetilde{\Lambda}-\Lambda H_1)'\mathsf{diag}(X_s)M_\alpha u_s\\
\leq\;& O_P(1)\frac{1}{N}\sum_j(\widetilde{\lambda}_j - H_1'\lambda_j)\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s e_{is}M_{\alpha,j}'u_s x_{js}\\
\leq\;& O_P(C_{NT}^{-1})(\frac{1}{N}\sum_j(\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s e_{is}M_{\alpha,j}'u_s x_{js})^2)^{1/2}\\
\leq\;& O_P(C_{NT}^{-1})(\frac{1}{N}\sum_j\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\mathsf{Var}_I(f_s e_{is}M_{\alpha,j}'u_s x_{js}))^{1/2}\\
=\;& O_P(C_{NT}^{-1})(\frac{1}{N}\sum_j\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\mathsf{E}_I f_s^2 x_{js}^2 e_{is}^2 M_{\alpha,j}'\mathsf{Var}_I(u_s|e_s,f_s)M_{\alpha,j})^{1/2}\\
\leq\;& O_P(C_{NT}^{-1})(\frac{1}{N}\sum_j\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\mathsf{E}_I f_s^2 x_{js}^2 e_{is}^2\|M_{\alpha,j}\|^2)^{1/2} = O_P(C_{NT}^{-1}) = O_P(\sqrt{T}C_{NT}^{-2}).
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s e_{is}(H_1'D_xH_1)^{-1}\frac{1}{N}H_1'\Lambda'\mathsf{diag}(X_s)M_\alpha u_s\\
\leq\;& O_P(N^{-1})(\mathsf{E}_I f_s^2 e_{is}^2\Lambda'\mathsf{diag}(X_s)M_\alpha\mathsf{E}_I(u_s u_s'|e_s,f_s)M_\alpha\mathsf{diag}(X_s)\Lambda)^{1/2}\\
\leq\;& O_P(N^{-1})(\mathsf{E}_I f_s^2 e_{is}^2\|\Lambda'\mathsf{diag}(X_s)\|^2)^{1/2}\\
\leq\;& O_P(N^{-1/2})(\frac{1}{N}\sum_j\mathsf{E}_I f_s^2 e_{is}^2 x_{js}^2)^{1/2} = O_P(N^{-1/2}) = O_P(\sqrt{T}C_{NT}^{-2}).
\end{aligned}
$$

Put together, we have $\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s e_{is}\widetilde{D}_{fs}^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}u_s = O_P(\sqrt{T}C_{NT}^{-2})$.

Thus $\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s(H_1^{-1}f_s-\widetilde{f}_s)e_{is}=O_P(\sqrt{T}C_{NT}^{-2})$.
Q.E.D.

Lemma E.3. *Suppose* $\max_{it} e_{it}^4=O_P(\min\{N,T\})$. *(i)* $\frac{1}{T}\|\widetilde{F}-FH_1^{-1'}\|_F^2=O_P(C_{NT}^{-2})=\frac{1}{T}\|\widetilde{G}-GH_2^{-1'}\|_F^2$, *and* $\frac{1}{T}\sum_{t\in I^c}\|\widetilde{f}_t-H_1^{-1}f_t\|^2e_{it}^2u_{it}^2=O_P(C_{NT}^{-2})$.
(ii) $\max_i\|\widetilde{D}_{\lambda i}^{-1}\|=O_P(1)$.
(iii) $\frac{1}{N}\sum_i\|\widetilde{D}_{\lambda i}^{-1}-(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1}\|^2=O_P(C_{NT}^{-2})$.

Proof. (i) The proof is straightforward given the expansion of (E.2) and $\max_s\|\widetilde{D}_{fs}^{-1}\|=O_P(1)$. So we omit the details for brevity. (ii) Note that

$$\begin{aligned}\widetilde{D}_{\lambda i}-H_1^{-1}D_{\lambda i}H_1^{-1'} &= \sum_l \delta_l, \quad \text{where}\\ \delta_1 &= \frac{1}{T}(\widetilde{F}-FH_1^{-1'})'\mathsf{diag}(X_i)M_{\widetilde{g}}\mathsf{diag}(X_i)\widetilde{F},\\ \delta_2 &= \frac{1}{T}H_1^{-1}F'\mathsf{diag}(X_i)M_{\widetilde{g}}\mathsf{diag}(X_i)(\widetilde{F}-FH_1^{-1'})\\ \delta_3 &= \frac{1}{T}H_1^{-1}F'\mathsf{diag}(X_i)(M_{\widetilde{g}}-M_g)\mathsf{diag}(X_i)FH_1^{-1'}. \quad \text{(E.3)}\end{aligned}$$

The proof is very similar to that of Lemma E.1.

(iii)

$$\begin{aligned}&\frac{1}{N}\sum_i\|\widetilde{D}_{\lambda i}^{-1}-(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1}\|^2\\ \leq\ &\frac{1}{N}\sum_i\|\widetilde{D}_{\lambda i}-(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})\|^2\max_i\|(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-2}\|\|\widetilde{D}_{\lambda i}^{-2}\|\\ \leq\ &O_P(1)\frac{1}{N}\sum_i\|\widetilde{D}_{\lambda i}-(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})\|^2\\ =\ &O_P(1)\sum_{l=1}^3\frac{1}{N}\sum_i\|\delta_l\|^2+O_P(1)\frac{1}{N}\sum_i\|D_{\lambda i}-\bar{D}_{\lambda i}\|^2.\end{aligned}$$

We now bound each term. With the assumption that $\max_{it} x_{it}^4=O_P(\min\{N,T\})$,

$$\begin{aligned}\frac{1}{N}\sum_i\|\delta_1\|^2 &\leq \frac{1}{T}\|\widetilde{F}-FH_1^{-1'}\|_F^2\|M_{\widetilde{g}}-M_g\|^2\frac{1}{T}\|\widetilde{F}\|_F^2\frac{1}{N}\sum_i\|\mathsf{diag}(X_i)\|^4\\ &+\frac{1}{T^2}\|\widetilde{F}-FH_1^{-1'}\|_F^4\frac{1}{N}\sum_i\|\mathsf{diag}(X_i)\|^4\\ &+\frac{1}{N}\sum_i\|\frac{1}{T}(\widetilde{F}-FH_1^{-1'})'\mathsf{diag}(X_i)M_g\mathsf{diag}(X_i)FH_1^{-1'}\|^2\\ &\leq O_P(C_{NT}^{-4})\max_{it}x_{it}^4+\frac{1}{NT}\sum_i\sum_t x_{it}^2(\sum_s M_{g,ts}x_{is}f_s)^2O_P(C_{NT}^{-2})\end{aligned}$$

$$\leq \quad O_P(C_{NT}^{-2}).$$

$\frac{1}{N}\sum_i \|\delta_2\|^2$ is bounded similarly. $\frac{1}{N}\sum_i \|\delta_3\|^2 \leq O_P(1)\|M_{\widetilde{g}} - M_g\|^2 = O_P(C_{NT}^{-2})$. Finally,

$$\frac{1}{N}\sum_i \|D_{\lambda i} - \bar{D}_{\lambda i}\|^2 \leq \frac{1}{N}\sum_i \|\frac{1}{T}\sum_{st} f_s f_t' M_{g,st}(x_{is}x_{it} - \mathsf{E}x_{is}x_{it})\|_F^2 = O_P(C_{NT}^{-2}).$$

The proof of above is essentially the same as that of $\frac{1}{T}\sum_s \|D_{fs} - D_f\|^2$ in Lemma E.1 (ii), provided that e_{it} is conditionally serially independent given f_t, so we omit the details for brevity.

Lemma E.4. *Suppose* $\mathsf{Var}(u_s|e_t, e_s) < C$.
(i) For each fixed $t \in I^c$, $\frac{1}{\sqrt{N}}\sum_i (H_1'\lambda_i - \dot{\lambda}_i)e_{it} = O_P(\sqrt{N}C_{NT}^{-2})$.
(ii) For each $i \leq N$, $\lambda_i - \dot{\lambda}_i = O_P(C_{NT}^{-1})$.
(iii) For each fixed $j \leq N$, $\frac{1}{T}\sum_{t\in I^c} \|e_{jt}f_t'\|^2 \|\frac{1}{N}\sum_i \lambda_i e_{it}\mu_i(\dot{\lambda}_i - H_1'\lambda_i)'\|^2 = O_P(C_{NT}^{-4})$.

Part (iii) is used for deriving the expansion of $\widehat{\lambda}_i$ later.

Proof. Given that the $|I^c| \times 1$ vector $y_i = G\alpha_i + \mathsf{diag}(X_i)F\lambda_i + u_i$ where u_i is a $|I^c| \times 1$ vector and G is an $|I^c| \times K_1$ matrix, we have

$$\begin{aligned} \dot{\lambda}_i &= \widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}y_i \\ &= H_1'\lambda_i + \widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}(GH_2^{-1'} - \widetilde{G})H_2'\alpha_i \\ &\quad + \widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}\mathsf{diag}(X_i)(FH_1^{-1'} - \widetilde{F})H_1'\lambda_i + \widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}u_i \end{aligned} \tag{E.4}$$

Part (ii) that $\lambda_i - \dot{\lambda}_i = O_P(C_{NT}^{-1})$. follows from a straightforward calculation. So we omit the details. We now prove the much harder part (i) by plugging in each term in the above expansion to $\frac{1}{\sqrt{N}}\sum_i (H_1'\lambda_i - \dot{\lambda}_i)e_{it}$.

First term: $\frac{1}{\sqrt{N}}\sum_i e_{it}\widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}(GH_2^{-1'} - \widetilde{G})H_2'\alpha_i$.

$$\begin{aligned} &\frac{1}{\sqrt{N}}\sum_i e_{it}(\widetilde{D}_{\lambda i}^{-1} - (H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1})\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}(GH_2^{-1'} - \widetilde{G})H_2'\alpha_i \\ \leq\ & O_P(\sqrt{N}C_{NT}^{-2})(\frac{1}{N}\sum_i e_{it}^2\frac{1}{T}\sum_s \widetilde{f}_s^2 x_{is}^2)^{1/2} \\ \leq\ & O_P(\sqrt{N}C_{NT}^{-2}) + O_P(\sqrt{N}C_{NT}^{-2})(\frac{1}{T}\sum_s (\widetilde{f}_s - H_1^{-1}f_s)^2)^{1/2}\max_{is}|x_{is}| \\ \leq\ & O_P(\sqrt{N}C_{NT}^{-2}) + O_P(\sqrt{N}C_{NT}^{-3})\max_{is}|x_{is}| = O_P(\sqrt{N}C_{NT}^{-2}). \end{aligned}$$

Given F, G, $\bar{D}_{\lambda i}^{-1}$ is nonrandom. So

$$\begin{aligned} & \frac{1}{\sqrt{N}}\sum_i e_{it}(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}(GH_2^{-1'}-\widetilde{G})H_2'\alpha_i \\ \leq\ & O_P(\sqrt{N}C_{NT}^{-1})(\frac{1}{T}\sum_s \widetilde{f}_s^2(\frac{1}{N}\sum_i \alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2)^{1/2} \\ \leq\ & O_P(\sqrt{N}C_{NT}^{-1})(a^{1/2}+b^{1/2}) \quad \text{where} \\ a\ =\ & \frac{1}{T}\sum_s(\widetilde{f}_s-H_1^{-1}f_s)^2(\frac{1}{N}\sum_i \alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2 \\ b\ =\ & \frac{1}{T}\sum_s f_s^2(\frac{1}{N}\sum_i \alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2. \end{aligned}$$

We now bound each term. As for b, note that for each fixed t,

$$\begin{aligned} & \mathsf{E}((\frac{1}{N}\sum_i \alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2|F,G,u_s) \\ \leq\ & \mathsf{E}(\frac{1}{N}\sum_i \alpha_i e_{it}\bar{D}_{\lambda i}^{-1}e_{is}|F,G,u_s)^2+\mathsf{E}(\frac{1}{N}\sum_i \alpha_i e_{it}\bar{D}_{\lambda i}^{-1}\mu_i|F,G,u_s)^2 \\ \leq\ & \mathsf{Var}(\frac{1}{N}\sum_i \alpha_i e_{it}e_{is}\bar{D}_{\lambda i}^{-1}|F,G,u_s)+(\mathsf{E}\frac{1}{N}\sum_i \alpha_i e_{it}e_{is}\bar{D}_{\lambda i}^{-1}|F,G,u_s)^2 \\ & +\mathsf{Var}(\frac{1}{N}\sum_i \alpha_i e_{it}\bar{D}_{\lambda i}^{-1}\mu_i|F,G,u_s) \\ \leq\ & \frac{1}{N}\max_i\|\bar{D}_{\lambda i}^{-2}\|\frac{1}{N}\sum_{ij}|\mathsf{Cov}(e_{jt}e_{js},e_{it}e_{is}|F,G,u_s)|+(\mathsf{E}\frac{1}{N}\sum_i \alpha_i e_{it}e_{is}\bar{D}_{\lambda i}^{-1}|F,G,u_s)^2 \\ & +\frac{1}{N}\max_i\|\bar{D}_{\lambda i}^{-2}\|\frac{1}{N}\sum_{ij}|\mathsf{Cov}(e_{it},e_{jt}|F,G,u_s)| \\ \leq\ & \frac{C}{N}+(\mathsf{E}\frac{1}{N}\sum_i \alpha_i e_{it}e_{is}\bar{D}_{\lambda i}^{-1}|F,G,u_s)^2 \end{aligned} \tag{E.5}$$

with the assumption $\frac{1}{N}\sum_{ij}|\mathsf{Cov}(e_{jt}e_{js},e_{it}e_{is}|F,G,u_s)|<C$. So

$$\begin{aligned} \mathsf{E}b\ \leq\ & \frac{1}{T}\sum_s \mathsf{E}f_s^2(\frac{1}{N}\sum_i \alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2 \\ \leq\ & \frac{1}{T}\sum_s \mathsf{E}f_s^2\frac{C}{N}+\frac{1}{T}\sum_s \mathsf{E}f_s^2(\mathsf{E}\frac{1}{N}\sum_i \alpha_i e_{it}e_{is}\bar{D}_{\lambda i}^{-1}|F,G)^2 \\ \leq\ & \frac{C}{N}+\frac{1}{T}\mathsf{E}f_t^2(\mathsf{E}\frac{1}{N}\sum_i \alpha_i e_{it}^2\bar{D}_{\lambda i}^{-1}|F,G)^2=O(C_{NT}^{-2}). \\ a\ =\ & \frac{1}{T}\sum_s f_s^2(\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)(\Lambda H_1-\widetilde{\Lambda}))^2(\frac{1}{N}\sum_i \alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2 \end{aligned}$$

$$+\frac{1}{T}\sum_s(\widetilde{D}_{fs}^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)(\Lambda H_1-\widetilde{\Lambda})H_1^{-1}f_s)^2(\frac{1}{N}\sum_i\alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2$$
$$+\frac{1}{T}\sum_s(\widetilde{D}_{fs}^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}u_s)^2(\frac{1}{N}\sum_i\alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2$$

The first line of a is bounded by, using the upper bound of $\mathsf{E}b$, and $\max\|\widetilde{D}_{fs}^{-1}\|=O_P(1)$,

$$\begin{aligned}&O_P(C_{NT}^{-2})\max_{is}x_{is}^4\frac{1}{T}\sum_s f_s^2(\frac{1}{N}\sum_i\alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2\leq O_P(1)\frac{1}{T}\sum_s\mathsf{E}f_s^2(\frac{1}{N}\sum_i\alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2\\&=\ O_P(C_{NT}^{-2}).\end{aligned}$$

The second line of a is bounded similarly. The third line is bounded by:

$$\begin{aligned}&O_P(1)\frac{1}{T}\sum_s(\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}u_s)^2(\frac{1}{N}\sum_i\alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2\\\leq\ &O_P(1)\frac{1}{T}\sum_s\mathsf{E}(\frac{1}{N}\Lambda'\mathsf{diag}(X_s)M_\alpha u_s)^2(\frac{1}{N}\sum_i\alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2\\&+O_P(C_{NT}^{-2})\max_{is}x_{is}^2\frac{1}{TN}\sum_s\mathsf{E}\|u_s\|^2\mathsf{E}((\frac{1}{N}\sum_i\alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2|u_s)\\:=\ &O_P(1)(a.1+a.2).\end{aligned}$$

We now respectively bound $a.1$ and $a.2$. As for $a.1$, note that $\mathsf{Var}(u_s|e_t,e_s)<C$ almost surely, thus

$$\begin{aligned}&\mathsf{E}(\frac{1}{N}\Lambda'\mathsf{diag}(X_s)M_\alpha u_s|e_t,e_s)^2=\frac{1}{N}\frac{1}{N}\Lambda'\mathsf{diag}(X_s)M_\alpha\mathsf{Var}(u_s|e_t,e_s)M_\alpha\mathsf{diag}(X_s)\Lambda\\\leq\ &C\frac{1}{N}\frac{1}{N}\Lambda'\mathsf{diag}(X_s)^2\Lambda.\end{aligned}$$

As for $a.2$, we use (E.5). Thus,

$$\begin{aligned}a.1\ &\leq\ \frac{1}{T}\sum_s\mathsf{E}(\frac{1}{N}\sum_i\alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2\mathsf{E}(\frac{1}{N}\Lambda'\mathsf{diag}(X_s)M_\alpha u_s|e_t,e_s)^2\\&\leq\ \frac{C}{N}\frac{1}{T}\sum_s\mathsf{E}(\frac{1}{N}\sum_i\alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2\ \frac{1}{N}\Lambda'\mathsf{diag}(X_s)^2\Lambda=O(N^{-1}).\\a.2\ &\leq\ C_{NT}^{-2}\max_{is}x_{is}^2\frac{1}{TN}\sum_s\mathsf{E}\|u_s\|^2\mathsf{E}((\frac{1}{N}\sum_i\alpha_i e_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2|u_s)\\&\leq\ C_{NT}^{-2}\max_{is}x_{is}^2\frac{1}{TN}\sum_s\mathsf{E}\|u_s\|^2\frac{C}{N}\\&\quad+C_{NT}^{-2}\max_{is}x_{is}^2\frac{1}{TN}\sum_s\mathsf{E}\|u_s\|^2(\mathsf{E}\frac{1}{N}\sum_i\alpha_i e_{it}e_{is}\bar{D}_{\lambda i}^{-1}|F,u_s)^2\end{aligned}$$

$$\leq \quad O(C_{NT}^{-2}).$$

Put together, $a^{1/2}+b^{1/2}=O(C_{NT}^{-1})$. So the first term in the expansion of $\frac{1}{\sqrt{N}}\sum_i(H_1'\lambda_i-\dot{\lambda}_i)e_{it}$ is

$$\frac{1}{\sqrt{N}}\sum_i e_{it}\widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}(GH_2^{-1'}-\widetilde{G})H_2'\alpha_i=O_P(\sqrt{N}C_{NT}^{-2}).$$

Second term: $\frac{1}{\sqrt{N}}\sum_i e_{it}\widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}\mathsf{diag}(X_i)(FH_1^{-1'}-\widetilde{F})H_1'\lambda_i$.

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_i e_{it}(\widetilde{D}_{\lambda i}^{-1}-(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1})\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}\mathsf{diag}(X_i)(FH_1^{-1'}-\widetilde{F})H_1'\lambda_i\\
\leq\ & O_P(\sqrt{N}C_{NT}^{-2})(\frac{1}{NT}\sum_i e_{it}^2\|F'\mathsf{diag}(X_i)M_g\mathsf{diag}(X_i)\|^2)^{1/2}\\
&+O_P(\sqrt{N}C_{NT}^{-2})(\frac{1}{N}\sum_i e_{it}^2\|\mathsf{diag}(X_i)\|^4)^{1/2}[\|M_g-M_{\widetilde{g}}\|\frac{1}{\sqrt{T}}\|\widetilde{F}\|+\frac{1}{\sqrt{T}}\|\widetilde{F}-FH_1^{-1}\|]\\
\leq\ & O_P(\sqrt{N}C_{NT}^{-2})(\frac{1}{NT}\sum_i\sum_s e_{it}^2f_s^2x_{is}^4)^{1/2}\\
&+O_P(\sqrt{N}C_{NT}^{-2})(\frac{1}{NT}\sum_i\sum_s x_{is}^2e_{it}^2g_s^2(\frac{1}{T}\sum_k f_kg_kx_{ik})^2)^{1/2}\\
&+O_P(\sqrt{N}C_{NT}^{-3})\max_{it}x_{it}^2=O_P(\sqrt{N}C_{NT}^{-2}).
\end{aligned}$$

Also, plug in (E.2) for $\widetilde{f}_s-H_1^{-1}f_s$, and note $\frac{1}{T}\sum_s(\frac{1}{N}\sum_i\lambda_ie_{it}\bar{D}_{\lambda i}^{-1}e_{is}^2)^2=O_P(C_{NT}^{-2})$,

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_i e_{it}(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1}\frac{1}{T}(\widetilde{F}-FH_1^{-1'})'\mathsf{diag}(X_i)M_{\widetilde{g}}\mathsf{diag}(X_i)(FH_1^{-1'}-\widetilde{F})H_1'\lambda_i\\
\leq\ & O_P(\sqrt{N}C_{NT}^{-3})\max_{it}x_{it}^2\\
&+\frac{1}{\sqrt{N}}\sum_i e_{it}(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1}\frac{1}{T}(\widetilde{F}-FH_1^{-1'})'\mathsf{diag}(X_i)M_g\mathsf{diag}(X_i)(FH_1^{-1'}-\widetilde{F})H_1'\lambda_i\\
\leq\ & O_P(\sqrt{N}C_{NT}^{-2})+\sqrt{N}\frac{1}{T}\sum_s(\widetilde{f}_s-H_1^{-1}f_s)^2\frac{1}{N}\sum_i\lambda_ie_{it}\bar{D}_{\lambda i}^{-1}e_{is}^2\\
\leq\ & O_P(\sqrt{N}C_{NT}^{-2})+\max_{is}x_{is}^2O_P(\sqrt{N}C_{NT}^{-2})\frac{1}{NT}\sum_s g_s^2\sum_i\lambda_ie_{it}\bar{D}_{\lambda i}^{-1}e_{is}^2\\
&+\max_{is}x_{is}^2O_P(\sqrt{N}C_{NT}^{-2})\frac{1}{T}\sum_s\frac{1}{N}\sum_j\lambda_j^2x_{js}^2f_s^2\frac{1}{N}\sum_i\lambda_ie_{it}\bar{D}_{\lambda i}^{-1}e_{is}^2\\
&+\sqrt{N}\frac{1}{T}\sum_s(\frac{1}{N}\Lambda'\mathsf{diag}(X_s)M_\alpha u_s)^2\frac{1}{N}\sum_i\lambda_ie_{it}\bar{D}_{\lambda i}^{-1}e_{is}^2\\
\leq\ & O_P(\sqrt{N}C_{NT}^{-2}).
\end{aligned}$$

Next,

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_i e_{it}(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1\prime})^{-1}\frac{1}{T}F'\mathsf{diag}(X_i)(M_{\widetilde{g}}-M_g)\mathsf{diag}(X_i)(FH_1^{-1\prime}-\widetilde{F})H_1'\lambda_i \\
\leq\ & O_P(\sqrt{N}C_{NT}^{-1})(\frac{1}{N}\sum_i \lambda_i^2 e_{it}^2\frac{1}{T}\sum_k f_k^2 x_{ik}^2)^{1/2}(\frac{1}{NT}\sum_i\sum_s(\widetilde{f}_s-H_1^{-1}f_s)^2e_{is}^2)^{1/2} \\
=\ & O_P(\sqrt{N}C_{NT}^{-2}),
\end{aligned}$$

where the last equality is due to $\frac{1}{NT}\sum_i\sum_{s\in I^c\cup\{t\}}(\widetilde{f}_s-H_1^{-1}f_s)^2e_{is}^2 = O_P(C_{NT}^{-2})$, proved as below. Use (E.2) for $\widetilde{f}_s - H_1^{-1}f_s$,

$$\begin{aligned}
&\frac{1}{NT}\sum_i\sum_s(\widetilde{f}_s-H_1^{-1}f_s)^2e_{is}^2 \\
=\ & \frac{1}{NT}\sum_i\sum_{s\in I^c\cup\{t\}} e_{is}^2\frac{1}{N}\|\Lambda'\mathsf{diag}(X_s)\|^2[g_s^2+\frac{1}{N}\|u_s\|^2]O_P(C_{NT}^{-2}) \\
&+O_P(1)\frac{1}{N}\sum_j(H_1\lambda_j-\widetilde{\lambda}_j)^2\max_j\frac{1}{NT}\sum_i\sum_{s\in I^c\cup\{t\}}\mathsf{E}_I e_{is}^2 f_s^2\frac{1}{N}\|\Lambda'\mathsf{diag}(X_s)\|^2x_{js}^2 \\
&+\frac{1}{NT}\sum_i\sum_s e_{is}^2(\frac{1}{N}\Lambda'\mathsf{diag}(X_s)M_\alpha u_s)^2 \\
&+\frac{1}{NT}\sum_i\sum_s e_{is}^2\frac{1}{N}\|\mathsf{diag}(X_s)M_\alpha u_s\|^2 O_P(C_{NT}^{-2}) \\
=\ & O_P(C_{NT}^{-2}).
\end{aligned}$$

Next,

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_i e_{it}(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1\prime})^{-1}\frac{1}{T}F'\mathsf{diag}(X_i)M_g\mathsf{diag}(X_i)(FH_1^{-1\prime}-\widetilde{F})H_1'\lambda_i \\
\leq\ & O_P(\sqrt{N}C_{NT}^{-1})\|\frac{1}{N}\sum_i\lambda_i e_{it}\bar{D}_{\lambda i}^{-1}\frac{1}{\sqrt{T}}F'\mathsf{diag}(X_i)M_g\mathsf{diag}(X_i)\| \\
\leq\ & O_P(\sqrt{N}C_{NT}^{-1})(\frac{1}{T}\sum_s f_s^2(\frac{1}{N}\sum_i\lambda_i x_{is}^2 e_{it}\bar{D}_{\lambda i}^{-1})^2)^{1/2} \\
&+O_P(\sqrt{N}C_{NT}^{-1})(\frac{1}{T}\sum_s g_s^2(\frac{1}{N}\sum_i x_{is}\lambda_i e_{it}\bar{D}_{\lambda i}^{-1}\frac{1}{T}\sum_k f_k x_{ik}g_k)^2)^{1/2} \\
=\ & O_P(\sqrt{N}C_{NT}^{-1})(a^{1/2}+b^{1/2}).
\end{aligned}$$

We aim to show $a = O_P(C_{NT}^{-2}) = b$.

$$\mathsf{E}a \ := \ \frac{1}{T}\sum_s \mathsf{E}f_s^2(\frac{1}{N}\sum_i\lambda_i x_{is}^2 e_{it}\bar{D}_{\lambda i}^{-1})^2$$

$$
\begin{aligned}
&\leq \frac{1}{T}\sum_s \frac{1}{N^2}\sum_{ij} \mathsf{E} f_s^2 \lambda_i \bar{D}_{\lambda i}^{-1}\lambda_j \bar{D}_{\lambda j}^{-1}\mathsf{E}(e_{is}^2 e_{js}^2|F)\mathsf{Cov}(e_{it}, e_{jt}|F) \\
&\quad + \frac{1}{T}\sum_s \mathsf{E} f_s^2 \frac{1}{N^2}\sum_{ij}\lambda_i\mu_i^2 \bar{D}_{\lambda i}^{-1}\lambda_j\mu_j^2\bar{D}_{\lambda j}^{-1}\mathsf{Cov}(e_{it}, e_{jt}|F) \\
&\quad + \frac{2}{T}\sum_s \mathsf{E} f_s^2 \frac{1}{N^2}\sum_{ij}\lambda_i\mu_i \bar{D}_{\lambda i}^{-1}\bar{D}_{\lambda j}^{-1}\lambda_i\mu_j \mathsf{E}(e_{js}e_{is}|F)\mathsf{Cov}(e_{jt}e_{it}|F) \\
&= O_P(N^{-1}). \\
\mathsf{E} b &:= \frac{1}{T}\sum_s \mathsf{E} g_s^2 (\frac{1}{N}\sum_i x_{is}\lambda_i e_{it}\bar{D}_{\lambda i}^{-1}\frac{1}{T}\sum_k f_k x_{ik} g_k)^2 \\
&= O(T^{-1}) + O(1)\frac{1}{T}\sum_s \mathsf{E} g_s^2 \frac{1}{N^2}\sum_{ij} |\mathsf{E}(x_{js}x_{is}|F)\mathsf{Cov}(e_{it}, e_{jt}|F)| \\
&= O_P(C_{NT}^{-2}).
\end{aligned}
$$

Therefore the second term is

$$
\frac{1}{\sqrt{N}}\sum_i e_{it}\widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}\mathsf{diag}(X_i)(FH_1^{-1'} - \widetilde{F})H_1'\lambda_i = O_P(\sqrt{N}C_{NT}^{-2}).
$$

The third term: $\frac{1}{\sqrt{N}}\sum_i e_{it}\widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}u_i$. First,

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_i e_{it}(\widetilde{D}_{\lambda i}^{-1} - (H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1})\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}u_i \\
\leq\ & O_P(\sqrt{N}C_{NT}^{-2}) + O_P(1)\frac{1}{\sqrt{N}}\sum_i e_{it}(\widetilde{D}_{\lambda i}^{-1} - (H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1})\frac{1}{T}F'\mathsf{diag}(X_i)M_g u_i \\
\leq\ & O_P(\sqrt{N}C_{NT}^{-2}) + O_P(\sqrt{N}C_{NT}^{-1})\frac{1}{N}\sum_i |e_{it}||\frac{1}{T}\sum_s f_s x_{is}u_{is}| \\
&+ O_P(\sqrt{N}C_{NT}^{-1})\frac{1}{N}\sum_i |e_{it}||\frac{1}{T}\sum_s f_s x_{is} g_s||\frac{1}{T}\sum_k g_k u_{ik}| \\
=\ & O_P(\sqrt{N}C_{NT}^{-2}).
\end{aligned}
$$

Second,

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_i e_{it}\bar{D}_{\lambda i}^{-1}\frac{1}{T}(\widetilde{F} - FH_1^{-1'})'\mathsf{diag}(X_i)M_{\widetilde{g}}u_i \\
\leq\ & \frac{1}{\sqrt{N}}\sum_i e_{it}\bar{D}_{\lambda i}^{-1}\frac{1}{T}(\widetilde{F} - FH_1^{-1'})'\mathsf{diag}(X_i)M_g u_i \\
&+ O_P(\sqrt{N}C_{NT}^{-1})(\frac{1}{NT}\sum_i e_{it}^2\|u_i\|^2)^{1/2}(\frac{1}{NT}\sum_i \|(\widetilde{F} - FH_1^{-1'})'\mathsf{diag}(X_i)\|^2)^{1/2} \\
\leq\ & O_P(\sqrt{N}C_{NT}^{-2}) + O_P(\sqrt{N}C_{NT}^{-1})\|\frac{1}{N}\sum_i \frac{1}{\sqrt{T}}\mathsf{diag}(X_i)M_g u_i e_{it}\bar{D}_{\lambda i}^{-1}\|
\end{aligned}
$$

$$
\begin{aligned}
&\leq O_P(\sqrt{N}C_{NT}^{-2}) + O_P(\sqrt{N}C_{NT}^{-1})(\frac{1}{T}\sum_s(\frac{1}{N}\sum_i x_{is}M_{g,s}'u_ie_{it}\bar{D}_{\lambda i}^{-1})^2)^{1/2} \\
&= O_P(\sqrt{N}C_{NT}^{-2}),
\end{aligned}
$$

where the last equality is due to $\frac{1}{T}\sum_s(\frac{1}{N}\sum_i x_{is}M_{g,s}'u_ie_{it}\bar{D}_{\lambda i}^{-1})^2 = O_P(C_{NT}^{-2})$, proved as follows:

$$
\begin{aligned}
&\frac{1}{T}\sum_s \mathsf{E}(\frac{1}{N}\sum_i x_{is}M_{g,s}'u_ie_{it}\bar{D}_{\lambda i}^{-1})^2 \\
\leq\ & \frac{1}{T}\sum_s \mathsf{E}(\frac{1}{N}\sum_i x_{is}u_{is}e_{it}\bar{D}_{\lambda i}^{-1})^2 + \frac{1}{T}\sum_s \mathsf{E}(\frac{1}{NT}\sum_i\sum_k x_{is}g_kg_su_{ik}e_{it}\bar{D}_{\lambda i}^{-1})^2 \\
\leq\ & O(T^{-1}) + \frac{1}{T}\sum_{s\neq t}\frac{1}{N^2}\sum_{ij}\mathsf{E}|\mathsf{E}(e_{jt}e_{it}|F)\mathsf{E}(u_{is}u_{js}x_{js}x_{is}|F)| \\
&+\frac{1}{T}\sum_s \mathsf{E}\frac{1}{NT}\sum_i\sum_k \bar{D}_{\lambda i}^{-1}\frac{1}{NT}\sum_j\sum_l \bar{D}_{\lambda j}^{-1}|x_{js}g_lg_su_{jl}x_{is}g_kg_su_{ik}||\mathsf{Cov}(e_{it}, e_{jt}|F)| \\
\leq\ & O(T^{-1}) + O(N^{-1})
\end{aligned}
$$

where the last equality is due to $\max_j\sum_i |\mathsf{Cov}(e_{it}, e_{jt}|F)| < C$.

