Estimation and Inference for Unbalanced Panel Data Models with Interactive Fixed Effects*

Liangjun Su[†] Fa Wang[‡] Yiren Wang[§]

March 7, 2025

Abstract

This paper establishes the inferential theory for unbalanced panel data models with interactive fixed effects. We propose a two-step estimation algorithm with the first step obtaining an initial consistent estimator followed by an alternating maximization procedure. We prove that the alternating maximization procedure is a contractionary mapping and the final estimator is asymptotically normal, as long as the initial estimator is consistent. We also develop analytical bias corrections according to the derived asymptotic bias expressions and the observed missing pattern. Our results cover important missing patterns such as completely exogenous missing, selection on regressors/factors/loadings and block/staggered missing, and we also show that our results can be readily extended to cases with a Heckman correction term or more general settings. An empirical analysis of the U.S. state-level tax rates from 1951 to 2000 with missing data reveals persistence in tax rates, while state income influences different taxes in varying ways.

Keywords: Dynamic Panel, Gauss-Seidel Algorithm, Interactive Effects, Missing Not At Random, Unbalanced Panel

JEL Classification: C33, C38, C55

*Su gratefully acknowledges the NSFC for financial support under the grant number 72133002.
[†]School of Economics and Management, Tsinghua University, China
[‡]Department of Finance, School of Economics, Peking University, China
[§]College of Finance and Statistics, Hunan University, China

1 Introduction

Factor model provides a parsimonious way to capture heterogeneous impacts of common shocks, time-varying effects of individual heterogeneity and cross-sectional dependence. In the past two decades panel data models with a factor error structure have experienced substantial development and have been widely applied in asset pricing, applied micro and empirical macro studies. In practice, it is quite common to encounter missing data. For example, the issues of attrition, sample selection and non-response are well known in applied micro studies. Participants of surveys may leave due to relocation, retirement or loss of interest. Participants may also skip sensitive questions such as income, health status and family planning. In financial and macro studies, observations may be missing due to company bankruptcy or mixed sampling frequency. In causal panel data, the untreated potential outcomes after treatment are unobservable and the treatment adoption could be simultaneous, staggered or switching on and off. See Verbeek and Nijman (1996), Baltagi and Song (2006), Bańbura and Modugno (2014), Athey et al. (2021), Bai and Ng (2021), Jin et al. (2021) and Agarwal et al. (2023) for detailed discussions.

While panel data models with interactive fixed effects (IFEs hereafter) are extensively studied and missing data is common in practice, there is almost no formal research on unbalanced panel with IFEs except for some simulation studies in Bai et al. (2015) and Czarnowske and Stammann (2020). To fill the void, this paper studies the numerical convergence and asymptotic properties of a two-step procedure for the estimation and inference of unbalanced panel with IFEs. In the first step we perform nuclear norm regularized (NNR) quasi maximum likelihood estimation. In the second step we use the first step estimator as initial value and maximize the quasi likelihood function iteratively by alternating maximization between the regression coefficients and the factors and loadings. We formally prove that the first step estimator is consistent and the second step as a block Gauss-Seidel (or Jacobi) procedure is a contractionary mapping within a local neighborhood of the true parameters. Then the convergence of the second step is guaranteed given the consistency of the first step estimator. We also provide the details of the Alternating Direction Method of Multipliers (ADMM) algorithm for computing the NNR estimator and prove the consistency of the estimated number of factors.

We establish the convergence rate and the limit distribution of the second step estimator in a unified framework that covers both static and dynamic panels and both random and block-type missing patterns. For random missing, we allow the missing probabilities to be cross-sectionally and serially heterogenous and correlated with the covariates, factors and loadings. We also allow the missing indicators to be serially dependent, which is crucial when the panel data model is dynamic and when the missing indicators of each individual follow a dynamic binary process, because in these cases the missing indicators are naturally serially dependent. Our assumptions on random missing are relevant for survey data, asset pricing panels, recommendation system, switchback digital platform experiments and other scenarios where the missingness arises from decision behavior and/or the missing probability is correlated with the missing value itself. For block-type missing, we only require that we have a positive fraction of completely observed individuals and a positive fraction of completely observed periods. This includes exact block missing, staggered missing, mixed frequency and regular missing as special cases, which are relevant for treatment effect estimation and macro/financial applications. A fundamental difference between the random missing and the block-type missing is that the latter does not treat the missing indicator as a random variable, since the analysis is conditioned on the block-type missing pattern.

For all the above missing patterns, the second step estimator of the regression coefficients is consistent and asymptotically normal but may have biases of the same asymptotic order as their standard deviations as N (the number of individuals) and T (the number of periods) tend to infinity jointly. The derived bias expressions show that the biases depend on the missing pattern and consequently bias correction needs to be done according to the specific observed missing pattern. Our simulation results indicate that when the missing probability is correlated with the factors and loadings, rank estimation needs larger sample size and the bias tends to be larger compared to other missing patterns. When there is no missing data, our bias expressions are the same as those in Bai (2009) for static panels and in Moon and Weidner (2017) for dynamic panels.

As in Bai (2009) and Moon and Weidner (2017), we need to tackle the incidental parameter problem resulting from estimating the factors and loadings. The main difficulty here is that explicit expressions of the estimated regression coefficients and the estimated factors and loadings are no longer available when there is missing data. Therefore we take asymptotic expansion of the first order conditions and analyze the high dimensional Hessian and third order terms directly. This is feasible since the IFEs model gives the Hessian and third order terms special structures that allow us to decompose them and analyze their asymptotic behavior. We show that the upper bound of the magnitude of third order terms is not affected by any missing pattern,[1] but the Hessian structure depends on the missing pattern. The assumptions we impose on the missing patterns are very general but sufficient to ensure that missingness does not destroy the local concavity of the Hessian.

Related literature and contributions of this paper. First, there is a large literature on panel data model with IFEs. Early studies consider GMM estimation under the large N fixed T asymptotics; see, e.g., Holtz-Eakin et al. (1988), Ahn et al. (2001) and Ahn et al. (2013). Recent developments can be broadly divided into two branches. One follows the common correlated effects (CCE) approach of Pesaran (2006); see, e.g., Chudik and Pesaran (2015), Chudik and Pesaran

[1]See Lemma E.4 in the appendix.

(2013), Westerlund and Urbain (2015), and Juodis and Sarafidis (2022). The other follows the principal component analysis (PCA) approach of Bai (2009); see, e.g., Lu and Su (2016), Moon and Weidner (2017), Shi and Lee (2017), Chen et al. (2021b) and Hong et al. (2023). As discussed in Chudik and Pesaran (2013), computationally missing data is not an issue for the CCE estimators since for unbalanced panels it is still straightforward to compute the cross-sectional averages of the dependent variable and the covariates. However, when the missing probability is correlated with the covariates/factors/loadings, it is no longer appropriate to put the cross-sectional averages in the regression as proxies of the factors, since the relationship between the factors and the cross-sectional averages are no longer stable over time.[2] As far as we know, this topic has not been fully studied in the literature. For the PCA approach, Bai (2009) remarks that one may modify the alternating maximization (AM hereafter) procedure to handle unbalanced panel where the EM algorithm is utilized to calculate the factors and loadings given the regression coefficients. However, the numerical convergence and asymptotic properties of this modified procedure are totally unknown. This paper proves that the modified AM procedure is guaranteed to converge and the regression coefficient estimators are consistent and asymptotically normal under a wide variety of missing patterns, as long as certain initial estimator is consistent. Also, the derived bias expressions show how the biases depend on the missing pattern and provide guidance for bias corrections and randomized experimental design.

When there is no missing data, iterative estimation of panel data models with IFEs has been studied by Jiang et al. (2021), Moon and Weidner (2023) and Hong et al. (2023).[3] In these papers, both the convergence proof of the alternating minimization and the asymptotic analyses rely on the explicit expressions of the estimated regression coefficients and the estimated factors and loadings, which are no longer available for unbalanced panels. Therefore, our convergence proof and asymptotic analyses are totally new and generally applicable to various alternating maximization/minimization problems.

Second, there is also a large literature on panel data models with missing data; see, e.g., Honoré (1992), Hu (2002) and Honoré and Hu (2004) for censored panels and Wooldridge (1995), Wooldridge (2019), Kyriazidou (1997), Kyriazidou (2001), Dustmann and Rochina-Barrachina (2007), Semykina and Wooldridge (2010) and Semykina and Wooldridge (2013) for sample selection panels. These studies focus on models with only individual effects under the large N and fixed T setup. However,

[2] See the footnote of Example 2 in Section 2 for details.

[3] Jiang et al. (2021) show that the regression coefficient estimators in all iterations are consistent and asymptotically normal if the initial estimator is consistent, but they may be inconsistent for arbitrary initial estimators unless the regressors satisfy some restrictive conditions. Moon and Weidner (2023) propose to use NNR estimator as the consistent initial estimator and show that alternating minimization in the second step converges to the local minimum. Hong et al. (2023) propose a two-step profile GMM estimation for panel data models with IFEs and endogenous regressors and also prove the convergence of the alternating minimization in the second step.

in addition to the individual effects, in many cases the sample selection process is also affected by common shocks such as business cycles, technology progresses and treatment adoption designs. For some technical reasons,[4] sample selection panels with both individual and time fixed effects are rarely studied. Instead, recent progresses on the two-way fixed effects (TWFE) panel regression focus on the block/staggered missing patterns to make inference on heterogenous treatment effect; see, e.g., de Chaisemartin and d'Haultfoeuille (2020), Goodman-Bacon (2021), Sun and Abraham (2021), Callaway and Sant'Anna (2021), Borusyak et al. (2024), Athey and Imbens (2022) and Arkhangelsky and Imbens (2024).[5] Given these previous studies, this paper shows that IFEs provide a more general way to capture time-varying latent confounders in panel data sample selection models. We prove that the regression coefficient estimator has no selection bias as long as the latent confounders that affect both the outcome and the treatment indicator can be captured by a factor structure. Based on our asymptotic analysis, it's not difficult to further extend the results to allow the error terms in the selection equation and the main equation to be correlated, say, by adding a Heckman correction term under some distributional assumptions. To crystallize our asymptotic analyses and stay focused, we do not pursue this extension formally in this paper. In addition, IFEs also help to relax the parallel trend assumption in the TWFE panel literature; see Callaway and Karami (2023) and the references therein.

Third, while we focus on linear panel with IFEs, our analyses for the numerical convergence and the asymptotic properties are prototypical for fixed effects estimation of unbalanced panels. We prove that the AM procedure is a contractionary mapping as long as the Hessian is locally concave, and our proof can be readily extended to other settings such as nonlinear panels, TWFE models, grouped fixed effects models and other missing patterns. The asymptotic distribution and bias of the QMLE under these settings can also be derived following similar steps as used in our analyses. As discussed in Fernández-Val and Weidner (2018), many panel data estimators have computation packages/solutions for the corresponding unbalanced cases, but there is no formal study that proves the asymptotic validity of these estimators for the unbalanced case under the relevant missing conditions, or extends the bias corrections to the unbalanced case. This paper seeks to fill this gap. We show that the asymptotic theory for the unbalanced panels is nontrivial, especially when we deviate from the case of completely random missing. The bias corrections also need to take the

[4]For example, as pointed out in Charbonneau (2017), under the large N fixed T asymptotics, Manski (1987)'s maximum score estimator for binary response panels cannot be extended to settings with both individual and time effects. The presence of time effects also makes it harder to difference out the sample selection correction term. The difficulites are partially circumvented by the large N large T asymptotics and recent progress are made in Hahn and Moon (2006), Moon and Weidner (2017), Fernández-Val and Weidner (2016) and Chen et al. (2021b) for dynamic/nonlinear panels with two way/interactive fixed effects. Fernández-Val and Vella (2011) study sample selection panels under the large N large T asymptotics but still only consider individual effects.

[5]The main focus of these papers is on the relationship between the TWFE and the DID estimator under heterogenous treatment effects.

specific missing patterns into account.

Outline. The rest of the paper is structured as follows. Section 2 introduces the notations and the missing patterns. Section 3 presents the estimation procedure and the relevant technical details. Section 4 presents the limit distribution of the proposed estimator and the analytical bias corrections. Section 5 discusses the Heckman correction for the sample selection with IFEs. Section 6 presents the simulation results. Section 7 presents an application to the US state tax data. Section 8 concludes. All proofs are relegated to the online appendix.

Notation. Let $(N,T) \to \infty$ denote N and T going to infinity jointly and $c_{NT} = \min\{N^{\frac{1}{2}}, T^{\frac{1}{2}}\}$. $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and distribution, respectively. "w.p.a.1" denotes with probability approaching 1. For real numbers a and b, $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. $\|\cdot\|$ denotes the Euclidean norm for vector and spectral norm for matrix. $\|\cdot\|_*$, $\|\cdot\|_{\max}$ and $\|\cdot\|_F$ denote the nuclear norm, max norm and Frobenius norm, respectively. $\sigma_{\min}(\cdot)$ denotes the smallest eigenvalue. Denote $[n] = \{1, \cdots, n\}$ for any positive integer n. M denotes some large positive constant that may vary throughout the paper.

2 Model Setup

Consider the following panel data model with IFEs and missing values:

$$\begin{aligned} y_{it} &= d_{it}(x_{it}'\beta^0 + f_t^{0\prime}\lambda_i^0 + v_{it}) \text{ for } i \in [N] \text{ and } t \in [T], \\ d_{it} &= 1\{y_{it} \text{ and } x_{it} \text{ are observable}\}, \end{aligned} \tag{2.1}$$

where f_t^0 is the r-dimensional unobservable factor at time t, λ_i^0 is the r-dimensional unobservable loading of unit i, r is the number of factors, $x_{it} = (x_{it1}, ..., x_{itK})'$ is the K-dimensional vector of regressors, β^0 is the K-dimensional regression coefficients, v_{it} is the error term, and $1\{\cdot\}$ denotes the indicator function. r and K are fixed as $(N,T) \to \infty$. We assume r is known for the moment and then propose a consistent estimator for r in Section 3.3. Our objective is to estimate β^0 based on the unbalanced panel data and study its asymptotic properties. The IFEs model allows us to incorporate heterogeneous impacts of common shocks f_t^0 and time-varying effects of individual heterogeneities λ_i^0, and the common shocks and the individual heterogeneities are allowed to be correlated with the regressors x_{it}. We allow x_{it} to contain lagged dependent variables, and thus our results are valid for both static and dynamic panels. We also allow x_{it} to contain low-rank regressors provided some regularity conditions are satisfied.

Let $\lambda = (\lambda_1', ..., \lambda_N')'$, $f = (f_1', ..., f_T')'$, $\Lambda = (\lambda_1, ..., \lambda_N)'$, $F = (f_1, ..., f_T)'$ and $\Theta = F\Lambda'$. For notational simplicity, we also define $\phi = (\lambda', f')'$ and $\gamma = (\beta', \lambda', f')'$. Note that the dimensions of

λ, f, Λ, F, Θ, ϕ and γ are $Nr\times 1$, $Tr\times 1$, $N\times r$, $T\times r$, $T\times N$, $(Nr+Tr)\times 1$ and $(K+Nr+Tr)\times 1$, respectively. We use β^0, λ^0, f^0, Λ^0, F^0, Θ^0, ϕ^0 and γ^0 to denote the true parameter values.

In addition, let $\mathbb{E}_\phi$ denote the expectation conditioning on the factors and loadings ϕ^0, $\Phi_{it}=\mathbb{E}_\phi(d_{it})$ and $\tilde{d}_{it}=d_{it}-\Phi_{it}$. For $k\in[K]$, let δ^0_{ki} and ω^0_{kt} be the solution of the following least squares problem:

$$\min_{\{\delta_{ki},\ \omega_{kt}\}}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\mathbb{E}_\phi(d_{it})\left\|\frac{\mathbb{E}_\phi(d_{it}x_{itk})}{\mathbb{E}_\phi(d_{it})}-\delta'_{ki}f^0_t-\omega'_{kt}\lambda^0_i\right\|^2, \tag{2.2}$$

and let δ^0_i and ω^0_t denote $K\times r$ matrices such that δ^0_{ki} and ω^0_{kt} are the transpose of the k-th row of δ^0_i and ω^0_t, respectively. Now we present the missing pattern conditions considered in this paper.

Assumption 1 *(i) For the random missing, $\tilde{d}_{it}$ is independent across i conditioning on ϕ^0 and $\tilde{d}_{it}$ is independent with v_{js} for all i,t,j,s; Φ_{it} is independent with v_{js} and $\min_{i,t}\Phi_{it}\geq\underline{d}>0$; $\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_{Nd}(t,s)\leq M$ where $\gamma_{Nd}(t,s)=\frac{1}{N}\sum_{i=1}^{N}|\mathbb{E}_\phi(\tilde{d}_{it}\tilde{d}_{is})|$; for some κ with $\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}}\to 0$ and $\frac{T^{\frac{1}{\kappa}}}{\sqrt{N}}\to 0$, $\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{d}_{it}f^0_tf^{0\prime}_t\right\|^\kappa)\leq M$, $\mathbb{E}(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{d}_{it}\lambda^0_i\lambda^{0\prime}_i\right\|^\kappa)\leq M$, $\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{d}_{it}f^0_t\omega^{0\prime}_{kt}\right\|^\kappa)\leq M$ and $\mathbb{E}(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{d}_{it}\lambda^0_i\delta^{0\prime}_{ki}\right\|^\kappa)\leq M$.*

(ii) For the block-type missing, $\{d_{it}:i\in[N]$ and $t\in[T]\}$ is nonrandom, $d_{it}=1$ 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 denotes the number of fully observed columns and rows, respectively. As $(N,T)\to\infty$, both N_o/N and T_o/T are bounded away from zero, and both $\sigma_{\min}(\frac{1}{T}\sum_{t=1}^{T_o}f^0_tf^{0\prime}_t)$ and $\sigma_{\min}(\frac{1}{N}\sum_{i=1}^{N_o}\lambda^0_i\lambda^{0\prime}_i)$ are bounded away from zero in probability.

Assumption 1(i)–(ii) covers the random missing and the block-type missing, respectively. Our results require either Assumption 1(i) or 1(ii) but not both. The main difference between these two types of missing is that d_{it} is a random variable in Assumption 1(i) while in Assumption 1(ii) d_{it} is treated as a given nonrandom variable. Consequently, $\mathbb{E}_\phi(d_{it})\geq\underline{d}>0$ for all i and t under random missing while under block-type missing $\mathbb{E}_\phi(d_{it})=0$ for the missing observations. For both types of missing patterns, d_{it} is assumed to be independent with v_{js} for all (i,j) and (t,s). If d_{it} is correlated with v_{js} conditioning on the factors and loadings, then additional procedures are needed to correct the sample selection bias, say, via Heckman corrections or propensity-score-based methods. Our results provide the tools for these further generalizations. See the discussion on missing patterns in Su and Wang (2024) from the perspective of a pure factor model.

Assumption 1(i) allows $\mathbb{E}_\phi(d_{it})$ to vary across i and t and correlate with λ^0_j and f^0_s for some (j,s), i.e., missingness can depend endogenously on multiple time-varying individual effects. For example, firms are more likely to bankrupt and drop out of the panel during economic recessions. Workers' decision to work and wage rates are correlated with both workers' individual characteristics and macroeconomic shocks. The estimated regression coefficients would be inconsistent if we only include individual effects in the regression while the true model contains IFEs. Thus the IFEs provides us a more general way to capture latent confounders. Assumption 1(i) also allows $\tilde{d}_{it}$ to be weakly dependent across t conditioning on ϕ^0 in the sense that the moment conditions in Assumption 1(i) are satisfied if $\tilde{d}_{it}$ is weakly dependent across t. These moment conditions are what we need in the proof and can be verified once we impose specific models on d_{it}. Some important examples of the models of d_{it} under Assumption 1(i) are listed below.

Example 1 (Completely random missing): *d_{it} is independent across i and t and independent with the factors, loadings and error terms.*

Example 2 (Selection on covariates/factors/loadings): *d_{it} follows a binary process for each i, where $\mathbb{E}_\phi(d_{it})$ may depend on the factors and loadings, $d_{i,t-1}$* and *some observable variables.*[6] *Specifically, consider the case where $d^*_{it} = z'_{it}\delta^0 + g_t^{0\prime}\alpha_i^0 + u_{it}$* and *$d_{it} = 1\{d^*_{it} > 0\}$, where z_{it} may contain x_{it}, $y_{i,t-1}$, $d_{i,t-1}$ and other observable variables, g_t^0 and α_i^0 denote some latent factors and loadings and u_{it} is the error term. z_{it}, g_t^0 and α_i^0 could be correlated with f_t^0 and λ_i^0; u_{it} is independent with v_{it} and both are independent with z_{it}, g_t^0 and α_i^0. Clearly, $\tilde{d}_{it}$ is dependent across t conditioning on ϕ^0 if u_{it} is independent across i and t and z_{it} contains $y_{i,t-1}$ or $d_{i,t-1}$. A special case is $d^*_{it} = d_{i,t-1}\delta^0 + f_t^{0\prime}\lambda_i^0 + u_{it}$ and u_{it} is independent across i and t. In this case d_{it} is a first order Markov process, which is relevant for the switchback missing pattern in the high-tech industry experiments.*

Example 3 (Dynamic panel): *Without loss of generality, consider the case $x_{it} = y_{i,t-1}$. Let $d^y_{it} = 1\{y_{it}$ is observable$\}$. Then we have $d_{it} = 1\{y_{it}$ and x_{it} are observable$\} = d^y_{it}d^y_{i,t-1}$. Suppose d^y_{it} is independent across i and t conditioning on ϕ, then d_{it} is correlated with $d_{i,t-1}$ but uncorrelated with $d_{i,t-l}$ for $l \geq 2$.*[7] *Thus for dynamic panels it is crucial to allow d_{it} to be weakly dependent across t conditioning on ϕ^0.*

Assumption 1(ii) requires that we have a positive fraction of completely observed individuals and a positive fraction of completely observed periods. Under Assumption 1(ii) our results are also valid for dynamic panels, since $\{d_{it} : i \in [N]$ and $t \in [T]\}$ are treated as nonrandom and the error v_{it} has mean zero for those (i, t) with $d_{it} = 1$. In the following we list some important special cases of the block-type missing.

Example 4 (Exact block missing): *$d_{it} = 1$ for $i \leq N_o$ or $t \leq T_o$, and $d_{it} = 0$ for $i > N_o$ and $t > T_o$. Both N_o/N and T_o/T are bounded away from zero.*

[6] For the CCE approach, it may not be appropriate to use the cross-sectional averages as proxies for the factors. Suppose $x_{it} = \lambda_i^{x\prime} f_t^0 + v_{it}^x$, then $\bar{x}'_t = \sum_{i=1}^N d_{it}x'_{it}/\sum_{i=1}^N d_{it} \xrightarrow{p} f_t^{0\prime}\sum_{i=1}^N \mathbb{E}_\phi(d_{it})\lambda_i^x/\sum_{i=1}^N \mathbb{E}_\phi(d_{it})$ under some conditions. If $\mathbb{E}_\phi(d_{it})$ is correlated with λ_i^x and heterogenous over t, the coefficient $\sum_{i=1}^N \mathbb{E}_\phi(d_{it})\lambda_i^x/\sum_{i=1}^N \mathbb{E}_\phi(d_{it})$ may not be stable over t.

[7] $\mathbb{E}_\phi(d_{it}d_{i,t-1}) = \mathbb{E}_\phi(d^y_{it})\mathbb{E}_\phi[(d^y_{i,t-1})^2]\mathbb{E}_\phi(d^y_{i,t-2}) \neq \mathbb{E}_\phi(d^y_{it})[\mathbb{E}_\phi(d^y_{i,t-1})]^2\mathbb{E}_\phi(d^y_{i,t-2}) = \mathbb{E}_\phi(d_{it})\mathbb{E}_\phi(d_{i,t-1})$. But for $l \geq 2$, $\mathbb{E}_\phi(d_{it}d_{i,t-l}) = \mathbb{E}_\phi(d^y_{it}d^y_{i,t-1}d^y_{i,t-l}d^y_{i,t-l-1}) = \mathbb{E}_\phi(d^y_{it}d^y_{i,t-1})\mathbb{E}_\phi(d^y_{i,t-l}d^y_{i,t-l-1}) = \mathbb{E}_\phi(d_{it})\mathbb{E}_\phi(d_{i,t-l})$.

Example 5 (Mixed frequency/regular missing): *$d_{it} = 1$ for $i \leq N_o$ and all t. $d_{it} = 0$ if $i > N_o$ and t/h is not an integer, where N_o is the number of high frequency series and h is the frequency ratio (e.g., $h = 3$* for the case of mixed monthly-quarterly data).

Example 6 (Staggered missing): *$d_{it} = 1$ for $i \leq N_o$ and all t. $d_{it} = 1$ for $i > N_o$ and $t \leq T_{oi}$, where T_{oi} is the largest t with $d_{it} = 1$. $T_o = \min T_{oi}$ and both N_o/N and T_o/T are bounded away from zero. Here $\{T_{oi}, i \in [N]\}$ is considered as fixed. In comparison, in Athey and Imbens (2022) the design based analysis for the TWFE estimator under staggered adoption considers T_{oi} as a random variable. They show that the DID estimator is an unbiased estimator of a particular weighted average causal effect under random assignment of T_{oi}. It would be interesting to extend our results to the design-based settings.*

3 The Estimation Procedure

To estimate the regression coefficient vector β^0, we consider a two-step procedure where in the first step we construct a consistent initial estimator of β^0 and in the second step we use this estimator as the initial value to maximize the quasi log-likelihood function

$$L(\beta, \lambda, f) = -\frac{1}{2} \sum\nolimits_{i=1}^{N} \sum\nolimits_{t=1}^{T} d_{it}(y_{it} - x_{it}'\beta - f_t'\lambda_i)^2 \tag{3.1}$$

by alternating maximization between the regression coefficients and the factors and loadings. We shall prove the numerical convergence of this two-step procedure in three parts. In Part 1 we show that there exists a unique solution for $S(\gamma) = 0$ within the local neighborhood of the true parameters $\gamma^0 = (\beta^{0\prime}, \lambda^{0\prime}, f^{0\prime})'$, where $S(\cdot)$ is the score function. In Part 2 we show that within the local neighborhood of γ^0 the alternating maximization procedure is a contractionary mapping towards the solution of $S(\gamma) = 0$. In Part 3 we show that the nuclear norm regularized estimator is consistent and hence lies in the local neighborhood of γ^0 w.p.a.1. These three parts jointly imply that the two-step procedure is guaranteed to converge to the solution of $S(\gamma) = 0$. To present the details for these three parts, we make some assumptions.

Below, we use $M > 0$ to denote a generic constant that may vary across places.

Assumption 2 *(i) $\frac{1}{T}F^{0\prime}F^0 \xrightarrow{p} \Sigma_F > 0$ and $\|f_t^0\| \leq M$ for all t; (ii) $\frac{1}{N}\Lambda^{0\prime}\Lambda^0 \xrightarrow{p} \Sigma_\Lambda > 0$ and $\|\lambda_i^0\| \leq M$ for all i; (iii) The eigenvalues of the $r \times r$ matrix $\Sigma_F\Sigma_\Lambda$ are different.*

Assumption 2(i)-(ii) assume that the factors are pervasive and the factors and loadings are uniformly bounded, as in Bai and Li (2014), Ando and Bai (2020) and Chen et al. (2021a). In the matrix completion literature, the uniform boundedness of f_t^0 and λ_i^0 is referred to as the incoherence condition and helps to verify the restricted strong convexity condition for the nuclear norm regularized estimation; see, e.g., Negahban and Wainwright (2011), Negahban and Wainwright (2012) and Chernozhukov et al. (2023). Assumption 2(iii) is a standard condition for identifying the factors and loadings from the common components.

Assumption 3 *(i) (a)* $plim \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}x_{it}'$ *is positive definite (p.d.) and (b)* $W_x \equiv plim \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}\dot{x}_{it}\dot{x}_{it}'$ *is p.d., where* $\dot{x}_{it} = (\dot{x}_{it1}, ..., \dot{x}_{itK})'$ *and* $\dot{x}_{itk} = x_{itk} - (\delta_{ki}^{0\prime} f_t^0 + \omega_{kt}^{0\prime}\lambda_i^0)$*;*

(ii) $\mathbb{E}(\|x_{it}\|^{\varrho}) \leq M$ *for all* i *and* t *and some* $\varrho \geq 8$*;*

(iii) $\mathbb{E}(||\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[d_{it}x_{it} - \mathbb{E}_\phi(d_{it}x_{it})]f_t^{0\prime}||_F^2) \leq M$ *for all* i*,* $\mathbb{E}(||\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[d_{it}x_{it} - \mathbb{E}_\phi(d_{it}x_{it})]\lambda_i^{0\prime}||_F^2) \leq M$ *for all* t*; and* $\mathbb{E}(||\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[d_{it}x_{it}x_{it}' - \mathbb{E}_\phi(d_{it}x_{it}x_{it}')]||_F^2) = o(1)$*.*

The first part of Assumption 3(i) is a standard noncollinearity condition and is implied by the second part. We state it separately since it is already sufficient in some steps of the proof. The second part of Assumption 3(i) requires the regressors to be noncollinear after projecting out the true factors and true loadings, where the projection is as defined in (2.2).[8] The second part of Assumption 3(i) is crucial for both the local identification of β^0 and the numerical convergence of the alternating maximization within the local neighborhood of the true parameters. Note that W_x can be p.d. even if x_{it} contains some low-rank regressors. Nevertheless, Assumption 3(i) is not enough for the global identification in the presence of low-rank regressors; see Appendix S.3 of Moon and Weidner (2017) for detailed explanations when there is no missing data. To achieve global identification, we use the nuclear norm regularized estimation to obtain an initial consistent estimator, which is a popular approach in the literature on low-rank estimation.

Assumption 3(ii) requires x_{it} to have bounded moments. Assumption 3(iii) requires that $d_{it}x_{it}$ and $d_{it}x_{it}x_{it}'$ be weakly dependent across i and t conditioning on ϕ^0. If d_{it} is nonrandom (see Assumption 1(ii)), we only need x_{it} to be weakly dependent across i and t conditioning on ϕ^0, which is also assumed in Assumption 5(iii)-(iv) of Moon and Weidner (2017) for dynamic panels.

Assumption 4 *(i)* $\mathbb{E}(|v_{it}|^{\zeta}) \leq M$ *for all* i *and* t *and some* $\zeta > 8$*;*

(ii) $\max_s \gamma_N(s,s) \leq M$ *and* $\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_N(s,t)| \leq M$ *where* $\gamma_N(s,t) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(d_{is}v_{is}d_{it}v_{it})$*;*

(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(i) requires v_{it} to have bounded ζ-th order moments. Assumption 4(ii)-(iii) imposes weak dependence condition on $\{d_{it}v_{it}\}$ along the time and cross-sectional dimensions. These conditions reduce to Assumption C in Bai (2003) when $d_{it} = 1$ for all i and t.

[8] When d_{it} is treated as nonrandom, we have $\mathbb{E}_\phi(d_{it}) = d_{it}$ and $\mathbb{E}_\phi(d_{it}x_{itk}) = d_{it}\mathbb{E}_\phi(x_{itk})$, then δ_{ki}^0 and ω_{kt}^0 are the solution of $\min_{\{\delta_{ki},\omega_{kt}\}}\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}(\mathbb{E}_\phi(x_{itk}) - \delta_{ki}' f_t^0 - \omega_{kt}'\lambda_i^0)^2$.

Assumption 5 *(i)* $\mathbb{E}(\frac{1}{NT}\sum_{t,s=1}^{T}\sum_{i=1}^{N}|\mathbb{E}_\phi(d_{it}v_{it}d_{is}v_{is})|) \le M$ *and* $\mathbb{E}(\frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t=1}^{T}|\mathbb{E}_\phi(d_{it}v_{it}d_{jt}v_{jt})|) \le M$*;*

(ii) $\mathbb{E}(v_{it}x_{it}d_{it}) = 0$ *and* $\mathbb{E}(||\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}d_{it}v_{it}||^2) \le M$*.*

Assumption 5(i) imposes additional weak dependence conditions on $\{d_{it}v_{it}\}$ along the time and cross-sectional dimensions. These conditions reduce to Assumption D Bai (2003) when $d_{it} = 1$ for all i and t. Assumption 5(ii) allows x_{it} to be either strictly exogenous or weakly exogenous as in dynamic panels, but we do not need to assume v_{it} to be independent across (i,t). In contrast, Moon and Weidner (2017) directly assumes that v_{it} is independent across (i,t) in dynamic panels.

3.1 Existence of the Local Critical Point

The criterion function for the estimation is $Q(\beta,\lambda,f) = L(\beta,\lambda,f) + G(\lambda,f)$, where

$$G(\lambda,f) = -\frac{c}{2}NT\sum\nolimits_{p=1}^{r}\sum\nolimits_{q=1}^{r}(\frac{1}{N}\sum\nolimits_{i=1}^{N}\lambda_{iq}^0\lambda_{ip} - \frac{1}{T}\sum\nolimits_{t=1}^{T}f_{tp}^0 f_{tq})^2, \tag{3.2}$$

c is an arbitrary positive constant and $L(\beta,\lambda,f)$ is the quasi log-likelihood function defined in (3.1). Note that the solution of maximizing $L(\beta,\lambda,f)$ is not unique since for any (λ,f) and any $r\times r$ invertible matrix Π, we have $f_t'\lambda_i = (\Pi' f_t)'\Pi^{-1}\lambda_i$ for all i and t. To ensure the uniqueness of the solution, we add the penalty $G(\lambda,f)$ to $L(\beta,\lambda,f)$ so that maximizing $Q(\beta,\lambda,f)$ is equivalent to picking up the solution that satisfies $\frac{1}{N}\sum_{i=1}^{N}\lambda_{iq}^0\lambda_{ip} = \frac{1}{T}\sum_{t=1}^{T}f_{tp}^0 f_{tq}$ for all $p,q\in[r]$ from the many solutions of maximizing $L(\beta,\lambda,f)$. Such a set of normalization conditions impose exactly r^2 restrictions for the identification of (λ^0,f^0) despite it is somewhat infeasible.[9] We assume that the true parameters (λ^0,f^0) satisfy these restrictions. If not, we can redefine the true parameters without changing the product $f_t^{0\prime}\lambda_i^0$ for all i and t.

Recall that $\phi = (\lambda', f')'$ and $\gamma = (\beta',\lambda',f')'$. Let $S(\gamma) = \partial_\gamma Q(\gamma)$ denote the score function of $Q(\gamma)$ and $Q_{\gamma\gamma'}(\gamma) = \partial_{\gamma\gamma'}Q(\gamma)$ denote the Hessian. Specifically, we introduce the following notations:

$$\begin{aligned}
S_\beta(\gamma) &= \partial_\beta Q(\gamma),\ S_{\lambda_i}(\gamma) = \partial_{\lambda_i}Q(\gamma),\ S_{f_t}(\gamma) = \partial_{f_t}Q(\gamma),\\
S_\lambda(\gamma) &= (S_{\lambda_1}(\gamma)',...,S_{\lambda_N}(\gamma)')',\ S_f(\gamma) = (S_{f_1}(\gamma)',...,S_{f_T}(\gamma)')',\\
S_\phi(\gamma) &= (S_\lambda(\gamma)', S_f(\gamma)')',\ S(\gamma) = (S_\beta(\gamma)', S_\phi(\gamma)')',\\
Q_{\beta\beta'}(\gamma) &= \partial_{\beta\beta'}Q(\gamma),\ Q_{\beta\phi'}(\gamma) = \partial_{\beta\phi'}Q(\gamma), \text{ and } Q_{\phi\phi'}(\gamma) = \partial_{\phi\phi'}Q(\gamma).
\end{aligned}$$

We suppress the argument when the score and the Hessian are evaluated at the true parameter

[9] In matrix form, the restrictions can be written as $\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\lambda_i' = \frac{1}{T}\sum_{t=1}^{T}f_t f_t^{0\prime}$. This type of restrictions is inspired by Chen et al. (2021b). Compared to the restrictions in Su and Wang (2024), the benefit of this type of restrictions is a simpler Hessian structure, while the cost is that it is computationally infeasible since f_t^0 and λ_i^0 are unobservable. See Section 3.2 for the computation details.

values, i.e., $S_\phi = S_\phi(\gamma^0)$, $Q_{\phi\phi'} = Q_{\phi\phi'}(\gamma^0)$, and so on. We shall show that $S(\gamma) = 0$ has a unique solution in the interior of the local region $\mathcal{B}_m(\gamma^0)$ of γ^0, where

$$\mathcal{B}_m(\gamma^0) = \{\gamma \in \mathbb{R}^{K+Nr+Tr} : \left\|\beta - \beta^0\right\| \leq m, \ \frac{1}{\sqrt{N}} \left\|\lambda - \lambda^0\right\| \leq m \text{ and } \frac{1}{\sqrt{T}} \left\|f - f^0\right\| \leq m\} \tag{3.3}$$

and $m > 0$ is a small but fixed constant. $\mathcal{B}_m(\gamma^0)$ is large enough to contain the initial consistent estimator without imposing any requirements on the convergence rate of the initial estimator. If $\left\|\hat{\beta}^{(0)} - \beta^0\right\| = o_p(1)$, $\left\|\hat{\lambda}^{(0)} - \lambda^0\right\| = o_p(\sqrt{N})$ and $\left\|\hat{f}^{(0)} - f^0\right\| = o_p(\sqrt{T})$, then obviously $(\hat{\beta}^{(0)\prime}, \hat{\lambda}^{(0)\prime}, \hat{f}^{(0)\prime})' \in \mathcal{B}_m(\gamma^0)$ w.p.a.1. In addition, $\mathcal{B}_m(\gamma^0)$ is small enough to guarantee that the Hessian matrix of $Q(\gamma)$ is well-behaved inside this region, as shown in the following proposition.

Proposition 3.1 *Let* $D_{NT} = \begin{bmatrix} N \times I_{Nr} & 0 \\ 0 & T \times I_{Tr} \end{bmatrix}$ *and* $D_{TN} = \begin{bmatrix} T \times I_{Nr} & 0 \\ 0 & N \times I_{Tr} \end{bmatrix}$. *Under Assumptions 1–4, there exist* $m > 0$ *and* $C > 0$ *such that as* $(N, T) \to \infty$,

(i) $\min_{\gamma \in \mathcal{B}_m(\gamma^0)} \sigma_{\min}(-D_{TN}^{-\frac{1}{2}} Q_{\phi\phi'}(\gamma) D_{TN}^{-\frac{1}{2}}) \geq C$ *w.p.a.1,*

(ii) $\min_{\gamma \in \mathcal{B}_m(\gamma^0)} \sigma_{\min}(-\frac{1}{NT}[Q_{\beta\beta'}(\gamma) - Q_{\beta\phi'}(\gamma)(Q_{\phi\phi'}(\gamma))^{-1} Q_{\phi\beta'}(\gamma)]) \geq C.$

Proposition 3.1 shows that the normalized Hessian matrix is negative definite uniformly within $\mathcal{B}_m(\gamma^0)$. This result could also be generalized to other missing patterns following a similar proof of Proposition 3.1. D_{NT} and D_{TN} are designed such that the eigenvalues of $-D_{TN}^{-\frac{1}{2}} Q_{\phi\phi'}(\gamma) D_{TN}^{-\frac{1}{2}}$ have the same asymptotic order even when N and T tend to infinity at different rates. Based on Proposition 3.1, we can prove the following important theorem.

Theorem 3.1 *Under Assumptions 1–5, there exists a unique solution for* $S(\gamma) = 0$ *in the interior of* $\mathcal{B}_m(\gamma^0)$. *Let* $\hat{\gamma} = (\hat{\beta}', \hat{\phi}')' = (\hat{\beta}', \hat{\lambda}', \hat{f}')'$ *denote this solution. Then we also have* $\left\|\hat{\beta} - \beta^0\right\| = O_p(\frac{1}{c_{NT}})$, $\left\|\frac{1}{\sqrt{N}}(\hat{\lambda} - \lambda^0)\right\| = O_p(\frac{1}{c_{NT}})$ *and* $\left\|\frac{1}{\sqrt{T}}(\hat{f} - f^0)\right\| = O_p(\frac{1}{c_{NT}})$, *where recall that* $c_{NT} = N^{\frac{1}{2}} \wedge T^{\frac{1}{2}}$.

Given Theorem 3.1, we define $\hat{\gamma}$ as our estimator for $\gamma^0 = (\beta^{0\prime}, \lambda^{0\prime}, f^{0\prime})'$. The condition $S(\hat{\gamma}) = 0$ allows us to take Taylor expansions for the asymptotic analysis in Section 4, and is also crucial for proving that $\hat{\gamma}$ is the convergence target of the alternating maximization procedure. Since Proposition 3.1 implies that $Q_{\gamma\gamma'}(\hat{\gamma})$ is negative definite, $\hat{\gamma}$ is also the local maximum in $\mathcal{B}_m(\gamma^0)$. Note that in general the local maximum could be on the boundary of the local region, Theorem 3.1 implies that the local maximum in $\mathcal{B}_m(\gamma^0)$ is an interior point.

3.2 Convergence of the Alternating Maximization Procedures

In this subsection we show that the alternating maximization algorithm is a contractionary mapping towards $\hat{\gamma}$ within $\mathcal{B}_m(\gamma^0)$. First, we present the algorithm as follows.

Algorithm 1

1. *Obtain an initial consistent estimate* $(\hat{\beta}^{(0)\prime}, \hat{\phi}^{(0)\prime})'$. *Let* $\tilde{\phi}^{(0)} = \hat{\phi}^{(0)}$.

2. *At step* $k \geq 0$, *given* $\hat{\beta}^{(k)}$, *use* $\tilde{\phi}^{(k)}$ *as the initial value and the EM algorithm to calculate* $\tilde{\phi}^{(k+1)} \equiv (\tilde{\lambda}^{(k+1)\prime}, \tilde{f}^{(k+1)\prime})' = \arg\max_{(\lambda,f)} L(\hat{\beta}^{(k)}, \lambda, f)$. *More specifically, for* $h = 0, 1, 2, ...$, *let* $\tilde{f}_t^{(k,0)} = \tilde{f}_t^{(k)}$ *and* $\tilde{\lambda}_i^{(k,0)} = \tilde{\lambda}_i^{(k)}$, *and let* $y_{it}^{(k,h)} = y_{it} - x_{it}'\hat{\beta}^{(k)}$ *if* $d_{it} = 1$, $y_{it}^{(k,h)} = \tilde{f}_t^{(k,h)\prime}\tilde{\lambda}_i^{(k,h)}$ *if* $d_{it} = 0$. *Let* $y^{(k,h)}$ *denote the* $T \times N$ *matrix with* $y_{it}^{(k,h)}$ *as the* (t,i)*-th element. Then* $\tilde{F}^{(k,h+1)} = (\tilde{f}_1^{(k,h+1)}, ..., \tilde{f}_T^{(k,h+1)})'$ *are* $\sqrt{T}$ *times the eigenvectors of* $y^{(k,h)}y^{(k,h)\prime}$ *corresponding to the* r*th largest eigenvalues and* $\tilde{\Lambda}^{(k,h+1)} = (\tilde{\lambda}_1^{(k,h+1)}, ..., \tilde{\lambda}_N^{(k,h+1)})' = \frac{1}{T}y^{(k,h)\prime}\tilde{F}^{(k,h+1)}$. *Iterate* $h = 0, 1, 2, ...$ *until convergence.*

3. *Given* $\tilde{\phi}^{(k+1)}$, *calculate* $\hat{\beta}^{(k+1)} = \arg\max_\beta L(\beta, \tilde{\lambda}^{(k+1)}, \tilde{f}^{(k+1)})$. *The solution is* $\hat{\beta}^{(k+1)} = (\sum_{i=1}^N \sum_{t=1}^T d_{it}x_{it}x_{it}')^{-1} \sum_{i=1}^N \sum_{t=1}^T d_{it}(y_{it} - \tilde{f}_t^{(k+1)\prime}\tilde{\lambda}_i^{(k+1)})$.

4. *Iterate between step 2 and step 3 until convergence and obtain the estimator* $(\hat{\beta}, \tilde{\phi})$.

Remark 3.1 *In Algorithm 1,* k *is the index for the outer loop while* h *is the index for the inner loop that carries out the EM. The difference between Algorithm 1 and the algorithm in Appendix B of Bai (2009) for unbalanced panels is that Algorithm 1 starts from consistent initial estimates while that in Bai (2009) starts from random initial values. This difference is crucial for the convergence analysis. In fact, Bai (2009) did not provide any convergence analysis or study the asymptotic properties of his estimator for unbalanced panels.*

Remark 3.2 *In Algorithm 1, the EM algorithm for calculating* $\tilde{\phi}^{(k+1)}$ *given* $(\hat{\beta}^{(k)}, \tilde{\phi}^{(k)})$ *can be replaced by a gradient descent algorithm or some other algorithm to handle ultra large data sets, where even singular value decomposition is slow.*

Remark 3.3 *Given* $\hat{\beta}^{(k)}$, *since* $L(\hat{\beta}^{(k)}, \lambda, f) = -\frac{1}{2}\sum_{i=1}^N \sum_{t=1}^T d_{it}(y_{it} - x_{it}'\hat{\beta}^{(k)} - f_t'\lambda_i)^2$ *only depends on the product* $f_t'\lambda_i$ *and* $f_t'\lambda_i = (\Pi' f_t)'\Pi^{-1}\lambda_i$ *for any invertible* $r \times r$ *matrix* Π, *the solution of maximizing* $L(\hat{\beta}^{(k)}, \lambda, f)$ *is not unique. Among all the solutions of maximizing* $L(\hat{\beta}^{(k)}, \lambda, f)$, *there is only one solution such that* $G(\lambda, f) = 0$, *denote this solution as*

$$\hat{\phi}^{(k+1)} = (\hat{\lambda}^{(k+1)\prime}, \hat{f}^{(k+1)\prime})' = \arg\max_{\lambda,f} L(\hat{\beta}^{(k)}, \lambda, f) + G(\lambda, f). \tag{3.4}$$

Since the objective function for the alternating maximization is $Q(\beta, \lambda, f) = L(\beta, \lambda, f) + G(\lambda, f)$ *in Step 2 of Algorithm 1, we are supposed to compute* $(\hat{\lambda}^{(k+1)}, \hat{f}^{(k+1)})$ *rather than* $(\tilde{\lambda}^{(k+1)}, \tilde{f}^{(k+1)})$. *However, this does not affect the computation of* $\hat{\beta}^{(k+1)}$ *in Step (3) of Algorithm 1 since* $\tilde{f}_t^{(k+1)\prime}\tilde{\lambda}_i^{(k+1)} =$

$\hat{f}_t^{(k+1)\prime}\hat{\lambda}_i^{(k+1)}$ *for all* (t,i)*. In summary, algorithm 1 produces* $(\hat{\beta},\tilde{\phi})$ *while the asymptotic analysis focuses on* $(\hat{\beta},\hat{\phi})$*.*

An important observation is that the alternating maximization in Algorithm 1 is a block Gauss-Seidel procedure and the Hessian is negative definite, symmetric and approximately constant within $\mathcal{B}_m(\gamma^0)$.[10] This observation allows us to prove the following theorem.

Theorem 3.2 *Suppose that Assumptions 1-5 hold and* $(\hat{\beta}^{(0)\prime},\hat{\phi}^{(0)\prime})' \in \mathcal{B}_m(\gamma^0)$*. Then* $\left\|\hat{\beta}^{(k+1)}-\hat{\beta}\right\| < \psi\left\|\hat{\beta}^{(k)}-\hat{\beta}\right\|$ *for some* $0<\psi<1$*, implying that* $\left\|\hat{\beta}^{(k)}-\hat{\beta}\right\| \to 0$ *and* $\left\|\hat{\phi}^{(k)}-\hat{\phi}\right\| \to 0$ *as* $k\to\infty$*.*

Theorem 3.2 shows that if the initial estimate $(\hat{\beta}^{(0)\prime},\hat{\phi}^{(0)\prime})'$ lie in the region $\mathcal{B}_m(\gamma^0)$, then $\left\|\hat{\beta}^{(k)}-\hat{\beta}\right\|$ converges to zero with a linear speed.[11] Note that the convergence of $(\hat{\beta}^{(k)\prime},\hat{\phi}^{(k)\prime})'$ does not rely on the convergence rate of $(\hat{\beta}^{(0)},\hat{\phi}^{(0)})$. This gives us plenty of freedom in constructing the initial estimator, since we can put aside the efficiency and inference concerns and focus on the consistency and computational convenience. For example, we can use a small but balanced part of the unbalanced panels to obtain the initial consistent estimates as in Bai and Ng (2021) for the case of pure factor models.

The proof of Theorem 3.2 applies in spirit to other settings such as nonlinear panels, two-way fixed effects models and other missing patterns. All we need to do is to verify the local concavity of the Hessian matrix in these settings. For example, the Hessian of nonlinear panel with two-way additive or interactive fixed effects is locally concave when there is no missing data, so we just need to verify this is still true under the missing patterns of Assumption 1 or other patterns of interest. Similarly, Theorem 3.1 also can be readily extended to these settings.

The convergence of the Gauss-Seidel procedure in the current context suggests that the Jacobi procedure should also converge. The details are as follows.

Algorithm 2

1. *Obtain an initial consistent estimate* $(\hat{\beta}^{(0)\prime},\hat{\phi}^{(0)\prime})'$*. Let* $\tilde{\phi}^{(0)}=\hat{\phi}^{(0)}$*.*

2. *At step* $k\geq 0$*, given* $\hat{\beta}^{(k)}$*, use* $\tilde{\phi}^{(k)}$ *as the initial value and the EM algorithm to calculate* $\tilde{\phi}^{(k+1)}=(\tilde{\lambda}^{(k+1)\prime},\tilde{f}^{(k+1)\prime})'=\arg\max_{(\lambda,f)}L(\hat{\beta}^{(k)},\lambda,f)$*.*

3. *Given* $\tilde{\phi}^{(k)}$*, obtain* $\hat{\beta}^{(k+1)}=\arg\max_{\beta}L(\beta,\tilde{\lambda}^{(k)},\tilde{f}^{(k)})$*. The solution is* $\hat{\beta}^{(k+1)}=(\sum_{i=1}^N\sum_{t=1}^T d_{it}x_{it}x_{it}')^{-1}\sum_{i=1}^N\sum_{t=1}^T d_{it}(y_{it}-\tilde{f}_t^{(k)\prime}\tilde{\lambda}_i^{(k)})$*.*

[10] See Golub and Van Loan (2013) and Hackbusch (2016) for textbook introductions for the Gauss-Seidel procedure.
[11] Iterative estimation is very common in the machine learning literature. A good initial value is sometimes called a warm start, which is crucial for large scale nonconvex optimization.

