Inference for Large Dimensional Factor Models under General Missing Data Patterns*

Liangjun Su† Fa Wang‡

February 1, 2024

Abstract

This paper establishes the inferential theory for the least squares estimation of large factor models with missing data. We propose a unified framework for asymptotic analysis of factor models that covers a wide range of missing patterns, including heterogenous random missing, selection on covariates/factors/loadings, block/staggered missing, mixed frequency and ragged edge. We establish the average convergence rates of the estimated factor space and loading space, the limit distributions of the estimated factors and loadings, as well as the limit distributions of the estimated average treatment effects and the parameter estimates in the factor-augmented regressions. These results allow us to impute the unbalanced panel appropriately or make inference for the heterogenous treatment effects. For computation, we can use the nuclear norm regularized estimator as the initial value for the EM algorithm and iterate until convergence. Empirically, we apply our method to test the average treatment effects of partisan alignment on grant allocation in UK.

JEL Classification: C13, C33, C38, C55

Keywords: Factor Models, Missing Data, EM Algorithm, Least Squares, Matrix Completion, Nuclear Norm, Causal Inference, Mixed Frequency

*Su gratefully acknowledges the support from the National Natural Science Foundation of China under Grant No. 72133002.

†School of Economics and Management, Tsinghua University, Beijing, China

‡Department of Finance, School of Economics, Peking University, Beijing, China

1 Introduction

The missing data issue in large factor models has gained considerable interest recently in various strands of the literature, including matrix completion, causal inference and factor analysis with unbalanced panel. See Banbura and Modugno (2014), Athey, Bayati, Doudchenko, Imbens and Khosravi (2021), Bai and Ng (2021), Jin, Miao and Su (2021), Chernozhukov, Hansen, Liao and Zhu (2023), Xiong and Pelger (2023), among others. The common objective of this literature is to estimate the factors and loadings consistently and establish the corresponding asymptotic properties. The difference is that depending on the targeted applications, the literature consider different missing patterns and different estimation algorithms. Although these methods are practically useful, the connections between them are unclear. Moreover, these methods are all closely related to the least squares estimation using the observed data, but the asymptotic theory for the least squares estimation itself is still unclear.

This paper studies the least squares estimation and develops a unified inferential theory for factor models with missing data that handles completely random heterogenous missing, selection on covariates/factors/loadings, block/staggered missing, mixed frequency, ragged edge and some other patterns in a single framework. These missing patterns include all patterns studied in the above papers and also some new patterns. We establish the average convergence rates of the estimated factor space and loading space, the limit distributions of the estimated factors and loadings, the limit distributions of linear combinations of the elements of the low rank matrix, and the limit distributions of the parameter estimates in the factor-augmented regressions, as N (the number of units) and T (the number of time periods) tend to infinity jointly. These results generalize those in Bai (2003) and Bai and Ng (2006) to factor models under general missing data patterns.

In terms of estimation, since principal component analysis (PCA) is no longer applicable under random missing, we propose to use the EM algorithm with the nuclear norm regularized estimator (NN) as the initial value and coin our estimator as a nuclear norm EM estimator (NN-EM). Since the EM algorithm is guaranteed to converge to local maximum (or minimum) and the nuclear norm regularized estimator is consistent and easy to compute, the proposed NN-EM algorithm is guaranteed to converge to the least squares solution. The EM algorithm is always a popular choice to handle missing data for factor models since Stock and Watson (2002). Nevertheless, it is unclear whether it always converges to the global maximum (minimum), and under what missing patterns the EM estimator is asymptotically valid, except for the homogenous random missing case studied by Jin et al. (2021). Our results show that actually the estimator calculated by the EM algorithm has the same asymptotic properties as the complete data case under a wide range of missing patterns, as long as the initial estimator is consistent on average in terms of Frobenius norm.

1.1 Intuition

Our asymptotic theory is built on the fact that the Hessian matrix for the objective function of the factor model is asymptotically diagonal and the tensor of the third order derivatives is sparse. More specifically, the diagonal elements (blocks) of the Hessian are of order $O_p(N)$ or $O_p(T)$ while the nondiagonal elements (blocks) are of order $O_p(1)$. This ensures that the estimator of a fixed dimensional target parameter is insensitive to the estimation errors of the remaining high dimensional parameters. Moreover, we show that the estimation errors in the least squares estimators of the factors and loadings are small enough on average. This, in conjunction with the asymptotically diagonal Hessian matrix, implies that the asymptotic distributions of the estimated factors (resp., loadings) are the same as if the loadings (resp., factors) were known, which is also observed in Bai's (2003) theory with complete data.

Since the complete data case can be considered as a special case of factor models with missing data, our asymptotic theory also explains Bai's (2003) results from the perspective of approximately diagonal Hessian. In fact, the general idea of diagonalization/orthogonality has been existing in the literature for a long time; see, e.g., Neyman (1979), Cox and Reid (1987), Lancaster (2002), Belloni, Chernozhukov and Hansen (2014), Belloni, Chernozhukov, Fernández-Val and Hansen (2017). However, the factor model literature rarely realizes the asymptotic diagonality structure in the Hessian and focuses almost entirely on the eigen-decomposition approach as pioneered by Bai (2003) with an exception by Wang (2022) who considers quasi-maximum likelihood estimation of nonlinear factor models. Almost all asymptotic analyses are explicitly or implicitly based on Bai's (2003) decomposition of the estimation error.[1] Unfortunately, Bai's (2003) decomposition is no longer applicable for factor models under more general setup, which includes the linear factor model with missing values studied here and nonlinear factor models with or without missing values. We show that the special structures in the Hessian and the third order derivative tensor enable us to go beyond the framework of Bai (2003) and derive a novel decomposition expression.

1.2 Related literature and contributions of this paper

First, this paper is closely related to the burgeoning matrix completion literature. Earlier works in this literature mainly focus on average (esp. Frobenius norm) convergence rate of the recovered missing

[1] See equation (A.1) in the Appendix A of Bai (2003).

values under homogenous or limited heterogenous random missing;[2] see Candes and Plan (2010), Koltchinskii, Lounici and Tsybakov (2011), Negahban and Wainwright (2011, 2012), Koltchinskii (2011), and Rohde and Tsybakov (2011), among others. Motivated by empirical applications in recommendation systems, causal inference and many social science studies, recent developments mainly focus on the derivation of the accurate convergence rates or asymptotic distributions for the estimators of elements or eigenvectors of certain low rank matrix in the context of heterogenous random missing or nonrandom missing. See Schnabel, Swaminathan, Singh, Chandak and Joachims (2016), Ma and Chen (2019), Sportisse, Boyer and Josse (2020), Athey et al. (2021), Bhattacharya and Chatterjee (2022), Zhu, Wang and Samworth (2022), Agarwal, Dahleh, Shah and Shen (2023) for heterogenous/nonrandom missing; see Chen, Fan, Ma and Yan (2019), Xia and Yuan (2021) and Chernozhukov et al. (2023) for asymptotic distribution results. Due to the convex nature of the nuclear norm, the nuclear norm regularization (NNR) approach has become one of the most popular approaches in the literature (Mazumder, Hastie and Tibshirani (2010)). However, due to the shrinkage bias caused by the nuclear norm regularization and the lack of explicit analytical expression for the estimator, post NNR inference is a difficult open question. Chen et al. (2019) and Xia and Yuan (2021) tackle this issue under the assumption of homogenous missing across both cross sectional and time dimensions, while Chernozhukov et al. (2023) allow heterogenous missing across either cross section dimension or time dimension, but not both. For more general missing patterns, e.g., where missing is heterogenous across both cross section and time dimensions or we have staggered missing, post regularization inference remains unknown.

In this paper, we provide a solution for post regularization inference under very general missing patterns, including heterogenous missing over both cross section and time dimensions, selection on covariates/factors/loadings, block/staggered missing, mixed frequency and ragged edge. In fact, we provide a unified framework for deriving the convergence rate and the limit distribution. In principle, other missing patterns may be also allowed as long as one can verify the restricted strong convexity condition and prove that the smallest eigenvalue of certain normalized Hessian matrix is bounded away from zero with probability approaching 1 (w.p.a.1).

Second, the paper is closely related to the flourishing causal inference literature. The causal inference literature mainly considers block missing or staggered missing (staggered treatment adoption), where some units are treated from possibly different initial dates to the end of the sample period. Under the potential outcome framework (e.g., Rubin (1974)), the untreated potential outcomes of

[2] Heterogenous missing means that the missing probabilities could vary across cross sectional units or/and time.

the treated observations are considered as missing and the objective is to impute the missing values using the control observations. The advantage of factor models is that the untreated potential outcomes of different units are allowed to follow unparallel trends and the treatment effects are allowed to be heterogenous over both cross section and time dimensions. See, e.g., Gobillon and Magnac (2016), Xu (2017), Chan and Kwok (2022) and Liu, Wang and Xu (2024). Important theoretical progresses have been made recently by Athey et al. (2021), Bai and Ng (2021) and Xiong and Pelger (2023). Athey et al. (2021) consider nuclear norm penalized least squares estimation and prove the average consistency of the imputed values. Bai and Ng (2021) propose a two-step estimation procedure for the block missing cases and provide an elegant inferential theory. Xiong and Pelger (2023) propose to apply PCA on the adjusted sample covariance matrix where each entry is adjusted by the inverse observation proportion. While these progresses are useful, the asymptotic properties of the fundamental least squares estimator based on the observed data remain unclear.

This paper provides the inference theory for the least squares estimator of factor models with missing values. Compared with Athey et al. (2021) who only establish an average convergence rate for the NNR estimator, we provide a complete set of inference theories. Compared with Bai and Ng (2021) who focus on block missing, our method applies to much more general missing patterns and may improve the efficiency if there are data outside of their "tall-wide" block. For example, we allow the treatment timing to be correlated with the factors and loadings in a block/staggered treatment design. Our method is also more general and more efficient than Xiong and Pelger (2023). In Xiong and Pelger (2023), the missing probabilities are allowed to be correlated with the factors or loadings but not both and the covariance matrix of the factor is required to be time-invariant. We show that the least squares estimation does not require such conditions. In addition, if we use Xiong and Pelger's (2023) estimator as the initial value for the EM algorithm and iterate until convergence, we can improve the efficiency by eliminating the additional variance term resulting from reweighting the entries of the sample covariance matrix.[3]

Third, the paper is closely related to the unbalanced panel literature because of the ragged edge problem,[4] the mixed frequency issue and random missing. Various algorithms have been proposed to handle these missing patterns; see, e.g., Stock and Watson (2002), Mariano and Murasawa (2003), Giannone, Reichlin and Small (2008), Aruoba, Diebold and Scotti (2009), Doz, Giannone and Reichlin (2011), Jungbacker, Koopman and van der Wel (2011), and Banbura and Modugno (2014).

[3] Xiong and Pelger (2023) point out in its simulation section that EM iterations could further improve the efficiency but there is no inferential theory for the iterative estimator under general missing patterns.

[4] In real-time data analyses, the ragged edge problem may mean that there is mising data at the end of the sample period or it arises because different series are released at different time.

Typically, these algorithms estimate the model parameters (mainly the loadings) by either PCA using only the balanced part of the panel or maximum likelihood estimation (MLE) using the EM algorithm, and then estimate the factors by Kalman smoother using the estimated parameters and the whole unbalanced panel. The PCA estimator using only the truncated balanced panel is easy to implement and well studied by Doz et al. (2011), but truncation may lead to serious efficiency loss or selection bias, especially when dealing with asset pricing panels and other high dimensional data; see Bryzgalova, Lerner, Lettau and Pelger (2022), Chen and McCoy (2022) and Freyberger, Höppner, Neuhierl and Weber (2022) for detailed discussions. The likelihood-based estimators (Stock and Watson (2002), Mariano and Murasawa (2003), Banbura and Modugno (2014)) do not suffer from the truncation issue, but their asymptotic properties and the corresponding missing pattern conditions are unknown.

This paper contributes to this literature by establishing the asymptotic theory of the least squares estimation without truncating the unbalanced panel into a balanced one or aggregating the high frequency series into low frequency series. Since least squares estimation is not equivalent to PCA estimation for mixed frequency factor models, how to analyze the least squares estimator of mixed frequency factor models is a well-recognized yet unsolved problem. This paper solves this problem. It is also worth noting that our results allow the missing probabilities to be correlated with the latent factors and loadings, which is particularly important for survey data and asset pricing panels, as discussed in Bryzgalova et al. (2022). In addition, our results also illuminate the (large N large T) asymptotic analysis of the MLE approaches in Mariano and Murasawa (2003), Banbura and Modugno (2014) and other related papers.[5] These approaches are quite popular in the nowcasting literature for handling mixed frequency data.

1.3 Roadmap

The rest of the paper is structured as follows. Section 2 introduces the notations, missing patterns and the estimation strategy. Section 3 discusses the roadmap for the asymptotic analyses by outlining the key steps and intuitions. Section 4 presents the assumptions and the asymptotic properties. Section 5 consider two potential applications of the theoretical results in the paper. Section 6 presents some simulation results. Section 7 presents an application to the UK grant allocation data to test the

[5] Mariano and Murasawa (2003) and Banbura and Modugno (2014) maximize essentially the same likelihood function using different algorithms. The former uses Kalman filter to evaluate the likehood and quasi-Newton method to maximize the likelihood, while the latter uses the EM algorithm. Note that these two papers treat the factors as missing data when calculating the likelihood, while the EM algorithm in Stock and Watson (2002) treats both the factors and the loadings as parameters.

average treatment effects of partisan alignment. Section 8 concludes. All proofs are relegated to the online appendix.

Notation. For a matrix A, we use $\|A\|$, $\|A\|_F$, $\|A\|_*$ and $\|A\|_\infty$ to denote its spectral norm, Euclidean norm, nuclear norm, and elementwise max form, respectively. $\sigma_{\min}(A)$ denotes the smallest eigenvalue of A. "$\circ$" denotes the Hadamard product of two vectors or matrices. $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and distribution, respectively. We use $(N,T) \to \infty$ to denote that N and T pass to infinity jointly. For a positive integer a, let $[a] \equiv \{1, 2, ..., a\}$, where $\equiv$ signifies a definitional relationship. Let I_a denote an $a \times a$ identity matrix. Let $c_1 \vee c_2 = \max(c_1, c_2)$ and $c_1 \wedge c_2 = \min(c_1, c_2)$. Let $c_{NT} = \sqrt{N \wedge T}$. Let M denote a generic large positive constant whose value may vary over places.

2 Missing Patterns and Estimation

Consider the following factor model with missing values:

$$\begin{aligned} y_{it} &= d_{it}(f_t^{0\prime}\lambda_i^0 + v_{it}) \text{ for } i \in [N] \text{ and } t \in [T], \\ d_{it} &= 1\{y_{it} \text{ is observable}\}, \end{aligned} \tag{2.1}$$

where $1\{\cdot\}$ denotes the usual indicator function, f_t^0 is the r-dimensional factor at time t and λ_i^0 is the r-dimensional loading for unit i, r is the number of factors, and v_{it} is the error term. In (2.1), we use 0 to denote y_{it} when it is not observable. Our objective is to estimate the factors and the loadings using the observed data, and establish the asymptotic theory for the proposed estimator.

2.1 Missing Patterns

Let $\phi^0 = (\lambda^{0\prime}, f^{0\prime})'$, where $\lambda^0 = (\lambda_1^{0\prime}, ..., \lambda_N^{0\prime})'$, and $f^0 = (f_1^{0\prime}, ..., f_T^{0\prime})'$. Let $\mathbb{E}_\phi(\cdot) = \mathbb{E}(\cdot|\phi^0)$. The missing data patterns allowed in this paper are listed below. These patterns can be divided into two types depending on whether $\mathbb{E}_\phi(d_{it}) = 0$ is allowed or not for some (i,t). The first type assumes $\mathbb{E}_\phi(d_{it}) \geq c > 0$ for all i and t, while the second type allows $\mathbb{E}_\phi(d_{it}) = 0$ for some i and t. For all of these patterns, d_{it} is assumed to be independent with v_{js} for all j and s, which corresponds to the unconfoundedness condition in Rosenbaum and Rubin (1983).

Example 1 ***(Completely random heterogenous missing):*** *d_{it} is independent across i and t and independent of f_s^0, λ_j^0 and v_{js} for all (j,s). The missing probability, $1 - \mathbb{E}(d_{it})$, is allowed to vary across both i and t, and $\min_{i,t} \mathbb{E}(d_{it}) \geq c > 0$.*

Note that here $\mathbb{E}(d_{it})$ can also be written as $\mathbb{E}_\phi(d_{it})$ because of independence between d_{it} and ϕ^0. In Rubin's (1976) terminology, this case is called missing completely at random (MCAR). The next example considers selection on observable covariates and/or factors and loadings, which is called missing not at random (MNAR).

Example 2 ***(Selection on covariates or factors and loadings):*** *$d_{it} - \mathbb{E}_\phi(d_{it})$ is independent across i and t, and independent with v_{js} for all j and s. $\mathbb{E}_\phi(d_{it})$ is independent with $\{v_{js}\}$ but could be correlated with λ_j^0 and f_s^0 for some j and s. The conditional missing probability, $1 - \mathbb{E}_\phi(d_{it})$, is allowed to vary across i and t, and $\mathbb{E}_\phi(d_{it}) \geq c > 0$.*

The key difference with Example 1 is that here we allow $\mathbb{E}_\phi(d_{it})$ to be correlated with the factors and loadings. This is particularly important as in recommendation system and many causal social science studies the missingness arises from treatment assignments or individual choices, and consequently the missing probability is correlated with certain elements of the matrix itself. For example, in the matrix consisting of movie ratings, the probability that a person submits his rating for a movie is positively correlated with how much she likes that movie. In survey data, high income respondents are less likely to answer questions that has tax consequences. In asset pricing panels consisting of characteristics of different firms, the missing probabilities of firm characteristics tend to be larger for small-cap firms, for extreme values of the characteristics and for certain time periods with common macroeconomic shock.

More formally, we can model the selection/assignment equation as

$$d_{it}^* = z_{it}\beta^0 + g_t^{0\prime}\alpha_i^0 + u_{it}, \;\; d_{it}^* \text{ is latent and } d_{it} = 1\left\{d_{it}^* > 0\right\}, \tag{2.2}$$

where z_{it} denotes some observable exogenous/predetermined variables (e.g., $z_{it} = d_{i,t-1}$), g_t^0 and α_i^0 denote some latent factors and loadings, and u_{it} denotes the error term. Thus each unit is allowed to switch between the treated status and the untreated status. Our asymptotic results imply that the least squares estimators of f_t^0 and λ_i^0 has no selection bias even if f_t^0 and λ_i^0 are correlated with z_{it}, g_t^0 and α_i^0, as long as v_{it} is uncorrelated with z_{it}, g_t^0, α_i^0 and u_{it} (strong ignorability). However, if there are missing covariates or factors in both y_{it} and d_{it}^*, then v_{it} is correlated with $\mathbb{E}_\phi(d_{it})$ and there is a selection bias.

To handle heterogenous $\mathbb{E}_\phi(d_{it})$, so far the literature only considers the case where $\mathbb{E}_\phi(d_{it})$ is heterogenous across i or t but not both (see, e.g., Chernozhukov et al. (2023), Zhu et al. (2022)), or assumes that $\mathbb{E}_\phi(d_{it})$ itself has an approximately low rank structure or depends only on observable exogenous covariates (see, e.g., Schnabel et al. (2016), Ma and Chen (2019), Bhattacharya and Chatterjee (2022), Sportisse et al. (2020)). Xiong and Pelger (2023) allow $\mathbb{E}_\phi(d_{it})$ to be heterogenous across both i and t, but do not allow $\mathbb{E}_\phi(d_{it})$ to be correlated with f_t^0; $\mathbb{E}(f_t^0 f_t^{0\prime})$ is not allowed to be different across t either. Example 2 includes all these patterns as special cases.

Example 3 ***(Block missing):*** *$d_{it} = 0$ for $i > N_o$ and $t > T_o$, where N_o and T_o denote the cardinality of $\{i \in [N] : d_{it} = 1$ for all $t\}$ and $\{t \in [T] : d_{it} = 1$ for all $i\}$, respectively. N_o/N and T_o/T are bounded away from zero as $(N,T) \to \infty$.*

Example 4 ***(Staggered treatment):*** *$d_{it} = 1$ for all $i \leq N_o$ and $t \in [T]$ and $d_{it} = 1$ for $i > N_o$ and $t \leq T_{oi}$; there are no restrictions on d_{it} for $i > N_o$ and $t > T_{oi}$. $T_o = \min_i T_{oi}$, and N_o/N and T_o/T are bounded away from zero as $(N,T) \to \infty$. In particular, the starting treatment time for individual $i > N_o$ can be written as $T_{oi} + 1$.*

Example 5 ***(Mixed frequency):*** *$d_{it} = 0$ if $i > N_o$ and t/h is not an integer, where N_o is the number of high frequency series, h is the frequency ratio. Note that after reorganizing the data across t, this case is just block missing when there are two different frequencies and staggered missing when there are more than two frequencies.*

These three examples allow d_{it} to be strongly correlated across i and t, the key difference from Example 2 is that here we allow $\mathbb{E}_\phi(d_{it}) = 0$ for some i and t. Note that here we focus on the strong signal case with $N_o/N > \underline{c}$ and $T_o/T > \underline{c}$ for some small $\underline{c} > 0$. Conceptually it is not difficult to extend our results to allow N_o/N and T_o/T to tend to zero at certain speed. But for notational simplicity we do not pursue this direction.

Examples 3–4 are relevant for program evaluation. For block missing, the units with $i \leq N_o$ would never get treated while the units with $i > N_o$ get treated simultaneously at time $T_o + 1$. For staggered treatment, the units with $i \leq N_o$ would never get treated, while the units with $i > N_o$ get treated from $T_{oi} + 1$ to the end of the sample and $T_o = \min_i\{T_{oi}\}$. For each i, the treatment timing T_{oi} is allowed to be correlated with the factors and loadings, which is relevant for the event study literature. Our asymptotic results only require that there exist N_o and T_o such that $d_{it} = 1$ for $i \leq N_o$ or $t \leq T_o$. Example 5 is relevant for the nowcasting literature. To estimate factor model from mixed frequency data, so far the literature uses either maximum likelihood (e.g., Mariano and Murasawa (2003), Banbura and Modugno (2014)), which is lack of asymptotic theory, or the PCA on the high frequency data, which is not efficient. Our results show that direct least squares estimation on the mixed frequency data is asymptotically normal and efficient.

Example 6 ***(Raggered edge/No missing):*** *$d_{it} = 1$ for $i \in [N]$ and $t \in [T-1]$ (or $[T]$).*

For raggered edge data, the missing observations may only arise at the end of the sample period, and typically principal component estimation is applied on the complete data excluding the data in the last period. Thus we put this case together with the no missing case. Our asymptotic theory also applies here, thus includes the results of Bai (2003) as a special case.

Other missing patterns could be allowed for as long as we can prove average consistency for the estimated factors and loadings (mainly verify the restricted strong convexity condition) and prove that the smallest eigenvalue of the normalized Hessian is bounded away from zero in probability.

2.2 Estimation

Let $\lambda = (\lambda_1', ..., \lambda_N')'$, $f = (f_1', ..., f_T')'$, $\Lambda = (\lambda_1, ..., \lambda_N)'$, and $F = (f_1, ..., f_T)'$. We propose to maximize the following penalized partial likelihood function:

$$Q(\lambda, f) = L(\lambda, f) + P(\lambda, f), \tag{2.3}$$

where

$$\begin{aligned} L(\lambda, f) &= -\frac{1}{2}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} d_{it}(y_{it} - f_t'\lambda_i)^2, \text{ and} &(2.4)\\ P(\lambda, f) &= -\frac{cNT}{8}\left\| dg(\frac{\Lambda'\Lambda}{N} - \frac{F'F}{T})\right\|_F^2 - \frac{cNT}{2}\left\| ndg(\frac{\Lambda'\Lambda}{N})\right\|_F^2 - \frac{cNT}{2}\left\| ndg(\frac{F'F}{T})\right\|_F^2. &(2.5)\end{aligned}$$

Below we explain the terms defined in (2.4) and (2.5) in order.

Here, $L(\lambda, f)$ is the partial quasi Gaussian likelihood of the outcome equation, ignoring the constant term and the likelihood function of d_{it} from the selection equation. Given ϕ^0, the probability of $d_{it} = 1$, viz., $\mathbb{E}_\phi(d_{it})$, may contain additional information about (f_t^0, λ_i^0), e.g., $\mathbb{E}_\phi(d_{it}) = \Phi(f_t^{0\prime}\lambda_i^0)$, where $\Phi(\cdot)$ denotes the CDF of the standard normal or logistic distribution function. However, utilizing such information requires us to assume a fully parametric model for d_{it}; see, e.g., equation (2.2) and the link function $\Phi(\cdot)$. We shall just focus on $L(\lambda, f)$, which avoids any parametric assumption on $\mathbb{E}_\phi(d_{it})$. In this case, maximizing the quasi Gaussian likelihood function is equivalent to minimizing the least squares objective function.

$P(\lambda, f)$ denotes a penalty function that accounts necessary restrictions on (λ, f) for the purpose of identification. $dg(\frac{\Lambda'\Lambda}{N} - \frac{F'F}{T})$ denotes a diagonal matrix with the same diagonal elements as $\frac{\Lambda'\Lambda}{N} - \frac{F'F}{T}$, $ndg(\frac{\Lambda'\Lambda}{N})$ denotes an upper-triangular matrix with the same elements as $\frac{\Lambda'\Lambda}{N}$ in the upper-triangular block, and $ndg(\frac{F'F}{T})$ is defined in the same way. c is an arbitrary positive constant. Thus adding the penalty $P(\lambda, f)$ is equivalent to imposing the following set of identification restrictions:

$$\frac{1}{N}\sum_{i=1}^{N}\lambda_i\lambda_i' = \frac{1}{T}\sum_{t=1}^{T} f_t f_t' \text{ and both are diagonal.} \tag{2.6}$$

Obviously, (2.6) imposes r^2 restrictions for identification. As a matter of fact, for any (F, Λ) and any $r \times r$ invertible matrix G, $L(FG, \Lambda G'^{-1}) = L(F, \Lambda)$, and there is a unique G such that $(FG, \Lambda G'^{-1})$ satisfies the r^2 restrictions in (2.6). Without loss of generality, we assume that after certain normalizations or redefinitions, the true value (Λ^0, F^0) also satisfies this restriction, i.e.,

$$\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime} = \frac{1}{T}\sum_{t=1}^{T} f_t^0 f_t^{0\prime} \text{ and both are diagonal.} \tag{2.7}$$

If (Λ^0, F^0) does not satisfy this restriction, there always exists an $r \times r$ normalization matrix G^0 such that $(F^0G^0, \Lambda^0(G^{0\prime})^{-1})$ satisfies this restriction, and we can redefine $(F^0G^0, \Lambda^0(G^{0\prime})^{-1})$ as the true value.

To account for as many missing patterns as possible, we restrict our attention to the case where the factors and loadings are uniformly bounded. The partial maximum likelihood estimator is obtained as follows:

$$(\hat{\lambda}, \hat{f}) = \arg\max_{\|\lambda\|_\infty \leq M,\ \|f\|_\infty \leq M} Q(\lambda, f),$$

where $\hat{\lambda} = (\hat{\lambda}_1', ..., \hat{\lambda}_N')'$ and $\hat{f} = (\hat{f}_1', ..., \hat{f}_T')'$. Let $\hat{\Lambda} = (\hat{\lambda}_1, ..., \hat{\lambda}_N)'$ and $\hat{F} = (\hat{f}_1, ..., \hat{f}_T)'$. In the above maximization, we impose the conditions that $\|\lambda\|_\infty \leq M$ and $\|f\|_\infty \leq M$ to help verify the *restricted strong convexity* (RSC) condition used in the proof of Theorem 4.1 of Section 4.2. It is also imposed in Negahban and Wainwright (2012), Chernozhukov et al. (2023), and many other papers in the matrix completion literature. But for the block missing cases, this condition is not needed as there are other ways to obtain initial consistent estimates of the factors and loadings.

Algorithm 2.1 ***Partial Maximum Likelihood Estimation***

1. *Obtain initial consistent estimates of the factors and loadings, $\tilde{f}$ and $\tilde{\lambda}$.*[6]

   (1) For the random missing cases (Examples 1-2), we use the nuclear norm regularized estimation. A popular algorithm for calculating the nuclear norm regularized estimator is the iterative singular value thresholding (ISVT) algorithm in Mazumder et al. (2010).

   (2) For examples 3-5, we can either use the nuclear norm regularized estimation, or apply PCA

[6] The estimators $\tilde{f}$ and $\tilde{\lambda}$ are consistent on average (in terms of Frobenius norm) if $\frac{1}{\sqrt{T}}\left\|\tilde{f} - f^0\right\|_F = o_p(1)$ and $\frac{1}{\sqrt{N}}\left\|\tilde{\lambda} - \lambda^0\right\|_F = o_p(1)$.

estimation on the two complete data blocks $(1 \le i \le N_o, 1 \le t \le T)$ *and* $(N_o + 1 \le i \le N, 1 \le t \le T_o)$ *separately as in Bai and Ng (2021). We can also use Xiong and Pelger's (2023) inverse observation-proportion weighted estimator.*

2. *Use the estimator in Step 1 as the initial value for the EM algorithm of Stock and Watson (2002) and iterate until convergence. That is, we use the estimated factors and loadings in the last iteration to impute the missing values and then update the estimated factors and loadings by the principal components of this imputed complete matrix, repeat this procedure until convergence.*

3 Roadmap and Discussion

In this section, we consider the roadmap that paves the way for formal derivation of the asymptotic properties of the partial maximum likelihood estimators $\hat{\lambda}$ and $\hat{f}$.

3.1 The Case of Random Missing

We focus on the case of random missing in Examples 1–2. It is well known that nuclear norm penalized least squares estimation is consistent on average (in terms of Frobenius norm) and computationally advantageous. On the other hand, the nuclear norm penalty also brings in shrinkage bias and makes it infeasible to derive explicit asymptotic expansion of the estimation error, which is crucial for proving the limit distributions of the estimators. Therefore, the crucial issue is how to eliminate the regularization bias and derive the asymptotic distributions.

As discussed in the introduction and in Example 2, existing methods typically rely on restrictive assumptions on $\mathbb{E}_\phi(d_{it})$ in order to ensure a delicately designed second step to eliminate the bias. Our solution is simple: just go back to the unpenalized least squares estimator but take into account the identification restrictions. In the following, we outline the key steps in the derivation of the accurate convergence rates and limit distributions for the estimated factors and loadings.

3.1.1 The Convergence Rates

Let $\phi = (\lambda', f')'$, $\phi^0 = (\lambda^{0\prime}, f^{0\prime})'$, and $\hat{\phi} = (\hat{\lambda}', \hat{f}')'$. Define the score and Hessian functions of $Q(\phi)$:

$$S_\phi(\phi) = \partial_\phi Q(\phi) \text{ and } H_{\phi\phi'}(\phi) = \partial_{\phi\phi'} Q(\phi).$$

When $S_\phi(\phi)$ and $H_{\phi\phi'}(\phi)$ are evaluated at ϕ^0, we simply write them as S_ϕ and $H_{\phi\phi'}$, respectively. The first order conditions (FOCs) of maximizing $Q(\phi)$ are given by $S_\phi(\hat{\phi}) = 0$.

First, we will show that $\hat{\phi}$ is consistent on average so that we can conduct the first order Taylor expansion of the above FOCs around ϕ^0 to obtain $0 = S_\phi + H_{\phi\phi'}(\hat{\phi} - \phi^0) + R_\phi$, or equivalently,

$$\hat{\phi} - \phi^0 = -H_{\phi\phi'}^{-1} S_\phi - H_{\phi\phi'}^{-1} R_\phi, \tag{3.1}$$

where R_ϕ denotes the $(Nr+Tr)\times 1$ vector of remainder terms: $R_\phi = (R'_\lambda, R'_f)'$, $R_\lambda = (R'_{\lambda_1}, ..., R'_{\lambda_N})'$, and $R_f = (R'_{f_1}, ..., R'_{f_T})'$. The above equivalence holds provided the inverse of $H_{\phi\phi'}$ is well defined asymptotically after suitable normalizations. Let

$$D_{NT} = \begin{pmatrix} N \times I_{Nr} & 0 \\ 0 & T \times I_{Tr} \end{pmatrix} \text{ and } D_{TN} = \begin{pmatrix} T \times I_{Nr} & 0 \\ 0 & N \times I_{Tr} \end{pmatrix}.$$

Then the normalized version of equation (3.1) is given by

$$\begin{aligned} \begin{pmatrix} \frac{1}{\sqrt{N}}(\hat{\lambda} - \lambda^0) \\ \frac{1}{\sqrt{T}}(\hat{f} - f^0) \end{pmatrix} &= D_{NT}^{-\frac{1}{2}}(\hat{\phi} - \phi^0) = -D_{NT}^{-\frac{1}{2}} H_{\phi\phi'}^{-1} S_\phi - D_{NT}^{-\frac{1}{2}} H_{\phi\phi'}^{-1} R_\phi \\ &= (-D_{TN}^{-\frac{1}{2}} H_{\phi\phi'} D_{TN}^{-\frac{1}{2}})^{-1} \frac{D_{TN}^{-\frac{1}{2}} S_\phi}{\sqrt{NT}} + (-D_{TN}^{-\frac{1}{2}} H_{\phi\phi'} D_{TN}^{-\frac{1}{2}})^{-1} \frac{D_{TN}^{-\frac{1}{2}} R_\phi}{\sqrt{NT}}. \end{aligned} \tag{3.2}$$

Second, noting that $S_\phi = (S'_{\lambda_1}, ..., S'_{\lambda_N}, S'_{f_1}, ..., S'_{f_T})'$ with $S_{\lambda_i} = \sum_{t=1}^T d_{it} v_{it} f_t^0$ and $S_{f_t} = \sum_{i=1}^N d_{it} v_{it} \lambda_i^0$, it is easy to see that

$$\left\| \frac{D_{TN}^{-\frac{1}{2}} S_\phi}{\sqrt{NT}} \right\| = O_p(\frac{1}{c_{NT}}). \tag{3.3}$$

Third, we show in Lemmas B.1–B.2 in the Appendix that under standard regularity conditions, as $(N, T) \to \infty$,

$$\left\| (-D_{TN}^{-\frac{1}{2}} H_{\phi\phi'} D_{TN}^{-\frac{1}{2}})^{-1} \right\| = O_p(1). \tag{3.4}$$

While this result appears simple, the proof is quite complicated because of the large dimension of the Hessian matrix $H_{\phi\phi'}$ as an $(Nr + Tr) \times (Nr + Tr)$ matrix. Our proof utilizes the special structure of $H_{\phi\phi'}$, which comes from the factor model itself. It is also crucial to normalize $H_{\phi\phi'}$ by $D_{TN}^{-\frac{1}{2}}$, since the eigenvalues of $H_{\phi\phi'}$ have different asymptotic orders when N and T pass to infinity at different speeds. For details, see the proofs of Lemmas B.1–B.2.

Fourth, for the remainder term $R_\phi = (R'_\lambda, R'_f)'$, we show in Lemma B.3 in the Appendix that

$$\begin{aligned} \|R_{\lambda_i}\| &= \left\|\hat{\lambda}_i - \lambda_i^0\right\| O_p(\sqrt{T}\left\|\hat{f} - f^0\right\| + \frac{T}{\sqrt{N}}\left\|\hat{\lambda} - \lambda^0\right\|) + O_p(\left\|\hat{f} - f^0\right\|^2 + \frac{T}{N}\left\|\hat{\lambda} - \lambda^0\right\|^2), \quad (3.5) \\ \|R_{f_t}\| &= \left\|\hat{f}_t - f_t^0\right\| O_p(\sqrt{N}\left\|\hat{\lambda} - \lambda^0\right\| + \frac{N}{\sqrt{T}}\left\|\hat{f} - f^0\right\|) + O_p(\left\|\hat{\lambda} - \lambda^0\right\|^2 + \frac{N}{T}\left\|\hat{f} - f^0\right\|^2), \quad (3.6) \end{aligned}$$

$$\left\| \frac{R_\lambda}{\sqrt{T}} \right\| = O_p(\left\|\hat{\lambda} - \lambda^0\right\| \left\|\hat{f} - f^0\right\| + \sqrt{\frac{N}{T}} \left\|\hat{f} - f^0\right\|^2 + \sqrt{\frac{T}{N}} \left\|\hat{\lambda} - \lambda^0\right\|^2), \quad (3.7)$$

$$\left\| \frac{R_f}{\sqrt{N}} \right\| = O_p(\left\|\hat{\lambda} - \lambda^0\right\| \left\|\hat{f} - f^0\right\| + \sqrt{\frac{N}{T}} \left\|\hat{f} - f^0\right\|^2 + \sqrt{\frac{T}{N}} \left\|\hat{\lambda} - \lambda^0\right\|^2). \quad (3.8)$$

In fact, the above results hold for arbitrary missing patterns discussed above. Our proof utilizes the fact that the third order derivatives of $Q(\lambda, f)$ are sparse; e.g., $\partial_{\lambda_i \lambda_j f_s} Q(\lambda, f) = 0$ for all $i \neq j$ and $\partial_{\lambda_i f_s f_t} Q(\lambda, f) = 0$ for all $s \neq t$. See the proof of Lemma B.3 for details.

Combining equations (3.2)-(3.8) yields that

$$\begin{pmatrix} \frac{1}{\sqrt{N}}(\hat{\lambda} - \lambda^0) \\ \frac{1}{\sqrt{T}}(\hat{f} - f^0) \end{pmatrix} = O_p(\frac{1}{c_{NT}}) + O_p(\frac{1}{T} \left\|\hat{f} - f^0\right\|^2 + \frac{1}{N} \left\|\hat{\lambda} - \lambda^0\right\|^2). \quad (3.9)$$

where the order on the right hand side on (3.9) holds elementwise for the left hand side object. This expression allows us to refine the convergence rates. We shall show $\frac{1}{\sqrt{N}} \left\|\hat{\lambda} - \lambda^0\right\| = O_p(\frac{1}{\sqrt{c_{NT}}})$ and $\frac{1}{\sqrt{T}} \left\|\hat{f} - f^0\right\| = O_p(\frac{1}{\sqrt{c_{NT}}})$ in Theorem 4.1 below. Then plugging these initial rates back into equation (3.9), we immediately obtain that $\frac{1}{\sqrt{N}} \left\|\hat{\lambda} - \lambda^0\right\| = O_p(\frac{1}{c_{NT}})$ and $\frac{1}{\sqrt{T}} \left\|\hat{f} - f^0\right\| = O_p(\frac{1}{c_{NT}})$, which are the same as the rates in Bai (2003) for the complete data case.