Next,

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_i e_{it}\bar{D}_{\lambda i}^{-1}\frac{1}{T}F'\mathsf{diag}(X_i)(M_{\widetilde{g}} - M_g)u_i \\
=\ & \mathsf{tr}\frac{1}{\sqrt{N}}\sum_i u_ie_{it}\bar{D}_{\lambda i}^{-1}\frac{1}{T}F'\mathsf{diag}(X_i)(M_{\widetilde{g}} - M_g) \\
\leq\ & O_P(\sqrt{N}C_{NT}^{-1})\frac{1}{T}\|\frac{1}{N}\sum_i u_ie_{it}\bar{D}_{\lambda i}^{-1}F'\mathsf{diag}(X_i)\|_F \\
\leq\ & O_P(\sqrt{N}C_{NT}^{-1})(\frac{1}{T^2}\sum_{sk}\mathsf{E}(\frac{1}{N}\sum_i u_{is}e_{it}\bar{D}_{\lambda i}^{-1}f_kx_{ik})^2)^{1/2} = O_P(\sqrt{N}C_{NT}^{-2})
\end{aligned}
$$

where the last equality is due to $\frac{1}{T^2}\sum_{sk}\mathsf{E}(\frac{1}{N}\sum_i u_{is}e_{it}\bar{D}_{\lambda i}^{-1}f_kx_{ik})^2 = O_P(C_{NT}^{-2})$.

$$
\begin{aligned}
&\frac{1}{T^2}\sum_{sk}\mathsf{E}(\frac{1}{N}\sum_i u_{is}e_{it}\bar{D}_{\lambda i}^{-1}f_kx_{ik})^2 \\
=\ & \frac{1}{T^2}\sum_{sk}\mathsf{E}\frac{1}{N}\sum_i \bar{D}_{\lambda i}^{-1}\frac{1}{N}\sum_j \bar{D}_{\lambda j}^{-1}f_k^2x_{jk}u_{is}x_{ik}u_{js}e_{jt}e_{it} \\
\leq\ & O(T^{-1}) + \frac{1}{T^2}\sum_{sk}\frac{1}{N^2}\sum_{ij}\mathsf{E}\bar{D}_{\lambda i}^{-1}\bar{D}_{\lambda j}^{-1}f_k^2\mathsf{E}(x_{jk}u_{is}x_{ik}u_{js}|F)\mathsf{Cov}(e_{jt}, e_{it}|F) \\
\leq\ & O_P(C_{NT}^{-2}).
\end{aligned}
$$

Finally, since u_{it} is conditionally serially independent given E, F,

$$\begin{aligned}
&\mathsf{E}(\frac{1}{\sqrt{N}}\sum_i e_{it}\bar{D}_{\lambda i}^{-1}\frac{1}{T}F'\mathsf{diag}(X_i)M_g u_i)^2 \\
=\;& \mathsf{E}\frac{1}{N}\sum_{ij}\sum_s \bar{D}_{\lambda j}^{-1}\bar{D}_{\lambda i}^{-1}\frac{1}{T}(F'\mathsf{diag}(X_i)M_{g,s})^2 e_{jt}e_{it}\frac{1}{T}u_{js}u_{is} \\
\leq\;& \frac{C}{T^2N}\sum_{ij}\sum_s \mathsf{E}u_{js}u_{is}\mathsf{E}((F'\mathsf{diag}(X_i)M_{g,s})^2 e_{jt}e_{it}|U) \\
\leq\;& \frac{C}{T^2N}\sum_{ij}\sum_s |\mathsf{Cov}(u_{js}, u_{is})| = O(T^{-1}) = O_P(C_{NT}^{-2}).
\end{aligned}$$

Together, $\frac{1}{\sqrt{N}}\sum_i(H_1'\lambda_i - \dot{\lambda}_i)e_{it} = O_P(\sqrt{T}C_{NT}^{-2})$. Q.E.D.

The proof of part (iii) follows from the same arguments as in part (i). While a rigorous proof still follows from substituting in the expansion of $\dot{\lambda}_i - H_1'\lambda_i$, the details would be mostly very similar. So we omit it for brevity.

E.2. Technical lemmas for $\widehat{f}_t$.

Lemma E.5. *Suppose* $\max_i|\bar{e}_i| = O_P(1)$. *For each fixed* t, *(i)* $\widehat{B}_t - B = O_P(C_{NT}^{-1})$. *(ii)* $\widehat{S}_t - S = O_P(C_{NT}^{-2})$.

Proof. Throughout the proof, we assume $\dim(\alpha_i) = \dim(\lambda_i) = 1$ without loss of generality.

(i) $\widehat{B}_t - B = b_1 + b_2$, where

$$\begin{aligned}
b_1 &= \frac{1}{N}\sum_i\begin{pmatrix}\widetilde{\lambda}_i\widetilde{\lambda}_i'\widehat{e}_{it}^2 - H_1'\lambda_i\lambda_i'H_1e_{it}^2 & \widetilde{\lambda}_i\widetilde{\alpha}_i'\widehat{e}_{it} - H_1'\lambda_i\alpha_iH_2'e_{it} \\ \widetilde{\alpha}_i\widetilde{\lambda}_i'\widehat{e}_{it} - H_2'\alpha_i\lambda_iH_1'e_{it} & \widetilde{\alpha}_i\widetilde{\alpha}_i' - H_2'\alpha_i\alpha_i'H_2\end{pmatrix} \\
b_2 &= \frac{1}{N}\sum_i\begin{pmatrix}H_1'\lambda_i\lambda_i'H_1(e_{it}^2 - \mathsf{E}e_{it}^2) & H_1'\lambda_i\alpha_iH_2'e_{it} \\ H_2'\alpha_i\lambda_iH_1'e_{it} & 0\end{pmatrix}.
\end{aligned}$$

To prove the convergence of b_1, first note that

$$\begin{aligned}
&\frac{1}{N}\sum_i\widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2 - e_{it}^2) = \frac{1}{N}\sum_i\widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it} - e_{it})^2 + \frac{2}{N}\sum_i\widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it} - e_{it})e_{it} \\
\leq\;& \frac{1}{N}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)(\widetilde{\lambda}_i - H_1'\lambda_i)'(\widehat{e}_{it} - e_{it})^2 + \frac{1}{N}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)\lambda_i'H_1'(\widehat{e}_{it} - e_{it})^2 \\
&+O_P(1)\frac{1}{N}\sum_i(\widehat{e}_{it} - e_{it})^2 + \frac{2}{N}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)(\widetilde{\lambda}_i - H_1'\lambda_i)'(\widehat{e}_{it} - e_{it})e_{it} \\
&+\frac{2}{N}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)\lambda_i'H_1'(\widehat{e}_{it} - e_{it})e_{it} + O_P(1)\frac{1}{N}\sum_i\lambda_i\lambda_i'(\widehat{e}_{it} - e_{it})e_{it}
\end{aligned}$$

$$
\begin{aligned}
&\leq O_P(C_{NT}^{-2}) \max_{it} |\widehat{e}_{it} - e_{it}| \max_{it} |e_{it}| + O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i (\widehat{e}_{it} - e_{it})^4)^{1/2} \\
&+O_P(1)\frac{1}{N}\sum_i (\widehat{e}_{it} - e_{it})^2 + O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i (\widehat{e}_{it} - e_{it})^2 e_{it}^2)^{1/2} \\
&+O_P(1)\frac{1}{N}\sum_i \lambda_i\lambda_i'(\widehat{e}_{it} - e_{it})e_{it}. \qquad (E.6)
\end{aligned}
$$

In addition,

$$
\begin{aligned}
&\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\alpha}_i'(\widehat{e}_{it} - e_{it}) = \frac{1}{N}\sum_i (\widetilde{\lambda}_i - H_1'\lambda_i)(\widetilde{\alpha}_i - H_2'\alpha_i)'(\widehat{e}_{it} - e_{it}) \\
&+\frac{1}{N}\sum_i (\widetilde{\lambda}_i - H_1'\lambda_i)\alpha_i' H_2(\widehat{e}_{it} - e_{it}) + O_P(1)\frac{1}{N}\sum_i \lambda_i\alpha_i'(\widehat{e}_{it} - e_{it}) \\
\leq\ & O_P(C_{NT}^{-2}) \max_{it} |\widehat{e}_{it} - e_{it}| + O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i (\widehat{e}_{it} - e_{it})^2)^{1/2} \\
&+O_P(1)\frac{1}{N}\sum_i \lambda_i\alpha_i'(\widehat{e}_{it} - e_{it}). \qquad (E.7)
\end{aligned}
$$

When $\widehat{e}_{it} - e_{it} = \mu_i - \bar{x}_i = -\bar{e}_i$, we assume that $C_{NT}^{-1}\max_{it}|\bar{e}_i|\max_{it}|e_{it}| = O_P(1)$. Also,
$\frac{1}{N}\sum_i (\widehat{e}_{it} - e_{it})^4 = \frac{1}{N}\sum_i \bar{e}_i^4 = O_P(T^{-1})$ and $\frac{1}{N}\sum_i (\widehat{e}_{it} - e_{it})^2 = \frac{1}{N}\sum_i \bar{e}_i^2 = O_P(T^{-1})$. Also, $\frac{1}{N}\sum_i (\widehat{e}_{it} - e_{it})^2 e_{it}^2 = O_P(T^{-1})$ and $\frac{1}{N}\sum_i \lambda_i\lambda_i'(\widehat{e}_{it} - e_{it})e_{it} = O_P(C_{NT}^{-2})$, and $\frac{1}{N}\sum_i \lambda_i\alpha_i'(\widehat{e}_{it} - e_{it}) = O_P(C_{NT}^{-2})$. Thus

$$
\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2 - e_{it}^2) = O_P(C_{NT}^{-2}), \text{ and } \frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\alpha}_i'(\widehat{e}_{it} - e_{it}) = O_P(C_{NT}^{-1}). \qquad (E.8)
$$

So the first term of b_1 is, due to the serial independence of e_{it}^2,

$$
\begin{aligned}
&\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\lambda}_i'\widehat{e}_{it}^2 - H_1'\lambda_i\lambda_i' H_1 e_{it}^2 \\
\leq\ & \frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2 - e_{it}^2) + \frac{1}{N}\sum_i (\widetilde{\lambda}_i\widetilde{\lambda}_i' - H_1'\lambda_i\lambda_i' H_1)e_{it}^2 \\
\leq\ & O_P(C_{NT}^{-2}) + O_P(1)\frac{1}{N}\sum_i \|\widetilde{\lambda}_i\widetilde{\lambda}_i' - H_1'\lambda_i\lambda_i' H_1\|_F \mathsf{E}_I e_{it}^2 \\
\leq\ & O_P(C_{NT}^{-2}) + O_P(1)\frac{1}{N}\sum_i \|\widetilde{\lambda}_i\widetilde{\lambda}_i' - H_1'\lambda_i\lambda_i' H_1\|_F = O_P(C_{NT}^{-1}).
\end{aligned}
$$

The second term is,

$$\begin{aligned}
&\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\alpha}_i'\widehat{e}_{it} - H_1'\lambda_i\alpha_i H_2' e_{it} \\
\leq\ & O_P(C_{NT}^{-1}) + O_P(1)\frac{1}{N}\sum_i \|\widetilde{\lambda}_i\widetilde{\alpha}_i' - H_1'\lambda_i\alpha_i H_2'\|_F \mathsf{E}_I|e_{it}| \leq O_P(C_{NT}^{-1}).
\end{aligned}$$

The third term of b_1 is bounded similarly. The last term of b_1 is easy to show to be $O_P(C_{NT}^{-1})$. As for b_2, by the assumption that $\frac{1}{N}\sum_{ij}|\mathsf{Cov}(e_{it}^2, e_{jt}^2)| < C$, thus $b_2 = O_P(N^{-1/2})$. Hence $\widehat{B}_t - B = O_P(C_{NT}^{-1})$.

(ii) $\widehat{S}_t - S = c_t + d_t$,

$$\begin{aligned}
c_t &= \frac{1}{N}\sum_i \begin{pmatrix} \widetilde{\lambda}_i\lambda_i' H_1 e_{it}(\widehat{e}_{it} - e_{it}) + \widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2 - e_{it}^2) & \widetilde{\lambda}_i(\widehat{e}_{it} - e_{it})(\alpha_i' H_2 - \widetilde{\alpha}_i') \\ \widetilde{\alpha}_i\lambda_i' H_1(e_{it} - \widehat{e}_{it}) + \widetilde{\alpha}_i(\lambda_i' H_1 - \widetilde{\lambda}_i')(\widehat{e}_{it} - e_{it}) & 0 \end{pmatrix} \\
d_t &= \frac{1}{N}\sum_i \begin{pmatrix} (\widetilde{\lambda}_i\lambda_i' H_1 - \widetilde{\lambda}_i\widetilde{\lambda}_i')e_{it}^2 - H_1'\lambda_i(\lambda_i' H_1 - \widetilde{\lambda}_i')\mathsf{E}e_{it}^2 & \widetilde{\lambda}_i e_{it}(\alpha_i' H_2 - \widetilde{\alpha}_i') \\ \widetilde{\alpha}_i(\lambda_i' H_1 - \widetilde{\lambda}_i')e_{it} & 0 \end{pmatrix}.
\end{aligned}$$

As for c_t, note that when $\widehat{e}_{it} - e_{it} = -\bar{e}_i$, we have

$$\begin{aligned}
&\frac{1}{N}\sum_i \widetilde{\lambda}_i\lambda_i' H_1 e_{it}(\widehat{e}_{it} - e_{it}) \leq \frac{1}{N}\sum_i (\widetilde{\lambda}_i - H_1'\lambda_i)\lambda_i' e_{it}(\widehat{e}_{it} - e_{it}) + O_P(1)\frac{1}{N}\sum_i \lambda_i\lambda_i e_{it}(\widehat{e}_{it} - e_{it}) \\
\leq\ & O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i e_{it}^2\bar{e}_i^2)^{1/2} + O_P(1)\frac{1}{N}\sum_i \lambda_i\lambda_i e_{it}\bar{e}_i = O_P(C_{NT}^{-2}).
\end{aligned}$$

Also, by (E.8), $\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2 - e_{it}^2) = O_P(C_{NT}^{-2})$. For the second term of c,

$$\begin{aligned}
&\frac{1}{N}\sum_i \widetilde{\lambda}_i(\widehat{e}_{it} - e_{it})(\alpha_i' H_2 - \widetilde{\alpha}_i') \\
\leq\ & \frac{1}{N}\sum_i (\widetilde{\lambda}_i - H_1'\lambda_i)(\widehat{e}_{it} - e_{it})(\alpha_i' H_2 - \widetilde{\alpha}_i') + O_P(1)\frac{1}{N}\sum_i \lambda_i(\widehat{e}_{it} - e_{it})(\alpha_i' H_2 - \widetilde{\alpha}_i') \\
\leq\ & O_P(C_{NT}^{-2})\max_i|\bar{e}_i| + O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i \bar{e}_i^2)^{1/2} = O_P(C_{NT}^{-2})
\end{aligned}$$

given that $\max_i|\bar{e}_i| = O_P(1)$. For the third term of c_t, similarly, $\frac{1}{N}\sum_i \widetilde{\alpha}_i(\lambda_i' H_1 - \widetilde{\lambda}_i')(\widehat{e}_{it} - e_{it}) = O_P(C_{NT}^{-2})$. Also,

$$\begin{aligned}
&\frac{1}{N}\sum_i \widetilde{\alpha}_i\lambda_i' H_1(e_{it} - \widehat{e}_{it}) \\
\leq\ & O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i (\widehat{e}_{it} - e_{it})^2)^{1/2} + O_P(1)\frac{1}{N}\sum_i \lambda_i\alpha_i'(\widehat{e}_{it} - e_{it}) \\
\leq\ & O_P(C_{NT}^{-2}).
\end{aligned}$$

So $c_t = O_P(C_{NT}^{-2})$.

As for d_t, we first prove that $\frac{1}{N}\sum_i \lambda_i(\lambda_i' H_1 - \widetilde{\lambda}_i')(e_{it}^2 - \mathsf{E}e_{it}^2) = O_P(C_{NT}^{-1}N^{-1/2})$. Note that $\mathsf{E}_I\frac{1}{N}\sum_i \lambda_i(\lambda_i' H_1 - \widetilde{\lambda}_i')(e_{it}^2 - \mathsf{E}e_{it}^2) = 0$. Let Υ_t be an $N\times 1$ vector of e_{it}^2, and $\mathsf{diag}(\Lambda)$ be diagonal matrix consisting of elements of Λ. Then

$$\|\mathsf{Var}(\Upsilon_t)\| \leq \max_i \sum_j |\mathsf{Cov}(e_{it}^2, e_{jt}^2)| < C,$$

and $\frac{1}{N}\sum_i \lambda_i(\lambda_i' H_1 - \widetilde{\lambda}_i')e_{it}^2 = \frac{1}{N}(H_1\Lambda - \widetilde{\Lambda})'\mathsf{diag}(\Lambda)\Upsilon_t$. So

$$\begin{aligned}
&\mathsf{Var}_I(\frac{1}{N}\sum_i \lambda_i(\lambda_i' H_1 - \widetilde{\lambda}_i')e_{it}^2) = \frac{1}{N^2}(H_1\Lambda - \widetilde{\Lambda})'\mathsf{diag}(\Lambda)\mathsf{Var}(\Upsilon_t)\mathsf{diag}(\Lambda)(H_1\Lambda - \widetilde{\Lambda}) \\
\leq\ & C\frac{1}{N^2}\|H_1\Lambda - \widetilde{\Lambda}\|_F^2 = O_P(C_{NT}^{-2}N^{-1}).
\end{aligned}$$

This implies $\frac{1}{N}\sum_i \lambda_i(\lambda_i' H_1 - \widetilde{\lambda}_i')(e_{it}^2 - \mathsf{E}e_{it}^2) = O_P(C_{NT}^{-1}N^{-1/2})$.

Thus the first term of d_t is

$$\begin{aligned}
& \frac{1}{N}\sum_i (\widetilde{\lambda}_i\lambda_i' H_1 - \widetilde{\lambda}_i\widetilde{\lambda}_i')e_{it}^2 - H_1'\lambda_i(\lambda_i' H_1 - \widetilde{\lambda}_i')\mathsf{E}e_{it}^2 \\
\leq\ & \frac{1}{N}\sum_i (\widetilde{\lambda}_i - H_1'\lambda_i)(\lambda_i' H_1 - \widetilde{\lambda}_i')e_{it}^2 + \frac{1}{N}\sum_i H_1'\lambda_i(\lambda_i' H_1 - \widetilde{\lambda}_i')(e_{it}^2 - \mathsf{E}e_{it}^2) \\
=\ & O_P(1)\frac{1}{N}\sum_i (\widetilde{\lambda}_i - H_1'\lambda_i)(\lambda_i' H_1 - \widetilde{\lambda}_i')\mathsf{E}_I e_{it}^2 + O_P(C_{NT}^{-1}N^{-1/2}) = O_P(C_{NT}^{-2}).
\end{aligned}$$

As for the second term of d_t, $\frac{1}{N}\sum_i \widetilde{\lambda}_i e_{it}(\alpha_i' H_2 - \widetilde{\alpha}_i')$, note that

$$\begin{aligned}
& \frac{1}{N}\sum_i (\widetilde{\lambda}_i - H_1'\lambda_i)e_{it}(\alpha_i' H_2 - \widetilde{\alpha}_i') = O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i e_{it}^2(\alpha_i' H_2 - \widetilde{\alpha}_i')^2)^{1/2} \\
\leq\ & O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i (\alpha_i' H_2 - \widetilde{\alpha}_i')^2\mathsf{E}_I e_{it}^2)^{1/2} = O_P(C_{NT}^{-2}).
\end{aligned}$$

And, $\mathsf{E}_I\frac{1}{N}\sum_i \widetilde{\lambda}_i e_{it}(\alpha_i' H_2 - \widetilde{\alpha}_i') = 0$.

$$\begin{aligned}
& \mathsf{Var}_I(\frac{1}{N}\sum_i \lambda_i e_{it}(\alpha_i' H_2 - \widetilde{\alpha}_i')) = \frac{1}{N^2}\mathsf{Var}_I((AH_2 - \widetilde{A})'\mathsf{diag}(\Lambda)e_t) \\
\leq\ & \frac{1}{N^2}(AH_2 - \widetilde{A})'\mathsf{diag}(\Lambda)\mathsf{Var}(e_t)\mathsf{diag}(\Lambda)(AH_2 - \widetilde{A}) = O_P(C_{NT}^{-2}N^{-1}),
\end{aligned}$$

implying $\frac{1}{N}\sum_i \widetilde{\lambda}_i e_{it}(\alpha_i' H_2 - \widetilde{\alpha}_i') = O_P(C_{NT}^{-2})$. Finally, the third term of d_t, $\frac{1}{N}\sum_i \widetilde{\lambda}_i e_{it}(\alpha_i' H_2 - \widetilde{\alpha}_i')$, is bounded similarly. So $d_t = O_P(C_{NT}^{-2})$.

Q.E.D.

Lemma E.6. *Suppose* $C_{NT}^{-1}\max_{it}|e_{it}|^2+\max_t\|\frac{1}{N}\sum_i\lambda_i\lambda_i'\bar{e}_ie_{it}\|_F=o_P(1)$.

(i) $\max_t\|\widehat{B}_t^{-1}\|=O_P(1)$.

(ii) $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|\widehat{B}_s^{-1}-B^{-1}\|^2=O_P(C_{NT}^{-2})$.

(iii) $\frac{1}{T}\sum_{t\in I^c}\|\widehat{S}_t-S\|^2=O_P(C_{NT}^{-4})$.

Proof. Define

$$B_t=\frac{1}{N}\sum_i\begin{pmatrix}H_1'\lambda_i\lambda_i'H_1e_{it}^2 & 0\\ 0 & H_2'\alpha_i\alpha_i'H_2\end{pmatrix}.$$

Then $\widehat{B}_t-B_t=b_{1t}+b_{2t}$,

$$\begin{aligned}
b_{1t} &= \frac{1}{N}\sum_i\begin{pmatrix}\widetilde{\lambda}_i\widetilde{\lambda}_i'\widehat{e}_{it}^2-H_1'\lambda_i\lambda_i'H_1e_{it}^2 & \widetilde{\lambda}_i\widetilde{\alpha}_i'\widehat{e}_{it}-H_1'\lambda_i\alpha_iH_2'e_{it}\\ \widetilde{\alpha}_i\widetilde{\lambda}_i'\widehat{e}_{it}-H_2'\alpha_i\lambda_iH_1'e_{it} & \widetilde{\alpha}_i\widetilde{\alpha}_i'-H_2'\alpha_i\alpha_i'H_2\end{pmatrix}\\
b_{2t} &= \frac{1}{N}\sum_i\begin{pmatrix}0 & H_1'\lambda_i\alpha_iH_2'e_{it}\\ H_2'\alpha_i\lambda_iH_1'e_{it} & 0\end{pmatrix}.
\end{aligned}$$

(i) We now show $\max_t|b_{1t}|=o_P(1)$. In addition, by assumption $\max_t|b_{2t}|=o_P(1)$. Thus $\max_t|\widehat{B}_t-B_t|=o_P(1)$, and thus $\max_t\|\widehat{B}_t^{-1}\|=o_P(1)+\max_t\|B_t^{-1}\|=O_P(1)$, by the assumption that $\|B_t^{-1}\|=O_P(1)$ uniformly in t. To show $\max_t|b_{1t}|=o_P(1)$, note that:

First term:

$$\begin{aligned}
&\max_t\|\frac{1}{N}\sum_i\widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2-e_{it}^2)\|\\
\leq\ & O_P(1)(\frac{1}{N}\sum_i\|\widetilde{\lambda}_i-H_1'\lambda_i\|^2)^{1/2}\max_t[\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^4+(\widehat{e}_{it}-e_{it})^2e_{it}^2]^{1/2}\\
&+\frac{1}{N}\sum_i\|\widetilde{\lambda}_i-H_1'\lambda_i\|^2[2\max_t|(\widehat{e}_{it}-e_{it})e_{it}|+\max_t(\widehat{e}_{it}-e_{it})^2]\\
&+O_P(1)\max_t\|\frac{1}{N}\sum_i\lambda_i\lambda_i'(\widehat{e}_{it}-e_{it})e_{it}\|_F\\
&+O_P(1)\max_t\|\frac{1}{N}\sum_i\lambda_i\lambda_i'(\widehat{e}_{it}-e_{it})^2\|_F. \qquad (E.9)
\end{aligned}$$

When $\widehat{e}_{it}-e_{it}=-\bar{e}_i$, we have

$$\max_t[\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^4+(\widehat{e}_{it}-e_{it})^2e_{it}^2]^{1/2}\leq[\frac{1}{N}\sum_i\bar{e}_i^4]^{1/2}+[\frac{1}{N}\sum_i\bar{e}_i^2]^{1/2}\max_{it}|e_{it}|$$

and $2\max_t |(\widehat{e}_{it}-e_{it})e_{it}| + \max_t(\widehat{e}_{it}-e_{it})^2 \leq 2|\bar{e}_i|\max_t |e_{it}| + \bar{e}_i^2$,

$$\begin{aligned}\max_t \|\frac{1}{N}\sum_i \lambda_i\lambda_i'(\widehat{e}_{it}-e_{it})^2\|_F &\leq \|\frac{1}{N}\sum_i \lambda_i\lambda_i'\bar{e}_i^2\|_F = O_P(T^{-1})\\ \max_t \|\frac{1}{N}\sum_i \lambda_i\lambda_i'(\widehat{e}_{it}-e_{it})e_{it}\|_F &\leq \max_t \|\frac{1}{N}\sum_i \lambda_i\lambda_i'\bar{e}_i e_{it}\|_F.\end{aligned}$$

Thus, with the assumption $C_{NT}^{-1}\max_{it}|e_{it}|^2 + \max_t \|\frac{1}{N}\sum_i \lambda_i\lambda_i'\bar{e}_i e_{it}\|_F = o_P(1)$,

$$\max_t \|\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2 - e_{it}^2)\| = o_P(1) + O_P(C_{NT}^{-2})\max_{it}|e_{it}| + \max_t \|\frac{1}{N}\sum_i \lambda_i\lambda_i'\bar{e}_i e_{it}\|_F = o_P(1).$$

In addition, $\max_t \|\frac{1}{N}\sum_i(\widetilde{\lambda}_i\widetilde{\lambda}_i' - H_1'\lambda_i\lambda_i'H_1)e_{it}^2\| \leq O_P(C_{NT}^{-1})\max_{it} e_{it}^2 = o_P(1)$. So the first term of $\max_t|b_{1t}|$ is $o_P(1)$.

Second term,

$$\begin{aligned}&\max_t \|\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\alpha}_i'(\widehat{e}_{it}-e_{it})\|_F\\ \leq\ &\max_{it}|\widehat{e}_{it}-e_{it}|\frac{1}{N}\sum_i \|(\widetilde{\lambda}_i - H_1'\lambda_i)(\widetilde{\alpha}_i - H_2'\alpha_i)'\|_F\\ &+O_P(1)(\frac{1}{N}\sum_i\|\widetilde{\lambda}_i - H_1'\lambda_i\|^2 + (\frac{1}{N}\sum_i\|\widetilde{\alpha}_i - H_2'\alpha_i\|^2)^{1/2}\max_t(\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2)^{1/2}\\ &+O_P(1)\max_t\|\frac{1}{N}\sum_i \lambda_i\alpha_i'(\widehat{e}_{it}-e_{it})\|_F. \qquad \text{(E.10)}\end{aligned}$$

When $\widehat{e}_{it} - e_{it} = -\bar{e}_i$, the above is $o_P(1)$ given that $C_{NT}^{-2}\max_i|\bar{e}_i| = o_P(1)$. Next,

$$\max_t \|\frac{1}{N}\sum_i(\widetilde{\lambda}_i\widetilde{\alpha}_i' - H_1'\lambda_i\alpha_i H_2')e_{it}\|_F = O_P(C_{NT}^{-1})\max_{it}|e_{it}| = o_P(1).$$

So the second term of $\max_t|b_{1t}|$ is $o_P(1)$. Similarly, the third term is also $o_P(1)$. Finally, the last term $\|\frac{1}{N}\sum_i \widetilde{\alpha}_i\widetilde{\alpha}_i' - H_2'\alpha_i\alpha_i'H_2\|_F = o_P(1)$.

(ii) Because we have proved $\max_t\|\widehat{B}_t^{-1}\| = O_P(1)$, it suffices to prove $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|\widehat{B}_s - B\|_F^2 = O_P(C_{NT}^{-2})$, or $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|b_{1t}\|_F^2 = O_P(C_{NT}^{-2}) = \frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|b_{2t}\|_F^2$.

First term of b_{1t}: by (E.6), the first term is bounded by

$$\begin{aligned}&\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\lambda}_i'\widehat{e}_{it}^2 - H_1'\lambda_i\lambda_i'H_1 e_{it}^2\|_F^2\\ \leq\ &\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2 - e_{it}^2)\|_F^2 + \frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i(\widetilde{\lambda}_i\widetilde{\lambda}_i' - H_1'\lambda_i\lambda_i'H_1)e_{it}^2\|_F^2\end{aligned}$$

$$\begin{aligned}
&\leq O_P(C_{NT}^{-4})\max_{it}|\widehat{e}_{it}-e_{it}|^2\max_{it}|e_{it}|^2+O_P(C_{NT}^{-2})\frac{1}{T}\sum_{t\in I^c}\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^4\\
&+O_P(1)\frac{1}{T}\sum_{t\in I^c}(\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2)^2+O_P(C_{NT}^{-2})\frac{1}{T}\sum_{t\in I^c}\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2e_{it}^2\\
&+O_P(1)\frac{1}{T}\sum_{t\in I^c}(\frac{1}{N}\sum_i\lambda_i\lambda_i'(\widehat{e}_{it}-e_{it})e_{it})^2\\
&+O_P(1)\mathsf{tr}(\frac{1}{N^2}\sum_{ij}(\widetilde{\lambda}_i\widetilde{\lambda}_i'-H_1'\lambda_i\lambda_i'H_1)(\widetilde{\lambda}_j\widetilde{\lambda}_j'-H_1'\lambda_j\lambda_i'H_1)\frac{1}{T}\sum_{t\in I^c}\mathsf{E}_Ie_{jt}^2e_{it}^2)\\
&\leq O_P(C_{NT}^{-2})+O_P(1)(\frac{1}{N}\sum_i\|(\widetilde{\lambda}_i\widetilde{\lambda}_i'-H_1'\lambda_i\lambda_i'H_1\|_F)^2=O_P(C_{NT}^{-2}). \qquad \text{(E.11)}
\end{aligned}$$

Second term of b_{1t}:

$$\begin{aligned}
&\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\widetilde{\lambda}_i\widetilde{\alpha}_i'\widehat{e}_{it}-H_1'\lambda_i\alpha_iH_2'e_{it}\|_F^2\\
&\leq \frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\widetilde{\lambda}_i\widetilde{\alpha}_i'(\widehat{e}_{it}-e_{it})\|_F^2+\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i(\widetilde{\lambda}_i\widetilde{\alpha}_i'-H_1'\lambda_i\alpha_iH_2')e_{it}\|_F^2\\
&\leq O_P(C_{NT}^{-4})\max_i\frac{1}{T}\sum_t(\widehat{e}_{it}-e_{it})^2+O_P(C_{NT}^{-2})\frac{1}{NT}\sum_{it}(\widehat{e}_{it}-e_{it})^2\\
&+\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\lambda_i\alpha_i'(\widehat{e}_{it}-e_{it})\|_F^2\\
&+O_P(1)\mathsf{tr}\frac{1}{N^2}\sum_{ij}(\widetilde{\lambda}_i\widetilde{\alpha}_i'-H_1'\lambda_i\alpha_iH_2')(\widetilde{\lambda}_j\widetilde{\alpha}_j'-H_1'\lambda_j\alpha_jH_2')\frac{1}{T}\sum_{t\in I^c}\mathsf{E}_Ie_{it}e_{jt}\\
&= O_P(C_{NT}^{-2}). \qquad \text{(E.12)}
\end{aligned}$$

The third term of b_{1t} is bounded similarly. Finally, it is straightforward to see that fourth term of b_{1t} is $\|\frac{1}{N}\sum_i\widetilde{\alpha}_i\widetilde{\alpha}_i'-H_2'\alpha_i\alpha_i'H_2\|_F^2=O_P(C_{NT}^{-2})$. Thus $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|b_{1t}\|_F^2=O_P(C_{NT}^{-2})$.

As for b_{2t}, it suffices to prove

$$\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\alpha_i\lambda_i'e_{it}\|_F^2=O_P(1)\frac{1}{T}\sum_{t\in I^c}\mathsf{E}\|\frac{1}{N}\sum_i\alpha_i\lambda_i'e_{it}\|_F^2=O_P(N^{-1}).$$

Thus $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|b_{2t}\|_F^2=O_P(C_{NT}^{-2})$. Q.E.D.

(iii) Note that $\frac{1}{T}\sum_{t\in I^c}\|\widehat{S}_t-S\|^2\leq\frac{2}{T}\sum_{t\in I^c}\|c_t\|^2+\frac{2}{T}\sum_{t\in I^c}\|d_t\|^2$, where

$$c_t = \frac{1}{N}\sum_i\begin{pmatrix}\widetilde{\lambda}_i\lambda_i'H_1e_{it}(\widehat{e}_{it}-e_{it})+\widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2-e_{it}^2) & \widetilde{\lambda}_i(\widehat{e}_{it}-e_{it})(\alpha_i'H_2-\widetilde{\alpha}_i')\\ \widetilde{\alpha}_i\lambda_i'H_1(e_{it}-\widehat{e}_{it})+\widetilde{\alpha}_i(\lambda_i'H_1-\widetilde{\lambda}_i')(\widehat{e}_{it}-e_{it}) & 0\end{pmatrix}$$

$$d_t = \frac{1}{N}\sum_i \begin{pmatrix} (\widetilde{\lambda}_i\lambda_i'H_1 - \widetilde{\lambda}_i\widetilde{\lambda}_i')e_{it}^2 - H_1'\lambda_i(\lambda_i'H_1 - \widetilde{\lambda}_i')\mathsf{E}e_{it}^2 & \widetilde{\lambda}_i e_{it}(\alpha_i'H_2 - \widetilde{\alpha}_i') \\ \widetilde{\alpha}_i(\lambda_i'H_1 - \widetilde{\lambda}_i')e_{it} & 0 \end{pmatrix}.$$

As for $\frac{1}{T}\sum_{t\in I^c}\|c_t\|_F^2$, we have

$$\begin{aligned}
&\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \widetilde{\lambda}_i\lambda_i'H_1 e_{it}(\widehat{e}_{it}-e_{it})\|_F^2 \\
\leq\;& O_P(1)\frac{1}{N}\sum_i\|\widetilde{\lambda}_i - H_1'\lambda_i\|^2\frac{1}{NT}\sum_i\sum_{t\in I^c}e_{it}^2(\widehat{e}_{it}-e_{it})^2 \\
&+O_P(1)\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\lambda_i\lambda_i e_{it}(\widehat{e}_{it}-e_{it})\|_F^2 \\
\leq\;& O_P(C_{NT}^{-2})\frac{1}{NT}\sum_i\sum_{t\in I^c}e_{it}^2\bar{e}_i^2 + O_P(1)\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\lambda_i\lambda_i e_{it}\bar{e}_i\|_F^2 \\
=\;& O_P(C_{NT}^{-4}) \\
&\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2 - e_{it}^2)\|_F^2 \\
\leq\;& (\frac{1}{N}\sum_i\|\widetilde{\lambda}_i - H_1'\lambda_i\|^2)^2\max_{it}[(\widehat{e}_{it}-e_{it})^4] \\
&+\frac{1}{T}\sum_{t\in I^c}(\frac{1}{N}\sum_i\|\widetilde{\lambda}_i - H_1'\lambda_i\|^2|e_{it}|)^2\max_{it}|\widehat{e}_{it}-e_{it}|^2 \\
&+\frac{1}{N}\sum_i\|\widetilde{\lambda}_i - H_1'\lambda_i\|^2\frac{1}{NT}\sum_{t\in I^c}\sum_i[(\widehat{e}_{it}-e_{it})^4 + (\widehat{e}_{it}-e_{it})^2e_{it}^2] \\
&+\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\lambda_i\lambda_i'(\widehat{e}_{it}-e_{it})e_{it}\|_F^2 + \frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\lambda_i\lambda_i'(\widehat{e}_{it}-e_{it})^2\|_F^2.
\end{aligned}$$

Note that

$$\begin{aligned}
&\frac{1}{T}\sum_{t\in I^c}(\frac{1}{N}\sum_i\|\widetilde{\lambda}_i - H_1'\lambda_i\|^2|e_{it}|)^2 \\
=\;& O_P(1)\frac{1}{N^2}\sum_{ij}\|\widetilde{\lambda}_i - H_1'\lambda_i\|^2\|\widetilde{\lambda}_j - H_1'\lambda_j\|^2\frac{1}{T}\sum_{t\in I^c}\mathsf{E}_I|e_{jt}e_{it}| \\
=\;& O_P(1)(\frac{1}{N}\sum_i\|\widetilde{\lambda}_i - H_1'\lambda_i\|^2)^2 = O_P(C_{NT}^{-4}).
\end{aligned}$$

So when $\widehat{e}_{it} - e_{it} = -\bar{e}_i$, with the assumption that $\max_{it}|\bar{e}_i| = O_P(1)$, we have

$$\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2 - e_{it}^2)\|_F^2 = O_P(C_{NT}^{-4}).$$

Also,

$$
\begin{aligned}
&\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \widetilde{\lambda}_i(\widehat{e}_{it}-e_{it})(\alpha_i'H_2-\widetilde{\alpha}_i')\|_F^2\\
\leq\ &\frac{1}{N}\sum_i\|\alpha_i'H_2-\widetilde{\alpha}_i'\|^2\frac{1}{N}\sum_i\|\widetilde{\lambda}_i-H_1'\lambda_i\|^2\max_{it}(\widehat{e}_{it}-e_{it})^2\\
&+O_P(1)\frac{1}{N}\sum_i\|\alpha_i'H_2-\widetilde{\alpha}_i'\|^2\frac{1}{N}\sum_i\frac{1}{T}\sum_{t\in I^c}(\widehat{e}_{it}-e_{it})^2=O_P(C_{NT}^{-4}).
\end{aligned}
$$

Similarly, $\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\widetilde{\alpha}_i(\lambda_i'H_1-\widetilde{\lambda}_i')(\widehat{e}_{it}-e_{it})\|_F^2=O_P(C_{NT}^{-4})$. Finally,

$$
\begin{aligned}
\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\widetilde{\alpha}_i\lambda_i'H_1(e_{it}-\widehat{e}_{it})\|_F^2 &= O_P(C_{NT}^{-4})+O_P(1)\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\alpha_i\lambda_i'(e_{it}-\widehat{e}_{it})\|_F^2\\
&= O_P(C_{NT}^{-4}).
\end{aligned}
$$

So $\frac{1}{T}\sum_{t\in I^c}\|c_t\|_F^2=O_P(C_{NT}^{-4})$.