4. Repeat step 2 and step 3 until $\hat{\beta}^{(k)}$ and $\tilde{\phi}^{(k)}$ converge.

The difference between Algorithm 1 and Algorithm 2 lies in the third step, where the former uses $\tilde{\phi}^{(k+1)}$ to calculate $\hat{\beta}^{(k+1)}$ while the latter uses $\tilde{\phi}^{(k)}$. As explained in Remark 3.3, the alternating maximization produces $(\hat{\beta}', \tilde{\phi}')'$ while the asymptotic analysis focuses on $(\hat{\beta}', \hat{\phi}')'$. Here, $(\hat{\beta}', \hat{\phi}')'$ is the convergence target of $(\hat{\beta}^{(k+1)\prime}, \hat{\phi}^{(k+1)\prime})'$:

$$\hat{\phi}^{(k+1)} = \arg\max_{\beta} L(\hat{\beta}^{(k)}, \phi) + G(\phi) \text{ and } \hat{\beta}^{(k+1)} = \arg\max_{\beta} L(\beta, \hat{\phi}^{(k)}). \tag{3.5}$$

Corollary 3.1 *Suppose Assumptions 1-5 hold and $(\hat{\beta}^{(0)\prime}, \hat{\phi}^{(0)\prime})' \in \mathcal{B}_m(\gamma^0)$. Then $(\hat{\beta}^{(k)\prime}, \hat{\phi}^{(k)\prime})'$ as defined in (3.5) converges to $(\hat{\beta}', \hat{\phi}')'$ as $k \to \infty$ in the sense $\left\|\hat{\beta}^{(k)} - \hat{\beta}\right\| \to 0$ and $\left\|\hat{\phi}^{(k)} - \hat{\phi}\right\| \to 0$.*

Similarly, other variants of alternating maximization such as the Richardson iteration and the successive over-relaxation also can be developed to accelerate the convergence speed or implement parallel computation. We may also iterate between β, λ and f by dividing the Hessian matrix into 3×3 blocks. These extensions are valuable for ultra large data sets and is left for future research.

3.3 The Initial Consistent Estimator

In this subsection we present the details of the nuclear norm regularized (NNR) approach for consistent initial estimation. Since the 2018 working paper version of Moon and Weidner (2023) introduced the NNR approach to the econometrics literature, similar regularizations have been used in many other contexts such as network structures, panel quantile regressions, grouped fixed effects panels, panel threshold models, high-dimensional VAR and conditional factor models; see the references in Moon and Weidner (2023). The popularity is mainly due to the global convexity of the NNR objective function. As emphasized in Moon and Weidner (2023), another advantage of the NNR approach is that it allows for low-rank regressors. In this paper we propose to apply the NNR approach to obtain the initial estimate for the unbalanced panels. It is useful to emphasize that the NNR approach is just one choice for consistent initial estimation. In practice, another potential choice is to obtain initial estimates from a small but balanced part of the unbalanced panel.

For the random missing case (Assumption 1(i)), we obtain the NNR estimator as

$$(\hat{\beta}^{(0)}, \hat{\Theta}^{(0)}) = \arg\min_{\beta,\Theta} \frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it}(y_{it} - \Theta_{it} - x_{it}'\beta)^2 + \nu_{NT} \left\|\Theta\right\|_*, \tag{3.6}$$

where ν_{NT} is a tuning parameter.[12] We choose to use the ADMM algorithm to solve this optimiza-

[12] The tuning parameter ν_{NT} is a critical component in model optimization, as it directly induces the low-rank

tion problem; see Appendix H in the online supplement for details. With $\hat{\Theta}^{(0)}$, let $\hat{\mathcal{U}}^{(0)}$, $\hat{\mathcal{V}}^{(0)}$ and $\hat{\Sigma}^{(0)}$ denote the first r left singular vectors, the first r right singular vectors and the $r \times r$ diagonal matrix that contains first r singular values of $\hat{\Theta}^{(0)}$ ordered in descending order along its diagonal line, respectively. Then we define $\hat{f}^{(0)} = \sqrt{T}\hat{\mathcal{U}}^{(0)}$ and $\hat{\lambda}^{(0)} = \sqrt{N}\hat{\mathcal{V}}^{(0)}\hat{\Sigma}^{(0)}$. The number of factors r is estimated via singular value thresholding:

$$\hat{r} = \sum_{s=1}^{N \wedge T} \mathbf{1}\left\{\sigma_s(\frac{\hat{\Theta}^{(0)}}{\sqrt{NT}}) \geq c_f \sqrt{c_{NT}^{-1/4}\sigma_1(\frac{\hat{\Theta}^{(0)}}{\sqrt{NT}})}\right\}, \tag{3.7}$$

where $\sigma_s(A)$ denotes the s-th largest singular value of A and $c_f > 0$. In the simulation, we set $c_f = 0.6$.

For the block missing case, we run the NNR estimation on the complete data block, e.g., the data block with $i \in [N_o]$ and $t \in [T]$, to obtain $\hat{r}$, $\hat{\beta}$, $\{\hat{f}_t^{(0)}, t \in [T]\}$ and $\{\hat{\lambda}_i^{(0)}, i \in [N_o]\}$. Then for each $i = N_o + 1, ..., N$, we regress $y_{it} - x_{it}'\hat{\beta}$ on $\hat{f}_t^{(0)}$ to obtain $\hat{\lambda}_i^{(0)}$. Let $\hat{F}^{(0)} = (\hat{f}_1^{(0)}, ..., \hat{f}_T^{(0)})'$ and $\hat{\Lambda}^{(0)} = (\hat{\lambda}_1^{(0)}, ..., \hat{\lambda}_N^{(0)})'$. Then we normalize $\hat{F}^{(0)}$ and $\hat{\Lambda}^{(0)}$ such that $\hat{F}^{(0)\prime}\hat{F}^{(0)}/T = \hat{\Lambda}^{(0)\prime}\hat{\Lambda}^{(0)}/N$ and both are diagonal.

To show the consistency of the NNR estimators, we define the following restricted set

$$\mathcal{R} = \{\Delta_\Theta \in \mathbb{R}^{T \times N} : \|\Delta_\Theta\|_{\max} \leq M \text{ and } \left\|\mathcal{P}^{\perp}(\Delta_\Theta)\right\|_* \leq 3\,\|\mathcal{P}(\Delta_\Theta)\|_*\}, \tag{3.8}$$

where $\Delta_\Theta = \Theta - \Theta^0$ for any Θ and $\mathcal{P}^{\perp}(\Delta_\Theta)$ and $\mathcal{P}(\Delta_\Theta)$ are defined as follows. Let $\Theta^0 = \mathcal{U}^0\Sigma^0\mathcal{V}^{0\prime}$ be the singular value decomposition of Θ^0 and decompose $\mathcal{U}^0 = (\mathcal{U}_r, \mathcal{U}_0)$ and $\mathcal{V}^0 = (\mathcal{V}_r, \mathcal{V}_0)$ with $(\mathcal{U}_r, \mathcal{V}_r)$ being the singular vectors corresponding to nonzero singular values and $(\mathcal{U}_0, \mathcal{V}_0)$ being the singular vectors corresponding to zero singular values. For any matrix $A \in \mathbb{R}^{T \times N}$, we define $\mathcal{P}^{\perp}(A) = \mathcal{U}_0\mathcal{U}_0' A \mathcal{V}_0\mathcal{V}_0'$ and $\mathcal{P}(A) = A - \mathcal{P}^{\perp}(A)$, i.e., $\mathcal{P}(A)$ can be seen as the linear projection of A onto the low-rank space with $\mathcal{P}^{\perp}(A)$ being its orthogonal space. Intuitively, the second restriction in (3.8) means the projection onto the high rank space can be controlled by the projection onto the low-rank space.

Let $M_{vec(x)} = I - P_{vec(x)}$ and $P_{vec(x)}$ denote the projection matrix of $[vec(x_1), ..., vec(x_K)]$, where $vec(x_k)$ is the $TN \times 1$ vector that vectorizes x_k. We add the following assumption.

Assumption 6 *(i) For the random missing, (a) d_{it} is independent across (i,t) conditioning on $(\phi^0,\ x_{it})$, $\min_{i,t} \mathbb{E}_{\phi x}(d_{it}) \geq \underline{d} > 0$ where $\mathbb{E}_{\phi x}(\cdot)$ the expectation conditioning on ϕ^0 and all x_{it}, (b) there exists $\mu > 0$, independent of (N,T), such that $vec(\Delta_\Theta)' M_{vec(x)} vec(\Delta_\Theta) \geq \mu \|\Delta_\Theta\|_F^2$ for any $\Delta_\Theta \in \mathcal{R}$ when N and T are sufficiently large.*

structure in the solution. In the simulation, we set $\nu_{NT} = c\sqrt{\max(N,T)}$ with $c \in \{10^{-2}, 10^{-1}, 1, 10\}$ and we choose the one giving the minimum value of objective function $L(\beta, \lambda, f)$.

(ii) For the block missing, Assumption 1(ii) holds and Assumptions 2, 3(i)a, 5(ii) and 6(i)b hold for at least one completely observed data block.

Here we need the conditional independence of d_{it} to verify the restricted strong convexity (RSC) condition as defined in Negahban and Wainwright (2011), Negahban and Wainwright (2012) and Klopp (2014). Assumption 6(i)b requires that the Euclidean norm of $vec(\Delta_\Theta)$ after projecting on $vec(x)$ be a positive proportion of its Euclidean norm before projection, if we restrict Δ_Θ to belong to $\mathcal{R}$. The same assumption also appears in Moon and Weidner (2023) and Mugnier (2024) and plays a similar role as the restricted eigenvalue condition common in the LASSO literature (e.g., Bickel et al. (2009)). Assumption 6(ii) presents the conditions for the consistency of the NNR estimator when applied to a balanced block of the unbalanced panel. Given these assumptions, the following theorem proves the consistency of $\hat{r}$, $\hat{\beta}^{(0)}$, $\hat{\lambda}^{(0)}$ and $\hat{f}^{(0)}$, which guarantees that $\hat{\gamma}^{(0)} = (\hat{\beta}^{(0)\prime}, \hat{\lambda}^{(0)\prime}, \hat{f}^{(0)\prime})'$ falls in $\mathcal{B}_m(\gamma^0)$ w.p.a.1.

Theorem 3.3 *Suppose that Assumptions 2, 3(i)-(ii), 4, 5(ii) and 6. Suppose that $\nu_{NT} = 2c_1(N^{1/2}T^{1/4} + N^{1/4}T^{1/2})$ for some constant $c_1 > 0$. Then as $(N,T) \to \infty$, $\mathbb{P}(\hat{r} = r) \to 1$, $\left\|\hat{\beta}^{(0)} - \beta^0\right\| = o_p(1)$, $N^{-1/2}\left\|\hat{\lambda}^{(0)} - \lambda^0\right\| = o_p(1)$ and $T^{-1/2}\left\|\hat{f}^{(0)} - f^0\right\| = o_p(1)$.*

The side condition on the tuning parameter ν_{NT} in Theorem 3.3 is sufficient but not necessary. Under the weak condition in Assumption 4, we show in Lemma D.1 that $\|d \circ v\| = O_p(N^{\frac{1}{2}}T^{\frac{1}{4}} + N^{\frac{1}{4}}T^{\frac{1}{2}})$, where $d \circ v$ is a $T \times N$ matrix with $d_{it}v_{it}$ as the (t,i)-th element. If we impose some stronger conditions than those in Assumption 4, this rate can be improved to $O_p(N^{\frac{1}{2}} + T^{\frac{1}{2}})$ as in Latała (2005) if one assumes some (conditional) independence conditions along both the individual and time dimensions or $O_p(N^{\frac{1}{2}} + (T\log T)^{\frac{1}{2}})$ as in Wang et al. (2022) if one assumes (conditional) independence along the individual dimension and (conditional) strong mixing along the time dimension. With such a better control on $\|d \circ v\|$, one can choose a smaller rate for ν_{NT} (as small as $2c_1\left(N^{1/2} + T^{1/2}\right)$) which will yield a better control on the rates at which $\left\|\hat{\beta}^{(0)} - \beta^0\right\|$, $N^{-1/2}\left\|\hat{\lambda}^{(0)} - \lambda^0\right\|$ and $T^{-1/2}\left\|\hat{f}^{(0)} - f^0\right\|$ converge to 0 in probability. Since we only aim at obtaining a consistent initial estimator, the choice of $\nu_{NT} = 2c_1\left(N^{1/2}T^{1/4} + N^{1/4}T^{1/2}\right)$ is sufficient.

4 Limit Distribution and Bias Correction

Theorem 3.1 shows $\left\|\hat{\beta} - \beta^0\right\| = O_p(\frac{1}{c_{NT}})$, $\left\|\hat{\lambda} - \lambda^0\right\| = O_p(\frac{\sqrt{N}}{c_{NT}})$ and $\left\|\hat{f} - f^0\right\| = O_p(\frac{\sqrt{T}}{c_{NT}})$. In this section we take higher order Taylor expansions of the first order conditions to refine the asymptotic expansions and to derive the asymptotic distribution and bias of $\hat{\beta}$:

$$0 \;=\; S_\beta(\hat{\beta}, \hat{\phi}) = S_\beta + Q_{\beta\beta'}(\hat{\beta} - \beta^0) + Q_{\beta\phi'}(\hat{\phi} - \phi^0) + R_\beta, \tag{4.1}$$

$$0 \quad = \quad S_\phi(\hat\beta, \hat\phi) = S_\phi + Q_{\phi\beta'}(\hat\beta - \beta^0) + Q_{\phi\phi'}(\hat\phi - \phi^0) + R_\phi. \tag{4.2}$$

Here, R_β (of dimension $K \times 1$) and R_ϕ (of dimension $(Nr + Tr) \times 1$) are the remainders in the expansions; see Lemma E.4 in the online supplement for their detailed expressions. Equations (4.1)-(4.2) provide a general framework for the asymptotic analyses of fixed effects estimation of large dimensional panels. Linear/nonlinear panels with one way/two way/interactive fixed effects are all covered by this framework, with each model corresponding to a specific structure on the Hessians and the remainders.[13] Moreover, equations (4.1)-(4.2) also cover these panel data models with missing data. We show in the online appendix that the upper bounds of $\|R_\beta\|$ and $\|R_\phi\|$ are not affected by any missing patterns and $Q_{\phi\phi'}$ is still negative definite under missing patterns of Assumption 1, but the asymptotic variances and biases of $\hat\beta$ do depend on the missing patterns. When there is no missing data, the most general framework for fixed effects estimation of large dimensional panels is Fernández-Val and Weidner (2016). Nevertheless, it is unclear how to extend their results to cases with missing patterns of Assumption 1. In the following we present the details of our asymptotic analyses.

4.1 Limit Distribution of the Local Critical Point

To study the limit distribution of our estimator, we add two assumptions.

Assumption 7 *(i)* $\max_{h\in[T]} \sum_{i=1}^N \sum_{t,s=1}^T \left|\mathbb{E}_\phi(d_{it}v_{it}d_{is}v_{is}d_{ih}v_{ih}^2)\right| \le MNT$ *a.s.,* $\max_{h\in[T]} \sum_{i,j=1}^N \sum_{t,s=1}^T |\mathbb{E}_\phi(d_{it}v_{it}d_{ih}v_{ih}d_{js}v_{js}d_{jh}v_{jh})| \le M(NT + N^2)$ *a.s.,* $\max_{k\in[N]} \sum_{t=1}^T \sum_{i,j=1}^N |\mathbb{E}_\phi(d_{kt}v_{kt}^2 d_{it}v_{it}d_{jt}v_{jt})| \le MNT$ *a.s., and* $\max_{k\in[N]} \sum_{t,s=1}^T \sum_{i,j=1}^N |\mathbb{E}_\phi(d_{kt}v_{kt}d_{it}v_{it}d_{ks}v_{ks}d_{js}v_{js})| \le M(NT + T^2)$ *a.s.;*

(ii) $\sum_{i,j=1}^N \sum_{t,s=1}^T |\mathbb{E}_\phi(d_{it}v_{it}d_{js}v_{js})| \le MNT$ *a.s.;*

(iii) $\max_{h\in[T]} \sum_{i=1}^N \sum_{t,s=1}^T |\mathbb{E}_\phi(d_{it}v_{it}d_{is}v_{is}d_{ih}x_{ih}'x_{ih})|) \le MNT$ *a.s., and* $\max_{h\in[T]} \sum_{i,j=1}^N \sum_{t,s=1}^T |\mathbb{E}_\phi(d_{it}v_{it}d_{ih}x_{ih}'x_{jh}d_{jh}d_{js}v_{js})| \le M(NT + N^2)$ *a.s..*

Assumption 7(i)-(ii) imposes weak dependence conditions for $\{d_{it}v_{it}\}$ along both the time and cross-sectional dimensions. These conditions generalize Assumption F(1)-(2) in Bai (2003) to the missing data setting, similar to Assumptions 3-4 in Gonçalves and Perron (2020) and Assumptions 2.4 and 3.8 in Fan and Liao (2022). Assumption 7(ii) strengthens Assumption 5. When d_{it} is independent of v_{js} for all i, j, t, s conditional on ϕ or d_{it} is treated as nonrandom, Assumption 7 hold as long as it holds when there is no missing data.

Assumption 8 *Define* $\widetilde{d_{it}x_{it}} = d_{it}x_{it} - \mathbb{E}_\phi(d_{it}x_{it})$*. Let* cov_ϕ *denote covariance conditional on* ϕ^0*.*

[13] Note that the Hessian and the remainder are high dimensional, but their structures allow us to decompose and analyze them accurately. The intuition is that the Hessian of all these models are approximately diagonal; see Appendix A and Lemma E.4 for the Hessian matrix and remainder terms for the current study.

(i) $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{x}_{it}d_{it}v_{it}\xrightarrow{d}\mathcal{N}(0,\Omega_x)$;

(ii) $\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\sum_{u,h=1}^{T}|cov_\phi(\widetilde{d_{it}x_{it}}d_{is}v_{is},\widetilde{d_{ju}x_{ju}}d_{jh}v_{jh})|\le MNT^2$, *and* $\sum_{t,s=1}^{T}\sum_{i,j=1}^{N}\sum_{k,l=1}^{N}|cov_\phi(\widetilde{d_{it}x_{it}}d_{jt}v_{jt},\widetilde{d_{ks}x_{ks}}d_{ls}v_{ls})|\le MN^2T$ *a.s.,*

(iii) $\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\sum_{u,h=1}^{T}|cov_\phi(\tilde{d}_{it}d_{is}v_{is},\tilde{d}_{ju}d_{jh}v_{jh})|\le MNT^2$, *and* $\sum_{t,s=1}^{T}\sum_{i,j=1}^{N}\sum_{k,l=1}^{N}|cov_\phi(\tilde{d}_{it}d_{jt}v_{jt},\tilde{d}_{ks}d_{ls}v_{ls})|\le MN^2T$ *a.s.;*

(iv) $\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\sum_{u,h=1}^{T}|cov_\phi(d_{it}v_{it}d_{is}v_{is},d_{ju}v_{ju}d_{jh}v_{jh})|\le MNT^2$, *and* $\sum_{t,s=1}^{T}\sum_{i,j=1}^{N}\sum_{k,l=1}^{N}|cov_\phi(d_{it}v_{it}d_{jt}v_{jt},d_{ks}v_{ks}d_{ls}v_{ls})|\le MN^2T$ *a.s..*

Assumption 8(i) is the central limit theorem. Assumption 8(ii) corresponds to Assumption 5(v) in Moon and Weidner (2017) but the latter authors assume v_{it} is independent across (i,t). In Bai (2009), Assumption 8(ii) is satisfied automatically since x_{it} is strictly exogenous there. Assumption 8(iii) is satisfied automatically when $\tilde{d}_{it}=0$, i.e., when there is no missing data or when d_{it} is treated as nonrandom. Assumption 8(iv) corresponds to Assumption C(iv) in Bai (2009) with "e_{it}" replaced by $d_{it}v_{it}$, and is also satisfied in Moon and Weidner (2017) in the case of no missing since they assume v_{it} is independent across (i,t).

Utilizing the asymptotic expansion (4.1)-(4.2) and Assumptions 7-8, the following proposition refines the results of Theorem 3.1.

Proposition 4.1 *Suppose Assumptions 1–5 and 7–8 hold and* $\frac{N^{\frac{1}{\kappa}}T^{\frac{1}{\kappa}}+N^{\frac{2}{\kappa}}}{\sqrt{T}}\to 0$ *and* $\frac{N^{\frac{1}{\kappa}}T^{\frac{1}{\kappa}}+T^{\frac{2}{\kappa}}}{\sqrt{N}}\to 0$ *as* $(N,T)\to\infty$. *Then (i)* $\left\|\hat{\beta}-\beta^0\right\|=O_p(\frac{1}{c_{NT}^2})$ *and (ii)* $\left\|D_{NT}^{-\frac{1}{2}}(\hat{\phi}-\phi^0)+D_{NT}^{-\frac{1}{2}}Q_{\phi\phi'}^{-1}S_\phi\right\|=O_p(\frac{1}{c_{NT}^2})$.

Proposition 4.1(i) establishes the accurate convergence rate of $\hat{\beta}-\beta^0$ and Proposition 4.1(ii) shows that the remainder of the asymptotic expansion of $D_{NT}^{-\frac{1}{2}}(\hat{\phi}-\phi^0)$ is $O_p(\frac{1}{c_{NT}^2})$. Note that Proposition 4.1 holds with $\kappa=\infty$ when d_{it} is treated as nonrandom. Proposition 4.1(ii) is crucial for calculating the effect of using estimated factors and loadings on the asymptotic distribution and bias of $\hat{\beta}^0$, which is presented in the following theorem.

Theorem 4.1 *Suppose that Assumptions 1–5 and 7–8 hold, and* $\frac{T^{[\frac{1}{2}+((\frac{3}{\varsigma\wedge\zeta}+\frac{3}{\varsigma}+\frac{1}{\varrho})\vee\frac{1}{\kappa})]\vee(\frac{2}{3}+\frac{4}{3\kappa})}}{N}\to 0$, $\frac{N^{[\frac{1}{2}+((\frac{3}{\varsigma\wedge\zeta}+\frac{3}{\varsigma}+\frac{1}{\varrho})\vee\frac{1}{\kappa})]\vee(\frac{2}{3}+\frac{4}{3\kappa})}}{T}\to 0$ *for some* $\kappa>4$ *as* $(N,T)\to\infty$. *Then*

$$\sqrt{NT}(\hat{\beta}-\beta^0)-W_x^{-1}b\xrightarrow{d}\mathcal{N}(0,W_x^{-1}\Omega_xW_x^{-1})$$

where $b=\sum_{l=1}^{6}b_l$, b_l*'s are all* $K\times 1$ *vectors whose k-th element are respectively given by*

$$\begin{aligned}
b_{1k} &= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}_\phi(d_{jt}v_{jt}d_{it}x_{itk})\lambda_i^{0\prime}[\bar{L}_{ff'}^{-1}]_t\lambda_j^0,\\
b_{2k} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}_\phi(d_{it}v_{it}d_{is}x_{isk})f_t^{0\prime}[\bar{L}_{\lambda\lambda'}^{-1}]_if_s^0,
\end{aligned}$$

$$
\begin{aligned}
b_{3k} &= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}_\phi(d_{it}v_{it}d_{jt}v_{jt})\delta_{ki}^{0\prime}[\bar{L}_{ff'}^{-1}]_t\lambda_j^0,\\
b_{4k} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}_\phi(d_{it}v_{it}d_{is}v_{is})\omega_{kt}^{0\prime}[\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0,\\
b_{5k} &= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}_\phi(d_{it}v_{it}d_{jt}v_{jt})\lambda_i^{0\prime}\Xi_{kt}\lambda_j^0,\\
b_{6k} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}_\phi(d_{it}v_{it}d_{is}v_{is})f_t^{0\prime}\Delta_{ki}f_s^0,
\end{aligned}
$$

$[\bar{L}_{ff'}^{-1}]_t = (-\sum_{i=1}^N \Phi_{it}\lambda_i^0\lambda_i^{0\prime})^{-1}$, $[\bar{L}_{\lambda\lambda'}^{-1}]_i = (-\sum_{t=1}^T \Phi_{it}f_t^0 f_t^{0\prime})^{-1}$, $\Delta_{ki} = [\bar{L}_{\lambda\lambda'}^{-1}]_i(\sum_{t=1}^T \Phi_{it}f_t^0\omega_{kt}^{0\prime})[\bar{L}_{\lambda\lambda'}^{-1}]_i$, $\Xi_{kt} = [\bar{L}_{ff'}^{-1}]_t(\sum_{i=1}^N \Phi_{it}\lambda_i^0\delta_{ki}^{0\prime})[\bar{L}_{ff'}^{-1}]_t$ *and* δ_{ki}^0 *and* ω_{kt}^0 *are defined in (2.2).*

Theorem 4.1 shows that $\hat{\beta}$ is consistent and asymptotically normal even if d_{it} is correlated with f_t^0 and λ_i^0 as long as d_{it} is independent with v_{it}. In other words, if the latent confounders that affect both the outcome y_{it} and the treatment indicator d_{it} can be modeled by a factor structure, then $\hat{\beta}$ has no selection bias once we perform the IFEs estimation. This covers a lot of empirically relevant cases since sample selection is likely to be affected by both individual heterogeneity and common shocks. While we assume independence between d_{it} and v_{it}, the techniques developed for Proposition 4.1 and Theorem 4.1 can be readily applied to cases where the latent confounders can not be fully captured by a factor structure; see Section 5 for details.

Theorem 4.1 also shows how missing patterns affect the asymptotic variances and biases. Both the bias $W_x^{-1}b$ and the variance $W_x^{-1}\Omega_x W_x^{-1}$ depend on the missing pattern, regressors, factors, loadings and error terms in a complicated way. Our simulation results show that the bias could be large when d_{it} is correlated with f_t^0 and λ_i^0. Moreover, the expressions of W_x, Ω_x and b allow us to construct analytical bias corrections and may also provide some guidance on experimental design.

Remark 4.1 *Theorem 4.1 holds with* $\kappa = \infty$ *for block missing. Note that* $\varsigma = \infty$ *for uniformly bounded factors and loadings under Assumption 2.* ϱ *and* ζ *also could be very large. Then the sufficient (but not necessary) conditions on* N *and* T *in Theorem 4.1 approximately become* $\frac{T^{\frac{2}{3}}}{N} \to 0$ *and* $\frac{N^{\frac{2}{3}}}{T} \to 0$.

Remark 4.2 *Before the* $\sqrt{NT}$*-normalizations,* b_l*'s are of order* $O_p(\frac{1}{N})$ *for* $l = 1, 3, 5$ *and* $O_p(\frac{1}{T})$ *for* $l = 2, 4, 6$ *under the usual weak cross-sectional and serial dependence conditions. Our bias expressions include those of the static panel Bai (2009) and the dynamic panel Moon and Weidner (2017) as special cases.*

(i) First, $b_{1k} = 0$ *for Bai (2009) since* $d_{it} = 1$ *for all* (i, t) *and* $\mathbb{E}_\phi(d_{jt}v_{jt}d_{it}x_{itk}) = 0$ *when* x_{itk} *is strictly exogenous.* $b_{1k} = 0$ *for Moon and Weidner (2017) since* $d_{it} = 1$ *for all* (i, t) *and they assume that* (x_{it}, v_{it}) *is independent across* i *and* $\mathbb{E}(v_{is} \,|x_{it}, v_{i,t-1}, \phi^0) = 0$ *for* $s \geq t$.

(ii) Second, $b_{2k} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=t+1}^{T} \mathbb{E}_\phi(d_{it}v_{it}d_{is}x_{isk}) f_t^{0\prime} [\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0$ *if* $d_{it}v_{it}$ *is uncorrelated with* $d_{is}x_{isk}$ *for* $s \leq t$ *as in dynamic panels, which corresponds to term B1 in Moon and Weidner (2017).* $b_{2k} = 0$ *in Bai (2009) for static panel.*

(iii) Third, $b_3 + b_5$ *corresponds to term B in expression (18) of Bai (2009) and term B2 in Moon and Weidner (2017). Both* b_3 *and* b_5 *are* $O_p(\frac{\sqrt{T}}{\sqrt{N}})$*, which is* $o_p(1)$ *if* $T/N \to 0$*.*

(iv) Fourth, $b_4 + b_6$ *corresponds to term C in expression (19) of Bai (2009) and term B3 in Moon and Weidner (2017). Both* b_4 *and* b_6 *are* $O_p(\frac{\sqrt{N}}{\sqrt{T}})$*, which is* $o_p(1)$ *if* $N/T \to 0$*.*

(v) Fifth, $b_4 + b_6 = 0$ *if* $d_{it}v_{it}$ *is uncorrelated across* t *and both* $\Phi_{it} = \mathbb{E}_\phi(d_{it})$ *and* $\mathbb{E}_\phi(d_{it}v_{it}^2)$ *are constant across* t *so that we have* $\Phi_{it} = \Phi_i$*. In this special case,* $[\bar{L}_{\lambda\lambda'}^{-1}]_i = -\frac{1}{\Phi_i}(\sum_{t=1}^{T} f_t^0 f_t^{0\prime})^{-1}$ *and*

$$\begin{aligned}
b_{6k} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}_\phi(d_{it}v_{it}^2) f_t^{0\prime} \Delta_{ki} f_t^0 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbb{E}_\phi(d_{i1}v_{i1}^2) \sum_{t=1}^{T} f_t^{0\prime} \Delta_{ki} f_t^0 \\
&= \frac{1}{\sqrt{NT}} tr\left(\sum_{i=1}^{N} \mathbb{E}_\phi(d_{i1}v_{i1}^2) \sum_{t=1}^{T} f_t^0 f_t^{0\prime} [\bar{L}_{\lambda\lambda'}^{-1}]_i (\Phi_i \sum_{t=1}^{T} f_t^0 \omega_{kt}^{0\prime}) [\bar{L}_{\lambda\lambda'}^{-1}]_i \right) \\
&= -\frac{1}{\sqrt{NT}} tr(\sum_{i=1}^{N} \mathbb{E}_\phi(d_{i1}v_{i1}^2)(\sum_{t=1}^{T} f_t^0 \omega_{kt}^{0\prime})[\bar{L}_{\lambda\lambda'}^{-1}]_i) \\
&= -\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}_\phi(d_{it}v_{it}^2) \omega_{kt}^{0\prime} [\bar{L}_{\lambda\lambda'}^{-1}]_i f_t^0 = -b_{4k}.
\end{aligned}$$

Similarly, $b_3 + b_5 = 0$ *if* v_{it} *is uncorrelated across* i *and both* $\mathbb{E}_\phi(d_{it})$ *and* $\mathbb{E}_\phi(d_{it}v_{it}^2)$ *are constant across* i*. But other than these special cases,* $b_4 + b_6$ *and* $b_3 + b_5$ *are generally nonzero. In particular, if* $\mathbb{E}_\phi(d_{it})$ *is not a constant across* i *and* t*, either* $b_3 + b_5 \neq 0$ *or* $b_4 + b_6 \neq 0$ *even if* $d_{it}v_{it}$ *is uncorrelated and homoskedastic across* i *and* t*.*

4.2 Analytical Bias Correction

In this subsection we utilize the expressions of W_x, Ω_x and b to construct analytical bias corrections for $\hat{\beta}$. Since the estimation of W_x, Ω_x and b depends on the serial and cross-sectional dependence, we impose the following assumption to simplify the asymptotics.

Assumption 9 *(i)* $\{(d_{it}, v_{it}, x_{it}), t = 1, ..., T\}$ *is independent across* i *conditioning on* ϕ^0*;*

(ii) $\mathbb{E}_\phi(d_{it}v_{it}d_{is}x_{isk}) \leq M\,|s-t|^{c_2}$ *and* $\mathbb{E}_\phi(d_{it}v_{it}d_{is}v_{is}) \leq M\,|s-t|^{c_2}$ *for all* i *and* $s \neq t$ *for some constant* $c_2 < -1$*.*

Assumption 9 rules out cross-sectional dependence of (d_{it}, v_{it}, x_{it}), but allows serial dependence of (d_{it}, v_{it}, x_{it}) and weakly exogenous regressors. As discussed in Bai (2009), the difficulty of bias correction under cross-sectional dependence is that there is no natural ordering of the data and large $|i - j|$ does not mean small correlation between v_{it} and v_{jt}. This issue has not been fully solved even without missing data. A promising solution is to extend the bootstrap bias correction for the individual effects panel to the IFEs panel, see Gonçalves and Perron (2014), Kim and Sun (2016) and Higgins and Jochmans (2024). We leave this for future research and simply focus on cases under Assumption 9.

When v_{it} is uncorrelated across i and t which can occur for both static and dynamic panels, W_x, Ω_x and b are estimated as follows:[14]

$$
\begin{aligned}
&\hat{W}_x = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}\hat{\dot{x}}_{it}\hat{\dot{x}}_{it}',\ \hat{\Omega}_x = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}\hat{v}_{it}^2\hat{\dot{x}}_{it}\hat{\dot{x}}_{it}',\\
&\hat{b}_{1k} = 0,\ \hat{b}_{2k} = 0 \text{ if } x_{it} \text{ is strictly exogenous},\\
&\hat{b}_{2k} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\Gamma(\frac{s-t}{L_T})d_{is}\hat{v}_{it}d_{it}x_{isk}\hat{f}_t'[\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\hat{f}_s \text{ if } x_{it} \text{ is weakly exogenous},\\
&\hat{b}_{3k} = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N} d_{it}\hat{v}_{it}^2\hat{\delta}_{ki}'[\hat{\bar{L}}_{ff'}^{-1}]_t\hat{\lambda}_i \text{ and } \hat{b}_{4k} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}\hat{v}_{it}^2\hat{\omega}_{kt}'[\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\hat{f}_t,\\
&\hat{b}_{5k} = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N} d_{it}\hat{v}_{it}^2\hat{\lambda}_i'\Xi_{kt}\hat{\lambda}_i \text{ and } \hat{b}_{6k} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}\hat{v}_{it}^2\hat{f}_t'\hat{\Delta}_{ki}\hat{f}_t,
\end{aligned}
$$

where $\Gamma(\cdot)$ is a kernel function such that $\Gamma(\frac{s-t}{L_T}) = \mathbf{1}\{|s-t| \le L_T\}$, L_T is a bandwidth parameter,

$$
\begin{aligned}
[\hat{\bar{L}}_{ff'}^{-1}]_t &= (-\sum_{i=1}^{N} d_{it}\hat{\lambda}_i\hat{\lambda}_i')^{-1},\ [\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i = (-\sum_{t=1}^{T} d_{it}\hat{f}_t\hat{f}_t')^{-1},\\
\hat{\Delta}_i &= [\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i(\sum_{t=1}^{T} d_{it}\hat{f}_t\hat{\omega}_{kt}')[\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i,\ \hat{\Xi}_t = [\hat{\bar{L}}_{ff'}^{-1}]_t(\sum_{i=1}^{N} d_{it}\hat{\lambda}_i\hat{\delta}_{ki}')[\hat{\bar{L}}_{ff'}^{-1}]_t,\\
\hat{v}_{it} &= y_{it} - x_{it}'\hat{\beta} - \hat{\lambda}_i'\hat{f}_t,\ \hat{\dot{x}}_{itk} = x_{itk} - \hat{\delta}_{ki}'\hat{f}_t - \hat{\omega}_{kt}'\hat{\lambda}_i,\\
(\{\hat{\delta}_{ki}\}_{i\in[N]}, \{\hat{\omega}_{kt}\}_{t\in[T]}) &= \underset{\delta_{ki},\omega_{kt}}{\arg\min}\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}(x_{itk} - \delta_{ki}'\hat{f}_t - \omega_{kt}'\hat{\lambda}_i)^2. \qquad (4.3)
\end{aligned}
$$

Eqn. (G.1) in the Appendix shows how to solve the minimization problem (4.3).

When v_{it} is correlated across t but uncorrelated across i under Assumption 9, we focus on the static panels with strict exogeneity to avoid the endogeneity issue. In this case, $\hat{W}_x$, $\hat{b}_{1k}$, $\hat{b}_{3k}$ and $\hat{b}_{5k}$ remain the same, $\hat{b}_{2k} = 0$, $\hat{\Omega}_x$, $\hat{b}_{4k}$ and $\hat{b}_{6k}$ are adjusted as follows:

$$
\begin{aligned}
\hat{\Omega}_x &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\Gamma(\frac{|s-t|}{L_T})d_{it}\hat{v}_{it}d_{is}\hat{v}_{is}\hat{\dot{x}}_{it}\hat{\dot{x}}_{is}',\\
\hat{b}_{4k} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\Gamma(\frac{|s-t|}{L_T})d_{it}\hat{v}_{it}d_{is}\hat{v}_{is}\hat{\omega}_{kt}'[\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\hat{f}_s,\\
\hat{b}_{6k} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\Gamma(\frac{|s-t|}{L_T})d_{it}\hat{v}_{it}d_{is}\hat{v}_{is}\hat{f}_t'\hat{\Delta}_{ki}\hat{f}_s.
\end{aligned}
$$

Let $\hat{\beta}^{abc} = \hat{\beta} - \frac{1}{\sqrt{NT}}\hat{W}_x^{-1}\hat{b}$ denote the analytically bias-corrected estimator where $\hat{b} = \sum_{l=1}^{6}\hat{b}_l$ and $\hat{b}_l = (\hat{b}_{l1}, ..., \hat{b}_{lK})'$, we have the following theorem for $\hat{\beta}^{abc}$.

[14] This is because $\mathbb{E}_\phi(d_{it}v_{it}d_{jt}v_{jt}) = \mathbb{E}_\phi(d_{it}d_{jt})\mathbb{E}_\phi(v_{it})\mathbb{E}_\phi(v_{jt}) = 0$, $\mathbb{E}_\phi(d_{jt}v_{jt}d_{it}x_{itk}) = \mathbb{E}_\phi(d_{jt}d_{it}x_{itk})\ \mathbb{E}_\phi(v_{jt}) = 0$ if $i \ne j$, $\mathbb{E}_\phi(d_{it}v_{it}d_{is}v_{is}) = \mathbb{E}_\phi(d_{it}d_{is})\mathbb{E}_\phi(v_{it})\mathbb{E}_\phi(v_{is}) = 0$ if $t \ne s$, and $\Omega_x = plim\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}_\phi(d_{it}v_{it}^2)\dot{x}_{it}\dot{x}_{it}'$.

Theorem 4.2 *Suppose Assumptions 1–5 and 7–9 hold. Suppose that* $\frac{N}{T} \to \varepsilon \in (0, \infty)$, $L_T/T^{\frac{1}{4}-\frac{4}{\varrho \wedge \varsigma}} \to 0$ *and* $L_T \to \infty$, *as* $(N, T) \to \infty$. *Then*

(i) $\hat{W}_x = W_x + o_p(1)$, $\hat{\Omega}_x = \Omega_x + o_p(1)$, $\hat{b}_l = b_l + o_p(1)$ *for* $l \in [6]$;

(ii) $\sqrt{NT}(\hat{\beta}^{abc} - \beta^0) \xrightarrow{d} \mathcal{N}(0, W_x^{-1}\Omega_x W_x^{-1})$.

Theorem 4.2 shows the analytically corrected estimator $\hat{\beta}^{abc}$ is asymptotically unbiased. To simplify the proof and the conditions on (N, T, L_T), we now focus on the case where N and T pass to infinity at the same speed. The confidence intervals of elements of β^0 can be constructed based on $\hat{\beta}^{abc}$ and $\hat{W}_x^{-1}\hat{\Omega}_x\hat{W}_x^{-1}$ as usual. Note that split panel jackknife (SPJ) bias corrections may not work for unbalanced panels since the "symmetry" utilized by the SPJ bias correction is likely to be destroyed by the missing data, unless the missing is uniform and completely random and certain stationarity conditions along both the cross-sectional and time dimensions are satisfied.

5 Discussion: Heckman Correction for Sample Selection with IFEs

When d_{it} is still correlated with the error term after conditioning on the factors and loadings, one possible solution is to employ the Heckman correction. Consider the following parametric model:

$$y_{it} = d_{it}(x_{it}'\beta_x^0 + f_t^{0\prime}\lambda_i^0 + \epsilon_{it}), \tag{5.1}$$

$$d_{it}^* = z_{it}'\delta^0 + \varphi(g_t^0, \alpha_i^0) + u_{it} \text{ and } d_{it} = 1\{d_{it}^* > 0\}, \tag{5.2}$$

where z_{it} may contain x_{it}, lagged y_{it}, lagged d_{it}, and other observable variables, $\varphi(g_t^0, \alpha_i^0)$ denotes the fixed effect component, u_{it} is the error term, and (u_{it}, ϵ_{it}) is jointly normal conditioning on the regressors, the factors and the loadings. $\varphi(g_t^0, \alpha_i^0)$ could be the one way/two way/interactive effects or have a more general form. To correct the selection endogeneity, we can add a Heckman correction term to eqn. (5.1). More specifically,

$$\mathbb{E}(y_{it} | d_{it} = 1) = x_{it}'\beta_x^0 + f_t^{0\prime}\lambda_i^0 + \mathbb{E}(u_{it} | d_{it} = 1)\frac{\mathbb{E}(u_{it}\epsilon_{it})}{\mathbb{E}(u_{it}^2)} = x_{it}'\beta_x^0 + f_t^{0\prime}\lambda_i^0 + m_{it}\beta_m^0, \tag{5.3}$$

where $m_{it} = \mathbb{E}(u_{it} | d_{it} = 1) = m(z_{it}'\delta^0 + \varphi(g_t^0, \alpha_i^0))$ and $m(\cdot)$ is the inverse Mills ratio function. Let $v_{it} = \epsilon_{it} - m_{it}\beta_m^0$. Then we have $y_{it} = x_{it}'\beta_x^0 + m_{it}\beta_m^0 + f_t^{0\prime}\lambda_i^0 + v_{it}$, i.e., we use $m_{it}\beta_m^0$ to correct the endogeneity.

Since m_{it} depends on unknown parameters $(\delta^0, g^0, \alpha^0)$, we can estimate the selection eqn. (5.2) first to obtain the estimators $(\hat{\delta}, \hat{\alpha}, \hat{g})$ and calculate $\hat{m}_{it} = m(z_{it}'\hat{\delta} + \varphi(\hat{g}_t, \hat{\alpha}_i))$. Then we replace m_{it} by $\hat{m}_{it}$ in the main eqn. and estimate the main eqn. by quasi-MLE: $(\hat{\beta}, \hat{\lambda}, \hat{f}) = \arg\max_{\gamma \in \mathcal{B}_m(\gamma^0)} Q(\beta, \lambda, f)$, where $\beta = (\beta_x', \beta_m)'$, $Q(\beta, \lambda, f) = L(\beta, \lambda, f) + G(\lambda, f)$, $G(\lambda, f)$ and $\mathcal{B}_m(\gamma^0)$ are the same as expressions (3.2) and (3.3), respectively, and $L(\beta, \lambda, f) = -\frac{1}{2}\sum_{i=1}^N \sum_{t=1}^T d_{it}(y_{it} - x_{it}'\beta_x - \hat{m}_{it}\beta_m - f_t'\lambda_i)^2$. Following the estimation procedure in Section 3, we get the initial estimates $(\hat{\beta}^{(0)}, \hat{\lambda}^{(0)}, \hat{f}^{(0)})$ first and then iterate by alternating maximization until convergence. The main difference here is that there is an extra error term due to replacing m_{it} by $\hat{m}_{it}$, i.e.,

$$y_{it} = x_{it}'\beta_x^0 + \hat{m}_{it}\beta_m^0 + f_t^{0\prime}\lambda_i^0 + v_{it} + (m_{it} - \hat{m}_{it})\beta_m^0. \tag{5.4}$$

Following the same roadmap of convergence analysis and asymptotic analysis in Section 3 and Section 4, Theorems 3.1-3.3, Proposition 4.1 and Theorem 4.1 can be reestablished in a similar way and we just need to calculate the effect of $(m_{it} - \hat{m}_{it})\beta_m^0$ on the asymptotic biases and variances. Depending on the model of $\varphi(g_t^0, \alpha_i^0)$, the calculations could be straightforward or tedious.

There is a large literature on panel data sample selection models, including both fixed effects and random effects approaches and both parametric and semi-non parametric approaches; see the references in the Introduction. Existing studies mainly focus on models with only individual effects under the large N fixed T setup, except for Fernández-Val and Vella (2011) who studies individual effects model under large N and large T. Equations (5.1)-(5.2) generalize the sample selection model to the case of IFEs, which is crucial for capturing both individual heterogeneities and common shocks in the selection process. Equations (5.1)-(5.2) is a fixed effects parametric approach. The advantage is avoiding distributional assumptions on the unobservable heterogeneities, while the cost is that the joint normality assumption may be too strong. Therefore, it is also promising to extend our results in Sections 3-4 to the propensity-score-based methods.

6 Simulations

In this section, we report simulation results for our proposed algorithms based on 1000 replications. We focus on the following panel data model $y_{it} = d_{it}(\sum_{r=1}^2 \lambda_{ir}^0 f_{tr}^0 + x_{it}'\beta^0 + v_{it})$, where $d_{it} = \mathbf{1}\{y_{it} \text{ and } x_{it} \text{ are observable}\}$, $x_{it} = (x_{1,it}, x_{2,it})'$ and $\beta^0 = (\beta_1^0, \beta_2^0)'$.

6.1 Data Generating Processes (DGPs)

The following two main DGPs are employed.

DGP 1: Static panel. For any $r \in \{1, 2\}$, we set $f_{tr}^0 \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\lambda_{ir}^0 \overset{i.i.d}{\sim} \mathcal{N}(1,1)$, $x_{1,it} = 1 + \sum_{r=1}^2(\lambda_{ir}^0 + \mu_{ir})(f_{tr}^0 + f_{t-1,r}^0) + \mathcal{N}(0,1)$ with $\mu_{ir} \overset{i.i.d}{\sim} \mathcal{N}(1,1)$. $x_{2,it} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\beta^0 = (1,1)'$. Errors $v_{it} = \frac{1}{\sqrt{2}}(e_{it} + e_{i,t-1})$ where $e_{it} \sim \mathcal{N}(0, \sigma_{e,it}^2)$ and $\sigma_{e,it} \overset{i.i.d}{\sim} U(0.5, 1.5)$.

DGP 2: Dynamic panel. In this case, $f_{tr}^0 = \rho_f f_{t-1,r}^0 + u_{tr}$ where $u_{tr} \stackrel{i.i.d}{\sim} \mathcal{N}(0,(1-\rho_f^2)\sigma_f^2)$ with $\rho_f = 0.5$, $\sigma_f = 0.5$. Moreover, $x_{1,it} = y_{i,t-1}$, and $v_{it} \sim \mathcal{N}(0,\sigma_{v,it}^2)$ with $\sigma_{v,it} \stackrel{i.i.d}{\sim} U(0.5, 1.5)$. $\beta^0 = (0.3, 1)'$. All others are same as DGP 1.

We define the missing patterns as follows.

Pattern 1: Completely random heterogeneous missing. d_{it} is binary, independent across (i,t) and independent of $(f_s^0, \lambda_j^0, v_{js})$ for all (i,j,t,s). $p_{it} := \mathbb{E}(d_{it})$ follows $i.i.d$ $U(0.5, 0.9)$ across (i,t).

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

Pattern 3: Mixed frequency. $d_{it} = 0$ if $i > 0.6N$ and $t/3$ is not an integer.

Pattern 4: Staggered missing. $d_{it} = 0$ when $(i,t) \in \{0.4N + 1 \le i \le 0.7N$ and $0.7T + 1 \le t \le T\}$ or $(i,t) \in \{0.7N + 1 \le i \le 0.7N$ and $0.4T + 1 \le t \le T\}$.

For the static and dynamic panel DGPs, we set the factor structure and the first regressor $x_{1,it}$ following Moon and Weidner (2015) and Moon and Weidner (2017). We allow the error term to be heteroskedastic in both DGPs. The number of factors are two in our simulations. We apply the proposed SVT approach to estimate the true number of factors, with the results presented in the next subsection. To derive the inference results, we adopt the commonly used Bartlett kernel for bias correction and covariance matrix estimation. For the bandwidth, we set $L_T = \lfloor cT^{1/5} \rfloor$ with $c = 2$ for the bias correction and the variance estimation.

6.2 Simulation Results

In this subsection, we first consider the performance of SVT in estimating the number of factors and then study the performance of the proposed AM algorithms. Table 1 reports the frequency of correctly estimating the number of factors using the SVT approach outlined in Section 3.3 with the true number of factors being 2. For all missing patterns across all DGPs, the accuracy is notably high, reflecting the robustness of the SVT approach across various scenarios.

Below we show the performance of the proposed AM algorithms. Define $\mathrm{RMSE}(\Theta) = \frac{1}{\sqrt{NT}} \left\| \Theta - \Theta^0 \right\|$ for any $\Theta \in \mathbb{R}^{T\times N}$. Similarly, define $\mathrm{RMSE}(\Lambda) = \frac{1}{\sqrt{N}} \left\| \Lambda - \Lambda^0 \right\|$ for any $\Lambda \in \mathbb{R}^{N\times r}$, $\mathrm{RMSE}(F) = \frac{1}{\sqrt{T}} \left\| F - F^0 \right\|$ for any $F \in \mathbb{R}^{T\times r}$, and $\mathrm{RMSE}(\beta) = \left\| \beta - \beta^0 \right\|$ for any $\beta \in \mathbb{R}^K$. Note that $K = 2$ in the simulations. We set the convergence accuracy of alternating maximization algorithm to be 10^{-4}.

Table 1: Frequency of estimating the correct number of factors by using SVT

Pattern	DGP 1				DGP 2			
	N=100			N=200	N=100			N=200
	T=25	T=50	T=100	T=100	T=25	T=50	T=100	T=100
1	1.00	1.00	0.957	1.00	1.00	0.992	1.00	1.00
2	0.868	0.888	0.986	0.998	0.955	0.969	0.993	1.00
3	1.00	1.00	1.00	1.00	1.00	0.999	1.00	1.00
4	1.00	1.00	1.00	1.00	1.00	0.969	1.00	1.00

Table 2 reports the number of iterations required for convergence in the two alternating maximization algorithms: "AM1", which corresponds to the Gauss-Seidel procedure from Algorithm 1, and "AM2", which corresponds to the Jacobi procedure from Algorithm 2. The table shows that the Gauss-Seidel procedure typically converges within 30 iterations, whereas the Jacobi procedure requires significantly more iterations, generally staying below 100. Across all DGPs and missing patterns, the Jacobi procedure consistently requires more iterations to achieve convergence compared to the Gauss-Seidel procedure. This difference arises because, in the Jacobi procedure, the slope estimator depends on factor estimates from the previous iteration, increasing its computational burden. Moreover, scenarios with serial correlation require substantially more iterations for both procedures compared to dynamic cases. Overall, "AM1", i.e., the Gauss-Seidel procedure, exhibits greater efficiency, achieving numerical convergence with fewer iterations across most settings.

Table 2: Number of iterations by AM algorithm

DGP	N	T	AM1				AM2			
			Pattern 1	Pattern 2	Pattern 3	Pattern 4	Pattern 1	Pattern 2	Pattern 3	Pattern 4
1	100	25	23.94	24.85	26.73	31.33	65.77	81.49	95.45	115.89
		50	18.38	15.50	17.43	24.58	61.80	35.46	59.20	98.37
		100	13.82	14.46	13.28	14.33	55.88	56.35	66.12	64.62
	200	100	12.87	13.05	11.25	13.17	58.91	65.46	55.84	56.66
2	100	25	7.31	15.65	7.48	12.24	11.55	28.37	13.32	19.42
		50	8.43	7.66	8.34	10.02	19.32	13.55	20.03	23.93
		100	6.39	6.58	6.25	6.11	15.58	16.21	16.67	14.40
	200	100	5.69	5.87	5.25	5.73	14.98	14.99	13.69	15.00

Tables 3 presents the RMSE results for the slope and intercept matrix estimators associated with factor and factor loading estimates across three algorithms: "NNR", "AM1", and "AM2". The results are divided into three panels, with the first panel (NNR) showing the RMSE results for initial estimators obtained using nuclear norm regularization. The findings reveal significant differences in estimation accuracy across the algorithms. As the first step in the estimation process, NNR produces higher RMSE values compared to AM1 and AM2. After applying the AM algorithms, in contrast, the RMSE values drop significantly. AM1 and AM2 demonstrate substantially lower RMSEs for all β, Θ, factor and factor loadings. For instance, under DGP 1 with $N = 200$ and $T = 100$, AM1 achieves RMSEs of 0.221 for Θ, which is approximately one-sixth of the RMSE of 1.473 achieved by NNR. Additionally, the performance of AM1 and AM2 is generally similar across most settings. Overall, AM1 and AM2 substantially outperform NNR in terms of precision as expected, highlighting the effectiveness of the alternating maximization approach in refining initial estimates.