3.1.2 The Limit Distributions

To derive the limit distribution of $\hat{\lambda}_i - \lambda_i^0$, from equation (3.1) we have

$$\hat{\lambda}_i - \lambda_i^0 = [\hat{\phi} - \phi^0]_i = -[H_{\phi\phi'}^{-1} S_\phi]_i - [H_{\phi\phi'}^{-1} R_\phi]_i,$$

where $[\cdot]_i$ denotes the i-th block of the vector inside the square brackets, each of length r. Utilizing the asymptotically block-diagonal structure of $H_{\phi\phi'}$, we can show that

$$[H_{\phi\phi'}^{-1} S_\phi]_i = ([L_{\lambda\lambda'}]_i)^{-1} S_{\lambda_i} + O_p(\frac{1}{\sqrt{NT}}), \quad (3.10)$$

$$[H_{\phi\phi'}^{-1} R_\phi]_i = O_p(\frac{1}{T}) \|R_{\lambda_i}\| + O_p(\frac{1}{T\sqrt{N}}) \|R_\lambda\| + O_p(\frac{1}{N\sqrt{T}}) \|R_f\|, \quad (3.11)$$

where $L_{\lambda\lambda'}$ is a block diagonal matrix, $[L_{\lambda\lambda'}]_i = \sum_{t=1}^T d_{it} f_t^0 f_t^{0\prime}$ is the (i,i)-th block of the square matrix $L_{\lambda\lambda'}$, each of size $r \times r$, and $S_{\lambda_i} = \sum_{t=1}^T d_{it} v_{it} f_t^0$. For detailed proof of these two expressions, see the proof of Theorem 4.3 in Section 4.2. Given the expressions of $\|R_{\lambda_i}\|$, $\|R_\lambda\|$ and $\|R_f\|$ in equations (3.5)–(3.8) and the fact that $\left\|\hat{\lambda} - \lambda^0\right\| = O_p(\frac{\sqrt{N}}{c_{NT}})$ and $\left\|\hat{f} - f^0\right\| = O_p(\frac{\sqrt{T}}{c_{NT}})$, it is easy to see

$$\|R_{\lambda_i}\| = \left\|\hat{\lambda}_i - \lambda_i^0\right\| O_p(\frac{T}{c_{NT}}) + O_p(\frac{T}{c_{NT}^2}), \ \ \|R_\lambda\| = O_p(\frac{T\sqrt{N}}{c_{NT}^2}), \text{ and } \|R_f\| = O_p(\frac{N\sqrt{T}}{c_{NT}^2}).$$

Thus we have $[H_{\phi\phi'}^{-1}R_\phi]_i = \left\|\hat{\lambda}_i - \lambda_i^0\right\| O_p(\frac{1}{c_{NT}}) + O_p(\frac{1}{c_{NT}^2})$ and it follows that

$$\hat{\lambda}_i - \lambda_i^0 = -\left([L_{\lambda\lambda'}]_i\right)^{-1} S_{\lambda_i} + O_p(\frac{1}{c_{NT}}) \left\|\hat{\lambda}_i - \lambda_i^0\right\| + O_p(\frac{1}{c_{NT}^2}).$$

Under some regularity conditions, $\left([\frac{1}{T}L_{\lambda\lambda'}]_i\right)^{-1} \frac{1}{\sqrt{T}} S_{\lambda_i}$ is asymptotically normal, thus $\sqrt{T}(\hat{\lambda}_i - \lambda_i^0)$ is also asymptotically normal if $\frac{\sqrt{T}}{c_{NT}^2} \to 0$. The limit distribution of $\sqrt{N}(\hat{f}_t - f_t^0)$ follows from similar arguments.

3.1.3 Summary

In summary, we first derive preliminary consistent but inaccurate initial convergences rates: $\left\|\hat{\lambda} - \lambda^0\right\| = O_p(\frac{\sqrt{N}}{\sqrt{c_{NT}}})$ and $\left\|\hat{f} - f^0\right\| = O_p(\frac{\sqrt{T}}{\sqrt{c_{NT}}})$. Then we use equation (3.9) to refine the rates to obtain $\left\|\hat{\lambda} - \lambda^0\right\| = O_p(\frac{\sqrt{N}}{c_{NT}})$ and $\left\|\hat{f} - f^0\right\| = O_p(\frac{\sqrt{T}}{c_{NT}})$. These rates together with the expressions of $\|R_{\lambda_i}\|$, $\|R_\lambda\|$ and $\|R_f\|$ in equations (3.5)-(3.8) imply that $[H_{\phi\phi'}^{-1}R_\phi]_i = O_p(\frac{1}{c_{NT}}) \left\|\hat{\lambda}_i - \lambda_i^0\right\| + O_p(\frac{1}{c_{NT}^2})$, which is asymptotically negligible compared with the leading term $-\left([L_{\lambda\lambda'}]_i\right)^{-1} S_{\lambda_i}$. Note that these results are also valid for the complete data case (see Example 6).

Equation (3.1) plays a similar role as the eigen-decomposition expression (equation A.1) in the Appendix A of Bai (2003). So far almost all asymptotic analyses of factor model are essentially based on/similar to Bai's (2003) eigen-decomposition, but it is applicable only for linear factor models with complete data. Equation (3.1) together with the structure of $H_{\phi\phi'}$ provides an alternative and more general way to decompose the estimation error $\hat{\lambda} - \lambda^0$ and $\hat{f} - f^0$. We believe that based on equation (3.1), we can generalize the existing results of factor models to many other setups such as nonlinear factor models or linear/nonlinear panel data models with missing values.

3.2 The Case of Block/Staggered Missing

For the block/staggered missing case (Examples 3–5), the roadmap is essentially the same as the random missing case. The major difficulty is how to show $(-D_{TN}^{-\frac{1}{2}} H_{\phi\phi'} D_{TN}^{-\frac{1}{2}})^{-1} = O_p(1)$ because the structure of $H_{\phi\phi'}$ under block/staggered missing and under random missing are quite different.

A key condition for proving $(-D_{TN}^{-\frac{1}{2}} H_{\phi\phi'} D_{TN}^{-\frac{1}{2}})^{-1} = O_p(1)$ in the random missing cases is that $\mathbb{E}_\phi(d_{it}) > 0$ for all i and t. Nevertheless, this condition is violated in the block/staggered missing case where $\mathbb{E}_\phi(d_{it}) = 0$ for some (i,t) when $i > N_o$ and $t > T_o$. Because of this fundamental difference, we prove $(-D_{TN}^{-\frac{1}{2}} H_{\phi\phi'} D_{TN}^{-\frac{1}{2}})^{-1} = O_p(1)$ separately in Lemma B.2 of the Appendix utilizing a totally different strategy. In fact, after some effort we successfully calculate all the eigenvalues and eigenvectors of $D_{TN}^{-\frac{1}{2}} H_{\phi\phi'} D_{TN}^{-\frac{1}{2}}$ for the block/staggered missing case!

4 Assumptions and Asymptotic Theories

In this section we first provide some assumptions and then formally study the asymptotic properties of the partial maximum likelihood estimators (PMLEs) $\hat{\lambda}$ and $\hat{f}$. Recall that M denote some generic large positive constant whose value may vary over places.

4.1 Assumptions

Assumption 1 *(i)* $\frac{1}{T} F^{0\prime} F^0 \xrightarrow{p} \Sigma_F > 0$, *and* $\max_{1 \le t \le T} \left\| f_t^0 \right\| \le M$.

(ii) $\frac{1}{N} \Lambda^{0\prime} \Lambda^0 \xrightarrow{p} \Sigma_\Lambda > 0$ *and* $\max_{1 \le i \le N} \left\| \lambda_i^0 \right\| \le M$.

Assumption 1(i)–(ii) corresponds to Assumptions A–B in Bai (2003). As in Bai (2003) we focus on the case of strong factors. Unlike Bai (2003) who only assumes uniformly bounded loadings, we require that both the factors and loadings are uniformly bounded. In the matrix completion literature, the uniform boundedness of both $\left\| f_t^0 \right\|$ and $\left\| \lambda_i^0 \right\|$ is sometimes referred to the incoherence condition; see. e.g., Candes and Recht (2009), Candes and Plan (2010), Keshavan, Montanari and Oh (2010), Negahban and Wainwright (2012), Chen et al. (2019), and Chernozhukov et al. (2023). This condition requires that the entries of the singular vectors of the latent signal matrix $F^0 \Lambda^{0\prime}$ should be approximately evenly distributed. Technically, the incoherent condition is crucial for verifying the restricted strong convexity (RSC) condition, which is a key step for proving the average consistency (in terms of Frobenius norm) of the imputed matrix using nuclear norm penalization. See, e.g., Negahban and Wainwright (2011, 2012) and Chernozhukov et al. (2023). In addition, the minimax result in Chernozhukov et al. (2023) argues that the incoherence condition is necessary. For the random missing cases (Examples 1–2), our proof for the initial convergence rates of $\left\| \hat{\lambda} - \lambda^0 \right\|$ and $\left\| \hat{f} - f^0 \right\|$ also requires verifying the RSC directly. This is why we also need the incoherence condition.

Let $\tilde{d}_{it} = d_{it} - \mathbb{E}_\phi(d_{it})$ and $\gamma_{Nd}(t,s) = \frac{1}{N} \sum_{i=1}^N |\mathbb{E}_\phi(\tilde{d}_{it}\tilde{d}_{is})|$. Let $\max_t = \max_{1 \le t \le T}$ and $\max_i = \max_{1 \le i \le N}$. Define $\min_t$ and $\min_i$ similarly.

Assumption 2 *(i) For random missing (Examples 1–2), given ϕ^0, $\tilde{d}_{it}$ is independent across i, and independent with $\{v_{js}\}$; $\mathbb{E}_\phi(d_{it})$ is independent with v_{js} and may vary across both i and t, and*

$\min_{i,t} \mathbb{E}_\phi(d_{it}) \geq c > 0$; $\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_{Nd}(t,s) \leq M$; $\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[d_{it} - \mathbb{E}_\phi(d_{it})]f_t^0 f_t^{0\prime}\right\|^{\kappa}) = O(1)$ *for some* $\kappa \geq 4$ *such that* $\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}} \to 0$ *and* $\frac{T^{\frac{1}{\kappa}}}{\sqrt{N}} \to 0$.

(ii) For block/staggered missing (Examples 3-4), $d_{it} = 1$ *for* $i \leq N_o$ *or* $t \leq T_o$ *and no restrictions on* d_{it} *for* $i > N_o$ *and* $t > T_o$*, where* N_o *and* T_o *is defined in Example 3. For mixed frequency (Example 5),* $d_{it} = 0$ *if* $i > N_o$ *and* t/h *is not an integer, where* N_o *is the number of high frequency series,* h *is the frequency ratio. In addition, as* $(N,T) \to \infty$*, both* N_o/N *and* T_o/T *are bounded away from zero, both* $\sigma_{\min}(\frac{1}{T}\sum_{t=1}^{T_o} f_t^0 f_t^{0\prime})$ *and* $\sigma_{\min}(\frac{1}{N}\sum_{i=1}^{N_o}\lambda_i^0\lambda_i^{0\prime})$ *are positive and bounded away from zero in probability.*

Assumption 2 summarizes the conditions on missing patterns discussed in Section 2.1. Note that our asymptotic results only require Assumption 2(i) or 2(ii), but not both. For random missing, Assumption 2(i) implies that $\mathbb{E}_\phi(d_{it})$ is allowed to vary across both i and t and be correlated with λ_j^0 and f_s^0 for some (j,s). The condition on $\gamma_{Nd}(t,s)$ ensures weak dependence of $\{\tilde{d}_{it}\}$ along the time dimension, which is comparable with Assumption 4(ii) below by noticing that $\max_t \gamma_{Nd}(t,t) \leq 1$.. As far as we know, Assumption 2 includes almost all the missing patterns considered in existing literature except for the case where the error term v_{js} is correlated with the missing mechanism.

Assumption 3 *The eigenvalues of the* $r \times r$ *matrix* $\Sigma_F \Sigma_\Lambda$ *are different.*

Assumption 3 is a standard identification condition and is the same as Assumption G in Bai (2003). It allows us to identify the factors and loadings from the common components.

Assumption 4 *Let* $\gamma_N(s,t) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(d_{is}v_{is}d_{it}v_{it})$.

(i) $\mathbb{E}(|d_{it}v_{it}|^4) \leq M$.

(ii) $\max_s \gamma_N(s,s) \leq M$ *and* $\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_N(s,t)| \leq M$.

(iii) For every (t,s), $\mathbb{E}\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[d_{is}v_{is}d_{it}v_{it} - \mathbb{E}(d_{is}v_{is}d_{it}v_{it})]\}^2 \leq M$.

Assumption 4 generalizes Assumption C in Bai (2003) to the missing data setting. When there is no missing data, $d_{it} = 1$ for all i and t, and these assumptions reduce to Bai's (2003) Assumption C with some slight modifications.

Assumption 5 $\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}||\frac{1}{\sqrt{T}}\sum_{t=1}^{T}d_{it}v_{it}f_t^0||^{\zeta}) \leq M$ *and* $\mathbb{E}(\frac{1}{T}\sum_{t=1}^{T}||\frac{1}{\sqrt{N}}\sum_{i=1}^{N}d_{it}v_{it}\lambda_i^0||^{\zeta}) \leq M$ *for some* $\zeta \geq 2$.

Assumption 5 generalizes Assumption D in Bai (2003) to the missing data setting. If $\zeta = 2$ and $d_{it} = 1$ for all i and t, the first part of the above assumption reduces to Assumption D in Bai (2003).

To introduce the next assumption, let $\bar{A}_{t\Lambda} = \frac{1}{N}\sum_{i=1}^{N} d_{it}\lambda_i^0\lambda_i^{0\prime}$, $\bar{A}_{iF} = \frac{1}{T}\sum_{t=1}^{T} d_{it}f_t^0 f_t^{0\prime}$, $A_{t\Lambda} = \mathbb{E}(\bar{A}_{t\Lambda})$, and $A_{iF} = \mathbb{E}(\bar{A}_{iF})$. Let $\xi_{1its} = f_t^0 d_{it}v_{it}d_{is}v_{is}$, $\xi_{2ijt} = \lambda_i^0 d_{it}v_{it}d_{jt}v_{jt}$, and $\xi_{it} = d_{it}v_{it}f_t^0\lambda_i^{0\prime}$.

Assumption 6 *(i)* $\frac{1}{T}\sum_{t=1}^{T}\left\|\bar{A}_{t\Lambda} - A_{t\Lambda}\right\|^2 = O_P(\frac{1}{N})$, $\frac{1}{N}\sum_{i=1}^{N}\left\|\bar{A}_{iF} - A_{iF}\right\|^2 = O_P(\frac{1}{T})$, $\min_t \sigma_{\min}(A_{t\Lambda}) \geq \underline{c} > 0$, *and* $\min_i \sigma_{\min}(A_{iF}) \geq \underline{c} > 0$.

(ii) $\max_s \mathbb{E}\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}[A_{iF}^{-1}\xi_{1its} - \mathbb{E}(A_{iF}^{-1}\xi_{1its})]\right\|^2 \leq M$, $\max_s \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\mathbb{E}(A_{iF}^{-1}\xi_{1its})\right\| \leq M$, $\max_s \mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\xi_{1its} - \mathbb{E}(\xi_{1its})]\right\|^2) \leq M$ *and* $\max_s \frac{1}{N}\sum_{i=1}^{N}\left\|\sum_{t=1}^{T}\mathbb{E}(\xi_{1its})\right\|^2 \leq M$;

$\max_j \mathbb{E}\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}[A_{t\Lambda}^{-1}\xi_{2ijt} - \mathbb{E}(A_{t\Lambda}^{-1}\xi_{2ijt})]\right\|^2 \leq M$, $\max_j \frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\mathbb{E}(A_{t\Lambda}^{-1}\xi_{2ijt})\right\| \leq M$, $\max_j \mathbb{E}(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\xi_{2ijt} - \mathbb{E}(\xi_{2ijt})]\right\|^2) \leq M$ *and* $\max_j \frac{1}{T}\sum_{t=1}^{T}\left\|\sum_{i=1}^{N}\mathbb{E}(\xi_{2ijt})\right\|^2 \leq M$.

(iii) $\mathbb{E}||\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}A_{iF}^{-1}\xi_{it}||^2 \leq M$ *and* $\mathbb{E}||\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}A_{t\Lambda}^{-1}\xi_{it}||^2 \leq M$;

$\max_s \mathbb{E}\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}A_{iF}^{-1}\xi_{it}d_{is}\right\|^2 \leq M$ *and* $\max_s \mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_{it}d_{is}\right\|^2) \leq M$;

$\max_j \mathbb{E}\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}A_{t\Lambda}^{-1}\xi_{it}d_{jt}\right\|^2 \leq M$ *and* $\max_j \mathbb{E}(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_{it}d_{jt}\right\|^2) \leq M$.

(iv) For any i, $\frac{1}{T}\sum_{t=1}^{T} d_{it}f_t^0 f_t^{0\prime} \xrightarrow{p} \Sigma_{iF}$ *and* $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} d_{it}v_{it}f_t^0 \xrightarrow{d} N(0,\Omega_{iF})$ *for some positive definite matrices* Σ_{iF} *and* Ω_{iF}.

(v) For any t, $\frac{1}{N}\sum_{i=1}^{N} d_{it}\lambda_i^0\lambda_i^{0\prime} \rightarrow \Sigma_{t\Lambda}$ *and* $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} d_{it}v_{it}\lambda_i^0 \xrightarrow{d} N(0,\Omega_{t\Lambda})$ *for some positive definite matrices* $\Sigma_{t\Lambda}$ *and* $\Omega_{t\Lambda}$.

Assumption 6 generalizes Assumption F in Bai (2003) to the missing data setting. Like Assumptions 4–5, it allows the error v_{it} to be heteroscedastic and weakly correlated across i and t. The matrix completion literature typically assumes that v_{it} is independent across i and t; see, e.g., Negahban and Wainwright (2012), Chen et al. (2019), Xia and Yuan (2021), Zhu et al. (2022), Bhattacharya and Chatterjee (2022) and Chernozhukov et al. (2023). In this sense, Assumptions 4–6 extend the matrix completion literature from the setup of strict factor models to that of approximate factor models, which is more suitable for asset pricing, economic forecasting and other non-experimental settings. For example, the asset returns may reflect the risk premium of both strong factors as defined in Assumption 1 and weak factors where the latter enter the error terms generating weak cross-sectional dependence.

4.2 Asymptotic Properties of the PMLEs

Now we are ready to formally present the asymptotic results.

Theorem 4.1 ***(Preliminary Consistency):*** *Suppose that Assumptions 1–4 hold. Then as* $(N,T) \rightarrow \infty$, $\frac{1}{\sqrt{N}}\left\|\hat{\lambda} - \lambda^0\right\| = O_p(\frac{1}{\sqrt{c_{NT}}})$ *and* $\frac{1}{\sqrt{T}}\left\|\hat{f} - f^0\right\| = O_p(\frac{1}{\sqrt{c_{NT}}})$.

Theorem 4.1 implies that the convergence rates of $\hat{\lambda}_i - \lambda_i^0$ and $\hat{f}_t - f_t^0$ are $O_p(\frac{1}{\sqrt{c_{NT}}})$ on average. Although the rate $O_p(\frac{1}{\sqrt{c_{NT}}})$ is not sharp, it is established allowing v_{it} to be weakly correlated across i and t. More importantly, once we plug the above results back into equation (3.9), we can obtain $\frac{1}{\sqrt{N}}\left\|\hat{\lambda} - \lambda^0\right\| = O_p(\frac{1}{c_{NT}})$ and $\frac{1}{\sqrt{T}}\left\|\hat{f} - f^0\right\| = O_p(\frac{1}{c_{NT}})$ as stated in Theorem 4.2 below. The latter rate is sharp and as accurate as the result of Bai (2003) for the complete data case.

Theorem 4.2 ***(Average Convergence Rate):*** *Suppose that Assumptions 1–5 hold. Then as* $(N,T) \to \infty$, $\frac{1}{\sqrt{N}}\left\|\hat{\lambda} - \lambda^0\right\| = O_p(\frac{1}{c_{NT}})$ *and* $\frac{1}{\sqrt{T}}\left\|\hat{f} - f^0\right\| = O_p(\frac{1}{c_{NT}})$.

Existing results in the matrix completion literature typically assume that v_{it} is independent across i and t, and the best rate proved (or implied) by these results is $O_p(\frac{\sqrt{\log(N+T)}}{c_{NT}})$; see, e.g., corollary 1 of Negahban and Wainwright (2012), Theorem 2 of Athey et al. (2021), and Theorem 1 of Zhu et al. (2022). Our two-step proof strategy, viz., first establishing the initial rate in Theorem 4.1 and then using equation (3.9) to refine the rate, allows us to establish the sharp rate in Theorem 4.2 even when v_{it}'s are weakly dependent across i and t.

Theorem 4.2 could be useful for characterizing the effect of using estimated factors or loadings as regressors in subsequent vector autoregression or forecasting equations, and the rate $O_p(\frac{1}{c_{NT}})$ is crucial to show that such effect is asymptotically negligible (e.g., Bai and Ng (2006)). In the current context, as discussed in Section 3, the rate $O_p(\frac{1}{c_{NT}})$, combined with equations (3.5)–(3.8), allows us to show that $[H_{\phi\phi'}^{-1}R_\phi]_i$ (the higher order term in the expansion of $\hat{\lambda}_i - \lambda_i^0$) equals $\left\|\hat{\lambda}_i - \lambda_i^0\right\| O_p(\frac{1}{c_{NT}}) + O_p(\frac{1}{c_{NT}^2})$, and is asymptotically negligible if $\sqrt{T}/N \to 0$.

Remark 4.1 ***(Convergence of the EM algorithm in Section 2.2)*** *Based on the structure of* $H_{\phi\phi'}(\phi)$ *presented in Appendix B, it is not difficult to show that there exist* $m > 0$ *and* $C > 0$ *such that* $\min_{\phi \in \mathcal{N}_m(\phi^0)} \sigma_{\min}(-D_{TN}^{-\frac{1}{2}} H_{\phi\phi'}(\phi) D_{TN}^{-\frac{1}{2}}) \geq C$ *w.p.a.1 as* $(N,T) \to \infty$, *where* $\sigma_{\min}(\cdot)$ *denotes the smallest eigenvalue and* $\mathcal{N}_m(\phi^0) \equiv \{\phi \in R^{(N+T)r} : ||D_{NT}^{-\frac{1}{2}}(\phi - \phi^0)|| \leq m\}$. *This implies that in the region* $\mathcal{N}_m(\phi^0)$, *the criterion function* $Q(\phi)$ *is concave and there exists a unique local maximum. By design, the initial value in step (1),* $\tilde{\phi} = (\tilde{\lambda}', \tilde{f}')'$, *lies in* $\mathcal{N}_m(\phi^0)$ *w.p.a.1, and it is well-known that the EM algorithm converges to the local maximum. Then the EM algorithm in step (2) would converge to the local maximum in* $\mathcal{N}_m(\phi^0)$. *In addition, Theorem 4.1 implies that the global maximum* $\hat{\phi} = (\hat{\lambda}', \hat{f}')'$ *lies in* $\mathcal{N}_m(\phi^0)$ *w.p.a.1. Then the local maximum in* $\mathcal{N}_m(\phi^0)$ *is also the global maximum and the EM algorithm in step (2) converges to the global maximum* $\hat{\phi}$.

Theorem 4.3 ***(Limit Distributions):*** *Suppose that Assumptions 1–6 hold. Then as* $(N,T) \to \infty$,

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i^0) \xrightarrow{d} \mathcal{N}(0, \Sigma_{iF}^{-1}\Omega_{iF}\Sigma_{iF}^{-1}) \text{ if } \sqrt{T}/N \to 0,$$
$$\sqrt{N}(\hat{f}_t - f_t^0) \xrightarrow{d} \mathcal{N}(0, \Sigma_{t\Lambda}^{-1}\Omega_{t\Lambda}\Sigma_{t\Lambda}^{-1}) \text{ if } \sqrt{N}/T \to 0.$$

Theorem 4.3 allows us to construct confidence intervals for the factors and loadings. This is useful since in various applications the factors represent economic indices and the loadings measure the exposure of stock or bond returns to risk factors. Note that the limit distribution of $\hat{\lambda}_i$ (resp. $\hat{f}_t$) is the same as if one runs the least squares regression $d_{it}y_{it}$ on $d_{it}f_t^0$ (resp. $d_{it}\lambda_i^0$), i.e., as if the factors (resp. loadings) were observable. The effect of using estimated factors (resp. loadings) is asymptotically negligible when $\sqrt{T}/N \to 0$ (resp. $\sqrt{N}/T \to 0$).

To make inferences on λ_i^0 and f_t^0, one needs to consistently estimate Σ_{iF}, $\Sigma_{t\Lambda}$, Ω_{iF}, and $\Omega_{t\Lambda}$. It is easy to see that we can estimate Σ_{iF} and $\Sigma_{t\Lambda}$ consistently by $\hat{\Sigma}_{tF} = \frac{1}{T}\sum_{t=1}^{T} d_{it}\hat{f}_t\hat{f}_t'$ and $\hat{\Sigma}_{t\Lambda} = \frac{1}{N}\sum_{i=1}^{N} d_{it}\hat{\lambda}_i\hat{\lambda}_i'$, respectively. Ω_{iF} can be also estimated by using the standard heterokedasticity-and-autocorrelation-consistent (HAC) formula with $d_{it}v_{it}f_t^0$ replaced by $d_{it}\hat{v}_{it}\hat{f}_t$, where $d_{it}\hat{v}_{it} = d_{it}(y_{it} - \hat{\lambda}_i'\hat{f}_t)$. For $\Omega_{t\Lambda}$, the estimation depends on whether we allow for cross-sectional correlation among $\left\{d_{it}v_{it}\lambda_i^0\right\}$. In the special case where $d_{it}v_{it}\lambda_i^0$ are uncorrelated across i, we have

$$\Omega_{t\Lambda} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}(d_{it}v_{it}^2\lambda_i^0\lambda_i^{0\prime}),$$

and then we can estimate $\Omega_{t\Lambda}$ by $\hat{\Omega}_{t\Lambda} = \frac{1}{N}\sum_{i=1}^{N} d_{it}\hat{v}_{it}^2\hat{\lambda}_i\hat{\lambda}_i'$. When cross-sectional correlations are present, we can follow Bai and Ng (2006) to estimate $\Omega_{t\Lambda}$ consistently.

Remark 4.2 ***(Mixed frequency)*** *Theorems 4.2–4.3 establish the asymptotic properties for the least squares estimator of large dimensional mixed frequency factor models via minimizing* $\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}(y_{it} - f_t'\lambda_i)^2$. *For the case of complete data, it is well-known that the PCA estimation is equivalent to the least squares estimation and the relevant asymptotic theory is well-studied in Bai (2003). However, the equivalence no longer applies for mixed frequency factor models, and consequently how to establish the asymptotic theory for the least squares estimation of mixed frequency factor models is a well-recognized yet unsolved problem. A popular method for mixed frequency time series is to aggregate the high frequency time series into low frequency series (e.g., Andreou, Gagliardini, Ghysels, and Rubin (2019)), but this is not efficient. Theorems 4.2–4.3 constitute the first theory that systematically solves this important problem.*

Remark 4.3 ***(Generality and efficiency)*** *Theorems 4.2–4.3 also show that the least squares estimator is more general, and it is either more efficient than or as efficient as the methods of Bai and Ng (2021), Xiong and Pelger (2023), Jin et al. (2021) and Chernozhukov et al. (2023). Bai and Ng (2021) provide an excellent practical solution for block missing (Examples 3–5), and their estimated factors and loadings are good enough after just one iteration. Our results show that if we use their estimates as initial values and iterate until convergence, we actually obtain the least squares estimates, and the least squares estimators are as asymptotically efficient as their estimators. Xiong and Pelger's (2023) method is quite general and applies to Examples 1–6 except that they only allow* $\mathbb{E}_\phi(d_{it})$ *to be correlated with* f_t^0 *or* λ_i^0 *but not both, and they also require* $\mathbb{E}(f_t^0 f_t^{0\prime})$ *be stable over time. Their estimator is less efficient than the least squares estimator due to the inverse observation-proportion weighting, but we can obtain the least squares estimate by using their estimate as the initial value and iterating until convergence. The EM algorithm has always been popular for dealing with missing data for factor models since Stock and Watson (2002). Jin et al. (2021) establish the asymptotic theory for the EM algorithm under homogenous random missing, which is more restrictive than Example 1. Our results provide the asymptotic theory for the EM algorithm under Examples 1–6, i.e., the EM algorithm is actually asymptotically valid for a very wide range of missing patterns. Chernozhukov et al. (2023) propose a debiasing procedure and rigorous theory for post nuclear norm regularization inference based on sample splitting. Their method allows* $\mathbb{E}_\phi(d_{it})$ *to vary across* i *or* t *but not both, which is more restrictive than Example 1.*

5 Applications: Average Treatment Effect Estimation and Factor-Augmented Regression

The asymptotic expansion in (3.1) also allows us to characterize the effect of using estimated factors/loadings on the limit distributions of the estimated average treatment effects or the estimated parameters of factor-augmented regressions. This section focuses on these two important applications.

5.1 Estimation of Average Treatment Effect

To proceed, we add the following assumption.

Assumption 7 *Let* $a = (a_1, ..., a_N)'$ *and* $b = (b_1, ..., b_T)'$ *denote some nonrandom or exogenous weighting vectors such that* $\|a\| = O_p(\sqrt{N})$ *and* $\|b\| = O_p(\sqrt{T})$. *Let* $\xi_{a,it} = d_{it} v_{it} f_t^0 a_i$ *and* $\xi_{b,it} = d_{it} v_{it} \lambda_i^0 b_t$.

(i) $\mathbb{E}\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\zeta_{it}\right\|^2 \leq M$ *for* $\zeta_{it} = A_{iF}^{-1}\xi_{a,it}$ *and* $A_{t\Lambda}^{-1}\xi_{b,it}$;
(ii) for any i, $\frac{1}{\sqrt{b'b}}\sum_{t=1}^{T} b_t v_{it} \xrightarrow{d} \mathcal{N}(0, \Omega_{ib})$ *for some* Ω_{ib};
(iii) for any t, $\frac{1}{\sqrt{a'a}}\sum_{i=1}^{N} a_i v_{it} \xrightarrow{d} \mathcal{N}(0, \Omega_{ta})$ *for some* Ω_{ta}.

Assumption 7 is similar to Assumption 6(iii), and it also allows the error v_{it} to be heteroscedastic and weakly correlated across i and t. Based on Assumption 7 and (3.1), we are able to prove the following theorem under the missing patterns in Examples 1–6.

Theorem 5.1 ***(Weighted convergence)****: Under Assumptions 1–7, as* $(N,T) \to \infty$,

$$\frac{1}{N}(\hat{\Lambda} - \Lambda^0)'a = O_p(\frac{1}{c_{NT}^2}) \text{ and } \frac{1}{T}(\hat{F} - F^0)'b = O_p(\frac{1}{c_{NT}^2}).$$

Now we apply Theorem 5.1 to analyze the average treatment effect. Let $y_{it}(1)$ and $y_{it}(0)$ denote the potential outcome of unit i at time t with and without treatment, respectively. The individual treatment effect is

$$\tau_{it} = y_{it}(1) - y_{it}(0) \text{ for } i > N_o \text{ and } t > T_{oi},$$

where N_o is the number of units that never get treated, and unit $i > N_o$ receives treatment from period $T_{oi}+1$ to the end of the sample period. We consider the case where the N units can be divided into K groups and all units in group k get treated from period T_k+1. Let $N_k = |\{i \in [N] : T_{oi} = T_k\}|$ with $|\cdot|$ denoting the cardinality of a set. Then the average treatment effect for group k at time $t > T_k$ is

$$ATT_{kt} = \frac{1}{N_k}\sum\nolimits_{i:T_{oi}=T_k} \tau_{it},$$

and the average treatment effect over t for unit $i > N_o$ is $ATT_i = \frac{1}{T-T_{oi}}\sum_{t=T_{oi}+1}^{T}\tau_{it}$.

Various methods have been proposed to estimate the treatment effects. As discussed in Athey et al. (2021) and Bai and Ng (2021), both the synthetic control approach and the unconfoundedness approach can be studied from a factor model perspective. Let x_{it} denote a vector of exogenous covariates. Following the literature, we assume

$$y_{it} = \tau_{it}(1 - d_{it}) + f_t^{0\prime}\lambda_i^0 + x_{it}'\beta^0 + \varepsilon_{it}, \tag{5.1}$$

where ε_{it} denotes the error term, $f_t^{0\prime}\lambda_i^0$ denotes the interactive fixed effects (IFEs), and β^0 is a vector of the slope coefficients of x_{it}. Note that here $\{(i,t) : d_{it} = 1\}$ and $\{(i,t) : d_{it} = 0\}$ are considered as the control group and treatment group, respectively. As a result, we have

$$y_{it} = y_{it}(0) \cdot 1\{d_{it} = 1\} + y_{it}(1) \cdot 1\{d_{it} = 0\},$$

where $y_{it}(0) = f_t^{0\prime}\lambda_i^0 + x_{it}'\beta^0 + \varepsilon_{it}$ and $y_{it}(1) = \tau_{it}(1-d_{it}) + y_{it}(0)$. If $\tau_{it} = \tau$ for all i and t, then (5.1) is exactly the panel data model with IFEs as studied by Bai (2009) and the least squares estimator $\hat{\tau}$ of τ is asymptotically normal. Here we allow the individual treatment effect τ_{it} to be heterogenous across both i and t, thus our model is more general than that in Bai (2009). Unlike Xiong and Pelger (2023) who model τ_{it} using the IFEs structure, we do not impose any structure on τ_{it} as in Lu, Miao and Su (2023). To estimate the treatment effects, we use the observations in the control group to impute the potential outcomes of the treated group if they were not treated. The procedure is stated in Algorithm 5.1 below.

Algorithm 5.1 ***Partial Maximum Likelihood Estimation***

1. *Obtain the estimator $\hat{\beta}$ of β^0 using the balanced part of the control group observations;*
2. *Obtain the estimators $\hat{f} = (\hat{f}_1', ..., \hat{f}_T')'$ and $\hat{\lambda} = (\hat{\lambda}_1', ..., \hat{\lambda}_T')'$ of $f^0 = \left(f_1^{0\prime}, ..., f_T^{0\prime}\right)'$ and $\lambda^0 = \left(\lambda_1^{0\prime}, ..., \lambda_T^{0\prime}\right)'$ using the observations in the control group ($y_{it} - x_{it}'\hat{\beta}$ with $d_{it} = 1$) and the algorithm in Section 2.2;*
3. *Calculate the individual treatment effect $\hat{\tau}_{it} = (y_{it} - x_{it}'\hat{\beta}) - \hat{f}_t'\hat{\lambda}_i$ for (i, t) in the treatment group.*

Note that $\hat{\tau}_{it} - \tau_{it} = -x_{it}'(\hat{\beta} - \beta^0) - (\hat{f}_t'\hat{\lambda}_i - f_t^{0\prime}\lambda_i^0) + \varepsilon_{it}$. For all $i > N_o$ and $t \geq T_{oi}$, $\hat{\tau}_{it}$ is generally inconsistent with τ_{it} due to the appearance of ε_{it}. For this reason, one typically considers the average treatment effect over i or t. Let $\tau_{\cdot t} = \frac{1}{a'a}\sum_{i=1}^N a_i\tau_{it}$ and $\tau_{i\cdot} = \frac{1}{b'b}\sum_{t=1}^T b_t\tau_{it}$ denote the average treatment effect weighted by $a = (a_1, ..., a_N)'$ and $b = (b_1, ..., b_N)'$, respectively. Define $\hat{\tau}_{\cdot t}$ and $\hat{\tau}_{i\cdot}$ analogously with τ_{it} replaced by $\hat{\tau}_{it}$.

For two scalars c_1 and c_2, $c_1 \asymp c_2$ denotes that both c_1/c_2 and c_2/c_1 are bounded away from 0 and infinity. To study the asymptotic properties of $\hat{\tau}_{\cdot t}$ and $\hat{\tau}_{i\cdot}$, we add the following assumption.

Assumption 8 *(i) $a'a \asymp N$ and $b'b \asymp T$;*
(ii) $\hat{\beta} - \beta^0 = O_p(c_{NT}^{-2})$;
(iii) $\max_{i,t}\mathbb{E}\left\|x_{it}\right\|^2 \leq M$.

The following theorem reports the asymptotic distributions of $\hat{\tau}_{\cdot t} - \tau_{\cdot t}$ and $\hat{\tau}_{i\cdot} - \tau_{i\cdot}$.

Proposition 5.1 *Suppose that Assumption 8 holds. Suppose that $\left\{d_{it}, f_t^0, \lambda_i^0, \varepsilon_{it}\right\}$ satisfy Assumptions 1, 2(ii), and 3–7 with v_{it} replaced by ε_{it}. Then as $(N, T) \to \infty$,*

$$\sqrt{N}(\hat{\tau}_{\cdot t} - \tau_{\cdot t})/\sigma_{N\tau_t} \xrightarrow{d} \mathcal{N}(0,1) \text{ if } \sqrt{N}/T \to 0, \text{ and}$$

$$\sqrt{T}(\hat{\tau}_{i\cdot} - \tau_{i\cdot})/\sigma_{T\tau_i} \xrightarrow{d} \mathcal{N}(0,1) \text{ if } \sqrt{T}/N \quad \to \quad 0,$$

where

$$\begin{aligned}
\sigma^2_{N\tau_t} &= C'_{a\Lambda}\Sigma^{-1}_{t\Lambda}\Omega_{t\Lambda}\Sigma^{-1}_{t\Lambda}C_{a\Lambda} + \frac{N}{a'a}\Omega_{ta} + 2C'_{a\Lambda}\Sigma^{-1}_{t\Lambda}\frac{1}{a'a}\sum\nolimits_{i,j=1}^{N} a_j\mathbb{E}(d_{it}\lambda_i^0\varepsilon_{it}\varepsilon_{jt}),\\
\sigma^2_{T\tau_i} &= C'_{bF}\Sigma^{-1}_{iF}\Omega_{iF}\Sigma^{-1}_{iF}C_{bF} + \frac{T}{b'b}\Omega_{ib} + 2C'_{bF}\Sigma^{-1}_{iF}\frac{1}{b'b}\sum\nolimits_{t,s=1}^{T} b_s\mathbb{E}(d_{it}f_t^0\varepsilon_{it}\varepsilon_{is}),\\
C_{a\Lambda} &= \frac{1}{a'a}\sum\nolimits_{i=1}^{N} a_i\lambda_i^0, \text{ and } C_{bF} = \frac{1}{b'b}\sum\nolimits_{t=1}^{T} b_t f_t^0.
\end{aligned}$$

The first two terms in $\sigma^2_{N\tau_t}$ (resp. $\sigma^2_{T\tau_i}$) can be consistently estimated by replacing λ_i^0, f_t^0 and ε_{it} by $\hat{\lambda}_i$, $\hat{f}_t$ and $\hat{\varepsilon}_{it}$, respectively. For the last term in $\sigma^2_{N\tau_t}$, the key is to estimate $V_\Lambda \equiv \frac{1}{a'a}\sum_{i,j=1}^{N} a_j\mathbb{E}(d_{it}\lambda_i^0\varepsilon_{it}\varepsilon_{jt})$ consistently. To do so, one can assume certain weak cross-sectional dependence condition in $\left(d_{it}, \lambda_i^0, \varepsilon_{it}\right)$. Alternatively, if one assumes that ε_{it}'s are independent over i conditional on ϕ^0, then

$$V_\Lambda = \frac{1}{a'a}\sum\nolimits_{i,j=1}^{N} a_j\mathbb{E}(d_{it}\lambda_i^0\varepsilon_{it}\varepsilon_{jt}) = \frac{1}{a'a}\sum\nolimits_{i=1}^{N} a_i\mathbb{E}(d_{it}\lambda_i^0\varepsilon_{it}^2),$$

which can be estimated consistently by its sample analogue with λ_i^0 and $d_{it}\varepsilon_{it}$ replaced by $\hat{\lambda}_i$ and $d_{it}\hat{\varepsilon}_{it}$, respectively, where $d_{it}\hat{\varepsilon}_{it} = d_{it}(y_{it} - \hat{f}_t'\hat{\lambda}_i - x_{it}'\hat{\beta})$. To estimate V_Λ under general weak cross-sectional dependence, we refer the readers directly to Bai and Ng (2006). Similarly, for the last term in $\sigma^2_{T\tau_i}$, the key is to estimate $V_F \equiv \frac{1}{b'b}\sum_{t,s=1}^{T} b_s\mathbb{E}(d_{it}f_t^0\varepsilon_{it}\varepsilon_{is})$ consistently, say by using the HAC procedure. The procedure is standard and thus omitted here for brevity.

For ATT_{kt}, we can simply take $a_i = 1$ if $T_{oi} = T_k$ and 0 if $T_{oi} \neq T_k$. Proposition 5.1 allows us to construct confidence intervals or perform hypothesis testing for the group-time average treatment effects ATT_{kt}. For example, in program evaluation studies, we want to know whether ATT_{kt} is heterogenous across groups, how ATT_{kt} evolves over the length of exposure to the treatment $t - T_k$, and what is the average of ATT_{kt} for all k and $t > T_k$.