As for $\frac{1}{T}\sum_{t\in I^c}\|d_t\|_F^2$, its first term depends on $\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\lambda_i(\lambda_i'H_1-\widetilde{\lambda}_i')(e_{it}^2-\mathsf{E}e_{it}^2)\|_F^2$. Let Υ_t be an $N\times 1$ vector of e_{it}^2, and $\mathsf{diag}(\Lambda)$ be diagonal matrix consisting of elements of Λ. Suppose $\dim(\lambda_i)=1$ (focus on each element), then the first term of $\frac{1}{T}\sum_{t\in I^c}\|d_t\|_F^2$ is bounded by

$$
\begin{aligned}
&\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i(\widetilde{\lambda}_i\lambda_i'H_1-\widetilde{\lambda}_i\widetilde{\lambda}_i')e_{it}^2-H_1'\lambda_i(\lambda_i'H_1-\widetilde{\lambda}_i')\mathsf{E}e_{it}^2\|_F^2\\
\leq\ &\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i(\widetilde{\lambda}_i-H_1'\lambda_i)(\lambda_i'H_1-\widetilde{\lambda}_i')e_{it}^2\|_F^2\\
&+\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i H_1'\lambda_i(\lambda_i'H_1-\widetilde{\lambda}_i')(e_{it}^2-\mathsf{E}e_{it}^2)\|_F^2\\
\leq\ &O_P(1)\frac{1}{N^2}\sum_{ij}\|\widetilde{\lambda}_i-H_1'\lambda_i\|^2\|\widetilde{\lambda}_j-H_1'\lambda_j\|^2\frac{1}{T}\sum_{t\in I^c}\mathsf{E}_I e_{jt}^2e_{it}^2\\
&+O_P(1)\frac{1}{T}\sum_{t\in I^c}\frac{1}{N^2}(\widetilde{\Lambda}-\Lambda H_1)'\mathsf{diag}(\Lambda)\mathsf{Var}_I(\Upsilon_t)\mathsf{diag}(\Lambda)(\widetilde{\Lambda}-\Lambda H_1)\\
=\ &O_P(C_{NT}^{-4}).
\end{aligned}
$$

In addition, using the same technique, it is easy to show

$$
\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\widetilde{\lambda}_ie_{it}(\alpha_i'H_2-\widetilde{\alpha}_i')\|_F^2=O_P(C_{NT}^{-4})=\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\widetilde{\alpha}_i(\lambda_i'H_1-\widetilde{\lambda}_i')e_{it}\|_F^2.
$$

Put together, $\frac{1}{T}\sum_{t\in I^c}\|\widehat{S}_t-S\|^2\leq\frac{2}{T}\sum_{t\in I^c}\|c_t\|^2+\frac{2}{T}\sum_{t\in I^c}\|d_t\|^2=O_P(C_{NT}^{-4})$. Q.E.D.

Lemma E.7. *For terms defined in (D.2), and for each fixed* $t \in I^c$, $\sum_{d=1}^5 A_{dt} = O_P(C_{NT}^{-2})$ *and for the "upper block" of* A_{6t}, $\frac{1}{N}\sum_i \lambda_i e_{it}(\mu_i \lambda_i' f_t - \bar{x}_i \dot{\lambda}_i' \widetilde{f}_t) = O_P(C_{NT}^{-2})$.

Proof. (i) Term A_{1t}. It suffices to show $\widehat{B}_t^{-1}\widehat{S}_t - B^{-1}S = O_P(C_{NT}^{-2})$. By Lemma E.5,

$$\begin{aligned} B_t^{-1}\widehat{S}_t - B^{-1}S &= (B_t^{-1} - B^{-1})(\widehat{S}_t - S) + (B_t^{-1} - B^{-1})S + B^{-1}(\widehat{S}_t - S) \\ &= O_P(C_{NT}^{-2}) + O_P(C_{NT}^{-1})S = O_P(C_{NT}^{-2}), \end{aligned}$$

where the last equality is due to $S = O_P(C_{NT}^{-1})$.

(ii) Term A_{2t}. Given $B_t^{-1} - B^{-1} = O_P(C_{NT}^{-1})$ and the cross-sectional weak correlations in u_{it}, it is easy to see $A_{2t} = O_P(C_{NT}^{-2})$.

(iii) Term A_{3t}. It suffices to prove:

$$\begin{aligned} &\frac{1}{N}\sum_i (\widetilde{\lambda}_i \widehat{e}_{it} - H_1'\lambda_i e_{it})u_{it} = O_P(C_{NT}^{-2}) \\ &\frac{1}{N}\sum_i (\widetilde{\alpha}_i - H_2'\alpha_i)u_{it} = O_P(C_{NT}^{-2}). \end{aligned} \tag{E.13}$$

First, let Υ_t be an $N \times 1$ vector of $e_{it}u_{it}$. Due to the serial independence of (u_{it}, e_{it}), we have

$$\begin{aligned} &\mathsf{E}_I \frac{1}{N}\sum_i (\widetilde{\lambda}_i - H_1'\lambda_i)e_{it}u_{it} = 0, \\ &\mathsf{Var}_I(\frac{1}{N}\sum_i (\widetilde{\lambda}_i - H_1'\lambda_i)e_{it}u_{it}) = \frac{1}{N^2}(\widetilde{\Lambda} - \Lambda H_1)'\mathsf{Var}_I(\Upsilon_t)(\widetilde{\Lambda} - \Lambda H_1) \\ \leq\ & O_P(C_{NT}^{-2}N^{-1}) \max_i \sum_j |\mathsf{Cov}(e_{it}u_{it}, e_{jt}u_{jt})| = O_P(C_{NT}^{-2}N^{-1}). \end{aligned}$$

Similarly, $\mathsf{E}_I \frac{1}{N}\sum_i(\widetilde{\alpha}_i - H_2'\alpha_i)u_{it} = 0$ and $\mathsf{Var}_I(\frac{1}{N}\sum_i(\widetilde{\alpha}_i - H_2'\alpha_i)u_{it}) = O_P(C_{NT}^{-2}N^{-1})$.

$$\begin{aligned} &\frac{1}{N}\sum_i (\widetilde{\lambda}_i - H_1'\lambda_i)e_{it}u_{it} = O_P(C_{NT}^{-1}N^{-1/2}) \\ &\frac{1}{N}\sum_i (\widetilde{\alpha}_i - H_2'\alpha_i)u_{it} = O_P(C_{NT}^{-1}N^{-1/2}) \\ &\frac{1}{N}\sum_i \|\widetilde{\lambda}_i - H_1'\lambda_i\|^2 u_{it}^2 = O_P(1)\frac{1}{N}\sum_i \|\widetilde{\lambda}_i - H_1'\lambda_i\|^2 \mathsf{E}_I u_{it}^2 = O_P(C_{NT}^{-2}) \end{aligned} \tag{E.14}$$

Thus, the first term of (E.13) is

$$\frac{1}{N}\sum_i (\widetilde{\lambda}_i \widehat{e}_{it} - H_1'\lambda_i e_{it})u_{it} \leq (\frac{1}{N}\sum_i \|\widetilde{\lambda}_i - H_1'\lambda_i\|^2 u_{it}^2)^{1/2}(\frac{1}{N}\sum_i (\widehat{e}_{it} - e_{it})^2)^{1/2}$$

$$
\begin{aligned}
&+H_1'\frac{1}{N}\sum_i \lambda_i(\widehat{e}_{it}-e_{it})u_{it}+\frac{1}{N}\sum_i(\widetilde{\lambda}_i-H_1'\lambda_i)e_{it}u_{it}\\
\leq\ & O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2)^{1/2}+O_P(C_{NT}^{-2})+O_P(1)\frac{1}{N}\sum_i\lambda_i(\widehat{e}_{it}-e_{it})u_{it}.
\end{aligned}\tag{E.15}
$$

When $\widehat{e}_{it}-e_{it}=-\bar{e}_i$ in the many mean model, the above is $O_P(C_{NT}^{-2})$.

(iv) Term A_{5t}. Given $B_t^{-1}-B^{-1}=O_P(C_{NT}^{-1})$, it suffices to prove the following terms are $O_P(C_{NT}^{-2})$.

$$
\begin{aligned}
B_{1t} &= \frac{1}{N}\sum_i \lambda_i e_{it}(\bar{x}_i-\mu_i)(\dot{\lambda}_i-H_1'\lambda_i)'\widetilde{f}_t\\
B_{2t} &= \frac{1}{N}\sum_i \lambda_i e_{it}(\bar{x}_i-\mu_i)\lambda_i'H_1\widetilde{f}_t\\
B_{3t} &= \frac{1}{N}\sum_i \lambda_i e_{it}\mu_i(\dot{\lambda}_i-H_1'\lambda_i)'\widetilde{f}_t\\
B_{4t} &= \frac{1}{N}\sum_i \lambda_i e_{it}\mu_i\lambda_i'H_1(\widetilde{f}_t-H_1^{-1}f_t)\\
B_{5t} &= \frac{1}{N}\sum_i \alpha_i(\bar{x}_i-\mu_i)(\dot{\lambda}_i-H_1'\lambda_i)'\widetilde{f}_t
\end{aligned}\tag{E.16}
$$

and that the following terms are $O_P(C_{NT}^{-1})$:

$$
\begin{aligned}
B_{6t} &= \frac{1}{N}\sum_i \alpha_i(\bar{x}_i-\mu_i)\lambda_i'H_1\widetilde{f}_t\\
B_{7t} &= \frac{1}{N}\sum_i \alpha_i\mu_i(\dot{\lambda}_i-H_1'\lambda_i)'\widetilde{f}_t\\
B_{8t} &= \frac{1}{N}\sum_i \alpha_i\mu_i\lambda_i'H_1(\widetilde{f}_t-H_1^{-1}f_t).
\end{aligned}\tag{E.17}
$$

In fact, we have,

$$
\begin{aligned}
B_{1t} &\leq O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i e_{it}^2\bar{e}_i^2)^{1/2}=O_P(C_{NT}^{-2}).\\
B_{2t} &= O_P(1)\frac{1}{NT}\sum_{is}\lambda_i\lambda_i'e_{it}e_{is}\\
&\leq O_P(1)(\frac{1}{N^2T^2}\sum_{ij}\sum_{sk}|\mathsf{E}e_{it}e_{jt}e_{is}e_{jk}|)^{1/2}=O_P(C_{NT}^{-2}).
\end{aligned}
$$

By Lemma E.4, note that $t \in I^c$, $B_{3t} = O_P(1)\frac{1}{N}\sum_i \lambda_i e_{it}\mu_i(\dot{\lambda}_i - H_1'\lambda_i)' = O_P(C_{NT}^{-2})$. Finally, still by Lemma E.4 that for each fixed t, $\widetilde{f}_t - H_1^{-1} f_t = O_P(C_{NT}^{-1})$. Thus

$$B_{4t} = O_P(C_{NT}^{-1})\frac{1}{N}\sum_i \lambda_i e_{it}\mu_i\lambda_i' = O_P(C_{NT}^{-2}).$$

The proofs for $B_{5t} \sim B_{8t}$ from the similar argument, using the Cauchy-Schwarz inequality. In addition, it is also straightforward to prove, for each fixed j,

$$\sum_{d=1}^{5}\frac{1}{T}\sum_{t\in I^c}\|B_{dt}\|^2(1+e_{jt}^4) = O_P(C_{NT}^{-4}), \quad \sum_{d=6}^{8}\frac{1}{T}\sum_{t\in I^c}\|B_{dt}\|^2(1+e_{jt}^4) = O_P(C_{NT}^{-2}). \tag{E.18}$$

(v) "upper block" of A_{6t}. From the proof of (iv), we have $B_{dt} = O_P(C_{NT}^{-2})$, for $d = 1, ..., 4$. It follows immediately that $\frac{1}{N}\sum_i \lambda_i e_{it}(\mu_i\lambda_i' f_t - \bar{x}_i\dot{\lambda}_i'\widetilde{f}_t) = O_P(C_{NT}^{-2})$.

(vi) Term A_{4t}. Note that $\bar{x}_i - \mu_i = e_{it} - \widehat{e}_{it}$. It suffices to prove each of the following terms is $O_P(C_{NT}^{-2})$: note that $\mu_{it} = \mu_i$.

$$\begin{aligned}
C_{1t} &= \frac{1}{N}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)e_{it}(\widehat{e}_{it} - e_{it})(\dot{\lambda}_i - H_1'\lambda_i)'\widetilde{f}_t \\
C_{2t} &= \frac{1}{N}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)(\widehat{e}_{it} - e_{it})^2(\dot{\lambda}_i - H_1'\lambda_i)'\widetilde{f}_t \\
C_{3t} &= \frac{1}{N}\sum_i H_1'\lambda_i(\widehat{e}_{it} - e_{it})^2(\dot{\lambda}_i - H_1'\lambda_i)'\widetilde{f}_t \\
C_{4t} &= \frac{1}{N}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)e_{it}(\widehat{e}_{it} - e_{it})\lambda_i' H_1\widetilde{f}_t \\
C_{5t} &= \frac{1}{N}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)(\widehat{e}_{it} - e_{it})^2\lambda_i' H_1\widetilde{f}_t \\
C_{6t} &= \frac{1}{N}\sum_i H_1'\lambda_i(\widehat{e}_{it} - e_{it})^2\lambda_i' H_1\widetilde{f}_t \\
C_{7t} &= \frac{1}{N}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)e_{it}\mu_{it}(\dot{\lambda}_i - H_1'\lambda_i)'\widetilde{f}_t \\
C_{8t} &= \frac{1}{N}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)(\widehat{e}_{it} - e_{it})\mu_{it}(\dot{\lambda}_i - H_1'\lambda_i)'\widetilde{f}_t \\
C_{9t} &= \frac{1}{N}\sum_i H_1'\lambda_i(\widehat{e}_{it} - e_{it})\mu_{it}(\dot{\lambda}_i - H_1'\lambda_i)'\widetilde{f}_t \\
C_{10t} &= \frac{1}{N}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)e_{it}\mu_{it}\lambda_i' H_1(\widetilde{f}_t - H_1^{-1}f_t) \\
C_{11t} &= \frac{1}{N}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)(\widehat{e}_{it} - e_{it})\mu_{it}\lambda_i' H_1(\widetilde{f}_t - H_1^{-1}f_t)
\end{aligned}$$

$$
\begin{aligned}
C_{12t} &= \frac{1}{N}\sum_i H_1'\lambda_i(\widehat{e}_{it}-e_{it})\mu_{it}\lambda_i'H_1(\widetilde{f}_t-H_1^{-1}f_t)\\
C_{13t} &= \frac{1}{N}\sum_i(\widetilde{\alpha}_i-H_2'\alpha_i)(\widehat{e}_{it}-e_{it})(\dot{\lambda}_i-H_1'\lambda_i)'\widetilde{f}_t\\
C_{14t} &= \frac{1}{N}\sum_i(\widetilde{\alpha}_i-H_2'\alpha_i)(\widehat{e}_{it}-e_{it})\lambda_i'H_1\widetilde{f}_t\\
C_{15t} &= \frac{1}{N}\sum_i(\widetilde{\alpha}_i-H_2'\alpha_i)\mu_{it}(\dot{\lambda}_i-H_1'\lambda_i)'\widetilde{f}_t\\
C_{16t} &= \frac{1}{N}\sum_i(\widetilde{\alpha}_i-H_2'\alpha_i)\mu_{it}\lambda_i'H_1(\widetilde{f}_t-H_1^{-1}f_t). \qquad \text{(E.19)}
\end{aligned}
$$

The proof follows from repeatedly applying the Cauchy-Schwarz inequality and is straightforward. In addition, it is also straightforward to apply Cauchy-Schwarz to prove that

$$
\frac{1}{T}\sum_{t\in I^c}\|C_{dt}\|^2=O_P(C_{NT}^{-4})=\frac{1}{T}\sum_{t\in I^c}\|C_{dt}\|^2e_{it}^2,\quad d=1,...,16.\text{ and fixed } i. \qquad \text{(E.20)}
$$

We omit the details for brevity. Q.E.D.

Lemma E.8. *(i)* $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|\widehat{f}_s-H_ff_s\|^2=O_P(C_{NT}^{-2})$.
(ii) For each fixed i, $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}(\widehat{f}_s-H_ff_s)e_{is}f_s'=O_P(C_{NT}^{-2})$.

This lemma is needed to prove the performance for $\widehat{\lambda}_i$, which controls the effect of $\widehat{f}_s-H_ff_s$ on the estimation of λ_i.

Proof. (i) Given (D.2), the proof is very similar to that of Lemma E.7. We omit the details for brevity. We now turn to the much harder part (ii).

(ii) Use (D.2), $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}(\widehat{f}_s-H_ff_s)e_{is}f_s'$ equals, up to a multiplier of order $O_P(1)$,

$$
\frac{1}{NT}\sum_{t\in I^c}\sum_i e_{it}^2u_{it}f_t'+\sum_{d=1}^5\frac{1}{T}\sum_{t\in I^c}a_{dt}e_{it}f_t'.
$$

where a_{dt} is the upper block of A_{dt}. Given that $\mathsf{E}(u_tu_t'|e_y,f_t)<C$ and that u_t is conditionally serially independent, we have $\frac{1}{NT}\sum_{t\in I^c}\sum_i e_{it}^2u_{it}f_t'=O_P(C_{NT}^{-2})$. Next, up to a multiplier of order $O_P(1)$, by Lemma E.6,

$$
\frac{1}{T}\sum_{t\in I^c}a_{1t}e_{it}f_t' \;\leq\; \frac{1}{T}\sum_{t\in I^c}\|\widehat{B}_t^{-1}\widehat{S}_t-B^{-1}S\|(\|f_t\|+\|g_t\|)\|f_t'\||e_{it}|
$$

$$
\begin{aligned}
&\leq (\frac{1}{T}\sum_{t\in I^c}\|\widehat{B}_t - B\|^2)^{1/2}\|S\| + (\frac{1}{T}\sum_{t\in I^c}\|\widehat{S}_t - S\|^2)^{1/2} \\
&= O_P(C_{NT}^{-2}) \\
\frac{1}{T}\sum_{t\in I^c} a_{2t}e_{it}f_t' &\leq \frac{1}{T}\sum_{t\in I^c}\|e_{it}f_t'\|\|\widehat{B}_t^{-1} - B^{-1}\|[\|\frac{1}{N}\sum_j \lambda_j e_{jt}u_{jt}\| + \|\frac{1}{N}\sum_j \alpha_j u_{jt}\|] \\
&\leq (\frac{1}{T}\sum_{t\in I^c}\|\widehat{B}_t - B\|^2)^{1/2}(\frac{1}{T}\sum_{t\in I^c}\|e_{it}f_t'\|^2[\|\frac{1}{N}\sum_j \lambda_j e_{jt}u_{jt}\| + \|\frac{1}{N}\sum_j \alpha_j u_{jt}\|]^2)^{1/2} \\
&\leq O_P(C_{NT}^{-1})(\frac{1}{T}\sum_{t\in I^c}\mathsf{E}\|e_{it}f_t'\|^2\mathsf{E}[\|\frac{1}{N}\sum_j \lambda_j e_{jt}u_{jt}\|^2 + \|\frac{1}{N}\sum_j \alpha_j u_{jt}\|^2|e_t, f_t])^{1/2} \\
&= O_P(C_{NT}^{-2}) \\
\frac{1}{T}\sum_{t\in I^c} a_{3t}e_{it}f_t' &\leq \max_t \|\widehat{B}_t^{-1}\|\frac{1}{T}\sum_{t\in I^c}\|e_{it}f_t'\|\|\frac{1}{N}\sum_j(\widetilde{\lambda}_j\widehat{e}_{jt} - H_1'\lambda_j e_{jt})u_{jt}\| \\
&\quad + \max_t \|\widehat{B}_t^{-1}\|\frac{1}{T}\sum_{t\in I^c}\|e_{it}f_t'\|\|\frac{1}{N}\sum_j(\widetilde{\alpha}_j - H_2'\alpha_j)u_{jt}\| \\
&\leq O_P(1)(\frac{1}{T}\sum_{t\in I^c}\mathsf{E}_I\|\frac{1}{N}\sum_j(\widetilde{\lambda}_j - H_1'\lambda_j)e_{jt}u_{jt}\|^2)^{1/2} \\
&\quad + O_P(1)(\frac{1}{T}\sum_{t\in I^c}\mathsf{E}_I\|\frac{1}{N}\sum_j(\widetilde{\alpha}_j - H_2'\alpha_j)u_{jt}\|^2)^{1/2} \\
&\quad + O_P(1)\frac{1}{T}\sum_{t\in I^c}\|e_{it}f_t'\|\|\frac{1}{N}\sum_j(\widetilde{\lambda}_j - H_1'\lambda_j)(\widehat{e}_{jt} - e_{jt})u_{jt}\| \\
&\quad + O_P(1)\frac{1}{T}\sum_{t\in I^c}\|e_{it}f_t'\|\|\frac{1}{N}\sum_j \lambda_j(\widehat{e}_{jt} - e_{jt})u_{jt}\| \\
&= O_P(C_{NT}^{-2}) \\
\frac{1}{T}\sum_{t\in I^c} a_{4t}e_{it}f_t' &\leq \max_t \|\widehat{B}_t^{-1}\|\frac{1}{T}\sum_{t\in I^c}\|e_{it}f_t'\|\|C_{dt}\|, \quad d = 1...16
\end{aligned}
$$

where C_{dt}'s are defined in the proof of Lemma E.7. By (E.20), $\frac{1}{T}\sum_{t\in I^c}\|C_{dt}\|^2 = O_P(C_{NT}^{-4})$ for $d \leq 16$. Thus $\frac{1}{T}\sum_{t\in I^c} a_{4t}e_{it}f_t' = O_P(C_{NT}^{-2})$, still following from Cauchy-Schwarz.

Next, for B_{dt} defined in the proof of Lemma E.7,

$$
\frac{1}{T}\sum_{t\in I^c} a_{5t}e_{it}f_t' \leq \sum_{d=1}^{8}\frac{1}{T}\sum_{t\in I^c}\|\widehat{B}_t - B\|\|B_{dt}e_{it}f_t'\|.
$$

Repeatedly using Cauchy-Schwarz, it can be shown that $\frac{1}{T}\sum_{t\in I^c} a_{5t}e_{it}f_t' = O_P(C_{NT}^{-2})$. Finally, using Cauchy-Schwarz,

$$\frac{1}{T}\sum_{t\in I^c} a_{6t}e_{it}f_t' \quad \leq \quad O_P(1)\sum_{d=1}^{4}\|\frac{1}{T}\sum_{t\in I^c} B_{dt}e_{it}f_t'\| = O_P(C_{NT}^{-2}).$$

Therefore, $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}(\widehat{f}_s - H_f f_s)e_{is}f_s' = O_P(C_{NT}^{-2})$.

Q.E.D.

E.3. **Technical lemmas for $\widehat{\lambda}_i$.**

Lemma E.9. *For each fixed i, $\widehat{D}_i - D_i = O_P(C_{NT}^{-1})$.*

Proof. $\widehat{D}_i - D_i$ is a two-by-two block matrix. The first block is

$$\begin{aligned}
&\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} \widehat{f}_s\widehat{f}_s'\widehat{e}_{is}^2 - H_f f_s f_s' H_f' \mathsf{E}e_{is}^2 \\
= \;& \frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} (\widehat{f}_s - H_f f_s)(\widehat{f}_s - H_f f_s)'[(\widehat{e}_{is} - e_{is})^2 + (\widehat{e}_{is} - e_{is})e_{is} + e_{is}^2] \\
&+\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} (\widehat{f}_s - H_f f_s)f_s'H_f'[(\widehat{e}_{is} - e_{is})^2 + (\widehat{e}_{is} - e_{is})e_{is} + e_{is}^2] \\
&+\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} H_f f_s(\widehat{f}_s - H_f f_s)'[(\widehat{e}_{is} - e_{is})^2 + (\widehat{e}_{is} - e_{is})e_{is} + e_{is}^2] \\
&+\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} H_f f_s f_s' H_f'[(\widehat{e}_{is} - e_{is})^2 + (\widehat{e}_{is} - e_{is})e_{is}] + \frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} H_f f_s f_s' H_f'(e_{is}^2 - \mathsf{E}e_{is}^2).
\end{aligned}$$

The second block is

$$\begin{aligned}
\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} \widehat{f}_s\widehat{g}_s'\widehat{e}_{is} \;=\; & \frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} (\widehat{f}_s - H_f f_s)(\widehat{g}_s - H_g g_s)'\widehat{e}_{is} + \frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} (\widehat{f}_s - H_f f_s)g_s'H_g'\widehat{e}_{is} \\
&+\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} H_f f_s(\widehat{g}_s - H_g g_s)'\widehat{e}_{is} + \frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} H_f f_s g_s' H_g'(\widehat{e}_{is} - e_{is}) \\
&+\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}} H_f f_s g_s' H_g' e_{is}
\end{aligned}$$

The third block is similar. The fourth block is $\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}}\widehat{g}_s\widehat{g}_s' - H_g g_s g_s' H_g'$. Given $\widehat{e}_{is} - e_{is} = -\bar{e}_i$ and Lemma E.9, it is straightforward to see each term in the above expansions is $O_P(C_{NT}^{-1})$. Q.E.D.

Lemma E.10. *For each fixed $i \leq N$,*

$$\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\widehat{f}_s\widehat{e}_{is}(\mu_i\lambda_i'f_s-\bar{x}_i\dot{\lambda}_i'\widetilde{f}_s)=O_P(C_{NT}^{-2}),\quad \frac{1}{T}\sum_{s\in I^c\cup\{t\}}\widehat{g}_s(\mu_i\lambda_i'f_s-\bar{x}_i\dot{\lambda}_i'\widetilde{f}_s)=O_P(C_{NT}^{-1}).$$

Note that the first term is $O_P(C_{NT}^{-2})$ while the second term is $O_P(C_{NT}^{-1})$.

Proof. It suffices to prove that the following statements.

$$\begin{aligned}
R_{3i,1} &:= \frac{1}{T}\sum_{s\in I^c\cup\{t\}}(\widehat{f}_s-H_ff_s)(\widehat{e}_{is}-e_{is})\mu_i\lambda_i'f_s=O_P(C_{NT}^{-2})\\
R_{3i,2} &:= \frac{1}{T}\sum_{s\in I^c\cup\{t\}}(\widehat{f}_s\widehat{e}_{is}-H_ff_se_{is})(\mu_{it}\lambda_i'f_s-\widehat{\mu}_{it}\dot{\lambda}_i'\widetilde{f}_s)=O_P(C_{NT}^{-2})\\
R_{3i,3} &:= \frac{1}{T}\sum_{s\in I^c\cup\{t\}}f_se_{is}(\widehat{e}_{is}-e_{is})\dot{\lambda}_i'\widetilde{f}_s=O_P(C_{NT}^{-2})\\
R_{3i,4} &:= \frac{1}{T}\sum_{s\in I^c\cup\{t\}}f_se_{is}\mu_i(\dot{\lambda}_i-H_1'\lambda_i)'\widetilde{f}_s=O_P(C_{NT}^{-2})\\
R_{3i,5} &:= \frac{1}{T}\sum_{s\in I^c\cup\{t\}}f_se_{is}\mu_i\lambda_i'(\widetilde{f}_s-H_1^{-1}f_s)=O_P(C_{NT}^{-2})\\
R_{3i,6} &:= \frac{1}{T}\sum_{s\in I^c\cup\{t\}}(\widehat{f}_s-H_ff_s)e_{is}\mu_i\lambda_i'f_s=O_P(C_{NT}^{-2})\\
R_{3i,7} &:= \frac{1}{T}\sum_{s\in I^c\cup\{t\}}\widehat{g}_s(\mu_i\lambda_i'f_s-\bar{x}_i\dot{\lambda}_i'\widetilde{f}_s)=O_P(C_{NT}^{-1})
\end{aligned}\tag{E.21}$$

where $\widehat{e}_{is}-e_{is}=\bar{e}_i$ and $\mu_{it}=\mu_i$, $\widehat{\mu}_{it}=\bar{x}_i$ in the above.

Note that all terms but the last one in the above are $O_P(C_{NT}^{-2})$. For the preliminary estimators, we have $\dot{\lambda}_i-H_1'\lambda_i=O_P(C_{NT}^{-1})$ and $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|\widetilde{f}_s-H_1^{-1}f_s\|^2=O_P(C_{NT}^{-2})$, by Lemmas E.3 and E.4. Then using the Cauchy-Schwarz inequality, it is easy to show that $R_{3i,1}$ through $R_{3i,4}$ are $O_P(C_{NT}^{-2})$ and $R_{3i,7}=O_P(C_{NT}^{-1})$. Next, By Lemmas E.2, E.9,

$$\frac{1}{T}\sum_{s\in I^c\cup\{t\}}f_s(H_1^{-1}f_s-\widetilde{f}_s)e_{is}=O_P(C_{NT}^{-2})=\frac{1}{T}\sum_{s\in I^c\cup\{t\}}(\widehat{f}_s-H_ff_s)e_{is}f_s'.$$

So $R_{3i,5}=O_P(C_{NT}^{-2})=R_{3i,6}$. (Note that the above two equalities involve in two factor estimators: the preliminary $\widetilde{f}_s$ and the final $\widehat{f}_s$.) This concludes that $R_{3i,d}=O_P(C_{NT}^{-2})$ for all $d\leq 7$ defined above, and completes the proof. Q.E.D.

Lemma E.11. *For B_{dt} defined in the proof of Lemma E.7, and a fixed $j \leq N$, for $d = 1...4$, $\frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt}^r B_{dt} = O_P(C_{NT}^{-2})$, $r = 0, 1, 2$. for $d = 5...8$, $\frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt} B_{dt} = O_P(C_{NT}^{-2})$.*

Proof. For $r = 0, 1, 2$,

$$
\begin{aligned}
\frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt}^r B_{1t} &= \frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt}^r \frac{1}{N}\sum_i \lambda_i e_{it}\bar{e}_i(\dot{\lambda}_i - H_1'\lambda_i)'\widetilde{f}_t \\
&\leq \frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt}^r \frac{1}{N}\sum_i \lambda_i e_{it}\bar{e}_i(\dot{\lambda}_i - H_1'\lambda_i)' f_t \\
&\quad + O_P(C_{NT}^{-2})(\frac{1}{N}\sum_i \bar{e}_i^2 \frac{1}{T_0}\sum_{t\in I^c} f_t^2 e_{jt}^{2r} e_{it}^2\|^2)^{1/2} \\
&\leq O_P(C_{NT}^{-2}) \\
\frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt}^r B_{2t} &= \frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt}^r \frac{1}{N}\sum_i \lambda_i e_{it}\bar{e}_i \lambda_i' H_1 \widetilde{f}_t \\
&= (\mathsf{E}(\frac{1}{T_0}\sum_{t\in I^c}\frac{1}{N}\sum_i \lambda_i^2 e_{it}\bar{e}_i f_t^2 e_{jt}^r)^2)^{1/2} + O_P(C_{NT}^{-2}) = O_P(C_{NT}^{-2}) \\
\frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt}^r B_{3t} &= \frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt}^r \frac{1}{N}\sum_i \lambda_i e_{it}\mu_i(\dot{\lambda}_i - H_1'\lambda_i)'\widetilde{f}_t \\
&\leq \frac{1}{T_0}\sum_{t\in I^c} f_t^2 e_{jt}^r \frac{1}{N}\sum_i \lambda_i e_{it}\mu_i(\dot{\lambda}_i - H_1'\lambda_i) + O_P(C_{NT}^{-2}) \\
&\leq O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i \mathsf{E}(\frac{1}{T_0}\sum_{t\in I^c} f_t^2 e_{jt}^r e_{it})^2)^{1/2} + O_P(C_{NT}^{-2}) \\
&\leq O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i \frac{1}{T_0^2}\sum_{s\neq t} \mathsf{E} f_t^2 f_s^2 \mathsf{E}(e_{jt}^r e_{it}|F)\mathsf{E} e_{js}^r e_{is}|F))^{1/2} + O_P(C_{NT}^{-2}) \\
&\leq O_P(C_{NT}^{-2})(\sum_i |\mathsf{Cov}(e_{js}^r, e_{is}|F)|)^{1/2} + O_P(C_{NT}^{-2}) = O_P(C_{NT}^{-2}) \\
\frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt}^r B_{4t} &= \frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt}^r \frac{1}{N}\sum_i \lambda_i e_{it}\mu_i \lambda_i' H_1(\widetilde{f}_t - H_1^{-1} f_t) = O_P(C_{NT}^{-2}).
\end{aligned}
$$

Next, for $d = 5 \sim 8$,

$$
\begin{aligned}
\frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt} B_{5t} &= \frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt} \frac{1}{N}\sum_i \alpha_i \bar{e}_i(\dot{\lambda}_i - H_1'\lambda_i)'\widetilde{f}_t \\
&= O_P(C_{NT}^{-2}), \quad \text{(similar to } \frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt}^r B_{1t}) \\
\frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt} B_{6t} &= \frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt}\widetilde{f}_t(\frac{1}{N}\sum_i \alpha_i \bar{e}_i \lambda_i' H_1) = O_P(C_{NT}^{-2})
\end{aligned}
$$

$$\begin{aligned}\frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt} B_{7t} &= \frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt}(\widetilde{f}_t - H_1^{-1}f_t)\frac{1}{N}\sum_i \alpha_i\mu_i(\dot{\lambda}_i - H_1'\lambda_i)' \\ &\quad +(\frac{1}{T_0}\sum_{t\in I^c} f_t^2 e_{jt})(\frac{1}{N}\sum_i \alpha_i\mu_i(\dot{\lambda}_i - H_1'\lambda_i)) = O_P(C_{NT}^{-2}) \\ \frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt} B_{8t} &= \frac{1}{T_0}\sum_{t\in I^c} f_t e_{jt}(\widetilde{f}_t - H_1^{-1}f_t)\frac{1}{N}\sum_i \alpha_i\mu_i\lambda_i' H_1 \\ &= O_P(C_{NT}^{-2}), \quad \text{by lemma E.2.}\end{aligned}$$

Q.E.D.