Table 4 presents the point estimation and inference results for β_1 and β_2 using the Gauss-Seidel algorithm across both DGPs and various missing patterns. For each missing pattern within each DGP, we report the bias, standard deviation (sd), and coverage probability (CV) of the 95% confidence intervals for both the uncorrected estimator and the bias-corrected estimator, denoted by "$\hat{\beta}$" and "$\hat{\beta}^{abc}$", respectively. We first examine the bias in $\hat{\beta}_2$. Due to the construction of $x_{2,it}$, which is independent of both factors and loadings, the bias remains generally small across all DGPs and missing patterns. In contrast, since $x_{1,it}$ is correlated with the factor structure, a noticeable bias is present. As N or T increases, this bias gradually decreases, aligning with the theoretical result in Theorem 4.1. Next, bias correction significantly reduces bias across all settings. When T is small (e.g., $T = 25$), the correction procedure eliminates approximately 40% of the bias in most cases. However, in more complex scenarios, such as missing pattern 2 in DGP 1 – where missingness is correlated with the factor structure – the correction is less effective, reducing the bias by only about 25%. As T increases, the effectiveness of the bias correction improves, with more substantial reductions observed for $T = 50$ and $T = 100$.

For the standard deviation of the estimators, as expected, increasing N or T reduces the standard deviation, as more observations provide more stable estimates. The bias correction procedure has minimal impact on standard deviation, as the values for $\hat{\beta}$ and $\hat{\beta}^{abc}$ remain almost identical across all settings. This suggests that the bias correction method effectively improves accuracy without inflating estimator variability, ensuring stable inference.

Finally, we assess the coverage probability of the 95% confidence intervals. For $\hat{\beta}_2$, the coverage probability is close to the nominal level across most cases, owing to its relatively small bias. However, for $\hat{\beta}_1$, the results highlight the critical role of bias correction. The coverage probability is lower when T is small, particularly in DGP 2, where the initial bias is larger. With bias correction, the coverage probability improves significantly and aligns more closely with 95%, especially in cases with moderate or large T. However, the coverage probabilities sometimes fall short of 0.95, even when T reaches 100; see the results for missing pattern 2 in the dynamic panel. This deviation may be attributed to the challenges in accurately estimating the bias for those more complex scenarios. Nonetheless, as the sample size increases, the coverage probability for $\hat{\beta}_1$ after bias correction

Table 3: RMSE results by NNR and AM algorithms

Pattern	Algorithm	N	T	DGP 1 β	DGP 1 Θ	DGP 1 F	DGP 1 Λ	DGP 2 β	DGP 2 Θ	DGP 2 F	DGP 2 Λ
1	NNR	100	25	0.249	1.755	2.156	2.157	0.289	1.099	1.854	1.457
			50	0.245	1.568	2.208	2.306	0.106	0.948	1.805	1.408
			100	0.227	1.594	2.703	2.794	0.081	0.951	1.795	1.398
		200	100	0.231	1.473	2.162	2.144	0.149	0.960	1.800	1.420
	AM1	100	25	0.029	0.441	0.189	0.354	0.039	0.470	0.366	0.543
			50	0.019	0.322	0.163	0.225	0.021	0.321	0.245	0.333
			100	0.012	0.254	0.158	0.157	0.014	0.260	0.247	0.244
		200	100	0.009	0.221	0.115	0.159	0.011	0.222	0.163	0.223
	AM2	100	25	0.029	0.442	0.189	0.354	0.039	0.471	0.368	0.544
			50	0.020	0.323	0.163	0.225	0.021	0.321	0.245	0.333
			100	0.013	0.254	0.159	0.158	0.014	0.260	0.247	0.244
		200	100	0.009	0.221	0.115	0.159	0.011	0.222	0.163	0.223
2	NNR	100	25	0.315	1.758	2.014	1.698	0.237	0.976	1.873	1.543
			50	0.278	1.622	2.133	1.821	0.119	0.893	2.144	1.584
			100	0.259	1.846	2.219	1.995	0.179	0.960	1.824	1.431
		200	100	0.227	1.739	2.144	1.967	0.173	0.918	1.806	1.370
	AM1	100	25	0.032	0.700	0.464	0.448	0.052	0.544	0.384	0.503
			50	0.022	0.541	0.409	0.306	0.027	0.602	0.680	0.571
			100	0.016	0.770	0.591	0.243	0.016	0.356	0.351	0.286
		200	100	0.011	0.395	0.261	0.211	0.012	0.292	0.257	0.277
	AM2	100	25	0.033	0.798	0.538	0.484	0.052	0.550	0.392	0.505
			50	0.022	0.569	0.420	0.315	0.027	0.623	0.705	0.585
			100	0.016	0.820	0.614	0.239	0.017	0.359	0.354	0.286
		200	100	0.011	0.384	0.247	0.209	0.012	0.297	0.264	0.277
3	NNR	100	25	0.248	1.838	2.295	2.490	0.113	0.970	1.695	1.233
			50	0.250	1.423	2.055	2.165	0.202	1.033	1.879	1.549
			100	0.242	1.567	2.237	2.244	0.143	0.984	1.856	1.511
		200	100	0.238	1.439	2.182	2.126	0.133	0.956	1.815	1.414
	AM1	100	25	0.027	0.486	0.183	0.369	0.031	0.489	0.319	0.570
			50	0.018	0.327	0.156	0.241	0.022	0.354	0.228	0.349
			100	0.012	0.274	0.154	0.177	0.014	0.279	0.222	0.263
		200	100	0.009	0.241	0.119	0.186	0.010	0.250	0.172	0.279
	AM2	100	25	0.027	0.487	0.184	0.369	0.031	0.492	0.322	0.574
			50	0.019	0.327	0.156	0.241	0.022	0.355	0.229	0.349
			100	0.012	0.275	0.155	0.177	0.014	0.279	0.222	0.264
		200	100	0.009	0.241	0.119	0.186	0.010	0.251	0.172	0.280
4	NNR	100	25	0.279	1.981	2.264	2.221	0.216	1.136	1.903	1.518
			50	0.237	1.795	2.229	2.201	0.239	1.050	1.900	1.603
			100	0.225	1.493	2.360	2.372	0.149	0.991	1.827	1.431
		200	100	0.240	1.486	2.220	2.313	0.154	0.962	1.802	1.409
	AM1	100	25	0.027	0.458	0.173	0.322	0.053	0.721	0.575	0.852
			50	0.019	0.352	0.176	0.251	0.027	0.330	0.220	0.307
			100	0.012	0.277	0.172	0.183	0.014	0.286	0.261	0.260
		200	100	0.009	0.220	0.117	0.153	0.011	0.229	0.176	0.236
	AM2	100	25	0.027	0.459	0.174	0.322	0.052	0.730	0.585	0.861
			50	0.019	0.353	0.177	0.251	0.027	0.330	0.220	0.307
			100	0.013	0.277	0.172	0.183	0.014	0.286	0.261	0.261
		200	100	0.009	0.220	0.117	0.153	0.011	0.229	0.176	0.236

gradually approaches the nominal level, demonstrating the asymptotic validity of the correction procedure.

Overall, the results in Table 4 indicate that our AM algorithms are robust and effective in managing a wide range of data structures. The estimation and inference results by using the Jacobi procedure are presented in Table 5, showing patterns and insights similar to those in Table 4.

7 Empirical Application

7.1 Background and Model

State excise taxes have long been a focal point in public finance and econometrics, providing a unique framework to study tax competition, policy diffusion, and temporal dynamics in the fiscal policy. Three key features of state excise taxes have consistently drawn attention in the literature. First, state excise taxes exhibit significant cross-sectional dependence, as tax decisions in one state are often influenced by those in neighboring states. This phenomenon aligns with the broader literature on tax competition and policy interdependence (see, e.g., Brueckner, 2003; Case et al., 1993; Besley and Case, 1995). Empirical studies such as Egger et al. (2005) and Devereux et al. (2007) quantify this cross-sectional dependence of state excise taxes using spatial panel models, which effectively captures the spatial spillover of tax policy. Second, state excise taxes exhibit strong persistence over time, suggesting that past taxes heavily influence current policy choices. This persistence can be attributed to the long-term fiscal planning and the stability of political preferences. Using data on cigarette and gasoline taxes for the U.S. states from 1977 to 1997, Devereux et al. (2007) demonstrate statistically significant and high persistence in excise taxes. Third, the level of income also plays a critical role in shaping state excise taxes. Some studies, such as Chaloupka and Warner (2000), find that higher income per capita can lead to higher excise taxes on specific goods, like alcohol, as wealthier populations may be more willing to tolerate such taxes due to increased awareness of public health considerations.

These three empirical regularities, viz., cross-sectional dependence, high persistence, and the influence of income, highlight the complex dynamics underlying state excise tax policies. In this section, we revisit the topic of U.S. state excise taxes. Unlike Devereux et al. (2007), who use a balanced panel dataset of cigarette and gasoline tax data for 20 years, the proposed algorithms for the unbalanced panel data model motivate us to study a longer period and include an additional excise tax: the wine tax. Wine taxes often exhibit more missing data and are frequently overlooked in the balanced panel data research. Beyond excise taxes, this section also extends the analysis to include income taxes, such as the state-level corporate top tax rate, to explore whether the persistence and the impact of income emerge across different types of state tax policies.

Table 4: Estimation and inference results by Gauss-Seidel algorithm

DGP	missing pattern			N=100, T=25		N=100, T=50		N=100, T=100		N=200, T=100	
				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
1	1	$\hat{\beta}$	bias	-0.0141	0.0008	-0.0069	-0.0006	-0.0042	0.0000	-0.0035	0.0007
			sd	0.0149	0.0264	0.0092	0.0189	0.0060	0.0125	0.0041	0.0089
			CV	0.818	0.935	0.881	0.939	0.874	0.949	0.899	0.952
		$\hat{\beta}^{abc}$	bias	-0.0090	0.0006	-0.0034	-0.0005	-0.0010	0.0000	-0.0008	0.0007
			sd	0.0149	0.0265	0.0092	0.0189	0.0060	0.0125	0.0041	0.0089
			CV	0.871	0.935	0.923	0.939	0.960	0.950	0.955	0.952
	2	$\hat{\beta}$	bias	-0.0112	-0.0004	-0.0065	-0.0013	-0.0032	-0.0002	-0.0023	0.0006
			sd	0.0145	0.0318	0.0113	0.0240	0.0072	0.0149	0.0046	0.0102
			CV	0.830	0.924	0.875	0.925	0.894	0.949	0.916	0.944
		$\hat{\beta}^{abc}$	bias	-0.0098	0.0001	-0.0039	-0.0012	-0.0006	0.0000	0.0000	0.0006
			sd	0.0144	0.0318	0.0114	0.0240	0.0072	0.0149	0.0047	0.0102
			CV	0.851	0.923	0.918	0.926	0.954	0.950	0.952	0.945
	3	$\hat{\beta}$	bias	-0.0104	-0.0015	-0.0059	0.0007	-0.0025	0.0004	-0.0023	0.0003
			sd	0.0124	0.0262	0.0102	0.0176	0.0063	0.0121	0.0043	0.0086
			CV	0.851	0.926	0.894	0.941	0.907	0.950	0.915	0.951
		$\hat{\beta}^{abc}$	bias	-0.0062	-0.0017	-0.0026	0.0003	-0.0004	0.0003	-0.0003	0.0003
			sd	0.0123	0.0262	0.0102	0.0176	0.0063	0.0121	0.0043	0.0086
			CV	0.906	0.925	0.937	0.942	0.947	0.950	0.960	0.950
	4	$\hat{\beta}$	bias	-0.0121	0.0023	-0.0079	-0.0015	-0.0038	0.0003	-0.0041	-0.0002
			sd	0.0127	0.0249	0.0084	0.0175	0.0060	0.0124	0.0044	0.0088
			CV	0.828	0.937	0.860	0.937	0.899	0.942	0.855	0.944
		$\hat{\beta}^{abc}$	bias	-0.0094	0.0022	-0.0036	-0.0013	-0.0014	0.0002	-0.0010	-0.0002
			sd	0.0127	0.0249	0.0084	0.0175	0.0060	0.0124	0.0044	0.0088
			CV	0.870	0.936	0.933	0.936	0.943	0.942	0.945	0.943
2	1	$\hat{\beta}$	bias	-0.0264	-0.0024	-0.0084	-0.0009	-0.0068	-0.0003	-0.0055	0.0002
			sd	0.0205	0.0270	0.0127	0.0178	0.0085	0.0125	0.0060	0.0089
			CV	0.631	0.931	0.863	0.940	0.857	0.943	0.835	0.946
		$\hat{\beta}^{abc}$	bias	-0.0150	-0.0014	-0.0029	-0.0008	-0.0021	-0.0002	-0.0015	0.0002
			sd	0.0196	0.0271	0.0125	0.0178	0.0084	0.0125	0.0059	0.0089
			CV	0.834	0.929	0.924	0.944	0.933	0.945	0.943	0.946
	2	$\hat{\beta}$	bias	-0.0382	-0.0012	-0.0121	-0.0001	-0.0070	-0.0017	-0.0062	-0.0002
			sd	0.0252	0.0337	0.0160	0.0233	0.0101	0.0147	0.0070	0.0098
			CV	0.519	0.912	0.830	0.920	0.862	0.942	0.842	0.955
		$\hat{\beta}^{abc}$	bias	-0.0226	0.0002	-0.0038	0.0001	-0.0023	-0.0017	-0.0014	-0.0001
			sd	0.0241	0.0339	0.0156	0.0233	0.0100	0.0147	0.0069	0.0098
			CV	0.750	0.908	0.917	0.919	0.939	0.941	0.935	0.955
	3	$\hat{\beta}$	bias	-0.0124	0.0009	-0.0113	-0.0006	-0.0038	0.0005	-0.0034	-0.0003
			sd	0.0183	0.0273	0.0125	0.0176	0.0086	0.0127	0.0058	0.0090
			CV	0.865	0.937	0.817	0.945	0.910	0.946	0.897	0.954
		$\hat{\beta}^{abc}$	bias	-0.0047	0.0011	-0.0031	-0.0004	-0.0009	0.0005	-0.0009	-0.0002
			sd	0.0178	0.0273	0.0124	0.0176	0.0086	0.0127	0.0058	0.0090
			CV	0.911	0.936	0.928	0.946	0.942	0.948	0.947	0.953
	4	$\hat{\beta}$	bias	-0.0449	-0.0048	-0.0190	-0.0006	-0.0071	-0.0007	-0.0066	0.0003
			sd	0.0201	0.0263	0.0129	0.0185	0.0087	0.0119	0.0059	0.0090
			CV	0.294	0.926	0.632	0.947	0.852	0.953	0.781	0.949
		$\hat{\beta}^{abc}$	bias	-0.0243	-0.0035	-0.0081	-0.0004	-0.0017	-0.0007	-0.0017	0.0003
			sd	0.0193	0.0264	0.0126	0.0186	0.0085	0.0119	0.0058	0.0090
			CV	0.683	0.932	0.878	0.947	0.930	0.951	0.941	0.949

Table 5: Estimation and inference results by Jacobi algorithm

DGP	missing pattern			N=100, T=25		N=100, T=50		N=100, T=100		N=200, T=100	
				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
1	1	$\hat{\beta}$	bias	-0.0138	0.0007	-0.0066	-0.0006	-0.0049	0.0000	-0.0034	0.0007
			sd	0.0151	0.0264	0.0096	0.0189	0.0064	0.0125	0.0045	0.0089
			CV	0.823	0.930	0.879	0.937	0.870	0.948	0.894	0.951
		$\hat{\beta}^{abc}$	bias	-0.0087	0.0006	-0.0032	-0.0005	-0.0009	0.0000	-0.0007	0.0007
			sd	0.0151	0.0265	0.0095	0.0189	0.0064	0.0125	0.0045	0.0089
			CV	0.870	0.933	0.916	0.937	0.949	0.948	0.951	0.951
	2	$\hat{\beta}$	bias	-0.0110	-0.0008	-0.0061	-0.0015	-0.0033	-0.0002	-0.0025	0.0003
			sd	0.0153	0.0319	0.0119	0.0240	0.0079	0.0150	0.0053	0.0102
			CV	0.821	0.921	0.885	0.924	0.892	0.948	0.919	0.943
		$\hat{\beta}^{abc}$	bias	-0.0095	-0.0004	-0.0034	-0.0014	-0.0007	0.0000	-0.0002	0.0003
			sd	0.0152	0.0318	0.0119	0.0240	0.0080	0.0150	0.0053	0.0102
			CV	0.843	0.920	0.917	0.925	0.940	0.948	0.942	0.942
	3	$\hat{\beta}$	bias	-0.0105	-0.0015	-0.0057	0.0007	-0.0023	0.0003	-0.0022	0.0003
			sd	0.0129	0.0262	0.0107	0.0176	0.0068	0.0121	0.0047	0.0086
			CV	0.844	0.925	0.897	0.941	0.902	0.949	0.915	0.950
		$\hat{\beta}^{abc}$	bias	-0.0063	-0.0017	-0.0024	0.0003	-0.0003	0.0003	-0.0003	0.0003
			sd	0.0128	0.0263	0.0107	0.0176	0.0068	0.0121	0.0047	0.0086
			CV	0.896	0.925	0.926	0.941	0.934	0.950	0.947	0.950
	4	$\hat{\beta}$	bias	-0.0124	0.0023	-0.0080	-0.0015	-0.0037	0.0003	-0.0041	-0.0002
			sd	0.0132	0.0249	0.0090	0.0176	0.0065	0.0124	0.0046	0.0088
			CV	0.819	0.936	0.843	0.936	0.894	0.942	0.846	0.944
		$\hat{\beta}^{abc}$	bias	-0.0097	0.0023	-0.0037	-0.0013	-0.0013	0.0002	-0.0011	-0.0002
			sd	0.0132	0.0249	0.0090	0.0176	0.0065	0.0124	0.0046	0.0088
			CV	0.857	0.936	0.924	0.936	0.934	0.942	0.930	0.943
2	1	$\hat{\beta}$	bias	-0.0261	-0.0023	-0.0083	-0.0009	-0.0067	-0.0003	-0.0054	0.0002
			sd	0.0206	0.0270	0.0126	0.0178	0.0085	0.0125	0.0060	0.0089
			CV	0.635	0.931	0.866	0.940	0.850	0.943	0.835	0.946
		$\hat{\beta}^{abc}$	bias	-0.0147	-0.0014	-0.0028	-0.0008	-0.0020	-0.0002	-0.0014	0.0002
			sd	0.0197	0.0271	0.0125	0.0178	0.0085	0.0125	0.0060	0.0089
			CV	0.833	0.928	0.926	0.943	0.932	0.945	0.945	0.946
	2	$\hat{\beta}$	bias	-0.0378	-0.0012	-0.0114	0.0000	-0.0079	-0.0016	-0.0060	-0.0002
			sd	0.0252	0.0337	0.0161	0.0235	0.0102	0.0147	0.0071	0.0098
			CV	0.524	0.910	0.842	0.917	0.851	0.942	0.843	0.955
		$\hat{\beta}^{abc}$	bias	-0.0222	0.0002	-0.0031	0.0001	-0.0021	-0.0016	-0.0013	-0.0002
			sd	0.0241	0.0340	0.0158	0.0235	0.0101	0.0147	0.0070	0.0098
			CV	0.754	0.909	0.922	0.918	0.935	0.941	0.933	0.954
	3	$\hat{\beta}$	bias	-0.0119	0.0010	-0.0111	-0.0006	-0.0037	0.0005	-0.0033	-0.0002
			sd	0.0183	0.0273	0.0126	0.0176	0.0087	0.0127	0.0059	0.0090
			CV	0.870	0.936	0.820	0.945	0.912	0.946	0.901	0.953
		$\hat{\beta}^{abc}$	bias	-0.0046	0.0012	-0.0026	-0.0004	-0.0008	0.0005	-0.0008	-0.0002
			sd	0.0178	0.0273	0.0125	0.0176	0.0086	0.0127	0.0058	0.0090
			CV	0.909	0.936	0.931	0.946	0.940	0.948	0.945	0.953
	4	$\hat{\beta}$	bias	-0.0446	-0.0047	-0.0189	-0.0006	-0.0070	-0.0007	-0.0065	0.0003
			sd	0.0199	0.0263	0.0129	0.0185	0.0087	0.0119	0.0059	0.0090
			CV	0.297	0.927	0.632	0.947	0.854	0.951	0.785	0.949
		$\hat{\beta}^{abc}$	bias	-0.0240	-0.0034	-0.0080	-0.0004	-0.0016	-0.0007	-0.0017	0.0003
			sd	0.0191	0.0264	0.0126	0.0186	0.0086	0.0119	0.0059	0.0090
			CV	0.694	0.932	0.879	0.947	0.929	0.951	0.940	0.948

To account for cross-sectional dependence, we employ IFEs. To capture the persistence in tax, we utilize dynamic panel models. The state-level income per capita is included as an independent variable. Specifically, we consider the following four panel models with IFEs and potential missing values:

$$\mathrm{lgas}_{it} = d_{it}(\lambda_i' f_t + \beta_1 \mathrm{lgas}_{i,t-1} + \beta_2 \mathrm{ginc}_{it} + v_{it}), \tag{7.1}$$

$$\mathrm{lcig}_{it} = d_{it}(\lambda_i' f_t + \beta_1 \mathrm{lcig}_{i,t-1} + \beta_2 \mathrm{ginc}_{it} + v_{it}), \tag{7.2}$$

$$\mathrm{lwine}_{it} = d_{it}(\lambda_i' f_t + \beta_1 \mathrm{lwine}_{i,t-1} + \beta_2 \mathrm{ginc}_{it} + v_{it}), \tag{7.3}$$

$$\mathrm{lcorp}_{it} = d_{it}(\lambda_i' f_t + \beta_1 \mathrm{lcorp}_{i,t-1} + \beta_2 \mathrm{ginc}_{it} + v_{it}). \tag{7.4}$$

In the models above, lgas_{it} is the logarithm of gasoline unit tax for state i at year t, $\mathrm{lgas}_{i,t-1}$ is the logarithm of time lag term for the gasoline unit tax, and similar definitions apply for lcig_{it}, $\mathrm{lcig}_{i,t-1}$, lwine_{it} and $\mathrm{lwine}_{i,t-1}$. Moreover, lcorp_{it} is the logarithm of state-level corporate top tax rate with $\mathrm{lcorp}_{i,t-1}$ being the corresponding time lag term. "ginc_{it}" is the growth rate of income per capita for each state, which is included as the state-level control variable. v_{it} is the error term, λ_i and f_t are the loadings and factors, respectively. d_{it} is the missing indicator. β_1 and β_2 are the corresponding slopes for each regression model, with β_2 indicating the income elasticity on the taxes.

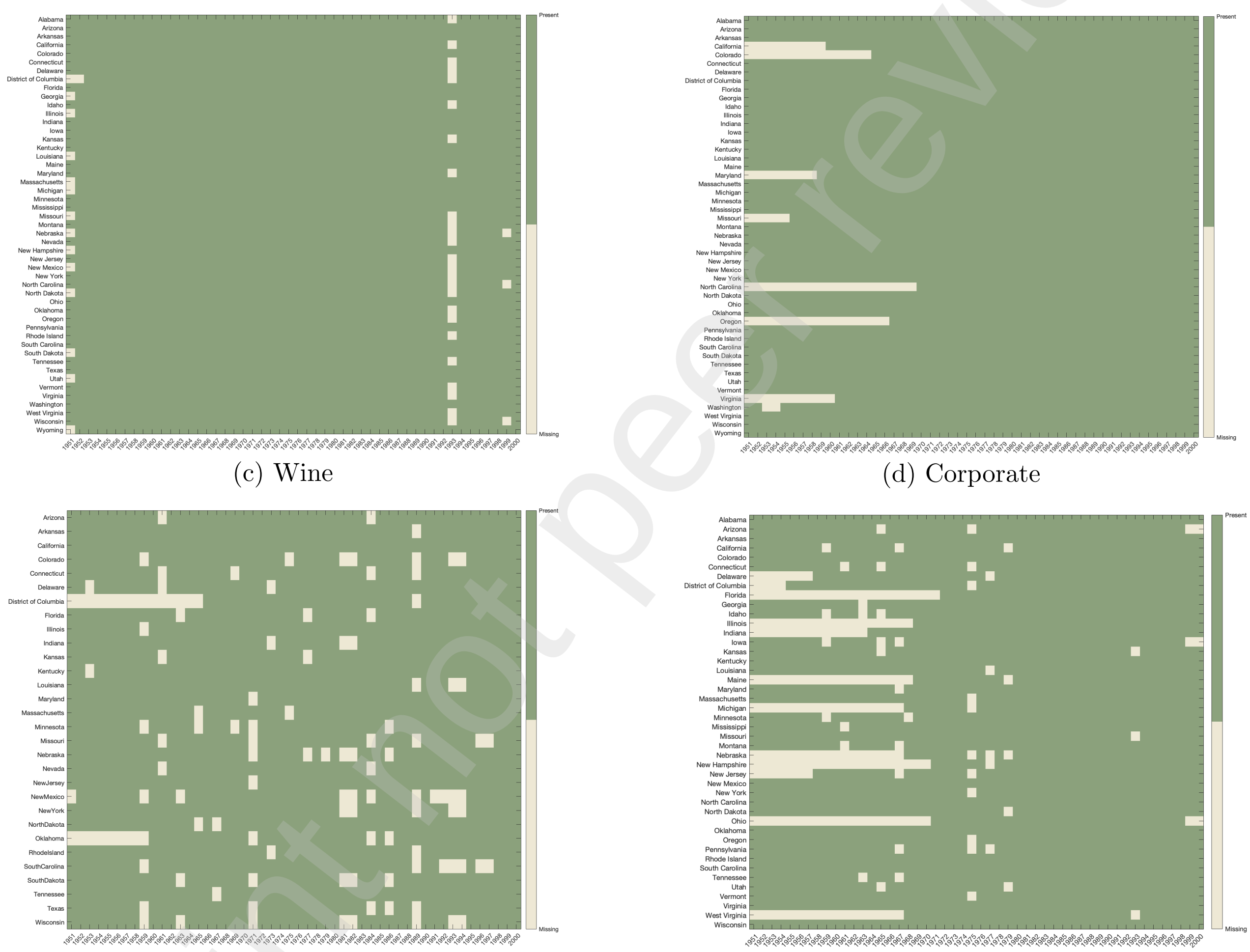
7.2 Data

We collect the tax data for gasoline, cigarettes, wine, and corporate from the World Tax Database maintained by the Office of Tax Policy Research at the University of Michigan. State-level income per capita data is sourced from the Bureau of Economic Analysis. Following Devereux et al. (2007), we deflate unit taxes using the 1982 CPI.

For the analysis of gasoline and cigarette taxes, we construct a dataset covering 49 U.S. states from 1951 to 2000, excluding Alaska and Hawaii, as they became states only after 1959. For wine and corporate taxes, after cleaning the data, the dataset includes 30 and 44 states over the same time period, as several states lack data for the entire time periods. The data reveal missing values for gasoline tax (1.76%), cigarette tax (3.39%), wine tax (8.47%) and corporate top tax (10.59%).

Figure 1 illustrates the missing data patterns for the four taxes, with missing values indicated in beige color. For gasoline taxes, missing data occur only in specific years: 1951, 1993, and 1999. Similarly, cigarette tax data exhibit minimal gaps, with only a few states (e.g., California, Colorado, Maryland) lacking data in the earlier years. In contrast, the wine tax and corporate top tax datasets exhibit significantly more missing data. For the wine tax, the gaps are both persistent throughout the timeline for certain states (e.g., the District of Columbia) and concentrated in specific time periods. The corporate tax dataset combines characteristics of both the cigarette and wine tax datasets, with more states missing data in the earlier years and certain specific years showing scattered missing data points. This makes the corporate tax dataset the most incomplete among the four.

Figure 1: Missing patterns for the tax data



(a) Gasoline (b) Cigarette (c) Wine (d) Corporate

7.3 Estimation and Inference Results for the Unbalanced Panel

Using the SVT approach described in Section 3.3, the estimated number of factors for all four models remain one. Table 6 presents the estimation results, where the columns correspond to the gasoline, cigarette, wine, and corporate tax models (7.1) – (7.4), respectively.

In this table, the row labeled "$\text{tax}_{i,t-1}$" represents the lagged dependent variable in each regression. Results are reported both with and without bias correction, as discussed in Section 4.

The estimated elasticity coefficients for the lagged tax are consistently high across all excise taxes, indicating strong persistence over time. While it becomes weaker for the corporate tax. The estimates range from 0.334 (corporate tax) to 0.971 (cigarette tax), all statistically significant at the 1% level. Bias correction has only minor effects on these estimates, suggesting the relatively small finite-sample bias.

Regarding the growth rate of state income, the estimated effects vary across tax types. Bias correction leads to more negative estimates for gasoline and cigarette taxes, with coefficients of -0.817 and -0.310, respectively, both statistically significant at the 1% and 5% levels. The wine tax exhibits a positive and significant relationship with income growth (0.579, significant at the 5% level), whereas the corporate tax shows a negative but statistically insignificant relationship. Especially, the bias correction helps to improve the significance level for the relationship of income and cigarette/wine taxes.

The differing effects of income growth across tax types likely reflect distinct economic and policy considerations. The negative relationship between income growth and gasoline/cigarette taxes may suggest that as a state economy grow, the state government may be less reliant on these two excise taxes, possibly shifting toward alternative revenue sources. In contrast, the positive effect on wine tax rates suggests a progressive taxation effect in states where wine consumption is more prevalent among higher-income populations. Moreover, the lack of significance for corporate tax rates implies that corporate tax policies may be driven by structural and political factors rather than short-term income fluctuations.

Table 6: Estimation results for US state tax with missing data

Variables	Bias correction	gas_{it}	cig_{it}	$wine_{it}$	$corp_{it}$
$tax_{i,t-1}$	yes	0.963 (214.41)***	0.971 (247.14)***	0.630 (21.87)***	0.334 (10.74)***
	no	0.964 (214.72)***	0.973 (247.53)***	0.625 (21.71)***	0.327 (10.53)***
ginc	yes	-0.817 $(-6.07)^{***}$	-0.310 $(-2.23)^{**}$	0.579 $(3.59)^{**}$	-0.163 (-0.75)
	no	-0.758 $(-5.63)^{***}$	-0.237 $(-1.71)^{*}$	0.524 $(3.21)^{*}$	-0.170 (-0.78)
	N	49	49	30	44
	T	50	50	50	50

Notes: Values in parentheses are the t-statistics. *, **, and ***denote significance at 10%, 5%, and 1%, respectively.

7.4 Comparison with the Results for the Balanced Sub-panel

To compare the results from our unbalanced dataset with a traditional balanced panel approach, we constructed a balanced panel by removing states with missing data from 1951 to 2000. After this data cleaning, the balanced dataset includes 15 states for gasoline tax, 41 states for cigarette tax, and 11 states for corporate tax. Due to data limitations after the truncation, the wine tax model is excluded from this balanced sample. For gasoline, cigarette, and corporate taxes, we apply the traditional panel data model with IFEs following the method in Moon and Weidner (2017). The estimation results are reported in Table 7.

For the lagged tax rate, the estimates in the balanced model are generally higher than those in the unbalanced model. The estimate for $\text{tax}_{i,t-1}$ increases from 0.963 and 0.971 in the unbalanced model to 0.99 and 0.981 in the balanced model for gasoline and cigarette taxes. However, for corporate tax, the estimate increases to 1.002, suggesting possible overestimation due to a smaller sample size ($N = 11$). The very high persistence of corporate tax rates (above 1) in the balanced model indicates potential estimation issues, reinforcing the advantage of using the missing data approach, which leverages a larger sample for more reliable estimates.

Regarding income growth, the balanced panel results differ significantly from the unbalanced model. The previously strong negative effect of income growth on gasoline tax (-0.817) becomes statistically insignificant (-0.164) in the balanced model. Similarly, for the cigarette tax, the negative income effect (-0.310, significant at 5%) disappears in the balanced model (0.035, insignificant). For corporate taxes, the unbalanced model shows a small negative effect (-0.163), but in the balanced panel, the estimate is nearly zero (0.007, insignificant). These differences suggest that restricting the data to a balanced panel can introduce bias and lead to misleading conclusions about the relationship between income growth and tax rates.

The comparison between our proposed AM algorithm for handling missing data and the traditional balanced panel approach highlights the advantages of incorporating all available data. The unbalanced panel approach captures meaningful economic relationships that are lost when using a traditional balanced panel. This suggests that removing data to create a balanced panel may introduce biases and obscure important economic dynamics, reinforcing the advantages of our AM algorithm.

8 Conclusions

This paper studies IFEs panel data models with missing data. A two-step procedure is proposed to estimate the regression coefficients and the factors and loadings, where in the first step we use nuclear norm regularization to obtain consistent initial estimates and in the second step we use

Table 7: Estimation results for US state tax with balanced data

Variables	Bias correction	gas_{it}	cig_{it}	corp_{it}
$\text{tax}_{i,t-1}$	yes	0.990 (62.60)***	0.981 (443.60)***	1.002 (498.56)***
	no	0.967 (61.16)***	0.981 (443.59)***	1.002 (498.54)***
ginc	yes	-0.164 (-1.04)	0.035 (0.55)	0.007 (0.11)
	no	-0.045 (-0.29)	0.035 (0.55)	0.010 (0.17)
	N	15	41	11
	T	50	50	50

Notes: Values in the parentheses are t-statistics. *** denote significance at 1%.

the alternating maximization (AM) to iterate until convergence. Under fairly general missing data patterns, we prove that the AM is a contractionary mapping towards the second step estimator and the second step estimator is asymptotically normal, as long as the initial estimator is consistent. We also show that the asymptotic biases and variances depend on the missing patterns and we develop analytical bias corrections according to the missing pattern. Monte Carlo simulations demonstrate excellent finite sample performance for the proposed estimation algorithm. An empirical application for the US state-level tax rates from year 1951 to 2000 with missing data shows that gasoline, cigarette, wine and corporate tax rate all exhibit persistence and state-level income growth affects various types of taxes differently.

Our results allow some important missing patterns, including block/staggered missing and selection on regressors/factors/loadings. Moreover, we also show that our results can be readily extended to cases with a Heckman correction term or other general settings such as nonlinear panels, two way fixed effects model and other missing patterns.

A series of work can be done based on our theoretical framework, e.g., testing the presence of sample selection bias in addition to the factor structure, extending our results to the fixed T cases or the nonlinear cases, allowing for nonstationarity in the data and allowing for potential slope heterogeneity. Our results are prototypical for the estimation and inference of unbalanced panels under general missing patterns and thus should be useful for these further studies.

References

Agarwal, A., Dahleh, M., Shah, D., and Shen, D. (2023). Causal matrix completion. In *The thirty sixth annual conference on learning theory*, pages 3821–3826. PMLR.

Ahn, S. C., Lee, Y. H., and Schmidt, P. (2001). GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics*, 101(2):219–255.

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

Ando, T. and Bai, J. (2020). Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity. *Journal of the American Statistical Association*, 115(529):266–279.

Arkhangelsky, D. and Imbens, G. (2024). Causal models for longitudinal and panel data: A survey. *Econometrics Journal*, 27(3):C1–C61.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536):1716–1730.

Athey, S. and Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, 226(1):62–79.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Bai, J. and Li, K. (2014). Theory and methods of panel data models with interactive effects. *Annals of Statistics*, 42(1):142 – 170.

Bai, J., Liao, Y., and Yang, J. (2015). Unbalanced panel data models with interactive effects. In *The Oxford Handbook of Panel Data*. Oxford University Press.

Bai, J. and Ng, S. (2021). Matrix completion, counterfactuals, and factor analysis of missing data. *Journal of the American Statistical Association*, 116(536):1746–1763.

Baltagi, B. H. and Song, S. H. (2006). Unbalanced panel data: A survey. *Statistical Papers*, 47(4):493–523.

Bańbura, M. and Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133–160.

Besley, T. and Case, A. (1995). Incumbent behavior: vote-seeking, tax-setting, and yardstick competition. *American Economic Review*, 85(1):25–45.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and dantzig selector. *Annals of Statistics*, 37(4):1705–1732.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *Review of Economic Studies*, 91(6):3253–3285.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122.

Brueckner, J. K. (2003). Strategic interaction among governments: An overview of empirical studies. *International Regional Science Review*, 26(2):175–188.

Callaway, B. and Karami, S. (2023). Treatment effects in interactive fixed effects models with a small number of time periods. *Journal of Econometrics*, 233(1):184–208.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Case, A. C., Rosen, H. S., and Hines Jr, J. R. (1993). Budget spillovers and fiscal policy interdependence: Evidence from the states. *Journal of Public Economics*, 52(3):285–307.

Chaloupka, F. J. and Warner, K. E. (2000). The economics of smoking. In *Handbook of Health Economics*.

Charbonneau, K. B. (2017). Multiple fixed effects in binary response panel data models. *Econometrics Journal*, 20(3):S1–S13.

Chen, L., Dolado, J. J., and Gonzalo, J. (2021a). Quantile factor models. *Econometrica*, 89(2):875–910.

Chen, M., Fernández-Val, I., and Weidner, M. (2021b). Nonlinear factor models for network and panel data. *Journal of Econometrics*, 220(2):296–324.

Chernozhukov, V., Hansen, C., Liao, Y., and Zhu, Y. (2023). Inference for low-rank models. *Annals of Statistics*, 51(3):1309–1330.

Chudik, A. and Pesaran, M. H. (2013). Large panel data models with cross-sectional dependence: a survey. *CAFE Research Paper*, (13.15).

Chudik, A. and Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, 188(2):393–420.

Czarnowske, D. and Stammann, A. (2020). Inference in unbalanced panel data models with interactive fixed effects. *arXiv preprint arXiv:2004.03414*.

de Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

Devereux, M. P., Lockwood, B., and Redoano, M. (2007). Horizontal and vertical indirect tax competition: Theory and some evidence from the USA. *Journal of Public Economics*, 91(3-4):451–479.

Dustmann, C. and Rochina-Barrachina, M. E. (2007). Selection correction in panel data models: An application to the estimation of females' wage equations. *Econometrics Journal*, 10(2):263–293.

Egger, P., Pfaffermayr, M., and Winner, H. (2005). An unbalanced spatial panel data approach to US state tax competition. *Economics Letters*, 88(3):329–335.

Fan, J. and Liao, Y. (2022). Learning latent factors from diversified projections and its applications to over-estimated and weak factors. *Journal of the American Statistical Association*, 117(538):909–924.

Fernández-Val, I. and Vella, F. (2011). Bias corrections for two-step fixed effects panel data estimators. *Journal of Econometrics*, 163(2):144–162.

Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192(1):291–312.

Fernández-Val, I. and Weidner, M. (2018). Fixed effects estimation of large-T panel data models. *Annual Review of Economics*, 10(1):109–138.

Golub, G. H. and Van Loan, C. F. (2013). *Matrix computations*. JHU press.

Gonçalves, S. and Perron, B. (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics*, 182(1):156–173.

Gonçalves, S. and Perron, B. (2020). Bootstrapping factor models with cross sectional dependence. *Journal of Econometrics*, 218(2):476–495.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Hackbusch, W. (2016). *Iterative Solution of Large Sparse Systems of Equations*, volume 95 of *Applied Mathematical Sciences*. Springer, second edition.

Hahn, J. and Moon, H. R. (2006). Reducing bias of mle in a dynamic panel model. *Econometric Theory*, 22(3):499–512.

Higgins, A. and Jochmans, K. (2024). Bootstrap inference for fixed-effect models. *Econometrica*, 92(2):411–427.

Holtz-Eakin, D., Newey, W., and Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica: Journal of the Econometric Society*, pages 1371–1395.

Hong, S., Su, L., and Jiang, T. (2023). Profile GMM estimation of panel data models with interactive fixed effects. *Journal of Econometrics*, 235(2):927–948.

Honoré, B. E. and Hu, L. (2004). Estimation of cross sectional and panel data censored regression models with endogeneity. *Journal of Econometrics*, 122(2):293–316.

Honoré, B. E. (1992). Trimmed lad and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica*, 60(3):533–565.

Hu, L. (2002). Estimation of a censored dynamic panel data model. *Econometrica*, 70(6):2499–2517.

Jiang, B., Yang, Y., Gao, J., and Hsiao, C. (2021). Recursive estimation in large panel data models: Theory and practice. *Journal of Econometrics*, 224(2):439–465.

Jin, S., Miao, K., and Su, L. (2021). On factor models with random missing: Em estimation, inference, and cross validation. *Journal of Econometrics*, 222(1):745–777.

Juodis, A. and Sarafidis, V. (2022). A linear estimator for factor-augmented fixed-T panels with endogenous regressors. *Journal of Business & Economic Statistics*, 40(1):1–15.

Kim, M. S. and Sun, Y. (2016). Bootstrap and k-step bootstrap bias corrections for the fixed effects estimator in nonlinear panel data models. *Econometric Theory*, 32(6):1523–1568.

Klopp, O. (2014). Noisy low-rank matrix completion with general sampling distribution. *Bernoulli*, 20(1):282–303.

Koltchinskii, V. (2011). Oracle inequalities in empirical risk minimization and sparse recovery problems, lecture notes in mathematics.

Kyriazidou, E. (1997). Estimation of a panel data sample selection model. *Econometrica: Journal of the Econometric Society*, pages 1335–1364.

Kyriazidou, E. (2001). Estimation of dynamic panel data sample selection models. *Review of Economic Studies*, 68(3):543–572.

Latała, R. (2005). Some estimates of norms of random matrices. *Proceedings of the American Mathematical Society*, 133(5):1273–1282.

Lu, X. and Su, L. (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics*, 190(1):148–175.

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica: Journal of the Econometric Society*, pages 357–362.

Moon, H. R. and Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4):1543–1579.

Moon, H. R. and Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1):158–195.

Moon, H. R. and Weidner, M. (2023). Nuclear norm regularized estimation of panel regression models.

Mugnier, M. (2024). A simple and computationally trivial estimator for grouped fixed effects models. *arXiv preprint arXiv:2203.08879*.

Negahban, S. and Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics*, 39(2):1069 – 1097.

Negahban, S. and Wainwright, M. J. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *Journal of Machine Learning Research*, 13(1):1665–1697.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Semykina, A. and Wooldridge, J. M. (2010). Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics*, 157(2):375–380.

Semykina, A. and Wooldridge, J. M. (2013). Estimation of dynamic panel data models with sample selection. *Journal of Applied Econometrics*, 28(1):47–61.

Shi, W. and Lee, L. F. (2017). Spatial dynamic panel data models with interactive fixed effects. *Journal of Econometrics*, 197(2):323–347.

Su, L. and Wang, F. (2024). Inference for large dimensional factor models under general missing data patterns. *Available at SSRN 4721159*.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Verbeek, M. and Nijman, T. (1996). Incomplete panels and selection bias. In *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, pages 449–490. Springer.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press.

Wang, Y., Su, L., and Zhang, Y. (2022). Low-rank panel quantile regression: Estimation and inference. *arXiv preprint arXiv:2210.11062*.

Westerlund, J. and Urbain, J.-P. (2015). Cross-sectional averages versus principal components. *Journal of Econometrics*, 185(2):372–377.

Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, 68(1):115–132.

Wooldridge, J. M. (2019). Correlated random effects models with unbalanced panels. *Journal of Econometrics*, 211(1):137–150.

Online Appendix for
"Estimation and Inference for Unbalanced Panel Data Models with Interactive Fixed Effects"

Liangjun Su[a], Fa Wang[b] and Yiren Wang[c]
[a]School of Economics and Management, Tsinghua University, China
[b]Department of Finance, School of Economics, Peking University, China
[c]College of Finance and Statistics, Hunan University, China

This online appendix contains 9 sections. Section A contains details for the Hessian matrices used in the estimation. Section B contains the proofs of Proposition 3.1 and Theorem 3.1. Section C contains the proofs of Theorem 3.2 and Corollary 3.1. Section D contains the proof Theorem 3.3. Section E contains the proofs of Proposition 4.1 and Theorem 4.1. Section F contains some supplementary lemmas used in the proofs of the results in Section E. Section G contains the proof of Theorem 4.2. Section H contains details for the Alternating Direction Method of Multipliers Algorithm (ADMM).

A Details for the Hessian Matrices

Recall that $\gamma = (\beta', \lambda', f')'$ and $Q(\gamma) = L(\gamma) + G(\gamma)$, where $L(\cdot)$ and $G(\cdot)$ are defined in (3.1) and (3.2), respectively. We use the following notations to define the Hessian matrix associated with $Q(\gamma)$.

$$Q_{\gamma\gamma'}(\gamma) = \partial_{\gamma\gamma'}Q(\gamma) = \begin{pmatrix} \partial_{\beta\beta'}Q(\gamma) & \partial_{\beta\phi'}Q(\gamma) \\ \partial_{\phi\beta'}Q(\gamma) & \partial_{\phi\phi'}Q(\gamma) \end{pmatrix} = \begin{pmatrix} Q_{\beta\beta'}(\gamma) & Q_{\beta\phi'}(\gamma) \\ Q_{\phi\beta'}(\gamma) & Q_{\phi\phi'}(\gamma) \end{pmatrix}, \tag{A.1}$$

$$Q_{\beta\beta'}(\gamma) = -\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}x_{it}x_{it}' \text{ for any } \gamma, \tag{A.2}$$

$$Q_{\beta\phi'}(\gamma) = (Q_{\beta\lambda_1'}(\gamma), ..., Q_{\beta\lambda_N'}(\gamma); Q_{\beta f_1'}(\gamma), ..., Q_{\beta f_T'}(\gamma)), \tag{A.3}$$

where $Q_{\beta\lambda_i'}(\gamma) = -\sum_{t=1}^{T} d_{it}x_{it}f_t'$ and $Q_{\beta f_t'}(\gamma) = -\sum_{i=1}^{N} d_{it}x_{it}\lambda_i'$. For $Q_{\phi\phi'}(\gamma)$, we make the following decomposition:

$$Q_{\phi\phi'}(\gamma) = \partial_{\phi\phi'}L(\gamma) + G_{\phi\phi'}(\gamma), \tag{A.4}$$

where $\partial_{\phi\phi'}L(\gamma) = L_{\phi\phi'}(\gamma) + J_{\phi\phi'}(\gamma)$ and $\partial_{\phi\phi'}G(\gamma) = G_{\phi\phi'}(\gamma)$. We make the following partitions:

$$Q_{\phi\phi'}(\gamma) = \begin{pmatrix} Q_{\lambda\lambda'}(\gamma) & Q_{\lambda f'}(\gamma) \\ Q_{f\lambda'}(\gamma) & Q_{ff'}(\gamma) \end{pmatrix}, \tag{A.5}$$

$$L_{\phi\phi'}(\gamma) = \begin{pmatrix} L_{\lambda\lambda'}(\gamma) & L_{\lambda f'}(\gamma) \\ L_{f\lambda'}(\gamma) & L_{ff'}(\gamma) \end{pmatrix} \text{ and } J_{\phi\phi'}(\gamma) = \begin{pmatrix} 0 & J_{\lambda f'}(\gamma) \\ J_{f\lambda'}(\gamma) & 0 \end{pmatrix}. \tag{A.6}$$

Here, $L_{\lambda\lambda'}(\gamma)$ is an $Nr \times Nr$ block-diagonal matrix and the i-th diagonal block is $-\sum_{t=1}^{T} d_{it}f_tf_t'$; $L_{ff'}(\gamma)$ is a $Tr \times Tr$ block-diagonal matrix and the t-th diagonal block is $-\sum_{i=1}^{N} d_{it}\lambda_i\lambda_i'$; $L_{\lambda f'}(\gamma)$ is of dimension

$Nr \times Tr$ and the (i,t)-th block is $-d_{it}f_t\lambda_i'$; $L_{f\lambda'}(\gamma)$ is the transpose of $L_{\lambda f'}(\gamma)$. $J_{\lambda f'}(\gamma)$ is also $Nr \times Tr$ and the (i,t)-th block is $d_{it}v_{it}(\gamma)I_r$ with $v_{it}(\gamma) = y_{it} - x_{it}\beta - f_t'\lambda_i$, and $J_{f\lambda'}(\gamma) = J_{\lambda f'}(\gamma)'$. When $\gamma = \gamma^0$, we suppress the argument of these matrices and write, e.g., $Q_{\phi\phi'} = Q_{\phi\phi'}(\gamma^0)$.

In addition, we define the following sets of $(Nr + Tr) \times 1$ vectors for $1 \leq p \leq q \leq r$.

w_{pp}^0 : For the first Nr elements, in the i-th block, the p-th element is λ_{ip}^0 and all the other elements are zeros; for the last Tr elements, in the t-th block, the p-th element is $-f_{tp}^0$ and all the other elements are zeros.

w_{pq}^0 : For the first Nr elements, in the i-th block, the p-th element is λ_{iq}^0 and all the other elements are zeros; for the last Tr elements, in the t-th block, the q-th element is $-f_{tp}^0$ and all the other elements are zeros.

w_{qp}^0 : For the first Nr elements, in the i-th block, the q-th element is λ_{ip}^0 and all the other elements are zero; for the last Tr elements, in the t-th block, the p-th element is $-f_{tq}^0$ and all the other elements are zero.

Let W^0 contain the vectors w_{pp}^0, w_{pq}^0 and w_{qp}^0, i.e.,

$$\begin{aligned} W^0 &= (w_{11}^0, ..., w_{rr}^0; w_{12}^0, ..., w_{1r}^0, w_{23}^0, ..., w_{2r}^0, ..., w_{(r-1)r}^0; w_{21}^0, ..., w_{r1}^0, w_{32}^0, ..., w_{r2}^0, ..., w_{r(r-1)}^0) \\ &\equiv (W_\lambda^{0\prime}, W_f^{0\prime})', \end{aligned} \tag{A.7}$$

where W_λ^0 (an $Nr \times r^2$ matrix) contains the first Nr rows of W^0 and W_f^0 ((a $Tr \times r^2$ matrix)) contains the last Tr rows. From expression (3.2) it's not difficult to see that

$$\begin{aligned} G(\lambda, f) &= -\frac{c}{2}\left\|\sqrt{NT}(D_{NT}^{-\frac{1}{2}}W^0)'D_{NT}^{-\frac{1}{2}}\phi\right\|^2 = -\frac{c}{2}\left\|W^{0\prime}D_{NT}^{-\frac{1}{2}}D_{TN}^{\frac{1}{2}}\phi\right\|^2 \\ &= -\frac{c}{2}\phi' D_{TN}^{\frac{1}{2}}D_{NT}^{-\frac{1}{2}}W^0W^{0\prime}D_{NT}^{-\frac{1}{2}}D_{TN}^{\frac{1}{2}}\phi, \text{ and hence} \\ G_{\phi\phi'} &= \partial_{\phi\phi'}G(\lambda, f) = -cD_{TN}^{\frac{1}{2}}D_{NT}^{-\frac{1}{2}}W^0W^{0\prime}D_{NT}^{-\frac{1}{2}}D_{TN}^{\frac{1}{2}}. \end{aligned} \tag{A.8}$$

A cautionary note. For a real square matrix A, we will A^{-1} to denote its usual inverse when A is of full rank, and the Moore-Penrose generalized inverse (A^+) if A is not of full rank. The Hessian matrix $L_{\phi\phi'}(\cdot)$ associated with $L(\cdot)$ is has rank $(N+T)r - r^2$, whereas the Hessian matrix $Q_{\phi\phi'}$ associated with $Q(\cdot)$ is full rank (see Lemma B.1 below).