5.2 Factor-Augmented Regressions

In this subsection, we consider the factor-augmented regression:

$$Y_{t+h} = \alpha^{0\prime} f_t^0 + \beta^{0\prime} W_t + \epsilon_{t+h}, \tag{5.2}$$

where W_t denotes the vector of exogenous variables and Y_{t+h} denotes the dependent variable at time $t+h$. Model (5.2) can be regarded as a predictive regression model when $h \geq 1$. As explained in Bai and Ng (2006), this is the diffusion index forecasting model when Y_{t+h} is a scalar; and when $h = 1$ and $Y_{t+1} = (f^{0\prime}_{t+1}, W'_{t+1})'$, α^0 and β^0 become the coefficient matrix and (5.2) becomes a factor-augmented vector autoregression model. Since f^0 is unobservable, we can use $\hat{f}$ as a proxy for f^0. Bai and Ng (2006) show that when $\hat{f}$ is estimated by the principal component analysis (PCA) and there is no missing data, using $\hat{f}$ for f^0 does not affect the limit distributions of the parameter estimates and the conditional mean of Y_{T+h} if $\sqrt{T}/N \to 0$. Theorem 5.1 allows us to extend the results of Bai and Ng (2006) to cases where $\hat{f}$ is estimated from panel data with missing observations.

Assumption 9 *Let $Z_t = (f^{0\prime}_t, W'_t)'$. $\mathbb{E}(\epsilon_{t+h} \,|Y_t, Z_t, Y_{t-1}, Z_{t-1}, \ldots) = 0$ for all $h > 0$. $\max_t \mathbb{E}\,\|Z_t\|^4 \leq M$. Z_t and ϵ_t are independent with v_{is} for all i and s. $\frac{1}{T}\sum_{t=1}^T Z_t Z'_t \xrightarrow{p} \Sigma_{ZZ} > 0$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^T Z_t \epsilon_{t+h} \xrightarrow{d} \mathcal{N}(0, \Sigma_{ZZ\epsilon})$, where $\Sigma_{ZZ\epsilon} = plim \frac{1}{T}\sum_{t=1}^T Z_t Z'_t \epsilon^2_{t+h}$.*

Assumption 9 is exactly the same as Assumption E in Bai and Ng (2006); see the discussion therein for details. Let $\omega^0 = (\alpha^{0\prime}, \beta^{0\prime})'$, $Y = (Y_{1+h}, ..., Y_T)'$, $Z = (Z_1, ..., Z_{T-h})'$, and $\epsilon = (\epsilon_{1+h}, ..., \epsilon_T)'$. Let $\hat{Z}_t = (\hat{f}'_t, W'_t)'$, where $\hat{f}$ is estimated using the incomplete panel (y_{it} with $d_{it} = 1$) and the algorithm in Section 2.2. Let $\hat{Z} = (\hat{Z}_1, ..., \hat{Z}_{T-h})'$. Let $\hat{\omega} = (\hat{\alpha}', \hat{\beta}')'$ be the least squares estimator of regressing Y on $\hat{Z}$. It follows that $Y = \hat{Z}\omega^0 + \epsilon + (F^0 - \hat{F})\alpha^0$ and

$$\hat{\omega} = (\hat{Z}'\hat{Z})^{-1}\hat{Z}'Y = \omega^0 + (\hat{Z}'\hat{Z})^{-1}[\hat{Z}'\epsilon + \hat{Z}'(F^0 - \hat{F})\alpha^0].$$

Theorem 4.2 implies that $\hat{Z}'\hat{Z} = Z'Z + O_p(\frac{T}{c_{NT}})$. If we take $b_t = \epsilon_{t+h}$ in Assumption 7, Theorem 5.1 implies that $\hat{Z}'\epsilon = Z'\epsilon + (\hat{F} - F^0)'\epsilon = Z'\epsilon + O_p(\frac{T}{c^2_{NT}})$ under Assumption 9. If take $b_t = Z_t$ in Assumption 7, then Theorem 5.1 implies $\hat{Z}'(F^0 - \hat{F}) = Z'(F^0 - \hat{F}) + (\hat{F} - F^0)'(F^0 - \hat{F}) = O_p(\frac{T}{c^2_{NT}})$ under Assumption 9. Thus we have the following theorem.

Proposition 5.2 *Suppose that Assumptions 1–7 and 9 hold. Then*

(i) As $(N, T) \to \infty$, $\sqrt{T}(\hat{\omega} - \omega) \xrightarrow{d} N(0, \Sigma^{-1}_{ZZ}\Sigma_{ZZ\epsilon}\Sigma^{-1}_{ZZ})$ if $\sqrt{T}/N \to 0$;

(ii) A consistent estimator of $\Sigma^{-1}_{ZZ}\Sigma_{ZZ\epsilon}\Sigma^{-1}_{ZZ}$ is $(\frac{1}{T}\sum_{t=1}^{T-h}\hat{Z}_t\hat{Z}'_t)^{-1}(\frac{1}{T}\sum_{t=1}^{T-h}\hat{\epsilon}^2_{t+j}\hat{Z}_t\hat{Z}'_t)(\frac{1}{T}\sum_{t=1}^{T-h}\hat{Z}_t\hat{Z}'_t)^{-1}$.

Proposition 5.2 allows us to derive the limit distributions of the conditional mean and the forecast. The conditional mean of Y_{T+h} at time T is $Y_{T+h|T} = \alpha^{0\prime} f^0_T + \beta^{0\prime} W_T$. Let $\hat{Y}_{T+h|T} = \hat{\alpha}'\hat{f}_T + \hat{\beta}' W_T$. Then $(\hat{Y}_{T+h|T} - Y_{T+h|T})/\sigma_Y \xrightarrow{d} N(0, 1)$, where $\sigma^2_Y = \frac{1}{T} Z'_T \Sigma^{-1}_{ZZ}\Sigma_{ZZ\epsilon}\Sigma^{-1}_{ZZ} Z_T + \frac{1}{N}\alpha^{0\prime}\Sigma^{-1}_{t\Lambda}\Omega_{t\Lambda}\Sigma^{-1}_{t\Lambda}\alpha^0$. Confidence intervals can be constructed accordingly. Compared with Bai and Ng (2006), Proposition 5.2 utilizes the incomplete panel data more efficiently, since we extract the factors through least squares directly rather than through aggregating the high frequency series into low frequency series or throwing away those series with missing data.

6 Simulations

In this section, we perform Monte Carlo simulations to access the adequacy of the limit distributions in approximating their finite sample counterparts and demonstrate the finite sample performance of our method. To facilitate graphical presentation for the distributions of the estimated factors and loadings, we focus on the case with one factor.

6.1 Data Generating Processes (DGPs)

The data are generated as follows. Generate f_t as *i.i.d.* $\mathcal{N}(0,1)$ for $t \in [T]$ and λ_i as *i.i.d.* $\mathcal{N}(0,1)$ for $i \in [N]$. Once the factors and loadings are independently generated, we find G (a scalar here) to normalize them such that $f_t^0 = G' f_t$, $\lambda_i^0 = G^{-1}\lambda_i$ and $\frac{1}{T}\sum_{t=1}^{T}(f_t^0)^2 = \frac{1}{N}\sum_{t=1}^{T}(\lambda_i^0)^2$. For each simulation, v_{it} is *i.i.d.* $\mathcal{N}(0,1)$ across i and t, d_{it} is generated according to the following four missing patterns, and $y_{it} = d_{it}(f_t^{0\prime}\lambda_i^0 + v_{it})$.

Pattern 1 (**Completely random heterogenous missing**): d_{it} is binary, independent across i and t and independent of f_t^0, λ_i^0 and v_{it}. $p_{it} \equiv \mathbb{E}(d_{it})$ follows *i.i.d.* *Uniform*(0.1,0.9) across i and t.

Pattern 2 (**Selection on factors and loadings**): Conditioning on ϕ^0, d_{it} is independent across i and t and independent of v_{it}; $\mathbb{E}_\phi(d_{it}) = \Phi(f_t^{0\prime}\lambda_i^0)$, where $\Phi(\cdot)$ denotes the CDF of the standard normal distribution.

Pattern 3 (**Mixed frequency**): $d_{it} = 0$ if $i > N_o = N/2$ and $t/3$ is not an integer, i.e., there are $N/2$ high frequency series and the frequency ratio is 3. In this case, we consider the mixed data with both monthly and quarterly observations.

Pattern 4 (**Staggered missing**): $N_o = 0.4N$ and $T_o = 0.4T$. $d_{it} = 0$ when (i,t) belongs to $\{N_o + 1 \leq i \leq 0.7N \text{ and } 0.7T + 1 \leq t \leq T\}$ or $\{0.7N + 1 \leq i \leq N \text{ and } T_o + 1 \leq t \leq T\}$, i.e., the first group has $0.4N$ units and never gets treated, the second group has $0.3N$ units and gets treated from $t = 0.7T + 1$ to $t = T$, and the third group has $0.3N$ units and gets treated from $t = 0.4T + 1$ to $t = T$.

The number of simulations is 2000.

6.2 Simulation Results

For all the above four patterns of missing, we use the iterative singular value thresholding (SVT) algorithm proposed in Mazumder et al. (2010) to calculate the nuclear norm regularized estimator,

$(\tilde{\lambda}, \tilde{f})$. Then we use $(\tilde{\lambda}, \tilde{f})$ as the initial value for the EM algorithm and iterate until convergence. That is, we use $(\tilde{\lambda}, \tilde{f})$ to impute the missing values and reestimate the factors and loadings by the principal components of this imputed complete matrix, and repeat this procedure until convergence.

Figures 1–4 present the histograms of standardized estimated factors at $t = T/2$ ($\hat{f}_{T/2} - f^0_{T/2}$ divided by its asymptotic standard deviation) and standardized estimated loadings at $i = N/2$ ($\hat{\lambda}_{N/2} - \lambda^0_{N/2}$ divided by its asymptotic standard deviation) for missing patterns 1–4, respectively. The standard normal density curve is overlaid on the histograms for comparison. Due to limited space, we only present the results for $(N, T) = (100, 100)$ and $(N, T) = (200, 200)$. We summarize some important findings from Figures 1–4. First, the two subfigures in the first row of Figures 1–4 are the distributions of the nuclear norm regularized estimators of the factors and loadings, respectively, when $(N, T) = (100, 100)$. Obviously, these preliminary estimators are biased and shrunk towards zero, which is due to the shrinkage effect of nuclear norm regularization. Second, the two subfigures in the second (resp. third) row of Figures 1–4 are the distributions of the NN-EM estimators of the factors and loadings, respectively, when $(N, T) = (100, 100)$ (resp. $(N, T) = (200, 200)$). The histograms in all these subfigures match very well with the standard normal density curve, although missing patterns 1–4 are quite different and the variances of the unnormalized estimation error ($\hat{f}_t - f^0_t$ and $\hat{\lambda}_i - \lambda^0_i$) also depend on (N, T). These results confirm our asymptotic results in finite samples.

Figures 5–6 graphically present the confidence intervals of the factors $\{f^0_t, t \in [50]\}$ constructed using the NN-EM estimator with $(N, T) = (100, 100)$ and $(200, 200)$, respectively. The solid curve in the middle denotes the true factor processes. Since v_{it} is *i.i.d.* $\mathcal{N}(0,1)$ in the simulations, $\Omega_{t\Lambda} = \Sigma_{t\Lambda}$ and the asymptotic variance of $\sqrt{N}(\hat{f}_t - f^0_t)$ is $\Sigma^{-1}_{t\Lambda}$, which is estimated by $\hat{\Sigma}_{t\Lambda} = \frac{1}{N}\sum_{i=1}^N d_{it}(\hat{\lambda}_i)^2$. It follows that the confidence interval for f^0_t can be constructed as $(\hat{f}_t - 1.96\frac{1}{\sqrt{N\hat{\Sigma}_{t\Lambda}}}, \hat{f}_t + 1.96\frac{1}{\sqrt{N\hat{\Sigma}_{t\Lambda}}})$ for $t \in [T]$. Comparing the results in Figures 5–6, we can see that the confidence intervals become narrower as N and T increase, and in all subfigures the true factor process is covered very well.

To evaluate the accuracy of the factor estimates, we report in Table 1 the correlation coefficients between the true factors and the factors estimated by NN and NN-EM, averaged over 2000 simulations. We can see that the correlation coefficients are all close to one, and the improvement is obvious when we compare NN with NN-EM or when N and T increase.

Tables 2 and 3 report the root mean squared errors (RMSEs) of the estimated factors and the estimated loadings (i.e., $\{\frac{1}{T}\sum_{t=1}^T(\hat{f}_t - f^0_t)^2\}^{1/2}$ and $\{\frac{1}{N}\sum_{i=1}^N(\hat{\lambda}_i - \lambda^0_i)^2\}^{1/2}$), averaged over 2000 simulations. The improvement of NN-EM over NN is also quite obvious. Moreover, the RMSEs of NN-EM are all very close to the theoretical standard deviations. For example, for pattern 3 in Table

Table 1: Average Correlation Coefficients of the Estimated Factors

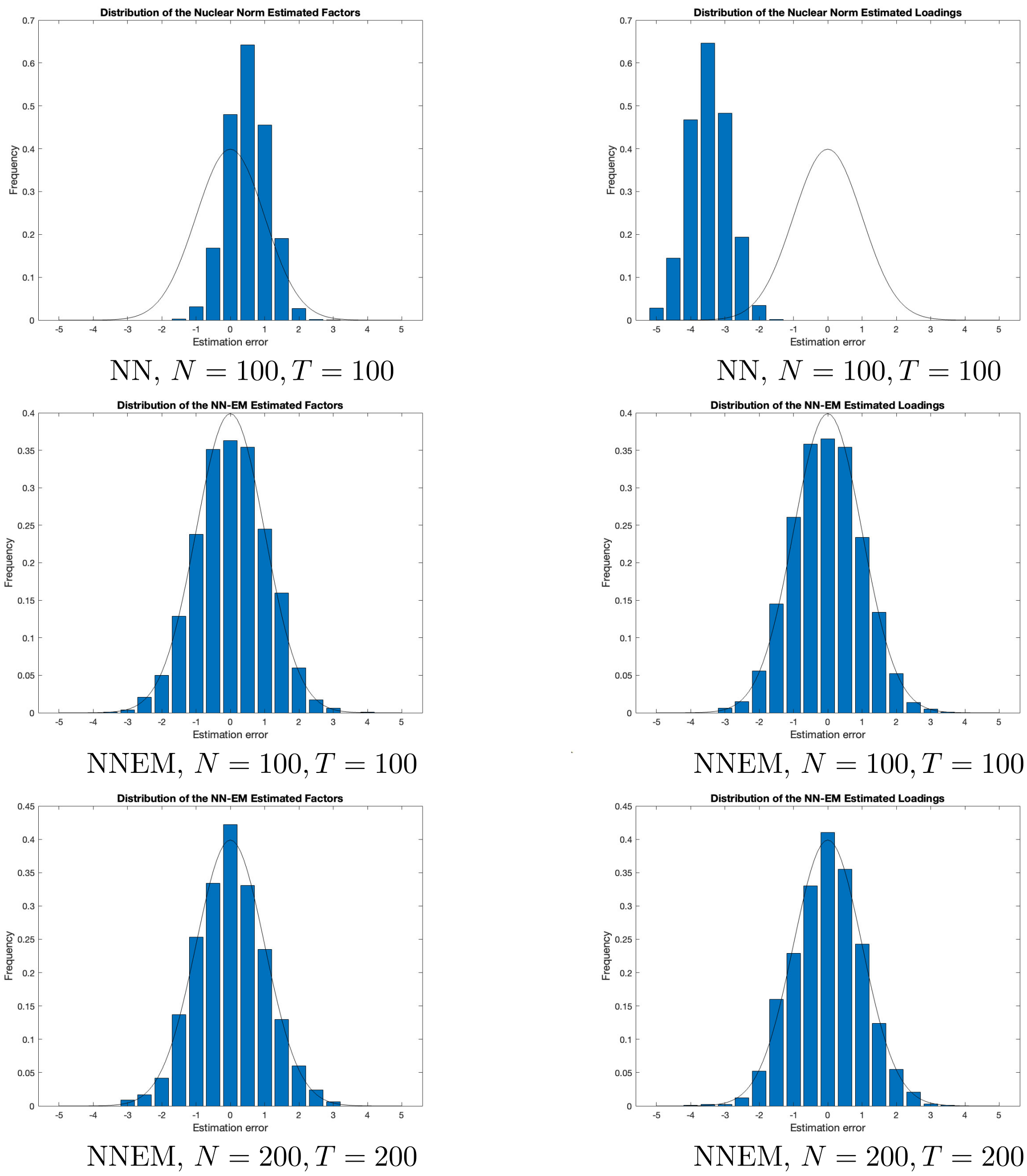
		Pattern 1		Pattern 2		Pattern 3		Pattern 4	
N	T	NN	NN-EM	NN	NN-EM	NN	NN-EM	NN	NN-EM
50	100	0.972	0.982	0.973	0.980	0.981	0.987	0.951	0.978
100	100	0.987	0.990	0.989	0.989	0.982	0.992	0.977	0.992
200	200	0.994	0.995	0.996	0.996	0.980	0.995	0.980	0.996
400	200	0.996	0.997	0.997	0.997	0.990	0.998	0.984	0.998

Notes: These are the correlation coefficients between the true factors and the factors estimated by NN or NN-EM, averaged over 2000 simulations.

2, "h/l" denotes the RMSEs of the estimated factors of integer periods and non-integer periods, respectively, viz, $\frac{1}{T/3}\sum_{t/3=\text{integer}}(\hat{f}_t - f_t^0)^2\}^{1/2}$ and $\{\frac{1}{2T/3}\sum_{t/3\neq\text{integer}}(\hat{f}_t - f_t^0)^2\}^{1/2}$. The effective sample size for the estimated factors in the non-integer periods and integer periods is N_o and N, respectively, since only the high frequency series are observable at non-integer periods. For pattern 3 in Table 3, "h/l" denotes the RMSEs of the estimated loadings of high frequency units and low frequency units, respectively, viz., $\{\frac{1}{N_o}\sum_{i=1}^{N_o}(\hat{\lambda}_i - \lambda_i^0)^2\}^{1/2}$ and $\{\frac{1}{N-N_o}\sum_{i=N_o+1}^{N}(\hat{\lambda}_i - \lambda_i^0)^2\}^{1/2}$. Note that the effective sample size for the estimated loadings of high frequency units and low frequency units is T and $T/3$, respectively. As we can tell from Tables 2–3, the RMSEs decrease as N and T increase at rates as predicted by the theory.

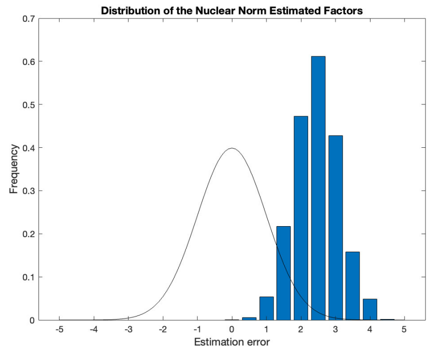
We next consider factor-augmented regressions and average treatment effect estimation. For factor-augmented regression, we generate the data by $Y_{t+1} = \alpha^0 f_t^0 + \beta^0 W_t + \epsilon_{t+1}$, where $\alpha^0 = \beta^0 = 1$, W_t is *i.i.d.* $\mathcal{N}(0,1)$, and ϵ_{t+1} is *i.i.d.* $\mathcal{N}(0,1)$. The estimated conditional mean is $\hat{Y}_{T+1|T} = \hat{\alpha}\hat{f}_T + \hat{\beta}W_T$, and the 95% confidence interval for $Y_{T+1|T}$ is $(\hat{Y}_{T+1|T} - 1.96\hat{\sigma}_Y, \hat{Y}_{T+1|T} + 1.96\hat{\sigma}_Y)$ where $\hat{\sigma}_Y$ is a consistent estimate of $\sigma_Y = \sqrt{\sigma_Y^2}$ with σ_Y^2 defined below Proposition 5.2. For patterns 3–4, we generate the individual treatment effects τ_{it} as *i.i.d.* *Uniform*(0.1,0.5) across i and t. Then the true average treatment effect at $t = T$ is $\tau_{\cdot T} = \frac{1}{N-N_o}\sum_{i=N_o+1}^{N}\tau_{iT}$ and the 95% confidence interval for τ_T is $(\hat{\tau}_T - 1.96\hat{\sigma}_{\tau_T}, \hat{\tau}_T + 1.96\hat{\sigma}_{\tau_T})$, where $\hat{\sigma}_{\tau_T}$ is a consistent estimator of $\sigma_{\tau_T} = \sqrt{\sigma_{\tau_T}^2}$ with $\sigma_{\tau_T}^2$ defined in Proposition 5.1. Table 4 reports the coverage rates for the 95% confidence intervals of $Y_{T+1|T}$ and τ_T under different patterns and different combinations of N and T based on 2000 replications. As we can tell from Table 4, the coverage rate is close to the nominal level 95% in all cases. This demonstrates the validity of using the NN-EM estimated factors and loadings for forecasting and treatment effect estimation.

Figure 1: Distributions of the Estimated Factors and Loadings: Pattern 1

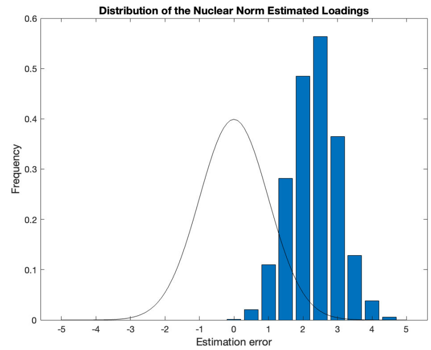


Notes: These are the histograms of the standardized estimated factors at $t = T/2$ and the standardized estimated loadings at $i = N/2$. The results are based on 2,000 simulations. The curve overlaid on the histograms is the standard normal density function.

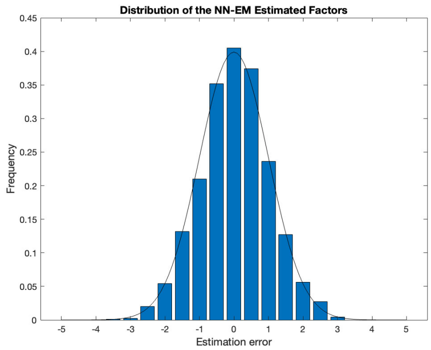
Figure 2: Distributions of the Estimated Factors and Loadings: Pattern 2



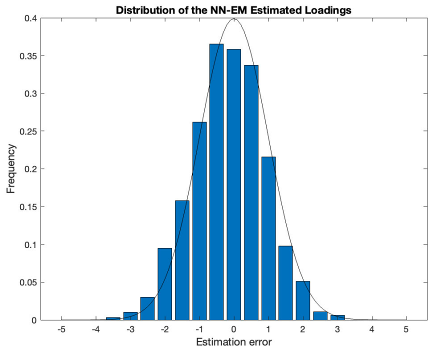
NN, $N = 100, T = 100$



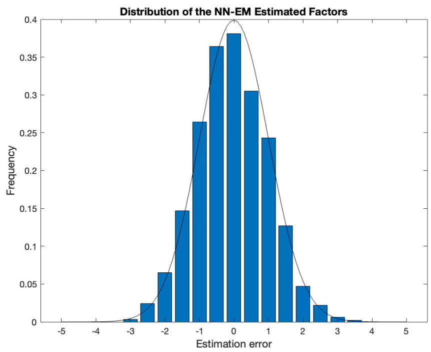
NN, $N = 100, T = 100$



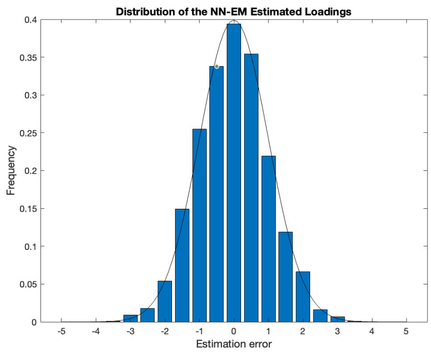
NNEM, $N = 100, T = 100$



NNEM, $N = 100, T = 100$



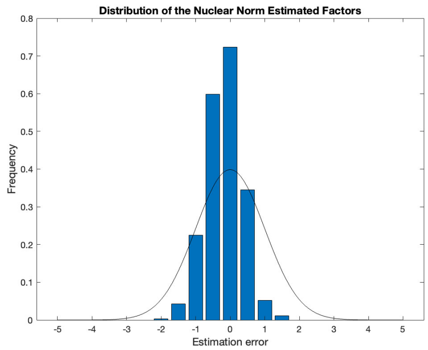
NNEM, $N = 200, T = 200$



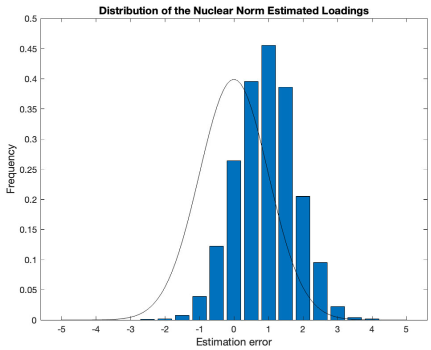
NNEM, $N = 200, T = 200$

Notes: These are the histograms of the standardized estimated factors at $t = T/2$ and the standardized estimated loadings at $i = N/2$. The results are based on 2,000 simulations. The curve overlaid on the histograms is the standard normal density function.

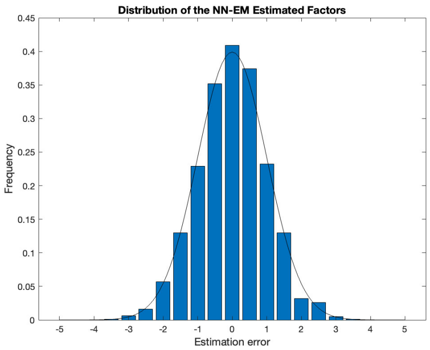
Figure 3: Distributions of the Estimated Factors and Loadings: Pattern 3



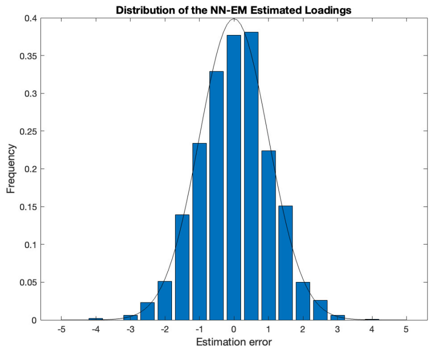
NN, $N = 100, T = 100$



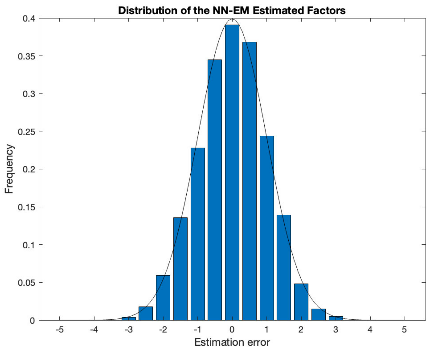
NN, $N = 100, T = 100$



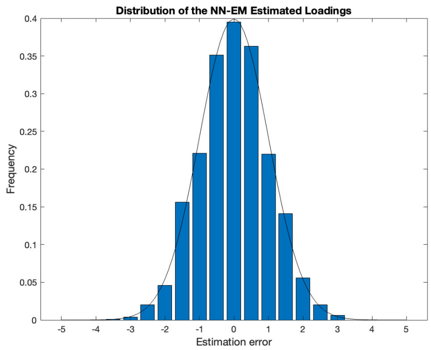
NNEM, $N = 100, T = 100$



NNEM, $N = 100, T = 100$



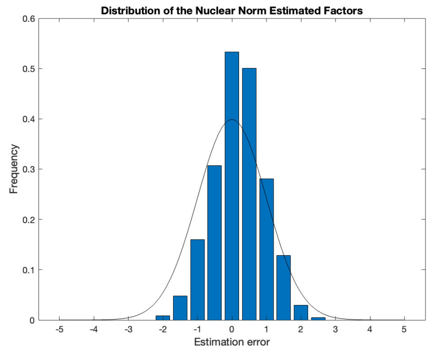
NNEM, $N = 200, T = 200$



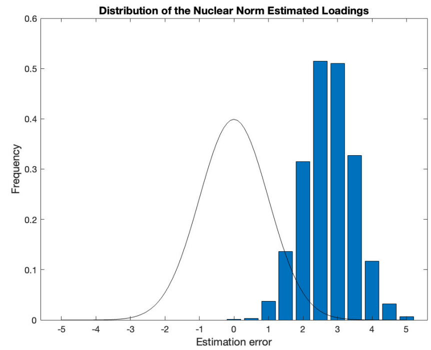
NNEM, $N = 200, T = 200$

Notes: These are the histograms of the standardized estimated factors at $t = T/2$ and the standardized estimated loadings at $i = N/2$. The results are based on 2,000 simulations. The curve overlaid on the histograms is the standard normal density function.

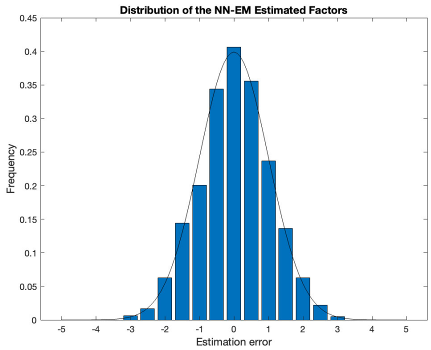
Figure 4: Distributions of the Estimated Factors and Loadings: Pattern 4



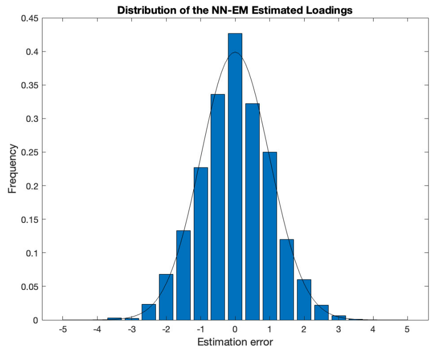
NN, $N = 100, T = 100$



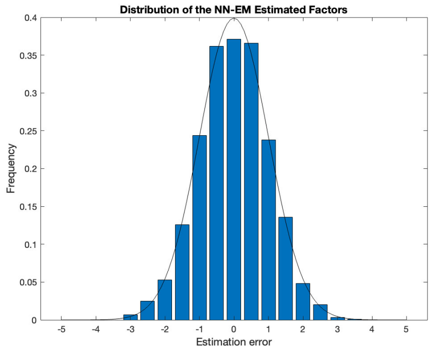
NN, $N = 100, T = 100$



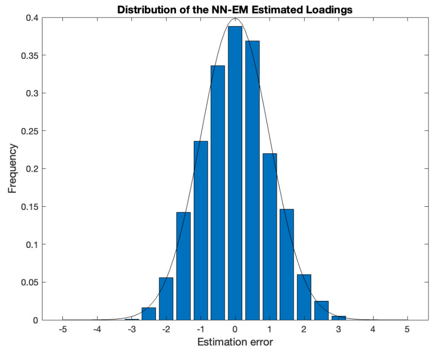
NNEM, $N = 100, T = 100$



NNEM, $N = 100, T = 100$



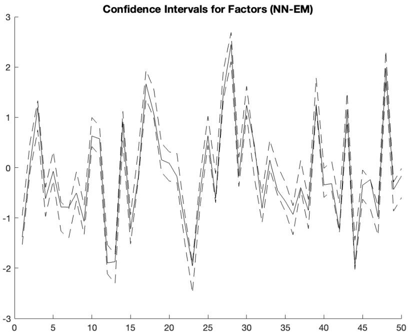
NNEM, $N = 200, T = 200$



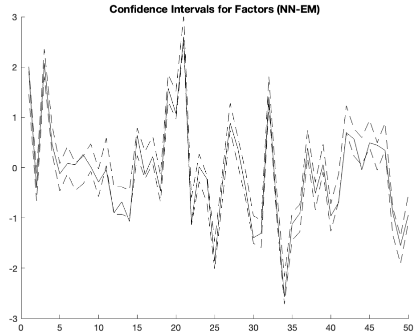
NNEM, $N = 200, T = 200$

Notes: These are the histograms of the standardized estimated factors at $t = T/2$ and the standardized estimated loadings at $i = N/2$. The results are based on 2,000 simulations. The curve overlaid on the histograms is the standard normal density function.

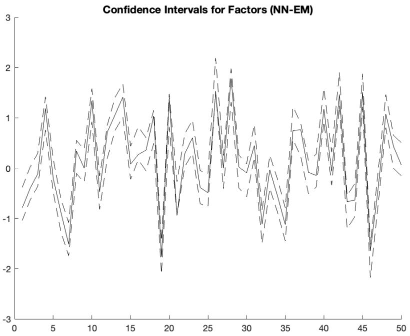
Figure 5: Confidence Intervals for the Factors by NNEM



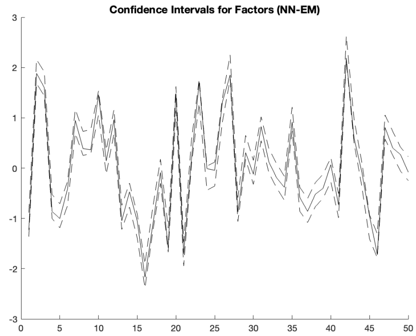
Pattern 1, $N = 100, T = 100$



Pattern 2, $N = 100, T = 100$



Pattern 3, $N = 100, T = 100$



Pattern 4, $N = 100, T = 100$

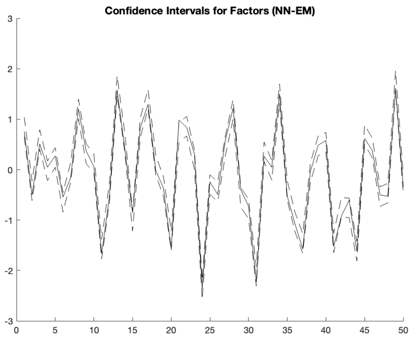
Notes: These are the 95% confidence intervals for the true factor process from $t = 1$ to $t = 50$. The confidence intervals are calculated by NN-EM. The solid curve in the middle is the true factor process.

Table 2: Root Mean Squared Error of the Estimated Factors

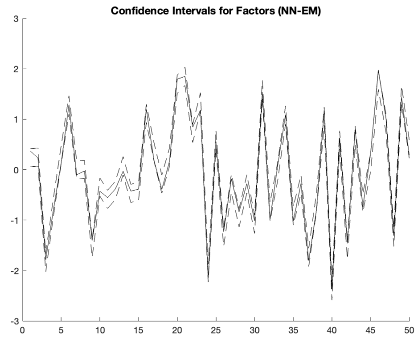
		Pattern 1		Pattern 2		Pattern 3		Pattern 4	
N	T	NN	NN-EM	NN	NN-EM	NN(h/l)	NN-EM(h/l)	NN	NN-EM
50	100	0.483	0.200	0.500	0.205	0.223/0.406	0.136/0.182	0.409	0.203
100	100	0.384	0.144	0.403	0.146	0.158/0.425	0.098/0.143	0.322	0.127
200	200	0.315	0.099	0.294	0.095	0.131/0.374	0.071/0.110	0.300	0.092
400	200	0.325	0.072	0.333	0.075	0.114/0.319	0.048/0.068	0.277	0.065

Notes: These are the root mean squared errors of the estimated factors calculated by NN or NN-EM averaged over 2000 simulations. For pattern 3 (mixed frequency), "h/l" denotes the root mean squared error of the estimated factors of integer periods and non-integer periods, respectively. Integer periods are those periods when both the high frequency series and the low frequency series are observable.

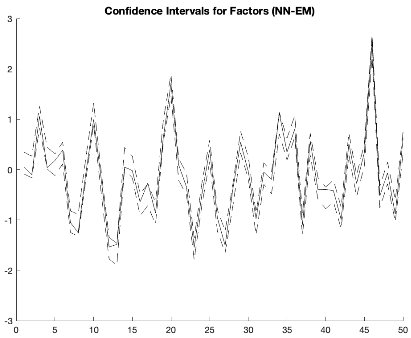
Figure 6: Confidence Intervals for the Factors by NNEM



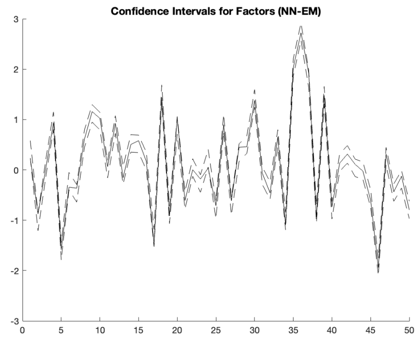
Pattern 1, $N = 200, T = 200$



Pattern 2, $N = 200, T = 200$



Pattern 3, $N = 200, T = 200$



Pattern 4, $N = 200, T = 200$

Notes: These are the 95% confidence intervals for the true factor process from $t = 1$ to $t = 50$. The confidence intervals are calculated by NN-EM. The solid curve in the middle is the true factor process.

Table 3: Root Mean Squared Error of the Estimated Loadings

		Pattern 1		Pattern 2		Pattern 3		Pattern 4	
N	T	NN	NN-EM	NN	NN-EM	NN(h/l)	NN-EM(h/l)	NN	NN-EM
50	100	0.461	0.141	0.482	0.148	0.199/0.548	0.098/0.180	0.360	0.125
100	100	0.382	0.142	0.421	0.149	0.14/0.614	0.098/0.199	0.315	0.125
200	200	0.316	0.099	0.295	0.094	0.114/0.458	0.072/0.119	0.292	0.090
400	200	0.333	0.101	0.343	0.105	0.111/0.418	0.068/0.112	0.266	0.086

Notes: These are the root mean squared errors of the estimated loadings calculated by NN or NN-EM averaged over 2000 simulations. For pattern 3 (mixed frequency), "h/l" denotes the root mean squared error of the estimated loadings of the high frequency units and the low frequency units, respectively.

Table 4: Coverage Rates of Confidence Intervals

		Pattern 1	Pattern 2	Pattern 3		Pattern 4	
N	T	$\hat{Y}_{T+1\|T}$	$\hat{Y}_{T+1\|T}$	$\hat{Y}_{T+1\|T}$	$\hat{\tau}_T$	$\hat{Y}_{T+1\|T}$	$\hat{\tau}_T$
50	100	0.955	0.950	0.957	0.945	0.957	0.956
100	100	0.950	0.952	0.954	0.932	0.956	0.948
200	200	0.953	0.953	0.956	0.944	0.958	0.950
400	200	0.942	0.947	0.954	0.938	0.949	0.951

Notes: These are the coverage rates of 95% confidence intervals for the conditional mean and the average treatment effect. The factors and loadings are estimated by NN-EM.

7 Empirical Illustration

In this section we apply our method to the grant allocation data of Fouirnaies and Mutlu-Eren (2015) to test the average treatment effects of partisan alignment.

7.1 Data

Fouirnaies and Mutlu-Eren (2015) argue that in England the government parties have incentives to allocate more resources to local councils that are controlled by their own party, since local governments are mainly funded by the central government and voters' assessment of the party at the local level has spillover effects on the assessment of the party at other levels of government. Fouirnaies and Mutlu-Eren (2015) collect the data of the partisan control of each local council and the specific grants per capita allocated to each local council for 460 local councils in England from 1992 to 2012, and they find that partisan alignment indeed brings local councils more resources and the alignment effect peaks in the third year after alignment.

In this application, the outcome variable y_{it} is the logarithm of specific grants per capita allocated to a local council i at period t, with $(N,T) = (460, 21)$. At time t, council i is considered as treated ($d_{it} = 0$) if the government party controls the majority of council i. Since both y_{it} and d_{it} have missing observations for some (i,t) and we do not know whether council i at period t is treated or not if d_{it} is missing, we focus on the data when y_{it} is observed and $d_{it} = 1$ (untreated).

7.2 Estimation

Given $\{y_{it}, d_{it}\}$, we can use the data with $d_{it} = 1$ to estimate the factors and loadings. Once we have the estimated factors and loadings ($\hat{f}_t$ and $\hat{\lambda}_i$), we impute the untreated potential outcome of the treated y_{it} (i.e., $d_{it} = 0$) by $\hat{y}_{it}(0) = \hat{\lambda}_i' \hat{f}_t$, and the individual treatment effect is estimated by $\hat{\tau}_{it} = y_{it} - \hat{y}_{it}(0)$ for all (i,t) with $d_{it} = 0$. Regarding the number of factors r, for each r we randomly cover y_{it} with probability 0.2, estimate the model using the uncovered y_{it}, and then use the estimated factors and loadings to impute the covered y_{it} and calculate the out-of-sample RMSE. This procedure is repeated 50 times and the average out-of-sample RMSE for $r = 1$, 2, 3, 4 and 5 is 0.454, 0.331, 0.332, 0.437 and 0.555, respectively. Based on the cross-validation method of Jin et al. (2021) for the determination of the number of factors, we can estimate r by 2, which yields the smallest out-of-sample RMSE. As a robustness check, we shall focus on the results for $r = 2$, 3 and 4.