Lemma E.12. *For R_{di} defined in (D.5), and for each fixed $j \leq N$, $R_{dj} = O_P(C_{NT}^{-2})$ for $d = 1, ..., 3$. The upper blocks of $R_{4j} \sim R_{6j}$ are $O_P(C_{NT}^{-2})$.*

Proof. (i) It is easy to see that $R_{1j} = O_P(C_{NT}^{-2}) = R_{2j}$. Also, it follows from Lemmas E.9, E.10 that both R_{3j} and the upper block of R_{4j} are $O_P(C_{NT}^{-2})$. Next, by (D.2),

$$\begin{aligned} R_{5j} &= \widehat{D}_j^{-1}\frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}\begin{pmatrix}\widehat{f}_t - H_f f_t\\ \widehat{g}_t - H_g g_t\end{pmatrix} u_{jt} \\ &= \widehat{D}_j^{-1}\frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix} B^{-1}\frac{1}{N}\sum_i \begin{pmatrix} H_1'\lambda_i e_{it}\\ H_2'\alpha_i\end{pmatrix} u_{it}u_{jt} \\ &\quad +\sum_{d=1}^{5}\widehat{D}_j^{-1}\frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix} A_{dt}u_{jt} \\ &\leq O_P(C_{NT}^{-1})\|\frac{1}{T_0}\sum_{t\in I^c}\widehat{e}_{jt}\frac{1}{N}\sum_i \lambda_i e_{it}u_{it}u_{jt}\| + O_P(C_{NT}^{-1})\|\frac{1}{T_0}\sum_{t\in I^c}\frac{1}{N}\sum_i \alpha_i u_{it}u_{jt}\| \\ &\quad + D_j^{-1}B^{-1}\frac{1}{T_0}\sum_{t\in I^c}\frac{1}{N}\sum_i\begin{pmatrix}\lambda_i e_{it}\widehat{e}_{jt}\\ \alpha_i\end{pmatrix} u_{it}u_{jt} + \sum_{d=1}^{5}\Upsilon_d, \end{aligned} \tag{E.22}$$

where, for A_{dt} defined in (D.2),

$$\Upsilon_d = \widehat{D}_j^{-1}\frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix} A_{dt}u_{jt}.$$

The first two terms of R_{5j} are $O_P(C_{NT}^{-2})$; the upper block of the third term is bounded by

$$O_P(1)\frac{1}{T_0}\sum_{t\in I^c}\frac{1}{N}\sum_i \lambda_i e_{it}\widehat{e}_{jt}u_{it}u_{jt} = O_P(C_{NT}^{-2})$$

granted by the assumption that $\mathsf{Var}(\frac{1}{\sqrt{N}}\sum_i w_{it}w_{jt}) < C$ and $\sum_i |\mathsf{Cov}(w_{it}, w_{jt})| < C$ for $w_{it} = u_{it}e_{it}$. To finish the proof for R_{5j}, we show $\Upsilon_d = O_P(C_{NT}^{-2})$ for $d = 1...5$.

Term Υ_1 is bounded by (Lemma E.6),

$$O_P(1)(\frac{1}{T_0}\sum_{t\in I^c}\|\widehat{B}_t^{-1}\widehat{S}_t - B^{-1}S\|^2)^{1/2} = O_P(C_{NT}^{-2}).$$

Term Υ_2 is bounded by

$$O_P(C_{NT}^{-2}) + O_P(C_{NT}^{-1})(\frac{1}{T_0}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \lambda_i w_{it}\|^4 + \|\frac{1}{N}\sum_i \alpha_i u_{it}\|^4)^{1/2} = O_P(C_{NT}^{-2})$$

with the assumption that $\max_t \mathsf{E}\|\frac{1}{\sqrt{N}}\sum_i \lambda_i w_{it}\|^4 < C$, and $\max_t \mathsf{E}\|\frac{1}{\sqrt{N}}\sum_i \alpha_i u_{it}\|^4 < C$.

Term Υ_3 is bounded by $O_P(C_{NT}^{-2})$ plus

$$O_P(1)(\frac{1}{T_0}\sum_{t\in I^c}\|\frac{1}{N}\sum_i(\widetilde{\lambda}_i\widehat{e}_{it} - H_1'\lambda_i e_{it})u_{it}\|^2 + \|\frac{1}{N}\sum_i(\widetilde{\alpha}_i - H_2'\alpha_i)u_{it}\|^2)^{1/2} = O_P(C_{NT}^{-2}).$$

Term Υ_4 is bounded by $O_P(C_{NT}^{-2})$ plus

$$\begin{aligned} & O_P(1)(\frac{1}{T_0}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \begin{pmatrix}\widetilde{\lambda}_i\widehat{e}_{it} - H_1'\lambda_i e_{it}\\ \widetilde{\alpha}_i - H_2'\alpha_i\end{pmatrix}(\mu_i\lambda_i' f_t - \bar{x}_i\dot{\lambda}_i'\widetilde{f}_t)\|^2)^{1/2} \\ \leq\ & O_P(1)(\sum_{d=1}^{16}\frac{1}{T_0}\sum_{t\in I^c}\|C_{dt}\|^2)^{1/2} = O_P(C_{NT}^{-2}) \end{aligned}$$

where C_{dt} are defined in the proof of Lemma E.7.

Term Υ_5 is bounded by, (B_{dt} is defined in the proof of Lemma E.7)

$$O_P(C_{NT}^{-2}) + O_P(1)\sum_{d=1}^{8}\frac{1}{T_0}\sum_{t\in I^c}\|\widehat{B}_t - B_t\|\|B_{dt}\||u_{jt}|(|e_{jt}| + 1),$$

which is $O_P(C_{NT}^{-2})$, using Lemma E.3 that $\frac{1}{T}\sum_{t\in I^c}\|\widetilde{f}_t - H_1^{-1}f_t\|^2 e_{it}^2 u_{it}^2 = O_P(C_{NT}^{-2})$.

Finally, the upper block of Υ_6 is bounded by

$$\begin{aligned} & \|\widehat{D}_j^{-1} - D_j^{-1}\|O_P(1)\sum_{d=1}^{8}\frac{1}{T_0}\sum_{t\in I^c}\|\widehat{B}_t - B_t\|\|B_{dt}\||u_{jt}|(|\widehat{e}_{jt}| + 1) \\ & + O_P(1)\sum_{d=1}^{4}\|\frac{1}{T_0}\sum_{t\in I^c}\widehat{e}_{jt}u_{jt}B_{dt}\| = O_P(C_{NT}^{-2}). \end{aligned}$$

Therefore, the upper block of R_{5j} is $O_P(C_{NT}^{-2})$.

(ii) We now show the upper block of R_{6j} is $O_P(C_{NT}^{-2})$. Let

$$\begin{aligned}\Gamma &:= \frac{1}{T_0}\sum_{s\in I^c\cup\{t\}}\begin{pmatrix}\widehat{f}_s\widehat{e}_{js}\\ \widehat{g}_s\end{pmatrix}(\lambda_j'H_f^{-1},\alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{js}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}\begin{pmatrix}\widehat{f}_s-H_ff_s\\ \widehat{g}_s-H_gg_s\end{pmatrix}\\ &= \Gamma_0+\Gamma_1+...+\Gamma_6,\end{aligned}$$

where, by (D.2),

$$\begin{aligned}\Gamma_0 &:= \frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{f}_t\widehat{e}_{jt}\\ \widehat{g}_t\end{pmatrix}(\lambda_j'H_f^{-1},\alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}B^{-1}\frac{1}{N}\sum_i\begin{pmatrix}H_1'\lambda_ie_{it}\\ H_2'\alpha_i\end{pmatrix}u_{it}\\ \Gamma_d &:= \sum_{d=1}^{6}\frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{f}_t\widehat{e}_{jt}\\ \widehat{g}_t\end{pmatrix}(\lambda_j'H_f^{-1},\alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}A_{dt},\quad d=1,...,6.\end{aligned}$$

Then $R_{6j}=\widehat{D}_j^{-1}\Gamma$. We aim to show that $\Gamma_0...\Gamma_5$ are each $O_P(C_{NT}^{-2})$. In addition, the upper block of Γ_6 is $O_P(C_{NT}^{-2})$, while its lower block is $O_P(C_{NT}^{-1})$. Once this is done, we then have: $\widehat{D}_j^{-1}(\Gamma_0+...+\Gamma_5)=O_P(C_{NT}^{-2})$. Also, due to $\widehat{D}_j^{-1}-D_j^{-1}=O_P(C_{NT}^{-1})$ and D_j is block diagonal (defined in (D.4)), the upper block of $\widehat{D}_j^{-1}\Gamma_6$ is also $O_P(C_{NT}^{-2})$. It then implies that the upper block of R_{6j} is $O_P(C_{NT}^{-2})$.

The upper block of Γ_0 is

$$\begin{aligned}&\frac{1}{T_0}\sum_{s\in I^c\cup\{t\}}\widehat{f}_t\widehat{e}_{jt}(\lambda_j'H_f^{-1},\alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}B^{-1}\frac{1}{N}\sum_i\begin{pmatrix}H_1'\lambda_ie_{it}\\ H_2'\alpha_i\end{pmatrix}u_{it}\\ =\ & O_P(1)\frac{1}{T_0}\sum_{t\in I^c}\widehat{f}_t\widehat{e}_{jt}^2\frac{1}{N}\sum_i\lambda_ie_{it}u_{it}+O_P(1)\frac{1}{T_0}\sum_{t\in I^c}\widehat{f}_t\widehat{e}_{jt}\frac{1}{N}\sum_i\alpha_iu_{it}\\ =\ & O_P(C_{NT}^{-2})+O_P(1)\frac{1}{T_0}\sum_{t\in I^c}f_te_{jt}^2\frac{1}{N}\sum_i\lambda_ie_{it}u_{it}+O_P(1)\frac{1}{T_0}\sum_{t\in I^c}f_te_{jt}\frac{1}{N}\sum_i\alpha_iu_{it}\\ =\ & O_P(C_{NT}^{-2}).\end{aligned}$$

Similarly, the lower block of Γ_0 is

$$O_P(1)\frac{1}{T_0}\sum_{t\in I^c}\widehat{g}_t\frac{1}{N}\sum_i\lambda_ie_{it}u_{it}+O_P(1)\frac{1}{T_0}\sum_{t\in I^c}\widehat{g}_t\frac{1}{N}\sum_i\alpha_iu_{it}=O_P(C_{NT}^{-2}).$$

$$\begin{aligned}\Gamma_1 &= \frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{f}_t\widehat{e}_{jt}\\ \widehat{g}_t\end{pmatrix}(\lambda_j'H_f^{-1},\alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}A_{1t}\\ &= \frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{f}_t\widehat{e}_{jt}\\ \widehat{g}_t\end{pmatrix}(\lambda_j'H_f^{-1},\alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}(\widehat{B}_t^{-1}\widehat{S}_t-B^{-1}S)\begin{pmatrix}H_1^{-1}f_t\\ H_2^{-1}g_t\end{pmatrix}\end{aligned}$$

$$
\begin{aligned}
&\leq O_P(C_{NT}^{-2})(\frac{1}{T_0}\sum_{t\in I^c}(\|\widehat{f}_t\widehat{e}_{jt}\|+\|\widehat{g}_t\|)^2(\widehat{e}_{jt}^2+1)(\|f_t\|+\|g_t\|)^2)^{1/2}\\
&= O_P(C_{NT}^{-2}).
\end{aligned}
$$

Now let

$$
\mathcal{D}=\left(\frac{1}{T_0}\sum_{t\in I^c}\|\widehat{f}_t-H_f f_t\|^2+\|\widehat{g}_t-H_g g_t\|^2\right)^{1/2}=O_P(C_{NT}^{-1}).
$$

Then

$$
\begin{aligned}
\Gamma_2 &= \frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{f}_t\widehat{e}_{jt}\\ \widehat{g}_t\end{pmatrix}(\lambda_j'H_f^{-1},\alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}A_{2t}\\
&= \frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{f}_t\widehat{e}_{jt}\\ \widehat{g}_t\end{pmatrix}(\lambda_j'H_f^{-1},\alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}(\widehat{B}_t^{-1}-B^{-1})\frac{1}{N}\sum_i\begin{pmatrix}H_1'\lambda_i e_{it}\\ H_2'\alpha_i\end{pmatrix}u_{it}\\
&\leq O_P(1)\mathcal{D}\left(\frac{1}{T_0}\sum_{t\in I^c}(|\widehat{e}_{jt}|^2+1)^2\left\|\frac{1}{N}\sum_i\begin{pmatrix}\lambda_i e_{it}\\ \alpha_i\end{pmatrix}u_{it}\right\|^2\right)^{1/2}\\
&\quad +O_P(C_{NT}^{-1})\left(\frac{1}{T_0}\sum_{t\in I^c}\left\|\begin{pmatrix}f_t\widehat{e}_{jt}\\ g_t\end{pmatrix}\right\|^2(|\widehat{e}_{jt}|+1)^2\left\|\frac{1}{N}\sum_i\begin{pmatrix}\lambda_i e_{it}\\ \alpha_i\end{pmatrix}u_{it}\right\|^2\right)^{1/2}\\
&= O_P(C_{NT}^{-2}).
\end{aligned}
$$

Next, let

$$
\mathcal{A}_t=\frac{1}{N}\sum_i\begin{pmatrix}\widetilde{\lambda}_i\widehat{e}_{it}-H_1'\lambda_i e_{it}\\ \widetilde{\alpha}_i-H_2'\alpha_i\end{pmatrix}u_{it}.
$$

Then

$$
\begin{aligned}
\Gamma_3 &= \frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{f}_t\widehat{e}_{jt}\\ \widehat{g}_t\end{pmatrix}(\lambda_j'H_f^{-1},\alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}\widehat{B}_t^{-1}\mathcal{A}_t\\
&\leq O_P(1)\mathcal{D}\left(\frac{1}{T_0}\sum_{t\in I^c}(|\widehat{e}_{jt}|^2+1)^2\|\mathcal{A}_t\|^2\right)^{1/2}+O_P(1)\frac{1}{T_0}\sum_{t\in I^c}\left\|\begin{pmatrix}f_t\widehat{e}_{jt}\\ g_t\end{pmatrix}\right\|(|\widehat{e}_{jt}|+1)\|\mathcal{A}_t\|.
\end{aligned}
\tag{E.23}
$$

Cauchy Schwarz implies $\frac{1}{T_0}\sum_{t\in I^c}(|\widehat{e}_{jt}|^2+1)^2\|\mathcal{A}_t\|^2=O_P(C_{NT}^{-2})$. Therefore, the first term in (E.23) is $O_P(C_{NT}^{-2})$. As for the second term in (E.23),

$$
\frac{1}{T_0}\sum_{t\in I^c}(\|f_t e_{jt}\|+\|g_t\|)(1+|e_{jt}|)|\frac{1}{N}\sum_i(\widetilde{\lambda}_i\widehat{e}_{it}-H_1'\lambda_i e_{it})u_{it}|
$$

$$\begin{aligned}
&\leq O_P(1)(\frac{1}{T_0}\sum_{t\in I^c}|\frac{1}{N}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)e_{it}u_{it}|^2)^{1/2} \\
&\quad + O_P(1)(\frac{1}{T_0}\sum_{t\in I^c}|\frac{1}{N}\sum_i \lambda_i(\widehat{e}_{it} - e_{it})u_{it}|^2)^{1/2} + O_P(C_{NT}^{-2}) = O_P(C_{NT}^{-2}),
\end{aligned}$$

where the last equality is due to:

$$\begin{aligned}
&\frac{1}{T_0}\sum_{t\in I^c}\mathsf{E}[|\frac{1}{N}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)e_{it}u_{it}|^2|E, I] \\
&= \frac{1}{T_0}\sum_{t\in I^c}\frac{1}{N^2}(\widetilde{\Lambda} - \Lambda H_1)'\mathsf{diag}(e_t)\mathsf{Var}(u_t|E, I)\mathsf{diag}(e_t)(\widetilde{\Lambda} - \Lambda H_1) \\
&\leq O_P(1)\frac{1}{N^2}\sum_i(\widetilde{\lambda}_i - H_1'\lambda_i)^2\frac{1}{T_0}\sum_{t\in I^c}\mathsf{E}_I e_{it}^2 = O_P(C_{NT}^{-4}).
\end{aligned}$$

Also similarly, $\frac{1}{T_0}\sum_{t\in I^c}\mathsf{E}(|\frac{1}{N}\sum_i(\widetilde{\alpha}_i - H_2'\alpha_i)u_{it}|^2|I) = O_P(C_{NT}^{-4})$. Therefore, the second term in (E.23) is $O_P(C_{NT}^{-2})$. Thus $\Gamma_3 = O_P(C_{NT}^{-2})$.

Next, for C_{dt} defined in the proof of Lemma E.7

$$\begin{aligned}
\Gamma_4 &= \sum_{d=1}^{16}\frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{f}_t\widehat{e}_{jt} \\ \widehat{g}_t\end{pmatrix}(\lambda_j'H_f^{-1}, \alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0 \\ 0 & \mathsf{I}\end{pmatrix}\widehat{B}_t^{-1}C_{dt} \\
&\leq O_P(1)\sum_{d=1}^{16}\mathcal{D}[\frac{1}{T_0}\sum_{t\in I^c}|e_{jt}|^4\|C_{dt}\|^2]^{1/2} + O_P(1)\sum_{d=1}^{16}(\frac{1}{T_0}\sum_{t\in I^c}\|C_{dt}\|^2)^{1/2} \\
&= O_P(C_{NT}^{-2}), \quad \text{by (E.20)}.
\end{aligned}$$

For B_{dt} defined in the proof of Lemma E.7

$$\begin{aligned}
\Gamma_5 &= \sum_{d=1}^{8}\frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{f}_t\widehat{e}_{jt} \\ \widehat{g}_t\end{pmatrix}(\lambda_j'H_f^{-1}, \alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0 \\ 0 & \mathsf{I}\end{pmatrix}(\widehat{B}_t^{-1} - B^{-1})B_{dt} \\
&\leq O_P(1)\max_t\|\widehat{B}_t^{-1} - B^{-1}\|\sum_{d=1}^{8}\mathcal{D}[\frac{1}{T_0}\sum_{t\in I^c}(|e_{jt}|^2+1)^2\|B_{dt}\|^2]^{1/2} \\
&\quad + O_P(1)(\frac{1}{T_0}\sum_{t\in I^c}\|\widehat{B}_t^{-1} - B^{-1}\|^2)^{1/2}\sum_{d=1}^{8}(\frac{1}{T_0}\sum_{t\in I^c}\|B_{dt}\|^2)^{1/2} \\
&= O_P(C_{NT}^{-2}), \quad \text{by (E.18)}.
\end{aligned}$$

Finally, the upper block of Γ_6 is

$$\frac{1}{T_0}\sum_{t\in I^c}\widehat{f}_t\widehat{e}_{jt}(\widehat{e}_{jt}\lambda_j'H_f^{-1}, \alpha_j'H_g^{-1})B^{-1}\frac{1}{N}\sum_i\begin{pmatrix}H_1'\lambda_i e_{it} \\ H_2'\alpha_i\end{pmatrix}(\mu_i\lambda_i'f_t - \bar{x}_i\dot{\lambda}_i'\widetilde{f}_t)$$

$$
\begin{aligned}
&= O_P(1)\sum_{d=1}^{4}\frac{1}{T_0}\sum_{t\in I^c}\widehat{f}_t\widehat{e}_{jt}^2B_{dt}+O_P(1)\sum_{d=5}^{8}\frac{1}{T_0}\sum_{t\in I^c}\widehat{f}_t\widehat{e}_{jt}B_{dt}\\
&\leq O_P(1)\sum_{d=1}^{4}\frac{1}{T_0}\sum_{t\in I^c}f_te_{jt}^2B_{dt}+O_P(1)\sum_{d=1}^{4}\frac{1}{T_0}\sum_{t\in I^c}f_t(\widehat{e}_{jt}-e_{jt})^2B_{dt}\\
&\quad+O_P(1)\sum_{d=1}^{4}\frac{1}{T_0}\sum_{t\in I^c}f_t(\widehat{e}_{jt}-e_{jt})e_{jt}B_{dt}\\
&\quad+O_P(1)\sum_{d=5}^{8}\frac{1}{T_0}\sum_{t\in I^c}f_t(\widehat{e}_{jt}-e_{jt})B_{dt}+O_P(1)\sum_{d=5}^{8}\frac{1}{T_0}\sum_{t\in I^c}f_te_{jt}B_{dt}\\
&\quad+O_P(C_{NT}^{-1})(\sum_{d=1}^{4}\frac{1}{T_0}\sum_{t\in I^c}e_{jt}^4\|B_{dt}\|^2)^{1/2}+O_P(C_{NT}^{-1})(\sum_{d=1}^{8}\frac{1}{T_0}\sum_{t\in I^c}e_{jt}^2\|B_{dt}\|^2)^{1/2}\\
&= O_P(C_{NT}^{-2}),\quad \text{by (E.18) and Lemma E.11 .}
\end{aligned}
$$

The lower block of Γ_6 is, by repeatedly using Cauchy-Schwarz,

$$
\sum_{d=1}^{4}\frac{1}{T_0}\sum_{t\in I^c}\widehat{g}_t\widehat{e}_{jt}B_{dt}+\sum_{d=5}^{8}\frac{1}{T_0}\sum_{t\in I^c}\widehat{g}_tB_{dt}=O_P(C_{NT}^{-1}).
$$

Q.E.D.

Appendix F. Technical lemmas in the factor model

Here we present the intermediate results when x_{it} admits a factor structure:

$$
x_{it}=l_i'w_t+e_{it}.
$$

The proof is similar to that of the many mean model, while the main difference lies on dealing with the effect of $\widehat{e}_{it}-e_{it}$.

F.1. **The effect of $\widehat{e}_{it}-e_{it}$ in the factor model.** Let $\widehat{w}_t$ be the PC estimator of w_t. Then $\widehat{l}_i'=\frac{1}{T}\sum_s x_{is}\widehat{w}_s'$ and $\widehat{e}_{it}-e_{it}=l_i'H_x(\widehat{w}_t-H_x^{-1}w_t)+(\widehat{l}_i'-l_i'H_x)\widehat{w}_t$.

Lemma F.1. *(i)* $\max_{it}|\widehat{e}_{it}-e_{it}|=O_P(\phi_{NT})$, *where*

$$
\phi_{NT}:=(C_{NT}^{-2}(\max_i\frac{1}{T}\sum_s e_{is}^2)^{1/2}+b_{NT,4}+b_{NT,5})(1+\max_t\|w_t\|)+(b_{NT,1}+C_{NT}^{-1}b_{NT,2}+b_{NT,3}).
$$

So $\max_{it}|\widehat{e}_{it}-e_{it}|\max_{it}|e_{it}|=O_P(1)$.

(ii) All terms below are $O_P(C_{NT}^{-2})$, *for a fixed* t: $\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^4$, $\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2$, *and* $\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2e_{it}^2$, $\frac{1}{N}\sum_i\lambda_i\lambda_i'(\widehat{e}_{it}-e_{it})e_{it}$, $\frac{1}{N}\sum_i\lambda_i(\widehat{e}_{it}-e_{it})u_{it}$.

(iii) $\frac{1}{N}\sum_i\lambda_i\alpha_i'(\widehat{e}_{it}-e_{it})=O_P(C_{NT}^{-1})$ *for a fixed* t.

(iv) All terms below are $O_P(C_{NT}^{-2})$, for a fixed $j \leq N$:
$\frac{1}{T}\sum_t \|\frac{1}{N}\sum_i \alpha_i\lambda_i'(e_{it}-\widehat{e}_{it})\|^2\|f_t\|^2(\|f_t\|+\|g_t\|)^2|e_{jt}|^{2(1+r)}$, $r\in\{0,1\}$,
$\frac{1}{T_0}\sum_{t\in I^c}\|f_t\|^2|\widehat{e}_{jt}-e_{jt}|^2$, $\frac{1}{T}\sum_{t\notin I} f_te_{jt}(\widehat{e}_{jt}-e_{jt})\widetilde{f}_t$, $\frac{1}{T}\sum_t f_t(\widehat{e}_{jt}-e_{jt})u_{jt}$

Proof. Below we first simplify the expansion of $\widehat{e}_{it}-e_{it}$. Let $K_3=\dim(l_i)$. Let $\mathcal{Q}$ be a diagonal matrix consisting of the reciprocal of the first K_3 eigenvalues of $XX'/(NT)$. Let

$$\begin{aligned}\zeta_{st} &= \frac{1}{N}\sum_i(e_{is}e_{it}-\mathsf{E}e_{is}e_{it}), \quad \eta_t=\frac{1}{N}\sum_i l_ie_{it},\\ \sigma^2 &= \frac{1}{N}\sum_i \mathsf{E}e_{it}^2.\end{aligned}$$

For the PC estimator, there is a rotation matrix $\bar{H}_x$, by (A.1) of Bai (2003), (which can be simplified due to the serial independence in e_{it})

$$\begin{aligned}\widehat{w}_t-\bar{H}_xw_t &= \mathcal{Q}\frac{\sigma^2}{T}(\widehat{w}_t-\bar{H}_xw_t)+\mathcal{Q}[\frac{\sigma^2}{T}\bar{H}_x+\frac{1}{TN}\sum_{is}\widehat{w}_sl_i'e_{is}]w_t\\ &\quad+\mathcal{Q}\frac{1}{T}\sum_s\widehat{w}_s\zeta_{st}+\mathcal{Q}\frac{1}{T}\sum_s\widehat{w}_sw_s'\eta_t.\end{aligned}$$

Move $\mathcal{Q}\frac{\sigma^2}{T}(\widehat{w}_t-\bar{H}_xw_t)$ to the left hand side (LHS); then LHS becomes $(\mathsf{I}-\mathcal{Q}\frac{\sigma^2}{T})(\widehat{w}_t-\bar{H}_xw_t)$. Note that $\|\mathcal{Q}\|=O_P(1)$ so $\mathcal{Q}_1:=(\mathsf{I}-\mathcal{Q}\frac{\sigma^2}{T})^{-1}$ exists whose eigenvalues all converge to one. Then multiply $\mathcal{Q}_1$ on both sides, we reach

$$\begin{aligned}\widehat{w}_t-\bar{H}_xw_t &= \mathcal{Q}_1\mathcal{Q}[\frac{\sigma^2}{T}\bar{H}_x+\frac{1}{TN}\sum_{is}\widehat{w}_sl_i'e_{is}]w_t\\ &\quad+\mathcal{Q}_1\mathcal{Q}\frac{1}{T}\sum_s\widehat{w}_s\zeta_{st}+\mathcal{Q}_1\mathcal{Q}\frac{1}{T}\sum_s\widehat{w}_sw_s'\eta_t.\end{aligned}$$

Next, move $\mathcal{Q}_1\mathcal{Q}[\frac{\sigma^2}{T}\bar{H}_x+\frac{1}{TN}\sum_{is}\widehat{w}_sl_i'e_{is}]w_t$ to LHS, combined with $-\bar{H}_xw_t$, then LHS becomes $\widehat{w}_t-H_x^{-1}w_t$, where $H_x^{-1}=(\mathsf{I}+\mathcal{Q}_1\mathcal{Q}\frac{\sigma^2}{T})\bar{H}_x+\mathcal{Q}_1\mathcal{Q}\frac{1}{TN}\sum_{is}\widehat{w}_sl_i'e_{is}$, where
$\mathcal{Q}_1\mathcal{Q}\frac{1}{TN}\sum_{is}\widehat{w}_sl_i'e_{is}=o_P(1)$. So the eigenvalues of H_x^{-1} converge to those of $\bar{H}_x$, which are well known to be bounded away from both zero and infinity (Bai, 2003). Finally, let $R_1=\mathcal{Q}_1\mathcal{Q}$ and $R_2=\mathcal{Q}_1\mathcal{Q}\frac{1}{T}\sum_s\widehat{w}_sw_s'$, we reach

$$\widehat{w}_t-H_x^{-1}w_t = R_1\frac{1}{T}\sum_s\widehat{w}_s\zeta_{st}+R_2\eta_t. \tag{F.1}$$

with $\|R_1\|+\|R_2\|=O_P(1)$.

Also, $\widehat{l}_i' = \frac{1}{T}\sum_s x_{is}\widehat{w}_s'$, for $\mathcal{Q}_3 = -H_x(R_1\frac{1}{T^2}\sum_{m,s\leq T}\zeta_{ms}\widehat{w}_m\widehat{w}_s' + R_2\frac{1}{T}\sum_s \eta_s\widehat{w}_s') = O_P(C_{NT}^{-2})$,

$$
\begin{aligned}
\widehat{e}_{it} - e_{it} &= l_i'H_x(\widehat{w}_t - H_x^{-1}w_t) + (\widehat{l}_i' - l_i'H_x)\widehat{w}_t \\
&= l_i'H_x(\widehat{w}_t - H_x^{-1}w_t) + \frac{1}{T}\sum_s e_{is}(\widehat{w}_s' - w_s'H_x^{-1'})\widehat{w}_t \\
&\quad + \frac{1}{T}\sum_s e_{is}w_s'H_x^{-1'}\widehat{w}_t + l_i'H_x\frac{1}{T}\sum_s(H_x^{-1}w_s - \widehat{w}_s)\widehat{w}_s'\widehat{w}_t \\
&= \frac{1}{T}\sum_s e_{is}w_s'H_x^{-1'}\widehat{w}_t + \frac{1}{T^2}\sum_{s,m\leq T} e_{is}\zeta_{ms}\widehat{w}_m'R_1'\widehat{w}_t + \frac{1}{T}\sum_s e_{is}\eta_s'R_2'\widehat{w}_t \\
&\quad + l_i'\mathcal{Q}_3\widehat{w}_t + l_i'H_xR_1\frac{1}{T}\sum_s \widehat{w}_s\zeta_{st} + l_i'H_xR_2\eta_t. \qquad \text{(F.2)}
\end{aligned}
$$

(i) We first show that $\max_t\|\widehat{w}_t - H_x^{-1}w_t\| = O_P(1)$. Define

$$
\begin{aligned}
b_{NT,1} &= \max_t \|\frac{1}{NT}\sum_{is} w_s(e_{is}e_{it} - \mathsf{E}e_{is}e_{it})\| \\
b_{NT,2} &= (\max_t \frac{1}{T}\sum_s(\frac{1}{N}\sum_i e_{is}e_{it} - \mathsf{E}e_{is}e_{it})^2)^{1/2} \\
b_{NT,3} &= \max_t \|\frac{1}{N}\sum_i l_ie_{it}\| \\
b_{NT,4} &= \max_i \|\frac{1}{T}\sum_s e_{is}w_s\| \\
b_{NT,5} &= \max_i \|\frac{1}{NT}\sum_{js} l_j(e_{js}e_{is} - \mathsf{E}e_{js}e_{is})\|
\end{aligned}
$$

Then

$$
\begin{aligned}
\max_t \|\frac{1}{T}\sum_s \widehat{w}_s\zeta_{st}\| &\leq O_P(1)\max_t\|\frac{1}{T}\sum_s w_s\zeta_{st}\| + O_P(C_{NT}^{-1})(\max_t\frac{1}{T}\sum_s\zeta_{st}^2)^{1/2} \\
&= O_P(b_{NT,1} + C_{NT}^{-1}b_{NT,2}).
\end{aligned}
$$

Then by assumption, $\max_t\|\widehat{w}_t - H_x^{-1}w_t\| = O_P(b_{NT,1} + C_{NT}^{-1}b_{NT,2} + b_{NT,3}) = O_P(1)$. As such $\max_t\|\widehat{w}_t\| = O_P(1) + \max_t\|w_t\|$.