B Proofs of Proposition 3.1 and Theorem 3.1

Lemma B.1 *Suppose Assumptions 1–2 and 4 hold. Then as* $(N,T) \to \infty$, $(-D_{TN}^{-\frac{1}{2}}Q_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1} = O_p(1)$.

Proof. For the random missing case, Lemma B.1 in Su and Wang (2024) (see steps (1.1)-(1.3) in particular) proves this result under their Assumptions 1, 2(i) and 4. Note that $Q_{\phi\phi'}$ here corresponds to $\check{H}_{\phi\phi'} + (L_{\phi\phi'} - \bar{L}_{\phi\phi'}) + J_{\phi\phi'}$ there. For the block-type missing case, Lemma B.2 in Su and Wang (2024) proves this result under their Assumptions 1, 2(ii) and 4.

The proofs are quite different under random missing and block-type missing because the structure of $Q_{\phi\phi'}$ under these two types of missing patterns are quite different. A key condition for proving $(-D_{TN}^{-\frac{1}{2}}Q_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1} =$

$O_p(1)$ in the random missing cases is that $\mathbb{E}_\phi(d_{it}) \geq \underline{d} > 0$ for all i and t. Nevertheless, this condition is violated in the block missing cases where d_{it} is always 1 if $i \leq N_o$ or $t \leq T_o$ and always 0 if $i > N_o$ and $t > T_o$. Because of this fundamental difference, Su and Wang (2024) prove $(-D_{TN}^{-\frac{1}{2}} Q_{\phi\phi'} D_{TN}^{-\frac{1}{2}})^{-1} = O_p(1)$ conditioning on the block missing pattern, utilizing a totally different strategy.

Given the factors, loadings and the block missing pattern are realized and fixed down, Su and Wang (2024) calculate exactly all the eigenvalues and eigenvectors of the normalized Hessian and show that the smallest eigenvalue is bounded away from zero in probability as long as both N_o/N and T_o/T are bounded away from zero. For the block-type (e.g., staggered) missing cases, they show that the eigenvalues of the normalized Hessian become smaller if some entries of the data matrix are thrown away so that it becomes an exact block missing matrix. Thus $(-D_{TN}^{-\frac{1}{2}} Q_{\phi\phi'} D_{TN}^{-\frac{1}{2}})^{-1} = O_p(1)$ holds for the block-type missing cases as long as it holds for the block missing case. ■

Proof of Proposition 3.1:

Part (i): Recall that the i-th diagonal block of $L_{\lambda\lambda'}(\gamma)$ is $-\sum_{t=1}^T d_{it} f_t f_t'$, the t-th diagonal block of $L_{ff'}(\gamma)$ is $-\sum_{i=1}^N d_{it}\lambda_i\lambda_i'$, the (i,t)-th block of $L_{\lambda f'}(\gamma)$ is $-d_{it} f_t \lambda_i'$ and the (i,t)-th block of $J_{\lambda f'}(\gamma)$ is $d_{it} v_{it}(\gamma) I_r$, where $v_{it}(\gamma) = y_{it} - x_{it}'\beta - f_t'\lambda_i$. When $\gamma \in \mathcal{B}_m(\gamma^0)$, we have $\left\| D_{NT}^{-\frac{1}{2}}(\phi - \phi^0) \right\| \leq m$. Then by using $a^2 - b^2 = 2(a-b)b + (a-b)^2$, $ab - a^0 b^0 = (a - a^0)b^0 + a^0(b - b^0) + (a - a^0)(b - b^0)$ and the Cauchy-Schwarz (CS hereafter) inequality, we obtain

$$\left| \frac{1}{N}\sum_{i=1}^N \lambda_{iq}^2 - \frac{1}{N}\sum_{i=1}^N (\lambda_{iq}^0)^2 \right| \leq 2m\sqrt{\frac{1}{N}\sum_{i=1}^N (\lambda_{iq}^0)^2} + m^2 \text{ for any } q, \tag{B.1}$$

$$\left| \frac{1}{N}\sum_{i=1}^N \lambda_{ip}\lambda_{iq} - \frac{1}{N}\sum_{i=1}^N \lambda_{ip}^0\lambda_{iq}^0 \right| \leq 2m\max_q\sqrt{\frac{1}{N}\sum_{i=1}^N (\lambda_{iq}^0)^2} + m^2 \text{ for } p \neq q, \tag{B.2}$$

$$\left| \frac{1}{T}\sum_{t=1}^T f_{tq}^2 - \frac{1}{T}\sum_{t=1}^T (f_{tq}^0)^2 \right| \leq 2m\sqrt{\frac{1}{T}\sum_{t=1}^T (f_{tq}^0)^2} + m^2 \text{ for any } q, \tag{B.3}$$

$$\left| \frac{1}{T}\sum_{t=1}^T f_{tp}f_{tq} - \frac{1}{T}\sum_{t=1}^T f_{tp}^0 f_{tq}^0 \right| \leq 2m\max_q\sqrt{\frac{1}{T}\sum_{t=1}^T (f_{tq}^0)^2} + m^2 \text{ for } p \neq q, \tag{B.4}$$

$$\sqrt{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T [x_{it}'(\beta - \beta^0)]^2} \leq m\sqrt{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \|x_{it}\|^2}. \tag{B.5}$$

Thus by Assumptions 2 and 3(ii), there exists $M > 0$ such that w.p.a.1,

$$\max_{\gamma\in\mathcal{B}_m(\gamma^0)} \left\| \frac{1}{T}(L_{\lambda\lambda'}(\gamma) - L_{\lambda\lambda'}) \right\| \leq \max_{\gamma\in\mathcal{B}_m(\gamma^0)} \left\| \frac{1}{T}\sum_{t=1}^T d_{it} f_t f_t' - \frac{1}{T}\sum_{t=1}^T d_{it} f_t^0 f_t^{0\prime} \right\| \leq mM,$$
$$\max_{\gamma\in\mathcal{B}_m(\gamma^0)} \left\| \frac{1}{N}(L_{ff'}(\gamma) - L_{ff'}) \right\| \leq \max_{\gamma\in\mathcal{B}_m(\gamma^0)} \left\| \frac{1}{N}\sum_{i=1}^N d_{it}\lambda_i\lambda_i' - \frac{1}{N}\sum_{i=1}^N d_{it}\lambda_i^0\lambda_i^{0\prime} \right\| \leq mM,$$
$$\max_{\gamma\in\mathcal{B}_m(\gamma^0)} \left\| \frac{1}{\sqrt{NT}}(L_{\lambda f'}(\gamma) - L_{\lambda f'}) \right\|_F \leq \max_{\gamma\in\mathcal{B}_m(\gamma^0)} \sqrt{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \left\| d_{it} f_t\lambda_i' - d_{it} f_t^0\lambda_i^{0\prime} \right\|_F^2} \leq mM,$$
$$\max_{\gamma\in\mathcal{B}_m(\gamma^0)} \left\| \frac{1}{\sqrt{NT}}(J_{\lambda f'}(\gamma) - J_{\lambda f'}) \right\|_F \leq \max_{\gamma\in\mathcal{B}_m(\gamma^0)} \sqrt{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \left\| d_{it} v_{it}(\gamma) I_r - d_{it} v_{it} I_r \right\|_F^2} \leq mM.$$

Since $Q_{\phi\phi'}(\gamma)=L_{\phi\phi'}(\gamma)+J_{\phi\phi'}(\gamma)+G_{\phi\phi'}$ and $Q_{\phi\phi'}=L_{\phi\phi'}+J_{\phi\phi'}+G_{\phi\phi'}$, we have

$$\begin{aligned}
& \max_{\gamma\in\mathcal{B}_m(\gamma^0)}\left\|D_{TN}^{-\frac{1}{2}}(Q_{\phi\phi'}(\gamma)-Q_{\phi\phi'})D_{TN}^{-\frac{1}{2}}\right\| \\
= & \max_{\gamma\in\mathcal{B}_m(\gamma^0)}\left\|D_{TN}^{-\frac{1}{2}}(L_{\phi\phi'}(\gamma)-L_{\phi\phi'}+J_{\phi\phi'}(\gamma)-J_{\phi\phi'})D_{TN}^{-\frac{1}{2}}\right\| \\
\leq & \max_{\gamma\in\mathcal{B}_m(\gamma^0)}\left\{\left\|\frac{1}{T}(L_{\lambda\lambda'}(\gamma)-L_{\lambda\lambda'})\right\|+\left\|\frac{1}{N}(L_{ff'}(\gamma)-L_{ff'})\right\|+\left\|\frac{2}{\sqrt{NT}}(L_{\lambda f'}(\gamma)-L_{\lambda f'})\right\|\right. \\
& \left.+\left\|\frac{2}{\sqrt{NT}}(J_{\lambda f'}(\gamma)-J_{\lambda f'})\right\|\right\} \\
\leq & \; mM \text{ w.p.a.1.}
\end{aligned} \tag{B.6}$$

If we take $m\leq\frac{C}{2M}$ with $0<2C\leq$plim$\sigma_{\min}(-D_{TN}^{-\frac{1}{2}}Q_{\phi\phi'}D_{TN}^{-\frac{1}{2}})$ by Lemma B.1, then w.p.a.1.

$$\min_{\gamma\in\mathcal{B}_m(\gamma^0)}\sigma_{\min}(-D_{TN}^{-\frac{1}{2}}Q_{\phi\phi'}(\gamma)D_{TN}^{-\frac{1}{2}})\geq\sigma_{\min}(-D_{TN}^{-\frac{1}{2}}Q_{\phi\phi'}D_{TN}^{-\frac{1}{2}})-\max_{\gamma\in\mathcal{B}_m(\gamma^0)}\left\|D_{TN}^{-\frac{1}{2}}(Q_{\phi\phi'}(\gamma)-Q_{\phi\phi'})D_{TN}^{-\frac{1}{2}}\right\|\geq C.$$

Part (ii): First note that $Q_{\beta\beta'}(\gamma)=Q_{\beta\beta'}$ since $\sum_{t=1}^T\sum_{i=1}^N d_{it}x_{it}x_{it}'$ is free of γ. Next, by the CS inequality,

$$\begin{aligned}
& \max_{\gamma\in\mathcal{B}_m(\gamma^0)}\left\|(Q_{\beta\phi'}(\gamma)-Q_{\beta\phi'})D_{TN}^{-\frac{1}{2}}\right\| \\
\leq & \max_{\gamma\in\mathcal{B}_m(\gamma^0)}\sqrt{\sum_{i=1}^N\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T d_{it}x_{it}(f_t-f_t^0)'\right\|_F^2+\sum_{t=1}^T\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N d_{it}x_{it}(\lambda_i-\lambda_i^0)'\right\|_F^2} \\
\leq & \max_{\gamma\in\mathcal{B}_m(\gamma^0)}\sqrt{\sum_{i=1}^N\sum_{t=1}^T\|x_{it}\|^2(\frac{\|f-f^0\|^2}{T}+\frac{\|\lambda-\lambda^0\|^2}{N})}\leq mM\sqrt{NT}\text{ w.p.a.1,}
\end{aligned} \tag{B.7}$$

and

$$\left\|Q_{\beta\phi'}D_{TN}^{-\frac{1}{2}}\right\|\leq\sqrt{\sum_{i=1}^N\sum_{t=1}^T\|x_{it}\|^2(\frac{\|f^0\|^2}{T}+\frac{\|\lambda^0\|^2}{N})}\leq M\sqrt{NT}\text{ w.p.a.1.} \tag{B.8}$$

In addition,

$$\begin{aligned}
& \max_{\gamma\in\mathcal{B}_m(\gamma^0)}\left\|D_{TN}^{\frac{1}{2}}[(Q_{\phi\phi'}(\gamma))^{-1}-Q_{\phi\phi'}^{-1}]D_{TN}^{\frac{1}{2}}\right\| \\
= & \max_{\gamma\in\mathcal{B}_m(\gamma^0)}\left\|D_{TN}^{\frac{1}{2}}(Q_{\phi\phi'}(\gamma))^{-1}(Q_{\phi\phi'}(\gamma)-Q_{\phi\phi'})Q_{\phi\phi'}^{-1}D_{TN}^{\frac{1}{2}}\right\| \\
\leq & \max_{\gamma\in\mathcal{B}_m(\gamma^0)}\left\|(D_{TN}^{-\frac{1}{2}}Q_{\phi\phi'}(\gamma)D_{TN}^{-\frac{1}{2}})^{-1}\right\|\left\|D_{TN}^{-\frac{1}{2}}(Q_{\phi\phi'}(\gamma)-Q_{\phi\phi'})D_{TN}^{-\frac{1}{2}}\right\|\left\|(D_{TN}^{-\frac{1}{2}}Q_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}\right\| \\
\leq & \frac{mM}{C^2}\text{ w.p.a.1,}
\end{aligned} \tag{B.9}$$

where the last inequality follows from Lemma B.1, Proposition 3.1(i), and eqn. (B.6). Combining (B.7)–(B.9) and using

$$ABA'-A^0B^0A^{0\prime} = (A-A^0)B^0A^{0\prime}+A^0(B-B^0)A^{0\prime}+A^0B^0(A-A^0)'+(A-A^0)(B-B^0)A^{0\prime}$$

$$+(A-A^0)B^0(A-A^0)'+A^0(B-B^0)(A-A^0)'+(A-A^0)(B-B^0)(A-A^0)',$$

we can readily show that for some large $M>0$,

$$\begin{aligned}
&\max_{\gamma\in\mathcal{B}_m(\gamma^0)}\left\|Q_{\beta\phi'}(\gamma)(Q_{\phi\phi'}(\gamma))^{-1}Q_{\phi\beta'}(\gamma)-\frac{1}{NT}Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'}\right\| \\
=\;&\max_{\gamma\in\mathcal{B}_m(\gamma^0)}\left\|Q_{\beta\phi'}(\gamma)D_{TN}^{-\frac{1}{2}}D_{TN}^{\frac{1}{2}}(Q_{\phi\phi'}(\gamma))^{-1}D_{TN}^{\frac{1}{2}}D_{TN}^{-\frac{1}{2}}Q_{\phi\beta'}(\gamma)-Q_{\beta\phi'}D_{TN}^{-\frac{1}{2}}D_{TN}^{\frac{1}{2}}Q_{\phi\phi'}^{-1}D_{TN}^{\frac{1}{2}}D_{TN}^{-\frac{1}{2}}Q_{\phi\beta'}\right\| \\
\leq\;& mMNT \text{ w.p.a.1.}
\end{aligned}$$

In addition, Assumption 3(i) and Lemma B.2 below imply $\sigma_{\min}(-\frac{1}{NT}[Q_{\beta\beta'}-Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'}])\geq C$ w.p.a.1. This finishes the proof of part (ii) after taking $m\leq\frac{C}{2M}$. ■

Lemma B.2 *(i) Let $\bar{Q}_{\phi\phi'}=\mathbb{E}_\phi(Q_{\phi\phi'})$ and define the other expected counterparts similarly, e.g., $\bar{Q}_{\beta\beta'}=\mathbb{E}_\phi(Q_{\beta\beta'})$, $\bar{Q}_{\beta\phi'}=\mathbb{E}_\phi(Q_{\beta\phi'})$, $\bar{L}_{\phi\phi'}=\mathbb{E}_\phi(L_{\phi\phi'})$ and $\bar{J}_{\phi\phi'}=\mathbb{E}_\phi(J_{\phi\phi'})$. Suppose the conditions in Proposition 3.1 hold. Then $\sum_{i=1}^N\sum_{t=1}^T\mathbb{E}_\phi(d_{it}\dot{x}_{it}\dot{x}_{it}')=-(\bar{Q}_{\beta\beta'}-\bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}\bar{Q}_{\phi\beta'})$.*

(ii) Under Assumptions 1-4, we have $\frac{1}{NT}(Q_{\beta\beta'}-Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'})=\frac{1}{NT}(\bar{Q}_{\beta\beta'}-\bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}\bar{Q}_{\phi\beta'})+o_p(1)$ as $(N,T)\to\infty$.

Proof. The following proof is presented in an unified way for both the random missing case and the block-type missing case. For the latter, since our analysis is conditioning on the missing pattern, we have $\mathbb{E}_\phi(d_{it})=d_{it}$, $\mathbb{E}_\phi(d_{it}x_{it})=d_{it}\mathbb{E}_\phi(x_{it})$, $\mathbb{E}_\phi(d_{it}x_{it}x_{it}')=d_{it}\mathbb{E}_\phi(x_{it}x_{it}')$, etc.

Part (i): Step (1): The first order conditions (FOCs) for the minimization problem in (2.2) is that for each k,

$$\begin{aligned}
\sum\nolimits_{t=1}^T\mathbb{E}_\phi(d_{it})\Big(\frac{\mathbb{E}_\phi(d_{it}x_{itk})}{\mathbb{E}_\phi(d_{it})}-\delta_{ki}^{0\prime}f_t^0-\omega_{kt}^{0\prime}\lambda_i^0\Big)f_t^{0\prime} &= 0 \text{ for any } i,\\
\sum\nolimits_{i=1}^N\mathbb{E}_\phi(d_{it})\Big(\frac{\mathbb{E}_\phi(d_{it}x_{itk})}{\mathbb{E}_\phi(d_{it})}-\delta_{ki}^{0\prime}f_t^0-\omega_{kt}^{0\prime}\lambda_i^0\Big)\lambda_i^{0\prime} &= 0 \text{ for any } t.
\end{aligned}$$

Recall that δ_i^0 and ω_t^0 are $K\times r$ matrices such that δ_{ki}^0 and ω_{kt}^0 are the transpose of the k-th row of δ_i^0 and ω_t^0, respectively. Define $\delta^0=(\delta_1^0,...,\delta_N^0)'$ and $\omega^0=(\omega_1^0,...,\omega_T^0)'$. Then the above FOCs imply

$$\bar{Q}_{\beta\phi'}=(\delta^{0\prime},\omega^{0\prime})\bar{L}_{\phi\phi'}. \tag{B.10}$$

They also imply $\sum_{i=1}^N\sum_{t=1}^T\mathbb{E}_\phi(d_{it})(\frac{\mathbb{E}_\phi(d_{it}x_{it})}{\mathbb{E}_\phi(d_{it})}-\delta_i^0f_t^0-\omega_t^0\lambda_i^0)(\delta_i^0f_t^0+\omega_t^0\lambda_i^0)'=0$. It follows that

$$\begin{aligned}
&\sum\nolimits_{i=1}^N\sum\nolimits_{t=1}^T\mathbb{E}_\phi(d_{it})(\delta_i^0f_t^0+\omega_t^0\lambda_i^0)(\delta_i^0f_t^0+\omega_t^0\lambda_i^0)' \\
=\;&\sum\nolimits_{i=1}^N\sum\nolimits_{t=1}^T\mathbb{E}_\phi(d_{it}x_{it})(\delta_i^0f_t^0+\omega_t^0\lambda_i^0)'=-\bar{Q}_{\beta\phi'}\bar{L}_{\phi\phi'}^{-1}\bar{Q}_{\phi\beta'},
\end{aligned} \tag{B.11}$$

where the second equality follows from eqn. (B.10), eqn. (A.3), and the fact that $\mathbb{E}_\phi(d_{it}x_{it}\lambda_i^{0\prime})=\mathbb{E}_\phi(d_{it}x_{it})\lambda_i^{0\prime}$ and $\mathbb{E}_\phi(d_{it}x_{it}f_t^{0\prime})=\mathbb{E}_\phi(d_{it}x_{it})f_t^{0\prime}$. Since $\dot{x}_{it}=x_{it}-(\delta_i^0f_t^0+\omega_t^0\lambda_i^0)$,

$$\sum\nolimits_{i=1}^N\sum\nolimits_{t=1}^T\mathbb{E}_\phi(d_{it}\dot{x}_{it}\dot{x}_{it}')$$

$$
\begin{aligned}
&= \sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}_{\phi}(d_{it}x_{it}x_{it}') - 2\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}_{\phi}(d_{it}x_{it})(\delta_i^0 f_t^0 + \omega_t^0\lambda_i^0)' \\
&\quad + \sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}_{\phi}(d_{it})(\delta_i^0 f_t^0 + \omega_t^0\lambda_i^0)(\delta_i^0 f_t^0 + \omega_t^0\lambda_i^0)' = -(\bar{Q}_{\beta\beta'} - \bar{Q}_{\beta\phi'}\bar{L}_{\phi\phi'}^{-1}\bar{Q}_{\phi\beta'}). \qquad (B.12)
\end{aligned}
$$

Thus it remains to show that $\bar{Q}_{\beta\phi'}\bar{L}_{\phi\phi'}^{-1}\bar{Q}_{\phi\beta'} = \bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}\bar{Q}_{\phi\beta'}$.

Step (2): Recall that $\bar{Q}_{\phi\phi'} = \bar{L}_{\phi\phi'} + G_{\phi\phi'} = \bar{L}_{\phi\phi'} - cD_{TN}^{\frac{1}{2}}D_{NT}^{-\frac{1}{2}}W^0W^{0\prime}D_{NT}^{-\frac{1}{2}}D_{TN}^{\frac{1}{2}}$ and W^0 is defined in Appendix A. Any two different columns of W^0 are orthogonal to each other, and the columns of W^0 are all orthogonal to the space spanned by the eigenvectors of $L_{\phi\phi'}$ and orthogonal to the rows of $Q_{\beta\phi'}$. In matrix form, we have

$$
\begin{aligned}
L_{\phi\phi'}W^0 &= 0,\ Q_{\beta\phi'}W^0 = 0, \\
\bar{L}_{\phi\phi'}W^0 &= 0,\ \text{and } \bar{Q}_{\beta\phi'}W^0 = 0. \qquad (B.13)
\end{aligned}
$$

It's easy to see that any two columns of $D_{NT}^{-\frac{1}{2}}W^0$ are also orthogonal to each other, and the columns of $D_{NT}^{-\frac{1}{2}}W^0$ are all orthogonal to the eigenvectors of $D_{TN}^{-\frac{1}{2}}\bar{L}_{\phi\phi'}D_{TN}^{-\frac{1}{2}}$ (since $D_{TN}^{-\frac{1}{2}}\bar{L}_{\phi\phi'}D_{TN}^{-\frac{1}{2}}D_{NT}^{-\frac{1}{2}}W^0 = \frac{1}{\sqrt{NT}}D_{TN}^{-\frac{1}{2}}\bar{L}_{\phi\phi'}W^0 = 0$). Thus we have

$$
(D_{TN}^{-\frac{1}{2}}\bar{Q}_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1} = (D_{TN}^{-\frac{1}{2}}\bar{L}_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1} - \frac{1}{c}D_{NT}^{-\frac{1}{2}}W^0(W^{0\prime}D_{NT}^{-1}W^0)^{-2}W^{0\prime}D_{NT}^{-\frac{1}{2}}. \qquad (B.14)
$$

We also have $\bar{Q}_{\beta\phi'}D_{TN}^{-\frac{1}{2}}D_{NT}^{-\frac{1}{2}}W^0 = \frac{1}{\sqrt{NT}}\bar{Q}_{\beta\phi'}W^0 = 0$, it follows that

$$
\begin{aligned}
\bar{Q}_{\beta\phi'}^{-1}\bar{Q}_{\phi\phi'}^{-1} &= \bar{Q}_{\beta\phi'}D_{TN}^{-\frac{1}{2}}(D_{TN}^{-\frac{1}{2}}\bar{Q}_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}D_{TN}^{-\frac{1}{2}} \\
&= \bar{Q}_{\beta\phi'}D_{TN}^{-\frac{1}{2}}(D_{TN}^{-\frac{1}{2}}\bar{L}_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}D_{TN}^{-\frac{1}{2}} = \bar{Q}_{\beta\phi'}\bar{L}_{\phi\phi'}^{-1}. \qquad (B.15)
\end{aligned}
$$

Then $\bar{Q}_{\beta\phi'}\bar{L}_{\phi\phi'}^{-1}\bar{Q}_{\phi\beta'} = \bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}\bar{Q}_{\phi\beta'}$. This concludes the proof of part (i).

Part (ii): It suffices to show that $\frac{1}{NT}(Q_{\beta\beta'} - \bar{Q}_{\beta\beta'}) = o_p(1)$ and $\frac{1}{NT}(Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'} - \bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}\bar{Q}_{\phi\beta'}) = o_p(1)$. We only prove the second claim as the first one is implied by Assumption 3(iii) (see (B.19) below). To show the second claim, given Lemma B.1, it suffices to show

$$
\begin{aligned}
\left\|(D_{TN}^{-\frac{1}{2}}Q_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1} - (D_{TN}^{-\frac{1}{2}}\bar{Q}_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}\right\| &= o_p(1), \qquad (B.16) \\
\left\|Q_{\beta\phi'}D_{TN}^{-\frac{1}{2}}\right\| &= O_p(\sqrt{NT}), \qquad (B.17) \\
\left\|(Q_{\beta\phi'} - \bar{Q}_{\beta\phi'})D_{TN}^{-\frac{1}{2}}\right\| &= O_p(\sqrt{N} + \sqrt{T}), \qquad (B.18) \\
\left\|Q_{\beta\beta'} - \bar{Q}_{\beta\beta'}\right\| &= o_p(NT). \qquad (B.19)
\end{aligned}
$$

Eqn. (B.17) holds by eqn. (B.8). Equations (B.18) and (B.19) follow from Assumption 3(iii). Now consider eqn. (B.16). Equations (B.16) and (B.15) in the appendix of Su and Wang (2024) show that $\left\|D_{TN}^{-\frac{1}{2}}(L_{\phi\phi'} - \bar{L}_{\phi\phi'})D_{TN}^{-\frac{1}{2}}\right\| = O_p(\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}} + \frac{T^{\frac{1}{\kappa}}}{\sqrt{N}} + \frac{1}{\sqrt{c_{NT}}})$ and $\left\|D_{TN}^{-\frac{1}{2}}J_{\phi\phi'}D_{TN}^{-\frac{1}{2}}\right\| = O_p(\frac{1}{\sqrt{c_{NT}}})$ for the random missing cases. Since $Q_{\phi\phi'} - \bar{Q}_{\phi\phi'} = L_{\phi\phi'} - \bar{L}_{\phi\phi'} + J_{\phi\phi'}$, we have

$$
\left\|D_{TN}^{-\frac{1}{2}}(Q_{\phi\phi'} - \bar{Q}_{\phi\phi'})D_{TN}^{-\frac{1}{2}}\right\| = O_p(\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}} + \frac{T^{\frac{1}{\kappa}}}{\sqrt{N}} + \frac{1}{\sqrt{c_{NT}}}), \qquad (B.20)
$$

which together with Lemma B.1 proves eqn. (B.16) for the random missing cases. For the block-type missing cases, $L_{\phi\phi'} = \bar{L}_{\phi\phi'}$ since $\mathbb{E}_\phi(d_{it}) = d_{it}$, i.e., eqn. (B.20) is still valid with $\kappa = \infty$. ■

Proof of Theorem 3.1:

Step (1): Let $\hat{\gamma} = (\hat{\beta}', \hat{\lambda}', \hat{f}')'$ be the solution of the problem $\min_{\gamma \in \mathcal{B}_m(\gamma^0)} \|S(\gamma)\|^2$. Since $\gamma^0 \in \mathcal{B}_m(\gamma^0)$, $\|S(\hat{\gamma})\| \leq \|S\|$ by definition. Take the Taylor expansion of $S(\hat{\gamma})$ at γ^0,

$$
\begin{aligned}
S_\beta(\hat{\beta}, \hat{\phi}) - S_\beta &= Q^*_{\beta\beta'}(\hat{\beta} - \beta^0) + Q^*_{\beta\phi'}(\hat{\phi} - \phi^0), &\text{(B.21)}\\
S_\phi(\hat{\beta}, \hat{\phi}) - S_\phi &= Q^*_{\phi\beta'}(\hat{\beta} - \beta^0) + Q^*_{\phi\phi'}(\hat{\phi} - \phi^0), &\text{(B.22)}
\end{aligned}
$$

where $Q^*_{\beta\beta'} = Q_{\beta\beta'}(\gamma^*)$, $Q^*_{\beta\phi'} = Q_{\beta\phi'}(\gamma^*)$, $Q^*_{\phi\beta'} = Q^{*\prime}_{\beta\phi'}$, $Q^*_{\phi\phi'} = Q_{\phi\phi'}(\gamma^*)$ and $\gamma^* = s\hat{\gamma} + (1-s)\gamma^0$ for some $0 < s < 1$. It follows that

$$
\begin{aligned}
\hat{\beta} - \beta^0 &= (Q^*_{\beta\beta'} - Q^*_{\beta\phi'}Q^{*-1}_{\phi\phi'}Q^*_{\phi\beta'})^{-1}(S_\beta(\hat{\beta}, \hat{\phi}) - S_\beta - Q^*_{\beta\phi'}Q^{*-1}_{\phi\phi'}(S_\phi(\hat{\beta}, \hat{\phi}) - S_\phi))\\
&= O_p(\frac{1}{NT})(O_p(\sqrt{NT}) + O_p(\sqrt{NT})O_p(1)O_p(\sqrt{T} + \sqrt{N})) = O_p(\frac{1}{c_{NT}}), &\text{(B.23)}
\end{aligned}
$$

where the second equality holds by facts (1)–(5) below.

(1) By Proposition 3.1(ii),

$$
\begin{aligned}
&\left\|(\frac{1}{NT}[Q^*_{\beta\beta'} - Q^*_{\beta\phi'}Q^{*-1}_{\phi\phi'}Q^*_{\phi\beta'}])^{-1}\right\| = 1/\sigma_{\min}(-\frac{1}{NT}[Q^*_{\beta\beta'} - Q^*_{\beta\phi'}Q^{*-1}_{\phi\phi'}Q^*_{\phi\beta'}])\\
\leq\ & 1/\min_{\gamma \in \mathcal{B}_m(\gamma^0)} \sigma_{\min}(-\frac{1}{NT}[Q_{\beta\beta'}(\gamma) - Q_{\beta\phi'}(\gamma)(Q_{\phi\phi'}(\gamma))^{-1}Q_{\phi\beta'}(\gamma)]) = O_p(1). &\text{(B.24)}
\end{aligned}
$$

(2) Note that $S_\beta = \sum_{i=1}^N \sum_{t=1}^T x_{it} d_{it} v_{it}$, $S_{\lambda_i} = \sum_{t=1}^T d_{it} v_{it} f_t^0$, $S_\lambda = (S'_{\lambda_1}, ..., S'_{\lambda_N})'$, $S_{f_t} = \sum_{i=1}^N d_{it} v_{it} \lambda_i^0$ and $S_f = (S'_{f_1}, ..., S'_{f_T})'$. By Assumption 5, $\|S_\beta\| = O_p(\sqrt{NT})$, $\|S_\lambda\| = O_p(\sqrt{NT})$, $\|S_f\| = O_p(\sqrt{NT})$ and $\|S\| \leq \|S_\beta\| + \|S_\lambda\| + \|S_f\| = O_p(\sqrt{NT})$. It follows that

$$
\left\|S_\beta(\hat{\beta}, \hat{\phi}) - S_\beta\right\| \leq \left\|S_\beta(\hat{\beta}, \hat{\phi})\right\| + \|S_\beta\| \leq \left\|S(\hat{\beta}, \hat{\phi})\right\| + \|S_\beta\| \leq \|S\| + \|S_\beta\| = O_p(\sqrt{NT}).
$$

(3) By the triangle inequality and equations (B.7) and (B.8),

$$
\left\|Q^*_{\beta\phi'}D_{TN}^{-1/2}\right\| \leq \max_{\gamma \in \mathcal{B}_m(\gamma^0)} \left\|(Q_{\beta\phi'}(\gamma) - Q_{\beta\phi'})D_{TN}^{-\frac{1}{2}}\right\| + \left\|Q_{\beta\phi'}D_{TN}^{-\frac{1}{2}}\right\| = O_p(\sqrt{NT}). \tag{B.25}
$$

(4) By Proposition 3.1(i),

$$
\begin{aligned}
\left\|D_{TN}^{1/2}Q^{*-1}_{\phi\phi'}D_{TN}^{1/2}\right\| &= \left\|(D_{TN}^{-1/2}Q^*_{\phi\phi'}D_{TN}^{-1/2})^{-1}\right\| = 1/\sigma_{\min}(-D_{TN}^{-1/2}Q^*_{\phi\phi'}D_{TN}^{-1/2})\\
&\leq 1/\min_{\gamma \in \mathcal{B}_m(\gamma^0)} \sigma_{\min}(-D_{TN}^{-1/2}Q_{\phi\phi'}(\gamma)D_{TN}^{-1/2}) = O_p(1). &\text{(B.26)}
\end{aligned}
$$

(5) Note that

$$
\left\|D_{TN}^{-1/2}(S_\phi(\hat{\beta}, \hat{\phi}) - S_\phi)\right\| \leq \frac{1}{\sqrt{T}}\left\|S_\lambda(\hat{\beta}, \hat{\phi}) - S_\lambda\right\| + \frac{1}{\sqrt{N}}\left\|S_f(\hat{\beta}, \hat{\phi}) - S_f\right\|
$$

$$\leq \frac{1}{\sqrt{T}}(\left\|S_\lambda(\hat\beta,\hat\phi)\right\| + \|S_\lambda\|) + \frac{1}{\sqrt{N}}(\left\|S_f(\hat\beta,\hat\phi)\right\| + \|S_f\|)$$
$$\leq 2(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{N}})\|S\| = O_p(\sqrt{T}+\sqrt{N}). \tag{B.27}$$

By eqn. (B.22) we also have $\hat\phi - \phi^0 = Q^{*-1}_{\phi\phi'}(S_\phi(\hat\beta,\hat\phi) - S_\phi) - Q^{*-1}_{\phi\phi'}Q^*_{\phi\beta'}(\hat\beta - \beta^0)$. Then

$$\begin{aligned}\left\|D^{1/2}_{TN}(\hat\phi-\phi^0)\right\| &= \left\|(D^{1/2}_{TN}Q^{*-1}_{\phi\phi'}D^{1/2}_{TN})\left[D^{-1/2}_{TN}(S_\phi(\hat\beta,\hat\phi)-S_\phi) - D^{-1/2}_{TN}Q^*_{\phi\beta'}(\hat\beta-\beta^0)\right]\right\| \\ &\leq \left\|(D^{1/2}_{TN}Q^{*-1}_{\phi\phi'}D^{1/2}_{TN})\right\|\left[\left\|D^{-1/2}_{TN}(S_\phi(\hat\beta,\hat\phi)-S_\phi)\right\| + \left\|D^{-1/2}_{TN}Q^*_{\phi\beta'}\right\|\left\|\hat\beta-\beta^0\right\|\right] \\ &= O_p(1)\left[O_p(\sqrt{T}+\sqrt{N}) + O_p(\sqrt{NT})O_p(1/c_{NT})\right] = O_p(\sqrt{T}+\sqrt{N}),\end{aligned}$$

where the second equality follows from equations (B.25), (B.26), (B.27) and (B.23). Thus $\left\|D^{-1/2}_{NT}(\hat\phi-\phi^0)\right\| = \frac{1}{\sqrt{NT}}\left\|D^{1/2}_{TN}(\hat\phi-\phi^0)\right\| = O_p(\frac{1}{c_{NT}})$, implying that $\left\|\frac{1}{\sqrt{N}}(\hat\lambda-\lambda^0)\right\| = O_p(\frac{1}{c_{NT}})$ and $\left\|\frac{1}{\sqrt{T}}(\hat f - f^0)\right\| = O_p(\frac{1}{c_{NT}})$.

Step (2): Combining the results that $\left\|\hat\beta-\beta^0\right\| = O_p(\frac{1}{c_{NT}})$, $\left\|\frac{1}{\sqrt{N}}(\hat\lambda-\lambda^0)\right\| = O_p(\frac{1}{c_{NT}})$ and $\left\|\frac{1}{\sqrt{T}}(\hat f - f^0)\right\| = O_p(\frac{1}{c_{NT}})$ together implies that $\hat\gamma = (\hat\beta', \hat\lambda', \hat f')'$ must be an interior point of $\mathcal{B}_m(\gamma^0)$ for any fixed m. It is not difficult to verify that the first order conditions of $\min_{\gamma\in\mathcal{B}_m(\gamma^0)}\|S(\gamma)\|^2$ is $2Q_{\gamma\gamma'}(\gamma)S(\gamma) = 0$. Since $\hat\gamma$ is an interior point, we must have $2Q_{\gamma\gamma'}(\hat\gamma)S(\hat\gamma) = 0$.

Since $\hat\gamma\in\mathcal{B}_m(\gamma^0)$, by Proposition 3.1, $\frac{1}{NT}[Q_{\beta\beta'}(\hat\gamma) - Q_{\beta\phi'}(\hat\gamma)(Q_{\phi\phi'}(\hat\gamma))^{-1}Q_{\phi\beta'}(\hat\gamma)]$ and $D^{-\frac12}_{TN}Q_{\phi\phi'}(\hat\gamma)D^{-\frac12}_{TN}$ are negative definite w.p.a.1. Then $Q_{\beta\beta'}(\hat\gamma) - Q_{\beta\phi'}(\hat\gamma)(Q_{\phi\phi'}(\hat\gamma))^{-1}Q_{\phi\beta'}(\hat\gamma)$ and $Q_{\phi\phi'}(\hat\gamma)$ are also negative definite w.p.a.1. This implies that $Q_{\gamma\gamma'}(\hat\gamma)$ is negative definite w.p.a.1.[1] It follows that $S(\hat\gamma) = 0$ w.p.a.1. ■

Remark. The intuition of Theorem 3.1 is that since the normalized Hessian is negative definite (the curvature is bounded away from zero) and $S(\hat\gamma)$ is close to $S(\gamma^0)$ (due to $S(\hat\gamma) = 0$ and $S(\gamma^0) = O_p(\sqrt{NT})$), $\hat\gamma$ should also be close to γ^0.

C Proofs of Theorem 3.2 and Corollary 3.1

Proof of Theorem 3.2:

Since $\hat\beta^{(k+1)} = \arg\max_\beta L(\beta, \tilde\phi^{(k+1)}) = \arg\max_\beta L(\beta,\hat\phi^{(k+1)})$ and $\hat\phi^{(k+1)} = \arg\max_\phi Q(\hat\beta^{(k)},\phi)$,[2] we have $S_\phi(\hat\beta^{(k)},\hat\phi^{(k+1)}) = 0$ and $S_\beta(\hat\beta^{(k+1)},\hat\phi^{(k+1)}) = 0$. By Theorem 3.1, we also have $S_\beta(\hat\beta,\hat\phi) = 0$ and $S_\phi(\hat\beta,\hat\phi) = 0$. By Taylor expansions, we have

$$Q_{\beta\beta'}(\hat\beta^{(k+1)} - \hat\beta) + Q^{(k+1)}_{\beta\phi'}(\hat\phi^{(k+1)} - \hat\phi) = S_\beta(\hat\beta^{(k+1)},\hat\phi^{(k+1)}) - S_\beta(\hat\beta,\hat\phi) = 0, \tag{C.1}$$
$$Q^{(k+\frac12)}_{\phi\beta'}(\hat\beta^{(k)} - \hat\beta) + Q^{(k+\frac12)}_{\phi\phi'}(\hat\phi^{(k+1)} - \hat\phi) = S_\phi(\hat\beta^{(k)},\hat\phi^{(k+1)}) - S_\phi(\hat\beta,\hat\phi) = 0, \tag{C.2}$$

where $Q_{\beta\beta'} = -\sum_{t=1}^T\sum_{i=1}^N d_{it}x_{it}x'_{it}$, $Q^{(k+1)}_{\beta\phi'} = Q_{\beta\phi'}(s\hat\gamma^{(k+1)} + (1-s)\hat\gamma)$ and $Q^{(k+\frac12)}_{\phi\phi'} = Q_{\phi\phi'}(s\hat\gamma^{(k+\frac12)} + (1-s)\hat\gamma)$ with $\hat\gamma^{(k+1)} = (\hat\beta^{(k+1)\prime},\hat\phi^{(k+1)\prime})'$, $\hat\gamma^{(k+\frac12)} = (\hat\beta^{(k)\prime},\hat\phi^{(k+1)\prime})'$. From eqn. (C.2) we have $\hat\phi^{(k+1)} - \hat\phi =$

[1] Note that $\begin{pmatrix} A & B \\ B' & D\end{pmatrix} = \begin{pmatrix} I & BD^{-1} \\ 0 & I\end{pmatrix}\begin{pmatrix} A - BD^{-1}B' & 0 \\ 0 & D\end{pmatrix}\begin{pmatrix} I & 0 \\ D^{-1}B' & I\end{pmatrix}$.

[2] As explained in Section 3.2, $L(\beta,\tilde\phi^{(k+1)}) = L(\beta,\hat\phi^{(k+1)})$ since $\hat f_t^{(k+1)\prime}\hat\lambda_i^{(k+1)} = \tilde f_t^{(k+1)\prime}\tilde\lambda_i^{(k+1)}$ for all (i,t).

$-(Q_{\phi\phi'}^{(k+\frac{1}{2})})^{-1}Q_{\phi\beta'}^{(k+\frac{1}{2})}(\hat{\beta}^{(k)}-\hat{\beta})$. Plugging this back into eqn. (C.1) yields

$$\hat{\beta}^{(k+1)}-\hat{\beta}=Q_{\beta\beta'}^{-1}Q_{\beta\phi'}^{(k+1)}(Q_{\phi\phi'}^{(k+\frac{1}{2})})^{-1}Q_{\phi\beta'}^{(k+\frac{1}{2})}(\hat{\beta}^{(k)}-\hat{\beta}). \tag{C.3}$$

We next show that

$$\left\|Q_{\beta\beta'}^{-1}Q_{\beta\phi'}^{(k+1)}(Q_{\phi\phi'}^{(k+\frac{1}{2})})^{-1}Q_{\phi\beta'}^{(k+\frac{1}{2})}\right\|<\psi \text{ for some constant } \psi\in(0,1), \tag{C.4}$$

which implies that $\left\|\hat{\beta}^{(k+1)}-\hat{\beta}\right\|<\psi\left\|\hat{\beta}^{(k)}-\hat{\beta}\right\|<...<\psi^{k+1}\left\|\hat{\beta}^{(0)}-\hat{\beta}\right\|$.

First, by eqn. (B.7) we have

$$\begin{aligned}
\left\|(Q_{\beta\phi'}^{(k+1)}-Q_{\beta\phi'})D_{TN}^{-1/2}\right\| &\leq \max_{\gamma\in\mathcal{B}_m(\gamma^0)}\left\|(Q_{\beta\phi'}(\gamma)-Q_{\beta\phi'})D_{TN}^{-\frac{1}{2}}\right\|\leq mM\sqrt{NT} \text{ w.p.a.1},\\
\left\|D_{TN}^{-1/2}(Q_{\phi\beta'}^{(k+\frac{1}{2})}-Q_{\phi\beta'})\right\| &\leq \max_{\gamma\in\mathcal{B}_m(\gamma^0)}\left\|D_{TN}^{-1/2}(Q_{\phi\beta'}(\gamma)-Q_{\phi\beta'})\right\|\leq mM\sqrt{NT} \text{ w.p.a.1},
\end{aligned}$$

and by eqn. (B.9) we have

$$D_{TN}^{1/2}((Q_{\phi\phi'}^{(k+\frac{1}{2})})^{-1}-Q_{\phi\phi'}^{-1})D_{TN}^{1/2}\leq \max_{\gamma\in\mathcal{B}_m(\gamma^0)}\left\|D_{TN}^{\frac{1}{2}}((Q_{\phi\phi'}(\gamma))^{-1}-Q_{\phi\phi'}^{-1})D_{TN}^{\frac{1}{2}}\right\|\leq\frac{mM}{C^2} \text{ w.p.a.1.}$$

Second, by Assumption 3(i) and Lemma B.2, $-\frac{1}{NT}(Q_{\beta\beta'}-Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'})$ is positive definite (p.d.) asymptotically. Then $\left\|Q_{\beta\beta'}^{-1}\right\|\leq\left\|(Q_{\beta\beta'}-Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'})^{-1}\right\|\leq\frac{1}{NT}M$ for some $M>0$ w.p.a.1. It follows that for some $M>0$,

$$\begin{aligned}
&\left\|Q_{\beta\beta'}^{-1}Q_{\beta\phi'}^{(k+1)}(Q_{\phi\phi'}^{(k+\frac{1}{2})})^{-1}Q_{\phi\beta'}^{(k+\frac{1}{2})}-Q_{\beta\beta'}^{-1}Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'}\right\|\\
\leq\ &\left\|Q_{\beta\beta'}^{-1}\right\|\left\|Q_{\beta\phi'}^{(k+1)}(Q_{\phi\phi'}^{(k+\frac{1}{2})})^{-1}Q_{\phi\beta'}^{(k+\frac{1}{2})}-Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'}\right\|\leq Mm \text{ w.p.a.1},
\end{aligned}\tag{C.5}$$

implying that $\left\|Q_{\beta\beta'}^{-1}Q_{\beta\phi'}^{(k+1)}(Q_{\phi\phi'}^{(k+\frac{1}{2})})^{-1}Q_{\phi\beta'}^{(k+\frac{1}{2})}\right\|\leq\left\|Q_{\beta\beta'}^{-1}Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'}\right\|+Mm$ w.p.a.1.

Third, note that

$$\begin{aligned}
\frac{-1}{NT}(Q_{\beta\beta'}-Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'}) &= \frac{1}{NT}(-Q_{\beta\beta'})^{1/2}\left[I_K-(-Q_{\beta\beta'}^{-1/2})Q_{\beta\phi'}(-Q_{\phi\phi'})^{-1}Q_{\phi\beta'}(-Q_{\beta\beta'}^{-1/2})\right](-Q_{\beta\beta'})^{1/2}\\
&= \frac{1}{NT}(-Q_{\beta\beta'})^{1/2}\left[I_K-\Xi\right](-Q_{\beta\beta'})^{1/2}
\end{aligned}$$

where $Q_{\beta\beta'}^{1/2}$ is the symmetric square root of $Q_{\beta\beta'}$ and $\Xi=(-Q_{\beta\beta'})^{-1/2}Q_{\beta\phi'}(-Q_{\phi\phi'})^{-1}Q_{\phi\beta'}(-Q_{\beta\beta'})^{-1/2}$. By the fact that $-\frac{1}{NT}(Q_{\beta\beta'}-Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'})$ is p.d. asymptotically, we know that $I_K-\Xi$ is p.d. asymptotically. Since $-\frac{1}{NT}Q_{\phi\phi'}$ is p.d., Ξ is also p.d., both asymptotically. Thus all eigenvalues of Ξ are between 0 and 1 w.p.a.1. Since $-Q_{\beta\beta'}^{-1}Q_{\beta\phi'}(-Q_{\phi\phi'})^{-1}Q_{\phi\beta'}$ and Ξ have the same eigenvalues, all eigenvalues of $-Q_{\beta\beta'}^{-1}Q_{\beta\phi'}(-Q_{\phi\phi'})^{-1}Q_{\phi\beta'}$ are strictly between 0 and 1 asymptotically, i.e., all eigenvalues of $Q_{\beta\beta'}^{-1}Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'}$ are strictly between 0 and 1 asymptotically. Therefore, if m is small enough, all eigenvalues of $Q_{\beta\beta'}^{-1}Q_{\beta\phi'}^{(k+1)}(Q_{\phi\phi'}^{(k+\frac{1}{2})})^{-1}Q_{\phi\beta'}^{(k+\frac{1}{2})}$ are strictly between 0 and 1 asymptotically. That is, (C.4) holds. ■

Proof of Corollary 3.1:

By eqn. (3.5), we have $S_\phi(\hat\beta^{(k)},\hat\phi^{(k+1)})=0$ and $S_\beta(\hat\beta^{(k+1)},\hat\phi^{(k)})=0$. This, in conjunction with the fact that $S_\beta(\hat\beta,\hat\phi)=0$ and $S_\phi(\hat\beta,\hat\phi)$ and Taylor expansions, implies that

$$\begin{aligned} Q_{\beta\beta'}(\hat\beta^{(k+1)}-\hat\beta)+Q_{\beta\phi'}^{(k+\frac12)}(\hat\phi^{(k)}-\hat\phi) &= S_\beta(\hat\beta^{(k+1)},\hat\phi^{(k)})-S_\beta(\hat\beta,\hat\phi)=0, &\text{(C.6)}\\ Q_{\phi\beta'}^{(k+\frac12)}(\hat\beta^{(k)}-\hat\beta)+Q_{\phi\phi'}^{(k+\frac12)}(\hat\phi^{(k+1)}-\hat\phi) &= S_\phi(\hat\beta^{(k)},\hat\phi^{(k+1)})-S_\phi(\hat\beta,\hat\phi)=0, &\text{(C.7)}\end{aligned}$$

where $Q_{\beta\phi'}^{(k+\frac12)}=Q_{\beta\phi'}(s(\hat\beta^{(k+1)},\hat\phi^{(k)})+(1-s)(\hat\beta,\hat\phi))$, $Q_{\phi\beta'}^{(k+\frac12)}=Q_{\phi\beta'}(s(\hat\beta^{(k)},\hat\phi^{(k+1)})+(1-s)(\hat\beta,\hat\phi))$, and $Q_{\phi\phi'}^{(k+\frac12)}=Q_{\phi\phi'}(s(\hat\beta^{(k)},\hat\phi^{(k+1)})+(1-s)(\hat\beta,\hat\phi))$ for $0<s<1$. It follows that

$$\begin{aligned}\begin{pmatrix}\hat\beta^{(k+1)}-\hat\beta\\ \hat\phi^{(k+1)}-\hat\phi\end{pmatrix} &= -\begin{pmatrix}Q_{\beta\beta'} & 0\\ 0 & Q_{\phi\phi'}^{(k+\frac12)}\end{pmatrix}^{-1}\begin{pmatrix}0 & Q_{\beta\phi'}^{(k+\frac12)}\\ Q_{\phi\beta'}^{(k+\frac12)} & 0\end{pmatrix}\begin{pmatrix}\hat\beta^{(k)}-\hat\beta\\ \hat\phi^{(k)}-\hat\phi\end{pmatrix}\\ &= -\begin{pmatrix}0 & Q_{\beta\beta'}^{-1}Q_{\beta\phi'}^{(k+\frac12)}\\ (Q_{\phi\phi'}^{(k+\frac12)})^{-1}Q_{\phi\beta'}^{(k+\frac12)} & 0\end{pmatrix}\begin{pmatrix}\hat\beta^{(k)}-\hat\beta\\ \hat\phi^{(k)}-\hat\phi\end{pmatrix}. \qquad\text{(C.8)}\end{aligned}$$

Thus $\left\|\hat\gamma^{(k+1)}-\hat\gamma\right\|\le\sqrt{\left\|Q_{\beta\beta'}^{-1}Q_{\beta\phi'}^{(k+\frac12)}(Q_{\phi\phi'}^{(k+\frac12)})^{-1}Q_{\phi\beta'}^{(k+\frac12)}\right\|}\left\|\hat\gamma^{(k)}-\hat\gamma\right\|$.[3] It's easy to see that eqn. (C.5) is also valid for $Q_{\beta\beta'}^{-1}Q_{\beta\phi'}^{(k+\frac12)}(Q_{\phi\phi'}^{(k+\frac12)})^{-1}Q_{\phi\beta'}^{(k+\frac12)}$ and we have already proved that $\left\|Q_{\beta\beta'}^{-1}Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'}\right\|<\psi\in(0,1)$. Then $\left\|Q_{\beta\beta'}^{-1}Q_{\beta\phi'}^{(k+\frac12)}(Q_{\phi\phi'}^{(k+\frac12)})^{-1}Q_{\phi\beta'}^{(k+\frac12)}\right\|$ is also strictly less than 1 asymptotically. ■

D Proof of Theorem 3.3

To prove Theorem 3.3, we state and prove five technical lemmas.