To see how the partisan alignment treatment effect evolves over time, we group $\hat{\tau}_{it}$ according to the number of periods relative to the onset of the treatment and then we calculate the average treatment effect of each group. For example, if council 1 starts treatment at $t = 5$, council 3 starts treatment at $t = 9$ and council 5 starts treatment at $t = 3$, then $\hat{\tau}_{15}$, $\hat{\tau}_{39}$ and $\hat{\tau}_{53}$ are in the same group. Let k denote the number of periods relative to the onset of the treatment and $\widehat{ATT}_k$ denote the corresponding group average of the treatment effects. An advantage of $\widehat{ATT}_k$ is that it clearly shows the dynamics of the average treatment effect over time while allowing the individual treatment effect to be different across both individuals and time.

The top-left block of Table 5 presents $\widehat{ATT}_k$ and its t-statistic for $k = 1, 2, 3, 4$ and $r = 2, 3, 4$. It is clear that the t-statistic of $\widehat{ATT}_k$ is significant at the 5% level in all cases, and $\widehat{ATT}_k$ increases with k initially and peaks at $k = 3$, which is consistent with the finding of Fouirnaies and Mutlu-Eren (2015). The latter authors consider a panel regression with council-specific linear time trends and a two-way fixed effects. The fact that $\widehat{ATT}_k$ peaks at $k = 3$ shows that there is an implementation delay or the government party strategically schedules the grant boost in the electoral cycle to maximize the election effect.

The middle-left block of Table 5 presents $\widehat{ATT}_k$ and its t-statistic for $k = -1$, -2, -3 and -4. $\widehat{ATT}_{-1}$ is the average treatment effect of the last untreated period before the onset of the treatment, and $\widehat{ATT}_{-2}$, $\widehat{ATT}_{-3}$, and $\widehat{ATT}_{-4}$ are defined analogously. It is clear that $\{\widehat{ATT}_k,\ k = -1, ..., -4\}$ are all close to zeros and mostly insignificant at the 5% level, which confirms that there is no pre-trend unaccounted by the factor structure. In general, $\{\widehat{ATT}_k,\ k = -1, ...\}$ could also help the researchers to test for the anticipation effects or evaluate the validity of the identification conditions. The bottom-left block of Table 5 presents $\widehat{ATT}_k$ and its t-statistic for $k = +1$, $+2$, $+3$ and $+4$. $\widehat{ATT}_{+1}$ is the average treatment effect of the first untreated period after the end of the treatment, and $\widehat{ATT}_{+2}$, $\widehat{ATT}_{+3}$, and $\widehat{ATT}_{+4}$ are defined analogously. $\{\widehat{ATT}_k,\ k = +1, ..., +4\}$ may deviate from zeros if the treatment has carryover effects or there are time-varying confounders unaccounted by the factor structure. From the bottom-left block of Table 5 we can see that $\{\widehat{ATT}_k,\ k = +1, ..., +4\}$ are also close to zeros and mostly insignificant at the 5% level, especially when $r = 3$ and 4.

Since the middle-left block and the bottom-left block of Table 5 are in-sample results, we also calculate $\{\widehat{ATT}_k,\ k = -1, ..., -4\}$ and $\{\widehat{ATT}_k,\ k = +1, ..., +4\}$ using the out-of-sample imputation errors. More specifically, for each k, we cover the data of y_{it} at the model estimation stage if t is the $|k|$-th period after the end (resp., before the onset) of the treatment of unit i, and then use the estimated factors and loadings to impute the covered y_{it} and calculate the out-of-sample imputation error $\hat{\tau}_{it}$. The middle-right block and the bottom-right block of Table 5 present the out-of-sample $\widehat{ATT}_k$ and its t-statistic for $k = -1, ..., -4$ and $k = +1, ..., +4$, respectively.

We summarize some important findings from the right panel in Table 5. First, we can see that the results of $r = 3$ and $r = 4$ are consistent with each other whereas the results of $r = 2$ tend to overestimate the treatment effects. Similar patterns also appear in the other three blocks of Table 5. This suggests that $r = 2$ may underestimate the number of factors and the results of $r = 3$ and 4 are more trustworthy. Second, we can see that overall $\widehat{ATT}_{-2}$, $\widehat{ATT}_{-3}$ and $\widehat{ATT}_{-4}$ are still insignificant while $\widehat{ATT}_{-1}$ becomes significant. The fact that $\widehat{ATT}_{-1}$ is significantly positive indicates that changes in the alignment status are related to grant allocation. For example, the government party may strategically allocate more grants to some swing councils before local elections even if those councils are controlled by different parties, and then the voters in those councils switch to the government party after the local elections. In other words, on average aligned councils receive more grants from the government party than unaligned councils, but unaligned swing councils also receive more grants. Third, for the post-treatment periods, the bottom-right block of Table 5 shows that $\widehat{ATT}_{+3}$ and $\widehat{ATT}_{+4}$ are clearly insignificant, while $\widehat{ATT}_{+1}$ and $\widehat{ATT}_{+2}$ are significant when $r = 2$, marginally significant when $r = 3$ and insignificant when $r = 4$. This suggests that the carryover effects of partisan alignment are not strong even if they exist. Overall, the out-of-sample results of the pre-treatment periods and the post-treatment periods are similar to their in-sample counterparts, except for $\widehat{ATT}_{-1}$.

8 Conclusions

This paper develops an inferential theory for the least squares estimators of the factors and loadings in a large dimensional factor model with missing data. To compute the least squares estimator, this paper proposes to use the nuclear norm regularized estimator as the initial value for the EM algorithm and iterate until convergence. Our results cover a wide range of missing patterns, includ-

Table 5: Testing the Average Treatment Effects

	Average Treatment Effect Dynamics							
$r \backslash k$	1	2	3	4				
2	0.085	0.089	0.116	0.050				
	(7.315)	(7.263)	(8.579)	(2.152)				
3	0.029	0.048	0.071	-0.087				
	(2.797)	(4.442)	(5.923)	(-4.215)				
4	0.055	0.072	0.094	-0.082				
	(6.066)	(7.646)	(8.937)	(-4.526)				
	Average Treatment Effect Dynamics				Out-of-sample Pre-treatment Results			
	-4	-3	-2	-1	-4	-3	-2	-1
2	-0.022	0.003	0.019	0.037	-0.059	-0.001	0.054	0.194
	(-1.573)	(0.186)	(1.407)	(2.878)	(-4.102)	(-0.057)	(4.041)	(15.27)
3	-0.014	-0.002	0.005	0.017	0.012	-0.001	0.006	0.136
	(-1.081)	(-0.144)	(0.395)	(1.448)	(0.925)	(-0.087)	(0.528)	(12.14)
4	-0.013	0.006	0.006	0.016	-0.011	0.010	-0.006	0.154
	(-1.139)	(0.059)	(0.624)	(1.600)	(-0.978)	(0.942)	(-0.575)	(15.69)
	Average Treatment Effect Dynamics				Out-of-sample Post-treatment Results			
	+1	+2	+3	+4	+1	+2	+3	+4
2	0.025	0.045	0.012	-0.005	0.072	0.061	0.014	-0.007
	(1.629)	(2.966)	(0.710)	(-0.249)	(4.918)	(4.032)	(0.855)	(-0.378)
3	-0.009	0.019	-0.003	-0.005	0.021	0.033	-0.001	-0.011
	(-0.683)	(1.412)	(-0.210)	(-0.298)	(1.596)	(2.49)	(-0.066)	(-0.721)
4	-0.008	0.010	-0.004	-0.002	-0.010	0.022	-0.006	0.003
	(-0.688)	(0.869)	(-0.320)	(-0.129)	(-0.824)	(1.890)	(-0.489)	(0.187)

Notes: The table presents the group average treatment effects and the corresponding t statistics (presented in the parenthesis) of partisan alignment where the groups are determined by the number of periods relative to the onset (end) of the treatment. The first column indicates the number of factors. The out-of-sample results in the middle-right block and the bottom-right block are calculated by covering the corresponding period of data.

ing heterogenous random missing, selection on covariates/factors/loadings, block missing, staggered missing, mixed frequency and ragged edge. For the matrix completion literature, our results provide a solution for the post nuclear norm regularization inference under missing patterns much more general than existing studies. For mixed frequency factor models, our results provide the asymptotic theory without aggregating the high frequency series into low frequency series. For panel data with missing observations, our methods allow us to impute the missing values appropriately even when the missing probability is correlated with the missing value. For causal inference, our results provide confidence intervals for the average treatment effects of different groups and time periods.

There are some interesting topics for further research. First, we may extend our results to allow for nonstationarity in the data. Second, our framework may also cover other missing patterns as long as we can verify the RSC condition and prove that the Hessian is well-behaved. Third, it is possible to extend our method to the framework of time-varying factor models as studied in Su and Wang (2017, 2024) and Pelger and Xiong (2022), among others. Fourth, so far we focus on the pure factor models, it would be interesting to add covariates and extend our theory to panel data with IFEs. Fifth, the technique developed in this paper is of independent interest and can be extended to more general setups that include nonlinear factor models with missing values and nonlinear panels with IFEs and missing values. We leave these topics for future research.

References

[1] Agarwal, A., Dahleh, M., Shah, D., and Shen, D., 2023. Causal Matrix Completion, *The Thirty Sixth Annual Conference on Learning Theory*, pp. 3821-3826.

[2] Andreou, E., Gagliardini, P., Ghysels, E., and Rubin, M., 2019. Inference in Group Factor Models With an Application to Mixed-Frequency Data, *Econometrica* 87, 1267-1305.

[3] Aruoba, S.B., Diebold, F.X., and Scotti, C., 2009. Real-Time Measurement of Business Conditions, *Journal of Business & Economic Statistics* 27, 417-427.

[4] Athey, S., Bayati, M., Doudchenko, N., Imbens, G. and Khosravi, K., 2021., Matrix Completion Methods for Causal Panel Data Models, *Journal of American Statistical Association* 116, 1716-1730.

[5] Bai, J., 2003. Inferential Theory for Factors Models of Large Dimensions, *Econometrica* 71, 135-173.

[6] Bai, J. and Ng, S., 2006. Confidence Intervals for Diffusion Index Forecasts and Inference with Factor-Augmented Regressions, *Econometrica* 74, 1133–1150.

[7] Bai, J. and Ng, S., 2021., Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data, *Journal of American Statistical Association* 116, 1746-1763.

[8] Banbura, M. and Modugno, M., 2014. Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data, *Journal of Applied Econometrics* 29, 133–160.

[9] Belloni, A., Chernozhukov, V., and Hansen, C., 2014., Inference on Treatment Effects after Selection amongst High-Dimensional Controls, *Review of Economic Studies* 81, 608–650.

[10] Belloni, A., Chernozhukov, V., Fernandez-Val, I., and Hansen, C., 2017. Program Evaluation and Causal Inference with High Dimensional Data. *Econometrica* 85, 233-298.

[11] Bhattacharya, S. and Chatterjee, S., 2022., Matrix Completion with Data-Dependent Missingness Probabilities, *IEEE Transactions on Information Theory* 68, 6762-6773.

[12] Bryzgalova, S., Lerner, S., Lettau, M., and Pelger, M., 2022. Missing Financial Data, available at SSRN 4106794.

[13] Candes, E. J. and Plan, Y., 2010. Matrix Completion with Noise, *Proceedings of the IEEE* 98, 925-936.

[14] Candes, E. J. and Recht, B., 2009. Exact Matrix Completion via Convex Optimization, *Foundations of Computational Mathematics* 9, 717–772.

[15] Chan, M.K. and Kwok, S.S., 2022. The PCDID Approach: Difference-in-Differences When Trends Are Potentially Unparallel and Stochastic, *Journal of Business & Economic Statistics* 40, 1216-1233.

[16] Chen, Y., Fan, J., Ma, C., and Yan, Y., 2019. Inference and Uncertainty Quantification for Noisy Matrix Completion, *Proceedings of the National Academy of Sciences* 116, 22931–22937.

[17] Chen, A.Y. and McCoy, J., 2022. Missing Values and the Dimensionality of Expected Returns, arXiv preprint arXiv:2207.13071.

[18] Chernozhukov, V., Hansen, C., Liao, Y., and Zhu, Y., 2023. Inference for Low-Rank Models, *Annals of Statistics* 51, 1309-1330.

[19] Cox, D.R. and Reid, N., 1987. Parameter Orthogonality and Approximate Conditional Inference, *Journal of the Royal Statistical Society: Series B* 49, 1-18.

[20] Doz, C., Giannone, D., and Reichlin, L., 2011. A Two-Step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering. *Journal of Econometrics* 164, 188-205.

[21] Fouirnaies, A. and Mutlu-Eren, H., 2015. English Bacon: Copartisan Bias in Intergovernmental Grant Allocation in England, *Journal of Politics* 77, 805–817.

[22] Freyberger, J., Höppner, B., Neuhierl, A., and Weber, M., 2022. Missing Data in Asset Pricing Panels, No. w30761. National Bureau of Economic Research.

[23] Giannone, D., Reichlin, L., and Small, D., 2008., Nowcasting: The Real-Time Informational Content of Macroeconomic Data. *Journal of Monetary Economics* 55, 665-676.

[24] Gobillon, L. and Magnac, T., 2016. Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls. *Review of Economics and Statistics* 98, 535–551.

[25] Jin, S., Miao, K., and Su, L., 2021. On Factor Models with Random Missing: EM Estimation, Inference, and Cross Validation. *Journal of Econometrics* 222, 745-777.

[26] Jungbacker, B., Koopman, S., and van der Wel, M., 2011. Maximum Likelihood Estimation for Dynamic Factor Models with Missing Data. *Journal of Economic Dynamics and Control* 35, 1358–1368.

[27] Keshavan, R.H., Montanari, A., and Oh, S., 2010. Matrix Completion from A Few Entries, *IEEE Transactions on Information Theory* 56, 2980–2998.

[28] Koltchinskii, V., 2011. Von Neumann Entropy Penalization and Low-Rank Matrix Estimation, *Annals of Statistics* 39, 2936-2973.

[29] Koltchinskii, V., Lounici, K., and Tsybakov, A.B., 2011. Nuclear-Norm Penalization and Optimal Rates for Noisy Low-Rank Matrix Completion, *Annals of Statistics* 39, 2302–2329.

[30] Lancaster, T., 2002. Orthogonal Parameters and Panel Data, *Review of Economic Studies* 69, 647-666.

[31] Liu, L., Wang, Y., and Xu, Y., 2024. A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data, *American Journal of Political Science*, 68, 160-176.

[32] Lu, X., K. Miao and L. Su, 2023. Estimation of Heterogeneous Panel Data Models with an Application to Program Evaluation. *Working Paper*, Chinese University of Hong Kong.

[33] Ma, W. and Chen, G.H. 2019, Missing not at Random in Matrix Completion: The Effectiveness of Estimating Missingness Probabilities under a Low Nuclear Norm Assumption, *Advances in Neural Information Processing Systems* 32.

[34] Mariano, R.S. and Murasawa, Y. 2003, A New Coincident Index of Business Cycles based on Monthly and Quarterly Series, *Journal of Applied Econometrics* 18, 427-443.

[35] Mazumder, R., Hastie, T., and Tibshirani, R. 2010, Spectral Regularization Algorithms for Learning Large Incomplete Matrices. *Journal of Machine Learning Research* 11, 2287-2322.

[36] Negahban, S. and Wainwright, M.J.. 2011. Estimation of (near) Low-Rank Matrices with Noise and High-Dimensional Scaling. *Annals of Statistics* 39, 1069–1097.

[37] Negahban, S. and Wainwright, M.J., 2012., Restricted Strong Convexity and Weighted Matrix Completion: Optimal Bounds with Noise, *Journal of Machine Learning Research* 13, 1665–1697.

[38] Neyman, J., 1979. C(α) Tests and Their Use, *Sankhya* 41, 1–21.

[39] Pelger, M., and Xiong, R., 2022. State-varying Factor Models of Large Dimensions. *Journal of Business & Economic Statistics*, 40(3), 1315-1333.

[40] Rohde, A. and Tsybakov, A. B., 2011. Estimation of High-Dimensional Low-Rank Matrices, *Annals of Statistics* 39, 887–930.

[41] Rosenbaum, P. R. and Rubin, D. B., 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika* 70, 41–55.

[42] Rubin, D. B., 1974. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, *Journal of Educational Psychology* 66, 688–701.

[43] Rubin, D. B., 1976. Inference and Missing Data, *Biometrika* 63, 581–592.

[44] Schnabel, T., Swaminathan, A., Singh, A., Chandak, N., and Joachims, T., 2016. Recommendations as Treatments: Debiasing Learning and Evaluation, *International Conference on Machine Learning*, pp. 1670–1679.

[45] Sportisse, A., Boyer, C., and Josse, J., 2020. Imputation and Low-Rank Estimation with Missing not at Random Data, *Statistics and Computing* 30, 1629–1643.

[46] Stock, J.H. and Watson, M.W.. 2002. Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics* 20, 147-162.

[47] Su, L., and Wang, X., 2017. On Time-varying Factor Models: Estimation and Testing. *Journal of Econometrics* 198(1), 84-101.

[48] Su, L. and Wang, X., 2024. Sieve Estimation of Time-Varying Factor Models. *Working Paper*, Tsinghua University.

[49] Wang, F., 2022. Maximum Likelihood Estimation and Inference for High Dimensional Generalized Factor Models with Application to Factor-Augmented Regressions. *Journal of Econometrics* 229, 180-200.

[50] Xia, D. and Yuan, M., 2021, Statistical Inferences of Linear Forms for Noisy Matrix Completion, *Journal of the Royal Statistical Society: Series B* 83, 58–77.

[51] Xiong, R. and Pelger, M., 2023. Large Dimensional Latent Factor Modeling with Missing Observations and Applications to Causal Inference, *Journal of Econometrics* 233, 271-301.

[52] Xu, Y., 2017. Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models, *Political Analysis* 25, 57–76.

[53] Zhu, Z., Wang, T., and Samworth, R. J., 2022. High-Dimensional Principal Component Analysis with Heterogeneous Missingness, *Journal of the Royal Statistical Society: Series B* 84, 2000-2031.

Online Supplement for "Inference for Large Dimensional Factor Models under General Missing Data Patterns"

Liangjun Su[a] and Fa Wang[b]

[a] School of Economics and Management, Tsinghua University

[b] School of Economics, Peking University

This online supplement is composed of five sections. Sections A–D contain the proofs of Theorems 4.1–5.1, respectively. Section E contains the proof of Proposition 5.1.

A Proof of Theorem 4.1

To prove Theorem 4.1, we introduce the following two lemmas.

Lemma A.1 *Suppose Assumption 4 holds. Then as* $(N,T)\to\infty$, $\|d\circ v\| = O_p(N^{\frac{1}{2}}T^{\frac{1}{4}} + N^{\frac{1}{4}}T^{\frac{1}{2}})$,[2] *where* $\circ$ *denotes the Hadamard product so that* $d\circ v$ *is a* $T\times N$ *matrix with* $d_{it}v_{it}$ *as the* (t,i)*th element.*

Proof. Note that

$$\begin{aligned}
\|d\circ v\|^4 &= \left\|(d\circ v)'(d\circ v)\right\|^2 \le \left\|(d\circ v)'(d\circ v)\right\|_F^2 = \sum\nolimits_{s,t=1}^T (\sum\nolimits_{i=1}^N d_{is}v_{is}d_{it}v_{it})^2 \\
&\le 2\sum\nolimits_{s,t=1}^T \mathbb{E}(\sum\nolimits_{i=1}^N [d_{is}v_{is}d_{it}v_{it} - \mathbb{E}(d_{is}v_{is}d_{it}v_{it})])^2 + 2\sum\nolimits_{s,t=1}^T (\sum\nolimits_{i=1}^N \mathbb{E}(d_{is}v_{is}d_{it}v_{it}))^2 \\
&\equiv 2I_{1,1} + 2I_{1,2},
\end{aligned}$$

where $\sum_{s,t=1}^T = \sum_{s=1}^T\sum_{t=1}^T$. By Assumption 4(ii)-(iii),

$$\begin{aligned}
I_{1,1} &= N\sum\nolimits_{s,t=1}^T \mathbb{E}\{\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^N [d_{is}v_{is}d_{it}v_{it} - \mathbb{E}(d_{is}v_{is}d_{it}v_{it})]\}^2 = O(NT^2), \\
I_{1,2} &= N^2\sum\nolimits_{s,t=1}^T [\gamma_N(s,t)]^2 \le MN^2\sum\nolimits_{s,t=1}^T |\gamma_N(s,t)| = O(N^2T),
\end{aligned}$$

where we also use the fact that $\max_{s,t}|\gamma_N(s,t)| \le \max_s \gamma_N(s,s) \le M$ by the Cauchy-Schwarz (CS) inequality. Then $\mathbb{E}\|d\circ v\|^4 = O(NT^2 + N^2T)$ and the result follows by the Markov inequality. ■

Lemma A.2 *Suppose Assumption 2(i) holds. Then as* $(N,T)\to\infty$, $\|d - \mathbb{E}_\phi(d)\| = O_p(N^{\frac{1}{2}}T^{\frac{1}{4}} + N^{\frac{1}{4}}T^{\frac{1}{2}})$, *where* $d - \mathbb{E}_\phi(d)$ *denotes the* $T\times N$ *matrix with* $d_{it} - \mathbb{E}_\phi(d_{it})$ *as the* (t,i) *element.*

[2] The rate $O_p(N^{\frac{1}{2}}T^{\frac{1}{4}} + N^{\frac{1}{4}}T^{\frac{1}{2}})$ is not sharp, but is enough for our purpose. If $d_{it}v_{it}$ is independent across i and t and its fourth moment is uniformly bounded over i and t, results in random matrix theory show that this rate can be improved to $O_p(N^{\frac{1}{2}} + T^{\frac{1}{2}})$.

Proof. Recall that $\tilde{d}_{it} = d_{it} - \mathbb{E}_\phi(d_{it})$. As in Lemma A.1, we have

$$\begin{aligned}\|d - \mathbb{E}_\phi(d)\|^4 &= \left\|(d - \mathbb{E}_\phi(d))'(d - \mathbb{E}_\phi(d))\right\|^2 \le \left\|(d - \mathbb{E}_\phi(d))'(d - \mathbb{E}_\phi(d))\right\|_F^2 \\ &= \sum\nolimits_{s,t=1}^T (\sum\nolimits_{i=1}^N \tilde{d}_{is}\tilde{d}_{it})^2.\end{aligned}$$

Since $\tilde{d}_{it}$ is independent across i conditional on ϕ^0, we have

$$\begin{aligned}\mathbb{E}_\phi \|d - \mathbb{E}_\phi(d)\|^4 &\le 2\sum\nolimits_{s,t=1}^T \mathbb{E}_\phi\{\sum\nolimits_{i=1}^N [\tilde{d}_{is}\tilde{d}_{it} - \mathbb{E}_\phi(\tilde{d}_{is}\tilde{d}_{it})]\}^2 + 2\sum\nolimits_{s,t=1}^T [\sum\nolimits_{i=1}^N \mathbb{E}_\phi(\tilde{d}_{is}\tilde{d}_{it})]^2 \\ &= 2\sum\nolimits_{s,t=1}^T \sum\nolimits_{i=1}^N \mathbb{E}_\phi[\tilde{d}_{is}\tilde{d}_{it} - \mathbb{E}_\phi(\tilde{d}_{is}\tilde{d}_{it})]^2 + 2\sum\nolimits_{s,t=1}^T [\sum\nolimits_{i=1}^N \mathbb{E}_\phi(\tilde{d}_{is}\tilde{d}_{it})]^2.\end{aligned}$$

The first term on the right hand side (RHS) of the above equation is bounded by $2\sum_{s,t=1}^T \sum_{i=1}^N 1 = 2T^2N$, and the second term is bounded above by

$$2N^2 \sum\nolimits_{s,t=1}^T |\gamma_{Nd}(s,t)| = O_p(N^2T)$$

under Assumption 2(i). It follows that $\mathbb{E}_\phi \|d - \mathbb{E}_\phi(d)\|^4 = O_p(T^2N + N^2T)$ and $\|d - \mathbb{E}(d)\| = O(N^{\frac{1}{2}}T^{\frac{1}{4}} + N^{\frac{1}{4}}T^{\frac{1}{2}})$. ■

Proof of Theorem 4.1.

(1) The random missing case. Let $\hat{\theta}_{it} = \hat{f}_t'\hat{\lambda}_i$ and $\theta_{it}^0 = f_t^{0\prime}\lambda_i^0$. Since $P(\hat{\lambda}, \hat{f}) = 0$ and $P(\lambda^0, f^0) = 0$,

$$Q(\hat{\lambda}, \hat{f}) - Q(\lambda^0, f^0) = \sum\nolimits_{i=1}^N \sum\nolimits_{t=1}^T d_{it}v_{it}(\hat{\theta}_{it} - \theta_{it}^0) - \frac{1}{2}\sum\nolimits_{i=1}^N \sum\nolimits_{t=1}^T d_{it}(\hat{\theta}_{it} - \theta_{it}^0)^2 \ge 0. \qquad \text{(A.1)}$$

Let $\hat{\Theta}$ (resp. Θ^0) denote the $T \times N$ matrix with (t,i)th element given by $\hat{\theta}_{it}$ (resp. θ_{it}^0). Since $\text{rank}(\hat{\Theta} - \Theta^0) \le 2r$ and $|tr(AB)| \le rank(B)\|A\| \|B\|_F$, we have

$$\left|\sum\nolimits_{i=1}^N \sum\nolimits_{t=1}^T d_{it}v_{it}(\hat{\theta}_{it} - \theta_{it}^0)\right| \le 2r\|d \circ v\| [\sum\nolimits_{i=1}^N \sum\nolimits_{t=1}^T (\hat{\theta}_{it} - \theta_{it}^0)^2]^{\frac{1}{2}}. \qquad \text{(A.2)}$$

By the submultiplicative property of the Hadamard product (see, e.g., Theorem 5.1.7 in Horn and Johnson (1991)), $\text{rank}((\hat{\Theta} - \Theta^0) \circ (\hat{\Theta} - \Theta^0)) \le 4r^2$. Then

$$\begin{aligned}\sum\nolimits_{i=1}^N \sum\nolimits_{t=1}^T (d_{it} - \mathbb{E}_\phi(d_{it}))(\hat{\theta}_{it} - \theta_{it}^0)^2 &\le 4r^2\|d - \mathbb{E}_\phi(d)\| [\sum\nolimits_{i=1}^N \sum\nolimits_{t=1}^T (\hat{\theta}_{it} - \theta_{it}^0)^4]^{\frac{1}{2}} \\ &\le 4r^2\|d - \mathbb{E}_\phi(d)\| M[\sum\nolimits_{i=1}^N \sum\nolimits_{t=1}^T (\hat{\theta}_{it} - \theta_{it}^0)^2]^{\frac{1}{2}}, \text{(A.3)}\end{aligned}$$

where the second inequality holds by the fact that θ_{it}^0 is bounded by Assumption 1 and $\hat{\theta}_{it}$ is bounded by design. It follows that

$$\begin{aligned}
& 2r\left\|d\circ v\right\|[\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\theta}_{it}-\theta_{it}^0)^2]^{\frac{1}{2}}+2r^2\left\|d-\mathbb{E}_\phi(d)\right\|M[\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\theta}_{it}-\theta_{it}^0)^2]^{\frac{1}{2}} \\
\geq & \sum_{i=1}^{N}\sum_{t=1}^{T}d_{it}v_{it}(\hat{\theta}_{it}-\theta_{it}^0)-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}(d_{it}-\mathbb{E}_\phi(d_{it}))(\hat{\theta}_{it}-\theta_{it}^0)^2 \\
\geq & \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}_\phi(d_{it})(\hat{\theta}_{it}-\theta_{it}^0)^2 \\
\geq & \frac{c}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\theta}_{it}-\theta_{it}^0)^2,
\end{aligned}$$

where the first inequality holds by (A.2) and (A.3), the second one holds by (A.1), and the last one holds by the assumption that $\mathbb{E}_\phi(d_{it})\geq c$ for all (i,t). Then by Lemmas A.1–A.2, we have

$$[\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\theta}_{it}-\theta_{it}^0)^2]^{\frac{1}{2}}\leq\frac{4r}{c}\left\|d\circ v\right\|+\frac{4r^2M}{c}\left\|d-\mathbb{E}_\phi(d)\right\|=O_p(\frac{\sqrt{NT}}{\sqrt{c_{NT}}}). \tag{A.4}$$

Let $\sigma_1,...,\sigma_r$ and $\hat{\sigma}_1,...,\hat{\sigma}_r$ denote the first r largest singular values of $\frac{F^0\Lambda^{0\prime}}{\sqrt{NT}}$ and $\frac{\hat{F}\hat{\Lambda}'}{\sqrt{NT}}$, respectively, ordered from the largest to the smallest. Let $e_1,...,e_r$ and $\hat{e}_1,...,\hat{e}_r$ denote the corresponding left-singular vectors. By the Davis-Kahan theorem (see, e.g., Yu, Wang and Samworth (2015)),

$$\left\|\hat{e}_j-e_j\right\|\leq\frac{\sqrt{2}}{\eta}\left\|\frac{\hat{F}\hat{\Lambda}'}{\sqrt{NT}}-\frac{F^0\Lambda^{0\prime}}{\sqrt{NT}}\right\|\leq\frac{\sqrt{2}}{\eta}\left\|\frac{\hat{F}\hat{\Lambda}'}{\sqrt{NT}}-\frac{F^0\Lambda^{0\prime}}{\sqrt{NT}}\right\|_F, \tag{A.5}$$

where $\eta=\min\{|\sigma_{j-1}-\hat{\sigma}_j|\wedge|\sigma_{j+1}-\hat{\sigma}_j|,j=1,...,r\}$. η is bounded and bounded away from zero in probability because (1) σ_j, $j\in[r]$, are all bounded and bounded away from zero in probability, and σ_j are different by Assumptions 3, (2) by Weyl's inequality, $|\sigma_j-\hat{\sigma}_j|\leq\left\|\frac{\hat{F}\hat{\Lambda}'}{\sqrt{NT}}-\frac{F^0\Lambda^{0\prime}}{\sqrt{NT}}\right\|_F=O_p(\frac{1}{\sqrt{c_{NT}}})$ for all $j\in[r]$. Thus $\max_{j\in[r]}\left\|\hat{e}_j-e_j\right\|=O_p(\frac{1}{\sqrt{c_{NT}}})$. From the conditions (2.7) and (2.6), we know that the j-th estimated factor is $\sqrt{T\hat{\sigma}_j}\hat{e}_j$, and the j-th factor is $\sqrt{T\sigma_j}e_j$. It follows that

$$\begin{aligned}
\left\|\hat{f}-f^0\right\| &\leq \max_{j\in[r]}\left\|\sqrt{T\hat{\sigma}_j}\hat{e}_j-\sqrt{T\sigma_j}e_j\right\| \\
&\leq \max_{j\in[r]}\left|\sqrt{T\hat{\sigma}_j}-\sqrt{T\sigma_j}\right|\left\|\hat{e}_j\right\|+\max_{j\in[r]}\sqrt{T\sigma_j}\left\|\hat{e}_j-e_j\right\|=O_p(\frac{\sqrt{T}}{\sqrt{c_{NT}}}).
\end{aligned}$$

By symmetry, we also have $\left\|\hat{\lambda}-\lambda^0\right\|=O_p(\frac{\sqrt{N}}{\sqrt{c_{NT}}})$.

(2) The block missing/staggered missing/mixed frequency case. Note that missing only occurs for $i>N_o$, which implies that $d_{it}=1$ for all $i\in[N_o]$ and $t\in[T]$. For $i>N_o$, define $\mathcal{T}_i=\{t\in[T],\ d_{it}=1\}$. Clearly, $\mathcal{T}_i=[T_o]$ for the block missing case, $\mathcal{T}_i\supseteq[T_{oi}]$ for the staggered treatment case, and $\mathcal{T}_i=\{t\in[T],\ t/h\text{ is an integer}\}$ where h denotes the frequency ratio (e.g., $h=3$

when we have monthly and quarterly data). Then equation (A.1) becomes

$$\begin{aligned}&\sum_{i=1}^{N_o}\sum_{t=1}^{T} v_{it}(\hat{\theta}_{it}-\theta_{it}^0)+\sum_{i=N_o+1}^{N}\sum_{t\in\mathcal{T}_i} v_{it}(\hat{\theta}_{it}-\theta_{it}^0)\\ &-\frac{1}{2}\sum_{i=1}^{N_o}\sum_{t=1}^{T}(\hat{\theta}_{it}-\theta_{it}^0)^2-\frac{1}{2}\sum_{i=N_o+1}^{N}\sum_{t\in\mathcal{T}_i}(\hat{\theta}_{it}-\theta_{it}^0)^2\geq 0.\end{aligned} \tag{A.6}$$

We consider two cases:

$$\text{Case (a):}\qquad \sum_{i=1}^{N_o}\sum_{t=1}^{T}(\hat{\theta}_{it}-\theta_{it}^0)^2\geq\sum_{i=N_o+1}^{N}\sum_{t\in\mathcal{T}_i}(\hat{\theta}_{it}-\theta_{it}^0)^2, \tag{A.7}$$

$$\text{Case (b):}\qquad \sum_{i=1}^{N_o}\sum_{t=1}^{T}(\hat{\theta}_{it}-\theta_{it}^0)^2<\sum_{i=N_o+1}^{N}\sum_{t\in\mathcal{T}_i}(\hat{\theta}_{it}-\theta_{it}^0)^2. \tag{A.8}$$

In Case (a), we have

$$\begin{aligned}&4r\,\|d\circ v\|\,[\sum_{i=1}^{N_o}\sum_{t=1}^{T}(\hat{\theta}_{it}-\theta_{it}^0)^2]^{\frac{1}{2}}\\ \geq\ &\sum_{i=1}^{N_o}\sum_{t=1}^{T} v_{it}(\hat{\theta}_{it}-\theta_{it}^0)+\sum_{i=N_o+1}^{N}\sum_{t\in\mathcal{T}_i} v_{it}(\hat{\theta}_{it}-\theta_{it}^0)\\ \geq\ &\frac{1}{2}\sum_{i=1}^{N_o}\sum_{t=1}^{T}(\hat{\theta}_{it}-\theta_{it}^0)^2+\frac{1}{2}\sum_{i=N_o+1}^{N}\sum_{t=1}^{T_{oi}}(\hat{\theta}_{it}-\theta_{it}^0)^2\\ \geq\ &\frac{1}{2}\sum_{i=1}^{N_o}\sum_{t=1}^{T}(\hat{\theta}_{it}-\theta_{it}^0)^2,\end{aligned} \tag{A.9}$$

where the second inequality follows from equation (A.6), and the first one holds by the fact that

$$\begin{aligned}\sum_{i=1}^{N_o}\sum_{t=1}^{T} v_{it}(\hat{\theta}_{it}-\theta_{it}^0) &= \sum_{i=1}^{N_o}\sum_{t=1}^{T} d_{it}v_{it}(\hat{\theta}_{it}-\theta_{it}^0)\\ &\leq 2r\,\|d\circ v\|\,[\sum_{i=1}^{N_o}\sum_{t=1}^{T}(\hat{\theta}_{it}-\theta_{it}^0)^2]^{\frac{1}{2}}\ \text{and}\\ \sum_{i=N_o+1}^{N}\sum_{t\in\mathcal{T}_i} v_{it}(\hat{\theta}_{it}-\theta_{it}^0) &= \sum_{i=N_o+1}^{N}\sum_{t\in\mathcal{T}_i} d_{it}v_{it}(\hat{\theta}_{it}-\theta_{it}^0)\\ &\leq 2r\,\|d\circ v\|\,[\sum_{i=N_o+1}^{N}\sum_{t\in\mathcal{T}_i}(\hat{\theta}_{it}-\theta_{it}^0)^2]^{\frac{1}{2}}\\ &\leq 2r\,\|d\circ v\|\,[\sum_{i=1}^{N_o}\sum_{t=1}^{T}(\hat{\theta}_{it}-\theta_{it}^0)^2]^{\frac{1}{2}}\end{aligned}$$

by (A.7) and similar arguments as used to obtain (A.3). Then by (A.9) and Lemma A.1

$$[\sum_{i=1}^{N_o}\sum_{t=1}^{T}(\hat{\theta}_{it}-\theta_{it}^0)^2]^{\frac{1}{2}}\leq 8r\,\|d\circ v\|=O_p(N^{\frac{1}{2}}T^{\frac{1}{4}}+N^{\frac{1}{4}}T^{\frac{1}{2}}). \tag{A.10}$$

which, in conjunction with (A.7), further implies that

$$[\sum_{i=N_o+1}^{N}\sum_{t\in\mathcal{T}_i}(\hat{\theta}_{it}-\theta_{it}^0)^2]^{\frac{1}{2}}\leq[\sum_{i=1}^{N_o}\sum_{t=1}^{T}(\hat{\theta}_{it}-\theta_{it}^0)^2]^{\frac{1}{2}}=O_p(N^{\frac{1}{2}}T^{\frac{1}{4}}+N^{\frac{1}{4}}T^{\frac{1}{2}}). \tag{A.11}$$

Given that N_o/N and T_o/T are bounded away from zero, equation (A.10) implies $||\hat{f}-f^0||=O_p(\frac{\sqrt{T}}{\sqrt{c_{NT}}})$ and $(\sum_{i=1}^{N_o}||\hat{\lambda}_i-\lambda_i^0||^2)^{\frac{1}{2}}=O_p(\frac{\sqrt{N_o}}{\sqrt{c_{NT}}})$, and equation (A.11) implies $(\sum_{i=N_o+1}^{N}||\hat{\lambda}_i-\lambda_i^0||^2)^{\frac{1}{2}}$

$= O_p(\frac{\sqrt{N-N_o}}{\sqrt{c_{NT}}})$. Then we have $||\hat{\lambda} - \lambda^0|| = O_p(\frac{\sqrt{N}}{\sqrt{c_{NT}}})$.

Similarly, we can show that $||\hat{f} - f^0|| = O_p(\frac{\sqrt{T}}{\sqrt{c_{NT}}})$ and $||\hat{\lambda} - \lambda^0|| = O_p(\frac{\sqrt{N}}{\sqrt{c_{NT}}})$ in Case (b). ■

B Proof of Theorem 4.2

To proceed, we introduce some notations associated with the Hessian matrix of $Q(\phi)$. Define

$$\partial_{\phi\phi'} L(\phi) = L_{\phi\phi'}(\phi) + J_{\phi\phi'}(\phi), \tag{B.1}$$

$$\partial_{\phi\phi'} P(\phi) = P_{\phi\phi'}(\phi), \tag{B.2}$$

$$\partial_{\phi\phi'} Q(\phi) = H_{\phi\phi'}(\phi) = L_{\phi\phi'}(\phi) + J_{\phi\phi'}(\phi) + P_{\phi\phi'}(\phi), \tag{B.3}$$

$$L_{\phi\phi'}(\phi) = \begin{bmatrix} L_{\lambda\lambda'}(\phi) & L_{\lambda f'}(\phi) \\ L_{f\lambda'}(\phi) & L_{ff'}(\phi) \end{bmatrix}, \; J_{\phi\phi'}(\phi) = \begin{bmatrix} 0 & J_{\lambda f'}(\phi) \\ J_{f\lambda'}(\phi) & 0 \end{bmatrix}, \tag{B.4}$$

$$H_{\phi\phi'}(\phi) = \begin{bmatrix} H_{\lambda\lambda'}(\phi) & H_{\lambda f'}(\phi) \\ H_{f\lambda'}(\phi) & H_{ff'}(\phi) \end{bmatrix}. \tag{B.5}$$

When these matrix are evaluated at the true value ϕ^0, we suppress the argument. That is, we simply write $L_{\phi\phi'}(\phi^0)$, $J_{\phi\phi'}(\phi^0)$, $P_{\phi\phi'}(\phi^0)$, and $H_{\phi\phi'}(\phi^0)$ as $L_{\phi\phi'}$, $J_{\phi\phi'}$, $P_{\phi\phi'}$, and $H_{\phi\phi'}$, respectively. Note that $L_{\lambda\lambda'}$ is an $Nr \times Nr$ block-diagonal matrix, with the ith diagonal block of size $r \times r$ given by $-\sum_{t=1}^{T} d_{it} f_t^0 f_t^{0\prime}$; $L_{ff'}$ is a $Tr \times Tr$ block-diagonal matrix with the t-th diagonal block of size $r \times r$ given by $-\sum_{i=1}^{N} d_{it} \lambda_i^0 \lambda_i^{0\prime}$; $L_{\lambda f'}$ is of dimension $Nr \times Tr$ with the (i,t)th block of size $r \times r$ given by $-d_{it} f_t^0 \lambda_i^{0\prime}$; $L_{f\lambda'}$ is the transpose of $L_{\lambda f'}$. $J_{\lambda f'}$ is also $Nr \times Tr$ with the (i,t)th block of size $r \times r$ given by $d_{it} v_{it} I_r$; $J_{f\lambda'}$ is the transpose of $J_{\lambda f'}$.