In addition,

$$
\begin{aligned}
\max_i \|\frac{1}{T^2}\sum_{m\leq T}\sum_{s\leq T} e_{is}\zeta_{ms}\widehat{w}_m'\| &\leq O_P(C_{NT}^{-1})(\max_i\frac{1}{T}\sum_{s\leq T}e_{is}^2)^{1/2}(\frac{1}{T^2}\sum_{s,t\leq T}\zeta_{st}^2)^{1/2} \\
&\quad + O_P(1)(\max_i\frac{1}{T}\sum_s e_{is}^2)^{1/2}(\frac{1}{T}\sum_s\|\frac{1}{TN}\sum_{jt}(e_{jt}e_{js} - \mathsf{E}e_{jt}e_{js})w_t\|^2)^{1/2}
\end{aligned}
$$

$$= O_P(C_{NT}^{-2})(\max_i \frac{1}{T}\sum_s e_{is}^2)^{1/2}.$$

So $\max_{it} |\widehat{e}_{it} - e_{it}| = O_P(\phi_{NT})$, where

$$\phi_{NT} := (C_{NT}^{-2}(\max_i \frac{1}{T}\sum_s e_{is}^2)^{1/2} + b_{NT,4} + b_{NT,5})(1 + \max_t \|w_t\|) + (b_{NT,1} + C_{NT}^{-1} b_{NT,2} + b_{NT,3}).$$

(ii) Let $a \in \{1,2,4\}$, and $b \in \{0,1,2\}$, and a bounded constant sequence c_i, consider, up to a $O_P(1)$ multiplier that is independent of (t,i),

$$\begin{aligned}
\frac{1}{N}\sum_i c_i e_{it}^b (\widehat{e}_{it} - e_{it})^a &= \frac{1}{N}\sum_i c_i e_{it}^b (\frac{1}{T}\sum_s e_{is} w_s)^a \widehat{w}_t^a \\
&+ \frac{1}{N}\sum_i c_i e_{it}^b (\frac{1}{T^2}\sum_{s,m\le T} e_{is}\zeta_{ms}\widehat{w}_m)^a \widehat{w}_t^a \\
&+ \frac{1}{N}\sum_i c_i e_{it}^b (\frac{1}{T}\sum_s e_{is}\eta_s')^a \widehat{w}_t^a + \frac{1}{N}\sum_i c_i e_{it}^b l_i^a \mathcal{Q}_3^a \widehat{w}_t^a \\
&+ \frac{1}{N}\sum_i c_i e_{it}^b l_i^a (\frac{1}{T}\sum_s \widehat{w}_s \zeta_{st})^a + \frac{1}{N}\sum_i c_i e_{it}^b l_i^a \eta_t^a. \quad \text{(F.3)}
\end{aligned}$$

Note that

$$\begin{aligned}
\frac{1}{T}\sum_s \widehat{w}_s \zeta_{st} &\le O_P(1)\frac{1}{TN}\sum_{is} w_s(e_{is}e_{it} - \mathsf{E} e_{is}e_{it}) \\
&\quad + (\frac{1}{T}\sum_s (\frac{1}{N}\sum_i (e_{is}e_{it} - \mathsf{E} e_{is}e_{it}))^2)^{1/2} O_P(C_{NT}^{-1}) = O_P(C_{NT}^{-2}) \\
\frac{1}{N}\sum_i e_{it}^b (\frac{1}{T}\sum_s e_{is} w_s)^a &= O(T^{-a/2}), \quad a = 2,4, \quad b = 0,2 \\
\frac{1}{N}\sum_i (\frac{1}{T^2}\sum_{s,m\le T} e_{is}\zeta_{ms}\widehat{w}_m)^4 &\le (\frac{1}{T^2}\sum_{m,s\le T}\zeta_{ms}^2)^2 (\frac{1}{T}\sum_m w_m^2)^2 \frac{1}{N}\sum_i (\frac{1}{T}\sum_s e_{is}^2)^2 \\
&\quad + \frac{1}{N}\sum_i (\frac{1}{T}\sum_{m\le T}(\frac{1}{T}\sum_{s\le T} e_{is}\zeta_{ms})^2)^2 O_P(C_{NT}^{-4}) = O_P(C_{NT}^{-4}) \\
\frac{1}{N}\sum_i (\frac{1}{T^2}\sum_{s,m\le T} e_{is}\zeta_{ms}\widehat{w}_m)^2 &\le \frac{1}{N}\sum_i \frac{1}{T}\sum_m (\frac{1}{T}\sum_s e_{is}\zeta_{ms})^2 \frac{1}{T}\sum_m w_m^2 \\
&\quad + \frac{1}{N}\sum_i \frac{1}{T}\sum_{m\le T}(\frac{1}{T}\sum_{s\le T} e_{is}\zeta_{ms})^2 O_P(C_{NT}^{-2}) = O_P(C_{NT}^{-2}) \\
\frac{1}{N}\sum_i (\frac{1}{T}\sum_s e_{is}\eta_s')^4 &\le \frac{1}{N}\sum_i (\frac{1}{T}\sum_s e_{is}^2)^2 (\frac{1}{T}\sum_t \eta_t^2)^2 = O_P(C_{NT}^{-4}) \\
\frac{1}{N}\sum_i e_{it}^2 (\frac{1}{T}\sum_s e_{is}\eta_s')^2 &= O_P(C_{NT}^{-2})
\end{aligned}$$

$$\begin{aligned}
\frac{1}{N}\sum_i e_{it}^2 l_i^2 (\frac{1}{T}\sum_s \widehat{w}_s\zeta_{st})^2 &\leq \frac{1}{N}\sum_i e_{it}^2 l_i^2(\frac{1}{T}\sum_s w_s\zeta_{st})^2 + O_P(C_{NT}^{-2}) = O_P(C_{NT}^{-2}) \\
\frac{1}{T}\sum_t(\frac{1}{T}\sum_s \widehat{w}_s\zeta_{st})^2 &\leq O_P(C_{NT}^{-4}) \\
\frac{1}{N}\sum_i c_i e_{it}(\frac{1}{T}\sum_s e_{is}w_s) &= O_P(1)(\frac{1}{N^2T^2}\sum_{ij\leq N}\sum_{sl\leq T} c_ic_j\mathsf{E}w_sw_l\mathsf{E}(e_{it}e_{jt}e_{is}e_{jl}|W))^{1/2} \\
&= O_P(C_{NT}^{-2}) \\
\frac{1}{N}\sum_i c_ie_{it}(\frac{1}{T^2}\sum_{s,m\leq T} e_{is}\zeta_{ms}\widehat{w}_m) &\leq O_P(C_{NT}^{-2}) + \frac{1}{N}\sum_i c_ie_{it}(\frac{1}{T}\sum_{s\leq T}\frac{1}{T}\sum_{m\leq T} e_{is}\zeta_{ms}w_m) \\
&\leq O_P(C_{NT}^{-2}) + O_P(1)(\frac{1}{T}\sum_{s\leq T}\mathsf{E}(\frac{1}{T}\sum_{m\leq T}\zeta_{ms}w_m)^2)^{1/2} \\
&= O_P(C_{NT}^{-2}).
\end{aligned}$$

where the last equality follows from the following:

$$\begin{aligned}
&\frac{1}{T}\sum_{s\leq T}\mathsf{E}(\frac{1}{T}\sum_{m\leq T}\zeta_{ms}w_m)^2 = O(T^{-2}) + \frac{1}{T}\sum_{s\leq T}\frac{1}{T^2}\sum_{t\neq s}\mathsf{E}w_sw_t\mathsf{Cov}(\zeta_{ss},\zeta_{ts}|W) \\
&+\frac{1}{T}\sum_{s\leq T}\frac{1}{T}\sum_{m\neq s}\frac{1}{T}\sum_{t\leq T}\mathsf{E}w_mw_t\mathsf{Cov}(\zeta_{ms},\zeta_{ts}|W) \\
= \; & O(T^{-2}) + \frac{1}{T}\sum_{s\leq T}\frac{1}{T}\sum_{t\neq s}\frac{1}{T}\mathsf{E}w_t^2\frac{1}{N^2}\sum_{ij}\mathsf{E}(e_{is}e_{js}|W)\mathsf{E}(e_{it}e_{jt}|W) = O(C_{NT}^{-4}).
\end{aligned}$$

With the above results ready, we can proceed proving (ii)(iii) as follows.

Now for $a=4, b=0$, $c_i=1$, up to a $O_P(1)$ multiplier

$$\begin{aligned}
\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^4 &= \frac{1}{N}\sum_i(\frac{1}{T}\sum_s e_{is}w_s)^4 + \frac{1}{N}\sum_i(\frac{1}{T^2}\sum_{s,m\leq T} e_{is}\zeta_{ms}\widehat{w}_m)^4 \\
&+\frac{1}{N}\sum_i(\frac{1}{T}\sum_s e_{is}\eta_s')^4 + \mathcal{Q}_3^4 + (\frac{1}{T}\sum_s \widehat{w}_s\zeta_{st})^4 + \eta_t^4 \\
&\leq O_P(C_{NT}^{-4}).
\end{aligned}$$

For $a=2$, $b=0$, $c_i=1$,

$$\begin{aligned}
\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2 &\leq \frac{1}{N}\sum_i(\frac{1}{T}\sum_s e_{is}w_s)^2 + \frac{1}{N}\sum_i(\frac{1}{T^2}\sum_{s,m\leq T} e_{is}\zeta_{ms}\widehat{w}_m)^2 \\
&+\frac{1}{N}\sum_i(\frac{1}{T}\sum_s e_{is}\eta_s')^2 + \mathcal{Q}_3^2 + (\frac{1}{T}\sum_s \widehat{w}_s\zeta_{st})^2 + \eta_t^2 \\
&\leq O_P(C_{NT}^{-2}).
\end{aligned}$$

For $a=2$, $b=2$, $c_i=1$,

$$\begin{aligned}\frac{1}{N}\sum_i e_{it}^2(\widehat{e}_{it}-e_{it})^2 &= \frac{1}{N}\sum_i e_{it}^2(\frac{1}{T}\sum_s e_{is}w_s)^2+\frac{1}{N}\sum_i e_{it}^2(\frac{1}{T^2}\sum_{s,m\leq T} e_{is}\zeta_{ms}\widehat{w}_m)^2\\ &\quad +\frac{1}{N}\sum_i e_{it}^2(\frac{1}{T}\sum_s e_{is}\eta_s')^2+\mathcal{Q}_3^2+(\frac{1}{T}\sum_s\widehat{w}_s\zeta_{st})^2+\eta_t^2\\ &\leq O_P(C_{NT}^{-2}).\end{aligned}$$

Next, let $a=b=1$ and c_i be any element of $\lambda_i\lambda_i'$,

$$\begin{aligned}\frac{1}{N}\sum_i c_ie_{it}(\widehat{e}_{it}-e_{it}) &= \frac{1}{N}\sum_i c_ie_{it}(\frac{1}{T}\sum_s e_{is}w_s)+\frac{1}{N}\sum_i c_ie_{it}(\frac{1}{T^2}\sum_{s,m\leq T} e_{is}\zeta_{ms}\widehat{w}_m)\\ &\quad +\frac{1}{N}\sum_i c_ie_{it}(\frac{1}{T}\sum_s e_{is}\eta_s')+\frac{1}{N}\sum_i c_ie_{it}l_i\mathcal{Q}_3\\ &\quad +\frac{1}{N}\sum_i c_ie_{it}l_i(\frac{1}{T}\sum_s\widehat{w}_s\zeta_{st})+\frac{1}{N}\sum_i c_ie_{it}l_i\eta_t\\ &\leq O_P(C_{NT}^{-2}).\end{aligned}$$

Next, ignoring an $O_P(1)$ multiplier,

$$\begin{aligned}\frac{1}{N}\sum_i \lambda_i(\widehat{e}_{it}-e_{it})u_{it} &\leq \frac{1}{N}\sum_i\lambda_i\frac{1}{T}\sum_s e_{is}w_su_{it}\\ &\quad +\frac{1}{N}\sum_i\lambda_i\frac{1}{T^2}\sum_{s,m\leq T}e_{is}\zeta_{ms}\widehat{w}_mu_{it}+\frac{1}{N}\sum_i\lambda_i\frac{1}{T}\sum_s e_{is}\eta_su_{it}\\ &\quad +\frac{1}{N}\sum_i\lambda_il_i'\mathcal{Q}_3u_{it}+\frac{1}{N}\sum_i\lambda_il_iu_{it}\frac{1}{T}\sum_s\widehat{w}_s\zeta_{st}+\frac{1}{N}\sum_i\lambda_il_i\eta_tu_{it}\\ &\leq O_P(C_{NT}^{-2}).\end{aligned}$$

(iii) Let $a=1, b=0$ and c_i be any element of $\lambda_i\alpha_i'$,

$$\begin{aligned}\frac{1}{N}\sum_i c_i(\widehat{e}_{it}-e_{it}) &= \frac{1}{N}\sum_i c_i(\frac{1}{T}\sum_s e_{is}w_s)+\frac{1}{N}\sum_i c_i(\frac{1}{T^2}\sum_{s,m\leq T} e_{is}\zeta_{ms}\widehat{w}_m)\\ &\quad +\frac{1}{N}\sum_i c_i(\frac{1}{T}\sum_s e_{is}\eta_s')+\mathcal{Q}_3+(\frac{1}{T}\sum_s\widehat{w}_s\zeta_{st})+\eta_t\\ &\leq O_P(C_{NT}^{-1}),\end{aligned}$$

where the dominating term is $\eta_t=O_P(C_{NT}^{-1})$.

(iv) Up to an $O_P(1)$ multiplier, for $r\in\{0,1\}$,

$$\frac{1}{T}\sum_t\|\frac{1}{N}\sum_i\alpha_i\lambda_i'(e_{it}-\widehat{e}_{it})\|^2\|f_t\|^2(\|f_t\|+\|g_t\|)^2|e_{jt}|^{2(1+r)}$$

$$
\begin{aligned}
&\leq \|\frac{1}{N}\sum_i \alpha_i\lambda_i'\frac{1}{T}\sum_s e_{is}w_s\|^2 O_P(1)+\|\frac{1}{N}\sum_i \alpha_i\lambda_i'\frac{1}{T}\sum_s e_{is}\eta_s\|^2 O_P(1)\\
&\quad +O_P(1)\mathcal{Q}_3^2+\frac{1}{T^2}\sum_{mt\leq T}\zeta_{mt}^2\frac{1}{T}\sum_s\|\frac{1}{N}\sum_i\alpha_i\lambda_i' e_{is}\|^2 O_P(1)\\
&\quad +\|\frac{1}{T}\sum_s \widehat{w}_s\zeta_{st}\|^2 O_P(1)+\frac{1}{T}\sum_t \eta_t^2\|f_t\|^2(\|f_t\|+\|g_t\|)^2|e_{jt}|^{2(1+r)}\\
&= O_P(C_{NT}^{-2})+(\frac{1}{T}\sum_t\eta_t^4)^{1/2}O_P(1)=O_P(C_{NT}^{-2}),
\end{aligned}
$$

with the assumptions that $\mathsf{E}e_{jt}^4\|f_t\|^8+\mathsf{E}e_{jt}^4\|f_t\|^4\|g_t\|^4+\mathsf{E}e_{jt}^2\|f_t\|^4\|w_t\|^2<C$ and $\mathsf{E}e_{jt}^2\|f_t\|^2\|w_t\|^2\|g_t\|^2<C$, $\mathsf{E}\frac{1}{\sqrt{N}}\sum_i l_i e_{it}\|^4<C$.

Similarly, $\frac{1}{T_0}\sum_{t\in I^c}\|f_t\|^2|\widehat{e}_{jt}-e_{jt}|^2=O_P(C_{NT}^{-2})$.

Next, for a fixed $i\leq N$,

$$
\begin{aligned}
\frac{1}{T}\sum_t f_t e_{it}\widehat{w}_t\widetilde{f}_t &= O_P(C_{NT}^{-1}), \quad \frac{1}{T}\sum_t f_t u_{it}\widehat{w}_t=O_P(C_{NT}^{-1})\\
\frac{1}{T}\sum_{t\notin I} f_t e_{it}\widetilde{f}_t &= O_P(C_{NT}^{-1}), \quad \frac{1}{T}\sum_{t\notin I} f_t e_{it}\eta_t\widetilde{f}_t=O_P(C_{NT}^{-2})
\end{aligned}
$$

Also, $h_i:=\frac{1}{T}\sum_s e_{is}w_s+\frac{1}{T}\sum_s e_{is}\eta_s+\mathcal{Q}_3+\frac{1}{T^2}\sum_{s,m\leq T}e_{is}\zeta_{ms}\widehat{w}_m=O_P(C_{NT}^{-1})$. So

$$
\begin{aligned}
\frac{1}{T}\sum_{t\notin I} f_t e_{it}(\widehat{e}_{it}-e_{it})\widetilde{f}_t &\leq \frac{1}{T}\sum_{t\notin I} f_t e_{it}\widehat{w}_t\widetilde{f}_t h_i+\frac{1}{T}\sum_{t\notin I} f_t e_{it}\widetilde{f}_t\frac{1}{T}\sum_s\widehat{w}_s\zeta_{st}+\frac{1}{T}\sum_{t\notin I} f_t e_{it}\eta_t\widetilde{f}_t\\
&= O_P(C_{NT}^{-2})\\
\frac{1}{T}\sum_t f_t(\widehat{e}_{it}-e_{it})u_{it} &\leq \frac{1}{T}\sum_t f_t u_{it}\widehat{w}_t h_i+\frac{1}{T}\sum_t f_t u_{it}\frac{1}{T}\sum_s\widehat{w}_s\zeta_{st}+\frac{1}{T}\sum_t f_t u_{it}\eta_t\\
&= O_P(C_{NT}^{-2}).
\end{aligned}
$$

Lemma F.2. *Assume* $\max_{it}|e_{it}|C_{NT}^{-1}=O_P(1)$ *and* $\mathsf{E}e_{it}^8<C$.

Let c_i *be a non-random bounded sequence.*

(i) $\max_t\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^4=O_P(1+\max_t\|w_t\|^4+b_{NT,2}^4)C_{NT}^{-4}+O_P(b_{NT,1}^4+b_{NT,3}^4)$.

$\max_t\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2e_{it}^2\leq O_P(1+\max_t\|w_t\|^2+b_{NT,2}^2)\max_{it}|e_{it}|^2C_{NT}^{-2}$
$+O_P(b_{NT,1}^2+b_{NT,3}^2)\max_t\frac{1}{N}\sum_i e_{it}^2$.

$\max_t|\frac{1}{N}\sum_i c_i(\widehat{e}_{it}-e_{it})e_{it}|\leq O_P(1+\max_t\|w_t\|+b_{NT,2})\max_{it}|e_{it}|C_{NT}^{-1}$
$+\max_t\|\frac{1}{N}\sum_i c_ie_{it}\|_F O_P(b_{NT,1}+b_{NT,3})$.

$\max_t|\frac{1}{N}\sum_i c_i(\widehat{e}_{it}-e_{it})^2|\leq O_P(1+\max_t\|w_t\|^2+b_{NT,2}^2)C_{NT}^{-2}+O_P(b_{NT,1}^2+b_{NT,3}^2)$.

$\max_t\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2\leq O_P(1+\max_t\|w_t\|^2+b_{NT,2}^2)C_{NT}^{-2}+O_P(b_{NT,1}^2+b_{NT,3}^2)$

$\max_t|\frac{1}{N}\sum_i c_i(\widehat{e}_{it}-e_{it})|\leq O_P(1+\max_t\|w_t\|)C_{NT}^{-2}+O_P(b_{NT,1}+b_{NT,3}+C_{NT}^{-1}b_{NT,2})$.

$\max_i \frac{1}{T}\sum_t(\widehat{e}_{it}-e_{it})^2 \leq O_P(b_{NT,4}^2+b_{NT,5}^2+C_{NT}^{-2})$.
(ii) All terms below are $O_P(C_{NT}^{-2})$: $\frac{1}{NT}\sum_{it}(\widehat{e}_{it}-e_{it})^4$
$\frac{1}{NT}\sum_{it}(\widehat{e}_{it}-e_{it})^2$, $\frac{1}{NT}\sum_{it}e_{it}^2(\widehat{e}_{it}-e_{it})^2$, $\frac{1}{T}\sum_t|\frac{1}{N}\sum_i c_i(e_{it}-\widehat{e}_{it})|^2$, *and*
$\frac{1}{TN}\sum_{t\in I^c}\sum_i \widetilde{f}_t f_t e_{jt}^r c_i e_{it}(\widehat{e}_{it}-e_{it})$ *for* $r=\{0,1,2\}$, *for a fixed* j.
$\frac{1}{TN}\sum_{t\in I^c}\sum_i f_t f_t e_{jt} c_i(\widehat{e}_{it}-e_{it})$ *for a fixed* j.
(iii) All terms below are $O_P(C_{NT}^{-4})$: $\frac{1}{T}\sum_t|\frac{1}{N}\sum_i c_i(\widehat{e}_{it}-e_{it})e_{it}|^2$,
$\frac{1}{T}\sum_t|\frac{1}{N}\sum_i c_i(\widehat{e}_{it}-e_{it})^2|^2$, $\frac{1}{T}\sum_t|\frac{1}{N}\sum_i c_i(\widehat{e}_{it}-e_{it})u_{it}|^2$,
For a fixed $j\leq N$, *and* $r\leq 2$,
$\frac{1}{T}\sum_t|\frac{1}{N}\sum_i c_i e_{it}e_{jt}^r(\widehat{e}_{it}-e_{it})f_t|^2+\frac{1}{T}\sum_{t\in I^c}(|e_{jt}|+1)^2u_{jt}^2|\frac{1}{N}\sum_i c_i e_{it}(\widehat{e}_{it}-e_{it})|^2$

Proof. (i) First, in the proof of Lemma F.1(i), we showed $\max_t\|\widehat{w}_t\|\leq O_P(1+\max_t\|w_t\|)$. By the proof of Lemma F.1(ii),

$$\begin{aligned}
&\max_t \frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^4 \leq \frac{1}{N}\sum_i(\frac{1}{T}\sum_s e_{is}w_s)^4\max_t\widehat{w}_t^4+\frac{1}{N}\sum_i(\frac{1}{T^2}\sum_{s,m\leq T}e_{is}\zeta_{ms}\widehat{w}_m)^4\max_t\widehat{w}_t^4\\
&+\frac{1}{N}\sum_i(\frac{1}{T}\sum_s e_{is}\eta_s')^4\max_t\widehat{w}_t^4+\mathcal{Q}_3^4\max_t\widehat{w}_t^4+\max_t(\frac{1}{T}\sum_s\widehat{w}_s\zeta_{st})^4+\max_t\eta_t^4\\
\leq\ & O_P(1+\max_t\|w_t\|^4+b_{NT,2}^4)(C_{NT}^{-4})+O_P(b_{NT,1}^4+b_{NT,3}^4).
\end{aligned}$$

Next,

$$\begin{aligned}
\max_t\frac{1}{N}\sum_i e_{it}^2(\widehat{e}_{it}-e_{it})^2 &= \max_t\frac{1}{N}\sum_i e_{it}^2(\frac{1}{T}\sum_s e_{is}w_s)^2\max_t\widehat{w}_t^2+\max_t\frac{1}{N}\sum_i e_{it}^2 l_i^4\mathcal{Q}_3^2\max_t\widehat{w}_t^2\\
&+\max_t\frac{1}{N}\sum_i e_{it}^2(\frac{1}{T^2}\sum_{s,m\leq T}e_{is}\zeta_{ms}\widehat{w}_m)^2\max_t\widehat{w}_t^2\\
&+\max_t\frac{1}{N}\sum_i e_{it}^2(\frac{1}{T}\sum_s e_{is}\eta_s')^2\max_t\widehat{w}_t^2\\
&+\max_t\frac{1}{N}\sum_i e_{it}^2 l_i^2(\frac{1}{T}\sum_s\widehat{w}_s\zeta_{st})^2+\max_t\frac{1}{N}\sum_i e_{it}^2 l_i^2\eta_t^2\\
&\leq O_P(1+\max_t\|w_t\|^2+b_{NT,2}^2)\max_{it}|e_{it}|^2C_{NT}^{-2}\\
&+O_P(b_{NT,1}^2+b_{NT,3}^2)\max_t\frac{1}{N}\sum_i e_{it}^2.\\
\frac{1}{N}\sum_i c_i e_{it}(\widehat{e}_{it}-e_{it}) &\leq \max_t\frac{1}{N}\sum_i c_i e_{it}(\frac{1}{T}\sum_s e_{is}w_s)\widehat{w}_t\\
&+\max_t\frac{1}{N}\sum_i c_i e_{it}(\frac{1}{T^2}\sum_{s,m\leq T}e_{is}\zeta_{ms}\widehat{w}_m)\widehat{w}_t\\
&+\max_t\frac{1}{N}\sum_i c_i e_{it}(\frac{1}{T}\sum_s e_{is}\eta_s')\widehat{w}_t+\max_t\frac{1}{N}\sum_i c_i e_{it}l_i\mathcal{Q}_3\max_t\widehat{w}_t
\end{aligned}$$

$$
\begin{aligned}
& +\max_t \frac{1}{N}\sum_i c_i e_{it} l_i (\frac{1}{T}\sum_s \widehat{w}_s \zeta_{st}) + \max_t \frac{1}{N}\sum_i c_i e_{it} l_i \max_t \eta_t \\
\leq\ & O_P(1+\max_t \|w_t\| + b_{NT,2}) \max_{it} |e_{it}| C_{NT}^{-1} \\
& +\max_t \|\frac{1}{N}\sum_i c_i e_{it} l_i\|_F O_P(b_{NT,1}+b_{NT,3}).
\end{aligned}
$$

$$
\begin{aligned}
\max_t \frac{1}{N}\sum_i c_i (\widehat{e}_{it}-e_{it})^2 \leq\ & \frac{1}{N}\sum_i c_i (\frac{1}{T}\sum_s e_{is} w_s)^2 \max_t \widehat{w}_t^2 + \max_t (\frac{1}{T}\sum_s \widehat{w}_s \zeta_{st})^2 + \max_t \eta_t^2 \\
& +\frac{1}{N}\sum_i c_i (\frac{1}{T^2}\sum_{s,m\leq T} e_{is}\zeta_{ms}\widehat{w}_m)^2 \max_t \widehat{w}_t^2 \\
& +\frac{1}{N}\sum_i c_i (\frac{1}{T}\sum_s e_{is}\eta_s')^2 \max_t \widehat{w}_t^2 + \max_t \mathcal{Q}_3^2 \widehat{w}_t^2 \\
\leq\ & O_P(1+\max_t \|w_t\|^2 + b_{NT,2}^2) C_{NT}^{-2} + O_P(b_{NT,1}^2 + b_{NT,3}^2),
\end{aligned}
$$

$$
\begin{aligned}
\max_t \frac{1}{N}\sum_i c_i (\widehat{e}_{it}-e_{it}) \leq\ & \frac{1}{N}\sum_i c_i (\frac{1}{T}\sum_s e_{is} w_s) \max_t \widehat{w}_t \\
& +\frac{1}{N}\sum_i c_i (\frac{1}{T^2}\sum_{s,m\leq T} e_{is}\zeta_{ms}\widehat{w}_m) \max_t \widehat{w}_t \\
& +\frac{1}{N}\sum_i c_i (\frac{1}{T}\sum_s e_{is}\eta_s') \max_t \widehat{w}_t + \mathcal{Q}_3 \max_t \widehat{w}_t \\
& +\max_t (\frac{1}{T}\sum_s \widehat{w}_s \zeta_{st}) + \max_t \eta_t \\
\leq\ & O_P(1+\max_t \|w_t\|) C_{NT}^{-2} + O_P(b_{NT,1}+b_{NT,3}+C_{NT}^{-1} b_{NT,2}).
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\max_i \frac{1}{T}\sum_t (\widehat{e}_{it}-e_{it})^2 \leq\ & \max_i \frac{1}{T}\sum_t \widehat{w}_t^2 (\frac{1}{T}\sum_s e_{is} w_s' H_x^{-1'})^2 + \max_i \frac{1}{T}\sum_t \widehat{w}_t^2 (\frac{1}{T^2}\sum_{s,m\leq T} e_{is}\zeta_{ms}\widehat{w}_m' R_1')^2 \\
& +\max_i \frac{1}{T}\sum_t \widehat{w}_t^2 (\frac{1}{T}\sum_s e_{is}\eta_s' R_2')^2 + \max_i \frac{1}{T}\sum_t \widehat{w}_t^2 (l_i'\mathcal{Q}_3)^2 \\
& +\max_i \frac{1}{T}\sum_t (l_i' H_x R_1 \frac{1}{T}\sum_s \widehat{w}_s \zeta_{st})^2 + \max_i \frac{1}{T}\sum_t \eta_t^2 (l_i' H_x R_2)^2 \\
\leq\ & O_P(b_{NT,4}^2 + b_{NT,5}^2 + C_{NT}^{-2}).
\end{aligned}
$$

(ii) Note that $\max_t \|\widehat{w}_t\|^2 = O_P(1) + O_P(1)\max_t \|w_t\|^2 \leq O_P(1) + o_P(C_{NT})$, where the last inequality follows from the assumption that $\max_t \|w_t\|^2 = o_P(C_{NT})$.

$$
\frac{1}{NT}\sum_{it}(\widehat{e}_{it}-e_{it})^4 \leq \frac{1}{T}\sum_t \|\widehat{w}_t\|^2 \max_t \|\widehat{w}_t\|^2 [\frac{1}{N}\sum_i (\frac{1}{T}\sum_s e_{is} w_s)^4 + O_P(1)(\frac{1}{T^2}\sum_{s,m\leq T} \zeta_{ms}^2)^2]
$$

$$
\begin{aligned}
&+\frac{1}{T}\sum_t \|\widehat{w}_t\|^2 \max_t \|\widehat{w}_t\|^2 [\frac{1}{N}\sum_i (\frac{1}{T}\sum_s e_{is}\eta_s)^4 + \mathcal{Q}_3^4] \\
&+\frac{1}{T}\sum_t (\frac{1}{T}\sum_s \zeta_{st}^2)^2 O_P(C_{NT}^{-4}) + \frac{1}{T}\sum_t \eta_t^4 + \frac{1}{T}\sum_t (\frac{1}{T}\sum_s w_s\zeta_{st})^4 \\
\leq\ & (O_P(1) + o_P(C_{NT}))C_{NT}^{-4} + O_P(C_{NT}^{-2}) \\
&+O_P(1)\frac{1}{T}\sum_t \frac{1}{T^4}\sum_{s,k,l,m\leq T} \mathsf{E} w_l w_m w_k w_s \mathsf{E}(\zeta_{st}\zeta_{kt}\zeta_{lt}\zeta_{mt}|W) \\
=\ & O_P(C_{NT}^{-2}).
\end{aligned}
$$

Similarly, $\frac{1}{NT}\sum_{it}(\widehat{e}_{it} - e_{it})^2 = O_P(C_{NT}^{-2})$.

$$
\begin{aligned}
\frac{1}{NT}\sum_{it} e_{it}^2(\widehat{e}_{it} - e_{it})^2 \leq\ & \frac{1}{NT}\sum_{it} e_{it}^2\widehat{w}_t^2(\frac{1}{T}\sum_s e_{is}\eta_s)^2 + \frac{1}{NT}\sum_{it} e_{it}^2\widehat{w}_t^2(\frac{1}{T}\sum_s e_{is}w_s)^2 \\
&+\frac{1}{NT}\sum_{it} e_{it}^2\widehat{w}_t^2(\frac{1}{T^2}\sum_{s,m\leq T} e_{is}\zeta_{ms}\widehat{w}_m)^2 + \frac{1}{NT}\sum_{it} e_{it}^2\widehat{w}_t^2(l_i'\mathcal{Q}_3)^2 \\
&+\frac{1}{NT}\sum_{it} e_{it}^2(l_i'\frac{1}{T}\sum_s \widehat{w}_s\zeta_{st})^2 + \frac{1}{NT}\sum_{it} e_{it}^2(l_i'\eta_t)^2 = O_P(C_{NT}^{-2}).
\end{aligned}
$$

Next,

$$
\begin{aligned}
\frac{1}{T}\sum_t(\frac{1}{N}\sum_i c_i(e_{it} - \widehat{e}_{it}))^2 \leq\ & \frac{1}{T}\sum_t \widehat{w}_t^2(\frac{1}{N}\sum_i c_i\frac{1}{T}\sum_s e_{is}w_s)^2 \\
&+\frac{1}{T}\sum_t \widehat{w}_t^2(\frac{1}{N}\sum_i c_i\frac{1}{T^2}\sum_{s,m\leq T} e_{is}\zeta_{ms}\widehat{w}_m)^2 \\
&+\frac{1}{T}\sum_t \widehat{w}_t^2(\frac{1}{N}\sum_i c_i\frac{1}{T}\sum_s e_{is}\eta_s)^2 \\
&+\frac{1}{T}\sum_t \widehat{w}_t^2\mathcal{Q}_3^2 + \frac{1}{T}\sum_t(\frac{1}{T}\sum_s \widehat{w}_s\zeta_{st})^2 + \frac{1}{T}\sum_t \eta_t^2 \\
=\ & O_P(C_{NT}^{-2}).
\end{aligned}
$$

Finally, for $r = 0, 1, 2$, and $p = 0, 1$,

$$
\begin{aligned}
&\frac{1}{TN}\sum_{t\in I^c}\sum_i f_t^2 e_{jt}^r c_i e_{it}^p(\widehat{e}_{it} - e_{it}) \\
=\ & \frac{1}{TN}\sum_{t\in I^c} f_t^2 e_{jt}^r \sum_i c_i e_{it}^p \widehat{w}_t \frac{1}{T^2}\sum_{s,m\leq T} e_{is}\zeta_{ms}\widehat{w}_m + \frac{1}{TN}\sum_{t\in I^c} f_t^2 e_{jt}^r \sum_i c_i e_{it}^p l_i \frac{1}{T}\sum_s \widehat{w}_s\zeta_{st} \\
&+\frac{1}{TN}\sum_{t\in I^c}\widehat{w}_t f_t^2 e_{jt}^r \sum_i c_i e_{it}^p l_i'\mathcal{Q}_3 + \frac{1}{TN}\sum_{t\in I^c} f_t\widehat{w}_t^2 e_{jt}^r \sum_i c_i e_{it}^p \frac{1}{T}\sum_s e_{is}(w_s + \eta_s) \\
&+\frac{1}{TN}\sum_{t\in I^c} f_t^2 e_{jt}^r \sum_i c_i e_{it}^p l_i\eta_t
\end{aligned}
$$

$$= \frac{1}{TN}\sum_{t\in I^c} f_t\widehat{w}_t^2 e_{jt}^r \sum_i c_i e_{it}^p \frac{1}{T}\sum_s e_{is}(w_s+\eta_s) + \frac{1}{TN}\sum_{t\in I^c} f_t^2 e_{jt}^r \sum_i c_i e_{it}^p l_i\eta_t + O_P(C_{NT}^{-2}),$$

where the last equality is due to: $\frac{1}{T}\sum_s(\frac{1}{T}\sum_t w_t\zeta_{ts})^2 = O_P(C_{NT}^{-4})$ and $\mathcal{Q}_3 = O_P(C_{NT}^{-2})$. Now for $p=1$,

$$\begin{aligned}
&\frac{1}{TN}\sum_{t\in I^c}\sum_i \widetilde{f}_t f_t e_{jt}^r c_i e_{it}(\widehat{e}_{it}-e_{it}) = O_P(C_{NT}^{-2}) + \frac{1}{TN}\sum_{t\in I^c}\sum_i f_t^2 e_{jt}^r c_i e_{it}(\widehat{e}_{it}-e_{it})\\
= &\frac{1}{TN}\sum_{t\in I^c} f_t\widehat{w}_t^2 e_{jt}^r \sum_i c_i e_{it}\frac{1}{T}\sum_s e_{is}(w_s+\eta_s) + \frac{1}{TN}\sum_{t\in I^c} f_t^2 e_{jt}^r \sum_i c_i e_{it} l_i\eta_t + O_P(C_{NT}^{-2})\\
= &O_P(C_{NT}^{-2})
\end{aligned}$$

where the last equality is due to $\frac{1}{T}\sum_t(\frac{1}{TN}\sum_{is}(w_s+\eta_s)e_{is}e_{it}c_i)^2 = O_P(C_{NT}^{-4})$; and $\frac{1}{T}\sum_t(\frac{1}{N}\sum_i e_{it}c_i)^2\eta_t^2 = O_P(C_{NT}^{-4})$.