Lemma D.1 *Under Assumption 4, as $(N,T)\to\infty$, $\|d\circ v\|=O_p(N^{\frac12}T^{\frac14}+N^{\frac14}T^{\frac12})$, where $d\circ v$ is a $T\times N$ matrix with $d_{it}v_{it}$ as the (t,i)-th element.*

Proof. This lemma is the same as Lemma A.1 in Su and Wang (2024). For completeness, we outline the proof here.

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

where the last equality follows from Assumption 4(ii)-(iii). Then the result hold by the Markov inequality. ■

[3] If c is an eigenvalue of $\begin{pmatrix}0 & B\\ C & 0\end{pmatrix}$, then $\begin{pmatrix}0 & B\\ C & 0\end{pmatrix}\begin{pmatrix}e_1\\ e_2\end{pmatrix}=\begin{pmatrix}Be_2\\ Ce_1\end{pmatrix}=\begin{pmatrix}ce_1\\ ce_2\end{pmatrix}$. It follows that $BCe_1=Bce_2=c^2e_1$. That is, c^2 is an eigenvalue of BC.

Lemma D.2 *Let $\hat{\Delta}_{\Theta}^{(0)} = \hat{\Theta}^{(0)} - \Theta^0$ and $\nu_{NT} = 2c_1(N^{\frac{1}{2}}T^{\frac{1}{4}} + N^{\frac{1}{4}}T^{\frac{1}{2}})$. Under Assumptions 3(i)-(ii), 4 and 5(ii), $\hat{\Delta}_{\Theta}^{(0)} \in \mathcal{R}$ w.p.a.1.*

Proof. Let $M_{vec(d\circ x)} = I_{NT} - P_{vec(d\circ x)}$ and $P_{vec(d\circ x)}$ denote the projection matrix of $[vec(d\circ x_1), ..., vec(d\circ x_K)]$, where $vec(d\circ x_k)$ is the $TN\times 1$ vector that vectorizes $d\circ x_k$. Also, let $mat(\cdot)$ denote the inverse operator of $vec(\cdot)$ by transforming a $TN\times 1$ vector back to a $T\times N$ matrix. Let y (resp. v) denote the $T\times N$ matrix with y_{it} (resp. v_{it}) as the (t,i)-th element. Then after concentrating out β, the sum of squared residuals (i.e., $-L(\cdot)$ with $L(\cdot)$ defined in (3.1)) can be rewritten as

$$SSR(\Theta) = \frac{1}{2} vec(d\circ(y-\Theta))' M_{vec(d\circ x)} vec(d\circ(y-\Theta)).$$

Then

$$\begin{aligned}
SSR(\Theta^0) - SSR(\hat{\Theta}^{(0)}) &= -vec(d\circ(y-\Theta^0))' M_{vec(d\circ x)} vec(d\circ(\Theta^0 - \hat{\Theta}^{(0)})) \\
&\quad -\frac{1}{2} vec(d\circ(\Theta^0 - \hat{\Theta}^{(0)}))' M_{vec(d\circ x)} vec(d\circ(\Theta^0 - \hat{\Theta}^{(0)})) \\
&\leq -vec(d\circ v)' M_{vec(d\circ x)} vec(d\circ(\Theta^0 - \hat{\Theta}^{(0)})) \\
&= -vec(d\circ v)' M_{vec(d\circ x)} vec(\Theta^0 - \hat{\Theta}^{(0)}) \\
&= \operatorname{tr}\left[\hat{\Delta}_{\Theta}^{(0)\prime} mat(M_{vec(d\circ x)} vec(d\circ v))\right] \leq \left\|mat(M_{vec(d\circ x)} vec(d\circ v))\right\| \left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_* \\
&= \left\|d\circ v - mat(P_{vec(d\circ x)} vec(d\circ v))\right\| \left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_* \\
&\leq c_1(N^{\frac{1}{2}}T^{\frac{1}{4}} + N^{\frac{1}{4}}T^{\frac{1}{2}}) \left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_* \text{ w.p.a.1,}
\end{aligned}$$

where c_1 is some positive constant, the first inequality holds by the fact that $M_{vec(d\circ x)} vec(d\circ x_k) = 0$ for all k, the second equality follows from the fact that $vec(d\circ x_k)' vec(d\circ(\Theta^0 - \hat{\Theta}^{(0)})) = vec(d\circ x_k)' vec(\Theta^0 - \hat{\Theta}^{(0)})$, the second inequality is due to the fact that $tr(A'B) \leq \|A\| \|B\|_*$, and the last inequality follows from Lemma D.1 and the fact that

$$\begin{aligned}
\left\|mat(P_{vec(d\circ x)} vec(d\circ v))\right\| &\leq \left\|mat(P_{vec(d\circ x)} vec(d\circ v))\right\|_F = \left\|P_{vec(d\circ x)} vec(d\circ v)\right\|_F \\
&\leq \sqrt{\sum_{k=1}^{K} \|d\circ x_k\|_F^2} \left\|(\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it} x_{it} x_{it}')^{-1}\right\| \left\|\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it} x_{it} v_{it}\right\| = O_p(1)
\end{aligned}$$

by Assumptions 3(ii), 3(i) and 5(ii).

Now, from the construction of $\mathcal{P}^{\perp}$ and $\mathcal{P}$, we have

$$\begin{aligned}
\left\|\hat{\Theta}^{(0)}\right\|_* &= \left\|\hat{\Delta}_{\Theta}^{(0)} + \Theta^0\right\|_* = \left\|\Theta^0 + \mathcal{P}^{\perp}(\hat{\Delta}_{\Theta}^{(0)}) + \mathcal{P}(\hat{\Delta}_{\Theta}^{(0)})\right\|_* \geq \left\|\Theta^0 + \mathcal{P}^{\perp}(\hat{\Delta}_{\Theta}^{(0)})\right\|_* - \left\|\mathcal{P}(\hat{\Delta}_{\Theta}^{(0)})\right\|_* \\
&= \left\|\Theta^0\right\|_* + \left\|\mathcal{P}^{\perp}(\hat{\Delta}_{\Theta}^{(0)})\right\|_* - \left\|\mathcal{P}(\hat{\Delta}_{\Theta}^{(0)})\right\|_*. \qquad \text{(D.1)}
\end{aligned}$$

Then $\left\|\Theta^0\right\|_* - \left\|\hat{\Theta}^{(0)}\right\|_* \leq \left\|\mathcal{P}(\hat{\Delta}_{\Theta}^{(0)})\right\| - \left\|\mathcal{P}^{\perp}(\hat{\Delta}_{\Theta}^{(0)})\right\|_*$. It follows that w.p.a.1

$$0 \leq SSR(\Theta^0) - SSR(\hat{\Theta}^{(0)}) + \nu_{NT}\left(\left\|\Theta^0\right\|_* - \left\|\hat{\Theta}^{(0)}\right\|_*\right)$$

$$
\begin{aligned}
&\leq c_1(N^{\frac{1}{2}}T^{\frac{1}{4}}+N^{\frac{1}{4}}T^{\frac{1}{2}})\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_*+\nu_{NT}\left(\left\|\mathcal{P}(\hat{\Delta}_{\Theta}^{(0)})\right\|_*-\left\|\mathcal{P}^{\perp}(\hat{\Delta}_{\Theta}^{(0)})\right\|_*\right)\\
&\leq c_1(N^{\frac{1}{2}}T^{\frac{1}{4}}+N^{\frac{1}{4}}T^{\frac{1}{2}})\left(\left\|\mathcal{P}^{\perp}(\hat{\Delta}_{\Theta}^{(0)})\right\|_*+\left\|\mathcal{P}(\hat{\Delta}_{\Theta}^{(0)})\right\|_*\right)+\nu_{NT}\left(\left\|\mathcal{P}(\hat{\Delta}_{\Theta}^{(0)})\right\|_*-\left\|\mathcal{P}^{\perp}(\hat{\Delta}_{\Theta}^{(0)})\right\|_*\right),
\end{aligned}
$$

and consequently $\left\|\mathcal{P}^{\perp}(\hat{\Delta}_{\Theta}^{(0)})\right\|_*\leq 3\left\|\mathcal{P}(\hat{\Delta}_{\Theta}^{(0)})\right\|_*$ w.p.a.1 if $\nu_{NT}=2c_1(N^{\frac{1}{2}}T^{\frac{1}{4}}+N^{\frac{1}{4}}T^{\frac{1}{2}})$. ■

Lemma D.3 *For any $\Delta_\Theta\in\mathcal{R}$, we have $\|\Delta_\Theta\|_*\leq 4\sqrt{2r}\,\|\Delta_\Theta\|_F$.*

Proof. For any $\Delta_\Theta\in\mathcal{R}$,

$$
\|\Delta_\Theta\|_*=\|\mathcal{P}(\Delta_\Theta)\|_*+\left\|\mathcal{P}^{\perp}(\Delta_\Theta)\right\|_*\leq 4\|\mathcal{P}(\Delta_\Theta)\|_*\leq 4\sqrt{2r}\,\|\mathcal{P}\left(\Delta_\Theta\right)\|_F\leq 4\sqrt{2r}\,\|\Delta_\Theta\|_F\,,
$$

where the first inequality follows from the definition of $\mathcal{R}$, the second inequality is due to $\|A\|_*\leq\sqrt{\mathrm{rank}(A)}\,\|A\|_F$ and $\mathrm{rank}\left(\mathcal{P}\left(\Delta_\Theta\right)\right)\leq 2r$, and the third inequality is due to the fact that $\|\Delta_\Theta\|_F^2=\mathrm{tr}(\Delta_\Theta'\Delta_\Theta)=tr([\mathcal{P}(\Delta_\Theta)+\mathcal{P}^{\perp}(\Delta_\Theta)]'[\mathcal{P}(\Delta_\Theta)+\mathcal{P}^{\perp}(\Delta_\Theta)])$, $\|\mathcal{P}\left(\Delta_\Theta\right)\|_F^2=tr(\mathcal{P}(\Delta_\Theta)'\mathcal{P}(\Delta_\Theta))$, $\left\|\mathcal{P}^{\perp}\left(\Delta_\Theta\right)\right\|_F^2=tr(\mathcal{P}^{\perp}(\Delta_\Theta)'\mathcal{P}^{\perp}(\Delta_\Theta))$ and $tr(\mathcal{P}(\Delta_\Theta)'\mathcal{P}^{\perp}(\Delta_\Theta))=0$. ■

Lemma D.4 *Suppose that Assumptions 3(ii) and 6(i) hold. Then*

(i) $\|[d-\mathbb{E}_{\phi x}(d)]\circ x_k\|=O_p(T^{\frac{1}{2}}N^{\frac{1}{4}})$, where $d-\mathbb{E}_{\phi x}(d)$ and x_k denotes the $T\times N$ matrix with $d_{it}-\mathbb{E}_{\phi x}(d_{it})$ and x_{itk} as the (t,i)-th element, respectively;

(ii) $\mathbb{E}_{\phi x}[\|\chi\circ d)\|]\leq T^{\frac{1}{2}}N^{\frac{1}{4}}$, where χ denotes the $T\times N$ matrix with χ_{it} as the (t,i)-th element and $\{\chi_{it}:i\in[N],t\in[T]\}$ are independent Rademacher random variables.

Proof. (i) Note that

$$
\begin{aligned}
\mathbb{E}_{\phi x}(\|[d-\mathbb{E}_{\phi x}(d)]\circ x_k\|^4) &= \mathbb{E}_{\phi x}(\|([d-\mathbb{E}_{\phi x}(d)]\circ x_k)([d-\mathbb{E}_{\phi x}(d)]\circ x_k)'\|^2)\\
&\leq \mathbb{E}_{\phi x}(\|([d-\mathbb{E}_{\phi x}(d)]\circ x_k)([d-\mathbb{E}_{\phi x}(d)]\circ x_k)'\|_F^2)\\
&= \sum\nolimits_{s,t=1}^{T}\mathbb{E}_{\phi x}\left\{\sum\nolimits_{i=1}^{N}[d_{is}-\mathbb{E}_{\phi x}(d_{is})][d_{it}-\mathbb{E}_{\phi x}(d_{it})]x_{itk}x_{isk}\right\}^2\\
&= \sum\nolimits_{s,t=1}^{T}\sum\nolimits_{i=1}^{N}\mathbb{E}_{\phi x}\left\{[d_{is}-\mathbb{E}_{\phi x}(d_{is})]^2[d_{it}-\mathbb{E}_{\phi x}(d_{it})]^2x_{itk}^2x_{isk}^2\right\}\\
&\leq \sum\nolimits_{s,t=1}^{T}\sum\nolimits_{i=1}^{N}x_{itk}^2x_{isk}^2,
\end{aligned}
$$

where the third equality follows from Assumption 6(i). Thus by Assumption 3(ii),

$$
\begin{aligned}
\mathbb{E}(\|[d-\mathbb{E}_{\phi x}(d)]\circ x_k\|^4) &= \mathbb{E}[\mathbb{E}_{\phi x}(\|[d-\mathbb{E}_{\phi x}(d)]\circ x_k\|^4)]\leq\sum\nolimits_{s,t=1}^{T}\sum\nolimits_{i=1}^{N}\mathbb{E}(x_{itk}^2x_{isk}^2)\\
&\leq \sum\nolimits_{s,t=1}^{T}\sum\nolimits_{i=1}^{N}\sqrt{\mathbb{E}(x_{itk}^4)\mathbb{E}(x_{isk}^4)}\leq T^2NM,
\end{aligned}
$$

and it follows that $\|[d-\mathbb{E}_{\phi x}(d)]\circ x_k\|=O_p(T^{\frac{1}{2}}N^{\frac{1}{4}})$.

(ii) The proof is similar to that of part (i) and thus omitted, given that $\{\chi_{it}:i\in[N],t\in[T]\}$ are independent Rademacher random variables. ■

Lemma D.5 *Let $\hat{\Delta}_\beta^{(0)} = \hat{\beta}^{(0)} - \beta^0$. Suppose that Assumptions 3(i)-(ii), 4, 5(ii) and 6(i) hold. Then w.p.a.1,*

$$\begin{aligned} & \left| \sum_{i=1}^N \sum_{t=1}^T [d_{it} - \mathbb{E}_{\phi x}(d_{it})] \, (\hat{\Delta}_{\Theta,it}^{(0)} + x_{it}' \hat{\Delta}_\beta^{(0)})^2 \right| \\ \leq \quad & M\sqrt{NT} \left\| \hat{\Delta}_\beta^{(0)} \right\|^2 + MT^{\frac{1}{2}} N^{\frac{1}{4}} \left\| \hat{\Delta}_\beta^{(0)} \right\| \left\| \hat{\Delta}_\Theta^{(0)} \right\|_* + A_{NT} + \mathcal{E}_{NT} + \frac{1}{\sqrt{c_{NT}}} \left\| \hat{\Delta}_\Theta^{(0)} \right\|_F^2, \end{aligned} \tag{D.2}$$

where $\mathcal{E}_{NT} = 3r(16M)^2 \sqrt{c_{NT}} (\mathbb{E}_{\phi x}[\|\chi \circ d)\|])^2$, $\{\chi_{it} : i \in [N], t \in [T]\}$ are independent Rademacher random variables and $A_{NT} = M\sqrt{NTc_{NT} \log(N+T)}$.

Proof. This lemma is crucial for extending the NNR estimator for balanced panels (Moon and Weidner (2023)) to the unbalanced case. The techniques are borrowed from the matrix completion literature (e.g., Klopp (2014) and Negahban and Wainwright (2012)) and compared to those previous papers, there are two main differences. First, we have $\hat{\Delta}_\beta^{(0)}$ here while the matrix completion literature typically does not. Second, the last term of expression (D.2) is $\frac{1}{\sqrt{c_{NT}}} \left\| \hat{\Delta}_\Theta^{(0)} \right\|_F^2$ while in previous literature the corresponding term is $\frac{1}{2} \left\| \hat{\Delta}_\Theta^{(0)} \right\|_F^2$. The design of expression (D.2) is tailored for the current setup. To prove this lemma, we first note that

$$\begin{aligned} & \left| \sum_{i=1}^N \sum_{t=1}^T [d_{it} - \mathbb{E}_{\phi x}(d_{it})] \, (\hat{\Delta}_{\Theta,it}^{(0)} + x_{it}' \hat{\Delta}_\beta^{(0)})^2 \right| \\ & \leq \left| \sum_{i=1}^N \sum_{t=1}^T [d_{it} - \mathbb{E}_{\phi x}(d_{it})] \, \hat{\Delta}_\beta^{(0)\prime} x_{it} x_{it}' \hat{\Delta}_\beta^{(0)} \right| \\ & + 2 \left| \sum_{i=1}^N \sum_{t=1}^T [d_{it} - \mathbb{E}_{\phi x}(d_{it})] \, \hat{\Delta}_{\Theta,it}^{(0)} x_{it}' \hat{\Delta}_\beta^{(0)} \right| + \left| \sum_{i=1}^N \sum_{t=1}^T [d_{it} - \mathbb{E}_{\phi x}(d_{it})] \, (\hat{\Delta}_{\Theta,it}^{(0)})^2 \right|. \end{aligned} \tag{D.3}$$

Below, we study the three terms on the right hand side of (D.3) one by one.

First, by Assumption 6(i), $d_{it} - \mathbb{E}_{\phi x}(d_{it})$ is independent across i and t conditioning on all x_{it}, thus

$$\begin{aligned} \mathbb{E}\left[\mathbb{E}_{\phi x} \left\{ \sum_{i=1}^N \sum_{t=1}^T [d_{it} - \mathbb{E}_{\phi x}(d_{it})] \, x_{itk} x_{itl} \right\}^2 \right] &= \mathbb{E} \left\{ \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}_{\phi x}([d_{it} - \mathbb{E}_{\phi x}(d_{it})]^2) x_{itk}^2 x_{itl}^2 \right\} \\ &\leq \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}(x_{itk}^2 x_{itl}^2) \leq \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}\{ \|x_{it}\|^4 \}, \end{aligned}$$

i.e., $\mathbb{E}(\left\| \sum_{i=1}^N \sum_{t=1}^T [d_{it} - \mathbb{E}_{\phi x}(d_{it})] \, x_{it} x_{it}' \right\|_F^2) = O(NT)$ by Assumption 3(ii). Thus

$$\begin{aligned} \left| \sum_{i=1}^N \sum_{t=1}^T [d_{it} - \mathbb{E}_{\phi x}(d_{it})] \, \hat{\Delta}_\beta^{(0)\prime} x_{it} x_{it}' \hat{\Delta}_\beta^{(0)} \right| &\leq \left\| \hat{\Delta}_\beta^{(0)} \right\|^2 \left\| \sum_{i=1}^N \sum_{t=1}^T [d_{it} - \mathbb{E}_{\phi x}(d_{it})] \, x_{it} x_{it}' \right\| \\ &= \left\| \hat{\Delta}_\beta^{(0)} \right\|^2 O_p(\sqrt{NT}). \end{aligned} \tag{D.4}$$

Second, by Lemma D.4,

$$\left|\sum_{i=1}^{N}\sum_{t=1}^{T}[d_{it}-\mathbb{E}_{\phi x}(d_{it})]\,\hat{\Delta}^{(0)}_{\Theta,it}x'_{it}\hat{\Delta}^{(0)}_{\beta}\right| \le \sum_{k=1}^{K}\left|\hat{\Delta}^{(0)}_{\beta,k}\right|\|[d-\mathbb{E}_{\phi x}(d)]\circ x_k\|\left\|\hat{\Delta}^{(0)}_{\Theta}\right\|_* \le \left\|\hat{\Delta}^{(0)}_{\beta}\right\|O_p(T^{\frac{1}{2}}N^{\frac{1}{4}})\left\|\hat{\Delta}^{(0)}_{\Theta}\right\|_*. \quad \text{(D.5)}$$

Third, we shall prove

$$\left|\sum_{i=1}^{N}\sum_{t=1}^{T}[d_{it}-\mathbb{E}_{\phi x}(d_{it})]\,(\hat{\Delta}^{(0)}_{\Theta,it})^2\right| \le A_{NT}+\frac{1}{\sqrt{c_{NT}}}\left\|\hat{\Delta}^{(0)}_{\Theta}\right\|_F^2+\mathcal{E}_{NT}\text{ w.p.a.1.} \quad \text{(D.6)}$$

To prove expression (D.6), we first define the events

$$\begin{aligned}
\mathcal{R} &= \{\Delta_\Theta\in\mathbb{R}^{T\times N}:\left\|\mathcal{P}^{\perp}(\Delta_\Theta)\right\|_*\le 3\left\|\mathcal{P}(\Delta_\Theta)\right\|_*\text{ and }\|\Delta_\Theta\|_{\max}\le M\},\\
\mathcal{R}_{A,l} &= \mathcal{R}\cap\{\Delta_\Theta:(\frac{6}{5})^{l-1}A_{NT}\le\|\Delta_\Theta\|_F^2\le(\frac{6}{5})^{l}A_{NT}\}\text{ and}\\
\mathcal{R}_{A} &= \mathcal{R}\cap\{\Delta_\Theta:\|\Delta_\Theta\|_F^2\ge A_{NT}\}=\cup_{l=1}^{\infty}\mathcal{R}_{A,l}.
\end{aligned}$$

and the functional

$$g_l(d)=\sup_{\Delta_\Theta\in\mathcal{R}_{A,l}}\left|\sum_{i=1}^{N}\sum_{t=1}^{T}[d_{it}-\mathbb{E}_{\phi x}(d_{it})]\,(\Delta_{\Theta,it})^2\right|.$$

Also, define the events

$$\begin{aligned}
\mathcal{A} &= \left\{\exists\Delta_\Theta\in\mathcal{R}_A\text{ s.t. }\left|\sum_{i=1}^{N}\sum_{t=1}^{T}[d_{it}-\mathbb{E}_{\phi x}(d_{it})]\,(\Delta_{\Theta,it})^2\right|\ge\frac{\|\Delta_\Theta\|_F^2}{\sqrt{c_{NT}}}+\mathcal{E}_{NT}\right\},\\
\mathcal{A}_l &= \left\{g_l(d)\ge\frac{1}{\sqrt{c_{NT}}}(\frac{6}{5})^{l-1}A_{NT}+\mathcal{E}_{NT}\right\}.
\end{aligned}$$

Note that $\mathcal{A}\subset\cup_{l=1}^{\infty}\mathcal{A}_l$ and

$$\begin{aligned}
\Pr(\mathcal{A}) &\le \Pr\left(\cup_{l=1}^{\infty}\mathcal{A}_l\right)\le\sum_{l=1}^{\infty}\Pr\left(|g_l(d)-\mathbb{E}_{\phi x}[g_l(d)]|+\mathbb{E}_{\phi x}[g_l(d)]\ge\frac{1}{\sqrt{c_{NT}}}(\frac{6}{5})^{l-1}A_{NT}+\mathcal{E}_{NT}\right)\\
&\le\sum_{l=1}^{\infty}\Pr\left(|g_l(d)-\mathbb{E}_{\phi x}[g_l(d)]|\ge\frac{(\frac{6}{5})^{l-1}}{\sqrt{c_{NT}}}A_{NT}+\mathcal{E}_{NT}-32M\mathbb{E}_{\phi x}[\|\chi\circ d\|]\sqrt{2r(\frac{6}{5})^{l}A_{NT}}\right)\\
&\le\sum_{l=1}^{\infty}\Pr\left(|g_l(d)-\mathbb{E}_{\phi x}[g_l(d)]|\ge\frac{1}{6\sqrt{c_{NT}}}(\frac{6}{5})^{l}A_{NT}\right)\\
&\le\sum_{l=1}^{\infty}2\exp\left(-\frac{1}{18M^4NTc_{NT}}(\frac{6}{5})^{2l}A_{NT}^2\right)\le\sum_{l=1}^{\infty}2\exp\left(-\frac{\log(\frac{6}{5})}{9M^4NTc_{NT}}A_{NT}^2l\right)\\
&=2\frac{\exp\left(-\frac{\log(\frac{6}{5})}{9M^4NTc_{NT}}A_{NT}^2\right)}{1-\exp\left(-\frac{\log(\frac{6}{5})}{9M^4NTc_{NT}}A_{NT}^2\right)}\le\frac{2}{\exp(\log(N+T))-1}\to 0,
\end{aligned} \quad \text{(D.7)}$$

where the third inequality follows from expression (D.10) below, the fourth inequality follows from the fact

$$\begin{aligned}32M\mathbb{E}_{\phi x}[\|\chi \circ d)\|]\sqrt{2r(\frac{6}{5})^l A_{NT}} &= 2\left\{\sqrt{3r}16Mc_{NT}^{1/4}\mathbb{E}_{\phi x}[\|\chi \circ d)\|]\right\} \cdot \frac{1}{\sqrt{3}c_{NT}^{1/4}}\sqrt{2(\frac{6}{5})^l A_{NT}} \\ &\le 3r(16M)^2\sqrt{c_{NT}}(\mathbb{E}_{\phi x}[\|\chi \circ d)\|])^2 + \frac{2}{3\sqrt{c_{NT}}}(\frac{6}{5})^l A_{NT} \\ &\equiv \mathcal{E}_{NT} + \frac{2}{3\sqrt{c_{NT}}}(\frac{6}{5})^l A_{NT}\end{aligned}$$

by the CS inequality, the fifth inequality follows from (D.9) below and the sixth inequality follows from $(\frac{6}{5})^{2l} > \log(\frac{6}{5})2l$. Now denote the event $\overline{\mathcal{A}} = \left\{\left|\sum_{i=1}^N \sum_{t=1}^T [d_{it} - \mathbb{E}_{\phi x}(d_{it})] (\hat{\Delta}_{\Theta,it}^{(0)})^2\right| > A_{NT} + \frac{\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F^2}{\sqrt{c_{NT}}} + \mathcal{E}_{NT}\right\}$.

$$\begin{aligned}\Pr(\overline{\mathcal{A}}) &= \Pr\left(\overline{\mathcal{A}} \cap \left\{\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F^2 < A_{NT}\right\}\right) + \Pr\left(\overline{\mathcal{A}} \cap \left\{\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F^2 \ge A_{NT}\right\}\right) \\ &= \Pr\left(\overline{\mathcal{A}} \cap \left\{\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F^2 \ge A_{NT}\right\}\right) \\ &\le \Pr\left(\overline{\mathcal{A}} \cap \left\{\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F^2 \ge A_{NT}\right\} \cap \left\{\hat{\Delta}_{\Theta}^{(0)} \in \mathcal{R}\right\}\right) + \Pr\left(\hat{\Delta}_{\Theta}^{(0)} \notin \mathcal{R}\right) \\ &\le \Pr\left(\overline{\mathcal{A}} \cap \left\{\hat{\Delta}_{\Theta}^{(0)} \in \mathcal{R}_A\right\}\right) + \Pr\left(\hat{\Delta}_{\Theta}^{(0)} \notin \mathcal{R}\right) \\ &\le \Pr(\mathcal{A}) + \Pr\left(\hat{\Delta}_{\Theta}^{(0)} \notin \mathcal{R}\right) \to 0 \text{ by expression (D.7) and Lemma D.2,}\end{aligned} \tag{D.8}$$

where the second equality follows from the fact that the events $\overline{\mathcal{A}}$ and $\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F^2 < A_{NT}$ are mutually exclusive since $|d_{it} - \mathbb{E}_{\phi x}(d_{it})| \le 1$. In summary, expressions (D.3)-(D.6) together proves this lemma.

(1) In the above proof, we call upon the following concentration inequality:

$$\Pr(|g_l(d) - \mathbb{E}_{\phi x}[g_l(d)]| \ge a) \le 2\exp\left(-\frac{2}{NTM^4}a^2\right). \tag{D.9}$$

This expression follows from the Azuma-Hoeffding inequality (see Corollary 2.21 in chapter 2 of Wainwright, 2019), since d_{it} is independent with each other conditioning on ϕ and x and $g_l(d)$ satisfies the bounded difference property with parameter M^2 as verified below.

Suppose $d_1 = \{d_{it1},\ i \in [N],\ t \in [T]\}$ and $d_2 = \{d_{it2},\ i \in [N],\ t \in [T]\}$ are the same except for the (i', t')-th element. Define $\Delta_{\Theta l,s} = \arg\max_{\Delta_\Theta \in \mathcal{R}_{A,l}} \left|\sum_{i=1}^N \sum_{t=1}^T [d_{its} - \mathbb{E}_{\phi x}(d_{it})] (\Delta_{\Theta,it})^2\right|$ for $s = 1, 2$, then

$$\begin{aligned}g_l(d_1) - g_l(d_2) &= \sup_{\Delta_\Theta \in \mathcal{R}_{A,l}} \left|\sum_{i=1}^N \sum_{t=1}^T [d_{it1} - \mathbb{E}_{\phi x}(d_{it})] (\Delta_{\Theta,it})^2\right| - \sup_{\Delta_\Theta \in \mathcal{R}_{A,l}} \left|\sum_{i=1}^N \sum_{t=1}^T [d_{it2} - \mathbb{E}_{\phi x}(d_{it})] (\Delta_{\Theta,it})^2\right| \\ &= \left|\sum_{i=1}^N \sum_{t=1}^T [d_{it1} - \mathbb{E}_{\phi x}(d_{it})] (\Delta_{\Theta l,1,it})^2\right| - \left|\sum_{i=1}^N \sum_{t=1}^T [d_{it2} - \mathbb{E}_{\phi x}(d_{it})] (\Delta_{\Theta l,2,it})^2\right| \\ &\le \left|\sum_{i=1}^N \sum_{t=1}^T [d_{it1} - \mathbb{E}_{\phi x}(d_{it})] (\Delta_{\Theta l,1,it})^2\right| - \left|\sum_{i=1}^N \sum_{t=1}^T [d_{it2} - \mathbb{E}_{\phi x}(d_{it})] (\Delta_{\Theta l,1,it})^2\right| \\ &\le \left|[d_{i't'1} - \mathbb{E}_{\phi x}(d_{i't'})] (\Delta_{\Theta l,1,i't'})^2\right| + \left|[d_{i't'2} - \mathbb{E}_{\phi x}(d_{i't'})] (\Delta_{\Theta l,1,i't'})^2\right| \le M^2.\end{aligned}$$

(2) In the above proof, we also call upon the following symmetrization and contraction inequality:

$$
\begin{aligned}
\mathbb{E}_{\phi x}[g_l(d)] &= \mathbb{E}_{\phi x}\left\{\sup_{\Delta_\Theta\in\mathcal{R}_{A,l}}\left|\sum_{i=1}^N\sum_{t=1}^T[d_{it}-\mathbb{E}_{\phi x}(d_{it})]\,(\Delta_{\Theta,it})^2\right|\right\} \\
&\le 2\mathbb{E}_{\phi x}\left\{\sup_{\Delta_\Theta\in\mathcal{R}_{A,l}}\left|\sum_{i=1}^N\sum_{t=1}^T\chi_{it}d_{it}(\Delta_{\Theta,it})^2\right|\right\} \le 8M\mathbb{E}_{\phi x}\left\{\sup_{\Delta_\Theta\in\mathcal{R}_{A,l}}\left|\sum_{i=1}^N\sum_{t=1}^T\chi_{it}d_{it}\Delta_{\Theta,it}\right|\right\} \\
&= 8M\mathbb{E}_{\phi x}\left\{\sup_{\Delta_\Theta\in\mathcal{R}_{A,l}}|\mathrm{tr}((\chi\circ d)'\Delta_\Theta)|\right\} \le 8M\mathbb{E}_{\phi x}\left\{\sup_{\Delta_\Theta\in\mathcal{R}_{A,l}}\|\chi\circ d)\|\,\|\Delta_\Theta\|_*\right\} \\
&\le 8M\mathbb{E}_{\phi x}[\|\chi\circ d)\|]\sup_{\Delta_\Theta\in\mathcal{R}_{A,l}}4\sqrt{2r}\,\|\Delta_\Theta\|_F = 32M\mathbb{E}_{\phi x}[\|\chi\circ d)\|]\sqrt{2r(\frac{6}{5})^l A_{NT}}, \qquad \text{(D.10)}
\end{aligned}
$$

where the first inequality follows from the symmetrization argument (e.g., Theorem 2.1 in chapter 2 of Koltchinskii, 2011), the second inequality follows from the Talagrand contraction inequality (Theorem 2.2 in chapter 2 of Koltchinskii, 2011), the fourth inequality follows from Lemma D.3 and the last equality follows from the definition of $\mathcal{R}_{A,l}$. ■

Proof of Theorem 3.3:

(1) The random missing cases.

Step 1: By (3.6), we have

$$
\begin{aligned}
0 &\le \frac{1}{2}\sum_{i=1}^N\sum_{t=1}^T d_{it}(y_{it}-x_{it}'\beta^0-\Theta_{it}^0)^2-\frac{1}{2}\sum_{i=1}^N\sum_{t=1}^T d_{it}(y_{it}-x_{it}'\hat\beta^{(0)}-\hat\Theta_{it}^{(0)})^2+\nu_{NT}\left(\left\|\Theta^0\right\|_*-\left\|\hat\Theta^{(0)}\right\|_*\right) \\
&= \sum_{i=1}^N\sum_{t=1}^T d_{it}v_{it}(\hat\Delta_{\Theta,it}^{(0)}+x_{it}'\hat\Delta_\beta^{(0)})-\frac{1}{2}\sum_{i=1}^N\sum_{t=1}^T d_{it}(\hat\Delta_{\Theta,it}^{(0)}+x_{it}'\hat\Delta_\beta^{(0)})^2+\nu_{NT}\left(\left\|\Theta^0\right\|_*-\left\|\hat\Theta^{(0)}\right\|_*\right). \qquad \text{(D.11)}
\end{aligned}
$$

It follows that w.p.a.1,

$$
\begin{aligned}
\frac{d\mu}{2}\left\|\hat\Delta_\Theta^{(0)}\right\|_F^2 &\le \frac{d}{2}vec(\hat\Delta_\Theta^{(0)})'M_{vec(x)}vec(\hat\Delta_\Theta^{(0)}) \le \frac{d}{2}\sum_{i=1}^N\sum_{t=1}^T(\hat\Delta_{\Theta,it}^{(0)}+x_{it}'\hat\Delta_\beta^{(0)})^2 \\
&\le \frac{1}{2}\sum_{i=1}^N\sum_{t=1}^T\mathbb{E}_{\phi x}(d_{it})(\hat\Delta_{\Theta,it}^{(0)}+x_{it}'\hat\Delta_\beta^{(0)})^2 \\
&\le \sum_{i=1}^N\sum_{t=1}^T d_{it}v_{it}\hat\Delta_{\Theta,it}^{(0)}+\sum_{i=1}^N\sum_{t=1}^T d_{it}v_{it}x_{it}'\hat\Delta_\beta^{(0)}+\nu_{NT}(\left\|\Theta^0\right\|_*-\left\|\hat\Theta^{(0)}\right\|_*) \\
&-\frac{1}{2}\sum_{i=1}^N\sum_{t=1}^T[d_{it}-\mathbb{E}_{\phi x}(d_{it})]\,(\hat\Delta_{\Theta,it}^{(0)}+x_{it}'\hat\Delta_\beta^{(0)})^2 \\
&\le (\|d\circ v\|+\nu_{NT})\left\|\hat\Delta_\Theta^{(0)}\right\|_*+M\sqrt{NT}\left\|\hat\Delta_\beta^{(0)}\right\|+\frac{1}{2}(M\sqrt{NT}\left\|\hat\Delta_\beta^{(0)}\right\|^2 \\
&+MT^{\frac{1}{2}}N^{\frac{1}{4}}\left\|\hat\Delta_\beta^{(0)}\right\|\left\|\hat\Delta_\Theta^{(0)}\right\|_*+A_{NT}+\frac{1}{\sqrt{c_{NT}}}\left\|\hat\Delta_\Theta^{(0)}\right\|_F^2+\mathcal{E}_{NT}) \\
&\le M(N^{\frac{1}{2}}T^{\frac{1}{4}}+N^{\frac{1}{4}}T^{\frac{1}{2}})4\sqrt{2r}\left\|\hat\Delta_\Theta^{(0)}\right\|_F+M\sqrt{NT}\left\|\hat\Delta_\beta^{(0)}\right\|+M\sqrt{NT}\left\|\hat\Delta_\beta^{(0)}\right\|^2
\end{aligned}
$$

$$
\begin{aligned}
&+ MT^{\frac{1}{2}}N^{\frac{1}{4}}\left\|\hat{\Delta}_{\beta}^{(0)}\right\| 4\sqrt{2r}\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F + A_{NT} + \frac{1}{\sqrt{c_{NT}}}\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F^2 + \mathcal{E}_{NT} \\
&\leq \frac{3[M(N^{\frac{1}{2}}T^{\frac{1}{4}} + N^{\frac{1}{4}}T^{\frac{1}{2}})4\sqrt{2r}]^2}{\underline{d}\mu} + \frac{\underline{d}\mu}{6}\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F^2 + M\sqrt{NT}\left\|\hat{\Delta}_{\beta}^{(0)}\right\| + M\sqrt{NT}\left\|\hat{\Delta}_{\beta}^{(0)}\right\|^2 \\
&+ \frac{3[MT^{\frac{1}{2}}N^{\frac{1}{4}}\left\|\hat{\Delta}_{\beta}^{(0)}\right\| 4\sqrt{2r}]^2}{\underline{d}\mu} + \frac{\underline{d}\mu}{6}\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F^2 + A_{NT} + \frac{1}{\sqrt{c_{NT}}}\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F^2 + \mathcal{E}_{NT},
\end{aligned} \tag{D.12}
$$

where the first inequality follows from Assumption 6(i), the second inequality follows from concentrating out β, the third inequality follows from $\mathbb{E}_{\phi x}(d_{it}) \geq \underline{d}$ for all (i,t), the fourth inequality follows from expression (D.11), the fifth inequality follows from Assumption 5(ii) and Lemma D.5, the sixth inequality follows from the following facts: (1) $\|d \circ v\| = O_p(N^{\frac{1}{2}}T^{\frac{1}{4}} + N^{\frac{1}{4}}T^{\frac{1}{2}})$ by Lemma D.1, (2) $\nu_{NT} = 2c_1(N^{\frac{1}{2}}T^{\frac{1}{4}} + N^{\frac{1}{4}}T^{\frac{1}{2}})$ and (3) $\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_* \leq 4\sqrt{2r}\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F$ by Lemma D.3, and the last inequality follows from the fact that $2ab \leq a^2 + b^2$.

From expression (D.12) it's not difficult to see that

$$
\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F^2 \leq M(N^{\frac{1}{2}}T^{\frac{1}{4}} + N^{\frac{1}{4}}T^{\frac{1}{2}})^2 + M\sqrt{NT}\left\|\hat{\Delta}_{\beta}^{(0)}\right\| + MT\sqrt{N}\left\|\hat{\Delta}_{\beta}^{(0)}\right\|^2 + A_{NT} + \mathcal{E}_{NT}. \tag{D.13}
$$

Plugging this back to (D.12) and noticing that $\hat{\Delta}_{\beta}^{(0)}\sum_{i=1}^N\sum_{t=1}^T x_{it}x_{it}'\hat{\Delta}_{\beta}^{(0)} \leq \sum_{i=1}^N\sum_{t=1}^T(\hat{\Delta}_{\Theta,it}^{(0)} + x_{it}'\hat{\Delta}_{\beta}^{(0)})^2 + \left|2\sum_{i=1}^N\sum_{t=1}^T\hat{\Delta}_{\Theta,it}^{(0)}x_{it}'\hat{\Delta}_{\beta}^{(0)}\right| + \left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F^2$ and $plim\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T x_{it}x_{it}'$ is positive definite under Assumption 3(i), after some calculations, we have

$$
\left\|\hat{\Delta}_{\beta}^{(0)}\right\| = \sqrt{\frac{1}{NT}O_p(M(N^{\frac{1}{2}}T^{\frac{1}{4}} + N^{\frac{1}{4}}T^{\frac{1}{2}})^2 + A_{NT} + \mathcal{E}_{NT})} = O_p(c_{NT}^{-1/4}),
$$

where the equality follows from $\mathcal{E}_{NT} = 3r(16M)^2\sqrt{c_{NT}}(\mathbb{E}_{\phi x}[\|\chi \circ d)\|])^2 = 3r(16M)^2\sqrt{c_{NT}}T\sqrt{N}$ by Lemma D.4 and $A_{NT} = M\sqrt{NTc_{NT}\log(N+T)}$. Plugging this back into (D.13), we have

$$
\frac{1}{\sqrt{NT}}\left\|\hat{\Delta}_{\Theta}^{(0)}\right\|_F = O_p(\sqrt{\frac{\mathcal{E}_{NT}}{NT}}) = O_p(c_{NT}^{-1/4}).
$$

Step 2: Now we prove the consistency of $\hat{r}$, $\hat{f}^{(0)}$ and $\hat{\lambda}^{(0)}$. This step is similar to Theorem 3.2 in Hong et al. (2023). Let $\{\hat{\sigma}_s,\ s \in [N \wedge T]\}$ and $\{\sigma_s,\ s \in [N \wedge T]\}$ denote the singular values of $\frac{\hat{\Theta}^{(0)}}{\sqrt{NT}}$ and $\frac{\Theta^0}{\sqrt{NT}}$, respectively. By Weyl's inequality for singular values, $|\hat{\sigma}_s - \sigma_s| \leq \left\|\frac{\hat{\Delta}_{\Theta}^{(0)}}{\sqrt{NT}}\right\| \leq \left\|\frac{\hat{\Delta}_{\Theta}^{(0)}}{\sqrt{NT}}\right\|_F = O_p(c_{NT}^{-1/4})$ for all s. Since σ_s is bounded away from zero in probability for $s \leq r$ and $\sigma_s = 0$ for $s > r$, we have $\hat{\sigma}_s = \sigma_s + o_p(1) \geq \sqrt{c_{NT}^{-1/4}}\hat{\sigma}_1 = O_p(c_{NT}^{-1/4})$ for $s \leq r$ and $\hat{\sigma}_s = 0 + O_p(c_{NT}^{-1/4}) < \sqrt{c_{NT}^{-1/4}}$ for $s > r$, thus $\Pr(\hat{r} = r) \to 1$.

Recall that $\{\hat{\mathcal{U}}_1^{(0)}, ..., \hat{\mathcal{U}}_r^{(0)}\}$ and $\{\mathcal{U}_1^0, ..., \mathcal{U}_r^0\}$ denote the left-singular vectors corresponding to $\{\hat{\sigma}_1, ..., \hat{\sigma}_r\}$ and $\{\sigma_1, ..., \sigma_r\}$, respectively. By the Davis-Kahan $\sin\Theta$ theorem,

$$
\left\|\hat{\mathcal{U}}_s^{(0)} - \mathcal{U}_s^0\right\| \leq \frac{\sqrt{2}}{\eta}\left\|\frac{\hat{\Delta}_{\Theta}^{(0)}}{\sqrt{NT}}\right\| \leq \frac{\sqrt{2}}{\eta}\left\|\frac{\hat{\Delta}_{\Theta}^{(0)}}{\sqrt{NT}}\right\|_F = O_p(c_{NT}^{-1/4}) \text{ for } s \in [r],
$$

where the equality is because $\eta = \min_s\{|\sigma_{s-1} - \hat{\sigma}_s| \wedge |\sigma_{s+1} - \hat{\sigma}_s|\}$ is bounded and bounded away from zero in probability. It follows that

$$\left\|\sqrt{T\hat{\sigma}_s}\hat{\mathcal{U}}_s^{(0)}-\sqrt{T\sigma_s}\mathcal{U}_s^0\right\|\leq\left|\sqrt{T\hat{\sigma}_s}-\sqrt{T\sigma_s}\right|\left\|\hat{\mathcal{U}}_s^{(0)}\right\|+\sqrt{T\sigma_s}\left\|\hat{\mathcal{U}}_s^{(0)}-\mathcal{U}_s^0\right\|=O_p(c_{NT}^{-1/4}\sqrt{T}),$$

i.e., $\left\|\hat{f}^{(0)}-f^0\right\|=O_p(c_{NT}^{-1/4}\sqrt{T})$. By symmetry, we also have $\left\|\hat{\lambda}^{(0)}-\lambda^0\right\|=O_p(c_{NT}^{-1/4}\sqrt{N})$.

(2) The block-type missing cases.

Given that Assumption 6(i)b holds for at least one completely observed data block, without loss of generality, we assume the block $\{i\in[N_o],t\in[T]\}$ satisfies Assumption 6(i)b. The nuclear norm regularized estimation applied on this block produces $\hat{r}$, $\hat{\beta}$, $\{\hat{f}_t^{(0)},t\in[T]\}$ and $\{\hat{\lambda}_i^{(0)},i\in[N_o]\}$. Since the data block is completely observed, Theorem 2 of Moon and Weidner (2023) for NNR estimation of panel without missing data is applicable here. For completeness, we outline the main steps (with slight adjustments).

Let (y_{bl},x_{bl}) denote the completely observed block of (y,x), and let v_{bl}, Θ_{bl}, $\hat{\Theta}_{bl}^{(0)}$ and $\hat{\Delta}_{\Theta bl}^{(0)}$ denote the corresponding blocks. The sum of squared residuals given Θ_{bl} and concentrating out β is $SSR(\Theta_{bl})=\frac{1}{2}vec(y_{bl}-\Theta_{bl})'M_{vec(x_{bl})}vec(y_{bl}-\Theta_{bl})$, and

$$\begin{aligned}0&\leq SSR(\Theta_{bl}^0)-SSR(\hat{\Theta}_{bl}^{(0)})+\nu_{N_oT}\left(\left\|\Theta_{bl}^0\right\|_*-\left\|\hat{\Theta}_{bl}^{(0)}\right\|_*\right)\\&=-vec(y_{bl}-\Theta_{bl}^0)'M_{vec(x_{bl})}vec(\Theta_{bl}^0-\hat{\Theta}_{bl}^{(0)})\\&\quad-\frac{1}{2}vec(\Theta_{bl}^0-\hat{\Theta}_{bl}^{(0)})'M_{vec(x_{bl})}vec(\Theta_{bl}^0-\hat{\Theta}_{bl}^{(0)})+\nu_{N_oT}\left(\left\|\Theta_{bl}^0\right\|_*-\left\|\hat{\Theta}_{bl}^{(0)}\right\|_*\right).\end{aligned}\tag{D.14}$$

It follows that

$$\begin{aligned}\frac{\mu}{2}\left\|\hat{\Delta}_{\Theta bl}^{(0)}\right\|_F^2&\leq\frac{1}{2}vec(\Theta_{bl}^0-\hat{\Theta}_{bl}^{(0)})'M_{vec(x_{bl})}vec(\Theta_{bl}^0-\hat{\Theta}_{bl}^{(0)})\\&\leq-vec(y_{bl}-\Theta_{bl}^0)'M_{vec(x_{bl})}vec(\Theta_{bl}^0-\hat{\Theta}_{bl}^{(0)})+\nu_{N_oT}\left(\left\|\Theta_{bl}^0\right\|_*-\left\|\hat{\Theta}_{bl}^{(0)}\right\|_*\right)\\&=vec(v_{bl})'M_{vec(x_{bl})}vec(\hat{\Theta}_{bl}^{(0)}-\Theta_{bl}^0)+\nu_{N_oT}\left(\left\|\Theta_{bl}^0\right\|_*-\left\|\hat{\Theta}_{bl}^{(0)}\right\|_*\right)\\&\leq\left\|mat(M_{vec(x_{bl})}vec(v_{bl})')\right\|\left\|\hat{\Delta}_{\Theta bl}^{(0)}\right\|_*+\nu_{N_oT}\left(\left\|\Theta_{bl}^0\right\|_*-\left\|\hat{\Theta}_{bl}^{(0)}\right\|_*\right)\\&\leq(\|v_{bl}\|+\left\|mat(P_{vec(x_{bl})}vec(v_{bl}))\right\|+\nu_{N_oT})\left\|\hat{\Delta}_{\Theta bl}^{(0)}\right\|_*\\&\leq 12c_1\sqrt{2r}(N^{\frac{1}{2}}T^{\frac{1}{4}}+N^{\frac{1}{4}}T^{\frac{1}{2}})\left\|\hat{\Delta}_{\Theta bl}^{(0)}\right\|_F\ \text{w.p.a.1},\end{aligned}$$

where the first inequality follows from Assumption 6(ii), the second inequality follows from expression (D.14), the fifth inequality follows from $\|v_{bl}\|\leq\|d\circ v\|$ because the $T\times N_o$ matrix v_{bl} is a submatrix of $d\circ v$, Lemma D.1, $\left\|mat(P_{vec(x_{bl})}vec(v_{bl}))\right\|=O_p(1)$ by Assumption 5(ii), $\nu_{NT}=2c_1(N^{\frac{1}{2}}T^{\frac{1}{4}}+N^{\frac{1}{4}}T^{\frac{1}{2}})$ and Lemma D.3. Thus $\frac{1}{\sqrt{NT}}\left\|\hat{\Delta}_{\Theta bl}^{(0)}\right\|_F=O_p(c_{NT}^{-1/2})$. This together with $\frac{1}{2}\sum_{i=1}^N\sum_{t=1}^T(\hat{\Delta}_{\Theta bl,it}^{(0)}+x_{it}'\hat{\Delta}_\beta^{(0)})^2\leq\sum_{i=1}^{N_0}\sum_{t=1}^T v_{it}\hat{\Delta}_{\Theta bl,it}^{(0)}+\sum_{i=1}^{N_0}\sum_{t=1}^T v_{it}x_{it}'\hat{\Delta}_\beta^{(0)}+\nu_{N_0T}(\left\|\Theta_{bl}^0\right\|_*-\left\|\hat{\Theta}_{bl}^{(0)}\right\|_*)$ implies that $\left\|\hat{\Delta}_\beta^{(0)}\right\|=O_p(\frac{N^{\frac{1}{2}}T^{\frac{1}{4}}+N^{\frac{1}{4}}T^{\frac{1}{2}}}{\sqrt{NT}})=O_p(c_{NT}^{-1/2})$. In addition, similar to Step (2) of the random missing case, $\frac{1}{\sqrt{NT}}\left\|\hat{\Delta}_{\Theta bl}^{(0)}\right\|_F=O_p(c_{NT}^{-1/2})$ implies that $\Pr(\hat{r}=r)\to 1$, $\left\|\hat{f}^{(0)}-f^0\right\|=O_p(c_{NT}^{-1/2}\sqrt{T})$ and $\sqrt{\sum_{i=1}^{N_0}\left\|\hat{\lambda}_i^{(0)}-\lambda_i^0\right\|^2}=O_p(c_{NT}^{-1/2}\sqrt{N})$. For

$N_o < i \leq N$,

$$\begin{aligned}\hat{\lambda}_i^{(0)} - \lambda_i^0 &= (\sum_{t=1}^{T} \hat{f}_t^{(0)} \hat{f}_t^{(0)\prime})^{-1} \sum_{t=1}^{T} \hat{f}_t^{(0)} (y_{it} - x_{it}\hat{\beta}) - \lambda_i^0 \\ &= (\sum_{t=1}^{T} \hat{f}_t^{(0)} \hat{f}_t^{(0)\prime})^{-1} \sum_{t=1}^{T} \hat{f}_t^{(0)} ((f_t^0 - \hat{f}_t^{(0)})'\lambda_i^0 + v_{it} - x_{it}\hat{\Delta}_\beta^{(0)}).\end{aligned}$$

Given that N_o/N and T_o/T are bounded away from zero, $\left\|\hat{f}^{(0)} - f^0\right\| = O_p(c_{NT}^{-1/2}\sqrt{T})$, $\sum_{t=1}^{T} f_t^0 v_{it} = O_p(\sqrt{T})$, $\sqrt{\sum_{i=N_0+1}^{N} \left\|\lambda_i^0\right\|^2} = O_p(\sqrt{N})$ and $\left\|\hat{\Delta}_\beta^{(0)}\right\| = O_p(c_{NT}^{-1/2})$, it is not difficult to see that $\sqrt{\sum_{i=N_0+1}^{N} \left\|\hat{\lambda}_i^{(0)} - \lambda_i^0\right\|^2} = O_p(c_{NT}^{-1/2}\sqrt{N})$. Thus $\sqrt{\sum_{i=1}^{N} \left\|\hat{\lambda}_i^{(0)} - \lambda_i^0\right\|^2}$ is also $O_p(c_{NT}^{-1/2}\sqrt{N})$. ■

E Proofs of Proposition 4.1 and Theorem 4.1

To prove Proposition 4.1, we state and prove four lemmas.