Define

$$D_1 = \begin{bmatrix} I_N \otimes \iota_{pq} & 0 \\ 0 & 0 \end{bmatrix}, \; D_2 = \begin{bmatrix} 0 & 0 \\ 0 & I_T \otimes \iota_{pq} \end{bmatrix}, \text{ and } D_3 = \begin{bmatrix} I_N \otimes \iota_q & 0 \\ 0 & -I_T \otimes \iota_q \end{bmatrix},$$

where ι_q is an $r \times r$ matrix with the q-th diagonal element being one and all the other elements being zero, and ι_{pq} is an $r \times r$ matrix with the (p,q) element and the (q,p) element being one and all the other elements being zero. Define the following three sets of $(Nr + Tr) \times 1$ vectors:

w_{pp} : For $1 \le p \le r$, w_{pp} is an $Nr + Tr$ dimensional vector; for the first Nr elements, in the i-th block, the p-th element is λ_{ip} and all the other elements are zeros; for the last Tr elements, in the t-th block, the p-th element is $-f_{tp}$ and all the other elements are zeros.

u_{pq} : For $1 \le p < q \le r$, u_{pq} is a $Nr + Tr$ dimensional vector; the last Tr elements are all zeros; for the first Nr elements, in the i-th block, the p-th element is λ_{iq}, the q-th element is λ_{ip} and all the other elements are zeros.

u_{qp} : For $1 \le p < q \le r$, u_{qp} is a $Nr + Tr$ dimensional vector; the first Nr elements are all zeros; for the last Tr elements, in the t-th block, the p-th element is f_{tq}, the q-th element is f_{tp} and all the other elements are zeros.

When $\lambda = \lambda^0$ and $f = f^0$ so that $\phi = \phi^0$, w_{pp}, u_{pq} and u_{qp} are denoted as w^0_{pp}, u^0_{pq} and u^0_{qp}, respectively. One can verify that

$$\begin{aligned}
\partial_{\phi\phi'}[(\frac{\sum_{i=1}^N \lambda_{ip}^2}{N} - \frac{\sum_{t=1}^T f_{tp}^2}{T})^2] &= 8D_{NT}^{-1} w_{pp} w'_{pp} D_{NT}^{-1} + 4(\frac{\sum_{i=1}^N \lambda_{ip}^2}{N} - \frac{\sum_{t=1}^T f_{tp}^2}{T}) D_{NT}^{-1} D_3, && \text{(B.6)}\\
\partial_{\phi\phi'}[(\sum_{i=1}^N \lambda_{ip}\lambda_{iq})^2] &= 2[u_{pq}u'_{pq} + (\sum_{i=1}^N \lambda_{ip}\lambda_{iq}) D_1], && \text{(B.7)}\\
\partial_{\phi\phi'}[(\sum_{t=1}^T f_{tp} f_{tq})^2] &= 2[u_{qp}u'_{qp} + (\sum_{t=1}^T f_{tp} f_{tq}) D_2]. && \text{(B.8)}
\end{aligned}$$

Since $\frac{1}{T}F^{0\prime}F^0 = \frac{1}{N}\Lambda^{0\prime}\Lambda^0$ and both are diagonal, the second term on the RHS of equations (B.6)-(B.8) are zeros when ϕ is evaluated at the true value ϕ^0. It follows that

$$\begin{aligned}
P_{\phi\phi'} &= -cD_{TN}^{\frac{1}{2}} D_{NT}^{-\frac{1}{2}} (\sum_{p=1}^r w^0_{pp} w^{0\prime}_{pp} + \sum_{p=1}^r \sum_{q=p+1}^r u^0_{pq} u^{0\prime}_{pq} + \sum_{p=1}^r \sum_{q=p+1}^r u^0_{qp} u^{0\prime}_{qp}) D_{NT}^{-\frac{1}{2}} D_{TN}^{\frac{1}{2}}\\
&= -cD_{TN}^{\frac{1}{2}} D_{NT}^{-\frac{1}{2}} U^0 U^{0\prime} D_{NT}^{-\frac{1}{2}} D_{TN}^{\frac{1}{2}}, && \text{(B.9)}
\end{aligned}$$

where

$$\begin{aligned}
U^0 &= (w^0_{11}, ..., w^0_{rr}; u^0_{12}, ..., u^0_{1r}, u^0_{23}, ..., u^0_{2r}, ..., u^0_{(r-1)r}; u^0_{21}, ..., u^0_{r1}, u^0_{32}, ..., u^0_{r2}, ..., u^0_{r(r-1)})\\
&\equiv (U^{0\prime}_\lambda, U^{0\prime}_f)'. && \text{(B.10)}
\end{aligned}$$

Here, U^0 is an $(Nr + Tr) \times r^2$ matrix, U^0_λ contains the first Nr rows of U^0 and U^0_f contains the last Tr rows.

Lemma B.1 *Suppose that Assumptions 1, 2(i) and 3 hold. Then* $||(-D_{TN}^{-\frac{1}{2}} H_{\phi\phi'} D_{TN}^{-\frac{1}{2}})^{-1}|| = O_p(1)$ *as* $(N, T) \to \infty$.

Proof. To proceed, we introduce an $(Nr + Tr) \times r^2$ matrix W^0, which specifies the null space of $L_{\phi\phi'}$ and plays an essential role in the following analysis. For $1 \le p < q \le r$, w^0_{pq} is a $(Nr+Tr) \times 1$ vector: for the first Nr elements, in the i-th block, the p-th element is λ^0_{iq} and all the other elements are zeros; for the last Tr elements, in the t-th block, the q-th element is $-f^0_{tp}$ and all the other elements are zeros. w^0_{qp} is also an $(Nr + Tr) \times 1$ vector: for the first Nr elements, in the i-th block, the q-th element is λ^0_{ip} and all the other elements are zero; for the last Tr elements, in the t-th block, the p-th element is $-f^0_{tq}$ and all the other elements are zero. Let

$$\begin{aligned} W^0 &\equiv (w^0_{11}, ..., w^0_{rr}; w^0_{12}, ..., w^0_{1r}, w^0_{23}, ..., w^0_{2r}, ..., w^0_{(r-1)r}; w^0_{21}, ..., w^0_{r1}, w^0_{32}, ..., w^0_{r2}, ..., w^0_{r(r-1)}) \\ &\equiv (W^{0\prime}_\lambda, W^{0\prime}_f)'. \end{aligned} \tag{B.11}$$

Here, W^0_λ contains the first Nr rows of W^0 and W^0_f contains the last Tr rows. It is not difficult to verify that any two different columns of W^0 are orthogonal to each other, the (i,t)-th block of $W^0_\lambda W^{0\prime}_f$ of size $r \times r$ is $-f^0_t \lambda^{0\prime}_i$,

$$L_{\phi\phi'} W^0 = 0 \text{ and } S'_\phi W^0 = 0. \tag{B.12}$$

Let $\bar{L}_{\phi\phi'} = \mathbb{E}_\phi(L_{\phi\phi'})$ and define $\bar{H}_{\phi\phi'}$, $\bar{L}_{\lambda\lambda'}$, $\bar{L}_{ff'}$ and $\bar{L}_{\lambda f'}$ similarly. Let 1^N_i denote $N \times 1$ vector with the i-th element being one and all the other elements being zero. Let $\breve{H}_{\phi\phi'} = \bar{L}_{\phi\phi'} - cD^{\frac{1}{2}}_{TN} D^{-\frac{1}{2}}_{NT} W^0 W^{0\prime} D^{-\frac{1}{2}}_{NT} D^{\frac{1}{2}}_{TN}$. Then by (B.3) we have

$$H_{\phi\phi'} = \breve{H}_{\phi\phi'} + [P_{\phi\phi'} + cD^{\frac{1}{2}}_{TN} D^{-\frac{1}{2}}_{NT} W^0 W^{0\prime} D^{-\frac{1}{2}}_{NT} D^{\frac{1}{2}}_{TN}] + (L_{\phi\phi'} - \bar{L}_{\phi\phi'}) + J_{\phi\phi'}, \tag{B.13}$$

where

$$\begin{aligned} -\breve{H}_{\phi\phi'} &= \sum_{i=1}^N \sum_{t=1}^T (\mathbb{E}_\phi(d_{it}) - c) \begin{pmatrix} 1^N_i \otimes f^0_t \\ 1^T_t \otimes \lambda^0_i \end{pmatrix} \begin{pmatrix} 1^N_i \otimes f^0_t \\ 1^T_t \otimes \lambda^0_i \end{pmatrix}' \\ &\quad + \begin{pmatrix} \frac{cT}{N} W^0_\lambda W^{0\prime}_\lambda & 0 \\ 0 & \frac{cN}{T} W^0_f W^{0\prime}_f \end{pmatrix} + \begin{pmatrix} I_N \otimes c\sum_{t=1}^T f^0_t f^{0\prime}_t & 0 \\ 0 & I_T \otimes c\sum_{i=1}^N \lambda^0_i \lambda^{0\prime}_i \end{pmatrix}. \end{aligned} \tag{B.14}$$

We will study each term on the RHS of (B.13). In Step (1.1)–(1.3), we study $\breve{H}_{\phi\phi'}$, $J_{\phi\phi'}$, and $L_{\phi\phi'} - \bar{L}_{\phi\phi'}$ in order. In Step (2), we focus on $P_{\phi\phi'} + cD^{\frac{1}{2}}_{TN} D^{-\frac{1}{2}}_{NT} W^0 W^{0\prime} D^{-\frac{1}{2}}_{NT} D^{\frac{1}{2}}_{TN}$.

Step (1.1). We study $\breve{H}_{\phi\phi'}$. It's easy to see that the first two terms on the RHS of (B.14) are positive semi-definite. This implies that $\sigma_{\min}(-D^{-\frac{1}{2}}_{TN} \breve{H}_{\phi\phi'} D^{-\frac{1}{2}}_{TN})$ is bounded away from zero in probability by Assumption 1.

Step (1.2). We study $J_{\phi\phi'}$. Since $||D^{-\frac{1}{2}}_{TN} J_{\phi\phi'} D^{-\frac{1}{2}}_{TN}||$ is bounded by $\frac{2}{\sqrt{NT}} \|d \circ v\|$, by Lemma A.1 we have

$$\left\| D^{-\frac{1}{2}}_{TN} J_{\phi\phi'} D^{-\frac{1}{2}}_{TN} \right\| = (NT)^{-1/2} O_p(N^{\frac{1}{2}} T^{\frac{1}{4}} + N^{\frac{1}{4}} T^{\frac{1}{2}})) = O_p(\frac{1}{\sqrt{c_{NT}}}). \tag{B.15}$$

Step (1.3). We study $L_{\phi\phi'} - \bar{L}_{\phi\phi'}$ and show

$$\left\| D^{-\frac{1}{2}}_{TN} (L_{\phi\phi'} - \bar{L}_{\phi\phi'}) D^{-\frac{1}{2}}_{TN} \right\| = O_p(\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}} + \frac{T^{\frac{1}{\kappa}}}{\sqrt{N}} + \frac{1}{\sqrt{c_{NT}}}). \tag{B.16}$$

By (B.4), it suffices to prove (B.16) by showing that

$$\begin{aligned}\left\|L_{\lambda\lambda'}-\bar{L}_{\lambda\lambda'}\right\| &= O_p(\sqrt{T}N^{\frac{1}{\kappa}}), &\text{(B.17)}\\ \left\|L_{ff'}-\bar{L}_{ff'}\right\| &= O_p(\sqrt{N}T^{\frac{1}{\kappa}}), &\text{(B.18)}\\ \left\|L_{\lambda f'}-\bar{L}_{\lambda f'}\right\| &= O_p(\frac{\sqrt{NT}}{\sqrt{c_{NT}}}). &\text{(B.19)}\end{aligned}$$

The proof of (B.19) is similar to that of Lemma A.2. The (i,t)th block of $L_{\lambda f'}-\bar{L}_{\lambda f'}$ of size $r\times r$ is $(d_{it}-\mathbb{E}_\phi(d_{it}))f_t^0\lambda_i^{0\prime}$, the (s,t)th block of $(L_{\lambda f'}-\bar{L}_{\lambda f'})'(L_{\lambda f'}-\bar{L}_{\lambda f'})$ of size $r\times r$ is $\sum_{i=1}^N \tilde{d}_{is}\tilde{d}_{it}a_{its}$, where $\tilde{d}_{it}=d_{it}-\mathbb{E}_\phi(d_{it})$ and $a_{ist}=\lambda_i^0 f_s^{0\prime} f_t^0\lambda_i^{0\prime}$. Let $a_{ist,pq}$ denotes the (p,q)th element of a_{its}. Let $\sum_{s,t=1}^T=\sum_{s=1}^T\sum_{t=1}^T$ and $\sum_{p,q=1}^r=\sum_{p=1}^r\sum_{q=1}^r$. Then

$$\begin{aligned}&\mathbb{E}_\phi(\left\|L_{\lambda f'}-\bar{L}_{\lambda f'}\right\|^4)\\ =\;& \mathbb{E}_\phi(\left\|(L_{\lambda f'}-\bar{L}_{\lambda f'})'(L_{\lambda f'}-\bar{L}_{\lambda f'})\right\|^2)\le \mathbb{E}_\phi(\left\|(L_{\lambda f'}-\bar{L}_{\lambda f'})'(L_{\lambda f'}-\bar{L}_{\lambda f'})\right\|_F^2)\\ =\;& \sum_{s,t=1}^T\sum_{p,q=1}^r \mathbb{E}_\phi(\sum_{i=1}^N \tilde{d}_{is}\tilde{d}_{it}a_{its,pq})^2\\ \le\;& 2\sum_{s,t=1}^T\sum_{p,q=1}^r \mathbb{E}_\phi\{\sum_{i=1}^N[\tilde{d}_{is}\tilde{d}_{it}-\mathbb{E}_\phi(\tilde{d}_{is}\tilde{d}_{it})]a_{its,pq}\}^2+2\sum_{s,t=1}^T\sum_{p,q=1}^r[\sum_{i=1}^N \mathbb{E}_\phi(\tilde{d}_{is}\tilde{d}_{it})a_{its,pq}]^2\\ =\;& 2\sum_{s,t=1}^T\sum_{p,q=1}^r\sum_{i=1}^N \mathbb{E}_\phi[\tilde{d}_{is}\tilde{d}_{it}-\mathbb{E}_\phi(\tilde{d}_{is}\tilde{d}_{it})]^2a_{its,pq}^2+2\sum_{s,t=1}^T\sum_{p,q=1}^r[\sum_{i=1}^N \mathbb{E}_\phi(\tilde{d}_{is}\tilde{d}_{it})a_{its,pq}]^2\\ \equiv\;& 2\Delta_{1,1}+2\Delta_{1,2}\end{aligned}$$

where the third equality follows from the conditional independence condition over i in Assumption 2(i). It is easy to see that

$$\begin{aligned}\mathbb{E}(\Delta_{1,1}) &\le \sum_{s,t=1}^T\sum_{p,q=1}^r\sum_{i=1}^N \mathbb{E}(a_{its,pq}^2)=O(T^2N),\ \text{and}\\ \Delta_{1,2} &= \sum_{s,t=1}^T\sum_{p,q=1}^r[\sum_{i=1}^N \mathbb{E}_\phi(\tilde{d}_{is}\tilde{d}_{it})\lambda_{ip}^0 f_s^{0\prime} f_t^0\lambda_{iq}^0]^2\\ &\le MN^2\sum_{s,t=1}^T\gamma_{Nd}(t,s)=O_p\left(N^2T\right).\end{aligned}$$

It follows that $\mathbb{E}_\phi(\left\|L_{\lambda f'}-\bar{L}_{\lambda f'}\right\|^4)=O_p(T^2N+N^2T)$ and $\left\|L_{\lambda f'}-\bar{L}_{\lambda f'}\right\|=O_p(T^{1/2}N^{1/4}+N^{1/2}T^{1/4})=O_p(\frac{\sqrt{NT}}{\sqrt{c_{NT}}})$. This proves (B.19). For equation (B.17), we have

$$\begin{aligned}\left\|L_{\lambda\lambda'}-\bar{L}_{\lambda\lambda'}\right\| &= (\max_i\left\|\sum_{t=1}^T[d_{it}-\mathbb{E}_\phi(d_{it})]f_t^0 f_t^{0\prime}\right\|^\kappa)^{\frac{1}{\kappa}}\\ &= (\sum_{i=1}^N\left\|\sum_{t=1}^T[d_{it}-\mathbb{E}_\phi(d_{it})]f_t^0 f_t^{0\prime}\right\|^\kappa)^{\frac{1}{\kappa}}=O_p(\sqrt{T}N^{\frac{1}{\kappa}}). &\text{(B.20)}\end{aligned}$$

The last equality is due to $\mathbb{E}(\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T[d_{it}-\mathbb{E}_\phi(d_{it})]f_t^0 f_t^{0\prime}\right\|^\kappa)=O(1)$ by Assumption 2(i). Analogously, we can prove (B.18). Then the result in (B.16) holds.

Step (2). We study $P_{\phi\phi'} + cD_{TN}^{\frac{1}{2}} D_{NT}^{-\frac{1}{2}} W^0 W^{0\prime} D_{NT}^{-\frac{1}{2}} D_{TN}^{\frac{1}{2}}$ on the RHS of (B.13) and show that it is asymptotically negligible in comparison with the other terms. The proof is similar to step (2) of Lemma 2 in Wang (2022) and we present it for completeness. Since the columns in W^0 are orthogonal to each other and also orthogonal to $\bar{L}_{\phi\phi'}$, the positive definiteness of $-D_{TN}^{-\frac{1}{2}} \breve{H}_{\phi\phi'} D_{TN}^{-\frac{1}{2}}$ implies that the eigenvectors of $D_{TN}^{-\frac{1}{2}} \bar{L}_{\phi\phi'} D_{TN}^{-\frac{1}{2}}$ together with $\{D_{NT}^{-\frac{1}{2}} w_{pp}^0 / ||D_{NT}^{-\frac{1}{2}} w_{pp}^0||,\ p \in [r]\}$, $\{D_{NT}^{-\frac{1}{2}} w_{pq}^0 / ||D_{NT}^{-\frac{1}{2}} w_{pq}^0||$, $p \in [r]$, $q = p+1, ..., r\}$ and $\{D_{NT}^{-\frac{1}{2}} w_{qp}^0 / ||D_{NT}^{-\frac{1}{2}} w_{qp}^0||, p \in [r], q = p+1, ..., r\}$ constitutes an orthonormal basis. Under this basis, for $j \in [r]$ and $k = j+1, ..., r$, let $(u_{jk,1}^0, ..., u_{jk,(N+T)r-r(r-1)}^0)$ denote the coordinates of u_{jk}^0 corresponding to the eigenvectors of $D_{TN}^{-\frac{1}{2}} \bar{L}_{\phi\phi'} D_{TN}^{-\frac{1}{2}}$ and $D_{NT}^{-\frac{1}{2}} w_{pp}^0 / ||D_{NT}^{-\frac{1}{2}} w_{pp}^0||$, and let $u_{jk,pq}^0$ and $u_{jk,qp}^0$ denote the coordinate of u_{jk}^0 corresponding to $D_{NT}^{-\frac{1}{2}} w_{pq}^0 / ||D_{NT}^{-\frac{1}{2}} w_{pq}^0||$ and $D_{NT}^{-\frac{1}{2}} w_{qp}^0 / ||D_{NT}^{-\frac{1}{2}} w_{qp}^0||$, respectively. The coordinates of u_{kj}^0 are defined in the same way.

To prove the lemma, it suffices to show that there exists $C > 0$ such that for any vector a with $\|a\| = 1$, $a'(-D_{TN}^{-\frac{1}{2}} H_{\phi\phi'} D_{TN}^{-\frac{1}{2}}) a \geq C > 0$ w.p.a.1 as $(N, T) \to \infty$. Let

$$(a_1, ..., a_{(N+T)r-r(r-1)}; a_{12}, ..., a_{1r}, a_{23}, ..., a_{2r}, ..., a_{(r-1)r}; a_{21}, ..., a_{r1}, a_{32}, ..., a_{r2}, ..., a_{r(r-1)})$$

be the coordinates of a. Plugging this into (B.13), we have

$$\begin{aligned} a'(-D_{TN}^{-\frac{1}{2}} H_{\phi\phi'} D_{TN}^{-\frac{1}{2}}) a &= a'(-D_{TN}^{-\frac{1}{2}} \bar{L}_{\phi\phi'} D_{TN}^{-\frac{1}{2}}) a + a'(-D_{TN}^{-\frac{1}{2}} P_{\phi\phi'} D_{TN}^{-\frac{1}{2}}) a \\ &\quad - a'(D_{TN}^{-\frac{1}{2}} (L_{\phi\phi'} - \bar{L}_{\phi\phi'}) D_{TN}^{-\frac{1}{2}}) a - a'(D_{TN}^{-\frac{1}{2}} J_{\phi\phi'} D_{TN}^{-\frac{1}{2}}) a \\ &= a'(-D_{TN}^{-\frac{1}{2}} \bar{L}_{\phi\phi'} D_{TN}^{-\frac{1}{2}} + c \sum_{p=1}^r D_{NT}^{-\frac{1}{2}} w_{pp}^0 w_{pp}^{0\prime} D_{NT}^{-\frac{1}{2}}) a \qquad \text{(B.21)} \\ &\quad + c a' [\sum_{j=1}^r \sum_{k=j+1}^r (\frac{u_{jk}^0 u_{jk}^{0\prime}}{N} + \frac{u_{kj}^0 u_{kj}^{0\prime}}{T})] a \qquad \text{(B.22)} \\ &\quad - a'(D_{TN}^{-\frac{1}{2}} (L_{\phi\phi'} - \bar{L}_{\phi\phi'}) D_{TN}^{-\frac{1}{2}}) a - a'(D_{TN}^{-\frac{1}{2}} J_{\phi\phi'} D_{TN}^{-\frac{1}{2}}) a \qquad \text{(B.23)} \end{aligned}$$

The term (B.21) is not smaller than $b \sum_{l=1}^{(N+T)r-r(r-1)} a_l^2$ w.p.a.1 because the smallest nonzero eigenvalue of $-D_{TN}^{-\frac{1}{2}} \bar{L}_{\phi\phi'} D_{TN}^{-\frac{1}{2}} + c \sum_{p=1}^r D_{NT}^{-\frac{1}{2}} w_{pp}^0 w_{pp}^{0\prime} D_{NT}^{-\frac{1}{2}}$ is not smaller than $\sigma_{\min}(-D_{TN}^{-\frac{1}{2}} \breve{H}_{\phi\phi'} D_{TN}^{-\frac{1}{2}})$. The term in (B.22) is not smaller than $c_1 \sum_{j=1}^r \sum_{k=j+1}^r [\frac{(a' u_{jk}^0)^2}{N} + \frac{(a' u_{kj}^0)^2}{T}]$ for some $0 < c_1 < c$. How to choose c_1 will be discussed later. For $a' u_{kj}^0$, we have

$$(a' u_{kj}^0)^2 \geq [\sum_{p=1}^r \sum_{q=p+1}^r (a_{pq} u_{kj,pq}^0 + a_{qp} u_{kj,qp}^0)]^2 - 2(\sum_{l=1}^{(N+T)r-r(r-1)} a_l^2)^{\frac{1}{2}} \left\|u_{kj}^0\right\|^2.$$

Thus the term (B.22) is not smaller than

$$c_1 \sum_{j=1}^r \sum_{k=j+1}^r \{ \frac{1}{N} [\sum_{p=1}^r \sum_{q=p+1}^r (a_{pq} u_{jk,pq}^0 + a_{qp} u_{jk,qp}^0)]^2$$

$$+\frac{1}{T}[\sum_{p=1}^{r}\sum_{q=p+1}^{r}(a_{pq}u_{kj,pq}^{0}+a_{qp}u_{kj,qp}^{0})]^{2}\} \tag{B.24}$$

$$-2c_{1}(\sum_{l=1}^{(N+T)r-r(r-1)}a_{l}^{2})^{\frac{1}{2}}\sum_{j=1}^{r}\sum_{k=j+1}^{r}(\frac{1}{N}\left\|u_{jk}^{0}\right\|^{2}+\frac{1}{T}\left\|u_{kj}^{0}\right\|^{2}). \tag{B.25}$$

By Assumption 1, expression (B.25) is not smaller than $-2c_1r(r-1)M(\sum_{l=1}^{(N+T)r-r(r-1)}a_l^2)^{\frac{1}{2}}$ for some $M<\infty$ w.p.a.1. To evaluate the expression in (B.24), let

$$u_{jk}^{*}=(u_{jk,12}^{0},...,u_{jk,1r}^{0};u_{jk,23}^{0},...,u_{jk,2r}^{0};...;u_{jk,(r-1)r}^{0};u_{jk,21}^{0},...,u_{jk,r1}^{0};u_{jk,32}^{0},...,u_{jk,r2}^{0};...;u_{jk,r(r-1)}^{0})'.$$

Define u_{kj}^* similarly. The dimension of u_{jk}^* is $\frac{r(r-1)}{2}+\frac{r(r-1)}{2}=r(r-1)$. Let

$$U^{*}=[N^{-\frac{1}{2}}(u_{12}^{*},...,u_{1r}^{*};u_{23}^{*},...,u_{2r}^{*};...;u_{(r-1)r}^{*});T^{-\frac{1}{2}}(u_{21}^{*},...,u_{r1}^{*};u_{32}^{*},...,u_{r2}^{*};...;u_{r(r-1)}^{*})].$$

Then the term in (B.24) is not smaller than $c_1\sigma_{\min}(U^*U^{*\prime})\sum_{p=1}^{r}\sum_{q=p+1}^{r}(a_{pq}^2+a_{qp}^2)$. Under Assumptions 1 and 3, $plimU^*$ is full rank. which is to be proved later. Thus $plimU^*U^{*\prime}$ is positive definite. This implies that there exists $\varpi>0$ such that $\sigma_{\min}(U^*U^{*\prime})\geq\varpi$ w.p.a.1 as $(N,T)\to\infty$. It follows that expression (B.24) is not smaller than $c_1\varpi\sum_{p=1}^{r}\sum_{q=p+1}^{r}(a_{pq}^2+a_{qp}^2)$ w.p.a.1. Finally, equations (B.15)-(B.16) imply that the term (B.23) is $O_p(\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}}+\frac{T^{\frac{1}{\kappa}}}{\sqrt{N}}+\frac{1}{\sqrt{c_{NT}}})$. Then the term term in (B.23) is not larger than $\frac{c_1\varpi}{3}$ w.p.a.1.

Combining the above analyses, we have that w.p.a.1,

$$\begin{aligned}
& a'(-D_{TN}^{-\frac{1}{2}}H_{\phi\phi'}D_{TN}^{-\frac{1}{2}})a \\
\geq\; & b\sum_{l=1}^{(N+T)r-r(r-1)}a_l^2+c_1\varpi\sum_{p=1}^{r}\sum_{q=p+1}^{r}(a_{pq}^2+a_{qp}^2) \\
& -2c_1r(r-1)M(\sum_{l=1}^{(N+T)r-r(r-1)}a_l^2)^{\frac{1}{2}}-\frac{c_1\varpi}{3} \\
=\; & (b-c_1\varpi)\sum_{l=1}^{(N+T)r-r(r-1)}a_l^2+c_1\varpi-2c_1r(r-1)M(\sum_{l=1}^{(N+T)r-r(r-1)}a_l^2)^{\frac{1}{2}}-\frac{c_1\varpi}{3} \\
\geq\; & c_1\varpi-\frac{c_1^2r^2(r-1)^2M^2}{b-c_1\varpi}-\frac{c_1\varpi}{3} \\
=\; & \frac{c_1(b\varpi-c_1\varpi^2-c_1r^2(r-1)^2M^2)}{b-c_1\varpi}-\frac{c_1\varpi}{3}.
\end{aligned} \tag{B.26}$$

When c_1 is small enough, $c_1\varpi^2+c_1r^2(r-1)^2M^2$ is smaller than $\frac{b\varpi}{2}$. Thus when c_1 is small enough, the term on the RHS of (B.26) is not smaller than $\frac{c_1\varpi}{6}$. Taking $C=\frac{c_1\varpi}{6}$, we have proved that $a'(-D_{TN}^{-\frac{1}{2}}H_{\phi\phi'}D_{TN}^{-\frac{1}{2}})a\geq C$ w.p.a.1.

Now, we prove the full rankness of $plimU^*$. We shall prove this explicitly for the case of $r=3$; the other cases can be shown similarly. When $r=3$, after some calculation, we find that U^* is given by

$$\begin{bmatrix} \frac{\frac{1}{N}\sum_{i=1}^{N}(\lambda_{i2}^0)^2}{\left\|D_{NT}^{-\frac{1}{2}}w_{12}^0\right\|} & 0 & 0 & \frac{-\frac{1}{T}\sum_{t=1}^{T}(f_{t1}^0)^2}{\left\|D_{NT}^{-\frac{1}{2}}w_{12}^0\right\|} & 0 & 0 \\ 0 & \frac{\frac{1}{N}\sum_{i=1}^{N}(\lambda_{i3}^0)^2}{\left\|D_{NT}^{-\frac{1}{2}}w_{13}^0\right\|} & 0 & 0 & \frac{-\frac{1}{T}\sum_{t=1}^{T}(f_{t1}^0)^2}{\left\|D_{NT}^{-\frac{1}{2}}w_{13}^0\right\|} & 0 \\ 0 & 0 & \frac{\frac{1}{N}\sum_{i=1}^{N}(\lambda_{i3}^0)^2}{\left\|D_{NT}^{-\frac{1}{2}}w_{23}^0\right\|} & 0 & 0 & \frac{-\frac{1}{T}\sum_{t=1}^{T}(f_{t2}^0)^2}{\left\|D_{NT}^{-\frac{1}{2}}w_{23}^0\right\|} \\ \frac{\frac{1}{N}\sum_{i=1}^{N}(\lambda_{i1}^0)^2}{\left\|D_{NT}^{-\frac{1}{2}}w_{21}^0\right\|} & 0 & 0 & \frac{-\frac{1}{T}\sum_{t=1}^{T}(f_{t2}^0)^2}{\left\|D_{NT}^{-\frac{1}{2}}w_{21}^0\right\|} & 0 & 0 \\ 0 & \frac{\frac{1}{N}\sum_{i=1}^{N}(\lambda_{i1}^0)^2}{\left\|D_{NT}^{-\frac{1}{2}}w_{31}^0\right\|} & 0 & 0 & \frac{-\frac{1}{T}\sum_{t=1}^{T}(f_{t3}^0)^2}{\left\|D_{NT}^{-\frac{1}{2}}w_{31}^0\right\|} & 0 \\ 0 & 0 & \frac{\frac{1}{N}\sum_{i=1}^{N}(\lambda_{i2}^0)^2}{\left\|D_{NT}^{-\frac{1}{2}}w_{32}^0\right\|} & 0 & 0 & \frac{-\frac{1}{T}\sum_{t=1}^{T}(f_{t3}^0)^2}{\left\|D_{NT}^{-\frac{1}{2}}w_{32}^0\right\|} \end{bmatrix}.$$

Note that $\frac{1}{T}\sum_{t=1}^{T}(f_{tp}^0)^2 = \frac{1}{N}\sum_{i=1}^{N}(\lambda_{ip}^0)^2$ for $p = 1, 2, 3$. Now consider $(plimU^*)g = 0$ for any vector g. If $plim\frac{1}{N}\sum_{i=1}^{N}(\lambda_{i1}^0)^2 \neq plim\frac{1}{N}\sum_{i=1}^{N}(\lambda_{i2}^0)^2$, then $g_1 = g_4 = 0$. If $plim\frac{1}{N}\sum_{i=1}^{N}(\lambda_{i1}^0)^2 \neq plim\frac{1}{N}\sum_{i=1}^{N}(\lambda_{i3}^0)^2$, then $g_2 = g_5 = 0$. If $plim\frac{1}{N}\sum_{i=1}^{N}(\lambda_{i2}^0)^2 \neq plim\frac{1}{N}\sum_{i=1}^{N}(\lambda_{i3}^0)^2$, then $g_3 = g_6 = 0$. Thus $g = 0$ by Assumption 3. ∎

Lemma B.2 *Suppose that Assumptions 1, 2(ii) and 3 hold. Then as* $(N, T) \to \infty$, *we have* $||(-D_{f\lambda}^{-\frac{1}{2}}H_{\phi\phi'}D_{f\lambda}^{-\frac{1}{2}})^{-1}|| = O_p(1)$ *and* $||(-D_{TN}^{-\frac{1}{2}}H_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}|| = O_p(1)$, *where*

$$D_{f\lambda} = \begin{bmatrix} I_N \otimes \sum_{t=1}^{T} f_t^0 f_t^{0\prime} & 0 \\ 0 & I_T \otimes \sum_{i=1}^{N} \lambda_i^0 \lambda_i^{0\prime} \end{bmatrix}. \tag{B.27}$$

Proof. For the random missing case, we directly use $D_{TN}^{-\frac{1}{2}}$ to normalize $H_{\phi\phi'}$; here it is crucial to use $D_{f\lambda}^{-\frac{1}{2}}$. For the random missing case, we first prove that $\sigma_{\min}(-D_{TN}^{-\frac{1}{2}}\breve{H}_{\phi\phi'}D_{TN}^{-\frac{1}{2}})$ is positive and bounded away from zero in probability, and then prove that the conclusion still holds when $D_{NT}^{-\frac{1}{2}}W^0W^{0\prime}D_{NT}^{-\frac{1}{2}}$ is replaced by $D_{NT}^{-\frac{1}{2}}U^0U^{0\prime}D_{NT}^{-\frac{1}{2}}$. For the block missing/staggered missing/mixed frequency case, the roadmap is similar, but the technical details are quite different. A key strategy utilized in the proof of Lemma B.1 is that $\mathbb{E}_\phi(d_{it}) > c > 0$ uniformly over i and t, which is no longer true for the block/staggered missing case because d_{it} is always zero for some i and t.

The key step is to calculate all the eigenvalues of $-D_{f\lambda}^{-\frac{1}{2}}L_{\phi\phi'}D_{f\lambda}^{-\frac{1}{2}}$. To do so, we distinguish between the block missing case and the staggered missing case. The results for the former case are summarized in equation (B.42) below and we have confirmed that these calculated eigenvalues are correct using MATLAB program. To help the readers to understand the proof, we also add the proof for the simple case with $r = 1$ after we present the proof for the general case with $r \geq 1$.

Step (1) We show that all of the nonzero eigenvalues of $-D_{f\lambda}^{-\frac{1}{2}}L_{\phi\phi'}D_{f\lambda}^{-\frac{1}{2}}$ are positive and bounded away from zero w.p.a.1.

(I) The block missing case

Step (1.1). Note that $-L_{\phi\phi'} = -L_{\phi\phi' full} - (-L_{\phi\phi' mis})$, where

$$-L_{\phi\phi' full} = \left[\begin{array}{cc} I_N \otimes \sum_{t=1}^T f_t^0 f_t^{0\prime} & (f_t^0\lambda_i^{0\prime})_{Nr\times Tr} \\ (\lambda_i^0 f_t^{0\prime})_{Nr\times Tr} & I_T \otimes \sum_{i=1}^N \lambda_i^0\lambda_i^{0\prime} \end{array}\right], \tag{B.28}$$

$$-L_{\phi\phi' mis} = \left[\begin{array}{cc|cc} 0_{N_or\times N_or} & 0_{N_or\times N_mr} & 0_{N_or\times T_or} & 0_{N_or\times T_mr} \\ 0_{N_mr\times N_or} & I_{N_m}\otimes \sum_{t=T_o+1}^T f_t^0 f_t^{0\prime} & 0_{N_mr\times T_or} & (f_t^0\lambda_i^{0\prime})_{N_mr\times T_mr} \\ \hline 0_{T_or\times N_or} & 0_{T_or\times N_mr} & 0_{T_or\times T_or} & 0_{T_or\times T_mr} \\ 0_{T_mr\times N_or} & (\lambda_i^0 f_t^{0\prime})_{T_mr\times N_mr} & 0_{T_mr\times T_or} & I_{T_m}\otimes \sum_{i=N_o+1}^N \lambda_i^0\lambda_i^{0\prime} \end{array}\right], \tag{B.29}$$

$N_m = N - N_o$, $T_m = T - T_o$, and $(f_t^0\lambda_i^{0\prime})_{N_mr\times T_mr}$ denotes the $N_mr \times T_mr$ matrix with $f_t^0\lambda_i^{0\prime}$ as the (i,t) block for $i = N_o+1,...,N$ and $t = T_o+1,...,T$. In the following we shall calculate all eigenvalues of $-D_{f\lambda}^{-\frac{1}{2}}L_{\phi\phi'}D_{f\lambda}^{-\frac{1}{2}}$.

First, let $f_t^n = (\sum_{t=1}^T f_t^0 f_t^{0\prime})^{-\frac{1}{2}} f_t^0$ and $\lambda_i^n = (\sum_{i=1}^N \lambda_i^0\lambda_i^{0\prime})^{-\frac{1}{2}}\lambda_i^0$ denote the normalized factors and loadings. $-D_{f\lambda}^{-\frac{1}{2}}L_{\phi\phi' full}D_{f\lambda}^{-\frac{1}{2}}$ and $-D_{f\lambda}^{-\frac{1}{2}}L_{\phi\phi' mis}D_{f\lambda}^{-\frac{1}{2}}$ has the same expressions as (B.28) and (B.29), respectively, once we replace f_t^0 by f_t^n and λ_i^0 by λ_i^n. Similarly, we define w_{pp}^n, w_{pq}^n and w_{qp}^n by replacing f_t^0 by f_t^n and λ_i^0 by λ_i^n in w_{pp}^0, w_{pq}^0 and w_{qp}^0. Let

$$\begin{aligned} W^n &= (w_{11}^n,...,w_{rr}^n; w_{12}^n,...,w_{1r}^n, w_{23}^n,...,w_{2r}^n,...,w_{(r-1)r}^n; w_{21}^n,...,w_{r1}^n, w_{32}^n,...,w_{r2}^n,...,w_{r(r-1)}^n) \\ &\equiv (W_\lambda^{n\prime}, W_f^{n\prime})'. \end{aligned} \tag{B.30}$$

It is not difficult to verify that any two different columns of W^n are orthogonal to each other, any column of W^n is orthogonal to both $-D_{f\lambda}^{-\frac{1}{2}}L_{\phi\phi' full}D_{f\lambda}^{-\frac{1}{2}}$ and $-D_{f\lambda}^{-\frac{1}{2}}L_{\phi\phi' mis}D_{f\lambda}^{-\frac{1}{2}}$. We can also verify that the (i,t)th block of $W_\lambda^n W_f^{n\prime}$ is $-f_t^n\lambda_i^{n\prime}$ and $W^nW^{n\prime} = \sum_{p=1}^r\sum_{q=1}^r w_{pq}^n w_{pq}^{n\prime}$.