For $p=0$ and $r=1$,

$$\begin{aligned}
&\frac{1}{TN}\sum_{t\in I^c}\sum_i f_t^2 e_{jt} c_i(\widehat{e}_{it}-e_{it})\\
= &O_P(1)\frac{1}{N}\sum_i c_i\frac{1}{T}\sum_s e_{is}(w_s+\eta_s) + O_P(1)\frac{1}{T}\sum_{t\in I^c} f_t^2 e_{jt}\eta_t + O_P(C_{NT}^{-2}),\\
= &O_P(C_{NT}^{-2}).
\end{aligned}$$

(iii)

$$\begin{aligned}
&\frac{1}{T}\sum_t(\frac{1}{N}\sum_i c_i(\widehat{e}_{it}-e_{it})e_{it})^2\\
\leq &\max_t\|\widehat{w}_t - H_x w_t\|^2\frac{1}{T}\sum_t[(\frac{1}{N}\sum_i c_i\frac{1}{T}\sum_s e_{is}w_s e_{it})^2 + \frac{1}{T}\sum_m(\frac{1}{N}\sum_i c_i e_{it}\frac{1}{T}\sum_{s\leq T} e_{is}\zeta_{ms})^2]\\
&+\frac{1}{T}\sum_t w_t^2[(\frac{1}{N}\sum_i c_i\frac{1}{T}\sum_s e_{is}w_s e_{it})^2 + \frac{1}{T}\sum_m(\frac{1}{N}\sum_i c_i e_{it}\frac{1}{T}\sum_{s\leq T} e_{is}\zeta_{ms})^2]\\
&+\frac{1}{T}\sum_t \widehat{w}_t^2(\frac{1}{N}\sum_i c_i\frac{1}{T}\sum_s e_{is}\eta_s e_{it})^2 + \frac{1}{T}\sum_t\widehat{w}_t^2(\frac{1}{N}\sum_i c_i l_i' e_{it})^2\mathcal{Q}_3^2\\
&+\frac{1}{T}\sum_t(\frac{1}{N}\sum_i c_i l_i e_{it})^2\frac{1}{T}\sum_s\zeta_{st}^2 + \frac{1}{T}\sum_t\eta_t^2(\frac{1}{N}\sum_i c_i l_i e_{it})^2 = O_P(C_{NT}^{-4})
\end{aligned}$$

Similarly, $\frac{1}{T}\sum_t(\frac{1}{N}\sum_j c_i(\widehat{e}_{it}-e_{it})u_{it})^2$, $\frac{1}{T}\sum_{t\in I^c}(|e_{jt}|+1)^2u_{jt}^2(\frac{1}{N}\sum_i c_i e_{it}(\widehat{e}_{it}-e_{it}))^2$ and

$\frac{1}{T}\sum_t e_{jt}^{2r} f_t^2(\frac{1}{N}\sum_i c_i e_{it}(\widehat{e}_{it}-e_{it}))^2$ are all $O_P(C_{NT}^{-4})$ (for fixed j and $r\leq 2$).

Next, to bound $\frac{1}{T}\sum_t(\frac{1}{N}\sum_i c_i(\widehat{e}_{it}-e_{it})^2)^2$, we first bound $\frac{1}{T}\sum_t \widehat{w}_t^4$. By (F.1),

$$\begin{aligned}
\frac{1}{T}\sum_t \widehat{w}_t^4 &\leq O_P(1)+\frac{1}{T}\sum_t\|\widehat{w}_t-H_x^{-1}w_t\|^4 \\
&\leq O_P(1)+\frac{1}{T}\sum_t\|\eta_t\|^4+\frac{1}{T}\sum_t\|\frac{1}{T}\sum_s w_s\zeta_{st}\|^4+\frac{1}{T}\sum_t(\frac{1}{T}\sum_s\zeta_{st}^2)^2 O_P(C_{NT}^{-4}) \\
&= O_P(1).
\end{aligned}$$

Thus

$$\begin{aligned}
&\frac{1}{T}\sum_{t\in I^c}(\frac{1}{N}\sum_i c_i(\widehat{e}_{it}-e_{it})^2)^2 \\
\leq\ & \frac{1}{T}\sum_{t\in I^c}\widehat{w}_t^4(\frac{1}{N}\sum_i c_i(\frac{1}{T}\sum_s e_{is}w_s)^2)^2+\frac{1}{T}\sum_{t\in I^c}\widehat{w}_t^4(\frac{1}{N}\sum_i c_i(\frac{1}{T}\sum_s e_{is}\eta_s)^2)^2 \\
&+\frac{1}{T}\sum_{t\in I^c}\widehat{w}_t^4\mathcal{Q}_3^4+\frac{1}{T}\sum_{t\in I^c}\eta_t^4+\frac{1}{T}\sum_{t\in I^c}(\frac{1}{T}\sum_s w_s\zeta_{st})^4+\frac{1}{T}\sum_{t\in I^c}(\frac{1}{T}\sum_s\zeta_{st}^2)^2 O_P(C_{NT}^{-4}) \\
&+\frac{1}{T}\sum_{t\in I^c}\widehat{w}_t^4(\frac{1}{N}\sum_i c_i\frac{1}{T}\sum_{m\leq T}(\frac{1}{T}\sum_{s\leq T}e_{is}\zeta_{ms})^2)^2=O_P(C_{NT}^{-4}).
\end{aligned}$$

F.2. **Behavior of the preliminary in the factor model.** Recall that

$$(\widetilde{f}_s,\widetilde{g}_s):=\arg\min_{f_s,g_s}\sum_{i=1}^N(y_{is}-\widetilde{\alpha}_i'g_s-x_{is}\widetilde{\lambda}_i'f_s)^2,\quad s\in I^c\cup\{t\}.$$

and

$$(\dot{\lambda}_i,\dot{\alpha}_i)=\arg\min_{\lambda_i,\alpha_i}\sum_{s\in I^c\cup\{t\}}(y_{is}-\alpha_i'\widetilde{g}_s-x_{is}\lambda_i'\widetilde{f}_s)^2,\quad i=1,...,N.$$

The goal of this section is still to show that the effect of the preliminary estimation is negligible. Specifically, we aim to show, for each fixed $t\in I^c$, fixed $i\leq N$,

$$\begin{aligned}
\frac{1}{\sqrt{N}}\sum_j(H_1'\lambda_j-\dot{\lambda}_j)e_{jt} &= O_P(\sqrt{N}C_{NT}^{-2}), \\
\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}}f_s(H_1^{-1}f_s-\widetilde{f}_s)e_{is} &= O_P(\sqrt{T}C_{NT}^{-2}).
\end{aligned}$$

Let L denote $N\times K_3$ matrix of l_i, so $X_s=Lw_s+e_s$. Also let W be $T\times K_3$ matrix of w_t.

Define

$$\widetilde{D}_{fs} = \frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)\widetilde{\Lambda}$$

$$
\begin{aligned}
D_{fs} &= \frac{1}{N}\Lambda'(\mathsf{diag}(X_s)M_\alpha\mathsf{diag}(X_s)\Lambda \\
\bar{D}_{fs} &= \frac{1}{N}\Lambda'\mathsf{E}((\mathsf{diag}(e_s)M_\alpha\mathsf{diag}(e_s))\Lambda+\frac{1}{N}\Lambda'(\mathsf{diag}(Lw_s)M_\alpha\mathsf{diag}(Lw_s)\Lambda \\
\widetilde{D}_{\lambda i} &= \frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}\mathsf{diag}(X_i)\widetilde{F} \\
D_{\lambda i} &= \frac{1}{T}F'(\mathsf{diag}(X_i)M_g\mathsf{diag}(X_i))F \\
\bar{D}_{\lambda i} &= \frac{1}{T}F'\mathsf{E}(\mathsf{diag}(E_i)M_g\mathsf{diag}(E_i))F+\frac{1}{T}F'(\mathsf{diag}(Wl_i)M_g\mathsf{diag}(Wl_i))F
\end{aligned}
$$

Lemma F.3. *Suppose* $\max_{it} e_{it}^2+\max_t\|w_t\|^2=o_P(C_{NT})$. *Also, there is* $c>0$, *so that*
$\min_s\min_j\psi_j(D_{fs})>c$. *Then*
(i) $\max_s\|\widetilde{D}_{fs}^{-1}\|=O_P(1)$.
(ii) $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|\widetilde{D}_{fs}^{-1}-(H_1'\bar{D}_{fs}H_1)^{-1}\|^2=O_P(C_{NT}^{-2})$.

Proof. The proof is mostly the same as that of Lemma E.1. The only difference is in the proof of (ii), where we need to show $\frac{1}{T}\sum_s\|D_{fs}-\bar{D}_{fs}\|^2=O_P(C_{NT}^{-2})$.

$$
\begin{aligned}
\frac{1}{T}\sum_s\|D_{fs}-\bar{D}_{fs}\|^2 &\leq \frac{1}{T}\sum_s\|\frac{1}{N}\sum_{ij}\lambda_i\lambda_j'M_{\alpha,ij}(x_{is}x_{js}-\mathsf{E}e_{is}e_{js}-l_i'w_sl_j'w_s)\|_F^2 \\
\leq\ & \frac{1}{T}\sum_s\|\frac{1}{N}\sum_{ij}\lambda_i\lambda_j'M_{\alpha,ij}(e_{is}e_{js}-\mathsf{E}e_{is}e_{js})\|_F^2 \\
&+\frac{2}{T}\sum_s\|\frac{1}{N}\sum_{ij}\lambda_i\lambda_j'M_{\alpha,ij}l_i'w_se_{js}\|_F^2
\end{aligned}
$$

We assume $\dim(\lambda_i)=\dim(p_i)=1$. As for the first term, it is less than

$$
\begin{aligned}
&\frac{1}{T}\sum_s\mathsf{Var}(\frac{1}{N}\sum_{ij}\lambda_i\lambda_j'M_{\alpha,ij}e_{is}e_{js}) \\
\leq\ &\frac{1}{T}\sum_s\frac{1}{N^4}\sum_{ijkl}|\mathsf{Cov}(e_{is}e_{js},e_{ks}e_{ls})|=O(N^{-1})
\end{aligned}
$$

provided that $\frac{1}{N^3}\sum_{ijkl}\mathsf{Cov}(e_{is}e_{js},e_{ks}e_{ls})<C$. As for the second term, it is less than

$$
O_P(1)\frac{1}{T}\sum_s\mathsf{E}w_s^2\frac{1}{N^2}\sum_{ij}|\mathsf{Cov}(e_{js},e_{ls}|w_s)|=O(N^{-1})
$$

provided that $\frac{1}{N}\sum_{ij}|\mathsf{Cov}(e_{is},e_{js}|w_s)|<\infty$ and $\|\mathsf{E}e_se_s'\|<\infty$. So $\frac{1}{T}\sum_s\|D_{fs}-\bar{D}_{fs}\|^2=O_P(N^{-1})$.

Lemma F.4. *(i) For each fixed* $t \in I^c$, $H_1^{-1} f_t - \widetilde{f}_t = O_P(C_{NT}^{-1})$.

(ii) For each fixed $i \leq N$, $\frac{1}{\sqrt{T}} \sum_{s \in I^c \cup \{t\}} f_s(H_1^{-1} f_s - \widetilde{f}_s) e_{is} = O_P(\sqrt{T} C_{NT}^{-2})$.

(iii) For each fixed $i \leq N$, $\frac{1}{\sqrt{T}} \sum_{s \in I^c \cup \{t\}} w_s f_s(H_1^{-1} f_s - \widetilde{f}_s) e_{is} = O_P(\sqrt{T} C_{NT}^{-2})$.

Proof. We highlight the similarity and main differences from the proof of Lemma E.2.

We have the same expansion.

$$\begin{aligned}
\widetilde{f}_s &= \widetilde{D}_{fs}^{-1} \frac{1}{N} \widetilde{\Lambda}' \mathsf{diag}(X_s) M_{\widetilde{\alpha}} y_s \\
&= H_1^{-1} f_s + \widetilde{D}_{fs}^{-1} \frac{1}{N} \widetilde{\Lambda}' \mathsf{diag}(X_s) M_{\widetilde{\alpha}} (A H_2 - \widetilde{A}) H_2^{-1} g_s \\
&\quad + \widetilde{D}_{fs}^{-1} \frac{1}{N} \widetilde{\Lambda}' \mathsf{diag}(X_s) M_{\widetilde{\alpha}} \mathsf{diag}(X_s)(\Lambda H_1 - \widetilde{\Lambda}) H_1^{-1} f_s \\
&\quad + \widetilde{D}_{fs}^{-1} \frac{1}{N} \widetilde{\Lambda}' \mathsf{diag}(X_s) M_{\widetilde{\alpha}} u_s. \qquad (F.4)
\end{aligned}$$

Step 1, the same proof yields

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \sum_s f_s e_{is} (\widetilde{D}_{fs}^{-1} - (H_1' D_x H_1)^{-1}) \frac{1}{N} \widetilde{\Lambda}' \mathsf{diag}(X_s) M_{\widetilde{\alpha}} (A H_2 - \widetilde{A}) H_2^{-1} g_s \\
\leq\;& O_P(\sqrt{T} C_{NT}^{-1}) [\frac{1}{N} \sum_j \widetilde{\lambda}_j^2 (\frac{1}{T} \sum_s f_s g_s e_{is} x_{js})^2]^{1/2} \\
\leq\;& O_P(\sqrt{T} C_{NT}^{-1}) [\frac{1}{N} \sum_j \widetilde{\lambda}_j^2 (\frac{1}{T} \sum_s f_s g_s e_{is} e_{js})^2]^{1/2} \\
&+ O_P(\sqrt{T} C_{NT}^{-1}) [\frac{1}{N} \sum_j \widetilde{\lambda}_j^2 l_j^2 (\frac{1}{T} \sum_s f_s g_s e_{is} w_s)^2]^{1/2} \\
\leq\;& O_P(\sqrt{T} C_{NT}^{-2}) + O_P(\sqrt{T} C_{NT}^{-1}) [\frac{1}{N} \sum_j \mathsf{E}_I (\frac{1}{T} \sum_s f_s g_s e_{is} e_{js})^2]^{1/2} \\
&+ O_P(\sqrt{T} C_{NT}^{-1}) [\frac{1}{N} \sum_j \mathsf{E}_I (\frac{1}{T} \sum_s f_s g_s e_{is} w_s)^2]^{1/2} \\
=\;& O_P(\sqrt{T} C_{NT}^{-2}).
\end{aligned}$$

Put together,

$$\frac{1}{\sqrt{T}} \sum_s f_s e_{is} \widetilde{D}_{fs}^{-1} \frac{1}{N} \widetilde{\Lambda}' \mathsf{diag}(X_s) M_{\widetilde{\alpha}} (A H_2 - \widetilde{A}) H_2^{-1} g_s = O_P(\sqrt{T} C_{NT}^{-2}).$$

Step 2: recall that $M_{\widetilde{\alpha},ij}$ and $P_{\widetilde{\alpha},ij}$ are the (i,j) th component of $M_{\widetilde{\alpha}}$ and $\widetilde{A}(\widetilde{A}'\widetilde{A})^{-1}\widetilde{A}'$ and write $P_{\widetilde{\alpha},ij} := \frac{1}{N}p_i'p_j$.

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_s f_s e_{is}(\widetilde{D}_{fs}^{-1} - (H_1'D_xH_1)^{-1})\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)(\Lambda H_1 - \widetilde{\Lambda})H_1^{-1}f_s \\
\leq\ & O_P(\sqrt{TN}C_{NT}^{-2})(\frac{1}{T}\sum_s\frac{1}{N^2}\sum_j\widetilde{\lambda}_j^2\mathsf{E}_I e_{js}^4 M_{\widetilde{\alpha},jj}^2\|f_s\|^2\|f_se_{is}\|^2)^{1/2} \\
&+O_P(\sqrt{TN}C_{NT}^{-2})(\frac{1}{T}\sum_s\frac{1}{N^2}\sum_j\widetilde{\lambda}_j^2 l_j^4 w_s^4 M_{\widetilde{\alpha},jj}^2\|f_s\|^2\|f_se_{is}\|^2)^{1/2} \\
&+O_P(\sqrt{T}C_{NT}^{-2})(\frac{1}{T}\sum_s\frac{1}{N^2}\sum_{k\neq j}x_{ks}^2(p_k'p_j)^2x_{js}^2\|f_s\|^2\|f_se_{is}\|^2)^{1/2} \\
=\ & O_P(\sqrt{T}C_{NT}^{-2}).
\end{aligned}
$$

Step 3: the same proof yields:

$$
\begin{aligned}
&(\frac{1}{\sqrt{T}}\sum_s f_s e_{is}(H_1'D_xH_1)^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)(M_{\widetilde{\alpha}} - M_\alpha)\mathsf{diag}(X_s)(\Lambda H_1 - \widetilde{\Lambda})H_1^{-1}f_s)^2 \\
\leq\ & O_P(TC_{NT}^{-4}).
\end{aligned}
$$

Step 4: let $z_{js} = \sum_k \widetilde{\lambda}_k M_{\alpha,kj}x_{ks}$,

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_s f_s e_{is}(H_1'D_xH_1)^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_\alpha\mathsf{diag}(X_s)(\Lambda H_1 - \widetilde{\Lambda})H_1^{-1}f_s \\
\leq\ & O_P(\sqrt{T}C_{NT}^{-1})[\frac{1}{N}\sum_j(\frac{1}{T}\sum_s f_s^2 e_{is}z_{js}x_{js})^2]^{1/2}
\end{aligned}
$$

To bound the last line, note that

$$
\begin{aligned}
&\frac{1}{N}\sum_j\mathsf{E}_I(\frac{1}{T}\sum_s f_s^2 e_{is}z_{js}x_{js})^2 \\
\leq\ & \frac{1}{N}\sum_j(\frac{1}{T}\sum_s\mathsf{E}_I(f_s^2e_{is}x_{js}^2)^2(\widetilde{\lambda}_jM_{\alpha,jj})^2 + O_P(1)\frac{1}{N}\sum_j\frac{1}{N}\sum_{k\neq j}(\frac{1}{T}\sum_s\mathsf{E}_I f_s^2x_{ks}e_{is}x_{js})^2 \\
&+\frac{1}{N}\sum_j\mathsf{Var}_I(\frac{1}{T}\sum_s f_s^2e_{is}z_{js}x_{js}) \\
\leq\ & \frac{1}{N}\sum_j(\frac{1}{T}\sum_s\mathsf{E}_I(f_s^2e_{is}x_{js}^2)^2(\widetilde{\lambda}_j - H_1'\lambda_j)^2 + O_P(1)\frac{1}{N}\sum_j(\frac{1}{T}\sum_s\mathsf{E}_I f_s^2e_{is}e_{js}^2)^2 \\
&+O_P(1)\frac{1}{N}\sum_j(\frac{1}{T}\sum_s\mathsf{E}_I f_s^2e_{is}l_jw_se_{js})^2 + O_P(1)\frac{1}{N}\sum_j\frac{1}{N}\sum_{k\neq j}(\frac{1}{T}\sum_s\mathsf{E}_I f_s^2e_{ks}e_{is}e_{js})^2 \\
&+O_P(1)\frac{1}{N}\sum_j\frac{1}{N}\sum_{k\neq j}(\frac{1}{T}\sum_s\mathsf{E}_I f_s^2e_{ks}e_{is}l_jw_s)^2
\end{aligned}
$$

$$+O_P(1)\frac{1}{N}\sum_j\frac{1}{N}\sum_{k\neq j}(\frac{1}{T}\sum_s\mathsf{E}_I f_s^2 l_k w_s e_{is}e_{js})^2+O_P(T^{-1})=O_P(C_{NT}^{-2})$$

given that $\sum_j|\mathsf{E}_I(e_{is}e_{js}|f_s,w_s)|+\frac{1}{N}\sum_{k\neq j}|\mathsf{E}_I(e_{ks}e_{is}e_{js}|f_s,w_s)|<\infty$. Put together,

$$\frac{1}{\sqrt{T}}\sum_s f_s e_{is}\widetilde{D}_{fs}^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)(\Lambda H_1-\widetilde{\Lambda})H_1^{-1}f_s=O_P(\sqrt{T}C_{NT}^{-2}).$$

Step 5:

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s e_{is}(\widetilde{D}_{fs}^{-1}-(H_1'D_xH_1)^{-1})\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}u_s\\
\leq\;& O_P(\sqrt{T}C_{NT}^{-1})(\frac{1}{T}\sum_{s\in I^c\cup\{t\}}|f_s e_{is}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}u_s|^2)^{1/2}\\
\leq\;& O_P(\sqrt{T}C_{NT}^{-2})(\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\mathsf{E}_I f_s^2e_{is}^2\frac{1}{N}\sum_j\widetilde{\lambda}_j^2x_{js}^2)^{1/2}\\
\leq\;& O_P(\sqrt{T}C_{NT}^{-2})(\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\frac{1}{N}\sum_j\widetilde{\lambda}_j^2\mathsf{E}_I f_s^2e_{is}^2e_{js}^2)^{1/2}\\
&+O_P(\sqrt{T}C_{NT}^{-2})(\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\mathsf{E}_I f_s^2e_{is}^2w_s^2)^{1/2}(\frac{1}{N}\sum_j\widetilde{\lambda}_j^2l_j^2)^{1/2}=O_P(\sqrt{T}C_{NT}^{-2}).
\end{aligned}$$

Step 6:

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s e_{is}(H_1'D_xH_1)^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)(M_{\widetilde{\alpha}}-M_\alpha)u_s\\
\leq\;& O_P(\sqrt{T}C_{NT}^{-1})(\frac{1}{N}\sum_j\|\frac{1}{T}\sum_{s\in I^c\cup\{t\}}u_sf_se_{is}\frac{1}{\sqrt{N}}x_{js}\|_F^2\widetilde{\lambda}_j^2)^{1/2}\\
\leq\;& O_P(\sqrt{T}C_{NT}^{-1})(\max_j\mathsf{E}_I\mathsf{E}_I(\|\frac{1}{T}\sum_{s\in I^c\cup\{t\}}u_sf_se_{is}\frac{1}{\sqrt{N}}x_{js}\|_F^2|X,E,F))^{1/2}\\
\leq\;& O_P(C_{NT}^{-1})(\max_{jks}\mathsf{Var}_I(u_{ks}f_se_{is}x_{js}))^{1/2}=O_P(C_{NT}^{-1})=O_P(\sqrt{T}C_{NT}^{-2}).
\end{aligned}$$

Step 7:

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s e_{is}(H_1'D_xH_1)^{-1}\frac{1}{N}(\widetilde{\Lambda}-\Lambda H_1)'\mathsf{diag}(X_s)M_\alpha u_s\\
\leq\;& O_P(C_{NT}^{-1})(\frac{1}{N}\sum_j\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\mathsf{E}_I f_s^2x_{js}^2e_{is}^2M_{\alpha,j}'\mathsf{Var}_I(u_s|e_s,w_s,f_s)M_{\alpha,j})^{1/2}\\
\leq\;& O_P(C_{NT}^{-1})(\frac{1}{N}\sum_j\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\mathsf{E}_I f_s^2x_{js}^2e_{is}^2\|M_{\alpha,j}\|^2)^{1/2}=O_P(C_{NT}^{-1})=O_P(\sqrt{T}C_{NT}^{-2}).
\end{aligned}$$

Finally,

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s e_{is}(H_1'D_xH_1)^{-1}\frac{1}{N}H_1'\Lambda'\mathsf{diag}(X_s)M_\alpha u_s \\
\leq\ & O_P(N^{-1})(\mathsf{E}_I f_s^2 e_{is}^2\Lambda'\mathsf{diag}(X_s)M_\alpha \mathsf{E}_I(u_su_s'|e_s,f_s)M_\alpha\mathsf{diag}(X_s)\Lambda)^{1/2} \\
\leq\ & O_P(N^{-1/2})(\frac{1}{N}\sum_j \mathsf{E}_I f_s^2e_{is}^2x_{js}^2)^{1/2} = O_P(N^{-1/2}) = O_P(\sqrt{T}C_{NT}^{-2}).
\end{aligned}
$$

Put together, we have $\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s e_{is}\widetilde{D}_{fs}^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}u_s = O_P(\sqrt{T}C_{NT}^{-2})$.

Thus $\frac{1}{\sqrt{T}}\sum_{s\in I^c\cup\{t\}} f_s(H_1^{-1}f_s-\widetilde{f}_s)e_{is} = O_P(\sqrt{T}C_{NT}^{-2})$. Q.E.D.

(iii) The proof is the same as that of (ii).

Lemma F.5. *Suppose* $\max_{it} e_{it}^4 = O_P(\min\{N,T\})$. *(i)* $\frac{1}{T}\|\widetilde{F}-FH_1^{-1'}\|_F^2 = O_P(C_{NT}^{-2}) = \frac{1}{T}\|\widetilde{G}-GH_2^{-1'}\|_F^2$, *and* $\frac{1}{T}\sum_{t\in I^c}\|\widetilde{f}_t - H_1^{-1}f_t\|^2 e_{it}^2u_{it}^2 = O_P(C_{NT}^{-2})$.
(ii) $\max_i\|\widetilde{D}_{\lambda i}^{-1}\| = O_P(1)$.
(iii) $\frac{1}{N}\sum_i\|\widetilde{D}_{\lambda i}^{-1} - (H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1}\|^2 = O_P(C_{NT}^{-2})$.

Proof. (i) The proof is very similar to that of Lemma E.1, hence we omit the details for (i)(ii). As for (iii), the same proof as in Lemma E.3 shows

$$
\begin{aligned}
&\frac{1}{N}\sum_i\|\widetilde{D}_{\lambda i}^{-1} - (H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1}\|^2 \\
\leq\ & O_P(C_{NT}^{-2}) + O_P(1)\frac{1}{N}\sum_i\|\frac{1}{T}\sum_{st} f_sf_t'M_{g,st}(x_{is}x_{it} - \mathsf{E}e_{is}e_{it} - l_i'w_sl_i'w_t)\|_F^2 \\
\leq\ & O_P(C_{NT}^{-2}) + O_P(1)\frac{1}{N}\sum_i\|\frac{1}{T}\sum_{st} f_sf_t'M_{g,st}(e_{is}e_{it} - \mathsf{E}e_{is}e_{it})\|_F^2 \\
&+O_P(1)\frac{1}{N}\sum_i\|\frac{1}{T}\sum_{st} f_sf_t'M_{g,st}w_se_{it}\|_F^2 \\
\leq\ & O_P(C_{NT}^{-2}) + O_P(1)\frac{1}{N}\sum_i\|\frac{1}{T}\sum_t f_t^2w_te_{it}\|_F^2 + O_P(1)\frac{1}{N}\sum_i\|\frac{1}{T}\sum_t f_tg_te_{it}\|_F^2 \\
=\ & O_P(C_{NT}^{-2}).
\end{aligned}
$$

Q.E.D.

Lemma F.6. *Suppose* $\mathsf{Var}(u_s|e_t,e_s,w_s) < C$ *and* $C_{NT}^{-1}\max_{is}|x_{is}|^2 = O_P(1)$.

(i) For each fixed $t\in I^c$, $\frac{1}{\sqrt{N}}\sum_i(H_1'\lambda_i - \dot{\lambda}_i)e_{it} = O_P(\sqrt{N}C_{NT}^{-2})$.

(ii) For each $i\leq N$, $\lambda_i - \dot{\lambda}_i = O_P(C_{NT}^{-1})$.

(iii) For each fixed $j\leq N$, $\frac{1}{T}\sum_{t\in I^c}\|w_te_{jt}f_t'\|^2\|\frac{1}{N}\sum_i\lambda_ie_{it}l_i(\dot{\lambda}_i - H_1'\lambda_i)'\|^2 = O_P(C_{NT}^{-4})$.

Proof. We have the same expansion

$$\begin{aligned}\dot{\lambda}_i &= \widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}y_i\\ &= H_1'\lambda_i+\widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}(GH_2^{-1'}-\widetilde{G})H_2'\alpha_i\\ &\quad+\widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}\mathsf{diag}(X_i)(FH_1^{-1'}-\widetilde{F})H_1'\lambda_i+\widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}u_i.\end{aligned}$$

We highlight the similarity and differences from the proof of Lemma E.4.

Step 1: given $\mathsf{E}e_{it}^4f_t^2+e_{it}^2f_t^2w_t^2<C$,

$$\begin{aligned}&\frac{1}{\sqrt{N}}\sum_i e_{it}(\widetilde{D}_{\lambda i}^{-1}-(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1})\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}(GH_2^{-1'}-\widetilde{G})H_2'\alpha_i\\ \leq\;& O_P(\sqrt{N}C_{NT}^{-2})+O_P(\sqrt{N}C_{NT}^{-2})(\frac{1}{T}\sum_s(\widetilde{f}_s-H_1^{-1}f_s)^2)^{1/2}\max_{is}|x_{is}|\\ \leq\;& O_P(\sqrt{N}C_{NT}^{-2})+O_P(\sqrt{N}C_{NT}^{-3})\max_{is}|x_{is}|=O_P(\sqrt{N}C_{NT}^{-2}).\end{aligned}$$

Step 2: $\bar{D}_{\lambda i}$ is nonrandom given W,G,F,

$$\begin{aligned}&\frac{1}{\sqrt{N}}\sum_i e_{it}(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}(GH_2^{-1'}-\widetilde{G})H_2'\alpha_i\\ &\leq O_P(\sqrt{N}C_{NT}^{-1})(a^{1/2}+b^{1/2})\quad\text{where}\\ a&=\frac{1}{T}\sum_s(\widetilde{f}_s-H_1^{-1}f_s)^2(\frac{1}{N}\sum_i\alpha_ie_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2\\ b&=\frac{1}{T}\sum_s f_s^2(\frac{1}{N}\sum_i\alpha_ie_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2.\end{aligned}$$

We now bound each term. As for b, note that for each fixed t,

$$\begin{aligned}&\mathsf{E}((\frac{1}{N}\sum_i\alpha_ie_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2|F,G,W,u_s)\\ \leq\;&\mathsf{E}(\frac{1}{N}\sum_i\alpha_ie_{it}\bar{D}_{\lambda i}^{-1}e_{is}|F,G,W,u_s)^2+\mathsf{E}(\frac{1}{N}\sum_i\alpha_ie_{it}\bar{D}_{\lambda i}^{-1}l_i'w_s|F,G,W,u_s)^2\\ \leq\;&\frac{C}{N}+\frac{C\|w_s\|^2}{N}+(\mathsf{E}\frac{1}{N}\sum_i\alpha_ie_{it}e_{is}\bar{D}_{\lambda i}^{-1}|F,G,W,u_s)^2 \qquad (F.5)\end{aligned}$$

with the assumption $\frac{1}{N}\sum_{ij}|\mathsf{Cov}(e_{jt}e_{js},e_{it}e_{is}|F,G,W,u_s)|<C$. So

$$\begin{aligned}\mathsf{E}b&\leq\frac{1}{T}\sum_s\mathsf{E}f_s^2(\frac{1}{N}\sum_i\alpha_ie_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2\leq O(C_{NT}^{-2}).\\ a&=\frac{1}{T}\sum_s f_s^2(\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)(\Lambda H_1-\widetilde{\Lambda}))^2(\frac{1}{N}\sum_i\alpha_ie_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{T}\sum_s(\widetilde{D}_{fs}^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}\mathsf{diag}(X_s)(\Lambda H_1-\widetilde{\Lambda})H_1^{-1}f_s)^2(\frac{1}{N}\sum_i\alpha_ie_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2\\
&+\frac{1}{T}\sum_s(\widetilde{D}_{fs}^{-1}\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}u_s)^2(\frac{1}{N}\sum_i\alpha_ie_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2\\
\leq\ & O_P(C_{NT}^{-2})\max_{is}x_{is}^4\frac{1}{T}\sum_s\mathsf{E}f_s^2(\frac{1}{N}\sum_i\alpha_ie_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2\\
&+O_P(1)\frac{1}{T}\sum_s(\frac{1}{N}\widetilde{\Lambda}'\mathsf{diag}(X_s)M_{\widetilde{\alpha}}u_s)^2(\frac{1}{N}\sum_i\alpha_ie_{it}\bar{D}_{\lambda i}^{-1}x_{is})^2\\
\leq\ & O_P(C_{NT}^{-2})
\end{aligned}
$$

where reaching the last inequality is similar to the proof in Lemma E.4, based on (F.5).