Lemma E.1 *Suppose that Assumptions 1, 2, 4(i), 5 and 7 hold. Then as* $(N,T) \to \infty$,

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

(ii) $\left\|W_\lambda^{0\prime} Q_{\lambda\lambda'}^{-1} S_\lambda\right\| = O_p(\sqrt{\frac{N}{T}} + \frac{N}{T})$ *and* $\left\|W_f^{0\prime} Q_{ff'}^{-1} S_f\right\| = O_p(\sqrt{\frac{T}{N}} + \frac{T}{N})$;

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

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

Proof. The proof is similar to that of Lemma C.3 in Su and Wang (2024) with slightly different notations. "W" and "Q" here correspond to "U" and "H" there, respectively. For the readers' convenience, we also provide the main details here. It suffices to prove the first half of parts (i)-(iv) in the above lemma as the second half can be done analogously. Below, we shall consider the random missing cases first.

(i) Recall that $S_{\lambda_i} = \sum_{t=1}^{T} d_{it} v_{it} f_t^0$, $S_\lambda = (S_{\lambda_1}', ..., S_{\lambda_N}')'$ and W_λ^0 is defined in (A.7). We need to show that for any (p,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 $O_p(\sqrt{\frac{N}{T}} + \frac{N}{T})$, where 1_q^r denotes the $r \times 1$ vector with the q-th element being one and the other elements being zeros. This is equivalent to the following expression, which will be discussed later.

$$\sum_{i=1}^{N} \sum_{t=1}^{T} (\sum_{t=1}^{T} d_{it} f_t^0 f_t^{0\prime})^{-1} d_{it} v_{it} f_t^0 \lambda_i^{0\prime} = O_p(\sqrt{\frac{N}{T}} + \frac{N}{T}). \tag{E.1}$$

(ii) Since $Q_{\lambda\lambda'} = L_{\lambda\lambda'} - \frac{cT}{N} W_\lambda^0 W_\lambda^{0\prime}$, by Woodbury identity we have

$$\begin{aligned} Q_{\lambda\lambda'}^{-1} &= L_{\lambda\lambda'}^{-1} - L_{\lambda\lambda'}^{-1} W_\lambda^0 (-\frac{N}{cT} I_{r^2} + W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} W_\lambda^0)^{-1} W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1}, \qquad \text{(E.2)} \\ \text{thus } W_\lambda^{0\prime} Q_{\lambda\lambda'}^{-1} S_\lambda &= W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} S_\lambda + W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} W_\lambda^0 (\frac{N}{cT} I_{r^2} - W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} W_\lambda^0)^{-1} W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} S_\lambda. \end{aligned}$$

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

$$\left\| (\frac{N}{cT} I_{r^2} - W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} W_\lambda^0)^{-1} \right\| \leq \frac{cT}{N}, \tag{E.3}$$

thus it remains to show $\left\| W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} W_\lambda^0 \right\| = O_p(\frac{N}{T})$. This follows from (1) by Assumption 2(ii),

$$\left\| W_\lambda^0 \right\| = O_p(\sqrt{N}), \tag{E.4}$$

and (2) by Lemma C.2(ii) in the appendix of Su and Wang (2024),

$$\left\| L_{\lambda\lambda'}^{-1} \right\| = O_p(\frac{1}{T}). \tag{E.5}$$

(iii) From eqn. (E.2), we have

$$L_{f\lambda'} Q_{\lambda\lambda'}^{-1} S_\lambda = L_{f\lambda'} L_{\lambda\lambda'}^{-1} S_\lambda + L_{f\lambda'} L_{\lambda\lambda'}^{-1} W_\lambda^0 (\frac{N}{cT} I_{r^2} - W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} W_\lambda^0)^{-1} W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} S_\lambda. \tag{E.6}$$

The first term on the right hand side is $O_p(\sqrt{N} + \frac{N}{\sqrt{T}})$ due to

$$\begin{aligned} [L_{f\lambda'} L_{\lambda\lambda'}^{-1} S_\lambda]_s &= \sum_{i=1}^N d_{is} \lambda_i^0 f_s^{0\prime} (\sum_{t=1}^T d_{it} f_t^0 f_t^{0\prime})^{-1} (\sum_{t=1}^T d_{it} v_{it} f_t^0) \\ &= [f_s^{0\prime} \sum_{i=1}^N \sum_{t=1}^T (\sum_{t=1}^T d_{it} f_t^0 f_t^{0\prime})^{-1} d_{it} v_{it} f_t^0 \lambda_i^{0\prime} d_{is}]' = O_p(\sqrt{\frac{N}{T}} + \frac{N}{T}), \end{aligned} \tag{E.7}$$

where the last equality will be proved later. From Assumption 2 it's easy to see that $\left\| L_{f\lambda'} \right\| = O_p(\sqrt{NT})$. This together with equations (E.3)-(E.5) and part (i) implies that the second term on the right hand of (E.6) is also $O_p(\sqrt{N} + \frac{N}{\sqrt{T}})$.

(iv) From eqn. (E.2), we have

$$J_{f\lambda'} Q_{\lambda\lambda'}^{-1} S_\lambda = J_{f\lambda'} L_{\lambda\lambda'}^{-1} S_\lambda + J_{f\lambda'} L_{\lambda\lambda'}^{-1} W_\lambda^0 (\frac{N}{cT} I_{r^2} - W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} W_\lambda^0)^{-1} W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} S_\lambda.$$

The first term on the right hand is $O_p(\sqrt{N} + \frac{N}{\sqrt{T}})$ due to the following expression, which will be discussed later.

$$[J_{f\lambda'} L_{\lambda\lambda'}^{-1} S_\lambda]_s = -\sum_{i=1}^N \sum_{t=1}^T (\sum_{t=1}^T d_{it} f_t^0 f_t^{0\prime})^{-1} d_{is} v_{is} d_{it} v_{it} f_t^0 = O_p(\sqrt{\frac{N}{T}} + \frac{N}{T}). \tag{E.8}$$

From Assumption 4(i) it's easy to see that $\left\| J_{f\lambda'} \right\| = O_p(\sqrt{NT})$. This together with equations (E.3)-(E.5) and part (i) implies that the second term on the right hand is also $O_p(\sqrt{N} + \frac{N}{\sqrt{T}})$.

Proof for expression (E.1): Let $A_{iF} = \frac{1}{T} \sum_{t=1}^T d_{it} f_t^0 f_t^{0\prime}$ and $\bar{A}_{iF} = \mathbb{E}_\phi(A_{iF})$. We have

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

$$
\begin{aligned}
&\leq O_p(\sqrt{\frac{N}{T}})+\sqrt{\sum\nolimits_{i=1}^{N}\left\|A_{iF}^{-1}-\bar{A}_{iF}^{-1}\right\|^2}\sqrt{\sum\nolimits_{i=1}^{N}\left\|\frac{1}{T}\sum\nolimits_{t=1}^{T}f_t^0 d_{it}v_{it}\lambda_i^{0\prime}\right\|_F^2}\\
&= O_p(\sqrt{\frac{N}{T}})+O_p(\sqrt{\frac{N}{T}})O_p(\sqrt{\frac{N}{T}})=O_p(\sqrt{\frac{N}{T}}+\frac{N}{T}).
\end{aligned}
\tag{E.9}
$$

The second inequality of expression (E.9) is due to

$$
\begin{aligned}
&\mathbb{E}(\left\|\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\bar{A}_{iF}^{-1}f_t^0\lambda_i^{0\prime}d_{it}v_{it}\right\|_F^2)\\
=\ &\mathbb{E}(\mathbb{E}_\phi(\left\|\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\bar{A}_{iF}^{-1}f_t^0\lambda_i^{0\prime}d_{it}v_{it}\right\|_F^2))\\
\leq\ &\mathbb{E}(r^2\sum\nolimits_{i,j=1}^{N}\sum\nolimits_{t,s=1}^{T}\left\|\bar{A}_{iF}^{-1}\right\|\left\|f_t^0\right\|\left\|\lambda_i^0\right\|\left\|\bar{A}_{jF}^{-1}\right\|\left\|f_s^0\right\|\left\|\lambda_j^0\right\||\mathbb{E}_\phi(d_{it}v_{it}d_{js}v_{js})|)\\
\leq\ &M\mathbb{E}(\sum\nolimits_{t,s=1}^{T}\sum\nolimits_{i,j=1}^{N}|\mathbb{E}_\phi(d_{it}v_{it}d_{js}v_{js})|)\leq NTM
\end{aligned}
\tag{E.10}
$$

by Assumptions 2(i) and 7(ii). The first equality of expression (E.9) is due to:

$$
\sum\nolimits_{i=1}^{N}\left\|A_{iF}^{-1}-\bar{A}_{iF}^{-1}\right\|^2\leq\sum\nolimits_{i=1}^{N}\left\|A_{iF}-\bar{A}_{iF}\right\|^2\sup_i\left\|A_{iF}^{-1}\right\|^2\sup_i\left\|\bar{A}_{iF}^{-1}\right\|^2=O_p(\frac{N}{T})
\tag{E.11}
$$

by Assumption 1 and eqn. (E.5), and by Assumption 5(i),

$$
\mathbb{E}(\sum\nolimits_{i=1}^{N}\left\|\sum\nolimits_{t=1}^{T}f_t^0\lambda_i^{0\prime}d_{it}v_{it}\right\|_F^2)\leq\mathbb{E}(\sum\nolimits_{t,s=1}^{T}\left\|f_t^0\right\|\left\|f_s^0\right\|\left\|\lambda_i^0\right\|^2\sum\nolimits_{i=1}^{N}|\mathbb{E}_\phi(d_{it}v_{it}d_{is}v_{is})|)\leq NTM.
$$

Proof for expression (E.7): Similar to eqn. (E.10), for all h we have

$$
\begin{aligned}
&\mathbb{E}_\phi(\left\|\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\bar{A}_{iF}^{-1}f_t^0\lambda_i^{0\prime}d_{it}v_{it}d_{ih}\right\|_F^2)\\
=\ &\mathbb{E}_\phi(\mathbb{E}_{\phi d}(\left\|\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\bar{A}_{iF}^{-1}f_t^0\lambda_i^{0\prime}d_{ih}d_{it}v_{it}\right\|_F^2))\\
\leq\ &\mathbb{E}_\phi(r^2\sum\nolimits_{i,j=1}^{N}\sum\nolimits_{t,s=1}^{T}\left\|\bar{A}_{iF}^{-1}\right\|\left\|f_t^0\right\|\left\|\lambda_i^0 d_{ih}\right\|\left\|\bar{A}_{jF}^{-1}\right\|\left\|f_s^0\right\|\left\|\lambda_j^0 d_{jh}\right\||\mathbb{E}_{\phi d}(d_{it}v_{it}d_{js}v_{js})|)\\
\leq\ &M\sum\nolimits_{t,s=1}^{T}\sum\nolimits_{i,j=1}^{N}|\mathbb{E}_\phi(d_{it}v_{it}d_{js}v_{js})|\leq NTM,
\end{aligned}
\tag{E.12}
$$

where $\mathbb{E}_{\phi d}(\cdot)$ denotes the expectation conditioning on ϕ and d, the first inequality is because v_{it} is independent with d, and the last two inequalities follow from Assumptions 2(i) and 7(ii). Also,

$$
\begin{aligned}
&\left\|\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}(A_{iF}^{-1}-\bar{A}_{iF}^{-1})f_t^0\lambda_i^{0\prime}d_{it}v_{it}d_{ih}\right\|_F^2\\
\leq\ &(\sum\nolimits_{i=1}^{N}\left\|A_{iF}^{-1}-\bar{A}_{iF}^{-1}\right\|^2)(\sum\nolimits_{i=1}^{N}\left\|\sum\nolimits_{t=1}^{T}f_t^0 d_{it}v_{it}\lambda_i^{0\prime}d_{ih}\right\|_F^2)=O_p(\frac{N}{T})O_p(NT)=O_p(N^2),
\end{aligned}
$$

where the equality is due to eqn. (E.11) and by Assumption 5

$$
\begin{aligned}
& \mathbb{E}_\phi(\sum_{i=1}^{N}\left\|\sum_{t=1}^{T} f_t^0\lambda_i^{0\prime} d_{it}v_{it}d_{ih}\right\|_F^2) \\
= \; & \mathbb{E}_\phi(\sum_{i=1}^{N}\sum_{t,s=1}^{T}\sum_{p,q=1}^{r} f_{tp}^0 f_{sp}^0(\lambda_{iq}^0 d_{ih})^2\mathbb{E}_{\phi d}(d_{it}v_{it}d_{is}v_{is})) \\
\leq \; & M\sum_{t,s=1}^{T}\left\|f_t^0\right\|\left\|f_s^0\right\|\sum_{i=1}^{N}\left|\mathbb{E}_\phi(d_{it}v_{it}d_{is}v_{is})\right|) \leq NTM.
\end{aligned} \tag{E.13}
$$

Proof for expression (E.8): Let $\xi_{ith} = d_{it}v_{it}d_{ih}v_{ih}$, then we have

$$
\begin{aligned}
-[J_{f\lambda'}L_{\lambda\lambda'}^{-1}S_\lambda]_h &= \sum_{i=1}^{N}\sum_{t=1}^{T}(\sum_{t=1}^{T} d_{it}f_t^0 f_t^{0\prime})^{-1}d_{ih}v_{ih}d_{it}v_{it}f_t^0 \\
&= \frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{A}_{iF}^{-1}\xi_{ith}f_t^0 + (A_{iF}^{-1}-\bar{A}_{iF}^{-1})\xi_{ith}f_t^0) \equiv II_{1h} + II_{2h}.
\end{aligned}
$$

By Assumptions 2(i) and 7(i), we have

$$
\begin{aligned}
\mathbb{E}_\phi(\left\|II_{1h}\right\|^2) &= \mathbb{E}_\phi\left\|\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{A}_{iF}^{-1}f_t^0\xi_{ith}\right\|^2 \\
&\leq \frac{r}{T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\left|\mathbb{E}_\phi(\xi_{ith}\xi_{jsh})\right|\left\|\bar{A}_{iF}^{-1}\right\|\left\|f_t^0\right\|\left\|\bar{A}_{jF}^{-1}\right\|\left\|f_s^0\right\| \\
&\leq \frac{M}{T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\left|\mathbb{E}_\phi(\xi_{ith}\xi_{jsh})\right| = O_p(\frac{NT+N^2}{T^2}).
\end{aligned}
$$

By Assumption 7(i) and eqn. (E.11), we have

$$
\begin{aligned}
\mathbb{E}_\phi\sum_{i=1}^{N}\left\|\sum_{t=1}^{T}\xi_{ith}f_t^0\right\|^2 &\leq \sum_{i=1}^{N}\sum_{t,s=1}^{T}\left\|f_t^0\right\|\left\|f_s^0\right\|\left|\mathbb{E}_\phi(\xi_{ith}\xi_{ish})\right| \leq NTM, \\
\text{and } \left\|II_{2h}\right\|^2 &\leq \sum_{i=1}^{N}\left\|A_{iF}^{-1}-\bar{A}_{iF}^{-1}\right\|^2\frac{1}{T^2}\sum_{i=1}^{N}\left\|\sum_{t=1}^{T}\xi_{ith}f_t^0\right\|^2 = O_p(\frac{N^2}{T^2}).
\end{aligned}
$$

The block-type missing case. For this case, $A_{iF} = \bar{A}_{iF}$, consequently all terms with $A_{iF}^{-1} - \bar{A}_{iF}^{-1}$ in the proof of expressions (E.1), (E.7) and (E.8) become zeros. Thus we have (i) $\left\|W_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\sqrt{\frac{N}{T}})$, (ii) $\left\|W_\lambda^{0\prime}Q_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\sqrt{\frac{N}{T}})$ and (iii) $\left\|L_{f\lambda'}Q_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\sqrt{N})$. However, $\left\|J_{f\lambda'}Q_{\lambda\lambda'}^{-1}S_\lambda\right\|$ is still $O_p(\sqrt{N}+\frac{N}{\sqrt{T}})$ since $\left\|II_{1h}\right\|^2$ is still $O_p(\frac{NT+N^2}{T^2})$. Therefore, $Q_{f\lambda'}Q_{\lambda\lambda'}^{-1}S_\lambda = (L_{f\lambda'} + J_{f\lambda'} - cW_f^0W_\lambda^{0\prime})Q_{\lambda\lambda'}^{-1}S_\lambda$ is still $O_p(\sqrt{N}+\frac{N}{\sqrt{T}})$ and this is what matters for Lemma E.2 below. ■

For the rest of the appendix, all lemmas, propositions and theorems hold for the random missing with $\kappa > 4$ and hold for the block-type missing with $\kappa = \infty$.

Lemma E.2 *Suppose that Assumptions 1-2, 4-5 and 7 hold. Let $\mathbb{S}_1$ be the $(Nr)\times(N+T)r$ selection matrix such that $\mathbb{S}_1 A$ selects the first Nr rows of a $(N+T)r\times(N+T)r$ matrix A. Let $\mathbb{S}_2$ be the $(Tr)\times(N+T)r$ selection matrix such that $\mathbb{S}_2 A$ selects the last Tr rows of A. Then as $(N,T)\to\infty$,*

(i) $\left\|\mathbb{S}_1 Q_{\phi\phi'}^{-1}S_\phi - \bar{L}_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}} + \frac{N^{\frac{1}{2}+\frac{1}{\kappa}}}{T}\right)$;

(ii) $\left\|\mathbb{S}_2 Q_{\phi\phi'}^{-1}S_\phi - \bar{L}_{ff'}^{-1}S_f\right\| = O_p\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{N}} + \frac{T^{\frac{1}{2}+\frac{1}{\kappa}}}{N}\right)$.

Proof. Note that $Q^{-1}_{\phi\phi'}S_\phi = \begin{pmatrix} (Q_{\lambda\lambda'} - Q_{\lambda f'}Q^{-1}_{ff'}Q_{f\lambda'})^{-1}(S_\lambda - Q_{\lambda f'}Q^{-1}_{ff'}S_f) \\ (Q_{ff'} - Q_{f\lambda'}Q^{-1}_{\lambda\lambda'}Q_{\lambda f'})^{-1}(S_f - Q_{f\lambda'}Q^{-1}_{\lambda\lambda'}S_\lambda) \end{pmatrix}$. We prove (i) by showing that

$$\begin{aligned}
\left\|\mathbb{S}_1 Q^{-1}_{\phi\phi'}S_\phi - (Q_{\lambda\lambda'} - Q_{\lambda f'}Q^{-1}_{ff'}Q_{f\lambda'})^{-1}S_\lambda\right\| &= O_p(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}) && \text{(E.14)}\\
\left\|(Q_{\lambda\lambda'} - Q_{\lambda f'}Q^{-1}_{ff'}Q_{f\lambda'})^{-1}S_\lambda - Q^{-1}_{\lambda\lambda'}S_\lambda\right\| &= O_p(\frac{1}{\sqrt{T}} + \frac{\sqrt{N}}{T}), && \text{(E.15)}\\
\left\|Q^{-1}_{\lambda\lambda'}S_\lambda - L^{-1}_{\lambda\lambda'}S_\lambda\right\| &= O_p(\frac{1}{\sqrt{T}} + \frac{\sqrt{N}}{T}) && \text{(E.16)}\\
\left\|L^{-1}_{\lambda\lambda'}S_\lambda - \bar{L}^{-1}_{\lambda\lambda'}S_\lambda\right\| &= O_p(\frac{N^{\frac{1}{2}+\frac{1}{\kappa}}}{T}). && \text{(E.17)}
\end{aligned}$$

Similarly, we prove (ii) by showing that

$$\begin{aligned}
\left\|\mathbb{S}_2 Q^{-1}_{\phi\phi'}S_\phi - (Q_{ff'} - Q_{f\lambda'}Q^{-1}_{\lambda\lambda'}Q_{\lambda f'})^{-1}S_f\right\| &= O_p(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}), && \text{(E.18)}\\
\left\|(Q_{ff'} - Q_{f\lambda'}Q^{-1}_{\lambda\lambda'}Q_{\lambda f'})^{-1}S_f - Q^{-1}_{ff'}S_f\right\| &= O_p(\frac{1}{\sqrt{N}} + \frac{\sqrt{T}}{N}), && \text{(E.19)}\\
\left\|Q^{-1}_{ff'}S_f - L^{-1}_{ff'}S_f\right\| &= O_p(\frac{1}{\sqrt{N}} + \frac{\sqrt{T}}{N}), && \text{(E.20)}\\
\left\|L^{-1}_{ff'}S_f - \bar{L}^{-1}_{ff'}S_f\right\| &= O_p(\frac{T^{\frac{1}{2}+\frac{1}{\kappa}}}{N}). && \text{(E.21)}
\end{aligned}$$

Proof for equations (E.14) and (E.18): $T(Q_{\lambda\lambda'} - Q_{\lambda f'}Q^{-1}_{ff'}Q_{f\lambda'})^{-1}$ and $N(Q_{ff'} - Q_{f\lambda'}Q^{-1}_{\lambda\lambda'}Q_{\lambda f'})$ is the upper-left and lower-right block of $(D^{-\frac{1}{2}}_{TN}Q_{\phi\phi'}D^{-\frac{1}{2}}_{TN})^{-1}$, respectively. By Lemma B.1,

$$\begin{aligned}
\left\|T(Q_{\lambda\lambda'} - Q_{\lambda f'}Q^{-1}_{ff'}Q_{f\lambda'})^{-1}\right\| &\leq \left\|(D^{-\frac{1}{2}}_{TN}Q_{\phi\phi'}D^{-\frac{1}{2}}_{TN})^{-1}\right\| = O_p(1), && \text{(E.22)}\\
\left\|N(Q_{ff'} - Q_{f\lambda'}Q^{-1}_{\lambda\lambda'}Q_{\lambda f'})^{-1}\right\| &\leq \left\|(D^{-\frac{1}{2}}_{TN}Q_{\phi\phi'}D^{-\frac{1}{2}}_{TN})^{-1}\right\| = O_p(1) && \text{(E.23)}
\end{aligned}$$

From equations (A.5)-(A.8), $Q_{\lambda f'} = L_{\lambda f'} + J_{\lambda f'} - cW^0_\lambda W^{0\prime}_f$. Then by Lemma E.1(ii)-(iv),

$$\left\|Q_{\lambda f'}Q^{-1}_{ff'}S_f\right\| \leq \left\|L_{\lambda f'}Q^{-1}_{ff'}S_f\right\| + \left\|J_{\lambda f'}Q^{-1}_{ff'}S_f\right\| + \left\|cW^0_\lambda W^{0\prime}_f Q^{-1}_{ff'}S_f\right\| = O_p(\sqrt{T} + \frac{T}{\sqrt{N}}). \quad \text{(E.24)}$$

It follows that

$$\begin{aligned}
\left\|\mathbb{S}_1 Q^{-1}_{\phi\phi'}S_\phi - (Q_{\lambda\lambda'} - Q_{\lambda f'}Q^{-1}_{ff'}Q_{f\lambda'})^{-1}S_\lambda\right\| &= \left\|(Q_{\lambda\lambda'} - Q_{\lambda f'}Q^{-1}_{ff'}Q_{f\lambda'})^{-1}Q_{\lambda f'}Q^{-1}_{ff'}S_f\right\|\\
&\leq \left\|(Q_{\lambda\lambda'} - Q_{\lambda f'}Q^{-1}_{ff'}Q_{f\lambda'})^{-1}\right\| \left\|Q_{\lambda f'}Q^{-1}_{ff'}S_f\right\|\\
&= O_p(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{N}})
\end{aligned}$$

Similarly,

$$\left\|Q_{f\lambda'}Q^{-1}_{\lambda\lambda'}S_\lambda\right\| = O_p(\sqrt{N} + \frac{N}{\sqrt{T}}), \quad \text{(E.25)}$$

and

$$\left\|\mathbb{S}_2 Q_{\phi\phi'}^{-1} S_\phi - (Q_{ff'} - Q_{f\lambda'}Q_{\lambda\lambda'}^{-1}Q_{\lambda f'})^{-1} S_f\right\| = \left\|(Q_{ff'} - Q_{f\lambda'}Q_{\lambda\lambda'}^{-1}Q_{\lambda f'})^{-1} Q_{f\lambda'}Q_{\lambda\lambda'}^{-1} S_\lambda\right\|$$
$$= O_p(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{T}}).$$

Proof for equations (E.15) and (E.19):

$$(Q_{\lambda\lambda'} - Q_{\lambda f'}Q_{ff'}^{-1}Q_{f\lambda'})^{-1} S_\lambda = Q_{\lambda\lambda'}^{-1} S_\lambda + Q_{\lambda\lambda'}^{-1} Q_{\lambda f'}(Q_{ff'} - Q_{f\lambda'}Q_{\lambda\lambda'}^{-1}Q_{\lambda f'})^{-1} Q_{f\lambda'}Q_{\lambda\lambda'}^{-1} S_\lambda \tag{E.26}$$

It is easy to see that $\|Q_{\lambda f'}\| = O_p(\sqrt{NT})$, and equations (E.2)-(E.5) imply that $\|Q_{\lambda\lambda'}^{-1}\| = O_p(\frac{1}{T})$. These results, along with equations (E.23) and (E.25), imply that the Euclidean norm of the second term on the r.h.s. of (E.26) is $O_p(\frac{1}{\sqrt{T}} + \frac{\sqrt{N}}{T})$. This implies (E.15). By symmetry, we also have (E.19).

Proof for equations (E.16) and (E.20): From eqn. (E.2) we have

$$Q_{\lambda\lambda'}^{-1} S_\lambda = L_{\lambda\lambda'}^{-1} S_\lambda + L_{\lambda\lambda'}^{-1} W_\lambda^0 (\frac{N}{cT} I_{r^2} - W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} W_\lambda^0)^{-1} W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} S_\lambda.$$

By equations (E.3)–(E.5) and Lemma E.1(i), the Euclidean norm of the second term on the right hand side is $O_p(\frac{1}{\sqrt{T}} + \frac{\sqrt{N}}{T})$. By symmetry, we also have $\left\|Q_{ff'}^{-1} S_f - L_{ff'}^{-1} S_f\right\| = O_p(\frac{1}{\sqrt{N}} + \frac{\sqrt{T}}{N})$.

Proof for equations (E.17) and (E.21): Eqn. B.17 in the appendix of Su and Wang (2024) shows that $\left\|L_{\lambda\lambda'} - \bar{L}_{\lambda\lambda'}\right\| = O_p(\sqrt{T} N^{\frac{1}{\kappa}})$. This together with eqn. (E.5) imply that

$$\left\|L_{\lambda\lambda'}^{-1} - \bar{L}_{\lambda\lambda'}^{-1}\right\| = \left\|-L_{\lambda\lambda'}^{-1}(L_{\lambda\lambda'} - \bar{L}_{\lambda\lambda'})\bar{L}_{\lambda\lambda'}^{-1}\right\| = O_p(\frac{N^{\frac{1}{\kappa}}}{T\sqrt{T}}). \tag{E.27}$$

By Assumption 5, $\|S_\lambda\| = O_p(\sqrt{NT})$, thus $\left\|L_{\lambda\lambda'}^{-1} S_\lambda - \bar{L}_{\lambda\lambda'}^{-1} S_\lambda\right\| = O_p(\frac{N^{\frac{1}{2}+\frac{1}{\kappa}}}{T})$. Similarly, $\left\|L_{ff'}^{-1} S_f - \bar{L}_{ff'}^{-1} S_f\right\| = O_p(\frac{T^{\frac{1}{2}+\frac{1}{\kappa}}}{N})$. For the block-type missing, this term is zero since $L_{\lambda\lambda'} = \bar{L}_{\lambda\lambda'}$ and $L_{ff'} = \bar{L}_{ff'}$. ■

Lemma E.3 *Suppose that Assumptions 1-2, 4-5 and 7-8 hold.*

(i) If $\frac{N^{\frac{1}{\kappa}}T^{\frac{1}{\kappa}} + N^{\frac{2}{\kappa}}}{\sqrt{T}} \to 0$ *and* $\frac{N^{\frac{1}{\kappa}}T^{\frac{1}{\kappa}} + T^{\frac{2}{\kappa}}}{\sqrt{N}} \to 0$, *then* $S_\beta - Q_{\beta\phi'}Q_{\phi\phi'}^{-1} S_\phi = O_p(\sqrt{NT} + N + T)$;

(ii) If $\frac{N^{\frac{2}{3}+\frac{4}{3\kappa}}}{T} \to 0$, $\frac{T^{\frac{2}{3}+\frac{4}{3\kappa}}}{N} \to 0$ *and* $\kappa > 4$, *then* $\frac{1}{\sqrt{NT}}(S_\beta - Q_{\beta\phi'}Q_{\phi\phi'}^{-1} S_\phi) - (b_1 + b_2 + b_3 + b_4) \xrightarrow{d} \mathcal{N}(0, \Omega_x)$, *where*

$$b_1 = \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{E}_\phi(d_{jt} v_{jt} d_{it} x_{it}) \lambda_i^{0\prime} [\bar{L}_{ff'}^{-1}]_t \lambda_j^0,$$
$$b_2 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}_\phi(d_{is} v_{is} d_{it} x_{it}) f_t^{0\prime} [\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0,$$
$$b_3 = \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{E}_\phi(d_{it} v_{it} d_{jt} v_{jt}) \delta_i^0 [\bar{L}_{ff'}^{-1}]_t \lambda_j^0,$$
$$b_4 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}_\phi(d_{is} v_{is} d_{it} v_{it}) \omega_t^0 [\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0.$$

Proof. First note that

$$S_\beta - Q_{\beta\phi'}Q_{\phi\phi'}^{-1} S_\phi = (S_\beta - \bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1} S_\phi) - (Q_{\beta\phi'} - \bar{Q}_{\beta\phi'})Q_{\phi\phi'}^{-1} S_\phi + \bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}(Q_{\phi\phi'} - \bar{Q}_{\phi\phi'})Q_{\phi\phi'}^{-1} S_\phi$$

$$\equiv \quad II_{1,1} - II_{1,2} + II_{1,3}. \tag{E.28}$$

We study $II_{1,1}$, $II_{1,2}$ and $II_{1,3}$ in turn.

(1) By equations (B.10) and (B.15), $\bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1} = \bar{Q}_{\beta\phi'}\bar{L}_{\phi\phi'}^{-1} = (\delta^{0\prime}, \omega^{0\prime})$. Then

$$\bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}S_\phi = (\delta^{0\prime}, \omega^{0\prime})S_\phi = \sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}(\delta_i^0 f_t^0 + \omega_t^0\lambda_i^0)d_{it}v_{it}.$$

Since $S_\beta = \sum_{i=1}^N\sum_{t=1}^T x_{it}d_{it}v_{it}$, we have $II_{1,1} = S_\beta - \bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}S_\phi = \sum_{i=1}^N\sum_{t=1}^T \dot{x}_{it}d_{it}v_{it}$. Then by Assumption 8(i),

$$\frac{1}{\sqrt{NT}}II_{1,1} = \frac{1}{\sqrt{NT}}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\dot{x}_{it}d_{it}v_{it} \xrightarrow{d} \mathcal{N}(0, \Omega_x). \tag{E.29}$$

(2) From Assumption 3(iii) and Lemma E.2, we have

$$(Q_{\beta\phi'} - \bar{Q}_{\beta\phi'})[Q_{\phi\phi'}^{-1}S_\phi - ((\bar{L}_{\lambda\lambda'}^{-1}S_\lambda)', (\bar{L}_{ff'}^{-1}S_f)')'] = \sqrt{NT}O_p(\frac{1}{\sqrt{T}} + \frac{N^{\frac{1}{2}+\frac{1}{\kappa}}}{T} + \frac{1}{\sqrt{N}} + \frac{T^{\frac{1}{2}+\frac{1}{\kappa}}}{N}),$$

which is $o_p(N+T)$ if $\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}} \to 0$ and $\frac{T^{\frac{1}{\kappa}}}{\sqrt{N}} \to 0$, or equals $o_p(\sqrt{NT})$ if $\frac{N^{\frac{1}{2}+\frac{1}{\kappa}}}{T} \to 0$ and $\frac{T^{\frac{1}{2}+\frac{1}{\kappa}}}{N} \to 0$. Next, note that

$$(Q_{\beta\phi'} - \bar{Q}_{\beta\phi'})((\bar{L}_{\lambda\lambda'}^{-1}S_\lambda)', (\bar{L}_{ff'}^{-1}S_f)')' = (Q_{\beta\lambda'} - \bar{Q}_{\beta\lambda'})\bar{L}_{\lambda\lambda'}^{-1}S_\lambda + (Q_{\beta f'} - \bar{Q}_{\beta f'})\bar{L}_{ff'}^{-1}S_f.$$

The i-th block of $Q_{\beta\lambda'} - \bar{Q}_{\beta\lambda'}$ is $-\sum_{t=1}^T(d_{it}x_{it} - \mathbb{E}_\phi(d_{it}x_{it}))f_t^{0\prime}$ and the t-th block of $Q_{\beta f'} - \bar{Q}_{\beta f'}$ is $-\sum_{i=1}^N(d_{it}x_{it} - \mathbb{E}_\phi(d_{it}x_{it}))\lambda_i^{0\prime}$. By Assumption 8(ii) and the fact that $\max_{i,t,s}\left|f_t^{0\prime}[T\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0\right| \le M$ and $\max_{i,j,t}\left|\lambda_i^{0\prime}[T\bar{L}_{ff'}^{-1}]_t\lambda_j^0\right| \le M$,

$$\begin{aligned}
& -(Q_{\beta\lambda'} - \bar{Q}_{\beta\lambda'})\bar{L}_{\lambda\lambda'}^{-1}S_\lambda \\
= \; & \sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\sum\nolimits_{s=1}^{T}(d_{it}x_{it} - \mathbb{E}_\phi(d_{it}x_{it}))d_{is}v_{is}f_t^{0\prime}[\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0 \\
= \; & \sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\sum\nolimits_{s=1}^{T}\mathbb{E}_\phi[(d_{it}x_{it} - \mathbb{E}_\phi(d_{it}x_{it}))d_{is}v_{is}]f_t^{0\prime}[\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0 + O_p(\sqrt{N}) \\
= \; & \sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\sum\nolimits_{s=1}^{T}\mathbb{E}_\phi(d_{it}x_{it}d_{is}v_{is})f_t^{0\prime}[\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0 + O_p(\sqrt{N}),
\end{aligned} \tag{E.30}$$

and

$$\begin{aligned}
& -(Q_{\beta f'} - \bar{Q}_{\beta f'})\bar{L}_{ff'}^{-1}S_f \\
= \; & \sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N}\sum\nolimits_{j=1}^{N}(d_{it}x_{it} - \mathbb{E}_\phi(d_{it}x_{it}))d_{jt}v_{jt}\lambda_i^{0\prime}[\bar{L}_{ff'}^{-1}]_t\lambda_j^0 \\
= \; & \sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N}\sum\nolimits_{j=1}^{N}\mathbb{E}_\phi[(d_{it}x_{it} - \mathbb{E}_\phi(d_{it}x_{it}))d_{jt}v_{jt}]\lambda_i^{0\prime}[\bar{L}_{ff'}^{-1}]_t\lambda_j^0 + O_p(\sqrt{T}) \\
= \; & \sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N}\sum\nolimits_{j=1}^{N}\mathbb{E}_\phi(d_{it}x_{it}d_{jt}v_{jt})\lambda_i^{0\prime}[\bar{L}_{ff'}^{-1}]_t\lambda_j^0 + O_p(\sqrt{T}).
\end{aligned} \tag{E.31}$$

It follows that

$$-\frac{1}{\sqrt{NT}}II_{1,2} \quad = \quad -\frac{1}{\sqrt{NT}}(Q_{\beta\phi'} - \bar{Q}_{\beta\phi'})Q_{\phi\phi'}^{-1}S_\phi$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}_\phi(d_{it}x_{it}d_{is}v_{is})f_t^{0\prime}[\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0 \\
&\quad +\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}_\phi(d_{it}x_{it}d_{jt}v_{jt})\lambda_i^{0\prime}[\bar{L}_{ff'}^{-1}]_t\lambda_j^0 + O_p(\frac{\sqrt{N}+\sqrt{T}}{\sqrt{NT}}) \\
&= b_2+b_1+O_p\left(\frac{1}{\sqrt{N}}+\frac{1}{\sqrt{T}}\right).
\end{aligned} \tag{E.32}
$$

(3) By equations (B.17) and (B.18), $\left\|\bar{Q}_{\beta\phi'}D_{TN}^{-\frac{1}{2}}\right\| = O_p(\sqrt{NT})$. By Lemma B.1 and eqn. (B.16), $(D_{TN}^{-\frac{1}{2}}\bar{Q}_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1} = O_p(1)$. These results, together with eqn. (B.20) and Lemma E.2, imply that

$$
\begin{aligned}
&\bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}(Q_{\phi\phi'}-\bar{Q}_{\phi\phi'})[Q_{\phi\phi'}^{-1}S_\phi-((\bar{L}_{\lambda\lambda'}^{-1}S_\lambda)',(\bar{L}_{ff'}^{-1}S_f)')'] \\
=\ &\bar{Q}_{\beta\phi'}D_{TN}^{-\frac{1}{2}}(D_{TN}^{-\frac{1}{2}}\bar{Q}_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}D_{TN}^{-\frac{1}{2}}(Q_{\phi\phi'}-\bar{Q}_{\phi\phi'})D_{TN}^{-\frac{1}{2}}D_{TN}^{\frac{1}{2}}[Q_{\phi\phi'}^{-1}S_\phi-((\bar{L}_{\lambda\lambda'}^{-1}S_\lambda)',(\bar{L}_{ff'}^{-1}S_f)')'] \\
=\ &O_p(\sqrt{NT})O_p(1)O_p\left(\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}}+\frac{T^{\frac{1}{\kappa}}}{\sqrt{N}}+\frac{1}{\sqrt{c_{NT}}}\right)O_p\left(1+\frac{N^{\frac{1}{2}+\frac{1}{\kappa}}}{\sqrt{T}}+\frac{T^{\frac{1}{2}+\frac{1}{\kappa}}}{\sqrt{N}}\right),
\end{aligned}
$$

which is $o_p(\sqrt{NT})$ if $\frac{N^{\frac{2}{3}+\frac{4}{3\kappa}}}{T}\to 0$, $\frac{T^{\frac{2}{3}+\frac{4}{3\kappa}}}{N}\to 0$ and $\kappa>4$, and $o_p(N+T)$ if $\frac{N^{\frac{1}{\kappa}}T^{\frac{1}{\kappa}}+N^{\frac{2}{\kappa}}}{\sqrt{T}}\to 0$ and $\frac{N^{\frac{1}{\kappa}}T^{\frac{1}{\kappa}}+T^{\frac{2}{\kappa}}}{\sqrt{N}}\to 0$.

Now, we consider $\bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}(Q_{\phi\phi'}-\bar{Q}_{\phi\phi'})((\bar{L}_{\lambda\lambda'}^{-1}S_\lambda)',(\bar{L}_{ff'}^{-1}S_f)')'$. Noting that $\bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}=\bar{Q}_{\beta\phi'}\bar{L}_{\phi\phi'}^{-1}$ and $Q_{\phi\phi'}-\bar{Q}_{\phi\phi'}=L_{\phi\phi'}-\bar{L}_{\phi\phi'}+J_{\phi\phi'}$, we have

$$
\begin{aligned}
&\bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}(Q_{\phi\phi'}-\bar{Q}_{\phi\phi'})((\bar{L}_{\lambda\lambda'}^{-1}S_\lambda)',(\bar{L}_{ff'}^{-1}S_f)')' \\
=\ &\bar{Q}_{\beta\phi'}\bar{L}_{\phi\phi'}^{-1}(L_{\phi\phi'}-\bar{L}_{\phi\phi'}+J_{\phi\phi'})((\bar{L}_{\lambda\lambda'}^{-1}S_\lambda)',(\bar{L}_{ff'}^{-1}S_f)')' \\
=\ &\delta^{0\prime}(L_{\lambda\lambda'}-\bar{L}_{\lambda\lambda'})\bar{L}_{\lambda\lambda'}^{-1}S_\lambda+\delta^{0\prime}(L_{\lambda f'}-\bar{L}_{\lambda f'})\bar{L}_{ff'}^{-1}S_f+\omega^{0\prime}(L_{f\lambda'}-\bar{L}_{f\lambda'})\bar{L}_{\lambda\lambda'}^{-1}S_\lambda \\
&+\omega^{0\prime}(L_{ff'}-\bar{L}_{ff'})\bar{L}_{ff'}^{-1}S_f+\delta^{0\prime}J_{\lambda f'}\bar{L}_{ff'}^{-1}S_f+\omega^{0\prime}J_{f\lambda'}\bar{L}_{\lambda\lambda'}^{-1}S_\lambda.
\end{aligned}
$$

By Assumption 8(iii) and the fact that $\mathbb{E}_\phi(\tilde{d}_{it}d_{is}v_{is})=0$, $\mathbb{E}_\phi(\tilde{d}_{it}d_{jt}v_{jt})=0$, $\max_{its}\left|f_t^{0\prime}[T\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0\right|\le M$, $\max_{ijt}\left|\lambda_i^{0\prime}[T\bar{L}_{ff'}^{-1}]_t\lambda_j^0\right|\le M$, $\max_{it}\left|\delta_{ki}^{0\prime}f_t^0\right|\le M(N\vee T)^{\frac{1}{\varrho}}$ and $\max_{it}\left|\omega_{kt}^{0\prime}\lambda_i^0\right|\le M(N\vee T)^{\frac{1}{\varrho}}$ (by Lemma F.3(ii)), we have

$$
\begin{aligned}
&\delta_k^{0\prime}(L_{\lambda\lambda'}-\bar{L}_{\lambda\lambda'})\bar{L}_{\lambda\lambda'}^{-1}S_\lambda \\
=\ &-\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{d}_{it}d_{is}v_{is}\delta_{ki}^{0\prime}f_t^0 f_t^{0\prime}[\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0 = O_p(\sqrt{N}(N\vee T)^{\frac{1}{\varrho}}),
\end{aligned} \tag{E.33}
$$

$$
\begin{aligned}
&\delta_k^{0\prime}(L_{\lambda f'}-\bar{L}_{\lambda f'})\bar{L}_{ff'}^{-1}S_f \\
=\ &-\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{d}_{it}d_{jt}v_{jt}\delta_{ki}^{0\prime}f_t^0\lambda_i^{0\prime}[\bar{L}_{ff'}^{-1}]_t\lambda_j^0 = O_p(\sqrt{T}(N\vee T)^{\frac{1}{\varrho}}),
\end{aligned} \tag{E.34}
$$

$$
\begin{aligned}
&\omega_k^{0\prime}(L_{f\lambda'}-\bar{L}_{f\lambda'})\bar{L}_{\lambda\lambda'}^{-1}S_\lambda \\
=\ &-\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{d}_{it}d_{is}v_{is}\omega_{kt}^{0\prime}\lambda_i^0 f_t^{0\prime}[\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0 = O_p(\sqrt{N}(N\vee T)^{\frac{1}{\varrho}}),
\end{aligned} \tag{E.35}
$$

and

$$\begin{aligned}
&\omega_k^{0\prime}(L_{ff'}-\bar{L}_{ff'})\bar{L}_{ff'}^{-1}S_f \\
= &-\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N}\sum\nolimits_{j=1}^{N}\tilde{d}_{it}d_{jt}v_{jt}\omega_{kt}^{0\prime}\lambda_i^0\lambda_i^{0\prime}[\bar{L}_{ff'}^{-1}]_t\lambda_j^0 = O_p(\sqrt{T}(N\vee T)^{\frac{1}{\varrho}}).
\end{aligned} \tag{E.36}$$

By Assumption 8(iv) and the fact that $\max_{ijt}\left|\delta_{ki}^{0\prime}[T\bar{L}_{ff'}^{-1}]_t\lambda_j^0\right| \leq M(N\vee T)^{\frac{1}{\varrho}}$ and $\max_{its}\left|\omega_{kt}^{0\prime}[T\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0\right| \leq M(N\vee T)^{\frac{1}{\varrho}}$, we have

$$\begin{aligned}
&\delta_k^{0\prime}J_{\lambda f'}\bar{L}_{ff'}^{-1}S_f \\
= &\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N}\sum\nolimits_{j=1}^{N}d_{it}v_{it}d_{jt}v_{jt}\delta_{ki}^{0\prime}[\bar{L}_{ff'}^{-1}]_t\lambda_j^0 \\
= &\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N}\sum\nolimits_{j=1}^{N}\mathbb{E}_\phi(d_{it}v_{it}d_{jt}v_{jt})\delta_{ki}^{0\prime}[\bar{L}_{ff'}^{-1}]_t\lambda_j^0 + O_p(\sqrt{T}(N\vee T)^{\frac{1}{\varrho}}),
\end{aligned} \tag{E.37}$$

and

$$\begin{aligned}
&\omega_k^{0\prime}J_{f\lambda'}\bar{L}_{\lambda\lambda'}^{-1}S_\lambda \\
= &\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\sum\nolimits_{s=1}^{T}d_{is}v_{is}d_{it}v_{it}\omega_{kt}^{0\prime}[\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0 \\
= &\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\sum\nolimits_{s=1}^{T}\mathbb{E}_\phi(d_{is}v_{is}d_{it}v_{it})\omega_{kt}^{0\prime}[\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0 + O_p(\sqrt{N}(N\vee T)^{\frac{1}{\varrho}}).
\end{aligned} \tag{E.38}$$

Then

$$\begin{aligned}
\frac{1}{\sqrt{NT}}II_{1,3} &= \frac{1}{\sqrt{NT}}\bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}(Q_{\phi\phi'}-\bar{Q}_{\phi\phi'})Q_{\phi\phi'}^{-1}S_\phi \\
&= \frac{1}{\sqrt{NT}}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N}\sum\nolimits_{j=1}^{N}\mathbb{E}_\phi(d_{it}v_{it}d_{jt}v_{jt})\delta_{ki}^{0\prime}[\bar{L}_{ff'}^{-1}]_t\lambda_j^0 + \\
&\quad +\frac{1}{\sqrt{NT}}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\sum\nolimits_{s=1}^{T}\mathbb{E}_\phi(d_{is}v_{is}d_{it}v_{it})\omega_{kt}^{0\prime}[\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0 \\
&\quad +O_p(N^{-1/2}(N\vee T)^{\frac{1}{\varrho}}) + T^{-1/2}(N\vee T)^{\frac{1}{\varrho}}) \\
&= b_3 + b_4 + O_p(N^{-1/2}(N\vee T)^{\frac{1}{\varrho}}) + T^{-1/2}(N\vee T)^{\frac{1}{\varrho}}),
\end{aligned}$$

where the remainder term is $o_p(1)$ if $\frac{T^{\frac{1}{\varrho}}}{\sqrt{N}} \to 0$ and $\frac{N^{\frac{1}{\varrho}}}{\sqrt{T}} \to 0$, which is easily satisfied when $\varrho \geq 8$.

Combining the above results, we obtain the desired result. ■

Lemma E.4 *Let R_β and R_ϕ be as defined in (4.1) and (4.2), respectively. Let $R_\phi = (R_\lambda', R_f')'$ where R_λ and R_f are $Nr\times 1$ and $Tr\times 1$, respectively. Suppose that Assumptions 2(i)-(ii) and 3(ii) hold. If $(\hat{\beta}', \hat{\lambda}', \hat{f}')' \in \mathcal{B}_m(\gamma^0)$ w.p.a.1 for any fixed $m>0$, then we have*

(i) $\|R_\beta\| = O_p(\sqrt{NT}\left\|\hat{\lambda}-\lambda^0\right\|\left\|\hat{f}-f^0\right\|)$;

(ii) $\|R_\lambda\| = O_p(\sqrt{NT}\left\|\hat{\beta}-\beta^0\right\|\left\|\hat{f}-f^0\right\| + \sqrt{T}\left\|\hat{\lambda}-\lambda^0\right\|\left\|\hat{f}-f^0\right\| + \sqrt{N}\left\|\hat{f}-f^0\right\|^2)$;

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

Proof. (i) Let R_{β_k} denote the k-th element of R_β, which denotes the remainder term in the first order Taylor expansion of $S_\beta(\hat{\beta},\hat{\phi})$ around $S_\beta = S_\beta(\beta^0,\phi^0)$ with a second order remainder term. Using the integral form of the mean value theorem for vector-valued functions to expand the first order conditions, we can express R_{β_k} as $R_{\beta_k} = R_{\beta\beta\beta_k} + R_{\beta\phi\beta_k} + R_{\phi\beta\beta_k} + R_{\phi\phi\beta_k}$,

$$\begin{aligned}
R_{\beta\beta\beta_k} &= (\hat{\beta}-\beta^0)'(\int_0^1\int_0^{s_1}\partial_{\beta\beta'\beta_k}Q(s_2)ds_2ds_1)(\hat{\beta}-\beta^0),\\
R_{\beta\phi\beta_k} &= R_{\phi\beta\beta_k} = (\hat{\beta}-\beta^0)'(\int_0^1\int_0^{s_1}\partial_{\beta\phi'\beta_k}Q(s_2)ds_2ds_1)(\hat{\phi}-\phi^0),\\
R_{\phi\phi\beta_k} &= (\hat{\phi}-\phi^0)'(\int_0^1\int_0^{s_1}\partial_{\phi\phi'\beta_k}Q(s_2)ds_2ds_1)(\hat{\phi}-\phi^0),
\end{aligned}$$

where $\partial_{\phi\phi'\beta_k}Q(s) = \partial_{\beta_k}\partial_{\phi\phi'}Q(s) = \partial_{\beta_k}\partial_{\phi\phi'}Q(\beta^0+s(\hat{\beta}-\beta^0), \phi^0+s(\hat{\phi}-\phi^0))$ and $\partial_{\beta\phi'\beta_k}Q(s)$ and $\partial_{\beta\beta'\beta_k}Q(s)$ are defined similarly. Since $Q(\cdot)$ is quadratic, we have

$$R_{\beta\beta\beta_k} = 0, \tag{E.39}$$
$$R_{\beta\phi\beta_k} = R_{\phi\beta\beta_k} = 0. \tag{E.40}$$

Since $\partial_{\phi\phi'}Q(\gamma) = L_{\phi\phi'}(\gamma) + J_{\phi\phi'}(\gamma) + G_{\phi\phi'}(\gamma)$, $\partial_{\beta_k}L_{\phi\phi'}(\gamma) = 0$, $\partial_{\beta_k}G_{\phi\phi'}(\gamma) = 0$ and the (i,t)-th block of $\partial_{\beta_k}J_{\lambda f'}(\gamma)$ is $-d_{it}x_{itk}I_r$,[4] we have

$$\begin{aligned}
\|R_{\phi\phi\beta_k}\| &= \left\|(\hat{\phi}-\phi^0)'(\int_0^1\int_0^{s_1}\partial_{\beta_k}J_{\phi\phi'}(s_2)ds_2ds_1)(\hat{\phi}-\phi^0)\right\|\\
&= \left\|-\sum_{i=1}^N\sum_{t=1}^T d_{it}x_{itk}(\hat{\lambda}_i-\lambda_i^0)'(\hat{f}_t-f_t^0)\right\|\\
&\le \sqrt{\sum_{i=1}^N\sum_{t=1}^T x_{itk}^2}\left\|\hat{\lambda}-\lambda^0\right\|\left\|\hat{f}-f^0\right\| = O_p(\sqrt{NT}\left\|\hat{\lambda}-\lambda^0\right\|\left\|\hat{f}-f^0\right\|),
\end{aligned}\tag{E.41}$$

where the last equality follows from Assumption 3(ii).