For $p,q \in [r]$, let $w_{pq,i}^n$ denote the i-th block of w_{pq}^n of size $r \times 1$. [Each block is an $r \times 1$ vector and there are $N + T$ blocks in total.] Let $w_{pq\lambda_o}^n = (w_{pq,1}^{n\prime},...,w_{pq,N_o}^{n\prime})'$, $w_{pq\lambda_m}^n = (w_{pq,N_o+1}^{n\prime},...,w_{pq,N}^{n\prime})'$, $w_{pqf_o}^n = (w_{pq,N+1}^{n\prime},...,w_{pq,N+T_o}^{n\prime})'$, $w_{pqf_m}^n = (w_{pq,N+T_o+1}^{n\prime},...,w_{pq,N+T}^{n\prime})'$, $w_{pq\lambda}^n = (w_{pq\lambda_o}^{n\prime}, w_{pq\lambda_m}^{n\prime})'$ and $w_{pqf}^n = (w_{pqf_o}^{n\prime}, w_{pqf_m}^{n\prime})'$. Then w_{pq}^n can be written as $(w_{pq\lambda}^{n\prime}, w_{pqf}^{n\prime})'$ and

$$\begin{aligned} &-D_{f\lambda}^{-\frac{1}{2}}L_{\phi\phi' full}D_{f\lambda}^{-\frac{1}{2}} - \begin{pmatrix} W_\lambda^n \\ -W_f^n \end{pmatrix}\begin{pmatrix} W_\lambda^n \\ -W_f^n \end{pmatrix}' \\ = &\left[\begin{array}{cc} I_{Nr} & (f_t^n\lambda_i^{n\prime})_{Nr\times Tr} \\ (\lambda_i^n f_t^{n\prime})_{Tr\times Nr} & I_{Tr} \end{array}\right] - \begin{pmatrix} W_\lambda^n \\ -W_f^n \end{pmatrix}\begin{pmatrix} W_\lambda^n \\ -W_f^n \end{pmatrix}' \end{aligned}$$

$$= \begin{bmatrix} I_{Nr} - \sum_{p=1}^r \sum_{q=1}^r w_{pq\lambda}^n w_{pq\lambda}^{n\prime} & 0_{Nr\times Tr} \\ 0_{Tr\times Nr} & I_{Tr} - \sum_{p=1}^r \sum_{q=1}^r w_{pqf}^n w_{pqf}^{n\prime} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{p=1}^r \sum_{j=1}^{N-r} (\xi_j \otimes 1_p^r)(\xi_j \otimes 1_p^r)' & 0_{Nr\times Tr} \\ 0_{Tr\times Nr} & \sum_{p=1}^r \sum_{j=1}^{T-r} (\chi_j \otimes 1_p^r)(\chi_j \otimes 1_p^r)' \end{bmatrix}, \quad \text{(B.31)}$$

where 1_p^r denotes the $r \times 1$ vector with the p-th element being one and all the other elements being zero, and the last equality is due to:

(1) $I_{Nr} = \sum_{p=1}^r I_N \otimes \iota_p$, where ι_p is an $r \times r$ matrix with the p-th diagonal element being one and all the other elements being zero;

(2) $\{\xi_j,\ j \in [N-r]\}$ and $\{\lambda_{\cdot q}^n \equiv (\lambda_{1q}^n, ..., \lambda_{Nq}^n)',\ q \in [r]\}$ together constitute an orthonormal basis for the N dimensional vector space; Since $I_N = \sum_{j=1}^{N-r} \xi_j \xi_j' + \sum_{q=1}^r \lambda_{\cdot q}^n \lambda_{\cdot q}^{n\prime}$ and $w_{pq\lambda}^n = \lambda_{\cdot q}^n \otimes 1_p^r$, we have

$$I_N \otimes \iota_p - \sum_{q=1}^r w_{pq\lambda}^n w_{pq\lambda}^{n\prime} = \sum_{j=1}^{N-r} (\xi_j \otimes 1_p^r)(\xi_j \otimes 1_p^r)';$$

(3) Similarly, $\{\chi_j, j \in [T-r]\}$ and $\{f_{\cdot q}^n \equiv (f_{1q}^n, ..., f_{Tq}^n)', q \in [r]\}$ together constitutes an orthonormal basis for the T dimensional vector space. Since $I_T = \sum_{j=1}^{T-r} \chi_j \chi_j' + \sum_{q=1}^r f_{\cdot q}^n f_{\cdot q}^{n\prime}$ and $w_{pqf}^n = f_{\cdot q}^n \otimes 1_p^r$, we have

$$I_T \otimes \iota_p - \sum_{q=1}^r w_{pqf}^n w_{pqf}^{n\prime} = \sum_{j=1}^{T-r} (\chi_j \otimes 1_p^r)(\chi_j \otimes 1_p^r)'.$$

How to choose ξ_j and χ_j will be discussed later.

Step (1.2). Similarly, we also have

$$-D_{f\lambda}^{-\frac{1}{2}} L_{\phi\phi' mis} D_{f\lambda}^{-\frac{1}{2}} - A_3$$

$$= \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & I_{N_m} \otimes \sum_{t=T_o+1}^T f_t^n f_t^{n\prime} & 0 & (f_t^n \lambda_i^{n\prime})_{N_m r \times T_m r} \\ \hline 0 & 0 & 0 & 0 \\ 0 & (\lambda_i^n f_t^{n\prime})_{T_m r \times N_m r} & 0 & I_{T_m} \otimes \sum_{i=N_o+1}^N \lambda_i^n \lambda_i^{n\prime} \end{array}\right] - A_3 = \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & A_1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & A_2 \end{array}\right], \quad \text{(B.32)}$$

where the dimension of the zero matrices are self-evident, the second equality is due to $-\sum_{p=1}^r \sum_{q=1}^r w_{pq\lambda_m}^n w_{pqf_m}^{n\prime} = (f_t^n \lambda_i^{n\prime})_{N_m r \times T_m r}$, and

$$A_1 = I_{N_m} \otimes \sum_{t=T_o+1}^T f_t^n f_t^{n\prime} - \sum_{p=1}^r \sum_{q=1}^r \left\| w_{pqf_m}^n \right\|^2 \frac{w_{pq\lambda_m}^n}{\left\| w_{pq\lambda_m}^n \right\|} \frac{w_{pq\lambda_m}^{n\prime}}{\left\| w_{pq\lambda_m}^n \right\|},$$

$$A_2 = I_{T_m} \otimes \sum_{i=N_o+1}^N \lambda_i^n \lambda_i^{n\prime} - \sum_{p=1}^r \sum_{q=1}^r \left\| w_{pq\lambda_m}^n \right\|^2 \frac{w_{pqf_m}^n}{\left\| w_{pqf_m}^n \right\|} \frac{w_{pqf_m}^{n\prime}}{\left\| w_{pqf_m}^n \right\|},$$

$$
\begin{aligned}
A_3 &= \sum\nolimits_{p=1}^{r}\sum\nolimits_{q=1}^{r} A_{3pq}A_{3pq}', \\
A_{3pq} &\equiv (0_{1\times N_o r}, \frac{\left\|w_{pqf_m}^{n}\right\|}{\left\|w_{pq\lambda_m}^{n}\right\|} w_{pq\lambda_m}^{n\prime}, 0_{1\times T_o r}, -\frac{\left\|w_{pq\lambda_m}^{n}\right\|}{\left\|w_{pqf_m}^{n}\right\|} w_{pqf_m}^{n\prime})'.
\end{aligned}
$$

Without loss of generality, we can consider $\sum_{t=T_o+1}^{T} f_t^n f_t^{n\prime}$ and $\sum_{i=N_o+1}^{N} \lambda_i^n \lambda_i^{n\prime}$ as diagonal. For example, if $\sum_{t=T_o+1}^{T} f_t^n f_t^{n\prime}$ is nondiagonal, we can replace f_t^n by $f_t^{nn} = \Gamma_{f_m f_m}' f_t^n = \Gamma_{f_m f_m}' (\sum_{t=1}^{T} f_t^0 f_t^{0\prime})^{-\frac{1}{2}} f_t^0$, where $\Gamma_{f_m f_m}$ is the eigenvector matrix of $\sum_{t=T_o+1}^{T} f_t^n f_t^{n\prime}$. This makes $\sum_{t=T_o+1}^{T} f_t^{nn} f_t^{nn\prime}$ diagonal and does not change the eigenvalues. When $\sum_{t=T_o+1}^{T} f_t^n f_t^{n\prime}$ is diagonal,

$$
I_{N_m} \otimes \sum\nolimits_{t=T_o+1}^{T} f_t^n f_t^{n\prime} = \sum\nolimits_{p=1}^{r} (\sum\nolimits_{t=T_o+1}^{T} (f_{tp}^n)^2) I_{N_m} \otimes \iota_p.
$$

For each given p, $w_{pq\lambda_m}^n = (\lambda_{(N_o+1)q}^n, ..., \lambda_{Nq}^n)' \otimes 1_p^r$ and $\left\|w_{pqf_m}^n\right\|^2 = \sum_{t=T_o+1}^{T} (f_{tp}^n)^2$ for all $q \in [r]$. It follows that

$$
\begin{aligned}
A_1 &= \sum\nolimits_{p=1}^{r} (\sum\nolimits_{t=T_o+1}^{T} (f_{tp}^n)^2) I_{N_m} \otimes \iota_p \\
&\quad - \sum\nolimits_{p=1}^{r}\sum\nolimits_{q=1}^{r} (\sum\nolimits_{t=T_o+1}^{T} (f_{tp}^n)^2) \frac{w_{pq\lambda_m}^n}{\left\|w_{pq\lambda_m}^n\right\|} \frac{w_{pq\lambda_m}^{n\prime}}{\left\|w_{pq\lambda_m}^n\right\|} \\
&= \sum\nolimits_{p=1}^{r} (\sum\nolimits_{t=T_o+1}^{T} (f_{tp}^n)^2) [I_{N_m} \otimes \iota_p - \sum\nolimits_{q=1}^{r} \frac{w_{pq\lambda_m}^n}{\left\|w_{pq\lambda_m}^n\right\|} \frac{w_{pq\lambda_m}^{n\prime}}{\left\|w_{pq\lambda_m}^n\right\|}] \\
&= \sum\nolimits_{p=1}^{r} (\sum\nolimits_{t=T_o+1}^{T} (f_{tp}^n)^2) [\sum\nolimits_{j=1}^{N_m - r} (\xi_{mj} \otimes 1_p^r)(\xi_{mj} \otimes 1_p^r)'], \qquad \text{(B.33)}
\end{aligned}
$$

where $\{\xi_{mj}, j \in [N_m - r]\}$ and $\{\lambda_{mq}^n \equiv (\lambda_{(N_o+1)q}^n, ..., \lambda_{Nq}^n)', q \in [r]\}$ together constitute an orthonormal basis for the N_m dimensional vector space. Similarly, when $\sum_{i=N_o+1}^{N} \lambda_i^n \lambda_i^{n\prime}$ is diagonal,

$$
I_{T_m} \otimes \sum\nolimits_{i=N_o+1}^{N} \lambda_i^n \lambda_i^{n\prime} = \sum\nolimits_{q=1}^{r} (\sum\nolimits_{i=N_o+1}^{N} (\lambda_{iq}^n)^2) I_{T_m} \otimes \iota_q.
$$

For each given q, $w_{pqf_m}^n = (f_{(T_o+1)p}^n, ..., f_{Tp}^n)' \otimes 1_q^r$ and $\left\|w_{pq\lambda_m}^n\right\|^2 = \sum_{i=N_o+1}^{N} (\lambda_{iq}^n)^2$ for all $p \in [r]$. Thus

$$
\begin{aligned}
A_2 &= \sum\nolimits_{q=1}^{r} (\sum\nolimits_{i=N_o+1}^{N} (\lambda_{iq}^n)^2) I_{T_m} \otimes \iota_q \\
&\quad - \sum\nolimits_{p=1}^{r}\sum\nolimits_{q=1}^{r} (\sum\nolimits_{i=N_o+1}^{N} (\lambda_{iq}^n)^2) \frac{w_{pqf_m}^n}{\left\|w_{pqf_m}^n\right\|} \frac{w_{pqf_m}^{n\prime}}{\left\|w_{pqf_m}^n\right\|} \\
&= \sum\nolimits_{q=1}^{r} (\sum\nolimits_{i=N_o+1}^{N} (\lambda_{iq}^n)^2) [I_{T_m} \otimes \iota_q - \sum\nolimits_{p=1}^{r} \frac{w_{pqf_m}^n}{\left\|w_{pqf_m}^n\right\|} \frac{w_{pqf_m}^{n\prime}}{\left\|w_{pqf_m}^n\right\|}]
\end{aligned}
$$

$$= \sum_{q=1}^{r}(\sum_{i=N_o+1}^{N}(\lambda_{iq}^n)^2)[\sum_{j=1}^{T_m-r}(\chi_{mj}\otimes 1_q^r)(\chi_{mj}\otimes 1_q^r)'], \quad \text{(B.34)}$$

where $\{\chi_{mj}, j\in[T_m-r]\}$ and $\{f_{mp}^n \equiv (f_{(T_o+1)p}^n,...,f_{Tp}^n)', p\in[r]\}$ together constitute an orthonormal basis for the T_m dimensional vector space.

Step (1.3). Since $\sum_{i=N_o+1}^{N}\lambda_i^n\lambda_i^{n\prime}$ is diagonal, $\lambda_{mp}^{n\prime}\lambda_{mq}^n=0$ for any $p\neq q$. Since $\sum_{i=1}^{N}\lambda_i^n\lambda_i^{n\prime}=I_r$, $\sum_{i=1}^{N_o}\lambda_i^n\lambda_i^{n\prime}$ is also diagonal, implying that $\lambda_{op}^{n\prime}\lambda_{oq}^n=0$ for any $p\neq q$ where $\lambda_{oq}^n\equiv(\lambda_{1q}^n,...,\lambda_{N_oq}^n)'$. Choose ξ_{oj} such that $\{\xi_{oj}, j\in[N_o-r]\}$ and $\{\lambda_{oq}^n, q\in[r]\}$ constitute an orthonormal basis for the N_o dimensional vector space. Then the following set of vectors are orthonormal,

$$\begin{pmatrix}\xi_{oj}\\ 0_{N_m\times 1}\end{pmatrix}_{j\in[N_o-r]}, \begin{pmatrix}0_{N_o\times 1}\\ \xi_{mj'}\end{pmatrix}_{j'\in[N_m-r]}, \begin{pmatrix}\frac{\|\lambda_{mq}^n\|}{\|\lambda_{oq}^n\|}\lambda_{oq}^n\\ -\frac{\|\lambda_{oq}^n\|}{\|\lambda_{mq}^n\|}\lambda_{mq}^n\end{pmatrix}_{q\in[r]}, \begin{pmatrix}\lambda_{oq'}^n\\ \lambda_{mq'}^n\end{pmatrix}_{q'\in[r]}. \quad \text{(B.35)}$$

The first three sets contain $N-r$ vectors in total, and we choose them as ξ_j for $j\in[N-r]$.

Since $\sum_{t=T_o+1}^{T}f_t^nf_t^{n\prime}$ is diagonal, $f_{mp}^{n\prime}f_{mq}^n=0$ for any $p\neq q$. Since $\sum_{t=1}^{T}f_t^nf_t^{n\prime}=I_r$, $\sum_{t=1}^{T_o}f_t^nf_t^{n\prime}$ is also diagonal, implying that $f_{op}^{n\prime}f_{oq}^n=0$ for any $p\neq q$, where $f_{oq}^n\equiv(f_{1q}^n,...,f_{T_oq}^n)'$. Choose χ_{oj} such that $\{\chi_{oj}, j\in[T_o-r]\}$ and $\{f_{oq}^n, q\in[r]\}$ constitute an orthonormal basis for the T_o dimensional vector space. Then the following set of vectors are orthonormal,

$$\begin{pmatrix}\chi_{oj}\\ 0_{T_m\times 1}\end{pmatrix}_{j\in[T_o-r]}, \begin{pmatrix}0_{T_o\times 1}\\ \chi_{mj'}\end{pmatrix}_{j'\in[T_m-r]}, \begin{pmatrix}\frac{\|f_{mq}^n\|}{\|f_{oq}^n\|}f_{oq}^n\\ -\frac{\|f_{oq}^n\|}{\|f_{mq}^n\|}f_{mq}^n\end{pmatrix}_{q\in[r]}, \begin{pmatrix}f_{oq'}^n\\ f_{mq'}^n\end{pmatrix}_{q'\in[r]}. \quad \text{(B.36)}$$

The first three sets contain $T-r$ vectors in total, and we choose them as χ_j for $j\in[N-r]$.

In addition, we can see that the columns of $\begin{pmatrix}W_\lambda^n\\ -W_f^n\end{pmatrix}$ and $\begin{pmatrix}W_\lambda^n\\ W_f^n\end{pmatrix}$ are

$$\begin{pmatrix}w_{pq\lambda}^n\\ -w_{pqf}^n\end{pmatrix} = \begin{pmatrix}\lambda_{\cdot q}^n\otimes 1_p^r\\ f_{\cdot p}^n\otimes 1_q^r\end{pmatrix} \text{ for } p,q\in[r], \text{ and} \quad \text{(B.37)}$$

$$\begin{pmatrix}w_{pq\lambda}^n\\ w_{pqf}^n\end{pmatrix} = \begin{pmatrix}\lambda_{\cdot q}^n\otimes 1_p^r\\ -f_{\cdot p}^n\otimes 1_q^r\end{pmatrix} \text{ for } p,q\in[r], \quad \text{(B.38)}$$

respectively. The $2r^2$ vectors listed in expressions (B.37)-(B.38) are orthogonal to each other and also orthogonal to the eigenvectors listed in expressions (B.35)-(B.36).

Let $\mathcal{A}\equiv -D_{f\lambda}^{-\frac{1}{2}}L_{\phi\phi' full}D_{f\lambda}^{-\frac{1}{2}}-(-D_{f\lambda}^{-\frac{1}{2}}L_{\phi\phi' mis}\ D_{f\lambda}^{-\frac{1}{2}}-A_3)$. From expressions (B.31)-(B.38), we can see that the $(N+T)\,r$ eigenvectors of $\mathcal{A}$ are given by

$$
\left.\begin{array}{c|c|c|c}
\underset{j\in[N_o-r],p\in[r]}{\begin{pmatrix} \xi_{oj} \\ 0_{N_m\times 1} \\ 0_{T_o\times 1} \\ 0_{T_m\times 1} \end{pmatrix}\otimes 1_p^r}, &
\underset{j'\in[N_m-r],p\in[r]}{\begin{pmatrix} 0_{N_o\times 1} \\ \xi_{mj'} \\ 0_{T_o\times 1} \\ 0_{T_m\times 1} \end{pmatrix}\otimes 1_p^r}, &
\underset{q\in[r],p\in[r]}{\begin{pmatrix} \frac{\|\lambda_{mq}^n\|}{\|\lambda_{oq}^n\|}\lambda_{oq}^n \\ \frac{-\|\lambda_{oq}^n\|}{\|\lambda_{mq}^n\|}\lambda_{mq}^n \\ 0_{T_o\times 1} \\ 0_{T_o\times 1} \end{pmatrix}\otimes 1_p^r}, &
\underset{q\in[r],p\in[r]}{\begin{pmatrix} \lambda_{\cdot q}^n\otimes 1_p^r \\ f_{\cdot p}^n\otimes 1_q^r \end{pmatrix}}, \\
\hline
\underset{j\in[T_o-r],p\in[r]}{\begin{pmatrix} 0_{N_o\times 1} \\ 0_{N_m\times 1} \\ \chi_{oj} \\ 0_{T_m\times 1} \end{pmatrix}\otimes 1_p^r}, &
\underset{j'\in[T_m-r],p\in[r]}{\begin{pmatrix} 0_{N_o\times 1} \\ 0_{N_m\times 1} \\ 0_{T_o\times 1} \\ \chi_{mj'} \end{pmatrix}\otimes 1_p^r}, &
\underset{q\in[r],p\in[r]}{\begin{pmatrix} 0_{N_o\times 1} \\ 0_{N_m\times 1} \\ \frac{\|f_{mq}^n\|}{\|f_{oq}^n\|}f_{oq}^n \\ \frac{-\|f_{oq}^n\|}{\|f_{mq}^n\|}f_{mq}^n \end{pmatrix}\otimes 1_p^r}, &
\underset{q\in[r],p\in[r]}{\begin{pmatrix} \lambda_{\cdot q}^n\otimes 1_p^r \\ -f_{\cdot p}^n\otimes 1_q^r \end{pmatrix}}
\end{array}\right. ; \tag{B.39}
$$

and the corresponding eigenvalues are

$$
\begin{array}{c|c|c|c}
\underset{((N_o-r)r\text{ times})}{1\text{ for all }j,p\ ,} &
\underset{(N_m-r\text{ times for each }p)}{\sum_{t=1}^{T_o}(f_{tp}^n)^2\text{ for all }j',} &
\underset{(r^2\ \textbf{times})}{1\ \textbf{for all}\ (p,q),} &
\underset{(r^2\ \textbf{times})}{2\ \textbf{for all}\ (p,q),} \\
\hline
\underset{((T_o-r)r\text{ times})}{1\text{ for all }j,p} &
\underset{(T_m-r\text{ times for each }p)}{\sum_{i=1}^{N_o}(\lambda_{ip}^n)^2\text{ for all }j'} &
\underset{(r^2\ \textbf{times})}{1\ \textbf{for all}\ (p,q)} &
\underset{(r^2\text{ times})}{0\text{ for all }(p,q)}
\end{array}. \tag{B.40}
$$

That is, among the $(N+T)\,r$ eigenvalues of $\mathcal{A}$, $(N_o+T_o)r$ of them are 1, r^2 of them 2, r^2 of them are 0, and the rest are also positive w.p.a.1.

Step (1.4). Now consider A_3. Given that $\left\|w_{pqf_m}^n\right\| = \|f_{mp}^n\|$, $\left\|w_{pq\lambda_m}^n\right\| = \|\lambda_{mq}^n\|$, $w_{pq\lambda_m}^n = \lambda_{mq}^n\otimes 1_p^r$ and $w_{pqf_m}^n = f_{mp}^n\otimes 1_q^r$, we have

$$
\begin{aligned}
A_{3pq} &= \begin{pmatrix} 0_{N_or\times 1} \\ \frac{\|w_{pqf_m}^n\|}{\|w_{pq\lambda_m}^n\|}w_{pq\lambda_m}^n \\ 0_{T_or\times 1} \\ -\frac{\|w_{pq\lambda_m}^n\|}{\|w_{pqf_m}^n\|}w_{pqf_m}^n \end{pmatrix} = \begin{pmatrix} 0_{N_or\times 1} \\ \frac{\|f_{mp}^n\|}{\|\lambda_{mq}^n\|}\lambda_{mq}^n\otimes 1_p^r \\ 0_{T_or\times 1} \\ \frac{\|\lambda_{mq}^n\|}{\|f_{mp}^n\|}f_{mp}^n\otimes 1_q^r \end{pmatrix} \\
&= \|f_{mp}^n\|\,\|\lambda_{mq}^n\| \\
&\quad\times\left[-\frac{\|\lambda_{oq}^n\|}{\|\lambda_{mq}^n\|}\begin{pmatrix} \frac{\|\lambda_{mq}^n\|}{\|\lambda_{oq}^n\|}\lambda_{oq}^n \\ -\frac{\|\lambda_{oq}^n\|}{\|\lambda_{mq}^n\|}\lambda_{mq}^n \\ 0_{T_o\times 1} \\ 0_{T_m\times 1} \end{pmatrix}\otimes 1_p^r - \frac{\|f_{op}^n\|}{\|f_{mp}^n\|}\begin{pmatrix} 0_{N_o\times 1} \\ 0_{N_m\times 1} \\ \frac{\|f_{mp}^n\|}{\|f_{op}^n\|}f_{op}^n \\ -\frac{\|f_{op}^n\|}{\|f_{mp}^n\|}f_{mp}^n \end{pmatrix}\otimes 1_q^r + \sqrt{2}\begin{pmatrix} \frac{\lambda_{\cdot q}^n\otimes 1_p^r}{\sqrt{2}} \\ \frac{f_{\cdot p}^n\otimes 1_q^r}{\sqrt{2}} \end{pmatrix}\right].
\end{aligned}
$$

The eigenvalues of $\mathcal{A}$ corresponding to the three vectors in the square brackets of the last displayed line are $1, 1$, and 2, respectively. After subtracting $A_{3pq}A_{3pq}'$, only these three eigenvalues are affected, and they become the eigenvalues of

$$\left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} - \vartheta_{pq}\vartheta_{pq}'\right], \tag{B.41}$$

where $\vartheta_{pq} \equiv \begin{pmatrix} -\left\|f_{mp}^n\right\|\left\|\lambda_{oq}^n\right\| \\ -\left\|f_{op}^n\right\|\left\|\lambda_{mq}^n\right\| \\ \sqrt{2}\left\|f_{mp}^n\right\|\left\|\lambda_{mq}^n\right\| \end{pmatrix}$. Obviously, the 3×1 vector that is orthogonal to both $(0,0,1)'$ and ϑ_{pq} is an eigenvector of (B.41), and the corresponding eigenvalue is 1. The remaining two eigenvectors of (B.41) should lie in the space orthogonal to this eigenvector, thus could be written as linear combinations of $(0,0,1)'$ and ϑ_{pq}, or (for simplicity) linear combinations of $(0,0,1)'$ and $\left(-\left\|f_{mp}^n\right\|\left\|\lambda_{oq}^n\right\|, -\left\|f_{op}^n\right\|\left\|\lambda_{mq}^n\right\|, 0\right)'$. Suppose vv is an eigenvector of (B.41) and $vv = x(-\left\|f_{mp}^n\right\|\left\|\lambda_{oq}^n\right\|, -\left\|f_{op}^n\right\|\left\|\lambda_{mq}^n\right\|, 0)' + y(0,0,1)'$. Pre-multiplying vv by (B.41), we have

$$\begin{aligned} x - [x(\left\|f_{mp}^n\right\|^2\left\|\lambda_{oq}^n\right\|^2 + \left\|f_{op}^n\right\|^2\left\|\lambda_{mq}^n\right\|^2) + y\sqrt{2}\left\|f_{mp}^n\right\|\left\|\lambda_{mq}^n\right\|] &= cx, \\ 2y - \sqrt{2}\left\|f_{mp}^n\right\|\left\|\lambda_{mq}^n\right\|[x(\left\|f_{mp}^n\right\|^2\left\|\lambda_{oq}^n\right\|^2 + \left\|f_{op}^n\right\|^2\left\|\lambda_{mq}^n\right\|^2) + y\sqrt{2}\left\|f_{mp}^n\right\|\left\|\lambda_{mq}^n\right\|] &= cy, \end{aligned}$$

where c represents the eigenvalue associated with the eigenvector vv. The solution of these two equations are

$$\begin{aligned} c_{pq1} &\equiv \frac{\left\|f_{op}^n\right\|^2 + \left\|\lambda_{oq}^n\right\|^2 + 1}{2} + \sqrt{\left(\frac{\left\|f_{op}^n\right\|^2 + \left\|\lambda_{oq}^n\right\|^2 + 1}{2}\right)^2 - 2\left\|f_{op}^n\right\|^2\left\|\lambda_{oq}^n\right\|^2}, \\ c_{pq2} &\equiv \frac{\left\|f_{op}^n\right\|^2 + \left\|\lambda_{oq}^n\right\|^2 + 1}{2} - \sqrt{\left(\frac{\left\|f_{op}^n\right\|^2 + \left\|\lambda_{oq}^n\right\|^2 + 1}{2}\right)^2 - 2\left\|f_{op}^n\right\|^2\left\|\lambda_{oq}^n\right\|^2}. \end{aligned}$$

Thus the three eigenvalues of (B.41) are $(1, c_{pq1}, c_{pq2})$. In summary, the eigenvalues of $-D_{f\lambda}^{-\frac{1}{2}} L_{\phi\phi'} D_{f\lambda}^{-\frac{1}{2}}$ are:

1 for all j, p $((N_o-r)r times)$	$\sum_{t=1}^{T_o}(f_{tp}^n)^2$ for all j' (N_m-r times for each p)	c_{pq2} **for each** (p,q)	c_{pq1} **for each** (p,q)
1 for all j, p $((T_o-r)r times)$	$\sum_{i=1}^{N_o}(\lambda_{ip}^n)^2$ for all j' (T_m-r times for each p)	1 **for all** (p,q) (r^2 **times**)	0 for all (p,q) (r^2 times)

. (B.42)

For each (p,q), due to subtracting $A_{3pq}A_{3pq}'$, the eigenvectors of the three bold blocks in expression (B.42) are orthogonal rotations of the eigenvectors of the three bold blocks in expression (B.40). All of the nonzero eigenvalues are positive and bounded away from zero w.p.a.1.

(II) The staggered missing case.

Let N_o and T_o denote the cardinality of $\{i : d_{it} = 1 \text{ for all } t \in [T]\}$ and $\{t : d_{it} = 1 \text{ for all } i \in [N]\}$, respectively.

$$
\begin{aligned}
-L_{\phi\phi'} &= \sum\nolimits_{i\leq N_o \text{ or } t\leq T_o} d_{it} \begin{pmatrix} 1_i^N \otimes f_t^0 \\ 1_t^T \otimes \lambda_i^0 \end{pmatrix} \begin{pmatrix} 1_i^N \otimes f_t^0 \\ 1_t^T \otimes \lambda_i^0 \end{pmatrix}' \\
&\quad + \sum\nolimits_{i> N_o \text{ and } t> T_o} d_{it} \begin{pmatrix} 1_i^N \otimes f_t^0 \\ 1_t^T \otimes \lambda_i^0 \end{pmatrix} \begin{pmatrix} 1_i^N \otimes f_t^0 \\ 1_t^T \otimes \lambda_i^0 \end{pmatrix}'. \qquad \text{(B.43)}
\end{aligned}
$$

If we throw away the entries of $\{(i,t) : i > N_o \text{ and } t > T_o\}$ so that the data matrix becomes a block missing matrix, $-L_{\phi\phi'}$ would be equal to the first term on the RHS. Since the second term on the RHS of (B.43) is positive semi-definite and we have proved the result for the block missing case, this case is also proved.

Step (2). Step (1) shows that all of the nonzero eigenvalues of $-D_{f\lambda}^{-\frac{1}{2}} L_{\phi\phi'} D_{f\lambda}^{-\frac{1}{2}}$ are positive and bounded away from zero w.p.a.1. From expression (B.39) we know that the eigenvectors corresponding to the r^2 zero eigenvalues of $-D_{f\lambda}^{-\frac{1}{2}} L_{\phi\phi'} D_{f\lambda}^{-\frac{1}{2}}$ are $\{\begin{pmatrix} \lambda_{\cdot q}^n \otimes 1_p^r \\ -f_{\cdot p}^n \otimes 1_q^r \end{pmatrix}, \ p,q \in [r]\}$, which are the columns of $D_{\lambda f}^{-\frac{1}{2}} W^0$. Thus if we define $\breve{H}_{\phi\phi'} = L_{\phi\phi'} - cD_{f\lambda}^{\frac{1}{2}} D_{\lambda f}^{-\frac{1}{2}} W^0 W^{0\prime} D_{\lambda f}^{-\frac{1}{2}} D_{f\lambda}^{\frac{1}{2}}$, then all eigenvalues of $-D_{f\lambda}^{-\frac{1}{2}} \breve{H}_{\phi\phi'} D_{f\lambda}^{-\frac{1}{2}}$ are positive and bounded away from zero w.p.a.1. The rest of the proof is the same as Step (2) of Lemma B.1 once we replace $\bar{L}_{\phi\phi'}$ by $L_{\phi\phi'}$, D_{TN} by $D_{f\lambda}$ and D_{NT} by $D_{\lambda f}$.

Combining the above results yields that $||(-D_{f\lambda}^{-\frac{1}{2}} H_{\phi\phi'} D_{f\lambda}^{-\frac{1}{2}})^{-1}|| = O_p(1)$. Now, let $\Xi_{NT} = -D_{f\lambda}^{-\frac{1}{2}} H_{\phi\phi'} D_{f\lambda}^{-\frac{1}{2}}$. Noting that $-D_{TN}^{-\frac{1}{2}} H_{\phi\phi'} D_{TN}^{-\frac{1}{2}} = D_{TN}^{-\frac{1}{2}} D_{f\lambda}^{\frac{1}{2}} \Xi_{NT} D_{f\lambda}^{\frac{1}{2}} D_{TN}^{-\frac{1}{2}}$, we have

$$
\sigma_{\min}(-D_{TN}^{-\frac{1}{2}} H_{\phi\phi'} D_{TN}^{-\frac{1}{2}}) \geq \sigma_{\min}\left(\Xi_{NT}\right) \sigma_{\min}(D_{TN}^{-\frac{1}{2}} D_{f\lambda} D_{TN}^{-\frac{1}{2}}).
$$

By Assumption 1, $\sigma_{\min}(D_{TN}^{-\frac{1}{2}} D_{f\lambda} D_{TN}^{-\frac{1}{2}})$ is bounded away from zero w.p.a.1. This, in conjunction with the above conclusion on Ξ_{NT}, implies that $\sigma_{\min}(-D_{TN}^{-\frac{1}{2}} H_{\phi\phi'} D_{TN}^{-\frac{1}{2}})$ is also bounded away from zero w.p.a.1. Thus $||(-D_{TN}^{-\frac{1}{2}} H_{\phi\phi'} D_{TN}^{-\frac{1}{2}})^{-1}|| = O_p\left(1\right)$. ■

Proof of Lemma B.2 for the simple case $r = 1$:

In this case, the above proof can be greatly simplified. Below we outline the key Steps (1.1)–(1.3).

Step (1.1). Note that $-L_{\phi\phi'} = (-L_{\phi\phi' full}) - (-L_{\phi\phi' mis})$, where

$$
-L_{\phi\phi' full} = \begin{bmatrix} I_N \times f^{0\prime} f^0 & \lambda^0 f^{0\prime} \\ f^0 \lambda^{0\prime} & I_T \times \lambda^{0\prime}\lambda^0 \end{bmatrix}, \qquad \text{(B.44)}
$$

$$
-L_{\phi\phi' mis} = \left[\begin{array}{cc|cc} 0_{N_o\times N_o} & 0_{N_o\times N_m} & 0_{N_o\times T_o} & 0_{N_o\times T_m} \\ 0_{N_m\times N_o} & I_{N_m} \times f_m^{0\prime} f_m^0 & 0_{N_m\times T_o} & \lambda_m^0 f_m^{0\prime} \\ \hline 0_{T_o\times N_o} & 0_{T_o\times N_m} & 0_{T_o\times T_o} & 0_{T_o\times T_m} \\ 0_{T_m\times N_o} & f_m^0 \lambda_m^{0\prime} & 0_{T_m\times T_o} & I_{T_m} \times \lambda_m^{0\prime}\lambda_m^0 \end{array}\right], \qquad \text{(B.45)}
$$

$N_m = N - N_o$, $T_m = T - T_o$, $\lambda^0_o = (\lambda^0_1, ..., \lambda^0_{N_o})'$, $\lambda^0_m = (\lambda^0_{N_o+1}, ..., \lambda^0_N)'$, $f^0_o = (f^0_1, ..., f^0_{T_o})$ and $f^0_m = (f^0_{T_o+1}, ..., f^0_T)$. In the following we shall calculate all eigenvalues of $-D_{f\lambda}^{-\frac{1}{2}} L_{\phi\phi'} D_{f\lambda}^{-\frac{1}{2}}$.

First, let $f^n_t = (\sum_{t=1}^T (f^0_t)^2)^{-\frac{1}{2}} f^0_t$ and $\lambda^n_i = (\sum_{i=1}^N (\lambda^0_i)^2)^{-\frac{1}{2}} \lambda^0_i$ denote the normalized factors and loadings, and let $\lambda^n_o = (\lambda^n_1, ..., \lambda^n_{N_o})'$, $\lambda^n_m = (\lambda^n_{N_o+1}, ..., \lambda^n_N)'$, $\lambda^n = (\lambda^{n\prime}_o, \lambda^{n\prime}_m)'$, $f^n_o = (f^n_1, ..., f^n_{T_o})$, $f^n_m = (f^n_{T_o+1}, ..., f^n_T)$ and $f^n = (f^{n\prime}_o, f^{n\prime}_m)'$. According to the definition of w_{pp} at the beginning of Appendix B, we have $w^0_{11} = (\lambda^{0\prime}, -f^{0\prime})'$ when $r = 1$. Similarly, let $w^n_{11} = (\lambda^{n\prime}, -f^{n\prime})'$. It is not difficult to verify that w^n_{11} is orthogonal to both $-D_{f\lambda}^{-\frac{1}{2}} L_{\phi\phi' full} D_{f\lambda}^{-\frac{1}{2}}$ and $-D_{f\lambda}^{-\frac{1}{2}} L_{\phi\phi' mis} D_{f\lambda}^{-\frac{1}{2}}$, and the (i,t)th element of $-\lambda^n f^{n\prime}$ is $-f^n_t \lambda^n_i$. It follows that

$$\begin{aligned} -D_{f\lambda}^{-\frac{1}{2}} L_{\phi\phi' full} D_{f\lambda}^{-\frac{1}{2}} - \begin{pmatrix} \lambda^n \\ f^n \end{pmatrix} \begin{pmatrix} \lambda^n \\ f^n \end{pmatrix}' &= \begin{bmatrix} I_N & \lambda^n f^{n\prime} \\ f^n \lambda^{n\prime} & I_T \end{bmatrix} - \begin{pmatrix} \lambda^n \\ f^n \end{pmatrix} \begin{pmatrix} \lambda^n \\ f^n \end{pmatrix}' \\ = \begin{bmatrix} I_N - \lambda^n \lambda^{n\prime} & 0_{N\times T} \\ 0_{T\times N} & I_T - f^n f^{n\prime} \end{bmatrix} &= \begin{bmatrix} \sum_{j=1}^{N-1} \xi_j \xi_j' & 0_{N\times T} \\ 0_{T\times N} & \sum_{j=1}^{T-1} \chi_j \chi_j' \end{bmatrix}, \end{aligned} \tag{B.46}$$

where $\{\xi_j, j \in [N-1]\}$ and λ^n together constitute an orthonormal basis for the N dimensional vector space, and $\{\chi_j, j \in [T-1]\}$ and f^n together constitutes an orthonormal basis for the T dimensional vector space. How to choose ξ_j and χ_j will be discussed later.