Put together, $a^{1/2}+b^{1/2}=O(C_{NT}^{-1})$. So the first term in the expansion of $\frac{1}{\sqrt{N}}\sum_i(H_1'\lambda_i-\dot{\lambda}_i)e_{it}$ is

$$
\frac{1}{\sqrt{N}}\sum_i e_{it}\widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}(GH_2^{-1'}-\widetilde{G})H_2'\alpha_i=O_P(\sqrt{N}C_{NT}^{-2}).
$$

Step 3: the same proof as in Lemma E.4 yields

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_i e_{it}(\widetilde{D}_{\lambda i}^{-1}-(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1})\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}\mathsf{diag}(X_i)(FH_1^{-1'}-\widetilde{F})H_1'\lambda_i\\
\leq\ & O_P(\sqrt{N}C_{NT}^{-2}).
\end{aligned}
$$

Step 4: similar to the proof in Lemma E.4

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_i e_{it}(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1'})^{-1}\frac{1}{T}(\widetilde{F}-FH_1^{-1'})'\mathsf{diag}(X_i)M_{\widetilde{g}}\mathsf{diag}(X_i)(FH_1^{-1'}-\widetilde{F})H_1'\lambda_i\\
\leq\ & O_P(\sqrt{N}C_{NT}^{-2})+\sqrt{N}\frac{1}{T}\sum_s(\widetilde{f}_s-H_1^{-1}f_s)^2\frac{1}{N}\sum_i\lambda_ie_{it}\bar{D}_{\lambda i}^{-1}x_{is}^2\\
\leq\ & O_P(\sqrt{N}C_{NT}^{-2})+\max_{is}x_{is}^2O_P(\sqrt{N}C_{NT}^{-2})\frac{1}{NT}\sum_s g_s^2w_s^2\sum_i\lambda_ie_{it}\bar{D}_{\lambda i}^{-1}l_i^2\\
&+\max_{is}x_{is}^2O_P(\sqrt{N}C_{NT}^{-2})\frac{1}{T}\sum_s w_s^2\frac{1}{N}\sum_j\lambda_j^2x_{js}^2f_s^2\frac{1}{N}\sum_i\lambda_ie_{it}\bar{D}_{\lambda i}^{-1}l_i^2\\
&+\sqrt{N}\frac{1}{T}\sum_s w_s^2(\frac{1}{N}\Lambda'\mathsf{diag}(X_s)M_\alpha u_s)^2\frac{1}{N}\sum_i\lambda_ie_{it}\bar{D}_{\lambda i}^{-1}l_i^2\\
\leq\ & O_P(\sqrt{N}C_{NT}^{-2}).
\end{aligned}
$$

Step 5: the same proof as in Lemma E.4 yields

$$\begin{aligned} &\frac{1}{\sqrt{N}}\sum_i e_{it}(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1\prime})^{-1}\frac{1}{T}F'\mathsf{diag}(X_i)(M_{\widetilde{g}}-M_g)\mathsf{diag}(X_i)(FH_1^{-1\prime}-\widetilde{F})H_1'\lambda_i \\ \leq\ & O_P(\sqrt{N}C_{NT}^{-2}), \end{aligned}$$

Step 6:

$$\begin{aligned} &\frac{1}{\sqrt{N}}\sum_i e_{it}(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1\prime})^{-1}\frac{1}{T}F'\mathsf{diag}(X_i)M_g\mathsf{diag}(X_i)(FH_1^{-1\prime}-\widetilde{F})H_1'\lambda_i \\ \leq\ & O_P(\sqrt{N}C_{NT}^{-1})(\frac{1}{T}\sum_s f_s^2(\frac{1}{N}\sum_i \lambda_i x_{is}^2 e_{it}\bar{D}_{\lambda i}^{-1})^2)^{1/2} \\ &+O_P(\sqrt{N}C_{NT}^{-1})(\frac{1}{T}\sum_s g_s^2(\frac{1}{N}\sum_i x_{is}\lambda_i e_{it}\bar{D}_{\lambda i}^{-1}\frac{1}{T}\sum_k f_k x_{ik} g_k)^2)^{1/2} \\ =\ & O_P(\sqrt{N}C_{NT}^{-1})(a^{1/2}+b^{1/2}). \end{aligned}$$

We aim to show $a = O_P(C_{NT}^{-2}) = b$.

$$\begin{aligned} \mathsf{E}a &:= \frac{1}{T}\sum_s \mathsf{E}f_s^2(\frac{1}{N}\sum_i \lambda_i x_{is}^2 e_{it}\bar{D}_{\lambda i}^{-1})^2 \\ &\leq \frac{1}{T}\sum_s \frac{1}{N^2}\sum_{ij}\mathsf{E}f_s^2\lambda_i\bar{D}_{\lambda i}^{-1}\lambda_j\bar{D}_{\lambda j}^{-1}\mathsf{E}(e_{is}^2e_{js}^2|F)\mathsf{Cov}(e_{it},e_{jt}|F,G,W) \\ &\quad+\frac{1}{T}\sum_s \mathsf{E}f_s^2\frac{1}{N^2}\sum_{ij}\lambda_i l_i^2 w_s^2\bar{D}_{\lambda i}^{-1}\lambda_j\mu_j^2\bar{D}_{\lambda j}^{-1}\mathsf{Cov}(e_{it},e_{jt}|F,G,W) \\ &\quad+\frac{2}{T}\sum_s \mathsf{E}f_s^2\frac{1}{N^2}\sum_{ij}\lambda_i l_i w_s\bar{D}_{\lambda i}^{-1}\bar{D}_{\lambda j}^{-1}\lambda_i l_j w_s\mathsf{E}(e_{js}e_{is}|F)\mathsf{Cov}(e_{jt},e_{it}|F,G,W) \\ &= O_P(N^{-1}). \\ \mathsf{E}b &:= \frac{1}{T}\sum_s \mathsf{E}g_s^2(\frac{1}{N}\sum_i x_{is}\lambda_i e_{it}\bar{D}_{\lambda i}^{-1}\frac{1}{T}\sum_k f_k x_{ik}g_k)^2 \\ &= O_P(C_{NT}^{-2}). \end{aligned}$$

Therefore the second term is

$$\frac{1}{\sqrt{N}}\sum_i e_{it}\widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}\mathsf{diag}(X_i)(FH_1^{-1\prime}-\widetilde{F})H_1'\lambda_i = O_P(\sqrt{N}C_{NT}^{-2}).$$

Step 7: $\frac{1}{\sqrt{N}}\sum_i e_{it}\widetilde{D}_{\lambda i}^{-1}\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}u_i$. The same proof yields

$$\frac{1}{\sqrt{N}}\sum_i e_{it}(\widetilde{D}_{\lambda i}^{-1}-(H_1^{-1}\bar{D}_{\lambda i}H_1^{-1\prime})^{-1})\frac{1}{T}\widetilde{F}'\mathsf{diag}(X_i)M_{\widetilde{g}}u_i \leq O_P(\sqrt{N}C_{NT}^{-2}).$$

$$\frac{1}{\sqrt{N}}\sum_i e_{it}\bar{D}_{\lambda i}^{-1}\frac{1}{T}(\widetilde{F}-FH_1^{-1'})'\mathsf{diag}(X_i)M_{\widetilde{g}}u_i \leq O_P(\sqrt{N}C_{NT}^{-2})$$
$$\frac{1}{\sqrt{N}}\sum_i e_{it}\bar{D}_{\lambda i}^{-1}\frac{1}{T}F'\mathsf{diag}(X_i)(M_{\widetilde{g}}-M_g)u_i \leq O_P(\sqrt{N}C_{NT}^{-2}).$$

Step 8:

$$\mathsf{E}(\frac{1}{\sqrt{N}}\sum_i e_{it}\bar{D}_{\lambda i}^{-1}\frac{1}{T}F'\mathsf{diag}(X_i)M_g u_i)^2 = O_P(C_{NT}^{-2}).$$

Together, $\frac{1}{\sqrt{N}}\sum_i(H_1'\lambda_i-\dot{\lambda}_i)e_{it} = O_P(\sqrt{T}C_{NT}^{-2})$. Q.E.D.

The proof of part (iii) follows from the same arguments as in part (i). While a rigorous proof still follows from substituting in the expansion of $\dot{\lambda}_i - H_1'\lambda_i$, the details would be mostly very similar. So we omit it for brevity.

F.3. **Technical lemmas for $\widehat{f}_t$ in the factor model.**

Lemma F.7. *For each fixed t, (i) $\widehat{B}_t - B = O_P(C_{NT}^{-1})$.*
(ii) The upper two blocks of $\widehat{B}_t^{-1}\widehat{S}_t - B^{-1}S$ are both $O_P(C_{NT}^{-2})$.

Proof. Throughout the proof, we assume $\dim(\alpha_i) = \dim(\lambda_i) = 1$ without loss of generality. We highlight the similarity and differences from the proofs of Lemma E.5.

(i) $\widehat{B}_t - B = b_1 + b_2$, where

$$\begin{aligned} b_1 &= \frac{1}{N}\sum_i \begin{pmatrix} \widetilde{\lambda}_i\widetilde{\lambda}_i'\widehat{e}_{it}^2 - H_1'\lambda_i\lambda_i'H_1 e_{it}^2 & \widetilde{\lambda}_i\widetilde{\alpha}_i'\widehat{e}_{it} - H_1'\lambda_i\alpha_i H_2' e_{it} \\ \widetilde{\alpha}_i\widetilde{\lambda}_i'\widehat{e}_{it} - H_2'\alpha_i\lambda_i H_1' e_{it} & \widetilde{\alpha}_i\widetilde{\alpha}_i' - H_2'\alpha_i\alpha_i' H_2 \end{pmatrix} \\ b_2 &= \frac{1}{N}\sum_i \begin{pmatrix} H_1'\lambda_i\lambda_i'H_1(e_{it}^2 - \mathsf{E}e_{it}^2) & H_1'\lambda_i\alpha_i H_2' e_{it} \\ H_2'\alpha_i\lambda_i H_1' e_{it} & 0 \end{pmatrix}. \end{aligned}$$

To prove the convergence of b_1, the same proof as in Lemma E.5 yields (by Lemma F.1),

$$\begin{aligned} &\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2 - e_{it}^2) \\ \leq\ & O_P(C_{NT}^{-2})\max_{it}|\widehat{e}_{it}-e_{it}|\max_{it}|e_{it}| + O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^4)^{1/2} \\ &+O_P(1)\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2 + O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2 e_{it}^2)^{1/2} \\ &+O_P(1)\frac{1}{N}\sum_i \lambda_i\lambda_i'(\widehat{e}_{it}-e_{it})e_{it} = O_P(C_{NT}^{-2}) \end{aligned}$$

In addition,

$$\begin{aligned}
&\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\alpha}_i'(\widehat{e}_{it}-e_{it}) \\
\leq\ & O_P(C_{NT}^{-2})\max_{it}|\widehat{e}_{it}-e_{it}| + O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2)^{1/2} \\
&+O_P(1)\frac{1}{N}\sum_i \lambda_i\alpha_i'(\widehat{e}_{it}-e_{it}) = O_P(C_{NT}^{-2})
\end{aligned}$$

So the same proof as in Lemma E.5 yields $b_1 = O_P(C_{NT}^{-1})$. In addition, $b_2 = O_P(N^{-1/2})$. Hence $\widehat{B}_t - B = O_P(C_{NT}^{-1})$.

(ii) We first bound the four blocks of $\widehat{S}_t - S$. We have $\widehat{S}_t - S = c_t + d_t$,

$$\begin{aligned}
c_t &= \frac{1}{N}\sum_i \begin{pmatrix} \widetilde{\lambda}_i\lambda_i' H_1 e_{it}(\widehat{e}_{it}-e_{it}) + \widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2 - e_{it}^2) & \widetilde{\lambda}_i(\widehat{e}_{it}-e_{it})(\alpha_i'H_2 - \widetilde{\alpha}_i') \\ \widetilde{\alpha}_i\lambda_i'H_1(e_{it}-\widehat{e}_{it}) + \widetilde{\alpha}_i(\lambda_i'H_1 - \widetilde{\lambda}_i')(\widehat{e}_{it}-e_{it}) & 0 \end{pmatrix} \\
d_t &= \frac{1}{N}\sum_i \begin{pmatrix} (\widetilde{\lambda}_i\lambda_i'H_1 - \widetilde{\lambda}_i\widetilde{\lambda}_i')e_{it}^2 - H_1'\lambda_i(\lambda_i'H_1 - \widetilde{\lambda}_i')\mathsf{E}e_{it}^2 & \widetilde{\lambda}_i e_{it}(\alpha_i'H_2 - \widetilde{\alpha}_i') \\ \widetilde{\alpha}_i(\lambda_i'H_1 - \widetilde{\lambda}_i')e_{it} & 0 \end{pmatrix}.
\end{aligned}$$

Call each block of c_t to be $c_{t,1}...c_{t,4}$ in the clockwise order. Note that $c_{t,4} = 0$.

As for $c_{t,1}$, it follows from Lemma F.1 that

$$\begin{aligned}
&\frac{1}{N}\sum_i \widetilde{\lambda}_i\lambda_i'H_1 e_{it}(\widehat{e}_{it}-e_{it}) \leq (\frac{1}{N}\sum_i e_{it}^2(\widehat{e}_{it}-e_{it})^2)^{1/2}O_P(C_{NT}^{-1}) \\
&+O_P(1)\frac{1}{N}\sum_i \lambda_i\lambda_i e_{it}(\widehat{e}_{it}-e_{it}) \leq O_P(C_{NT}^{-2}).
\end{aligned}$$

We have also shown $\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2 - e_{it}^2) = O_P(C_{NT}^{-2})$. Thus $c_{t,1} = O_P(C_{NT}^{-2})$.

For $c_{t,2}$, from Lemma F.1,

$$\begin{aligned}
&\frac{1}{N}\sum_i \widetilde{\lambda}_i(\widehat{e}_{it}-e_{it})(\alpha_i'H_2 - \widetilde{\alpha}_i') \\
\leq\ & O_P(C_{NT}^{-2})\max_i|e_{it}-\widehat{e}_{it}| + O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2)^{1/2} = O_P(C_{NT}^{-2}).
\end{aligned}$$

For the third term of c_t, similarly, $\frac{1}{N}\sum_i \widetilde{\alpha}_i(\lambda_i'H_1 - \widetilde{\lambda}_i')(\widehat{e}_{it}-e_{it}) = O_P(C_{NT}^{-2})$. Also, by Lemma F.1,

$$\begin{aligned}
&\frac{1}{N}\sum_i \widetilde{\alpha}_i\lambda_i'H_1(e_{it}-\widehat{e}_{it}) \\
\leq\ & O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2)^{1/2} + O_P(1)\frac{1}{N}\sum_i \lambda_i\alpha_i'(\widehat{e}_{it}-e_{it})
\end{aligned}$$

$$\leq \quad O_P(C_{NT}^{-1}).$$

So $c_{t,3} = O_P(C_{NT}^{-1})$. As for d_t, the same proof of Lemma E.5 shows that $d_t = O_P(C_{NT}^{-2})$.

Put together, we have: $\widehat{S}_t - S = c_t + d_t = O_P(C_{NT}^{-1})$. On the other hand, the upper two blocks of $\widehat{S}_t - S$ are $O_P(C_{NT}^{-2})$, determined by $c_{t,1}, c_{t,2}$ and the upper blocks of d_t. In addition, note that both B, S are block diagonal matrices, and the diagonal blocks of S are $O_P(C_{NT}^{-1})$. Due to

$$\widehat{B}_t^{-1}\widehat{S}_t - B^{-1}S \quad = \quad (\widehat{B}_t^{-1} - B^{-1})(\widehat{S}_t - S) + (\widehat{B}_t^{-1} - B^{-1})S + B^{-1}(\widehat{S}_t - S),$$

hence the upper blocks of $\widehat{B}_t^{-1}\widehat{S}_t - B^{-1}S$ are both $O_P(C_{NT}^{-2})$.

Q.E.D.

Lemma F.8. *Suppose* $C_{NT}^{-1}\max_{it}|e_{it}|^2 + \max_t \|\frac{1}{N}\sum_i \lambda_i\lambda_i'\bar{e}_i e_{it}\|_F = o_P(1)$.

(i) $\max_t \|\widehat{B}_t^{-1}\| = O_P(1)$.

(ii) $\frac{1}{T}\sum_{s\notin I}\|\widehat{B}_s^{-1} - B^{-1}\|^2 = O_P(C_{NT}^{-2})$,

(iii) Write

$$\widehat{S}_t - S = \begin{pmatrix} \Delta_{t1} \\ \Delta_{t2} \end{pmatrix},$$

whose partition matches with that of $(f_t', g_t')'$. *Then* $\frac{1}{T}\sum_{t\in I^c}\|\Delta_{t1}\|^2 = O_P(C_{NT}^{-4})$ *and* $\frac{1}{T}\sum_{t\in I^c}\|\Delta_{t2}\|^2 = O_P(C_{NT}^{-2})$.

Proof. Define

$$B_t = \frac{1}{N}\sum_i \begin{pmatrix} H_1'\lambda_i\lambda_i'H_1 e_{it}^2 & 0 \\ 0 & H_2'\alpha_i\alpha_i'H_2 \end{pmatrix}.$$

Then $\widehat{B}_t - B_t = b_{1t} + b_{2t}$,

$$b_{1t} \quad = \quad \frac{1}{N}\sum_i \begin{pmatrix} \widetilde{\lambda}_i\widetilde{\lambda}_i'\widehat{e}_{it}^2 - H_1'\lambda_i\lambda_i'H_1 e_{it}^2 & \widetilde{\lambda}_i\widetilde{\alpha}_i'\widehat{e}_{it} - H_1'\lambda_i\alpha_i H_2' e_{it} \\ \widetilde{\alpha}_i\widetilde{\lambda}_i'\widehat{e}_{it} - H_2'\alpha_i\lambda_i H_1' e_{it} & \widetilde{\alpha}_i\widetilde{\alpha}_i' - H_2'\alpha_i\alpha_i'H_2 \end{pmatrix}$$

$$b_{2t} \quad = \quad \frac{1}{N}\sum_i \begin{pmatrix} 0 & H_1'\lambda_i\alpha_i H_2' e_{it} \\ H_2'\alpha_i\lambda_i H_1' e_{it} & 0 \end{pmatrix}.$$

The proof is similar to that of Lemma E.6, with the main difference from dealing with terms involving $\widehat{e}_{it} - e_{it}$. Those are bounded by Lemma F.2.

(i) By assumption $\max_t |b_{2t}| = o_P(1)$. It suffices to show $\max_t |b_{1t}| = o_P(1)$. First term: it follows from Lemma F.2 that

$$
\begin{aligned}
& \max_t \|\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2 - e_{it}^2)\| \\
\leq\ & O_P(1)(\frac{1}{N}\sum_i \|\widetilde{\lambda}_i - H_1'\lambda_i\|^2)^{1/2}\max_t[\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^4 + (\widehat{e}_{it}-e_{it})^2e_{it}^2]^{1/2} \\
& +\frac{1}{N}\sum_i \|\widetilde{\lambda}_i - H_1'\lambda_i\|^2[2\max_t |(\widehat{e}_{it}-e_{it})e_{it}| + \max_t(\widehat{e}_{it}-e_{it})^2] \\
& +O_P(1)\max_t \|\frac{1}{N}\sum_i \lambda_i\lambda_i'(\widehat{e}_{it}-e_{it})e_{it}\|_F \\
& +O_P(1)\max_t \|\frac{1}{N}\sum_i \lambda_i\lambda_i'(\widehat{e}_{it}-e_{it})^2\|_F \\
\leq\ & O_P(1+\max_t \|w_t\|^2 + b_{NT,2}^2)C_{NT}^{-2} + O_P(b_{NT,1}^2 + b_{NT,3}^2) \\
& +O_P(1+\max_t \|w_t\| + b_{NT,2})\max_{it}|e_{it}|C_{NT}^{-1} + O_P(b_{NT,1}+b_{NT,3})(\max_t \frac{1}{N}\sum_i e_{it}^2)^{1/2}C_{NT}^{-1} \\
& +O_P(\phi_{NT}^2 C_{NT}^{-2}) + \phi_{NT}\max_{it}|e_{it}|C_{NT}^{-2} + \max_t \|\frac{1}{N}\sum_i c_i e_{it} l_i\|_F O_P(b_{NT,1}+b_{NT,3}) \\
=\ & o_P(1),
\end{aligned}
$$

given assumptions $C_{NT}^{-1}(b_{NT,2} + \max_t \|w_t\|)\max_{it}|e_{it}| = o_P(1)$, $\phi_{NT}\max_{it}|e_{it}| = O_P(1)$, $\max_t \|\frac{1}{N}\sum_i e_{it}\alpha_i\lambda_i'\|_F = o_P(1)$, and $(b_{NT,1}+b_{NT,3})[(\max_t \frac{1}{N}\sum_i e_{it}^2)^{1/2}C_{NT}^{-1} + 1] = o_P(1)$.

In addition, $\max_t \|\frac{1}{N}\sum_i(\widetilde{\lambda}_i\widetilde{\lambda}_i' - H_1'\lambda_i\lambda_i'H_1)e_{it}^2\| \leq O_P(C_{NT}^{-1})\max_{it} e_{it}^2 = o_P(1)$. So the first term of $\max_t |b_{1t}|$ is $o_P(1)$.

Second term,

$$
\begin{aligned}
& \max_t \|\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\alpha}_i'(\widehat{e}_{it}-e_{it})\|_F \\
\leq\ & \max_{it}|\widehat{e}_{it}-e_{it}|\frac{1}{N}\sum_i \|(\widetilde{\lambda}_i - H_1'\lambda_i)(\widetilde{\alpha}_i - H_2'\alpha_i)'\|_F \\
& +O_P(1)(\frac{1}{N}\sum_i \|\widetilde{\lambda}_i - H_1'\lambda_i\|^2 + (\frac{1}{N}\sum_i \|\widetilde{\alpha}_i - H_2'\alpha_i\|^2)^{1/2}\max_t(\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2)^{1/2} \\
& +O_P(1)\max_t \|\frac{1}{N}\sum_i \lambda_i\alpha_i'(\widehat{e}_{it}-e_{it})\|_F \\
\leq\ & O_P(\phi_{NT} + 1 + \max_t \|w_t\|)C_{NT}^{-2} + O_P(b_{NT,1} + b_{NT,3} + C_{NT}^{-1}b_{NT,2}) = o_P(1).
\end{aligned}
$$

The rest of the proof is the same as that of Lemma E.6.

(ii) It suffices to prove $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|b_{1t}\|_F^2=O_P(C_{NT}^{-2})=\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|b_{2t}\|_F^2$.
First term of b_{1t}: by Lemma F.2,

$$
\begin{aligned}
&\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\widetilde\lambda_i\widetilde\lambda_i'\widehat e_{it}^2-H_1'\lambda_i\lambda_i'H_1e_{it}^2\|_F^2\\
\leq\;& O_P(C_{NT}^{-4})\max_{it}|\widehat e_{it}-e_{it}|^2\max_{it}|e_{it}|^2+O_P(C_{NT}^{-2})\frac{1}{T}\sum_{t\in I^c}\frac{1}{N}\sum_i(\widehat e_{it}-e_{it})^4\\
&+O_P(1)\frac{1}{T}\sum_{t\in I^c}(\frac{1}{N}\sum_i(\widehat e_{it}-e_{it})^2)^2+O_P(C_{NT}^{-2})\frac{1}{T}\sum_{t\in I^c}\frac{1}{N}\sum_i(\widehat e_{it}-e_{it})^2e_{it}^2\\
&+O_P(1)\frac{1}{T}\sum_{t\in I^c}(\frac{1}{N}\sum_i\lambda_i\lambda_i'(\widehat e_{it}-e_{it})e_{it})^2+O_P(C_{NT}^{-2})\\
\leq\;& O_P(C_{NT}^{-4}\phi_{NT}^2)\max_{it}|e_{it}|^2+O_P(C_{NT}^{-2})=O_P(C_{NT}^{-2}).
\end{aligned}
$$

Second term of b_{1t}: still by Lemma F.2,

$$
\begin{aligned}
&\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\widetilde\lambda_i\widetilde\alpha_i'\widehat e_{it}-H_1'\lambda_i\alpha_iH_2'e_{it}\|_F^2\\
\leq\;& O_P(C_{NT}^{-4})\max_i\frac{1}{T}\sum_t(\widehat e_{it}-e_{it})^2+O_P(C_{NT}^{-2})\frac{1}{NT}\sum_{it}(\widehat e_{it}-e_{it})^2\\
&+\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\lambda_i\alpha_i'(\widehat e_{it}-e_{it})\|_F^2+O_P(C_{NT}^{-2})\\
\leq\;& O_P(b_{NT,4}^2+b_{NT,5}^2+C_{NT}^{-2})C_{NT}^{-4}+O_P(C_{NT}^{-2})=O_P(C_{NT}^{-2}).
\end{aligned}
$$

The rest of the proof for $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|b_{1t}\|_F^2=O_P(C_{NT}^{-2})=\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|b_{2t}\|_F^2$ is the same as that of Lemma E.6.

(iii) Note that $\widehat S_t-S=c_t+d_t$, where

$$
\begin{aligned}
c_t&=\frac{1}{N}\sum_i\begin{pmatrix}\widetilde\lambda_i\lambda_i'H_1e_{it}(\widehat e_{it}-e_{it})+\widetilde\lambda_i\widetilde\lambda_i'(\widehat e_{it}^2-e_{it}^2) & \widetilde\lambda_i(\widehat e_{it}-e_{it})(\alpha_i'H_2-\widetilde\alpha_i')\\ \widetilde\alpha_i\lambda_i'H_1(e_{it}-\widehat e_{it})+\widetilde\alpha_i(\lambda_i'H_1-\widetilde\lambda_i')(\widehat e_{it}-e_{it}) & 0\end{pmatrix}\\
d_t&=\frac{1}{N}\sum_i\begin{pmatrix}(\widetilde\lambda_i\lambda_i'H_1-\widetilde\lambda_i\widetilde\lambda_i')e_{it}^2-H_1'\lambda_i(\lambda_i'H_1-\widetilde\lambda_i')\mathsf{E}e_{it}^2 & \widetilde\lambda_ie_{it}(\alpha_i'H_2-\widetilde\alpha_i')\\ \widetilde\alpha_i(\lambda_i'H_1-\widetilde\lambda_i')e_{it} & 0\end{pmatrix}.
\end{aligned}
$$

As for the upper blocks of c_t, by Lemma F.2,

(ii) All terms below are $O_P(C_{NT}^{-2})$: $\frac{1}{NT}\sum_{it}(\widehat e_{it}-e_{it})^4$
$\frac{1}{NT}\sum_{it}(\widehat e_{it}-e_{it})^2$, $\frac{1}{NT}\sum_{it}e_{it}^2(\widehat e_{it}-e_{it})^2$ and $\frac{1}{T}\sum_t\|\frac{1}{N}\sum_i\alpha_i\lambda_i'(e_{it}-\widehat e_{it})\|_F^2$.
(iii) All terms below are $O_P(C_{NT}^{-4})$: $\frac{1}{T}\sum_t\|\frac{1}{N}\sum_i\lambda_i\lambda_i'(\widehat e_{it}-e_{it})e_{it}\|_F^2$,
$\frac{1}{T}\sum_t\|\frac{1}{N}\sum_i\lambda_i\lambda_i'(\widehat e_{it}-e_{it})^2\|_F^2$.

$$
\begin{aligned}
& \frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \widetilde{\lambda}_i\lambda_i' H_1 e_{it}(\widehat{e}_{it}-e_{it})\|_F^2 \\
\leq\; & O_P(1)\frac{1}{N}\sum_i \|\widetilde{\lambda}_i - H_1'\lambda_i\|^2 \frac{1}{NT}\sum_i\sum_{t\in I^c} e_{it}^2(\widehat{e}_{it}-e_{it})^2 \\
& +O_P(1)\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \lambda_i\lambda_i e_{it}(\widehat{e}_{it}-e_{it})\|_F^2 \\
=\; & O_P(C_{NT}^{-4}) \\
& \frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \widetilde{\lambda}_i\widetilde{\lambda}_i'(\widehat{e}_{it}^2-e_{it}^2)\|_F^2 \\
\leq\; & (\frac{1}{N}\sum_i \|\widetilde{\lambda}_i - H_1'\lambda_i\|^2)^2 \max_{it}[(\widehat{e}_{it}-e_{it})^4] \\
& +\frac{1}{T}\sum_{t\in I^c}(\frac{1}{N}\sum_i \|\widetilde{\lambda}_i - H_1'\lambda_i\|^2 |e_{it}|)^2 \max_{it}|\widehat{e}_{it}-e_{it}|^2 \\
& +\frac{1}{N}\sum_i \|\widetilde{\lambda}_i - H_1'\lambda_i\|^2 \frac{1}{NT}\sum_{t\in I^c}\sum_i [(\widehat{e}_{it}-e_{it})^4 + (\widehat{e}_{it}-e_{it})^2 e_{it}^2] \\
& +\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \lambda_i\lambda_i'(\widehat{e}_{it}-e_{it})e_{it}\|_F^2 + \frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \lambda_i\lambda_i'(\widehat{e}_{it}-e_{it})^2\|_F^2 \\
=\; & O_P(C_{NT}^{-4}).
\end{aligned}
$$

Also,

$$
\begin{aligned}
& \frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \widetilde{\lambda}_i(\widehat{e}_{it}-e_{it})(\alpha_i' H_2 - \widetilde{\alpha}_i')\|_F^2 \\
\leq\; & \frac{1}{N}\sum_i \|\alpha_i' H_2 - \widetilde{\alpha}_i'\|^2 \frac{1}{N}\sum_i \|\widetilde{\lambda}_i - H_1'\lambda_i\|^2 \max_{it}(\widehat{e}_{it}-e_{it})^2 \\
& +O_P(1)\frac{1}{N}\sum_i \|\alpha_i' H_2 - \widetilde{\alpha}_i'\|^2 \frac{1}{N}\sum_i \frac{1}{T}\sum_{t\in I^c}(\widehat{e}_{it}-e_{it})^2 = O_P(C_{NT}^{-4}).
\end{aligned}
$$

Similarly, $\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \widetilde{\alpha}_i(\lambda_i' H_1 - \widetilde{\lambda}_i')(\widehat{e}_{it}-e_{it})\|_F^2 = O_P(C_{NT}^{-4})$. Finally,

$$
\begin{aligned}
\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \widetilde{\alpha}_i\lambda_i' H_1(e_{it}-\widehat{e}_{it})\|_F^2 &= O_P(C_{NT}^{-4}) + O_P(1)\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \alpha_i\lambda_i'(e_{it}-\widehat{e}_{it})\|_F^2 \\
&= O_P(C_{NT}^{-2}).
\end{aligned}
$$

So if we let the two upper blocks of c_t be $c_{t,1}, c_{t,2}$, and the third block of c_t be $c_{t,3}$, then $\frac{1}{T}\sum_{t\in I^c}\|c_{t,1}\|_F^2 + \frac{1}{T}\sum_{t\in I^c}\|c_{t,2}\|_F^2 = O_P(C_{NT}^{-4})$ while $\frac{1}{T}\sum_{t\in I^c}\|c_{t,3}\|_F^2 = O_P(C_{NT}^{-2})$.

As for the blocks of d_t, let $d_{t,1},...,d_{t,3}$ denote the nonzero blocks, then the same proof as in Lemma E.6 shows $\frac{1}{T}\sum_{t\in I^c}\|d_{t,k}\|_F^2 = O_P(C_{NT}^{-4})$ for $k=1,2,3$.

Together, we have

$$\frac{1}{T}\sum_{t\in I^c}\|\Delta_{t1}\|^2 \leq \frac{1}{T}\sum_{t\in I^c}\|c_{t,1}+d_{t,1}\|^2 + \frac{1}{T}\sum_{t\notin I}\|c_{t,2}+d_{t,2}\|^2 = O_P(C_{NT}^{-4})$$

and $\frac{1}{T}\sum_{t\notin I}\|\Delta_{t2}\|^2 = \frac{1}{T}\sum_{t\notin I}\|c_{t,3}+d_{t,3}\|^2 = O_P(C_{NT}^{-2})$.

Q.E.D.

Lemma F.9. *For terms defined in (D.2), and for each fixed $t\in I^c$,*

(i) $\sum_{d=2}^5 A_{dt} = O_P(C_{NT}^{-2})$

(ii) For the "upper block" of A_{6t}, $\frac{1}{N}\sum_i \lambda_i e_{it}(l_i' w_t \lambda_i' f_t - \widehat{l_i' w_t}\dot{\lambda}_i'\widetilde{f}_t) = O_P(C_{NT}^{-2})$.

(iii) The upper block of A_{1t} is $O_P(C_{NT}^{-2})$.

Proof. Term A_{2t} is the same as that of Lemma E.7.

Term A_{3t}: based on the proof of Lemma E.7, the only difference is to bound the following term, which involves $\widehat{e}_{it}-e_{it}$:

$$\begin{aligned} & \frac{1}{N}\sum_i(\widetilde{\lambda}_i\widehat{e}_{it} - H_1'\lambda_i e_{it})u_{it} \\ \leq\ & O_P(C_{NT}^{-1})(\frac{1}{N}\sum_i(\widehat{e}_{it}-e_{it})^2)^{1/2} + O_P(C_{NT}^{-2}) + O_P(1)\frac{1}{N}\sum_i \lambda_i(\widehat{e}_{it}-e_{it})u_{it} \\ =\ & O_P(C_{NT}^{-2}) \end{aligned}$$

where the last equality follows from Lemma F.1.

Term A_{4t}. It suffices to prove each of the terms defined in (E.19), with $\mu_{it} = l_i' w_t$, is $O_P(C_{NT}^{-2})$. Given Lemma F.1, the proof follows from repeatedly applying the Cauchy-Schwarz inequality and is straightforward.

Term A_{5t}. Given $B_t^{-1} - B^{-1} = O_P(C_{NT}^{-1})$, it suffices to prove the following terms are $O_P(C_{NT}^{-2})$.

$$\begin{aligned} B_{1t} &= \frac{1}{N}\sum_i \lambda_i e_{it}(\widehat{e}_{it}-e_{it})(\dot{\lambda}_i - H_1'\lambda_i)'\widetilde{f}_t \\ B_{2t} &= \frac{1}{N}\sum_i \lambda_i e_{it}(\widehat{e}_{it}-e_{it})\lambda_i' H_1\widetilde{f}_t \\ B_{3t} &= \frac{1}{N}\sum_i \lambda_i e_{it} l_i' w_t(\dot{\lambda}_i - H_1'\lambda_i)'\widetilde{f}_t \\ B_{4t} &= \frac{1}{N}\sum_i \lambda_i e_{it} l_i' w_t \lambda_i' H_1(\widetilde{f}_t - H_1^{-1}f_t) \end{aligned}$$

$$B_{5t} \;=\; \frac{1}{N}\sum_i \alpha_i(\widehat{e}_{it}-e_{it})(\dot{\lambda}_i-H_1'\lambda_i)'\widetilde{f}_t \tag{F.6}$$

and that the following terms are $O_P(C_{NT}^{-1})$:

$$\begin{aligned} B_{6t} &= \frac{1}{N}\sum_i \alpha_i(\widehat{e}_{it}-e_{it})\lambda_i'H_1\widetilde{f}_t \\ B_{7t} &= \frac{1}{N}\sum_i \alpha_i l_i'w_t(\dot{\lambda}_i-H_1'\lambda_i)'\widetilde{f}_t \\ B_{8t} &= \frac{1}{N}\sum_i \alpha_i l_i'w_t\lambda_i'H_1(\widetilde{f}_t-H_1^{-1}f_t). \end{aligned} \tag{F.7}$$

In fact, $B_{1t}, B_{5,t} \sim B_{8t}$ follow immediately from the Cauchy-Schwarz inequality. $B_{2t} = O_P(C_{NT}^{-2})$ due to Lemma F.1. Term $B_{3,t}$ follows from Lemma F.6. Finally, $B_{4t} = O_P(C_{NT}^{-1})\frac{1}{N}\sum_i \lambda_i e_{it} l_i \lambda_i' = O_P(C_{NT}^{-2})$.

(ii) “upper block” of A_{6t}. Note that $l_i'w_t - \widehat{l_i'w_t} = \widehat{e}_{it} - e_{it}$.

From the proof of (i), we have $B_{dt} = O_P(C_{NT}^{-2})$, for $d = 1, ..., 4$. It follows immediately that $\frac{1}{N}\sum_i \lambda_i e_{it}(l_i'w_t\lambda_i'f_t - \widehat{l_i'w_t}\dot{\lambda}_i'\widetilde{f}_t) = O_P(C_{NT}^{-2})$.

(iii) Lastly, note that the upper block of A_{1t} is determined by the upper blocks of $\widehat{B}_t^{-1}\widehat{S}_t - B^{-1}S$, and are both $O_P(C_{NT}^{-2})$ by Lemma F.7.