(ii)-(iii) Now we consider $R_\phi = (R_\lambda', R_f')' = (\phi_1, ..., \phi_{(N+T)r})'$, where ϕ_j denotes the jth element of ϕ for $j \in [(N+T)\,r]$. Note that

$$R_{\phi_j} = R_{\beta\beta\phi_j} + R_{\beta\phi\phi_j} + R_{\phi\beta\phi_j} + R_{\phi\phi\phi_j}, \tag{E.42}$$

where

$$\begin{aligned}
R_{\beta\beta\phi_j} &= (\hat{\beta}-\beta^0)'(\int_0^1\int_0^{s_1}\partial_{\beta\beta'\phi_j}Q(s_2)ds_2ds_1)(\hat{\beta}-\beta^0),\\
R_{\phi\beta\phi_j} &= R_{\beta\phi\phi_j} = (\hat{\beta}-\beta^0)'(\int_0^1\int_0^{s_1}\partial_{\beta\phi'\phi_j}Q(s_2)ds_2ds_1)(\hat{\phi}-\phi^0),\\
R_{\phi\phi\phi_j} &= (\hat{\phi}-\phi^0)'(\int_0^1\int_0^{s_1}\partial_{\phi\phi'\phi_j}Q(s_2)ds_2ds_1)(\hat{\phi}-\phi^0).
\end{aligned}$$

[4] Recall from expression (A.8) that $G_{\phi\phi'}(\gamma)$ is a constant matrix for all γ.

First, noting that $Q(\cdot)$ is quadratic, we have

$$R_{\beta\beta\phi_j} = 0_{K\times K}. \tag{E.43}$$

Next, we consider $R_{\beta\phi\phi} = (R_{\beta\beta\phi_1}, ..., R_{\beta\beta\phi_{(N+T)r}})' = (R'_{\beta\phi\lambda}, R'_{\beta\phi f})'$, where

$$\begin{aligned}
R_{\beta\phi\lambda} &= (R_{\beta\beta\phi_1}, ..., R_{\beta\beta\phi_{Nr}})' \equiv (R'_{\beta\beta\lambda_1}, ..., R'_{\beta\beta\lambda_N})', \text{ and} \\
R_{\beta\phi f} &= (R_{\beta\beta\phi_{Nr+1}}, ..., R_{\beta\beta\phi_{(N+T)r}})' \equiv (R'_{\beta\phi f_1}, ..., R'_{\beta\phi f_T})'.
\end{aligned}$$

Note that

$$\partial_{\beta\phi'\lambda_{iq}} Q(\gamma) = \partial_{\lambda_{iq}} \partial_{\beta\phi'} Q(\gamma) = \partial_{\lambda_{iq}} Q_{\beta\phi'}(\gamma) = (0_{K\times r}, ..., 0_{K\times r}, d_{i1}x_{i1}1_q^{r\prime}, ...d_{iT}x_{iT}1_q^{r\prime}),$$

which is a $K \times (N+T)\, r$ matrix. Here, recall that $0_{K\times r}$ denotes a $K \times r$ matrix of zeros and 1_q^r denotes the $r \times 1$ vector with the q-th element being one and the other elements being zeros. Thus we have

$$R_{\beta\phi\lambda_i} = -\frac{1}{2}\sum\nolimits_{t=1}^{T} (\hat{\beta} - \beta^0)' d_{it} x_{it}\, (\hat{f}_t - f_t^0), \text{ and} \tag{E.44}$$

$$R_{\beta\phi f_t} = -\frac{1}{2}\sum\nolimits_{i=1}^{N} (\hat{\beta} - \beta^0)' d_{it} x_{it}\, (\hat{\lambda}_i - \lambda_i^0). \tag{E.45}$$

It follows that

$$\begin{aligned}
\|R_{\beta\phi\lambda}\| &\leq \frac{1}{2}\left\|\hat{\beta} - \beta^0\right\| \sqrt{\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} \|x_{it}\|^2} \left\|\hat{f} - f^0\right\| \\
&= O_p(\sqrt{NT}\left\|\hat{\beta} - \beta^0\right\|\left\|\hat{f} - f^0\right\|), \qquad \text{(E.46)} \\
\|R_{\beta\phi f}\| &\leq \frac{1}{2}\left\|\hat{\beta} - \beta^0\right\| \sqrt{\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} \|x_{it}\|^2} \left\|\hat{\lambda} - \lambda^0\right\| \\
&= O_p(\sqrt{NT}\left\|\hat{\beta} - \beta^0\right\|\left\|\hat{\lambda} - \lambda^0\right\|). \qquad \text{(E.47)}
\end{aligned}$$

Now, we consider $R_{\phi\phi\phi} = (R_{\phi\phi\phi_1}, ..., R_{\phi\phi\phi_{(N+T)r}})' = (R'_{\phi\phi\lambda}, R'_{\phi\phi f})'$, where

$$\begin{aligned}
R_{\phi\phi\lambda} &= (R_{\phi\phi\phi_1}, ..., R_{\phi\phi\phi_{Nr}})' \equiv (R'_{\beta\beta\lambda_1}, ..., R'_{\beta\beta\lambda_N})', \text{ and} \\
R_{\phi\phi f} &= (R_{\phi\phi\phi_{Nr+1}}, ..., R_{\phi\phi\phi_{(N+T)r}})' \equiv (R'_{\phi\phi f_1}, ..., R'_{\phi\phi f_T})'.
\end{aligned}$$

Note that (1) $\partial_{\lambda_{iq}} L_{\lambda\lambda'}(\gamma) = 0$, (2) the t-th block of $\partial_{\lambda_{iq}} L_{ff'}(\gamma)$ is $-d_{it}(\lambda_i 1_q^{r\prime} + 1_q^r \lambda_i')$, (3) the (i,t)-th block of $\partial_{\lambda_{iq}} L_{\lambda f'}(\gamma)$ is $-d_{it} f_t 1_q^{r\prime}$ and the (j,s)-th block is zero if $j \neq i$, (4) the (i,t)-th block of $\partial_{\lambda_{iq}} J_{\lambda f'}(\gamma)$ is $-d_{it} f_{tq} I_r$ and the (j,s)-th block is zero if $j \neq i$, and (5) $\partial_{\lambda_{iq}} G_{\phi\phi'} = 0$. It follows that

$$\begin{aligned}
& (\hat{\phi} - \phi^0)' \int_0^1 \int_0^{s_1} \partial_{\lambda_{iq}} \partial_{\phi\phi'} Q(s_2) ds_2 ds_1 (\hat{\phi} - \phi^0) \\
=\; & -2\sum\nolimits_{t=1}^{T} d_{it}(\hat{\lambda}_i - \lambda_i^0)' \int_0^1 \int_0^{s_1} (f_t(s_2) 1_q^{r\prime} + f_{tq}(s_2) I_r) ds_2 ds_1 (\hat{f}_t - f_t^0) \\
& -2\sum\nolimits_{t=1}^{T} d_{it}(\hat{f}_t - f_t^0)' \int_0^1 \int_0^{s_1} \lambda_i(s_2) 1_q^{r\prime} ds_2 ds_1 (\hat{f}_t - f_t^0), \qquad \text{(E.48)}
\end{aligned}$$

where $f_t(s) = f_t^0 + s(\hat{f}_t - f_t^0)$ and $\lambda_i(s) = \lambda_i^0 + s(\hat{\lambda}_i - \lambda_i^0)$. Then

$$\begin{aligned}\|R_{\phi\phi\lambda}\| &\leq M\left\|\hat{\lambda}-\lambda^0\right\|\sup_{0\leq s\leq 1}\|f(s)\|\left\|\hat{f}-f^0\right\| + M\sup_{0\leq s\leq 1}\|\lambda(s)\|\left\|\hat{f}-f^0\right\|^2 \\ &= O_p(\sqrt{T}\left\|\hat{\lambda}-\lambda^0\right\|\left\|\hat{f}-f^0\right\| + \sqrt{N}\left\|\hat{f}-f^0\right\|^2),\end{aligned} \tag{E.49}$$

where the equality follows from $\sup_{0\leq s\leq 1}\|f(s)\| \leq \|f^0\| + m\sqrt{T} = O_p(\sqrt{T})$ and $\sup_{0\leq s\leq 1}\|\lambda(s)\| \leq \|\lambda^0\| + m\sqrt{N} = O_p(\sqrt{N})$ by Assumption 2(i)-(ii) and the fact that $(\hat{\beta}', \hat{\lambda}', \hat{f}')' \in \mathcal{B}_m(\gamma^0)$ for a fixed m. By symmetry, we also have

$$\|R_{\phi\phi f}\| = O_p(\sqrt{N}\left\|\hat{\lambda}-\lambda^0\right\|\left\|\hat{f}-f^0\right\| + \sqrt{T}\left\|\hat{\lambda}-\lambda^0\right\|^2). \tag{E.50}$$

Combining equations (E.43), (E.46)-(E.47) and (E.49)-(E.50), we conclude the proof of the lemma. ■

Proof of Proposition 4.1

From equations (4.1)-(4.2), we have

$$\begin{aligned}\hat{\beta}-\beta^0 &= -(Q_{\beta\beta'} - Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'})^{-1}(S_\beta - Q_{\beta\phi'}Q_{\phi\phi'}^{-1}S_\phi) \\ &\quad -(Q_{\beta\beta'} - Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'})^{-1}(R_\beta - Q_{\beta\phi'}Q_{\phi\phi'}^{-1}R_\phi), \text{ and} \end{aligned} \tag{E.51}$$

$$\hat{\phi}-\phi^0 = -Q_{\phi\phi'}^{-1}S_\phi - Q_{\phi\phi'}^{-1}R_\phi - Q_{\phi\phi'}^{-1}Q_{\phi\beta'}(\hat{\beta}-\beta^0). \tag{E.52}$$

By Assumption 3(i) and Lemma B.2, $(Q_{\beta\beta'} - Q_{\beta\phi'}Q_{\phi\phi'}^{-1}Q_{\phi\beta'})^{-1} = O_p(\frac{1}{NT})$. By Lemma E.3, $S_\beta - Q_{\beta\phi'}Q_{\phi\phi'}^{-1}S_\phi = O_p(\sqrt{NT}+N+T)$. By Lemma E.4, Lemma B.1 and eqn. (B.17), we have

$$\begin{aligned}\left\|Q_{\beta\phi'}Q_{\phi\phi'}^{-1}R_\phi\right\| &= \left\|Q_{\beta\phi'}D_{TN}^{-\frac{1}{2}}(D_{TN}^{-\frac{1}{2}}Q_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}D_{TN}^{-\frac{1}{2}}R_\phi\right\| \\ &\leq \left\|Q_{\beta\phi'}D_{TN}^{-\frac{1}{2}}\right\|\left\|(D_{TN}^{-\frac{1}{2}}Q_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}\right\|\left\|D_{TN}^{-\frac{1}{2}}R_\phi\right\| \\ &= \sqrt{NT}O_p\Big\{\sqrt{N}\left\|\hat{\beta}-\beta^0\right\|\left\|\hat{f}-f^0\right\| + \left\|\hat{\lambda}-\lambda^0\right\|\left\|\hat{f}-f^0\right\| \\ &\quad +\sqrt{\frac{N}{T}}\left\|\hat{f}-f^0\right\|^2 + \sqrt{T}\left\|\hat{\beta}-\beta^0\right\|\left\|\hat{\lambda}-\lambda^0\right\| + \sqrt{\frac{T}{N}}\left\|\hat{\lambda}-\lambda^0\right\|^2\Big\}.\end{aligned}$$

Plug these rates into eqn. (E.51), we have

$$\begin{aligned}&\left\|\hat{\beta}-\beta^0\right\| \\ = \ & O_p\left(\frac{1}{\sqrt{NT}}+\frac{1}{N}+\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{NT}}\left\|\hat{\lambda}-\lambda^0\right\|\left\|\hat{f}-f^0\right\|\right) \\ &+O_p\left\{\frac{1}{\sqrt{T}}\left\|\hat{f}-f^0\right\|\left\|\hat{\beta}-\beta^0\right\| + \frac{1}{\sqrt{N}}\left\|\hat{\lambda}-\lambda^0\right\|\left\|\hat{\beta}-\beta^0\right\| + \frac{1}{T}\left\|\hat{f}-f^0\right\|^2 + \frac{1}{N}\left\|\hat{\lambda}-\lambda^0\right\|^2\right\} \\ &= O_p(\frac{1}{c_{NT}^2}),\end{aligned} \tag{E.53}$$

where the second equality follows from Theorem 3.1. By eqn. (E.52) and Lemma B.1, we have

$$\left\|D_{NT}^{-\frac{1}{2}}(\hat{\phi}-\phi^0+Q_{\phi\phi'}^{-1}S_\phi)\right\| \leq (NT)^{-1/2}\left\|(-D_{TN}^{-\frac{1}{2}}Q_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}\left[D_{TN}^{-\frac{1}{2}}R_\phi + D_{TN}^{-\frac{1}{2}}Q_{\phi\beta'}(\hat{\beta}-\beta^0)\right]\right\|$$

$$\begin{aligned}
&\lesssim (NT)^{-1/2}\left\{\left\|D_{TN}^{-\frac{1}{2}}R_\phi\right\|+\left\|D_{TN}^{-\frac{1}{2}}Q_{\phi\beta'}\right\|\left\|(\hat{\beta}-\beta^0)\right\|\right\} \\
&= O_p(\frac{1}{c_{NT}^2}), \qquad \text{(E.54)}
\end{aligned}$$

where the last equality holds by the fact that $\left\|D_{TN}^{-\frac{1}{2}}R_\phi\right\|=O_p(\frac{\sqrt{NT}}{c_{NT}^2})$, $\left\|D_{TN}^{-\frac{1}{2}}Q_{\phi\beta'}\right\|=O_p(\sqrt{NT})$ by eqn. (B.17) and $\left\|\hat{\beta}-\beta^0\right\|=O_p(\frac{1}{c_{NT}^2})$ by eqn. (E.53). To see the first fact, note that by Theorem 3.1, $\left\|\hat{\beta}-\beta^0\right\|=O_p(\frac{1}{c_{NT}})$, $\frac{1}{\sqrt{N}}\left\|\hat{\lambda}-\lambda^0\right\|=O_p(\frac{1}{c_{NT}})$ and $\frac{1}{\sqrt{T}}\left\|\hat{f}-f^0\right\|=O_p(\frac{1}{c_{NT}})$; plugging these rates into Lemma E.4 yields $\left\|D_{TN}^{-\frac{1}{2}}R_\phi\right\|=O_p(\frac{\sqrt{NT}}{c_{NT}^2})$. ■

To prove Theorem 4.1, we state and prove the next lemma where refine our calculations of R_β and $Q_{\beta\phi'}Q_{\phi\phi'}^{-1}R_\phi$.

Lemma E.5 *Suppose that Assumptions 1-5, 7 and 8(ii)-(iv) hold.*

(i) $R_\beta=o_p(\sqrt{NT})$ if $\frac{\sqrt{T}}{N}\to 0$ and $\frac{\sqrt{N}}{T}\to 0$;

(ii) $-Q_{\beta\phi'}Q_{\phi\phi'}^{-1}R_\phi=\sqrt{NT}(b_5+b_6)+o_p(\sqrt{NT})$ if $\frac{T^{[\frac{1}{2}+((\frac{3}{\varsigma\wedge\zeta}+\frac{3}{\varsigma}+\frac{1}{\varrho})\vee\frac{1}{\kappa})]\vee\frac{2}{3}}}{N}\to 0$ and $\frac{N^{[\frac{1}{2}+((\frac{3}{\varsigma\wedge\zeta}+\frac{3}{\varsigma}+\frac{1}{\varrho})\vee\frac{1}{\kappa})]\vee\frac{2}{3}}}{T}\to 0$, where b_5 and b_6 are $K\times 1$ vectors with the respective k-th elements given by

$$\begin{aligned}
b_{5k} &= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}_\phi(d_{it}v_{it}d_{jt}v_{jt})\lambda_i^{0\prime}[\bar{L}_{ff'}^{-1}]_t(\sum_{i=1}^{N}\Phi_{it}\lambda_i^0\delta_{ki}^{0\prime})[\bar{L}_{ff'}^{-1}]_t\lambda_j^0, \\
b_{6k} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}_\phi(d_{it}v_{it}d_{is}v_{is})f_t^{0\prime}[\bar{L}_{\lambda\lambda'}^{-1}]_i(\sum_{t=1}^{T}\Phi_{it}f_t^0\omega_{kt}^{0\prime})[\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0.
\end{aligned}$$

Proof. (i) From equations (E.39)–(E.41), we have

$$\begin{aligned}
\|R_\beta\| &= \|R_{\phi\phi\beta}\|=\left\|-\sum_{i=1}^{N}\sum_{t=1}^{T}d_{it}x_{it}(\hat{\lambda}_i-\lambda_i^0)'(\hat{f}_t-f_t^0)\right\| \\
&\le \sqrt{\sum_{t=1}^{T}\left\|\hat{f}_t-f_t^0\right\|^2}\sqrt{\sum_{t=1}^{T}\left\|\sum_{i=1}^{N}d_{it}x_{it}(\hat{\lambda}_i-\lambda_i^0)'\right\|^2} \\
&= O_p(\frac{TN}{c_{NT}^3})=o_p(\sqrt{NT}) \text{ if } \frac{\sqrt{T}}{N}\to 0 \text{ and } \frac{\sqrt{N}}{T}\to 0,
\end{aligned}$$

where the third equality follows from Theorem 3.1 and Lemma F.1(iii).

(ii) By (E.42), we can write

$$R_\phi=R_{\beta\beta\phi}+R_{\beta\phi\phi}+R_{\phi\beta\phi}+R_{\phi\phi\phi}, \qquad \text{(E.55)}$$

where $R_{\beta\beta\phi}$, $R_{\beta\phi\phi}$, $R_{\phi\beta\phi}$, and $R_{\phi\phi\phi}$ are all $(N+T)r\times 1$ vector with typical elements given by $R_{\beta\beta\phi_j}$, $R_{\beta\phi\phi_j}$, $R_{\phi\beta\phi_j}$, and $R_{\phi\phi\phi_j}$, respectively.

First note that $R_{\beta\beta\phi}=0_{(N+T)r\times 1}$ by (E.43). Next, using $R_{\beta\phi\phi}=(R_{\beta\phi\lambda}',R_{\beta\phi f}')'$ and (E.46)–(E.47), we have

$$\|R_{\beta\phi\lambda}\|=O_p(\sqrt{NT}\left\|\hat{\beta}-\beta^0\right\|\left\|\hat{f}-f^0\right\|) \text{ and } \|R_{\beta\phi f}\|=O_p(\sqrt{NT}\left\|\hat{\beta}-\beta^0\right\|\left\|\hat{\lambda}-\lambda^0\right\|).$$

It follows that

$$\left\|D_{TN}^{-\frac{1}{2}}R_{\beta\phi\phi}\right\|=O_p\left\{\sqrt{NT}\left\|\hat{\beta}-\beta^0\right\|\left(\frac{1}{\sqrt{T}}\left\|\hat{f}-f^0\right\|+\frac{1}{\sqrt{N}}\left\|\hat{\lambda}-\lambda^0\right\|\right)\right\}.$$

Then by eqn. (B.17), Lemma B.1 and Proposition 4.1,

$$\begin{aligned} Q_{\beta\phi'}Q_{\phi\phi'}^{-1}R_{\beta\phi\phi} &= Q_{\beta\phi'}D_{TN}^{-\frac{1}{2}}(-D_{TN}^{-\frac{1}{2}}Q_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}D_{TN}^{-\frac{1}{2}}R_{\beta\phi\phi} = O_p(\frac{NT}{c_{NT}^3}) \\ &= o_p(\sqrt{NT}) \text{ if } \frac{\sqrt{T}}{N} \to 0 \text{ and } \frac{\sqrt{N}}{T} \to 0. \end{aligned} \tag{E.56}$$

Now, we consider $Q_{\beta\phi'}Q_{\phi\phi'}^{-1}R_{\phi\phi\phi}$. Define $Q_{\beta\phi'}Q_{\phi\phi'}^{-1} = (\delta^{*\prime}, \omega^{*\prime}) = (\delta_1^*, ..., \delta_N^*, \omega_1^*, ..., \omega_T^*)$, where δ_i^*'s and ω_t^*'s are all $K \times r$ matrices. Let δ_{ki}^* and ω_{kt}^* denote the transpose of the k-th row of δ_i^* and ω_t^*, respectively. From eqn. (E.48), the k-th element of $Q_{\beta\phi'}Q_{\phi\phi'}^{-1}R_{\phi\phi\phi}$ is:

$$\begin{aligned} Q_{\beta_k\phi'}Q_{\phi\phi'}^{-1}R_{\phi\phi\phi} &= -2\sum_{t=1}^{T}(\hat{f}_t - f_t^0)'\int_0^1\int_0^{s_1}\sum_{i=1}^{N} d_{it}\lambda_i(s_2)\delta_{ki}^{*\prime}ds_2ds_1(\hat{f}_t - f_t^0) \\ &\quad -2\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\lambda}_i - \lambda_i^0)'\int_0^1\int_0^{s_1} d_{it}(f_t(s_2)\delta_{ki}^{*\prime} + \delta_{ki}^{*\prime}f_t(s_2)I_r)ds_2ds_1(\hat{f}_t - f_t^0) \\ &\quad -2\sum_{t=1}^{T}\sum_{i=1}^{N}(\hat{f}_t - f_t^0)'\int_0^1\int_0^{s_1} d_{it}(\lambda_i(s_2)\omega_{kt}^{*\prime} + \omega_{kt}^{*\prime}\lambda_i(s_2)I_r)ds_2ds_1(\hat{\lambda}_i - \lambda_i^0) \\ &\quad -2\sum_{i=1}^{N}(\hat{\lambda}_i - \lambda_i^0)'\int_0^1\int_0^{s_1}\sum_{t=1}^{T} d_{it}f_t(s_2)\omega_{kt}^{*\prime}ds_2ds_1(\hat{\lambda}_i - \lambda_i^0). \end{aligned}$$

Thus by Lemma F.2 and Theorem 3.1, we have

$$\begin{aligned} Q_{\beta_k\phi'}Q_{\phi\phi'}^{-1}R_{\phi\phi\phi} &= -\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\lambda}_i - \lambda_i^0)'d_{it}(f_t^0\delta_{ki}^{*\prime} + \delta_{ki}^{*\prime}f_t^0I_r)(\hat{f}_t - f_t^0) \\ &\quad -\sum_{t=1}^{T}\sum_{i=1}^{N}(\hat{f}_t - f_t^0)'d_{it}(\lambda_i^0\omega_{kt}^{*\prime} + \omega_{kt}^{*\prime}\lambda_i^0I_r)(\hat{\lambda}_i - \lambda_i^0) \\ &\quad -\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{f}_t - f_t^0)'(d_{it}\lambda_i^0\delta_{ki}^{*\prime})(\hat{f}_t - f_t^0) \\ &\quad -\sum_{t=1}^{T}\sum_{i=1}^{N}(\hat{\lambda}_i - \lambda_i^0)'(d_{it}f_t^0\omega_{kt}^{*\prime})(\hat{\lambda}_i - \lambda_i^0) + O_p(\frac{NT}{c_{NT}^3}) \\ &\equiv -II_{2,1} - II_{2,2} - II_{2,3} - II_{2,4} + O_p(\frac{NT}{c_{NT}^3}). \end{aligned} \tag{E.57}$$

Noting that $\sum_{t=1}^{T} d_{it}f_t^0\delta_{ki}^{*\prime}(\hat{f}_t - f_t^0) = \sum_{t=1}^{T} d_{it}f_t^0(\hat{f}_t - f_t^0)'\delta_{ki}^*$ and $\sum_{t=1}^{T} d_{it}\delta_{ki}^{*\prime}f_t^0(\hat{f}_t - f_t^0) = \sum_{t=1}^{T} d_{it}(\hat{f}_t - f_t^0)f_t^{0\prime}\delta_{ki}^*$,

$$\begin{aligned} \|II_{2,1}\| &\leq 2\sum_{i=1}^{N}\left\|\hat{\lambda}_i - \lambda_i^0\right\|\|\delta_{ki}^*\|\left\|\sum_{t=1}^{T} d_{it}f_t^0(\hat{f}_t - f_t^0)'\right\| \\ &= O_p(\frac{\sqrt{N}}{c_{NT}})O_p((N \vee T)^{\frac{3}{\varsigma \wedge \zeta} + \frac{3}{\varsigma} + \frac{1}{\varrho}})O_p(\frac{T\sqrt{N}}{c_{NT}^2}) = O_p((N \vee T)^{\frac{3}{\varsigma \wedge \zeta} + \frac{3}{\varsigma} + \frac{1}{\varrho}})O_p(\frac{NT}{c_{NT}^3}) \\ &= o_p(\sqrt{NT}) \text{ if } \frac{T^{\frac{1}{2} + \frac{3}{\varsigma \wedge \zeta} + \frac{3}{\varsigma} + \frac{1}{\varrho}}}{N} \to 0 \text{ and } \frac{N^{\frac{1}{2} + \frac{3}{\varsigma \wedge \zeta} + \frac{3}{\varsigma} + \frac{1}{\varrho}}}{T} \to 0, \end{aligned} \tag{E.58}$$

where the first equality follows from Theorem 3.1, Lemma F.1(i) and Lemma F.3. Similarly,

$$\|II_{2,2}\| \leq 2\sum_{t=1}^{T}\left\|\hat{f}_t - f_t^0\right\|\|\omega_{kt}^*\|\left\|\sum_{i=1}^{N} d_{it}\lambda_i^0(\hat{\lambda}_i - \lambda_i^0)'\right\|$$

$$
\begin{aligned}
&= O_p(\frac{\sqrt{T}}{c_{NT}})O_p((N\vee T)^{\frac{3}{\varsigma\wedge\varsigma}+\frac{3}{\varsigma}+\frac{1}{\varrho}})O_p(\frac{N\sqrt{T}}{c_{NT}^2}) \leq O_p((N\vee T)^{\frac{3}{\varsigma\wedge\varsigma}+\frac{3}{\varsigma}+\frac{1}{\varrho}})O_p(\frac{TN}{c_{NT}^3}) \\
&= o_p(\sqrt{NT}) \text{ if } \frac{T^{\frac{1}{2}+\frac{3}{\varsigma\wedge\varsigma}+\frac{3}{\varsigma}+\frac{1}{\varrho}}}{N} \to 0 \text{ and } \frac{N^{\frac{1}{2}+\frac{3}{\varsigma\wedge\varsigma}+\frac{3}{\varsigma}+\frac{1}{\varrho}}}{T} \to 0, \qquad \text{(E.59)}
\end{aligned}
$$

where the first equality follows from Theorem 3.1, Lemma F.1(ii) and Lemma F.3. Next, by Lemma F.4(i)-(iv) and Theorem 3.1, we have

$$
\begin{aligned}
II_{2,3} &= O_p(\frac{T}{c_{NT}^2})O_p(\sqrt{N}T^{\frac{1}{\kappa}} + \frac{N^{1+\frac{1}{\kappa}}}{\sqrt{T}} + \frac{N}{\sqrt{c_{NT}}}) - \sum_{t=1}^T (\hat{f}_t - f_t^0)'(\sum_{i=1}^N \Phi_{it}\lambda_i^0\delta_{ki}^{0\prime})(\hat{f}_t - f_t^0), \qquad \text{(E.60)} \\
II_{2,4} &= O_p(\frac{N}{c_{NT}^2})O_p(\sqrt{T}N^{\frac{1}{\kappa}} + \frac{T^{1+\frac{1}{\kappa}}}{\sqrt{N}} + \frac{T}{\sqrt{c_{NT}}}) - \sum_{i=1}^N (\hat{\lambda}_i - \lambda_i^0)'(\sum_{t=1}^T \Phi_{it}f_t^0\omega_{kt}^{0\prime})(\hat{\lambda}_i - \lambda_i^0), \qquad \text{(E.61)}
\end{aligned}
$$

where $O_p(\frac{NT}{c_{NT}^2})O_p(\frac{T^{\frac{1}{\kappa}}}{\sqrt{N}} + \frac{N^{\frac{1}{\kappa}}}{\sqrt{T}} + \frac{1}{\sqrt{c_{NT}}}) = o_p(\sqrt{NT})$ if $\frac{T^{(\frac{1}{2}+\frac{1}{\kappa})\vee\frac{2}{3}}}{N} \to 0$ and $\frac{N^{(\frac{1}{2}+\frac{1}{\kappa})\vee\frac{2}{3}}}{T} \to 0$. By eqn. (E.54) and Lemma E.2, we have $\sqrt{\sum_{t=1}^T \left\|\hat{f}_t - f_t^0 + [\bar{L}_{ff'}^{-1}]_t S_{f_t}\right\|^2} = O_p(\frac{\sqrt{T}}{c_{NT}^2} + \frac{T^{\frac{1}{2}+\frac{1}{\kappa}}}{N})$. By Theorem 3.1, $\sqrt{\sum_{t=1}^T \left\|\hat{f}_t - f_t^0\right\|^2} = O_p(\frac{\sqrt{T}}{c_{NT}})$. These, together with Lemma F.4(v)-(vi), imply that expressions in (E.60) and (E.61) can be further simplified to obtain

$$
\begin{aligned}
II_{2,3} &= -\sum_{t=1}^T S_{f_t}'[\bar{L}_{ff'}^{-1}]_t(\sum_{i=1}^N \Phi_{it}\lambda_i^0\delta_{ki}^{0\prime})[\bar{L}_{ff'}^{-1}]_t S_{f_t} + O_p(\frac{NT}{c_{NT}^3} + \frac{T^{1+\frac{1}{\kappa}}}{c_{NT}}), \qquad \text{(E.62)} \\
II_{2,4} &= -\sum_{i=1}^N S_{\lambda_i}'[\bar{L}_{\lambda\lambda'}^{-1}]_i(\sum_{t=1}^T \Phi_{it}f_t^0\omega_{kt}^{0\prime})[\bar{L}_{\lambda\lambda'}^{-1}]_i S_{\lambda_i} + O_p(\frac{NT}{c_{NT}^3} + \frac{N^{1+\frac{1}{\kappa}}}{c_{NT}}), \qquad \text{(E.63)}
\end{aligned}
$$

where $O_p(\frac{NT}{c_{NT}^3} + \frac{T^{1+\frac{1}{\kappa}}}{c_{NT}}) = o_p(\sqrt{NT})$ and $O_p(\frac{NT}{c_{NT}^3} + \frac{N^{1+\frac{1}{\kappa}}}{c_{NT}}) = o_p(\sqrt{NT})$ if $\frac{T^{\frac{1}{2}+\frac{1}{\kappa}}}{N} \to 0$, $\frac{T^{\frac{1}{\kappa}}}{\sqrt{N}} \to 0$, $\frac{N^{\frac{1}{2}+\frac{1}{\kappa}}}{T} \to 0$ and $\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}} \to 0$. Plugging in $S_{f_t} = \sum_{i=1}^N \lambda_i^0 d_{it}v_{it}$ and $S_{\lambda_i} = \sum_{t=1}^T f_t^0 d_{it}v_{it}$, we have

$$
\begin{aligned}
II_{2,3} &= -\sum_{t=1}^T S_{f_t}'[\bar{L}_{ff'}^{-1}]_t(\sum_{i=1}^N \Phi_{it}\lambda_i^0\delta_{ki}^{0\prime})[\bar{L}_{ff'}^{-1}]_t S_{f_t} + o_p(\sqrt{NT}) \\
&= -\sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N d_{it}v_{it}d_{jt}v_{jt}\lambda_j^{0\prime}[\bar{L}_{ff'}^{-1}]_t(\sum_{i=1}^N \Phi_{it}\lambda_i^0\delta_{ki}^{0\prime})[\bar{L}_{ff'}^{-1}]_t\lambda_i^0 + o_p(\sqrt{NT}) \\
&= -\sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N \mathbb{E}_\phi(d_{it}v_{it}d_{jt}v_{jt})\lambda_j^{0\prime}[\bar{L}_{ff'}^{-1}]_t(\sum_{i=1}^N \Phi_{it}\lambda_i^0\delta_{ki}^{0\prime})[\bar{L}_{ff'}^{-1}]_t\lambda_i^0 + O_p(\sqrt{T}) + o_p(\sqrt{NT}),
\end{aligned}
$$

and

$$
\begin{aligned}
II_{2,4} &= -\sum_{i=1}^N S_{\lambda_i}'[\bar{L}_{\lambda\lambda'}^{-1}]_i(\sum_{t=1}^T \Phi_{it}f_t^0\omega_{kt}^{0\prime})[\bar{L}_{\lambda\lambda'}^{-1}]_i S_{\lambda_i} + o_p(\sqrt{NT}) \\
&= -\sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T d_{it}v_{it}d_{is}v_{is}f_t^{0\prime}[\bar{L}_{\lambda\lambda'}^{-1}]_i(\sum_{t=1}^T \Phi_{it}f_t^0\omega_{kt}^{0\prime})[\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0 + o_p(\sqrt{NT}) \\
&= -\sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \mathbb{E}_\phi(d_{it}v_{it}d_{is}v_{is})f_t^{0\prime}[\bar{L}_{\lambda\lambda'}^{-1}]_i(\sum_{t=1}^T \Phi_{it}f_t^0\omega_{kt}^{0\prime})[\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0 + O_p(\sqrt{N}) + o_p(\sqrt{NT}).
\end{aligned}
$$

where we use Assumption 8(iv) and the fact that $\max_{i,j,t}|\lambda_j^{0\prime}[N\bar{L}_{ff'}^{-1}]_t(\frac{1}{N}\sum_{i=1}^N \Phi_{it}\lambda_i^0\delta_{ki}^{0\prime})[N\bar{L}_{ff'}^{-1}]_t\lambda_i^0| \leq M$ and $\max_{i,t,s}|f_t^{0\prime}[T\bar{L}_{\lambda\lambda'}^{-1}]_i(\frac{1}{T}\sum_{t=1}^T \Phi_{it}f_t^0\omega_{kt}^{0\prime})[T\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0| \leq M$ by Lemma F.4(v)-(vi).

Combining the above results on $II_{2,j}$'s yields the desired result. ■

Proof of Theorem 4.1

Given eqn. (E.51), it's easy to see that Theorem 4.1 follows from Lemma B.2, Lemma E.3 and Lemma E.5. ■

F Supplementary Lemmas Used in the Proof of Lemma E.5

In this section, we state and prove four technical lemmas used in the proof of Lemma E.5.

Lemma F.1 *Under Assumptions 1-5 and 7, as* $(N,T)\to\infty$,

(i) $\sum_{i=1}^N \left\|\sum_{t=1}^T d_{it} f_t^0(\hat{f}_t - f_t^0)'\right\|^2 = O_p(\frac{T^2N}{c_{NT}^4})$;

(ii) $\sum_{t=1}^T \left\|\sum_{i=1}^N d_{it}\lambda_i^0(\hat{\lambda}_i - \lambda_i^0)'\right\|^2 = O_p(\frac{TN^2}{c_{NT}^4})$;

(iii) $\sum_{t=1}^T \left\|\sum_{i=1}^N d_{it}x_{it}(\hat{\lambda}_i - \lambda_i^0)'\right\|^2 = O_p(\frac{TN^2}{c_{NT}^4})$.

Proof. (i)-(ii) The proof of (i) is similar to that of (ii). We focus on the proof of (ii) here. By equations (E.14)–(E.16), (E.18)–(E.20), and (E.54), we have

$$\frac{1}{\sqrt{N}}\left\|(\hat{\lambda}-\lambda^0)+L_{\lambda\lambda'}^{-1}S_\lambda\right\| = O_p(\frac{1}{c_{NT}^2}), \text{ and} \tag{F.1}$$

$$\frac{1}{\sqrt{T}}\left\|(\hat{f}-f^0)+L_{ff'}^{-1}S_f\right\| = O_p(\frac{1}{c_{NT}^2}). \tag{F.2}$$

Noting that the i-th block of $L_{\lambda\lambda'}^{-1}S_\lambda$ is $[L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i}$, we have by the CS inequality

$$\begin{aligned}
&\sum_{s=1}^T \left\|\sum_{i=1}^N d_{is}\lambda_i^0(\hat{\lambda}_i-\lambda_i^0)'\right\|^2 \\
\le\ & 2\sum_{s=1}^T \left\|\sum_{i=1}^N d_{is}\lambda_i^0(\hat{\lambda}_i-\lambda_i^0+[L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i})'\right\|^2 + 2\sum_{s=1}^T\left\|\sum_{i=1}^N d_{is}\lambda_i^0 S_{\lambda_i}'[L_{\lambda\lambda'}^{-1}]_i\right\|^2 \\
\le\ & 2T\sum_{i=1}^N \left\|\lambda_i^0\right\|^2 \sum_{i=1}^N \left\|\hat{\lambda}_i-\lambda_i^0+[L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i}\right\|^2 \\
& +2\sum_{s=1}^T\left\|\sum_{i=1}^N\sum_{t=1}^T d_{is}\lambda_i^0 f_t^{0\prime} d_{it}v_{it}(\sum_{t=1}^T d_{it}f_t^0 f_t^{0\prime})^{-1}\right\|^2 \\
=\ & O_p(\frac{N^2T}{c_{NT}^4}) + O_p(\frac{NT+N^2}{T}) = O_p(\frac{N^2T}{c_{NT}^4}),
\end{aligned}$$

where the equality follows from eqn. (F.1) and expression (E.7).

(iii) The proof is similar to that of part (ii). Note that

$$\begin{aligned}
&\sum_{s=1}^T \left\|\sum_{i=1}^N d_{is}x_{is}(\hat{\lambda}_i-\lambda_i^0)'\right\|^2 \\
\le\ & 2\sum_{s=1}^T \left\|\sum_{i=1}^N d_{is}x_{is}(\hat{\lambda}_i-\lambda_i^0+[L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i})'\right\|^2 + 2\sum_{s=1}^T\left\|\sum_{i=1}^N d_{is}x_{is} S_{\lambda_i}'[L_{\lambda\lambda'}^{-1}]_i\right\|^2 \\
\le\ & 2T\sum_{i=1}^N \left\|x_{is}\right\|^2 \sum_{i=1}^N \left\|\hat{\lambda}_i-\lambda_i^0+[L_{\lambda\lambda'}^{-1}]_i S_{\lambda_i}\right\|^2 \\
& +2\sum_{s=1}^T\left\|\sum_{i=1}^N\sum_{t=1}^T d_{is}x_{is} f_t^{0\prime} d_{it}v_{it}(\sum_{t=1}^T d_{it}f_t^0 f_t^{0\prime})^{-1}\right\|^2
\end{aligned}$$

$$= O_p(\frac{N^2T}{c_{NT}^4}) + 2\sum_{s=1}^{T}\left\|\sum_{i=1}^{N}\sum_{t=1}^{T} d_{is}x_{is}f_t^{0\prime}d_{it}v_{it}(\sum_{t=1}^{T} d_{it}f_t^0 f_t^{0\prime})^{-1}\right\|^2$$
$$= O_p(\frac{N^2T}{c_{NT}^4}) + O_p(\frac{NT^2+N^2T}{T^2}) = O_p(\frac{N^2T}{c_{NT}^4}),$$

where the first equality follows from eqn. (F.2) and Assumption 3(ii), and the second equality is proved below. Recalling that $A_{iF} = \frac{1}{T}\sum_{t=1}^{T} d_{it}f_t^0 f_t^{0\prime}$ and $\bar{A}_{iF} = \mathbb{E}_\phi(A_{iF})$, we have

$$\mathbb{E}_\phi(\sum_{h=1}^{T}\left\|\sum_{i=1}^{N}\sum_{t=1}^{T} \bar{A}_{iF}^{-1}f_t^0 x_{ih}' d_{it}v_{it}d_{ih}\right\|_F^2)$$
$$\leq \sum_{h=1}^{T}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\left\|\bar{A}_{iF}^{-1}\right\|\left\|f_t^0\right\|\left\|\bar{A}_{jF}^{-1}\right\|\left\|f_s^0\right\|\left|\mathbb{E}_\phi(d_{it}v_{it}d_{ih}d_{jh}x_{ih}'x_{jh}d_{js}v_{js})\right|$$
$$\leq T(NT+N^2)M,$$

where the last inequality follows from Assumptions 2(i) and 7 (iii) and $\left\|\bar{A}_{iF}^{-1}\right\|$ is bounded for all i by Assumption 1. Also,

$$\sum_{h=1}^{T}\left\|\sum_{i=1}^{N}\sum_{t=1}^{T}(A_{iF}^{-1} - \bar{A}_{iF}^{-1})f_t^0 x_{ih}' d_{it}v_{it}d_{ih}\right\|_F^2$$
$$\leq (\sum_{i=1}^{N}\left\|A_{iF}^{-1} - \bar{A}_{iF}^{-1}\right\|^2)(\sum_{h=1}^{T}\sum_{i=1}^{N}\left\|\sum_{t=1}^{T} f_t^0 d_{it}v_{it}x_{ih}'d_{ih}\right\|_F^2)$$
$$= O_p(\frac{N}{T})O_p(NT^2) = O_p(N^2T),$$

where the first equality is due to eqn. (E.11) and

$$\sum_{h=1}^{T}\mathbb{E}_\phi(\sum_{i=1}^{N}\left\|\sum_{t=1}^{T} f_t^0 x_{ih}' d_{it}v_{it}d_{ih}\right\|_F^2)$$
$$= \sum_{h=1}^{T}\mathbb{E}_\phi(\sum_{i=1}^{N}\sum_{t,s=1}^{T}\sum_{p=1}^{r}\sum_{k=1}^{K} f_{tp}^0 f_{sp}^0 (x_{ihk}d_{ih})^2 d_{it}v_{it}d_{is}v_{is})$$
$$\leq \sum_{h=1}^{T}\sum_{t,s=1}^{T}\left\|f_t^0\right\|\left\|f_s^0\right\|\sum_{i=1}^{N}\left|\mathbb{E}_\phi(d_{it}v_{it}d_{is}v_{is}x_{ih}'x_{ih}d_{ih})\right|) \leq NT^2M,$$

by Assumption 7(iii). ■

Lemma F.2 *Let $B_{\lambda f}(s)$ denote the $Nr\times Tr$ matrix with $d_{it}(f_t(s)\delta_{ki}^{*\prime}+\delta_{ki}^{*\prime}f_t(s)I_r)$ as the (i,t)-th block, $B_{f\lambda}(s)$ denote the $Tr\times Nr$ matrix with $d_{it}(\lambda_i(s)\omega_{kt}^{*\prime}+\omega_{kt}^{*\prime}\lambda_i(s)I_r)$ as the (t,i)-th block, $B_{\lambda\lambda}(s)$ denote the $Nr\times Nr$ block diagonal matrix with $\sum_{t=1}^{T} d_{it}f_t(s)\omega_{kt}^{*\prime}$ as the i-th block, and $B_{ff}(s)$ denote the $Tr\times Tr$ block diagonal matrix with $\sum_{i=1}^{N} d_{it}\lambda_i(s)\delta_{ki}^{*\prime}$ as the t-th block. Suppose that Assumptions 1-5 hold. Then as $(N,T)\to\infty$,*

(i) $\sup_{0\leq s\leq 1}\left\|B_{\lambda f}(s) - B_{\lambda f}(0)\right\|_F = O_p(\frac{\sqrt{NT}}{c_{NT}})$;

(ii) $\sup_{0\leq s\leq 1}\left\|B_{f\lambda}(s) - B_{f\lambda}(0)\right\|_F = O_p(\frac{\sqrt{NT}}{c_{NT}})$;

(iii) $\sup_{0\leq s\leq 1}\left\|B_{\lambda\lambda}(s) - B_{\lambda\lambda}(0)\right\| = O_p(\frac{T}{c_{NT}})$;

(iv) $\sup_{0\leq s\leq 1}\left\|B_{ff}(s) - B_{ff}(0)\right\| = O_p(\frac{N}{c_{NT}})$.

Proof. (i) Noting that $f(s) = f^0 + s(\hat{f} - f^0)$, we notice that the (i,t)-th block of $B_{\lambda f}(s) - B_{\lambda f}(0)$ is

$s(\hat{f}_t - f_t^0)\delta_{ki}^{*\prime} d_{it} + s\delta_{ki}^{*\prime}(\hat{f}_t - f_t^0) d_{it} I_r$, and

$$\sup_{0\le s\le 1} \|B_{\lambda f}(s) - B_{\lambda f}(0)\|_F^2 \le (r+1)\sum\nolimits_{i=1}^N \|\delta_{ki}^*\|^2 \sum\nolimits_{t=1}^T \left\|\hat{f}_t - f_t^0\right\|^2 = O_p(\frac{NT}{c_{NT}^2}),$$

where the equality follows from Theorem 3.1 and

$$T\sum\nolimits_{i=1}^N \|\delta_{ki}^*\|^2 \le \left\|Q_{\beta\phi'}Q_{\phi\phi'}^{-1}D_{TN}^{\frac{1}{2}}\right\|_F^2 \le \left\|Q_{\beta\phi'}D_{TN}^{-\frac{1}{2}}\right\|_F^2 \left\|(D_{TN}^{-\frac{1}{2}}Q_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}\right\|^2 = O_p(NT).$$

The last equality is due to eqn. (B.17) and Lemma B.1.

(ii) Noting that the (t,i)-th block of $B_{f\lambda}(s) - B_{f\lambda}(0)$ is $s(\hat{\lambda}_i - \lambda_i^0)\omega_{kt}^{*\prime} d_{it} + s\omega_{kt}^{*\prime}(\hat{\lambda}_i - \lambda_i^0) d_{it} I_r$, the proof is similar to that of part (i).

(iii) The i-th block of $B_{\lambda\lambda}(s) - B_{\lambda\lambda}(0)$ is $s\sum_{t=1}^T d_{it}(\hat{f}_t - f_t^0)\omega_{kt}^{*\prime}$. Then

$$\begin{aligned}\sup_{0\le s\le 1}\|B_{\lambda\lambda}(s) - B_{\lambda\lambda}(0)\| &= \max_i \left\|\sum\nolimits_{t=1}^T d_{it}(\hat{f}_t - f_t^0)\omega_{kt}^{*\prime}\right\| \\ &\le \sqrt{\sum\nolimits_{t=1}^T \left\|\hat{f}_t - f_t^0\right\|^2}\sqrt{\sum\nolimits_{t=1}^T \|\omega_{kt}^*\|^2} = O_p(\frac{T}{c_{NT}}),\end{aligned}$$

where the last equality follows from Theorem 3.1 and

$$N\sum\nolimits_{t=1}^T \|\omega_{kt}^*\|^2 \le \left\|Q_{\beta\phi'}Q_{\phi\phi'}^{-1}D_{TN}^{\frac{1}{2}}\right\|_F^2 = O_p(NT).$$

(iv) The proof is similar to that of part (iii). ■

Lemma F.3 *Suppose that Assumptions 1, 2, 3(ii) and 4 hold. Then as $(N,T)\to\infty$,*

(i) $\max_i \|\delta_{ki}^*\| = O_p((N\vee T)^{\frac{3}{\varsigma\wedge\zeta}+\frac{3}{\varsigma}+\frac{1}{\varrho}})$ and $\max_t \|\omega_{kt}^*\| = O_p((N\vee T)^{\frac{3}{\varsigma\wedge\zeta}+\frac{3}{\varsigma}+\frac{1}{\varrho}})$;

(ii) $\max_i \left\|\delta_{ki}^0\right\| = O_p((N\vee T)^{\frac{6}{\varsigma}+\frac{1}{\varrho}})$ and $\max_t \left\|\omega_{kt}^0\right\| = O_p((N\vee T)^{\frac{6}{\varsigma}+\frac{1}{\varrho}})$.

Note that $\varsigma=\infty$ when $\left\|\lambda_i^0\right\|$ and $\left\|f_t^0\right\|$ are uniformly bounded.

Proof. (i) Let $\|\cdot\|_1$ denote the 1-norm of a matrix that is given by the maximum absolute column sum: for an $m\times n$ matrix $A=\{a_{ij}\}$, $\|A\|_1 = \max_{j\in[n]}\sum_{i=1}^m |a_{ij}|$.

Step (1): We first bound $\left\|Q_{\lambda\lambda'}^{-1}\right\|_1$.

By eqn. (E.2), $Q_{\lambda\lambda'}^{-1} = L_{\lambda\lambda'}^{-1} - L_{\lambda\lambda'}^{-1}W_\lambda^0(-\frac{N}{cT}I_{r^2} + W_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}W_\lambda^0)^{-1}W_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}$. Then

$$\begin{aligned}\left\|Q_{\lambda\lambda'}^{-1}\right\|_1 &\le \left\|L_{\lambda\lambda'}^{-1}\right\|_1 + \left\|L_{\lambda\lambda'}^{-1}\right\|_1 \left\|W_\lambda^0\right\|_1 \left\|(-\frac{N}{cT}I_{r^2} + W_\lambda^{0\prime}L_{\lambda\lambda'}^{-1}W_\lambda^0)^{-1}\right\|_1 \left\|W_\lambda^{0\prime}\right\|_1 \left\|L_{\lambda\lambda'}^{-1}\right\|_1 \\ &= O_p(\frac{1}{T}) + O_p(\frac{1}{T})O_p(N)O(\frac{rcT}{N})O_p(N^{\frac{1}{\varsigma}})O_p(\frac{1}{T}) = O_p(\frac{N^{\frac{1}{\varsigma}}}{T}),\end{aligned}$$

where the first equality by the results in (1.1)–(1.3) below.