Step (1.2). Similarly, we also have

$$-D_{f\lambda}^{-\frac{1}{2}} L_{\phi\phi' mis} D_{f\lambda}^{-\frac{1}{2}} = \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & I_{N_m} \times f^{n\prime}_m f^n_m & 0 & \lambda^n_m f^{n\prime}_m \\ \hline 0 & 0 & 0 & 0 \\ 0 & f^n_m \lambda^{n\prime}_m & 0 & I_{T_m} \times \lambda^{n\prime}_m \lambda^n_m \end{array}\right] = \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & A_1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & A_2 \end{array}\right] + A_3, \tag{B.47}$$

where the dimension of the zero matrices are self-evident and

$$\begin{aligned} A_1 &= I_{N_m} \times f^{n\prime}_m f^n_m - \|f^n_m\|^2 \frac{\lambda^n_m}{\|\lambda^n_m\|} \frac{\lambda^{n\prime}_m}{\|\lambda^n_m\|}, \\ A_2 &= I_{T_m} \times \lambda^{n\prime}_m \lambda^n_m - \|\lambda^n_m\|^2 \frac{f^n_m}{\|f^n_m\|} \frac{f^{n\prime}_m}{\|f^n_m\|}, \\ A_3 &= A_{311} A_{311}', \; A_{311} \equiv (0_{1\times N_o}, \|f^n_m\| \frac{\lambda^n_m}{\|\lambda^n_m\|}, 0_{1\times T_o}, \|\lambda^n_m\| \frac{f^{n\prime}_m}{\|f^n_m\|})'. \end{aligned} \tag{B.48}$$

Similar to equation (B.46), let $\{\xi_{mj}, j \in [N_m - 1]\}$ and $\frac{\lambda^n_m}{\|\lambda^n_m\|}$ constitute an orthonormal basis for the N_m dimensional vector space, and $\{\chi_{mj}, j \in [T_m - 1]\}$ and $\frac{f^n_m}{\|f^n_m\|}$ constitute an orthonormal basis for the T_m dimensional vector space. Then we have

$$A_1 = f^{n\prime}_m f^n_m \sum_{j=1}^{N_m - 1} \xi_{mj} \xi_{mj}', \tag{B.49}$$

$$A_2 = \lambda^{n\prime}_m \lambda^n_m \sum_{j=1}^{T_m - 1} \chi_{mj} \chi_{mj}', \tag{B.50}$$

Step (1.3). Choose ξ_{oj} such that $\{\xi_{oj},\, j\in[N_o-1]\}$ and $\frac{\lambda_o^n}{\|\lambda_o^n\|}$ constitute an orthonormal basis for the N_o dimensional vector space. Then the following set of vectors are orthonormal,

$$\begin{pmatrix} \xi_{oj} \\ 0_{N_m\times 1} \end{pmatrix}_{j\in[N_o-1]}, \begin{pmatrix} 0_{N_o\times 1} \\ \xi_{mj'} \end{pmatrix}_{j'\in[N_m-1]}, \begin{pmatrix} \frac{\|\lambda_m^n\|}{\|\lambda_o^n\|}\lambda_o^n \\ -\frac{\|\lambda_o^n\|}{\|\lambda_m^n\|}\lambda_m^n \end{pmatrix}, \begin{pmatrix} \lambda_o^n \\ \lambda_m^n \end{pmatrix}. \tag{B.51}$$

The first three sets contain $N-1$ vectors in total, and we choose them as ξ_j for $j\in[N-1]$. Similarly, choose χ_{oj} such that $\{\chi_{oj},\, j\in[T_o-1]\}$ and $\frac{f_o^n}{\|f_o^n\|}$ constitute an orthonormal basis for the T_o dimensional vector space. Then the following set of vectors are orthonormal,

$$\begin{pmatrix} \chi_{oj} \\ 0_{T_m\times 1} \end{pmatrix}_{j\in[T_o-1]}, \begin{pmatrix} 0_{T_o\times 1} \\ \chi_{mj'} \end{pmatrix}_{j'\in[T_m-1]}, \begin{pmatrix} \frac{\|f_m^n\|}{\|f_o^n\|}f_o^n \\ -\frac{\|f_o^n\|}{\|f_m^n\|}f_m^n \end{pmatrix}, \begin{pmatrix} f_o^n \\ f_m^n \end{pmatrix}. \tag{B.52}$$

The first three sets contain $T-1$ vectors in total, and we choose them as χ_j for $j\in[T-1]$.

Let $\mathcal{A}\equiv -D_{f\lambda}^{-\frac{1}{2}}L_{\phi\phi' full}D_{f\lambda}^{-\frac{1}{2}}-(-D_{f\lambda}^{-\frac{1}{2}}L_{\phi\phi' mis}\,D_{f\lambda}^{-\frac{1}{2}}-A_3)$. From expressions (B.46)-(B.52) we can see that the $N+T$ eigenvectors of $\mathcal{A}$ are given by

$$\begin{array}{c|c|c|c} \begin{pmatrix} \xi_{oj} \\ 0_{N_m\times 1} \\ 0_{T_o\times 1} \\ 0_{T_m\times 1} \end{pmatrix}_{j\in[N_o-1]} & \begin{pmatrix} 0_{N_o\times 1} \\ \xi_{mj'} \\ 0_{T_o\times 1} \\ 0_{T_m\times 1} \end{pmatrix}_{j'\in[N_m-1]} & \begin{pmatrix} \frac{\|\lambda_m^n\|}{\|\lambda_o^n\|}\lambda_o^n \\ \frac{-\|\lambda_o^n\|}{\|\lambda_m^n\|}\lambda_m^n \\ 0_{T_o\times 1} \\ 0_{T_o\times 1} \end{pmatrix} & \begin{pmatrix} \lambda^n \\ f^n \end{pmatrix}/\sqrt{2} \\ \hline \begin{pmatrix} 0_{N_o\times 1} \\ 0_{N_m\times 1} \\ \chi_{oj} \\ 0_{T_m\times 1} \end{pmatrix}_{j\in[T_o-1]} & \begin{pmatrix} 0_{N_o\times 1} \\ 0_{N_m\times 1} \\ 0_{T_o\times 1} \\ \chi_{mj'} \end{pmatrix}_{j'\in[T_m-1]} & \begin{pmatrix} 0_{N_o\times 1} \\ 0_{N_m\times 1} \\ \frac{\|f_m^n\|}{\|f_o^n\|}f_o^n \\ \frac{-\|f_o^n\|}{\|f_m^n\|}f_m^n \end{pmatrix} & \begin{pmatrix} \lambda^n \\ -f^n \end{pmatrix}/\sqrt{2} \end{array}, \tag{B.53}$$

and the corresponding eigenvalues are

$$\begin{array}{c|c|c|c} \underset{(N_o-1\text{ times})}{1\text{ for all } j} & \underset{(N_m-1\text{ times})}{1-f_m^{n\prime}f_m^n\text{ for all } j',} & \mathbf{1} & \mathbf{2} \\ \hline \underset{(T_o-1\text{ times})}{1\text{ for all } j} & \underset{(T_m-1\text{ times})}{1-\lambda_m^{n\prime}\lambda_m^n\text{ for all } j'} & \mathbf{1} & 0 \end{array}. \tag{B.54}$$

Step (1.4). Now consider A_3. From equation (B.48), it is not difficult to verify that

$$A_{311}=\|f_m^n\|\,\|\lambda_m^n\|\left[-\frac{\|\lambda_o^n\|}{\|\lambda_m^n\|}\begin{pmatrix} \frac{\|\lambda_m^n\|}{\|\lambda_o^n\|}\lambda_o^n \\ -\frac{\|\lambda_o^n\|}{\|\lambda_m^n\|}\lambda_m^n \\ 0_{T_o\times 1} \\ 0_{T_m\times 1} \end{pmatrix}-\frac{\|f_o^n\|}{\|f_m^n\|}\begin{pmatrix} 0_{N_o\times 1} \\ 0_{N_m\times 1} \\ \frac{\|f_m^n\|}{\|f_o^n\|}f_o^n \\ -\frac{\|f_o^n\|}{\|f_m^n\|}f_m^n \end{pmatrix}+\sqrt{2}\begin{pmatrix} \frac{\lambda^n}{\sqrt{2}} \\ \frac{f^n}{\sqrt{2}} \end{pmatrix}\right].$$

The eigenvalues of $\mathcal{A}$ corresponding to the three vectors in the square brackets are $1,1$ and 2, respectively. After subtracting $A_{311}A_{311}'$, only these three eigenvalues are affected, and they become the eigenvalues of

$$\left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} - \vartheta\vartheta'\right], \tag{B.55}$$

where $\vartheta \equiv \|f_m^n\| \|\lambda_m^n\| \begin{pmatrix} -\frac{\|\lambda_o^n\|}{\|\lambda_m^n\|} \\ -\frac{\|f_o^n\|}{\|f_m^n\|} \\ \sqrt{2} \end{pmatrix} = \begin{pmatrix} -\|f_m^n\| \|\lambda_o^n\| \\ -\|f_o^n\| \|\lambda_m^n\| \\ \sqrt{2}\|f_m^n\| \|\lambda_m^n\| \end{pmatrix}$. Obviously, the 3×1 vector that is orthogonal to both $(0,0,1)'$ and ϑ is an eigenvector of (B.55), and the corresponding eigenvalue is 1. The remaining two eigenvectors of (B.55) should lie in the space orthogonal to this eigenvector, thus could be written as linear combinations of $(0,0,1)'$ and ϑ, or (for simplicity) linear combinations of $(0,0,1)'$ and $(-\|f_m^n\| \|\lambda_o^n\|, -\|f_o^n\| \|\lambda_m^n\|, 0)'$. Suppose vv is an eigenvector of (B.55) and $vv = x(-\|f_m^n\| \|\lambda_o^n\|, -\|f_o^n\| \|\lambda_m^n\|, 0)' + y(0,0,1)'$. Pre-multiplying vv by (B.55), we have

$$\begin{aligned} x - [x(\|f_m^n\|^2 \|\lambda_o^n\|^2 + \|f_o^n\|^2 \|\lambda_m^n\|^2) + y\sqrt{2}\|f_m^n\| \|\lambda_m^n\|] &= cx, \\ 2y - \sqrt{2}\|f_m^n\| \|\lambda_m^n\| [x(\|f_m^n\|^2 \|\lambda_o^n\|^2 + \|f_o^n\|^2 \|\lambda_m^n\|^2) + y\sqrt{2}\|f_m^n\| \|\lambda_m^n\|] &= cy, \end{aligned}$$

where c represents the eigenvalue of vv. The solution of these two equations are

$$\begin{aligned} c_1 &\equiv \frac{\|f_o^n\|^2 + \|\lambda_o^n\|^2 + 1}{2} + \sqrt{\left(\frac{\|f_o^n\|^2 + \|\lambda_o^n\|^2 + 1}{2}\right)^2 - 2\|f_o^n\|^2 \|\lambda_o^n\|^2}, \\ c_2 &\equiv \frac{\|f_o^n\|^2 + \|\lambda_o^n\|^2 + 1}{2} - \sqrt{\left(\frac{\|f_o^n\|^2 + \|\lambda_o^n\|^2 + 1}{2}\right)^2 - 2\|f_o^n\|^2 \|\lambda_o^n\|^2}. \end{aligned}$$

Thus the three eigenvalues of (B.55) are $(1, c_1, c_2)$. In sum, the eigenvalues of $-D_{f\lambda}^{-\frac{1}{2}} L_{\phi\phi'} D_{f\lambda}^{-\frac{1}{2}}$ are:

$$\begin{array}{c|c|c|c} \underset{(N_o-1 \text{ times})}{1 \text{ for all } j} & \underset{(N_m-1 \text{ times})}{1 - f_m^{n\prime} f_m^n \text{ for all } j'}, & \mathbf{c_2} & \mathbf{c_1} \\ \hline \underset{(T_o-1 \text{ times})}{1 \text{ for all } j} & \underset{(T_m-1 \text{ times})}{1 - \lambda_m^{n\prime}\lambda_m^n \text{ for all } j'} & \mathbf{1} & 0 \end{array}. \tag{B.56}$$

Due to subtracting $A_{311}A_{311}'$, the eigenvectors of the three bold blocks in expression (B.56) are orthogonal rotations of the eigenvectors of the three bold blocks in expression (B.54). All of the nonzero eigenvalues are positive and bounded away from zero w.p.a.1. ■

Lemma B.3 *Suppose that the conditions in Theorem 4.1 hold. Then as* $(N,T) \to \infty$, *both* $\frac{1}{\sqrt{T}} R_\lambda$ *and* $\frac{1}{\sqrt{N}} R_f$ *are* $O_p(\left\|\hat{\lambda} - \lambda^0\right\| \left\|\hat{f} - f^0\right\| + \sqrt{\frac{N}{T}}\left\|\hat{f} - f^0\right\|^2 + \sqrt{\frac{T}{N}}\left\|\hat{\lambda} - \lambda^0\right\|^2)$.

Proof. First note that (1) $\partial_{\lambda_{iq}} L_{\lambda\lambda'} = 0$, (2) the t-th block of $\partial_{\lambda_{iq}} L_{ff'}$ is $-d_{it}(\lambda_i 1_q^{r\prime} + 1_q^r \lambda_i')$ where 1_q^r

denotes the $r\times 1$ vector with the q-th element being 1 and all the other elements being zeros, (3) the (i,t)th block of $\partial_{\lambda_{iq}}L_{\lambda f'}$ is $-d_{it}f_t 1_q^{r\prime}$, and the (j,s)-th blocks are zeros if $j\neq i$, and (4) the (i,t)th block of $\partial_{\lambda_{iq}}J_{\lambda f'}$ is $-d_{it}f_{tq}I_r$, and the (j,s)th blocks are zeros if $j\neq i$. By (B.6)-(B.8), we have

$$\begin{aligned}
\partial_{\phi\phi'\lambda_{iq}}[(\frac{\sum_{i=1}^N\lambda_{iq}^2}{N}-\frac{\sum_{t=1}^T f_{tq}^2}{T})^2] &= 8D_{NT}^{-1}(1_{iq}w_{qq}'+w_{qq}1_{iq}')D_{NT}^{-1}+\frac{8}{N}\lambda_{iq}D_{NT}^{-1}(I_{N+T}\otimes\iota_q),\\
\partial_{\phi\phi'\lambda_{iq}}[\sum_{p=1}^r\sum_{q>p}(\sum_{t=1}^T f_{tp}f_{tq})^2] &= 0,\\
\partial_{\phi\phi'\lambda_{iq}}[\sum_{p=1}^r\sum_{q>p}(\sum_{i=1}^N\lambda_{ip}\lambda_{iq})^2] &= 2(\sum_{p\neq q}\lambda_{ip}D_1+\sum_{p\neq q}1_{ip}u_{pq}'+\sum_{p\neq q}u_{pq}1_{ip}'),
\end{aligned}$$

where 1_{iq} is an $Nr+Tr$ dimensional vector with the q-th element in the i-th block being one and all the other elements being zero. Let $R_{\lambda_{iq}}$ and $R_{f_{tq}}$ denote the q-th element of R_{λ_i} and R_{f_t}, respectively. It follows that

$$\begin{aligned}
R_{\lambda_{iq}} &= (\hat\phi-\phi^0)'\partial_{\lambda_{iq}}\partial_{\phi\phi'}Q(s)(\hat\phi-\phi^0)\\
&= -\sum_{t=1}^T d_{it}(\hat\lambda_i-\lambda_i^0)'(f_t(s)1_q^{r\prime}+f_{tq}(s)I_r)(\hat f_t-f_t^0)\\
&\quad -\sum_{t=1}^T d_{it}(\hat f_t-f_t^0)'\lambda_i(s)1_q^{r\prime}(\hat f_t-f_t^0)\\
&\quad -cT(\hat\lambda_{iq}-\lambda_{iq}^0)[\frac{1}{N}\sum_{j=1}^N\lambda_{jq}(s)(\hat\lambda_{jq}-\lambda_{jq}^0)-\frac{1}{T}\sum_{t=1}^T f_{tq}(s)(\hat f_{tq}-f_{tq}^0)]\\
&\quad -\frac{c}{2}T\lambda_{iq}(s)[\frac{1}{N}\sum_{j=1}^N(\hat\lambda_{jq}-\lambda_{jq}^0)^2+\frac{1}{T}\sum_{t=1}^T(\hat f_{tq}-f_{tq}^0)^2]\\
&\quad -\frac{cT}{N}\sum_{p\neq q}\lambda_{ip}(s)\sum_{j=1}^N(\hat\lambda_{jp}-\lambda_{jp}^0)(\hat\lambda_{jq}-\lambda_{jq}^0)\\
&\quad -\frac{cT}{N}\sum_{p\neq q}(\hat\lambda_{ip}-\lambda_{ip}^0)[\sum_{j=1}^N\lambda_{jq}(s)(\hat\lambda_{jp}-\lambda_{jp}^0)+\sum_{j=1}^N\lambda_{jp}(s)(\hat\lambda_{jq}-\lambda_{jq}^0)]\\
&\equiv L1i+L2i+P1i+P2i+P3i+P4i, \qquad\qquad (B.57)
\end{aligned}$$

where $f(s)=f^0+s(\hat f-f^0)$ and $\lambda(s)=\lambda^0+s(\hat\lambda-\lambda^0)$. It's easy to see

$$\begin{aligned}
|L1i| &\le M\left\|\hat\lambda_i-\lambda_i^0\right\|\|f(s)\|\left\|\hat f-f^0\right\|, &(B.58)\\
|L2i| &\le M\|\lambda_i(s)\|\left\|\hat f-f^0\right\|^2, &(B.59)\\
|P1i| &\le MT\left\|\hat\lambda_i-\lambda_i^0\right\|(\frac{1}{N}\|\lambda(s)\|\left\|\hat\lambda-\lambda^0\right\|+\frac{1}{T}\|f(s)\|\left\|\hat f-f^0\right\|), &(B.60)\\
|P2i| &\le MT\|\lambda_i(s)\|(\frac{1}{N}\left\|\hat\lambda-\lambda^0\right\|^2+\frac{1}{T}\left\|\hat f-f^0\right\|^2), &(B.61)\\
|P3i| &\le M\frac{T}{N}\|\lambda_i(s)\|\left\|\hat\lambda-\lambda^0\right\|^2, &(B.62)\\
|P4i| &\le M\frac{T}{N}\left\|\hat\lambda_i-\lambda_i^0\right\|\|\lambda(s)\|\left\|\hat\lambda-\lambda^0\right\|. &(B.63)
\end{aligned}$$

Since $\sup_{0\le s\le 1}\|f_t(s)\| \le \|f_t^0\| + \left\|\hat{f}_t - f_t^0\right\|$ and $\sup_{0\le s\le 1}\|\lambda_i(s)\| \le \|\lambda_i^0\| + \left\|\hat{\lambda}_i - \lambda_i^0\right\|$, Assumption 1 and Theorem 4.1 implies that $\sup_{0\le s\le 1}\|f(s)\| = O_p(\sqrt{T})$ and $\sup_{0\le s\le 1}\|\lambda(s)\| = O_p(\sqrt{N})$, it follows that

$$\|R_{\lambda_i}\| = \left\|\hat{\lambda}_i - \lambda_i^0\right\| O_p(\sqrt{T}\left\|\hat{f} - f^0\right\| + \frac{T}{\sqrt{N}}\left\|\hat{\lambda} - \lambda^0\right\|) + O_p(\left\|\hat{f} - f^0\right\|^2 + \frac{T}{N}\left\|\hat{\lambda} - \lambda^0\right\|^2), \quad (B.64)$$

$$\|R_{\lambda}\| = O_p(\sqrt{T}\left\|\hat{\lambda} - \lambda^0\right\|\left\|\hat{f} - f^0\right\| + \sqrt{N}\left\|\hat{f} - f^0\right\|^2 + \frac{T}{\sqrt{N}}\left\|\hat{\lambda} - \lambda^0\right\|^2). \quad (B.65)$$

By symmetric arguments, we also have

$$\|R_{f_t}\| = \left\|\hat{f}_t - f_t^0\right\| O_p(\sqrt{N}\left\|\hat{\lambda} - \lambda^0\right\| + \frac{N}{\sqrt{T}}\left\|\hat{f} - f^0\right\|) + O_p(\left\|\hat{\lambda} - \lambda^0\right\|^2 + \frac{N}{T}\left\|\hat{f} - f^0\right\|^2), \quad (B.66)$$

$$\|R_f\| = O_p(\sqrt{N}\left\|\hat{\lambda} - \lambda^0\right\|\left\|\hat{f} - f^0\right\| + \sqrt{T}\left\|\hat{\lambda} - \lambda^0\right\|^2 + \frac{N}{\sqrt{T}}\left\|\hat{f} - f^0\right\|^2). \quad (B.67)$$

This completes the proof of the lemma. ■

Proof of Theorem 4.2.

Noting that $\partial_\phi P(\lambda^0, f^0) = 0$, we can write $S_\phi \equiv \partial_\phi Q\left(\phi^0\right) = (S_{\lambda_1}', ..., S_{\lambda_N}', S_{f_1}', ..., S_{f_T}')'$, where $S_{\lambda_i} = \sum_{t=1}^T d_{it}v_{it}f_t^0$ and $S_{f_t} = \sum_{i=1}^N d_{it}v_{it}\lambda_i^0$. Then

$$\left\|D_{TN}^{-\frac{1}{2}}S_\phi\right\|^2 = \sum_{i=1}^N \left\|\frac{\sum_{t=1}^T d_{it}v_{it}f_t^0}{\sqrt{T}}\right\|^2 + \sum_{t=1}^T \left\|\frac{\sum_{i=1}^N d_{it}v_{it}\lambda_i^0}{\sqrt{N}}\right\|^2 = O_p(N+T), \quad (B.68)$$

where the second equality holds by Assumption 5 and Markov and Jensen inequalities. It follows that

$$\begin{aligned}
\begin{pmatrix} \frac{1}{\sqrt{N}}(\hat{\lambda} - \lambda^0) \\ \frac{1}{\sqrt{T}}(\hat{f} - f^0) \end{pmatrix} &= D_{NT}^{-\frac{1}{2}}(\hat{\phi} - \phi^0) = -D_{NT}^{-\frac{1}{2}}H_{\phi\phi'}^{-1}S_\phi - D_{NT}^{-\frac{1}{2}}H_{\phi\phi'}^{-1}R_\phi \\
&= (-D_{TN}^{-\frac{1}{2}}H_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}\frac{D_{TN}^{-\frac{1}{2}}S_\phi}{\sqrt{NT}} + (-D_{TN}^{-\frac{1}{2}}H_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}\frac{D_{TN}^{-\frac{1}{2}}R_\phi}{\sqrt{NT}} \\
&= O_p(\frac{1}{c_{NT}}) + O_p(\frac{1}{T}\left\|\hat{f} - f^0\right\|^2 + \frac{1}{N}\left\|\hat{\lambda} - \lambda^0\right\|^2), \quad (B.69)
\end{aligned}$$

where the last equality holds by (B.68) and Lemmas B.1, B.2 and B.3. By Theorem 4.1, $||\hat{\lambda} - \lambda^0|| = O_p(\sqrt{\frac{N}{c_{NT}}})$ and $||\hat{f} - f^0|| = O_p(\sqrt{\frac{T}{c_{NT}}})$. Plugging this back into equation (B.69) yields that $\frac{1}{\sqrt{N}}||\hat{\lambda} - \lambda^0|| = O_p(\frac{1}{c_{NT}})$ and $\frac{1}{\sqrt{T}}||\hat{f} - f^0|| = O_p(\frac{1}{c_{NT}})$. ■

C Proof of Theorem 4.3

To prove Theorem 4.3, we need Lemmas C.1–C.3 below.

Lemma C.1 *Suppose that the conditions in Theorem 4.2 hold. Then as* $(N,T)\to\infty$,
(i) $\|R_{\lambda_i}\| = \left\|\hat{\lambda}_i - \lambda_i^0\right\| O_p(\frac{T}{c_{NT}}) + O_p(\frac{T}{c_{NT}^2})$ *and* $\|R_\lambda\| = O_p(\frac{T\sqrt{N}}{c_{NT}^2})$;
(ii) $\|R_{f_t}\| = \left\|\hat{f}_t - f_t^0\right\| O_p(\frac{N}{c_{NT}}) + O_p(\frac{N}{c_{NT}^2})$ *and* $\|R_f\| = O_p(\frac{N\sqrt{T}}{c_{NT}^2})$.

Proof. By Theorem 4.2, we have $||\hat{\lambda}-\lambda^0|| = O_p(\frac{\sqrt{N}}{c_{NT}})$ and $||\hat{f}-f^0|| = O_p(\frac{\sqrt{T}}{c_{NT}})$. Plugging these back into (B.64)-(B.67), we prove the above lemma. ∎

Lemma C.2 *Suppose that Assumptions 1, 2 and 4(i) hold. Then as* $(N,T)\to\infty$,
(i) $\left\|[U_\lambda^0]_i\right\| = O_p(1)$ *and* $\left\|U_\lambda^0\right\| = O_p(\sqrt{N})$;
(ii) $\left\|[L_{\lambda\lambda'}^{-1}]_i\right\| = O_p(\frac{1}{T})$ *and* $\left\|L_{\lambda\lambda'}^{-1}\right\| = O_p(\frac{1}{T})$;
(iii) $\left\|H_{\lambda\lambda'}^{-1}\right\| = O_p(\frac{1}{T})$ *and* $\left\|H_{ff'}^{-1}\right\| = O_p(\frac{1}{N})$;
(iv) $\left\|[H_{\lambda f'}]_i\right\| = O_p(\sqrt{T})$ *and* $\left\|H_{\lambda f'}\right\| = O_p(\sqrt{NT})$.

Proof. (i) The results are obvious by noting that $\left\|\lambda_i^0\right\| \leq M$ under Assumption 1(ii).

(ii) For the random missing case, since $\mathbb{E}_\phi(d_{it}) > c > 0$ for all i and t by Assumption 2(i), $\min_i \sigma_{\min}(\frac{1}{T}\sum_{t=1}^T \mathbb{E}_\phi(d_{it}) f_t^0 f_t^{0\prime})$ is bounded away from zero in probability by Assumption 1(i). (B.17) implies $\left\|\frac{1}{T}(L_{\lambda\lambda'} - \bar{L}_{\lambda\lambda'})\right\| = O_p(\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}}) = o_p(1)$ when $\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}} \to 0$. It follows that

$$
\begin{aligned}
\left\|(\frac{1}{T}L_{\lambda\lambda'})^{-1}\right\| &= \left[\min_i \sigma_{\min}(\frac{1}{T}\sum\nolimits_{t=1}^T d_{it} f_t^0 f_t^{0\prime})\right]^{-1} \\
&\leq 1/\left[\min_i \sigma_{\min}(\frac{1}{T}\sum\nolimits_{t=1}^T \mathbb{E}_\phi(d_{it}) f_t^0 f_t^{0\prime}) - \left\|\frac{1}{T}(L_{\lambda\lambda'} - \bar{L}_{\lambda\lambda'})\right\|\right] = O_p(1).
\end{aligned}
$$

For the block missing/staggered missing/mixed frequency case, the result follows from Assumptions 1 and 2(ii).

(iii) By equations (B.9)–(B.10), $H_{\lambda\lambda'} = L_{\lambda\lambda'} - \frac{cT}{N} U_\lambda^0 U_\lambda^{0\prime}$. Then by Woodbury matrix identity (see, e.g., Fact 6.4.31 in Bernstein (2005) or p.309 in Seber (2008)),

$$
H_{\lambda\lambda'}^{-1} = L_{\lambda\lambda'}^{-1} - L_{\lambda\lambda'}^{-1} U_\lambda^0 (-\frac{N}{cT} I_{r^2} + U_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} U_\lambda^0)^{-1} U_\lambda^{0\prime} L_{\lambda\lambda'}^{-1}. \tag{C.1}
$$

Since $U_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} U_\lambda^0$ is negative definite, we have

$$
\left\|[-\frac{N}{cT} I_{r^2} + U_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} U_\lambda^0]^{-1}\right\| \leq \frac{cT}{N}. \tag{C.2}
$$

Then by (i)-(ii), we have

$$
\begin{aligned}
\left\|H_{\lambda\lambda'}^{-1}\right\| &\leq \left\|L_{\lambda\lambda'}^{-1}\right\| + \left\|L_{\lambda\lambda'}^{-1}\right\|^2 \left\|U_\lambda^0\right\|^2 \left\|(-\frac{N}{cT} I_{r^2} + U_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} U_\lambda^0)^{-1}\right\| \\
&= O_p(\frac{1}{T}) + O_p(\frac{1}{T^2}) O_p(N) O_p(\frac{T}{N}) = O_p(\frac{1}{T}).
\end{aligned}
$$

The second part can be proved analogously.

(iv) By equations (B.9)–(B.10), $H_{\lambda f'} = L_{\lambda f'} + J_{\lambda f'} - cU_\lambda^0 U_f^{0\prime}$. Noting that the (i,t)th block of $L_{\lambda f'}$ is $-d_{it}f_t^0\lambda_i^{0\prime}$, $||[L_{\lambda f'}]_i|| \leq ||\lambda_i^0||||f^0|| = O_p(\sqrt{T})$ and $||L_{\lambda f'}|| \leq ||\lambda^0||||f^0|| = O_p(\sqrt{NT})$ by Assumption 1. Similarly to $L_{\lambda f'}$, we also have $||[U_\lambda^0 U_f^{0\prime}]_i|| = O_p(\sqrt{T})$ and $||U_\lambda^0 U_f^{0\prime}|| = O_p(\sqrt{NT})$. Noting that the (i,t)th block of $J_{\lambda f'}$ is $d_{it}v_{it}I_r$, we have $||[J_{\lambda f'}]_i|| \leq (r\sum_{t=1}^T d_{it}^2 v_{it}^2)^{\frac{1}{2}} = O_p(\sqrt{T})$ and $\left\|J_{\lambda f'}\right\| \leq (r\sum_{i=1}^N\sum_{t=1}^T d_{it}^2 v_{it}^2)^{\frac{1}{2}} = O_p(\sqrt{NT})$ by Assumption 4(i). ■

Lemma C.3 *Suppose that Assumptions 1, 2 and 6 hold. Then as* $(N,T)\to\infty$*,*

(i) $\left\|U_\lambda^{0\prime}H_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\sqrt{\frac{N}{T}} + \frac{N}{T})$ *and* $\left\|U_f^{0\prime}H_{ff'}^{-1}S_f\right\| = O_p(\sqrt{\frac{T}{N}} + \frac{T}{N})$*;*

(ii) $\left\|L_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\sqrt{N} + \frac{N}{\sqrt{T}})$ *and* $\left\|L_{\lambda f'}H_{ff'}^{-1}S_f\right\| = O_p(\sqrt{T} + \frac{T}{\sqrt{N}})$*;*

(iii) $\left\|J_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\sqrt{N} + \frac{N}{\sqrt{T}})$ *and* $\left\|J_{\lambda f'}H_{ff'}^{-1}S_f\right\| = O_p(\sqrt{T} + \frac{T}{\sqrt{N}})$*;*

(iv) $\left\|H_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\sqrt{N} + \frac{N}{\sqrt{T}})$ *and* $\left\|H_{\lambda f'}H_{ff'}^{-1}S_f\right\| = O_p(\sqrt{T} + \frac{T}{\sqrt{N}})$*;*

(v) $\left\|[H_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda]_t\right\| = O_p(\sqrt{\frac{N}{T}} + \frac{N}{T})$ *and* $\left\|[H_{\lambda f'}H_{ff'}^{-1}S_f]_i\right\| = O_p(\sqrt{\frac{T}{N}} + \frac{T}{N})$*.*

Proof. Recall that $S_\lambda = (S_{\lambda_1}', ..., S_{\lambda_N}')'$ and $S_f = (S_{f_1}', ..., S_{f_T}')'$, where $S_{\lambda_i} = \sum_{t=1}^T d_{it}v_{it}f_t^0$ and $S_{f_t} = \sum_{i=1}^N d_{it}v_{it}\lambda_i^0$. In the following, we shall only prove the first half of parts (i)–(v) as the second half follows from symmetry.

(i) Note that $U_\lambda^{0\prime}H_{\lambda\lambda'}^{-1}S_\lambda = U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}S_\lambda - U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0(-\frac{N}{cT}I_{r^2} + U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0)^{-1}U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}S_\lambda$ by (C.1). We first show

$$\left\|U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\sqrt{\frac{N}{T}} + \frac{N}{T}). \tag{C.3}$$

$U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}S_\lambda$ is an r^2-dimensional vector. From the definition of U_λ^0, we need to show that for any p and q, $\sum_{i=1}^N \lambda_{ip}^0 1_q^{(r)\prime}(\sum_{t=1}^T d_{it}f_t^0 f_t^{0\prime})^{-1}(\sum_{t=1}^T d_{it}v_{it}f_t^0)$ is $O_p(\sqrt{\frac{N}{T}} + \frac{N}{T})$. This follows because by Assumptions 5, 6(i) and 6(iii),

$$\begin{aligned}
&\left\|\sum_{i=1}^N\sum_{t=1}^T(\sum_{t=1}^T d_{it}f_t^0 f_t^{0\prime})^{-1}f_t^0\lambda_i^{0\prime}d_{it}v_{it}\right\|_F \\
\leq\ &\left\|\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T A_{iF}^{-1}f_t^0\lambda_i^{0\prime}d_{it}v_{it}\right\|_F + \left\|\sum_{i=1}^N(\bar{A}_{iF}^{-1} - A_{iF}^{-1})\frac{1}{T}\sum_{t=1}^T f_t^0\lambda_i^{0\prime}d_{it}v_{it}\right\|_F \\
\leq\ &O_p(\sqrt{\frac{N}{T}}) + \left\{\sum_{i=1}^N\left\|\bar{A}_{iF}^{-1} - A_{iF}^{-1}\right\|^2\right\}^{1/2}\left\{\sum_{i=1}^N\left\|\frac{1}{T}\sum_{t=1}^T f_t^0 d_{it}v_{it}\lambda_i^{0\prime}\right\|_F^2\right\}^{1/2} \\
=\ &O_p(\sqrt{\frac{N}{T}}) + O_p(\sqrt{\frac{N}{T}})O_p(\sqrt{\frac{N}{T}}) = O_p(\sqrt{\frac{N}{T}} + \frac{N}{T}),
\end{aligned} \tag{C.4}$$

where $\sum_{i=1}^N\left\|\bar{A}_{iF}^{-1} - A_{iF}^{-1}\right\|^2 \leq \sum_{i=1}^N\left\|A_{iF} - \bar{A}_{iF}\right\|^2 \sup_i\left\|A_{iF}^{-1}\right\|^2 \sup_i\left\|\bar{A}_{iF}^{-1}\right\|^2 = O_p(\frac{N}{T})$. Then

$$\left\|U_\lambda^{0\prime}H_{\lambda\lambda'}^{-1}S_\lambda\right\| \leq \left\|U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}S_\lambda\right\| + \left\|U_\lambda^0\right\|^2\left\|L_{\lambda\lambda'}^{-1}\right\|\left\|(\frac{N}{cT}I_{r^2} - U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0)^{-1}\right\|\left\|U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}S_\lambda\right\|$$

$$= O_p(\sqrt{\frac{N}{T}}) + O_p(N)O_p(\frac{1}{T})O_p(\frac{T}{N})O_p(\sqrt{\frac{N}{T}} + \frac{N}{T}) = O_p(\sqrt{\frac{N}{T}} + \frac{N}{T}),$$

where the first equality holds by (C.3), (C.2), and Lemma C.2(i)–(ii).

(ii) Note that $L_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda = L_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda - L_{f\lambda'}L_{\lambda\lambda'}^{-1}U_\lambda^0(-\frac{N}{cT}I_{r^2} + U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0)^{-1}U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}S_\lambda$ by (C.1). We first show that $\left\|L_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\sqrt{N} + \frac{N}{\sqrt{T}})$. Note that

$$\begin{aligned}
[L_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda]_s &= \sum\nolimits_{i=1}^N d_{is}\lambda_i^0 f_s^{0\prime}(\sum\nolimits_{t=1}^T d_{it}f_t^0 f_t^{0\prime})^{-1}(\sum\nolimits_{t=1}^T d_{it}v_{it}f_t^0) \\
&= [f_s^{0\prime}\sum\nolimits_{i=1}^N\sum\nolimits_{t=1}^T(\sum\nolimits_{t=1}^T d_{it}f_t^0 f_t^{0\prime})^{-1}d_{it}v_{it}f_t^0\lambda_i^{0\prime}d_{is}]' = \frac{1}{T}[f_s^{0\prime}\sum\nolimits_{i=1}^N\sum\nolimits_{t=1}^T A_{iF}^{-1}\xi_{it}d_{is}]',
\end{aligned}$$

where $\xi_{it} = d_{it}v_{it}f_t^0\lambda_i^{0\prime}$. We make the following decomposition.

$$\frac{1}{T}\sum\nolimits_{i=1}^N\sum\nolimits_{t=1}^T \bar{A}_{iF}^{-1}\xi_{it}d_{is} = \frac{1}{T}\sum\nolimits_{i=1}^N\sum\nolimits_{t=1}^T\{A_{iF}^{-1}\xi_{it}d_{is} + \left(\bar{A}_{iF}^{-1} - A_{iF}^{-1}\right)\xi_{it}d_{is}\} \equiv I_{1s} + I_{2s},$$

Under Assumption 1 that $\left\|L_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda\right\|^2 \leq M^2\sum_{l=1}^2\sum_{s=1}^T\|I_{ls}\|^2 \equiv \sum_{l=1}^2 I_l$. By Assumption 6(i) and (iii) and the CS inequality, we have $\mathbb{E}\|I_1\| = \sum_{s=1}^T\mathbb{E}\left\|\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T A_{iF}^{-1}\xi_{it}d_{is}\right\|^2 = O(N)$, and

$$\begin{aligned}
I_2 &= \sum_{s=1}^T\left\|\sum\nolimits_{i=1}^N\left(\bar{A}_{iF}^{-1} - A_{iF}^{-1}\right)\frac{1}{T}\sum\nolimits_{t=1}^T\xi_{it}d_{is}\right\|^2 \\
&\leq \sum\nolimits_{i=1}^N\left\|\bar{A}_{iF}^{-1} - A_{iF}^{-1}\right\|^2\sum_{s=1}^T\sum\nolimits_{i=1}^N\left\|\frac{1}{T}\sum\nolimits_{t=1}^T\xi_{it}d_{is}\right\|^2 = O(\frac{N^2}{T}). \qquad \text{(C.5)}
\end{aligned}$$

Then $\left\|L_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\sqrt{N} + \frac{N}{\sqrt{T}})$. It follows that

$$\begin{aligned}
\left\|L_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda\right\| &\leq \left\|L_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda\right\| + \left\|L_{f\lambda'}\right\|\left\|L_{\lambda\lambda'}^{-1}\right\|\left\|U_\lambda^0\right\|\left\|(\frac{N}{cT}I_{r^2} - U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0)^{-1}\right\|\left\|U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}S_\lambda\right\| \\
&= O_p(\sqrt{N} + \frac{N}{\sqrt{T}}) + O_p(\sqrt{NT})O_p(\frac{1}{T})O_p(\sqrt{N})O_p(\frac{T}{N})O_p(\sqrt{\frac{N}{T}} + \frac{N}{\sqrt{T}}) \\
&= O_p(\sqrt{N} + \frac{N}{\sqrt{T}}),
\end{aligned}$$

where the first equality follows from (C.2), (C.3), Lemma C.2(i)-(ii), and the fact that $\left\|L_{f\lambda'}\right\| = O_p(\sqrt{NT})$. In addition, it is easy to see that $[L_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda]_s = O_p(\sqrt{\frac{N}{T}} + \frac{N}{T})$.

(iii) Note that $J_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda = J_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda - J_{f\lambda'}L_{\lambda\lambda'}^{-1}U_\lambda^0(-\frac{N}{cT}I_{r^2} + U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0)^{-1}U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}S_\lambda$ by (C.1). We first show that $\left\|J_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\sqrt{N} + \frac{N}{\sqrt{T}})$. Note that

$$\begin{aligned}
-[J_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda]_s &= \sum\nolimits_{i=1}^N\sum\nolimits_{t=1}^T(\sum\nolimits_{t=1}^T d_{it}f_t^0 f_t^{0\prime})^{-1}d_{is}v_{is}d_{it}v_{it}f_t^0 \\
&= \frac{1}{T}\sum\nolimits_{i=1}^N\sum\nolimits_{t=1}^T\bar{A}_{iF}^{-1}d_{is}v_{is}d_{it}v_{it}f_t^0
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{T}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\{A_{iF}^{-1}[\xi_{1its}-\mathbb{E}(\xi_{1its})]+A_{iF}^{-1}\mathbb{E}(\xi_{1its}) \\
&\quad +(\bar{A}_{iF}^{-1}-A_{iF}^{-1})[\xi_{1its}-\mathbb{E}(\xi_{1its})]+(\bar{A}_{iF}^{-1}-A_{iF}^{-1})\mathbb{E}(\xi_{1its})\} \\
&\equiv II_{1s}+II_{2s}+II_{3s}+II_{4s}.
\end{aligned}
$$

where recall that $\xi_{1its}=d_{is}v_{is}d_{it}v_{it}f_t^0$. Then $\left\|J_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda\right\|^2 \le 4\sum_{l=1}^4\sum_{s=1}^T\|II_{ls}\|^2 \equiv 4\sum_{l=1}^4 II_l$. By Assumption 6(i)–(ii) and the CS inequality, we have

$$
\begin{aligned}
\mathbb{E}\|II_1\| &= \sum_{s=1}^{T}\mathbb{E}\left\|\frac{1}{T}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}A_{iF}^{-1}[\xi_{1its}-\mathbb{E}(\xi_{1its})]\right\|^2 = O(N), \\
II_2 &= \sum_{s=1}^{T}\left\|\frac{1}{T}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}A_{iF}^{-1}\mathbb{E}(\xi_{1its})]\right\|^2 = O(\frac{N^2}{T}), \\
II_3 &\le \sum\nolimits_{i=1}^{N}\|\bar{A}_{iF}^{-1}-A_{iF}^{-1}\|^2\sum\nolimits_{s=1}^{T}\sum\nolimits_{i=1}^{N}\left\|\frac{1}{T}\sum\nolimits_{t=1}^{T}[\xi_{1its}-\mathbb{E}(\xi_{1its})]\right\|^2 = O(\frac{N^2}{T}) \\
II_4 &\le \sum\nolimits_{i=1}^{N}\|\bar{A}_{iF}^{-1}-A_{iF}^{-1}\|^2\sum\nolimits_{s=1}^{T}\sum\nolimits_{i=1}^{N}\left\|\frac{1}{T}\sum\nolimits_{t=1}^{T}\mathbb{E}(\xi_{1its})\right\|^2 = O(\frac{N^2}{T^2}).
\end{aligned}
$$

Then $\left\|J_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\sqrt{N}+\frac{N}{\sqrt{T}})$. Then

$$
\begin{aligned}
J_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda &\le \left\|J_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda\right\|+\left\|J_{f\lambda'}\right\|\left\|L_{\lambda\lambda'}^{-1}\right\|\left\|U_\lambda^0\right\|\left\|(\frac{N}{cT}I_{r^2}-U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0)^{-1}\right\|\left\|U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}S_\lambda\right\| \\
&= O_p(\sqrt{N})+O_p(\sqrt{NT})O_p(\frac{1}{T})O_p(\sqrt{N})O_p(\frac{T}{N})O_p(\sqrt{\frac{N}{T}}+\frac{N}{T}) = O_p(\sqrt{N}+\frac{N}{T}),
\end{aligned}
$$

by (C.2), (C.3), Lemma C.2(i)–(ii), and the fact that $\left\|J_{f\lambda'}\right\| = O_p(\sqrt{NT})$. In addition, it is easy to see that $[J_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda]_s = O_p(\sqrt{\frac{N}{T}}+\frac{N}{T})$.