Q.E.D.

Lemma F.10. *For terms B_{dt} defined in (F.6) (F.7), and C_{dt} defined in (E.19), with $\mu_{it} = l_i'w_t$, we have for a fixed $j \le N$,*

(i) $\frac{1}{T}\sum_{t\in I^c}\|B_{dt}\|^2 = O_P(C_{NT}^{-4})$ *for* $d = 1, 3, 4, 5$.

$\frac{1}{T}\sum_{t\in I^c}\|B_{dt}\|^2 = O_P(C_{NT}^{-2})$ *for* $d = 2, 6, 7, 8$.

(ii) $\sum_{d=1}^{5}\frac{1}{T}\sum_{t\in I^c}\|B_{dt}e_{jt}^r f_t'\| = O_P(C_{NT}^{-2})$ *for* $r = 0, 1, 2$.

$\sum_{d=1}^{5}\frac{1}{T}\sum_{t\in I^c}\|B_{dt}\||u_{jt}|(|e_{jt}|+1) + \sum_{d=6}^{8}\|\frac{1}{T_0}\sum_{t\in I^c}B_{dt}e_{jt}f_t\| = O_P(C_{NT}^{-2})$

(iii) $\sum_{d=6}^{8}\frac{1}{T}\sum_{t\in I^c}\|B_{dt}\|^2 e_{jt}^2\|f_t\|^2 + \sum_{d=1}^{8}\frac{1}{T}\sum_{t\in I^c}\|B_{dt}\|^2 e_{jt}^4 = O_P(C_{NT}^{-2})$,

$\sum_{d=1}^{8}\frac{1}{T}\sum_{t\in I^c}\|B_{dt}\|^2 u_{jt}^2(|e_{jt}|+1)^2 = O_P(C_{NT}^{-2})$, $\sum_{d=1}^{16}\frac{1}{T}\sum_{t\in I^c}\|C_{dt}\|^2 e_{jt}^4 = O_P(C_{NT}^{-2})$, *and* $\sum_{d=1}^{16}\frac{1}{T}\sum_{t\in I^c}\|C_{dt}\|^2 = O_P(C_{NT}^{-4})$.

Proof. (i) Note that $\frac{1}{T}\sum_{t\in I^c}\|B_{dt}\|^2 = O_P(C_{NT}^{-4})$ for $d = 1, 4, 5$, following from Cauchy Schwarz. Also, Cauchy-Schwarz and Lemma F.6 imply that $\frac{1}{T}\sum_{t\in I^c}\|B_{3t}\|^2 = O_P(C_{NT}^{-4})$.

Next, $\max_{it}|e_{it}(\widehat{e}_{it}-e_{it})| = O_P(1)$ by the assumption that $\phi_{NT}\max_{it}|e_{it}| = O_P(1)$.

$$\begin{aligned}\frac{1}{T}\sum_{t\in I^c}\|B_{2t}\|^2 &\leq O_P(1)\frac{1}{T}\sum_{t\in I^c}\|\widetilde{f}_t - H_1^{-1}f_t\|^2\max_{it}|e_{it}(\widehat{e}_{it}-e_{it})|^2\\ &+O_P(1)\frac{1}{T}\sum_{t\in I^c}\|f_t\|^2\|\frac{1}{N}\sum_i\lambda_i e_{it}(\widehat{e}_{it}-e_{it})\lambda_i'\|^2\\ &= O_P(C_{NT}^{-2}).\end{aligned}$$

Also, $\sum_{d=6}^{8}\frac{1}{T}\sum_{t\in I^c}\|B_{dt}\|^2 = O_P(C_{NT}^{-2})$ following from Cauchy-Schwarz.

(ii) By (i), $\frac{1}{T}\sum_{t\in I^c}\|B_{dt}e_{jt}^r f_t'\| + \frac{1}{T}\sum_{t\in I^c}\|B_{dt}\||u_{jt}|(|e_{jt}|+1) = O_P(C_{NT}^{-2})$ for $d=1,3,4,5$, using Cauchy Schwarz. In addition, it follows from Lemma F.2 that

$$\begin{aligned}\frac{1}{T}\sum_{t\in I^c}\|B_{2t}e_{jt}^r f_t'\| &\leq O_P(1)(\frac{1}{T}\sum_{t\in I^c}\|\frac{1}{N}\sum_i\lambda_i\lambda_i' e_{jt}^r e_{it}(\widehat{e}_{it}-e_{it})f_t\|^2)^{1/2} = O_P(C_{NT}^{-2})\\ \frac{1}{T}\sum_{t\in I^c}\|B_{2t}\||u_{jt}|(|e_{jt}|+1) &\leq O_P(1)(\frac{1}{T}\sum_{t\in I^c}(|e_{jt}|+1)^2u_{jt}^2\|\frac{1}{N}\sum_i\lambda_i\lambda_i' e_{it}(\widehat{e}_{it}-e_{it})\|^2)^{1/2}\\ &= O_P(C_{NT}^{-2}).\end{aligned}$$

For $d=6\sim 8$,

$$\begin{aligned}\frac{1}{T_0}\sum_{t\in I^c}f_t e_{jt}B_{6t} &= \frac{1}{T_0}\sum_{t\in I^c}f_t e_{jt}\frac{1}{N}\sum_i\alpha_i(\widehat{e}_{it}-e_{it})\lambda_i' H_1\widetilde{f}_t = O_P(C_{NT}^{-2})\\ &\quad(\text{ by Lemma F.2})\\ \frac{1}{T_0}\sum_{t\in I^c}f_t e_{jt}B_{7t} &= \frac{1}{T_0}\sum_{t\in I^c}f_t e_{jt}(\widetilde{f}_t - H_1^{-1}f_t)\frac{1}{N}\sum_i\alpha_i l_i' w_t(\dot{\lambda}_i - H_1'\lambda_i)'\\ &+(\frac{1}{T_0}\sum_{t\in I^c}f_t^2 e_{jt}w_t)(\frac{1}{N}\sum_i\alpha_i l_i(\dot{\lambda}_i - H_1'\lambda_i)) = O_P(C_{NT}^{-2})\\ \frac{1}{T_0}\sum_{t\in I^c}f_t e_{jt}B_{8t} &= \frac{1}{T_0}\sum_{t\in I^c}f_t e_{jt}(\widetilde{f}_t - H_1^{-1}f_t)w_t\frac{1}{N}\sum_i\alpha_i l_i\lambda_i' H_1\\ &= O_P(C_{NT}^{-2}),\quad \text{by lemma F.4.}\end{aligned}$$

(iii) follows from applying the Cauchy-Schwarz inequality.

Q.E.D.

Lemma F.11. *(i)* $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|\widehat{f}_s - H_f f_s\|^2(1+e_{jt}^4) = O_P(C_{NT}^{-2})$ *for a fixed* $j\leq N$. *(ii) For each fixed* i, $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}(\widehat{f}_s - H_f f_s)e_{is}f_s' = O_P(C_{NT}^{-2})$.

Proof. We only prove the harder part (ii). The proof is similar to that of Lemma E.9. By the proof of Lemma E.9, it suffices to prove $\sum_{d=1}^{5}\frac{1}{T}\sum_{t\in I^c}a_{dt}e_{it}f_t' = O_P(C_{NT}^{-2})$, where a_{dt} is the upper block of A_{dt}.

By Lemma F.8, $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}\|\widehat{B}_s^{-1}-B^{-1}\|^2=O_P(C_{NT}^{-2})$. So the proof of $\frac{1}{T}\sum_{t\in I^c}a_{2t}e_{it}f_t'\leq O_P(C_{NT}^{-2})$ follows from the same argument as that of Lemma E.9. Next, for B_{dt} defined in the proof of Lemma E.7, by Lemma F.10,

$$\begin{aligned}\frac{1}{T}\sum_{t\in I^c}a_{5t}e_{it}f_t' &\leq \sum_{d=1}^{8}\frac{1}{T}\sum_{t\in I^c}\|\widehat{B}_t^{-1}-B^{-1}\|\|B_{dt}e_{it}f_t'\| \\ &\leq \max_t\|\widehat{B}_t^{-1}-B^{-1}\|\sum_{d=1}^{5}\frac{1}{T}\sum_{t\in I^c}\|B_{dt}e_{it}f_t'\| \\ &\quad +(\frac{1}{T}\sum_t\|\widehat{B}_t^{-1}-B^{-1}\|^2)^{1/2}\sum_{d=6}^{8}(\frac{1}{T}\sum_t\|B_{dt}e_{it}f_t'\|^2)^{1/2} \\ &= O_P(C_{NT}^{-2}).\end{aligned}$$

Next, $\frac{1}{T}\sum_t\|\frac{1}{N}\sum_j\lambda_j(\widehat{e}_{jt}-e_{jt})u_{jt}\|^2=O_P(C_{NT}^{-4})$ by Lemma F.2,

$$\begin{aligned}\frac{1}{T}\sum_{t\in I^c}a_{3t}e_{it}f_t' &\leq O_P(1)(\frac{1}{T}\sum_{t\in I^c}\mathsf{E}_I\|\frac{1}{N}\sum_j(\widetilde{\lambda}_j-H_1'\lambda_j)e_{jt}u_{jt}\|^2)^{1/2} \\ &\quad +O_P(1)(\frac{1}{T}\sum_{t\in I^c}\mathsf{E}_I\|\frac{1}{N}\sum_j(\widetilde{\alpha}_j-H_2'\alpha_j)u_{jt}\|^2)^{1/2} \\ &\quad +O_P(1)\frac{1}{T}\sum_{t\in I^c}\|e_{it}f_t'\|\|\frac{1}{N}\sum_j(\widetilde{\lambda}_j-H_1'\lambda_j)(\widehat{e}_{jt}-e_{jt})u_{jt}\| \\ &\quad +O_P(1)(\frac{1}{T}\sum_t\|\frac{1}{N}\sum_j\lambda_j(\widehat{e}_{jt}-e_{jt})u_{jt}\|^2)^{1/2} \\ &= O_P(C_{NT}^{-2}) \\ \frac{1}{T}\sum_{t\in I^c}a_{4t}e_{it}f_t' &\leq \max_t\|\widehat{B}_t^{-1}\|\frac{1}{T}\sum_{t\in I^c}\|e_{it}f_t'\|\sum_{d=1}^{16}\|C_{dt}\|,\end{aligned}$$

where C_{dt}'s are defined in the proof of Lemma E.7. By Lemma F.10, $\frac{1}{T}\sum_{t\in I^c}\|C_{dt}\|^2=O_P(C_{NT}^{-4})$ for $d\leq 16$. Thus $\frac{1}{T}\sum_{t\in I^c}a_{4t}e_{it}f_t'=O_P(C_{NT}^{-2})$, still following from Cauchy-Schwarz.

Below we present the proof for $\frac{1}{T}\sum_{t\in I^c}a_{1t}e_{it}f_t'$, which is different from that of Lemma E.9. Note that

$$A_{1t}=(\widehat{B}_t^{-1}\widehat{S}_t-B^{-1}S)\begin{pmatrix}H_1^{-1}f_t\\H_2^{-1}g_t\end{pmatrix}:=(A_{1t,a}+A_{1t,b}+A_{1t,c})\begin{pmatrix}H_1^{-1}f_t\\H_2^{-1}g_t\end{pmatrix},$$

where

$$A_{1t,a} = (\widehat{B}_t^{-1}-B^{-1})(\widehat{S}_t-S)$$

$$\begin{aligned} A_{1t,b} &= (\widehat{B}_t^{-1} - B^{-1})S \\ A_{1t,c} &= B^{-1}(\widehat{S}_t - S). \end{aligned}$$

Let $(a_{1t,a}, a_{1t,b}, a_{1t,c})$ respectively be the upper blocks of $(A_{1t,a}, A_{1t,b}, A_{1t,c})$. Then

$$\frac{1}{T}\sum_{t\in I^c} a_{1t}e_{it}f_t' \leq \frac{1}{T}\sum_{t\in I^c}(\|a_{1t,a}\| + \|a_{1t,b}\| + \|a_{1t,c}\|)\|e_{it}f_t\|(\|f_t\| + \|g_t\|).$$

By the Cauchy-Schwarz and Lemma F.8, and B is a block diagonal matrix,

$$\begin{aligned} \frac{1}{T}\sum_{t\in I^c}\|a_{1t,b}\|\|e_{it}f_t\|(\|f_t\| + \|g_t\|) &\leq O_P(1)(\frac{1}{T}\sum_{t\in I^c}\|A_{1t,b}\|^2)^{1/2} \\ &\leq O_P(\|S\|)(\frac{1}{T}\sum_{t\in I^c}\|\widehat{B}_t - B\|^2)^{1/2} = O_P(C_{NT}^{-2}) \\ \frac{1}{T}\sum_{t\in I^c}\|a_{1t,c}\|\|e_{it}f_t\|(\|f_t\| + \|g_t\|) &\leq O_P(1)(\frac{1}{T}\sum_{t\in I^c}\|a_{1t,c}\|^2)^{1/2} \\ &\leq O_P(1)(\frac{1}{T}\sum_{t\in I^c}\|\Delta_{t1}\|^2)^{1/2} = O_P(C_{NT}^{-2}), \end{aligned}$$

where Δ_{t1} is defined in Lemma F.8, the upper block of $\widehat{S}_t - S$. The treatment of $\frac{1}{T}\sum_{t\in I^c}\|a_{1t,a}\|\|e_{it}f_t\|(\|f_t\| + \|g_t\|)$ is slightly different. Note that $\max_t\|\widehat{B}_t^{-1}\| + \|B^{-1}\| = O_P(1)$, shown in Lemma F.8. Partition

$$\widehat{S}_t - S = \begin{pmatrix} \Delta_{t1} \\ (\Delta_{t2,1} + \Delta_{t2,2}, 0) \end{pmatrix}$$

where the notation Δ_{t1} is defined in the proof of Lemma F.8. The proof of Lemma F.8 also gives

$$\begin{aligned} \Delta_{t2,1} &= \frac{1}{N}\sum_i \widetilde{\alpha}_i(\lambda_i'H_1 - \widetilde{\lambda}_i')e_{it} + \frac{1}{N}\sum_i \widetilde{\alpha}_i(\lambda_i'H_1 - \widetilde{\lambda}_i')(\widehat{e}_{it} - e_{it}) \\ &\quad + \frac{1}{N}\sum_i(\widetilde{\alpha}_i - H_2'\alpha_i)\lambda_i'H_1(e_{it} - \widehat{e}_{it}) \\ \Delta_{t2,2} &= H_2'\frac{1}{N}\sum_i \alpha_i\lambda_i'(e_{it} - \widehat{e}_{it})H_1. \end{aligned}$$

Therefore,

$$\begin{aligned} \|a_{1t,a}\| &\leq \|(\widehat{B}_t^{-1} - B^{-1})(\widehat{S}_t - S)\| \\ &\leq (\max_t\|\widehat{B}_t^{-1}\| + \|B\|^{-1})(\|\Delta_{t1}\| + \|\Delta_{t2,1}\|) + \|\widehat{B}_t^{-1} - B^{-1}\|\|\Delta_{t2,2}\|. \end{aligned}$$

Note that the above bound treats Δ_{t1} and Δ_{t2} differently because by the proof of Lemma F.8, $\frac{1}{T}\sum_{t\in I^c}\|\Delta_{t1}\|^2 = O_P(C_{NT}^{-4}) = \frac{1}{T}\sum_{t\in I^c}\|\Delta_{t2,1}\|^2$ but the rate of convergence for $\frac{1}{T}\sum_{t\in I^c}\|\Delta_{t2,2}\|^2$ is slower ($= O_P(C_{NT}^{-2})$). Hence

$$\begin{aligned}
&\frac{1}{T}\sum_{t\in I^c}\|a_{1t,a}\|\|e_{it}f_t\|(\|f_t\|+\|g_t\|) \leq O_P(1)(\frac{1}{T}\sum_{t\in I^c}\|\Delta_{t1}\|^2+\|\Delta_{t2,1}\|^2)^{1/2}\\
&+(\frac{1}{T}\sum_{t\in I^c}\|\widehat{B}_t^{-1}-B^{-1}\|^2)^{1/2}(\frac{1}{T}\sum_{t\in I^c}\|\Delta_{t2,2}\|^2\|\|e_{it}f_t\|^2(\|f_t\|+\|g_t\|)^2)^{1/2}\\
\leq^{(a)}\quad & O_P(C_{NT}^{-2})+O_P(C_{NT}^{-1})(\frac{1}{T}\sum_{t\in I^c}\|\Delta_{t2,2}\|^2\|\|e_{it}f_t\|^2(\|f_t\|+\|g_t\|)^2)^{1/2}\\
\leq^{(b)}\quad & O_P(C_{NT}^{-2}).
\end{aligned}$$

where (a) follows from the proof of Lemma F.8, while (b) follows from Lemma F.1. Thus

$$\frac{1}{T}\sum_{t\in I^c}a_{1t}e_{it}f_t' \leq O_P(C_{NT}^{-2}).$$

Therefore, $\frac{1}{T}\sum_{s\in I^c\cup\{t\}}(\widehat{f}_s - H_f f_s)e_{is}f_s' = O_P(C_{NT}^{-2})$.
Q.E.D.

Lemma F.12. *$\frac{1}{T_0}\sum_{t\in I^c}\widehat{f}_t\widehat{e}_{jt}^2B_{dt} = O_P(C_{NT}^{-2})$, for $d = 1...4$, $\frac{1}{T_0}\sum_{t\in I^c}\widehat{f}_t\widehat{e}_{jt}B_{dt} = O_P(C_{NT}^{-2})$ for $d = 5...8$.*

Proof. By Lemmas F.10 and F.11, for $r = 1, 2$ and $d \leq 8$,

$$\begin{aligned}
\frac{1}{T_0}\sum_{t\in I^c}\widehat{f}_t\widehat{e}_{jt}^rB_{dt} \;\leq\;& \max_{jt}|\widehat{e}_{jt}^r - e_{jt}^r|(\frac{1}{T_0}\sum_{t\in I^c}\|\widehat{f}_t - H_f f_t\|^2)^{1/2}(\frac{1}{T_0}\sum_{t\in I^c}\|B_{dt}\|^2)^{1/2}\\
&+(\frac{1}{T_0}\sum_{t\in I^c}\|\widehat{f}_t - H_f f_t\|^2e_{jt}^{2r})^{1/2}(\frac{1}{T_0}\sum_{t\in I^c}\|B_{dt}\|^2)^{1/2}\\
&+\frac{1}{T_0}\sum_{t\in I^c}f_te_{jt}^rB_{dt}+\frac{1}{T_0}\sum_{t\in I^c}\|f_tB_{dt}\||\widehat{e}_{jt}^r - e_{jt}^r|\\
\leq\;& O_P(C_{NT}^{-2})+\frac{1}{T_0}\sum_{t\in I^c}f_te_{jt}^rB_{dt}+\frac{1}{T_0}\sum_{t\in I^c}\|f_tB_{dt}\||\widehat{e}_{jt}^r - e_{jt}^r|.
\end{aligned}$$

Hence by Lemmas F.10, the above is further bounded by

$$\begin{aligned}
\sum_{d=1}^4\|\frac{1}{T_0}\sum_{t\in I^c}\widehat{f}_t\widehat{e}_{jt}^rB_{dt}\| \;\leq\;& O_P(C_{NT}^{-2})+\max_t|\widehat{e}_{jt}^r - e_{jt}^r|\sum_{d=1}^4\frac{1}{T_0}\sum_{t\in I^c}\|f_tB_{dt}\| = O_P(C_{NT}^{-2})\\
\sum_{d=5}^8\frac{1}{T_0}\sum_{t\in I^c}\widehat{f}_t\widehat{e}_{jt}B_{dt} \;\leq\;& O_P(C_{NT}^{-2})+\sum_{d=5}^8(\frac{1}{T_0}\sum_{t\in I^c}\|B_{dt}\|^2)^{1/2}(\frac{1}{T_0}\sum_{t\in I^c}\|f_t\|^2|\widehat{e}_{jt} - e_{jt}|^2)^{1/2}
\end{aligned}$$

$$\leq \quad O_P(C_{NT}^{-2}).$$

Q.E.D.

F.4. **Technical lemmas for $\widehat{\lambda}_i$.**

Lemma F.13. *For each fixed i, $\widehat{D}_i - D_i = O_P(C_{NT}^{-1})$.*

Proof. The proof is very similar to that of Lemma E.9. The only difference arises from bounding $\widehat{e}_{is} - e_{is}$ in the factor model. Examining the proof of Lemma E.9, in the current context, the proof still carries over by applying Cauchy Schwarz.

Q.E.D.

Lemma F.14. *For each fixed $i \leq N$,*

$$\frac{1}{T}\sum_{s\in I^c\cup\{t\}} \widehat{f}_s\widehat{e}_{is}(l_i'w_t\lambda_i'f_s - \widehat{l_i'w_t}\dot{\lambda}_i'\widetilde{f}_s) = O_P(C_{NT}^{-2}), \quad \frac{1}{T}\sum_{s\in I^c\cup\{t\}} \widehat{g}_s(l_i'w_t\lambda_i'f_s - \widehat{l_i'w_t}\dot{\lambda}_i'\widetilde{f}_s) = O_P(C_{NT}^{-1}).$$

Proof. It suffices to show (E.21) as in the proof of Lemma E.10, with $\mu_{is} = l_i'w_s$ and $\widehat{\mu}_{is} = \widehat{l_i'w_s}$. Using the Cauchy-Schwarz inequality, it is easy to show that $R_{3i,1}, R_{3i,2}$ and $R_{3i,4}$ are $O_P(C_{NT}^{-2})$ and $R_{3i,7} = O_P(C_{NT}^{-1})$. Next, by Lemma F.1, $R_{3i,3} = O_P(C_{NT}^{-2})$. In addition, by Lemmas F.4 and F.11, $R_{3i,5} = O_P(C_{NT}^{-2})$ and $R_{3i,6} = O_P(C_{NT}^{-2})$.

This concludes that $R_{3i,d} = O_P(C_{NT}^{-2})$ for all $d \leq 7$ defined in Lemma E.10, and completes the proof. Q.E.D.

Lemma F.15. *For R_{di} defined in (D.5), and for each fixed $j \leq N$, $R_{dj} = O_P(C_{NT}^{-2})$ for $d = 1, ..., 3$. The upper blocks of $R_{4j} \sim R_{6j}$ are $O_P(C_{NT}^{-2})$.*

Proof. (i) It is easy to see that $R_{1j} = O_P(C_{NT}^{-2})$. Lemma F.1 implies $R_{2j} = O_P(C_{NT}^{-2})$. Also, it follows from Lemmas F.13, F.14 that both R_{3j} and the upper block of R_{4j} are $O_P(C_{NT}^{-2})$.

(ii) We now show the upper blocks of R_{5j} are $O_P(C_{NT}^{-2})$. Similar to the proof of Lemma E.12,

$$R_{5j} \quad = \quad \widehat{D}_j^{-1}\frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}\begin{pmatrix}\widehat{f}_t - H_f f_t\\ \widehat{g}_t - H_g g_t\end{pmatrix}u_{jt} \leq \sum_{d=1}^{5}\Upsilon_d + O_P(C_{NT}^{-2}),$$

where, for A_{dt} defined in (D.2), with $\mu_{it} = l_i'w_t$ and $\widehat{\mu}_{it} = \widehat{l_i'w_t}$,

$$\Upsilon_d = \widehat{D}_j^{-1}\frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}A_{dt}u_{jt}.$$

We now show $\Upsilon_d = O_P(C_{NT}^{-2})$ for $d = 1...5$.

The proof for terms Υ_1 and Υ_2 is the same as in Lemma E.12. Next,

$$\Upsilon_3 \leq O_P(C_{NT}^{-2}) + (\frac{1}{T_0}\sum_{t\in I^c}\|\frac{1}{N}\sum_i \widetilde{\lambda}_i(\widehat{e}_{it} - e_{it})u_{it}\|^2)^{1/2}$$

where the last equality follows from Lemma F.2 and the Cauchy Schwarz.

$$\Upsilon_4 \leq O_P(C_{NT}^{-2}) + O_P(1)(\sum_{d=1}^{16}\frac{1}{T_0}\sum_{t\in I^c}\|C_{dt}\|^2)^{1/2} = O_P(C_{NT}^{-2})$$

where the last equality is due to Lemma F.10.

$$\begin{aligned}
\Upsilon_5 &\leq O_P(C_{NT}^{-2}) + O_P(1)\sum_{d=1}^{8}\frac{1}{T_0}\sum_{t\in I^c}\|\widehat{B}_t - B_t\|\|B_{dt}\||u_{jt}|(|e_{jt}|+1) \\
&\leq O_P(C_{NT}^{-2}) + O_P(1)\sum_{d=1}^{5}\frac{1}{T_0}\sum_{t\in I^c}\|B_{dt}\||u_{jt}|(|e_{jt}|+1) \\
&\quad + O_P(C_{NT}^{-1})\sum_{d=6}^{8}(\frac{1}{T_0}\sum_{t\in I^c}\|B_{dt}\|^2u_{jt}^2(|e_{jt}|+1)^2)^{1/2} = O_P(C_{NT}^{-2}),
\end{aligned}$$

where the last equality follows from Lemma F.10.

Finally, the upper block of Υ_6 is bounded by, still by Lemma F.10,

$$O_P(1)\sum_{d=1}^{4}\|\frac{1}{T_0}\sum_{t\in I^c}\widehat{e}_{jt}u_{jt}B_{dt}\| + O_P(C_{NT}^{-2}).$$

Therefore, the upper block of R_{5j} is $O_P(C_{NT}^{-2})$.

(iii) We now show the upper block of R_{6j} is $O_P(C_{NT}^{-2})$. Let

$$\begin{aligned}
\Gamma &:= \Gamma_0 + \Gamma_1 + ... + \Gamma_6, \\
\Gamma_0 &:= \frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{f}_t\widehat{e}_{jt}\\ \widehat{g}_t\end{pmatrix}(\lambda_j'H_f^{-1}, \alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}B^{-1}\frac{1}{N}\sum_i\begin{pmatrix}H_1'\lambda_i e_{it}\\ H_2'\alpha_i\end{pmatrix}u_{it} \\
\Gamma_d &:= \sum_{d=1}^{6}\frac{1}{T_0}\sum_{t\in I^c}\begin{pmatrix}\widehat{f}_t\widehat{e}_{jt}\\ \widehat{g}_t\end{pmatrix}(\lambda_j'H_f^{-1}, \alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}A_{dt}, \quad d = 1, ..., 6.
\end{aligned}$$

Then $R_{6j} = \widehat{D}_j^{-1}\Gamma$. Similar to the proof of Lemma E.12, we aim to show that $\Gamma_0, \Gamma_2 ... \Gamma_5$ are each $O_P(C_{NT}^{-2})$. In addition, the upper blocks of Γ_1 and Γ_6 are $O_P(C_{NT}^{-2})$, while their lower blocks are $O_P(C_{NT}^{-1})$.

First, $\Gamma_0 + \Gamma_2 = O_P(C_{NT}^{-2})$ follows from the same proof as in Lemma E.12.

Next, the upper block of Γ_1 is

$$\begin{aligned}
&\frac{1}{T_0}\sum_{t\in I^c}\widehat{f}_t\widehat{e}_{jt}(\lambda_j'H_f^{-1},\alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}(\widehat{B}_t^{-1}\widehat{S}_t-B^{-1}S)\begin{pmatrix}H_1^{-1}f_t\\ H_2^{-1}g_t\end{pmatrix}\\
\leq\ & a1+a2+O_P(C_{NT}^{-2})\\
a1\ =\ & \frac{1}{T_0}\sum_{t\in I^c}f_te_{jt}(\lambda_j'H_f^{-1}e_{jt},\alpha_j'H_g^{-1})(\widehat{B}_t^{-1}-B^{-1})(\widehat{S}_t-S)\begin{pmatrix}H_1^{-1}f_t\\ H_2^{-1}g_t\end{pmatrix}\\
a2\ =\ & \frac{1}{T_0}\sum_{t\in I^c}f_te_{jt}(\lambda_j'H_f^{-1}e_{jt},\alpha_j'H_g^{-1})B^{-1}(\widehat{S}_t-S)\begin{pmatrix}H_1^{-1}f_t\\ H_2^{-1}g_t\end{pmatrix}.
\end{aligned}$$

By the proof of lemma F.11, we have

$$\widehat{S}_t-S=\begin{pmatrix}\Delta_{t1}\\ (\Delta_{t2,1}+\Delta_{t2,2},0)\end{pmatrix},\quad \Delta_{t2,2}=H_2'\frac{1}{N}\sum_i\alpha_i\lambda_i'(e_{it}-\widehat{e}_{it})H_1.$$

As for $a1$, by the proof of Lemma F.8, $\frac{1}{T}\sum_{t\in I^c}\|\Delta_{t1}\|^2=O_P(C_{NT}^{-4})=\frac{1}{T}\sum_{t\in I^c}\|\Delta_{t2,1}\|^2$,

$$\begin{aligned}
a1\ \leq\ & O_P(1)(\max_t\|\widehat{B}_t^{-1}\|+\|B^{-1}\|)(\frac{1}{T}\sum_t\|\Delta_{t1}\|^2+\|\Delta_{t2,1}\|^2)^{1/2}\\
&+O_P(1)(\frac{1}{T}\sum_t\|\widehat{B}_t^{-1}-B^{-1}\|^2)^{1/2}(\frac{1}{T}\sum_t\|\Delta_{t2,2}\|^2\|f_t\|^2(\|f_t\|+\|g_t\|)^2|e_{jt}|^4)^{1/2}\\
=\ & O_P(C_{NT}^{-2})
\end{aligned}$$

where the last equality uses Lemma F.8 that $\max_t\|\widehat{B}_t^{-1}\|=O_P(1)$, Lemma F.1 that $\frac{1}{T}\sum_t\|\Delta_{t2,2}\|^2\|f_t\|^2(\|f_t\|+\|g_t\|)^2|e_{jt}|^4=O_P(C_{NT}^{-2})$ and $\frac{1}{T}\sum_{s\notin I}\|\widehat{B}_s^{-1}-B^{-1}\|^2=O_P(C_{NT}^{-2})$.

As for $a2$, still by $\frac{1}{T}\sum_{t\in I^c}\|\Delta_{t1}\|^2=O_P(C_{NT}^{-4})=\frac{1}{T}\sum_{t\in I^c}\|\Delta_{t2,1}\|^2$, and by Lemma F.2,

$$\begin{aligned}
a2\ \leq\ & O_P(1)(\frac{1}{T}\sum_t\|\Delta_{t1}\|^2+\|\Delta_{t2,1}\|^2)^{1/2}+O_P(1)\frac{1}{T_0}\sum_{t\in I^c}f_te_{jt}\Delta_{t2,2}f_t\\
\leq\ & O_P(C_{NT}^{-2})+O_P(1)\frac{1}{TN}\sum_{it}f_te_{jt}f_t\alpha_i\lambda_i'(e_{it}-\widehat{e}_{it})=O_P(C_{NT}^{-2}).
\end{aligned}$$

Together, the upper block of Γ_1 is $O_P(C_{NT}^{-2})$.

The lower block of Γ_1 is

$$\frac{1}{T_0}\sum_{t\in I^c}\widehat{g}_t(\lambda_j'H_f^{-1},\alpha_j'H_g^{-1})\begin{pmatrix}\widehat{e}_{jt}\mathsf{I} & 0\\ 0 & \mathsf{I}\end{pmatrix}(\widehat{B}_t^{-1}\widehat{S}_t-B^{-1}S)\begin{pmatrix}H_1^{-1}f_t\\ H_2^{-1}g_t\end{pmatrix}$$

$$\leq \quad O_P(C_{NT}^{-1})(\frac{1}{T_0}\sum_{t\in I^c}|\widehat{g}_t\|^2(\widehat{e}_{jt}^2+1)(\|f_t\|+\|g_t\|)^2)^{1/2}=O_P(C_{NT}^{-1}).$$

Next, $\Gamma_3 = O_P(C_{NT}^{-2})$ follows from the same proof as in Lemma E.12. The only difference is to bound $\frac{1}{T}\sum_t|\frac{1}{N}\sum_i \lambda_i(\widehat{e}_{it}-e_{it})u_{it}|^2 = O_P(C_{NT}^{-4})$ by Lemma F.2.

Next, $\Gamma_4+\Gamma_5 = O_P(C_{NT}^{-2})$ follows from the same proof as in Lemma E.12, whose proof in the current context uses Lemma F.10 that $\sum_{d=1}^{16}\frac{1}{T_0}\sum_{t\in I^c}|e_{jt}|^4\|C_{dt}\|^2 = O_P(C_{NT}^{-2})$, $\sum_{d=1}^{16}\frac{1}{T_0}\sum_{t\in I^c}\|C_{dt}\|^2 = O_P(C_{NT}^{-4})$, and $\sum_{d=1}^{8}[\frac{1}{T_0}\sum_{t\in I^c}(|e_{jt}|^2+1)^2\|B_{dt}\|^2 = O_P(C_{NT}^{-2})$.

Finally, the upper block of Γ_6 is

$$\frac{1}{T_0}\sum_{t\in I^c}\widehat{f}_t\widehat{e}_{jt}(\widehat{e}_{jt}\lambda_j'H_f^{-1},\alpha_j'H_g^{-1})B^{-1}\begin{pmatrix}H_1' & 0\\ 0 & H_2\end{pmatrix}\begin{pmatrix}\sum_{d=1}^4 B_{dt}\\ \sum_{d=5}^8 B_{dt}\end{pmatrix}.$$

Now Lemma F.12 implies $\frac{1}{T_0}\sum_{t\in I^c}\widehat{f}_t\widehat{e}_{jt}^2B_{dt} = O_P(C_{NT}^{-2})$ for $d=1\sim 4$, and $\frac{1}{T_0}\sum_{t\in I^c}\widehat{f}_t\widehat{e}_{jt}B_{dt} = O_P(C_{NT}^{-2})$ for $d=5\sim 8$. So the above is $O_P(C_{NT}^{-2})$.

The lower block of Γ_6 is, by repeatedly using Cauchy-Schwarz,

$$\sum_{d=1}^{4}\frac{1}{T_0}\sum_{t\in I^c}\widehat{g}_t\widehat{e}_{jt}B_{dt}+\sum_{d=5}^{8}\frac{1}{T_0}\sum_{t\in I^c}\widehat{g}_tB_{dt} = O_P(C_{NT}^{-1}).$$

Q.E.D.

Department of Economics, MIT, Cambridge, MA 02139
E-mail address: vchern@mit.edu

Booth School of Business, University of Chicago, Chicago, IL 60637
E-mail address: Christian.Hansen@chicagobooth.edu

Department of Economics, Rutgers University, New Brunswick, NJ 08901
E-mail address: yuan.liao@rutgers.edu

Lundquist College of Business, Oregon State University, 1208 University St, Eugene, OR 97403
E-mail address: yzhu6@uoregon.edu