(iv) Noting $H_{f\lambda'} = L_{f\lambda'}+J_{f\lambda'}-cU_f^0U_\lambda^{0\prime}$, we have by the results in parts (i)-(iii) that

$$
\begin{aligned}
\left\|H_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda\right\| &\le \left\|L_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda\right\|+\left\|J_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda\right\|+c\left\|U_f^0\right\|\left\|U_\lambda^{0\prime}H_{\lambda\lambda'}^{-1}S_\lambda\right\| \\
&= O_p(\sqrt{N}+\frac{N}{\sqrt{T}})+O_p(\sqrt{N}+\frac{N}{\sqrt{T}})+O_p(\sqrt{T})O_p(\sqrt{\frac{N}{T}}+\frac{N}{T}) \\
&= O_p(\sqrt{N}+\frac{N}{\sqrt{T}}).
\end{aligned}
$$

(v) This part is implicitly proved in parts (i)–(iv). ■

Proof for Theorem 4.3. We focus on the analysis of $\hat{\lambda}_i-\lambda_i^0$. Noting that $\hat{\phi}-\phi^0 = -H_{\phi\phi'}^{-1}S_\phi-H_{\phi\phi'}^{-1}R_\phi$ by (3.1), we have

$$
\hat{\lambda}_i-\lambda_i^0 = -[H_{\phi\phi'}^{-1}S_\phi]_i-[H_{\phi\phi'}^{-1}R_\phi]_i, \tag{C.6}
$$

where recall that $[A]_i$ denotes the ith block of vector A (of size $r\times 1$).

First, we study $[H_{\phi\phi'}^{-1}R_\phi]_i$. The upper-left block of $H_{\phi\phi'}^{-1}$ is $[H_{\lambda\lambda'} - H_{\lambda f'}H_{ff'}^{-1}H_{f\lambda'}]^{-1}$, the upper-right block is $-[H_{\lambda\lambda'} - H_{\lambda f'}H_{ff'}^{-1}H_{f\lambda'}]^{-1}H_{\lambda f'}H_{ff'}^{-1}$, and

$$[H_{\lambda\lambda'} - H_{\lambda f'}H_{ff'}^{-1}H_{f\lambda'}]^{-1} = H_{\lambda\lambda'}^{-1} + H_{\lambda\lambda'}^{-1}H_{\lambda f'}[H_{ff'} - H_{f\lambda'}H_{\lambda\lambda'}^{-1}H_{\lambda f'}]^{-1}H_{f\lambda'}H_{\lambda\lambda'}^{-1}. \tag{C.7}$$

It follows that

$$\begin{aligned}
[H_{\phi\phi'}^{-1}R_\phi]_i &= [H_{\lambda\lambda'}^{-1}R_\lambda]_i + [H_{\lambda\lambda'}^{-1}H_{\lambda f'}(H_{ff'} - H_{f\lambda'}H_{\lambda\lambda'}^{-1}H_{\lambda f'})^{-1}H_{f\lambda'}H_{\lambda\lambda'}^{-1}R_\lambda]_i \\
&\quad -[H_{\lambda\lambda'}^{-1}H_{\lambda f'}H_{ff'}^{-1}R_f]_i - [H_{\lambda\lambda'}^{-1}H_{\lambda f'}(H_{ff'} - H_{f\lambda'}H_{\lambda\lambda'}^{-1}H_{\lambda f'})^{-1}H_{f\lambda'}H_{\lambda\lambda'}^{-1}H_{\lambda f'}H_{ff'}^{-1}R_f]_i \\
&\equiv R1i + R2i - R3i - R4i.
\end{aligned} \tag{C.8}$$

We study $R1i, ..., R4i$ in turn.

(R1i) By (C.1) we have $[H_{\lambda\lambda'}^{-1}R_\lambda]_i = [L_{\lambda\lambda'}^{-1}]_i R_{\lambda_i} - [L_{\lambda\lambda'}^{-1}]_i[U_\lambda^0]_i(-\frac{N}{cT}I_{r^2} + U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0)^{-1}U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}R_\lambda$. By Lemmas C.1(i) and C.2(ii),

$$\left\|[L_{\lambda\lambda'}^{-1}]_i R_{\lambda_i}\right\| = O_p(\frac{1}{c_{NT}})\left\|\hat{\lambda}_i - \lambda_i^0\right\| + O_p(\frac{1}{c_{NT}^2}).$$

By (C.2), Lemma C.1(i) and Lemma C.2(i)-(ii),

$$\begin{aligned}
&\left\|[L_{\lambda\lambda'}^{-1}]_i[U_\lambda^0]_i(-\frac{N}{cT}I_{r^2} + U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0)^{-1}U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}R_\lambda\right\| \\
= \;& O_p(\frac{1}{T})O_p(1)O_p(\frac{T}{N})O_p(\sqrt{N})O_p(\frac{1}{T})O_p(\frac{T\sqrt{N}}{c_{NT}^2}) = O_p(\frac{1}{c_{NT}^2}).
\end{aligned}$$

It follows that $R1i = O_p(\frac{1}{c_{NT}})\left\|\hat{\lambda}_i - \lambda_i^0\right\| + O_p(\frac{1}{c_{NT}^2})$.

(R2i) By (C.1) we have

$$\begin{aligned}
R2i &= [L_{\lambda\lambda'}^{-1}]_i[H_{\lambda f'}]_i(H_{ff'} - H_{f\lambda'}H_{\lambda\lambda'}^{-1}H_{\lambda f'})^{-1}H_{f\lambda'}H_{\lambda\lambda'}^{-1}R_\lambda \\
&\quad -[L_{\lambda\lambda'}^{-1}]_i[U_\lambda^0]_i[-\frac{N}{cT}I_{r^2} + U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0]^{-1}U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}H_{\lambda f'}(H_{ff'} - H_{f\lambda'}H_{\lambda\lambda'}^{-1}H_{\lambda f'})^{-1}H_{f\lambda'}H_{\lambda\lambda'}^{-1}R_\lambda \\
&\equiv R21i - R22i.
\end{aligned}$$

Note that $(H_{ff'} - H_{f\lambda'}H_{\lambda\lambda'}^{-1}H_{\lambda f'})^{-1}$ equals the lower-right block of $D_{TN}^{-\frac{1}{2}}(D_{TN}^{-\frac{1}{2}}H_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}D_{TN}^{-\frac{1}{2}}$. By Lemmas B.1 and B.2,

$$\left\|(H_{ff'} - H_{f\lambda'}H_{\lambda\lambda'}^{-1}H_{\lambda f'})^{-1}\right\| = O_p(\frac{1}{N}). \tag{C.9}$$

This together with (C.2), Lemma C.2 and Lemma C.1(i) implies that

$$R21i = O_p(\frac{1}{T})O_p(\sqrt{T})O_p(\frac{1}{N})O_p(\sqrt{NT})O_p(\frac{1}{T})O_p(\frac{T\sqrt{N}}{c_{NT}^2}) = O_p(\frac{1}{c_{NT}^2}),$$

$$\begin{aligned} R22i &= O_p(\frac{1}{T})O_p(1)O_p(\frac{T}{N})O_p(\sqrt{N})O_p(\frac{1}{T})O_p(\sqrt{NT})O_p(\frac{1}{N})O_p(\sqrt{NT})O_p(\frac{1}{T})O_p(\frac{T\sqrt{N}}{c_{NT}^2}) \\ &= O_p(\frac{1}{c_{NT}^2}). \end{aligned}$$

It follows that $R2i = O_p(\frac{1}{c_{NT}^2})$.

(R3i) The analysis is similar to that of $R1i$ and the main difference is that R_λ is replaced by $H_{\lambda f'}H_{ff'}^{-1}R_f$. Part (R1i) uses $\|R_{\lambda_i}\| = \left\|\hat{\lambda}_i - \lambda_i^0\right\| O_p(\frac{T}{c_{NT}}) + O_p(\frac{T}{c_{NT}^2})$ and $\|R_\lambda\| = O_p(\frac{T\sqrt{N}}{c_{NT}^2})$. Here by Lemma C.2(iii)-(iv) and Lemma C.1(ii), we have

$$\begin{aligned} \left\|[H_{\lambda f'}H_{ff'}^{-1}R_f]_i\right\| &\leq \|[H_{\lambda f'}]_i\| \left\|H_{ff'}^{-1}\right\| \|R_f\| = O_p(\sqrt{T})O_p(\frac{1}{N})O_p(\frac{N\sqrt{T}}{c_{NT}^2}) = O_p(\frac{T}{c_{NT}^2}), \\ \left\|H_{\lambda f'}H_{ff'}^{-1}R_f\right\| &\leq \|H_{\lambda f'}\| \left\|H_{ff'}^{-1}\right\| \|R_f\| = O_p(\sqrt{NT})O_p(\frac{1}{N})O_p(\frac{N\sqrt{T}}{c_{NT}^2}) = O_p(\frac{T\sqrt{N}}{c_{NT}^2}). \end{aligned}$$

Thus $R3i = [H_{\lambda\lambda'}^{-1}H_{\lambda f'}H_{ff'}^{-1}R_f]_i = O_p(\frac{1}{c_{NT}^2})$.

(R4i) The analysis is similar to that of $R2i$, and the main difference is that R_λ is replaced by $H_{\lambda f'}H_{ff'}^{-1}R_f$. Part (R2i) uses $\|R_\lambda\| = O_p(\frac{T\sqrt{N}}{c_{NT}^2})$, and here we use $\left\|H_{\lambda f'}H_{ff'}^{-1}R_f\right\| = O_p(\frac{T\sqrt{N}}{c_{NT}^2})$. Then $R4i$ is also $O_p(\frac{1}{c_{NT}^2})$.

Combining the above results for $R1i, ..., R4i$, we have

$$[H_{\phi\phi'}^{-1}R_\phi]_i = O_p(\frac{1}{c_{NT}}) \left\|\hat{\lambda}_i - \lambda_i^0\right\| + O_p(\frac{1}{c_{NT}^2}). \tag{C.10}$$

Now, we study $[H_{\phi\phi'}^{-1}S_\phi]_i$. As in (C.8), we have

$$\begin{aligned} [H_{\phi\phi'}^{-1}S_\phi]_i &= [H_{\lambda\lambda'}^{-1}S_\lambda]_i + [H_{\lambda\lambda'}^{-1}H_{\lambda f'}(H_{ff'} - H_{f\lambda'}H_{\lambda\lambda'}^{-1}H_{\lambda f'})^{-1}H_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda]_i \\ &\quad -[H_{\lambda\lambda'}^{-1}H_{\lambda f'}H_{ff'}^{-1}S_f]_i - [H_{\lambda\lambda'}^{-1}H_{\lambda f'}(H_{ff'} - H_{f\lambda'}H_{\lambda\lambda'}^{-1}H_{\lambda f'})^{-1}H_{f\lambda'}H_{\lambda\lambda'}^{-1}H_{\lambda f'}H_{ff'}^{-1}S_f]_i \\ &\equiv S1i + S2i - S3i - S4i \end{aligned} \tag{C.11}$$

We study $S1i, ..., S4i$ in turn.

(S1i) By (C.1) we have

$$[H_{\lambda\lambda'}^{-1}S_\lambda]_i = [L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i} - [L_{\lambda\lambda'}^{-1}]_i[U_\lambda^0]_i[-\frac{N}{cT}I_{r^2} + U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0]^{-1}U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}S_\lambda. \tag{C.12}$$

By (C.2), (C.3) and Lemma C.2(i)-(ii),

$$\left\|[L_{\lambda\lambda'}^{-1}]_i[U_\lambda^0]_i[-\frac{N}{cT}I_{r^2} + U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0]^{-1}U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}S_\lambda\right\|$$

$$= O_p(\frac{1}{T})O_p(1)O_p(\frac{T}{N})O_p(\sqrt{\frac{N}{T}}+\frac{N}{T}) = O_p(\frac{1}{\sqrt{NT}}+\frac{1}{T}).$$

It follows that $S1i = [L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i} + O_p(\frac{1}{\sqrt{NT}}+\frac{1}{T})$.

(S2i) By (C.1) we have

$$\begin{aligned} S2i &= [L_{\lambda\lambda'}^{-1}]_i[H_{\lambda f'}]_i(H_{ff'}-H_{f\lambda'}H_{\lambda\lambda'}^{-1}H_{\lambda f'})^{-1}H_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda \\ &\quad -[L_{\lambda\lambda'}^{-1}]_i[U_\lambda^0]_i[-\frac{N}{cT}I_{r^2}+U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0]^{-1}U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}H_{\lambda f'}(H_{ff'}-H_{f\lambda'}H_{\lambda\lambda'}^{-1}H_{\lambda f'})^{-1}H_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda \\ &\equiv S21i - S22i. \end{aligned} \tag{C.13}$$

By (C.9), (C.2), Lemma C.2, and Lemma C.3(iv),

$$\begin{aligned} S21i &= O_p(\frac{1}{T})O_p(\sqrt{T})O_p(\frac{1}{N})O_p(\sqrt{N}+\frac{N}{\sqrt{T}}) = O_p(\frac{1}{\sqrt{NT}}+\frac{1}{T}), \\ S22i &= O_p(\frac{1}{T})O_p(1)O_p(\frac{T}{N})O_p(\sqrt{N})O_p(\frac{1}{T})O_p(\sqrt{NT})O_p(\frac{1}{N})O_p(\sqrt{N}+\frac{N}{\sqrt{T}}) \\ &= O_p(\frac{1}{\sqrt{NT}}+\frac{1}{T}). \end{aligned}$$

Then $S2i = O_p(\frac{1}{\sqrt{NT}}+\frac{1}{T})$.

(S3i) As in part (S1i), we have

$$\begin{aligned} &[H_{\lambda\lambda'}^{-1}H_{\lambda f'}H_{ff'}^{-1}S_f]_i \\ =\ & [L_{\lambda\lambda'}^{-1}]_i[H_{\lambda f'}H_{ff'}^{-1}S_f]_i - [L_{\lambda\lambda'}^{-1}]_i[U_\lambda^0]_i[-\frac{N}{cT}I_{r^2}+U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0]^{-1}U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}H_{\lambda f'}H_{ff'}^{-1}S_f. \end{aligned} \tag{C.14}$$

The difference is that S_λ is replaced by $H_{\lambda f'}H_{ff'}^{-1}S_f$ and S_{λ_i} is replaced by $[H_{\lambda f'}H_{ff'}^{-1}S_f]_i$. Part (S1i) uses $\left\|U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\sqrt{\frac{N}{T}}+\frac{N}{T})$ whereas here by Lemmas C.2(i)-(ii) and C.3(iv), we have

$$\begin{aligned} \left\|U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}H_{\lambda f'}H_{ff'}^{-1}S_f\right\| &\leq \left\|U_\lambda^0\right\|\left\|L_{\lambda\lambda'}^{-1}\right\|\left\|H_{\lambda f'}H_{ff'}^{-1}S_f\right\| \\ &= O_p(\sqrt{N})O_p(\frac{1}{T})O_p(\sqrt{T}+\frac{T}{\sqrt{N}}) = O_p(\sqrt{\frac{N}{T}}+1). \end{aligned}$$

In addition, by Lemma C.2(ii) and Lemma C.3(v),

$$[L_{\lambda\lambda'}^{-1}]_i[H_{\lambda f'}H_{ff'}^{-1}S_f]_i = O_p(\frac{1}{T})O_p(\sqrt{\frac{T}{N}}+\frac{T}{N}) = O_p(\frac{1}{\sqrt{NT}}+\frac{1}{N}).$$

Then by the above results, (C.2), and Lemma C.2(i)–(ii),

$$\|S3i\| \leq \left\|[L_{\lambda\lambda'}^{-1}]_i[H_{\lambda f'}H_{ff'}^{-1}S_f]_i\right\| + \left\|[L_{\lambda\lambda'}^{-1}]_i\right\|\left\|[U_\lambda^0]_i\right\|\left\|[-\frac{N}{cT}I_{r^2}+U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0]^{-1}\right\|\left\|U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}H_{\lambda f'}H_{ff'}^{-1}S_f\right\|$$

$$= O_p(\frac{1}{\sqrt{NT}} + \frac{1}{N}) + O_p(\frac{1}{T})O_p(1)O_p(\frac{T}{N})O_p(\sqrt{\frac{N}{T}} + 1) = O_p(\frac{1}{\sqrt{NT}} + \frac{1}{N}).$$

(S4i) The analysis is similar to that in part (S2i). The main difference is that S_λ is replaced by $H_{\lambda f'} H_{ff'}^{-1} S_f$. Part (S2i) uses $\left\|H_{f\lambda'} H_{\lambda\lambda'}^{-1} S_\lambda\right\| = O_p(\sqrt{N} + \frac{N}{\sqrt{T}})$. Here, by Lemma C.2(iii)-(iv) and Lemma C.3(iv),

$$\begin{aligned}\left\|H_{f\lambda'} H_{\lambda\lambda'}^{-1} H_{\lambda f'} H_{ff'}^{-1} S_f\right\| &\leq \left\|H_{f\lambda'}\right\| \left\|H_{\lambda\lambda'}^{-1}\right\| \left\|H_{\lambda f'} H_{ff'}^{-1} S_f\right\| \\ &= O_p(\sqrt{NT}) O_p(\frac{1}{T}) O_p(\sqrt{T} + \frac{T}{\sqrt{N}}) = O_p(\sqrt{N} + \sqrt{T}).\end{aligned}$$

With this change, we can follow the analysis of $S2i$ and showing that $S4i = O_p(\frac{1}{\sqrt{NT}} + \frac{1}{N})$.

Combining the above results for $S1i, ..., S4i$ yields

$$[H_{\phi\phi'}^{-1} S_\phi]_i = [L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i} + O_p(\frac{1}{\sqrt{NT}} + \frac{1}{N} + \frac{1}{T}). \tag{C.15}$$

This, in conjunction with (C.6) and (C.10), implies that

$$\hat{\lambda}_i - \lambda_i^0 = -[L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i} + O_p(\frac{1}{c_{NT}}) \left\|\hat{\lambda}_i - \lambda_i^0\right\| + O_p(\frac{1}{c_{NT}^2}). \tag{C.16}$$

Since $O_p(\frac{1}{c_{NT}}) \left\|\hat{\lambda}_i - \lambda_i^0\right\| = o_p(\left\|\hat{\lambda}_i - \lambda_i^0\right\|)$ and $\left\|[L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i}\right\| = O_p(\frac{1}{\sqrt{T}})$, we have $\left\|\hat{\lambda}_i - \lambda_i^0\right\| = O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{c_{NT}^2})$. Plug this back, we have

$$\hat{\lambda}_i - \lambda_i^0 = -[L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i} + O_p(\frac{1}{\sqrt{T} c_{NT}}) + O_p(\frac{1}{c_{NT}^2}) = -[L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i} + O_p(\frac{1}{c_{NT}^2}).$$

By Assumption 6(iv), we have $-T[L_{\lambda\lambda'}^{-1}]_i \xrightarrow{p} \Sigma_{iF}^{-1}$ and $\frac{1}{\sqrt{T}} S_{\lambda_i} \xrightarrow{d} \mathcal{N}(0, \Omega_{iF})$. Thus if $\frac{T^{\frac{1}{2}}}{c_{NT}^2} \to 0$, we have

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i^0) = [\frac{1}{T} \sum\nolimits_{t=1}^T d_{it} f_t^0 f_t^{0\prime}]^{-1} \frac{1}{\sqrt{T}} \sum\nolimits_{t=1}^T d_{it} f_t^0 v_{it} + O_p(\frac{\sqrt{T}}{c_{NT}^2}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{iF}^{-1} \Omega_{iF} \Sigma_{iF}^{-1}).$$

The limit distributions of the estimated factors follow from symmetric arguments. In particular, we have $\hat{f}_t - f_t^0 = -[L_{ff'}^{-1}]_t S_{f_t} + O_p(\frac{1}{c_{NT}^2})$ and $\sqrt{N}(\hat{f}_t - f_t^0) = [\frac{1}{N} \sum_{i=1}^N d_{it} \lambda_i^0 \lambda_i^{0\prime}]^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N d_{it} \lambda_t^0 v_{it} + O_p(\frac{\sqrt{N}}{c_{NT}^2}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{t\Lambda}^{-1} \Omega_{t\Lambda} \Sigma_{t\Lambda}^{-1})$. ■

D Proof of Theorem 5.1

We shall only prove $(\hat{\Lambda} - \Lambda^0)' a = O_p(\frac{N}{c_{NT}^2})$ as the other claim follows by symmetric arguments. By (3.1) we have

$$(\hat{\Lambda} - \Lambda^0)' a = -\sum\nolimits_{i=1}^N [H_{\phi\phi'}^{-1} S_\phi]_i a_i - \sum\nolimits_{i=1}^N [H_{\phi\phi'}^{-1} R_\phi]_i a_i.$$

It suffices to study $\sum_{i=1}^N [H_{\phi\phi'}^{-1} S_\phi]_i a_i$ and $\sum_{i=1}^N [H_{\phi\phi'}^{-1} R_\phi]_i a_i$.

First, we show $\sum_{i=1}^N [H_{\phi\phi'}^{-1} S_\phi]_i a_i = O_p(\frac{\sqrt{N}}{\sqrt{T}})$. By (C.11) we have

$$\begin{aligned}
[H_{\phi\phi'}^{-1} S_\phi]_i &= [H_{\lambda\lambda'}^{-1} S_\lambda]_i + [H_{\lambda\lambda'}^{-1} H_{\lambda f'}(H_{ff'} - H_{f\lambda'} H_{\lambda\lambda'}^{-1} H_{\lambda f'})^{-1} H_{f\lambda'} H_{\lambda\lambda'}^{-1} S_\lambda]_i \\
&\quad -[H_{\lambda\lambda'}^{-1} H_{\lambda f'} H_{ff'}^{-1} S_f]_i - [H_{\lambda\lambda'}^{-1} H_{\lambda f'}(H_{ff'} - H_{f\lambda'} H_{\lambda\lambda'}^{-1} H_{\lambda f'})^{-1} H_{f\lambda'} H_{\lambda\lambda'}^{-1} H_{\lambda f'} H_{ff'}^{-1} S_f]_i \\
&= S1i + S2i - S3i - S4i. \qquad \text{(D.1)}
\end{aligned}$$

We study $\sum_{i=1}^N S1i \cdot a_i, ..., \sum_{i=1}^N S4i \cdot a_i$ in turn.

(S1i) By (C.12), we have $S1i = [L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i} - [L_{\lambda\lambda'}^{-1}]_i [U_\lambda^0]_i [-\frac{N}{cT} I_{r^2} + U_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} U_\lambda^0]^{-1} U_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} S_\lambda = [L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i} - S12i$. Then

$$\begin{aligned}
\sum_{i=1}^N [L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i} a_i &= \sum_{i=1}^N \sum_{s=1}^T (\sum_{s=1}^T d_{is} f_s^0 f_s^{0\prime})^{-1} d_{is} v_{is} f_s^0 a_i = \frac{1}{T} \sum_{i=1}^N \sum_{s=1}^T \bar{A}_{iF}^{-1} d_{is} v_{is} f_s^0 a_i \\
&= \frac{1}{T} \sum_{i=1}^N \sum_{s=1}^T A_{iF}^{-1} \xi_{a,is} + \frac{1}{T} \sum_{i=1}^N \sum_{s=1}^T (\bar{A}_{iF}^{-1} - A_{iF}^{-1}) \xi_{a,is} \equiv III_{1,1} + III_{1,2}.
\end{aligned}$$

By Assumption 7(i), $III_{1,1} = \frac{\sqrt{N}}{\sqrt{T}} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{s=1}^T A_{iF}^{-1} \xi_{a,is} = O_p(\frac{\sqrt{N}}{\sqrt{T}})$. By Assumptions 6(i) and 5(i)

$$|III_{1,2}| \leq \left\{ \sum_{i=1}^N \left\| \bar{A}_{iF}^{-1} - A_{iF}^{-1} \right\|^2 \right\}^{1/2} \left\{ \sum_{i=1}^N \left\| \frac{1}{T} \sum_{s=1}^T \xi_{a,is} \right\|^2 \right\}^{1/2} = O_p\left(\frac{\sqrt{N}}{\sqrt{T}} \right) O_p(\frac{\sqrt{N}}{\sqrt{T}}).$$

Then $\sum_{i=1}^N [L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i} a_i = O_p(\frac{\sqrt{N}}{\sqrt{T}} + \frac{N}{T})$. For $S12i$, we have by (C.2), (C.3), and Lemmas C.2(i)–(ii) and C.3(i),

$$\begin{aligned}
\sqrt{\sum_{i=1}^N \|S12i\|^2} &= \left\| L_{\lambda\lambda'}^{-1} U_\lambda^0 [-\frac{N}{cT} I_{r^2} + U_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} U_\lambda^0]^{-1} U_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} S_\lambda \right\| \\
&\leq \left\| L_{\lambda\lambda'}^{-1} \right\| \left\| U_\lambda^0 \right\| \left\| [-\frac{N}{cT} I_{r^2} + U_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} U_\lambda^0]^{-1} \right\| \left\| U_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} S_\lambda \right\| \\
&= O_p(\frac{1}{T}) O_p(\sqrt{N}) O_p(\frac{T}{N}) O_p(\sqrt{\frac{N}{T}} + \frac{N}{T}) = O_p(\frac{1}{\sqrt{T}} + \frac{\sqrt{N}}{T}).
\end{aligned}$$

Then $\left\| \sum_{i=1}^N S12i a_i \right\| \leq \sqrt{\sum_{i=1}^N \|S12i\|^2} \|a\| = O_p(\frac{\sqrt{N}}{\sqrt{T}} + \frac{N}{T})$ by the CS inequality. It follows that $\sum_{i=1}^N S1i \cdot a_i = O_p(\frac{\sqrt{N}}{\sqrt{T}} + \frac{N}{T})$.

(S2i) By (C.13) we have

$$\begin{aligned}
S2i &= [L_{\lambda\lambda'}^{-1}]_i [H_{\lambda f'}]_i (H_{ff'} - H_{f\lambda'} H_{\lambda\lambda'}^{-1} H_{\lambda f'})^{-1} H_{f\lambda'} H_{\lambda\lambda'}^{-1} S_\lambda \\
&\quad -[L_{\lambda\lambda'}^{-1}]_i [U_\lambda^0]_i [-\frac{N}{cT} I_{r^2} + U_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} U_\lambda^0]^{-1} U_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} H_{\lambda f'} (H_{ff'} - H_{f\lambda'} H_{\lambda\lambda'}^{-1} H_{\lambda f'})^{-1} H_{f\lambda'} H_{\lambda\lambda'}^{-1} S_\lambda \\
&\equiv S21i + S22i.
\end{aligned}$$

By Lemmas C.2 and C.3(iv), (C.2), and (C.9), we have

$$\begin{aligned}\sqrt{\sum_{i=1}^N \|S21i\|^2} &= O_p(\frac{1}{T})O_p(\sqrt{NT})O_p(\frac{1}{N})O_p(\sqrt{N}+\frac{N}{\sqrt{T}}) = O_p(\frac{1}{\sqrt{T}}+\frac{\sqrt{N}}{T}),\\ \sqrt{\sum_{i=1}^N \|S22i\|^2} &= O_p(\frac{1}{T})O_p(\sqrt{N})O_p(\frac{T}{N})O_p(\sqrt{N})O_p(\frac{1}{T})O_p(\sqrt{NT})O_p(\frac{1}{N})O_p(\sqrt{N}+\frac{N}{\sqrt{T}})\\ &= O_p(\frac{1}{\sqrt{T}}+\frac{\sqrt{N}}{T}).\end{aligned}$$

It follows that $\left\|\sum_{i=1}^N S2ia_i\right\| \leq (\sqrt{\sum_{i=1}^N \|S21i\|^2}+\sqrt{\sum_{i=1}^N \|S22i\|^2})\,\|a\| = O_p(\frac{\sqrt{N}}{\sqrt{T}}+\frac{N}{T})$.

(S3i) By (C.14) we have

$$\begin{aligned}[H_{\lambda\lambda'}^{-1}H_{\lambda f'}H_{ff'}^{-1}S_f]_i &= [L_{\lambda\lambda'}^{-1}]_i[H_{\lambda f'}H_{ff'}^{-1}S_f]_i - [L_{\lambda\lambda'}^{-1}]_i[U_\lambda^0]_i[-\frac{N}{cT}I_{r^2}+U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0]^{-1}U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}H_{\lambda f'}H_{ff'}^{-1}S_f\\ &\equiv S31i - S32i.\end{aligned}$$

Compared to part (S1i), the difference is that S_λ is replaced by $H_{\lambda f'}H_{ff'}^{-1}S_f$ and S_{λ_i} is replaced by $[H_{\lambda f'}H_{ff'}^{-1}S_f]_i$. By Lemma C.2(ii) and Lemma C.3(iv),

$$\sqrt{\sum_{i=1}^N \|S31i\|^2} = \left\|L_{\lambda\lambda'}^{-1}\right\|\left\|H_{\lambda f'}H_{ff'}^{-1}S_f\right\| = O_p(\frac{1}{T})O_p(\sqrt{T}+\frac{T}{\sqrt{N}}) = O_p(\frac{1}{\sqrt{N}}+\frac{1}{\sqrt{T}}).$$

By (C.2), and Lemmas C.2(i)-(ii) and C.3(iv), we have

$$\begin{aligned}\sqrt{\sum_{i=1}^N \|S32i\|^2} &= \left\|L_{\lambda\lambda'}^{-1}U_\lambda^0[-\frac{N}{cT}I_{r^2}+U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0]^{-1}U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}H_{\lambda f'}H_{ff'}^{-1}S_f\right\|\\ &\leq \left\|L_{\lambda\lambda'}^{-1}\right\|^2\left\|U_\lambda^0\right\|^2\left\|[-\frac{N}{cT}I_{r^2}+U_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}U_\lambda^0]^{-1}\right\|\left\|H_{\lambda f'}H_{ff'}^{-1}S_f\right\|\\ &= O_p(\frac{1}{T^2})O_p(N)O_p(\frac{T}{N})O_p(\sqrt{T}+\frac{T}{\sqrt{N}}) = O_p(\frac{1}{\sqrt{N}}+\frac{1}{\sqrt{T}}).\end{aligned}$$

It follows that $\left\|\sum_{i=1}^N S3ia_i\right\| \leq (\sqrt{\sum_{i=1}^N \|S31i\|^2}+\sqrt{\sum_{i=1}^N \|S32i\|^2})\,\|a\| = O_p(\frac{\sqrt{N}}{\sqrt{T}}+1)$.

(S4i) The analysis is similar to case of part (S2i). The difference is that S_λ is replaced by $H_{\lambda f'}H_{ff'}^{-1}S_f$. Part (S2i) uses $\left\|H_{f\lambda'}H_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\sqrt{N}+\frac{N}{\sqrt{T}})$. Now, by Lemma C.2(ii) and (iv) and Lemma C.3(iv),

$$\begin{aligned}\left\|H_{f\lambda'}H_{\lambda\lambda'}^{-1}H_{\lambda f'}H_{ff'}^{-1}S_f\right\| &\leq \left\|H_{f\lambda'}\right\|\left\|H_{\lambda\lambda'}^{-1}\right\|\left\|H_{\lambda f'}H_{ff'}^{-1}S_f\right\| = O_p(\sqrt{NT})O_p(\frac{1}{T})O_p(\sqrt{T}+\frac{T}{\sqrt{N}})\\ &= O_p(\sqrt{N}+\sqrt{T}).\end{aligned}$$

Then we also have $\left\|\sum_{i=1}^N S4i\cdot a_i\right\| = O_p(\frac{\sqrt{N}}{\sqrt{T}}+1)$. In sum, we have $\sum_{i=1}^N[H_{\phi\phi'}^{-1}S_\phi]_i a_i = O_p(\frac{\sqrt{N}}{\sqrt{T}}+\frac{N}{T}+1)$.

Now, we consider $\sum_{i=1}^{N}[H_{\phi\phi'}^{-1}R_\phi]_i a_i$. Note that

$$\begin{aligned}\left\|\sum_{i=1}^{N}[H_{\phi\phi'}^{-1}R_\phi]_i a_i\right\| &\leq \sqrt{\sum_{i=1}^{N}\left\|[H_{\phi\phi'}^{-1}R_\phi]_i\right\|^2}\,\|a\| \leq \frac{\|a\|}{\sqrt{T}}\left\|(D_{TN}^{-\frac{1}{2}}H_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}D_{TN}^{-\frac{1}{2}}R_\phi\right\| \\ &\leq \frac{\|a\|}{\sqrt{T}}\left\|\left\|(D_{TN}^{-\frac{1}{2}}H_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}\right\|D_{TN}^{-\frac{1}{2}}R_\phi\right\| \\ &= O_p(\frac{\sqrt{N}}{\sqrt{T}})O_p(1)\,O_p(\frac{\sqrt{NT}}{c_{NT}^2}) = O_p(\frac{N}{c_{NT}^2}), \qquad \text{(D.2)}\end{aligned}$$

by Lemmas B.1, B.2 and B.3. It follows that $\frac{1}{N}(\hat{\Lambda}-\Lambda^0)'a = \frac{1}{N}O_p(\frac{\sqrt{N}}{\sqrt{T}}+\frac{N}{T}+1+\frac{N}{c_{NT}^2}) = O_p(\frac{1}{c_{NT}^2})$. ■

E Proof of Proposition 5.1

Proof of Proposition 5.1. We focus on the asymptotic distribution of $\hat{\tau}_{\cdot t}-\tau_{\cdot t}$ as the asymptotic distribution of $\hat{\tau}_{i\cdot}-\tau_{i\cdot}$ follows by analogous arguments.

Noting that $\hat{\tau}_{it}-\tau_{it} = -x_{it}'(\hat{\beta}-\beta^0)-(\hat{f}_t'\hat{\lambda}_i - f_t^{0\prime}\lambda_i^0)+\varepsilon_{it}$, we have

$$\begin{aligned}\hat{\tau}_{\cdot t}-\tau_{\cdot t} &= -\frac{1}{a'a}\sum_{i=1}^{N}a_i x_{it}'(\hat{\beta}-\beta^0) - \frac{1}{a'a}\sum_{i=1}^{N}a_i(\hat{\lambda}_i-\lambda_i^0)'\hat{f}_t \\ &\quad -\frac{1}{a'a}\sum_{i=1}^{N}a_i\lambda_i^{0\prime}(\hat{f}_t-f_t^0)+\frac{1}{a'a}\sum_{i=1}^{N}a_i\varepsilon_{it} \\ &\equiv -\Delta_{1,t}-\Delta_{2,t}-\Delta_{3,t}+\Delta_{4,t}. \qquad \text{(E.1)}\end{aligned}$$

Noting that $\frac{1}{a'a}||\sum_{i=1}^{N}a_i x_{it}'|| \leq \frac{1}{a'a}\sqrt{(\sum_{i=1}^{N}a_i^2)(\sum_{i=1}^{N}\|x_{it}\|^2)} = \sqrt{\frac{1}{a'a}\sum_{i=1}^{N}\|x_{it}\|^2} = O_p(1)$, $|\Delta_{1,t}| \leq O_p(1)\,||\hat{\beta}-\beta^0|| = O_p(\frac{1}{c_{NT}^2})$ by Assumption 8(i)–(ii). By Theorem 5.1(i) and Assumption 8(i), $|\Delta_{2,t}| = \frac{1}{a'a}O_p(\frac{N}{c_{NT}^2}) = O_p(\frac{1}{c_{NT}^2})$. It is easy to check that given the fast convergence rate of $\hat{\beta}$, the estimators $\hat{f}_t$ and $\hat{\lambda}_i$ in Step 2 of Algorithm 5.1 share the same asymptotic properties as stated in Theorem 4.3. By the proof of Theorem 4.3, we now have

$$\sqrt{N}(\hat{f}_t-f_t^0) = [\frac{1}{N}\sum_{i=1}^{N}d_{it}\lambda_i^0\lambda_i^{0\prime}]^{-1}\frac{1}{N^{1/2}}\sum_{i=1}^{N}d_{it}\lambda_i^0\varepsilon_{it}+O_p(\frac{\sqrt{N}}{c_{NT}^2}) \xrightarrow{d} \mathcal{N}(0,\Sigma_{t\Lambda}^{-1}\Omega_{t\Lambda}\Sigma_{t\Lambda}^{-1}).$$

By Assumption 7 with v_{it} replaced by ε_{it}, $\frac{1}{\sqrt{a'a}}\sum_{i=1}^{N}a_i\varepsilon_{it} \xrightarrow{d} \mathcal{N}(0,\Omega_{ta})$. Let $C_{a\Lambda} = \frac{1}{a'a}\sum_{i=1}^{N}a_i\lambda_i^0$. Then by the Cramér-Wold device and Slutsky theorem, we have

$$\begin{aligned}\sqrt{N}(\hat{\tau}_{\cdot t}-\tau_{\cdot t}) &= C_{a\Lambda}'[\frac{1}{N}\sum_{i=1}^{N}d_{it}\lambda_i^0\lambda_i^{0\prime}]^{-1}\frac{1}{N^{1/2}}\sum_{i=1}^{N}d_{it}\lambda_i^0\varepsilon_{it}+\sqrt{\frac{N}{a'a}}\frac{1}{\sqrt{a'a}}\sum_{i=1}^{N}a_i\varepsilon_{it}+o_P(1) \\ &\xrightarrow{d} \mathcal{N}(0,\ \text{plim}_{N\to\infty}\sigma_{N\tau_t}^2),\end{aligned}$$

where $\sigma_{N\tau_t}^2 = C_{a\Lambda}'\Sigma_{t\Lambda}^{-1}\Omega_{t\Lambda}\Sigma_{t\Lambda}^{-1}C_{a\Lambda}+\frac{N}{a'a}\Omega_{ta}+2C_{a\Lambda}'\Sigma_{t\Lambda}^{-1}\frac{1}{a'a}\sum_{i,j=1}^{N}\mathbb{E}(d_{it}\lambda_i^0\varepsilon_{it}\varepsilon_{jt}a_j)$. ■

References

[1] Bernstein, D.S., 2005. Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory. Princeton University Press, New Jersey.

[2] Horn, R.A. and Johnson, C. R., 1991. Topics in Matrix Analysis. Cambridge University Press, Cambridge.

[3] Seber, G.A.F., 2008. A Matrix Handbook for Statisticians. John Wiley & Sons, New Jersey.

[4] Wang, F., 2022. Maximum Likelihood Estimation and Inference for High Dimensional Generalized Factor Models with Application to Factor-Augmented Regressions. *Journal of Econometrics* 229, 180-200.

[5] Yu, Y., Wang, T., and Samworth, R.J., 2015. A Useful Variant of the Davis-Kahan Theorem for Statisticians. *Biometrika* 102, 315–323.