(1.1) by eqn. (E.5), $\left\|L_{\lambda\lambda'}^{-1}\right\|_1 \le \sqrt{r}\left\|L_{\lambda\lambda'}^{-1}\right\| = O_p(\frac{1}{T})$ when $\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}}\to 0$;

(1.2) by Assumption 2, $\left\|W_\lambda^0\right\|_1 = \max_{q\in[r]}\sum_{i=1}^N |\lambda_{iq}^0| \le \sqrt{N}\left\|\lambda^0\right\| = O_p(N)$ and $\left\|W_\lambda^{0\prime}\right\|_1 = \max_{i\in[N]}\sum_{q=1}^r |\lambda_{iq}^0| = O(N^{\frac{1}{\varsigma}})$;

(1.3) by negative definiteness of $W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} W_\lambda^0$, we have

$$\left\|(-\frac{N}{cT} I_{r^2} + W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} W_\lambda^0)^{-1}\right\|_1 \leq r \left\|(-\frac{N}{cT} I_{r^2} + W_\lambda^{0\prime} L_{\lambda\lambda'}^{-1} W_\lambda^0)^{-1}\right\| \leq \frac{rcT}{N}. \tag{F.3}$$

Step (2): Let $(Q_{\phi\phi'}^{-1})_{ul}$, $(Q_{\phi\phi'}^{-1})_{lr}$, $(Q_{\phi\phi'}^{-1})_{ur}$ and $(Q_{\phi\phi'}^{-1})_{ll}$ denote the upper-left, lower-right, upper-right and lower-left block of $Q_{\phi\phi'}^{-1}$ of sizes $Nr \times Nr$, $Tr \times Tr$, $Nr \times Tr$, $Tr \times Nr$, respectively. We shall prove that the 1-norm of these four terms are respectively $O_p(\frac{N^{\frac{2}{\varsigma\wedge\zeta}+\frac{2}{\varsigma}}}{T})$, $O_p(\frac{T^{\frac{2}{\varsigma\wedge\zeta}+\frac{2}{\varsigma}}}{N})$, $O_p(\frac{N^{\frac{2}{\varsigma\wedge\zeta}+\frac{2}{\varsigma}} T^{\frac{1}{\varsigma\wedge\zeta}+\frac{1}{\varsigma}}}{T})$, and $O_p(\frac{T^{\frac{2}{\varsigma\wedge\zeta}+\frac{2}{\varsigma}} N^{\frac{1}{\varsigma\wedge\zeta}+\frac{1}{\varsigma}}}{N})$ in (2.1)–(2.4) below.

(2.1) The upper-left block is $(Q_{\phi\phi'}^{-1})_{ul} = [Q_{\lambda\lambda'} - Q_{\lambda f'} Q_{ff'}^{-1} Q_{f\lambda'}]^{-1}$, thus

$$\begin{aligned} & \left\|[Q_{\lambda\lambda'} - Q_{\lambda f'} Q_{ff'}^{-1} Q_{f\lambda'}]^{-1}\right\|_1 \\ = \ & \left\|Q_{\lambda\lambda'}^{-1} + Q_{\lambda\lambda'}^{-1} Q_{\lambda f'} [Q_{ff'} - Q_{f\lambda'} Q_{\lambda\lambda'}^{-1} Q_{\lambda f'}]^{-1} Q_{f\lambda'} Q_{\lambda\lambda'}^{-1}\right\|_1 \\ = \ & O_p(\frac{N^{\frac{1}{\varsigma}}}{T}) + O_p(\frac{N^{\frac{1}{\varsigma}}}{T}) O_p(N^{\frac{2}{\varsigma\wedge\zeta}} T) O_p(\frac{N^{\frac{1}{\varsigma}}}{T}) = O_p(\frac{N^{\frac{2}{\varsigma\wedge\zeta}+\frac{2}{\varsigma}}}{T}). \end{aligned} \tag{F.4}$$

We next show $\left\|Q_{\lambda f'} [Q_{ff'} - Q_{f\lambda'} Q_{\lambda\lambda'}^{-1} Q_{\lambda f'}]^{-1} Q_{f\lambda'}\right\|_1 = O_p(N^{\frac{2}{\varsigma\wedge\zeta}} T)$. It suffices to show

$$\left\|Q_{\lambda f'} [Q_{ff'} - Q_{f\lambda'} Q_{\lambda\lambda'}^{-1} Q_{\lambda f'}]^{-1} Q_{f\lambda'}\right\|_{\max} = O_p(\frac{T}{N} N^{\frac{2}{\varsigma\wedge\zeta}}). \tag{F.5}$$

For a matrix A of $Nr \times Tr$ which is written as an $N \times T$ block partitioned matrix with each block of size $r \times r$, we use the row index ip to denote the pth element of the ith row block, similarly for the column index. The (ip, jq)th element is $Q_{\lambda f'} [Q_{ff'} - Q_{f\lambda'} Q_{\lambda\lambda'}^{-1} Q_{\lambda f'}]^{-1} Q_{f\lambda'}$ is

$$[Q_{\lambda f'}]_{ip} [Q_{ff'} - Q_{f\lambda'} Q_{\lambda\lambda'}^{-1} Q_{\lambda f'}]^{-1} [Q_{\lambda f'}]'_{jq},$$

where $[Q_{\lambda f'}]_{ip}$ denotes the ip-th row of $Q_{\lambda f'}$. $[Q_{ff'} - Q_{f\lambda'} Q_{\lambda\lambda'}^{-1} Q_{\lambda f'}]^{-1}$ equals the lower right block of $Q_{\phi\phi'}^{-1} = D_{TN}^{-\frac{1}{2}} (D_{TN}^{-\frac{1}{2}} Q_{\phi\phi'} D_{TN}^{-\frac{1}{2}})^{-1} D_{TN}^{-\frac{1}{2}}$. Then by Lemma B.1,

$$\left\|[Q_{ff'} - Q_{f\lambda'} Q_{\lambda\lambda'}^{-1} Q_{\lambda f'}]^{-1}\right\| \leq \frac{1}{N} \left\|(D_{TN}^{-\frac{1}{2}} Q_{\phi\phi'} D_{TN}^{-\frac{1}{2}})^{-1}\right\| = O_p(\frac{1}{N}). \tag{F.6}$$

Recall that $Q_{\lambda f'} = L_{\lambda f'} + J_{\lambda f'} + G_{\lambda f'}$ and the (i,t)-th block of $L_{\lambda f'}$, $J_{\lambda f'}$ and $G_{\lambda f'}$ is $-d_{it} f_t^0 \lambda_i^{0\prime}$, $d_{it} v_{it} I_r$ and $c f_t^0 \lambda_i^{0\prime}$, respectively. It follows that

$$\begin{aligned} \max_{ip} \|[Q_{\lambda f'}]_{ip}\| &= \max_{ip} \|[L_{\lambda f'}]_{ip}\| + \max_{ip} \|[G_{\lambda f'}]_{ip}\| + \max_{ip} \|[J_{\lambda f'}]_{ip}\| \\ &= O_p(\sqrt{T} N^{\frac{1}{\varsigma}} + \sqrt{T} N^{\frac{1}{\zeta}}) = O_p(\sqrt{T} N^{\frac{1}{\varsigma\wedge\zeta}}), \end{aligned} \tag{F.7}$$

where the second equality is due to Assumptions 2 and 4(i).

(2.2) By symmetry, $\left\|(Q_{\phi\phi'}^{-1})_{lr}\right\|_1 = \left\|[Q_{ff'} - Q_{f\lambda'} Q_{\lambda\lambda'}^{-1} Q_{\lambda f'}]^{-1}\right\|_1 = O_p(\frac{T^{\frac{2}{\varsigma\wedge\zeta}+\frac{2}{\varsigma}}}{N})$.

(2.3) The upper-right block is $(Q_{\phi\phi'}^{-1})_{ur} = -(Q_{\phi\phi'}^{-1})_{ul} Q_{\lambda f'} Q_{ff'}^{-1}$. In step (2.1) we proved that $\left\|Q_{\lambda\lambda'}^{-1}\right\|_1 = O_p(\frac{N^{\frac{1}{\varsigma}}}{T})$. By symmetry, $\left\|Q_{ff'}^{-1}\right\|_1 = O_p(\frac{T^{\frac{1}{\varsigma}}}{N})$. By Assumptions 2 and 4(i), $\|Q_{\lambda f'}\|_1 \leq \|L_{\lambda f'}\|_1 + \|J_{\lambda f'}\|_1 +$

$\|G_{\lambda f'}\|_1 = O_p(NT^{\frac{1}{\varsigma}} + NT^{\frac{1}{\zeta}})$. These results, together with part (2.1), imply that $\left\|(Q_{\phi\phi'}^{-1})_{ur}\right\|_1 = O_p(\frac{N^{\frac{2}{\varsigma\wedge\zeta}+\frac{2}{\varsigma}}T^{\frac{1}{\varsigma\wedge\zeta}+\frac{1}{\varsigma}}}{T})$.

(2.4) The lower-left block is the transpose of the upper-right block, thus by symmetry, $\left\|(Q_{\phi\phi'}^{-1})_{ll}\right\|_1 = O_p(\frac{T^{\frac{2}{\varsigma\wedge\zeta}+\frac{2}{\varsigma}}N^{\frac{1}{\varsigma\wedge\zeta}+\frac{1}{\varsigma}}}{N})$.

Step (3): Recall that $Q_{\beta\phi'}Q_{\phi\phi'}^{-1} = (\delta^{*\prime}, \omega^{*\prime}) = (\delta_1^*, ..., \delta_N^*, \omega_1^*, ..., \omega_T^*)$, where δ_i^* and ω_t^* are $K \times r$ matrices and δ_{ki}^* and ω_{kt}^* denote the transpose of the k-th row of δ_i^* and ω_t^* respectively. Let $\delta_{k\cdot}^*$ and $\omega_{k\cdot}^*$ denote the k-th row of $\delta^{*\prime}$ and $\omega^{*\prime}$, respectively. It follows that

$$\begin{aligned}
\max_{i\in[N]}\|\delta_{ki}^*\| &\leq \sqrt{r}\,\|\delta_{k\cdot}^*\|_{\max} = \sqrt{r}\left\|Q_{\beta_k\lambda'}(Q_{\phi\phi'}^{-1})_{ul} + Q_{\beta_k f'}(Q_{\phi\phi'}^{-1})_{ll}\right\|_{\max} \\
&\leq \sqrt{r}\left\|Q_{\beta_k\lambda'}(Q_{\phi\phi'}^{-1})_{ul}\right\|_{\max} + \sqrt{r}\left\|Q_{\beta_k f'}(Q_{\phi\phi'}^{-1})_{ll}\right\|_{\max} \\
&\leq \sqrt{r}\,\|Q_{\beta_k\lambda'}\|_{\max}\left\|(Q_{\phi\phi'}^{-1})_{ul}\right\|_1 + \sqrt{r}\,\|Q_{\beta_k f'}\|_{\max}\left\|(Q_{\phi\phi'}^{-1})_{ll}\right\|_1 \\
&= O_p(TN^{\frac{1}{\varrho}})O_p(\frac{N^{\frac{2}{\varsigma\wedge\zeta}+\frac{2}{\varsigma}}}{T}) + O_p(NT^{\frac{1}{\varrho}})O_p(\frac{T^{\frac{2}{\varsigma\wedge\zeta}+\frac{2}{\varsigma}}N^{\frac{1}{\varsigma\wedge\zeta}+\frac{1}{\varsigma}}}{N}) \\
&= O_p(N^{\frac{2}{\varsigma\wedge\zeta}+\frac{2}{\varsigma}+\frac{1}{\varrho}}) + O_p(T^{\frac{2}{\varsigma\wedge\zeta}+\frac{2}{\varsigma}+\frac{1}{\varrho}}N^{\frac{1}{\varsigma\wedge\zeta}+\frac{1}{\varsigma}}) = O_p((N\vee T)^{\frac{3}{\varsigma\wedge\zeta}+\frac{3}{\varsigma}+\frac{1}{\varrho}}),
\end{aligned}$$

where the second equality is due to

$$\begin{aligned}
\|Q_{\beta_k\lambda'}\|_{\max} &= \max_{i\in[N]}\left\|\sum_{t=1}^T d_{it}x_{itk}f_t^{0\prime}\right\| \leq \max_{i\in[N]}\sqrt{\sum_{t=1}^T x_{itk}^2}\sqrt{\sum_{t=1}^T\|f_t^0\|^2} = O_p(TN^{\frac{1}{\varrho}}), \\
\|Q_{\beta_k f'}\|_{\max} &= \max_{t\in[T]}\left\|\sum_{i=1}^N d_{it}x_{itk}\lambda_i^{0\prime}\right\| \leq \max_{t\in[T]}\sqrt{\sum_{i=1}^N x_{itk}^2}\sqrt{\sum_{i=1}^N\|\lambda_i^0\|^2} = O_p(NT^{\frac{1}{\varrho}}).
\end{aligned}$$

By symmetry, we also have $\max_t\|\omega_{kt}^*\| = O_p(T^{\frac{2}{\varsigma\wedge\zeta}+\frac{2}{\varsigma}+\frac{1}{\varrho}}) + O_p(N^{\frac{2}{\varsigma\wedge\zeta}+\frac{2}{\varsigma}+\frac{1}{\varrho}}T^{\frac{1}{\varsigma\wedge\zeta}+\frac{1}{\varsigma}})$.

(ii) From equations (B.10) and (B.15), we have $(\delta^{0\prime}, \omega^{0\prime}) = \bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}$. Then the proof is the same as part (i) with Q, L and J replaced by $\bar{Q}$, $\bar{L}$ and $\bar{J}$, respectively. Since $\bar{J}_{\lambda f'} = 0$, the term $\max_{ip}\|[J_{\lambda f'}]_{ip}\|$ in eqn. (F.7) would disappear when J is replaced by $\bar{J}$. Consequently, ζ disappears throughout the whole proof. ■

Lemma F.4 *Suppose that Assumptions 1-2, 3(ii)-(iii) and 4 hold. Then as $(N,T)\to\infty$,*

(i) $\max_i\left\|\sum_{t=1}^T d_{it}f_t^0(\omega_{kt}^{*\prime} - \omega_{kt}^{0\prime})\right\| = O_p(\sqrt{T}N^{\frac{1}{\kappa}} + \frac{T^{1+\frac{1}{\kappa}}}{\sqrt{N}} + \frac{T}{\sqrt{c_{NT}}})$;

(ii) $\max_t\left\|\sum_{i=1}^N d_{it}\lambda_i^0(\delta_{ki}^{*\prime} - \delta_{ki}^{0\prime})\right\| = O_p(\sqrt{N}T^{\frac{1}{\kappa}} + \frac{N^{1+\frac{1}{\kappa}}}{\sqrt{T}} + \frac{N}{\sqrt{c_{NT}}})$;

(iii) $\max_i\left\|\sum_{t=1}^T(d_{it} - \Phi_{it})f_t^0\omega_{kt}^{0\prime}\right\| = O_p(\sqrt{T}N^{\frac{1}{\kappa}})$;

(iv) $\max_t\left\|\sum_{i=1}^N(d_{it} - \Phi_{it})\lambda_i^0\delta_{ki}^{0\prime}\right\| = O_p(\sqrt{N}T^{\frac{1}{\kappa}})$;

(v) $\max_t\left\|\sum_{i=1}^N \Phi_{it}\lambda_i^0\delta_{ki}^{0\prime}\right\| = O_p(N)$;

(vi) $\max_i\left\|\sum_{t=1}^T \Phi_{it}f_t^0\omega_{kt}^{0\prime}\right\| = O_p(T)$.

Proof. (i)–(ii). First note that

$$\sqrt{T\sum_{i=1}^N\left\|\delta_{ki}^* - \delta_{ki}^0\right\|^2 + N\sum_{t=1}^T\left\|\omega_{kt}^* - \omega_{kt}^0\right\|^2}$$

$$
\begin{aligned}
&\leq \left\|(Q_{\beta\phi'}Q_{\phi\phi'}^{-1}-\bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1})D_{TN}^{\frac{1}{2}}\right\| \\
&\leq \left\|\bar{Q}_{\beta\phi'}D_{TN}^{-\frac{1}{2}}\right\|\left\|D_{TN}^{\frac{1}{2}}(Q_{\phi\phi'}^{-1}-\bar{Q}_{\phi\phi'}^{-1})D_{TN}^{\frac{1}{2}}\right\|+\left\|(Q_{\beta\phi'}-\bar{Q}_{\beta\phi'})D_{TN}^{-\frac{1}{2}}\right\|\left\|(D_{TN}^{-\frac{1}{2}}Q_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}\right\| \\
&= O_p(\sqrt{NT})O_p(\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}}+\frac{T^{\frac{1}{\kappa}}}{\sqrt{N}}+\frac{1}{\sqrt{c_{NT}}})+O_p(\sqrt{N+T})=O_p(N^{\frac{1}{2}+\frac{1}{\kappa}}+T^{\frac{1}{2}+\frac{1}{\kappa}}+\frac{\sqrt{NT}}{\sqrt{c_{NT}}}),
\end{aligned}
$$

where the first equality follows from equations (B.17)-(B.18), (B.20) and Lemma B.1. It follows that $\sqrt{\sum_{t=1}^{T}\|\omega_{kt}^{*}-\omega_{kt}^{0}\|^{2}}=O_p(N^{\frac{1}{\kappa}}+\frac{T^{\frac{1}{2}+\frac{1}{\kappa}}}{\sqrt{N}}+\frac{\sqrt{T}}{\sqrt{c_{NT}}})$ and $\sqrt{\sum_{i=1}^{N}\|\delta_{ki}^{*}-\delta_{ki}^{0}\|^{2}}=O_p(T^{\frac{1}{\kappa}}+\frac{N^{\frac{1}{2}+\frac{1}{\kappa}}}{\sqrt{T}}+\frac{\sqrt{N}}{\sqrt{c_{NT}}})$. Then the results follow under Assumption 2.

(iii)–(iv). The proof follows from Assumption 1. Note that we assume $\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}}\to 0$ and $\frac{T^{\frac{1}{\kappa}}}{\sqrt{N}}\to 0$ in Assumption 1(i), and $d_{it}=\Phi_{it}$ under Assumption 1(ii) since our analysis is conditioning on the missing pattern under Assumption 1(ii).

(v)–(vi). First note that

$$
\begin{aligned}
&\sqrt{T\sum_{i=1}^{N}\left\|\delta_{ki}^{0}\right\|^{2}+N\sum_{t=1}^{T}\|\omega_{kt}^{0}\|^{2}} \\
=\ &\left\|\bar{Q}_{\beta\phi'}\bar{Q}_{\phi\phi'}^{-1}D_{TN}^{\frac{1}{2}}\right\|_F\leq\left\|\bar{Q}_{\beta\phi'}D_{TN}^{-\frac{1}{2}}\right\|_F\left\|(D_{TN}^{-\frac{1}{2}}\bar{Q}_{\phi\phi'}D_{TN}^{-\frac{1}{2}})^{-1}\right\|=O_p(\sqrt{NT}).
\end{aligned}
$$

Thus $\sum_{i=1}^{N}\left\|\delta_{ki}^{0}\right\|^{2}=O_p(N)$ and $\sum_{t=1}^{T}\left\|\omega_{kt}^{0}\right\|^{2}=O_p(T)$. Then the results follow under Assumption 2. ■

G Proof of Theorem 4.2

To prove Theorem 4.2, we first state and prove two technical lemmas.

Lemma G.1 *Under Assumptions 1-5, as* $(N,T)\to\infty$, $\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\delta}_{ki}-\delta_{ki}^{0}\right\|^{2}}=O_p(\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}}+\frac{T^{\frac{1}{\kappa}}}{\sqrt{N}}+\frac{1}{\sqrt{c_{NT}}})=o_p(1)$ *and* $\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\hat{\omega}_{kt}-\omega_{kt}^{0}\|^{2}}=O_p(\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}}+\frac{T^{\frac{1}{\kappa}}}{\sqrt{N}}+\frac{1}{\sqrt{c_{NT}}})=o_p(1)$.

Proof. The FOCs for the minimization problem in (4.3) are: for each $k\in[K]$,

$$
\begin{aligned}
\sum_{t=1}^{T}d_{it}(x_{itk}-\hat{\delta}_{ki}'\hat{f}_t-\hat{\omega}_{kt}'\hat{\lambda}_i)\hat{f}_t' &= 0 \text{ for all } i\in[N]\,, \\
\sum_{i=1}^{N}d_{it}(x_{itk}-\hat{\delta}_{ki}'\hat{f}_t-\hat{\omega}_{kt}'\hat{\lambda}_i)\hat{\lambda}_i' &= 0 \text{ for all } t\in[T]\,.
\end{aligned}
$$

As in step (1) of the proof of Lemma B.2(i), let $\hat{\delta}_i$ and $\hat{\omega}_t$ denote $K\times r$ matrices such that $\hat{\delta}_{ki}$ and $\hat{\omega}_{kt}$ are the transpose of the k-th row of $\hat{\delta}_i$ and $\hat{\omega}_t$, respectively. Let $\hat{\delta}=(\hat{\delta}_1,...,\hat{\delta}_N)'$ and $\hat{\omega}=(\hat{\omega}_1,...,\hat{\omega}_T)'$. Then the above FOCs imply

$$
Q_{\beta\phi'}^{(\infty)}=(\hat{\delta}',\hat{\omega}')L_{\phi\phi'}^{(\infty)}, \tag{G.1}
$$

where $Q_{\beta\phi'}^{(\infty)}=Q_{\beta\phi'}(\hat{\gamma})$, $L_{\phi\phi'}^{(\infty)}=L_{\phi\phi'}(\hat{\gamma})$, and $Q_{\beta\phi'}(\hat{\gamma})$ and $L_{\phi\phi'}(\hat{\gamma})$ are defined in expressions (A.3) and (A.6) with $\gamma=\hat{\gamma}$, respectively. Note that here we use the supscript $^{(\infty)}$ because in the proof of Theorem 3.2 we defined $Q_{\beta\phi'}^{(k+1)}=Q_{\beta\phi'}(s\hat{\gamma}^{(k+1)}+(1-s)\hat{\gamma})$ and $\hat{\gamma}^{(k)}\to\hat{\gamma}$ as $k\to\infty$.

From eqn. (G.1), we have $(\hat{\delta}',\hat{\omega}')=Q_{\beta\phi'}^{(\infty)}L_{\phi\phi'}^{(\infty)-1}$. From eqn. (B.10), we have $(\delta^{0\prime},\omega^{0\prime})=\bar{Q}_{\beta\phi'}\bar{L}_{\phi\phi'}^{-1}$. As remarked at the end of Section A, the inverse here actually denotes the Moore-Penrose pseudo-inverse because $L^{(\infty)}_{\phi\phi'}$ and $\bar{L}_{\phi\phi'}$ are not of full rank. Thus

$$
\begin{aligned}
&\sqrt{T\sum_{i=1}^{N}\left\|\hat{\delta}_{ki}-\delta^0_{ki}\right\|^2+N\sum_{t=1}^{T}\left\|\hat{\omega}_{kt}-\omega^0_{kt}\right\|^2}\\
=\ &\left\|(Q^{(\infty)}_{\beta_k\phi'}L^{(\infty)-1}_{\phi\phi'}-\bar{Q}_{\beta_k\phi'}\bar{L}^{-1}_{\phi\phi'})D^{\frac{1}{2}}_{TN}\right\|_F\le\left\|(Q^{(\infty)}_{\beta\phi'}L^{(\infty)-1}_{\phi\phi'}-\bar{Q}_{\beta\phi'}\bar{L}^{-1}_{\phi\phi'})D^{\frac{1}{2}}_{TN}\right\|\\
\le\ &\left\|Q^{(\infty)}_{\beta\phi'}(L^{(\infty)-1}_{\phi\phi'}-\bar{L}^{-1}_{\phi\phi'})D^{\frac{1}{2}}_{TN}\right\|+\left\|(Q^{(\infty)}_{\beta\phi'}-\bar{Q}_{\beta\phi'})\bar{L}^{-1}_{\phi\phi'}D^{\frac{1}{2}}_{TN}\right\|\\
\le\ &\left\|Q^{(\infty)}_{\beta\phi'}D^{-\frac{1}{2}}_{TN}\right\|\left\|D^{\frac{1}{2}}_{TN}(L^{(\infty)-1}_{\phi\phi'}-\bar{L}^{-1}_{\phi\phi'})D^{\frac{1}{2}}_{TN}\right\|+\left\|(Q^{(\infty)}_{\beta\phi'}-\bar{Q}_{\beta\phi'})D^{-\frac{1}{2}}_{TN}\right\|\left\|(D^{-\frac{1}{2}}_{TN}\bar{L}_{\phi\phi'}D^{-\frac{1}{2}}_{TN})^{-1}\right\|\\
=\ &O_p(\sqrt{NT})O_p(\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}}+\frac{T^{\frac{1}{\kappa}}}{\sqrt{N}}+\frac{1}{\sqrt{c_{NT}}})+O_p(\sqrt{N+T})=O_p(N^{\frac{1}{2}+\frac{1}{\kappa}}+T^{\frac{1}{2}+\frac{1}{\kappa}}+\frac{\sqrt{NT}}{\sqrt{c_{NT}}}),
\end{aligned}
$$

where the second equality holds by the following arguments.

(1) Similarly to expression (B.25), $\left\|Q^{(\infty)}_{\beta\phi'}D^{-\frac{1}{2}}_{TN}\right\|=O_p(\sqrt{NT})$.

(2) We want to show that $\left\|(Q^{(\infty)}_{\beta\phi'}-\bar{Q}_{\beta\phi'})D^{-\frac{1}{2}}_{TN}\right\|=O_p(\sqrt{N+T})$. First, $\left\|(Q_{\beta\phi'}-\bar{Q}_{\beta\phi'})D^{-\frac{1}{2}}_{TN}\right\|=O_p(\sqrt{N+T})$ by expression (B.18). Second, by Assumption 3(ii) and Theorem 3.1,

$$
\begin{aligned}
\left\|(Q^{(\infty)}_{\beta\phi'}-Q_{\beta\phi'})D^{-\frac{1}{2}}_{TN}\right\| &\le \sqrt{\sum_{i=1}^{N}\left\|\frac{\sum_{t=1}^{T}d_{it}x_{it}(\hat{f}_t-f^0_t)'}{\sqrt{T}}\right\|_F^2+\sum_{t=1}^{T}\left\|\frac{\sum_{i=1}^{N}d_{it}x_{it}(\hat{\lambda}_i-\lambda^0_i)'}{\sqrt{N}}\right\|_F^2}\\
&\le \sqrt{\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|x_{it}\right\|^2(\frac{\left\|f-f^0\right\|^2}{T}+\frac{\left\|\lambda-\lambda^0\right\|^2}{N})}=O_p(\frac{\sqrt{NT}}{c_{NT}}).
\end{aligned}
$$

Combining these results yields the claim.

(3) We want to show $\left\|(D^{-\frac{1}{2}}_{TN}\bar{L}_{\phi\phi'}D^{-\frac{1}{2}}_{TN})^{-1}\right\|=O_p(1)$. As explained in step (2) of the proof of Lemma B.2, any two different columns of $D^{-\frac{1}{2}}_{NT}W^0$ are orthogonal to each other, the columns of $D^{-\frac{1}{2}}_{NT}W^0$ are all orthogonal to the eigenvectors of $D^{-\frac{1}{2}}_{TN}\bar{L}_{\phi\phi'}D^{-\frac{1}{2}}_{TN}$, and $D^{-\frac{1}{2}}_{TN}\bar{Q}_{\phi\phi'}D^{-\frac{1}{2}}_{TN}=D^{-\frac{1}{2}}_{TN}\bar{L}_{\phi\phi'}D^{-\frac{1}{2}}_{TN}-cD^{-\frac{1}{2}}_{NT}W^0W^{0\prime}D^{-\frac{1}{2}}_{NT}$. All nonzero eigenvalues of $-D^{-\frac{1}{2}}_{TN}\bar{L}_{\phi\phi'}D^{-\frac{1}{2}}_{TN}$ are positive. Thus the smallest nonzero eigenvalue of $-D^{-\frac{1}{2}}_{TN}\bar{L}_{\phi\phi'}D^{-\frac{1}{2}}_{TN}$ is not smaller than the smallest eigenvalue of $-D^{-\frac{1}{2}}_{TN}\bar{Q}_{\phi\phi'}D^{-\frac{1}{2}}_{TN}$. It follows that

$$
\left\|(D^{-\frac{1}{2}}_{TN}\bar{L}_{\phi\phi'}D^{-\frac{1}{2}}_{TN})^{-1}\right\|\le\left\|(D^{-\frac{1}{2}}_{TN}\bar{Q}_{\phi\phi'}D^{-\frac{1}{2}}_{TN})^{-1}\right\|=O_p(1),
$$

where the equality follows from step (1.1) of Lemma B.1 in Su and Wang (2024) with $\bar{Q}_{\phi\phi'}$ corresponding to $\breve{H}_{\phi\phi'}$ there. For the block-type missing, the equality follows from Lemma B.1 (or Lemma B.2 in Su and Wang, 2024).

(4) $\left\|D^{-\frac{1}{2}}_{TN}(L_{\phi\phi'}-\bar{L}_{\phi\phi'})D^{-\frac{1}{2}}_{TN}\right\|=O_p(\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}}+\frac{T^{\frac{1}{\kappa}}}{\sqrt{N}}+\frac{1}{\sqrt{c_{NT}}})$ by step (1.3) of Lemma B.1 in Su and Wang (2024). We also have $\left\|D^{-\frac{1}{2}}_{TN}(L^{(\infty)}_{\phi\phi'}-L_{\phi\phi'})D^{-\frac{1}{2}}_{TN}\right\|=O_p(\frac{1}{c_{NT}})$ due to the following:

$$
\begin{aligned}
\left\|\frac{1}{T}(L_{\lambda\lambda'}(\hat{\gamma})-L_{\lambda\lambda'})\right\|&\le\left\|\frac{1}{T}\sum_{t=1}^{T}d_{it}\hat{f}_t\hat{f}_t'-\frac{1}{T}\sum_{t=1}^{T}d_{it}f^0_tf^{0\prime}_t\right\|=O_p(\frac{1}{c_{NT}}),\\
\left\|\frac{1}{N}(L_{ff'}(\hat{\gamma})-L_{ff'})\right\|&\le\left\|\frac{1}{N}\sum_{i=1}^{N}d_{it}\hat{\lambda}_i'\hat{\lambda}_i-\frac{1}{N}\sum_{i=1}^{N}d_{it}\lambda^0_i\lambda^{0\prime}_i\right\|=O_p(\frac{1}{c_{NT}}),
\end{aligned}
$$

$$\left\| \frac{1}{\sqrt{NT}} (L_{\lambda f'}(\hat{\gamma}) - L_{\lambda f'}) \right\|_F \le \sqrt{\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\| d_{it} \hat{f}_t \hat{\lambda}_i' - d_{it} f_t^0 \lambda_i^{0\prime} \right\|_F^2} = O_p(\frac{1}{c_{NT}}),$$

by a simple application of the results in Theorem 3.1. It follows that

$$\begin{aligned}
\left\| D_{TN}^{\frac{1}{2}} (L_{\phi\phi'}^{(\infty)-1} - \bar{L}_{\phi\phi'}^{-1}) D_{TN}^{\frac{1}{2}} \right\| &= \left\| (D_{TN}^{\frac{1}{2}} \bar{L}_{\phi\phi'}^{-1} D_{TN}^{\frac{1}{2}}) D_{TN}^{-\frac{1}{2}} (L_{\phi\phi'}^{(\infty)} - \bar{L}_{\phi\phi'}) D_{TN}^{-\frac{1}{2}} (D_{TN}^{\frac{1}{2}} L_{\phi\phi'}^{(\infty)-1} D_{TN}^{\frac{1}{2}}) \right\| \\
&\le \left\| (D_{TN}^{-\frac{1}{2}} \bar{L}_{\phi\phi'} D_{TN}^{-\frac{1}{2}})^{-1} \right\| \left\| D_{TN}^{-\frac{1}{2}} (L_{\phi\phi'}^{(\infty)} - \bar{L}_{\phi\phi'}) D_{TN}^{-\frac{1}{2}} \right\| \left\| (D_{TN}^{-\frac{1}{2}} L_{\phi\phi'}^{(\infty)} D_{TN}^{-\frac{1}{2}})^{-1} \right\| \\
&= O_p(\frac{N^{\frac{1}{\kappa}}}{\sqrt{T}} + \frac{T^{\frac{1}{\kappa}}}{\sqrt{N}} + \frac{1}{\sqrt{c_{NT}}}).
\end{aligned}$$

■

Lemma G.2 *Suppose that Assumptions 1-5 hold. Then* $\max_{it} |\hat{v}_{it}| = O_p(N^{\frac{1}{\zeta}} T^{\frac{1}{\zeta}}) + O_p(\frac{\sqrt{NT}}{c_{NT}^2})$.

Proof. First, by Theorem 3.1 and Assumption 2, we have $\max_i \left\| \hat{\lambda}_i \right\| \le \max_i \left\| \lambda_i^0 \right\| + \max_i \left\| \hat{\lambda}_i - \lambda_i^0 \right\| \le \max_i \left\| \lambda_i^0 \right\| + \sqrt{\sum_{i=1}^{N} \left\| \hat{\lambda}_i - \lambda_i^0 \right\|^2} = O_p(\frac{\sqrt{N}}{c_{NT}})$. Similarly, we also have $\max_t \left\| \hat{f}_t \right\| = O_p(\frac{\sqrt{T}}{c_{NT}})$. It follows that

$$\begin{aligned}
\max_{it} |\hat{v}_{it} - v_{it}| &\le \max_{it} \|x_{it}\| \left\| \hat{\beta} - \beta^0 \right\| + \max_{it} \left\| \hat{\lambda}_i \right\| \left\| \hat{f}_t \right\| + \max_{it} \left\| \lambda_i^0 \right\| \left\| f_t^0 \right\| \\
&= O_p(\frac{N^{\frac{1}{\varrho}} T^{\frac{1}{\varrho}}}{c_{NT}}) + O_p(\frac{\sqrt{NT}}{c_{NT}^2}) = O_p(\frac{\sqrt{NT}}{c_{NT}^2}) \text{ when } \varrho \ge 4, \qquad \text{(G.2)}
\end{aligned}$$

where the equality also uses $\max_{it} \|x_{it}\| \le (\sum_{i=1}^{N} \sum_{t=1}^{T} \|x_{it}\|^{\varrho})^{\frac{1}{\varrho}} = O_p(N^{\frac{1}{\varrho}} T^{\frac{1}{\varrho}})$ by Assumption 3(ii) and $\left\| \hat{\beta} - \beta^0 \right\| = O_p(\frac{1}{c_{NT}})$ by Theorem 3.1. In addition, $\max_{it} |v_{it}| = O_p(N^{\frac{1}{\zeta}} T^{\frac{1}{\zeta}})$ by Assumption 4(i), thus $\max_{it} |\hat{v}_{it}| \le \max_{it} |v_{it}| + \max_{it} |\hat{v}_{it} - v_{it}| = O_p(N^{\frac{1}{\zeta}} T^{\frac{1}{\zeta}}) + O_p(\frac{\sqrt{NT}}{c_{NT}^2})$. ■

Proof of Theorem 4.2

(i) (1) Consistency of $\hat{W}_x$: Since $\dot{x}_{it} = x_{it} - (\delta_i^0 f_t^0 + \omega_t^0 \lambda_i^0)$,

$$\begin{aligned}
W_x &= plim \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it} \dot{x}_{it} \dot{x}_{it}' \\
&= plim \frac{1}{NT} \Big\{ \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it} x_{it} x_{it}' - \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it} x_{it} (\delta_i^0 f_t^0 + \omega_t^0 \lambda_i^0)' \\
&\quad - \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it} (\delta_i^0 f_t^0 + \omega_t^0 \lambda_i^0) x_{it}' + \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it} (\delta_i^0 f_t^0 + \omega_t^0 \lambda_i^0)(\delta_i^0 f_t^0 + \omega_t^0 \lambda_i^0)' \Big\} \\
&= -(\bar{Q}_{\beta\beta'} - \bar{Q}_{\beta\phi'} \bar{L}_{\phi\phi'}^{-1} \bar{Q}_{\phi\beta'}),
\end{aligned}$$

where the last equality follows from Assumptions 3(iii) and 1 and equations (B.11)–(B.12). Then it suffices to show $\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\| d_{it} \hat{x}_{it} \hat{x}_{it}' - d_{it} \dot{x}_{it} \dot{x}_{it}' \right\| = O_p(\frac{1}{c_{NT}})$. Since $\hat{x}_{it} = x_{it} - \hat{\delta}_i \hat{f}_t - \hat{\omega}_t \hat{\lambda}_i$, it suffices to show

$$\begin{aligned}
\sum_{i=1}^{N} \sum_{t=1}^{T} \left\| d_{it} x_{it} (\hat{\delta}_i \hat{f}_t + \hat{\omega}_t \hat{\lambda}_i)' - d_{it} x_{it} (\delta_i^0 f_t^0 + \omega_t^0 \lambda_i^0)' \right\| &= O_p(\frac{NT}{c_{NT}}) \text{ and} \\
\sum_{i=1}^{N} \sum_{t=1}^{T} \left\| d_{it} (\hat{\delta}_i \hat{f}_t + \hat{\omega}_t \hat{\lambda}_i)(\hat{\delta}_i \hat{f}_t + \hat{\omega}_t \hat{\lambda}_i)' - d_{it} (\delta_i^0 f_t^0 + \omega_t^0 \lambda_i^0)(\delta_i^0 f_t^0 + \omega_t^0 \lambda_i^0)' \right\| &= O_p(\frac{NT}{c_{NT}}).
\end{aligned}$$

It's easy to see that these follow from Theorem 3.1 and Lemma G.1, e.g.,

$$\begin{aligned}
&\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\left\|d_{it}x_{it}(\hat{\delta}_i\hat{f}_t)' - d_{it}x_{it}(\delta_i^0 f_t^0)'\right\| \\
\leq\ &\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\left\|d_{it}x_{it}(\hat{\delta}_i\hat{f}_t)' - d_{it}x_{it}(\delta_i^0 \hat{f}_t)'\right\| + \sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\left\|d_{it}x_{it}(\delta_i^0\hat{f}_t)' - d_{it}x_{it}(\delta_i^0 f_t^0)'\right\| \\
\leq\ &\sqrt{\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} d_{it}\left\|x_{it}\right\|^2}\sqrt{\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\left\|(\hat{\delta}_i-\delta_i^0)\right\|^2\left\|\hat{f}_t\right\|^2} \\
&+\sqrt{\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T} d_{it}\left\|x_{it}\right\|^2}\sqrt{\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\left\|\delta_i^0\right\|^2\left\|\hat{f}_t-f_t^0\right\|^2} \\
=\ &O_p(\sqrt{NT})O_p(\frac{\sqrt{NT}}{c_{NT}}) = O_p(\frac{NT}{c_{NT}}).
\end{aligned}$$

(2) Consistency of $\hat{\Omega}_x$ without cross-sectional and serial dependence:

First, part (1) shows $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|d_{it}\hat{\dot{x}}_{it}\hat{\dot{x}}_{it}' - d_{it}\dot{x}_{it}\dot{x}_{it}'\right\| = O_p(\frac{1}{c_{NT}})$, which, along with Lemma G.2, implies that if $N^{\frac{2}{\zeta}}/T^{\frac{1}{2}-\frac{2}{\zeta}} \to 0$, $T^{\frac{2}{\zeta}}/N^{\frac{1}{2}-\frac{2}{\zeta}} \to 0$, $N/T^{\frac{3}{2}} \to 0$ and $T/N^{\frac{3}{2}} \to 0$ (which are satisfied when $\zeta > 8$ and $T/N \to \varepsilon \in (0,\infty)$),

$$\begin{aligned}
&\left\|\hat{\Omega}_x - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}\hat{v}_{it}^2\dot{x}_{it}\dot{x}_{it}'\right\| = \left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}\hat{v}_{it}^2(\hat{\dot{x}}_{it}\hat{\dot{x}}_{it}' - \dot{x}_{it}\dot{x}_{it}')\right\| \\
\leq\ &\max_{i,t}\hat{v}_{it}^2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}\left\|\hat{\dot{x}}_{it}\hat{\dot{x}}_{it}' - \dot{x}_{it}\dot{x}_{it}'\right\| = [O_p(N^{\frac{2}{\zeta}}T^{\frac{2}{\zeta}}) + O_p(\frac{NT}{c_{NT}^4})]O_p(\frac{1}{c_{NT}}) \\
=\ &o_p(1).
\end{aligned} \tag{G.3}$$

Second, given that $\hat{v}_{it} = y_{it} - x_{it}'\hat{\beta} - \hat{\lambda}_i'\hat{f}_t$, by Assumption 3(ii) and Theorem 3.1,

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{v}_{it}-v_{it})^2 = \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}'(\hat{\beta}-\beta^0) + \hat{\lambda}_i'\hat{f}_t - \lambda_i^{0\prime}f_t^0)^2 \\
\leq\ &2\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|x_{it}\right\|^2\left\|\hat{\beta}-\beta^0\right\|^2 + 4\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\lambda}_i'\hat{f}_t - \lambda_i^{0\prime}\hat{f}_t)^2 + 4\sum_{i=1}^{N}\sum_{t=1}^{T}(\lambda_i^{0\prime}\hat{f}_t - \lambda_i^{0\prime}f_t^0)^2 \\
=\ &O_p(NT)O_p(\frac{1}{c_{NT}^2}) + O_p(\frac{NT}{c_{NT}^2}) = O_p(\frac{NT}{c_{NT}^2}).
\end{aligned} \tag{G.4}$$

Note that $\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2 = O_p(NT)$ by Assumption 4 (i). In addition,

$$\begin{aligned}
\max_{i,t}\left\|\dot{x}_{itk}\right\| &\leq \max_{i,t}\left\|x_{itk}\right\| + \max_{i,t}\left\|\delta_{ki}^{0\prime}f_t^0 + \omega_{kt}^{0\prime}\lambda_i^0\right\| \\
&= O_p(N^{\frac{1}{\varrho}}T^{\frac{1}{\varrho}}) + O_p((N\vee T)^{\frac{6}{\varsigma}+\frac{1}{\varrho}}) = O_p(N^{\frac{1}{\varrho}}T^{\frac{1}{\varrho}})
\end{aligned} \tag{G.5}$$

by Lemma F.3(ii), Assumption 3(ii) and Assumption 2 . Note that $\varsigma = \infty$ when $\left\|f_t^0\right\|$ and $\left\|\lambda_i^0\right\|$ are uniformly bounded. It follows that

$$\left\|\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}(\hat{v}_{it}^2 - v_{it}^2)\dot{x}_{it}\dot{x}_{it}'\right\| = \left\|\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}(\hat{v}_{it}-v_{it})^2\dot{x}_{it}\dot{x}_{it}' + \sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}(\hat{v}_{it}-v_{it})2v_{it}\dot{x}_{it}\dot{x}_{it}'\right\|$$

$$\leq \max_{i,t} \|\dot{x}_{it}\|^2 [\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{v}_{it}-v_{it})^2 + 2\sqrt{\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{v}_{it}-v_{it})^2}\sqrt{\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}^2}]$$
$$= O_p(N^{\frac{2}{\varrho}}T^{\frac{2}{\varrho}})O_p(\frac{NT}{c_{NT}}) = o_p(NT) \text{ if } N^{\frac{2}{\zeta}}/T^{\frac{1}{2}-\frac{2}{\zeta}} \to 0 \text{ and } T^{\frac{2}{\zeta}}/N^{\frac{1}{2}-\frac{2}{\zeta}} \to 0. \quad (G.6)$$

In sum, we have shown $\left\|\hat{\Omega}_x - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}v_{it}^2\dot{x}_{it}\dot{x}_{it}'\right\| = o_p(1)$. This implies $\left\|\hat{\Omega}_x - \Omega_x\right\| = o_p(1)$ since $\Omega_x = plim\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}v_{it}^2\dot{x}_{it}\dot{x}_{it}'$.

(3) Consistency of $\hat{b}_{2k}, ..., \hat{b}_{6k}$. We focus on the analysis of $\hat{b}_{2k}$ as the proof of the consistency of the other terms is analogous. We shall show that $\hat{b}_{2k} = b_{2k} + o_p(1)$ in four steps below.

Step (1): First,

$$\left\|\hat{b}_{2k} - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\frac{1}{T}\Gamma(\frac{s-t}{L_T})d_{it}v_{it}d_{is}x_{isk}\hat{f}_t'[T\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\hat{f}_s\right\|$$
$$= \left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T\wedge(t+L_T)}\frac{1}{T}d_{it}(\hat{v}_{it}-v_{it})d_{is}x_{isk}\hat{f}_t'[T\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\hat{f}_s\right\|$$
$$\leq \max_{i,t,s}\left\|\hat{f}_t'[T\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\hat{f}_s\right\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T\wedge(t+L_T)}\left\|\frac{1}{T}d_{it}(\hat{v}_{it}-v_{it})d_{is}x_{isk}\right\|$$
$$\leq \max_{i,t,s}\left\|\hat{f}_t'[T\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\hat{f}_s\right\| L_T\max_{i,s}\|x_{is}\|\frac{1}{T\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\|d_{it}(\hat{v}_{it}-v_{it})\|$$
$$\leq \max_{i,t,s}\left\|\hat{f}_t'[T\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\hat{f}_s\right\| L_T\max_{i,s}\|x_{is}\|\frac{1}{T}\sqrt{\sum_{i=1}^{N}\sum_{t=1}^{T-1}(\hat{v}_{it}-v_{it})^2}$$
$$= O_p(\frac{T}{c_{NT}^2})L_TO_p(N^{\frac{1}{\varrho}}T^{\frac{1}{\varrho}})\frac{1}{T}O_p(\frac{\sqrt{NT}}{c_{NT}})$$
$$= O_p(L_T\frac{N^{\frac{1}{2}+\frac{1}{\varrho}}T^{\frac{1}{2}+\frac{1}{\varrho}}}{c_{NT}^3}) = o_p(1) \text{ if } \frac{N}{T}\to\varepsilon \text{ and } \frac{L_T}{T^{\frac{1}{2}-\frac{2}{\varrho}}}\to 0, \quad (G.7)$$

where the second equality follows from $\sum_{i=1}^{N}\sum_{t=1}^{T-1}(\hat{v}_{it}-v_{it})^2 = O_p(\frac{NT}{c_{NT}^2})$ by eqn. (G.4), $\max_{i,s}\|x_{is}\| = O_p(N^{\frac{1}{\varrho}}T^{\frac{1}{\varrho}})$ by Assumption 3(ii) and

$$\max_{i,t,s}\left\|\hat{f}_t'[T\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\hat{f}_s\right\| \leq \max_t\left\|\hat{f}_t\right\|\max_i\left\|T[\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\right\|\max_s\left\|\hat{f}_s\right\| = O_p(\frac{T}{c_{NT}^2}). \quad (G.8)$$

Expression(G.8) is due to $\max_t\left\|\hat{f}_t\right\| = O_p(\frac{\sqrt{T}}{c_{NT}})$ as explained at the beginning of the proof of Lemma G.2 and $\max_i\left\|T[\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\right\| \leq \frac{1}{\underline{d}}\left\|(\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t\hat{f}_t')^{-1}\right\| = O_p(1)$.

Step (2): Next,

$$\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\frac{1}{T}\Gamma(\frac{s-t}{L_T})d_{it}v_{it}d_{is}x_{isk}(\hat{f}_t'[T\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\hat{f}_s - f_t^{0\prime}[T\bar{L}_{\lambda\lambda'}^{-1}]_if_s^0)\right\|$$
$$\leq \max_{its}\|v_{it}\|\|x_{is}\|\frac{1}{T\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T\wedge(t+L_T)}\left\|\hat{f}_t'[T\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\hat{f}_s - f_t^{0\prime}[T\bar{L}_{\lambda\lambda'}^{-1}]_if_s^0\right\|$$
$$= O_p(N^{\frac{1}{\zeta}+\frac{1}{\varrho}}T^{\frac{1}{\zeta}+\frac{1}{\varrho}})\frac{N}{\sqrt{NT}}L_TO_p(\frac{1}{c_{NT}}) = o_p(1) \text{ if } \frac{N}{T}\to\varepsilon \text{ and } \frac{L_T}{T^{\frac{1}{2}-\frac{2}{\varrho}-\frac{2}{\zeta}}}\to 0, \quad (G.9)$$

where $\max_{it}|v_{it}| = O_p(N^{\frac{1}{\zeta}}T^{\frac{1}{\zeta}})$ by Assumption 4(i), $\max_{is}\|x_{is}\| = O_p(N^{\frac{1}{\varrho}}T^{\frac{1}{\varrho}})$ by Assumption 3(ii) and

$\frac{1}{T\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T\wedge(t+L_T)}\left\|\hat{f}_t'[T\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\hat{f}_s - f_t^{0\prime}[T\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0\right\| = \frac{NL_T}{\sqrt{NT}}O_p(\frac{1}{c_{NT}})$ is proved in the following. First,

$$\begin{aligned}
&\frac{1}{T\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T\wedge(t+L_T)}\left\|\hat{f}_t'[T\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\hat{f}_s - \hat{f}_t'[T\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i f_s^0\right\| \\
\leq\; & \frac{1}{T\sqrt{NT}}N\max_i\left\|T[\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\right\|\sum_{t=1}^{T-1}\sum_{s=t+1}^{T\wedge(t+L_T)}\left\|\hat{f}_t'\right\|\left\|\hat{f}_s - f_s^0\right\| \\
\leq\; & \frac{1}{T\sqrt{NT}}N\max_i\left\|T[\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\right\|L_T\sqrt{\sum_{t=1}^{T}\left\|\hat{f}_t\right\|^2}\sqrt{\sum_{s=1}^{T}\left\|\hat{f}_s - f_s^0\right\|^2} \\
=\; & \frac{1}{T\sqrt{NT}}NO_p(1)L_T O_p(\sqrt{T})O_p(\frac{\sqrt{T}}{c_{NT}}) = \frac{N}{\sqrt{NT}}L_T O_p(\frac{1}{c_{NT}}),
\end{aligned}$$

where the second inequality is due to $\sum_{t=1}^{T-1}\left\|\hat{f}_t\right\|\left\|\hat{f}_{t+l} - f_{t+l}^0\right\| \leq \sqrt{\sum_{t=1}^{T}\left\|\hat{f}_t\right\|^2\sum_{s=1}^{T}\left\|\hat{f}_s - f_s^0\right\|^2}$ for $l \in [L_T]$. Similarly, $\frac{1}{T\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T\wedge(t+L_T)}\left\|\hat{f}_t'[T\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0 - f_t^{0\prime}[T\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0\right\| = \frac{NL_T}{\sqrt{NT}}O_p(\frac{1}{c_{NT}})$. Second,

$$\begin{aligned}
&\frac{1}{T\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T\wedge(t+L_T)}\left\|\hat{f}_t'[T\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i f_s^0 - \hat{f}_t'[T\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0\right\| \\
\leq\; & \frac{1}{T\sqrt{NT}}\sum_{i=1}^{N}\left\|[T\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i - [T\bar{L}_{\lambda\lambda'}^{-1}]_i\right\|\sum_{t=1}^{T-1}\sum_{s=t+1}^{T\wedge(t+L_T)}\left\|\hat{f}_t\right\|\left\|f_s^0\right\| \\
\leq\; & \frac{N}{T\sqrt{NT}}\max_i\left\|[T\hat{\bar{L}}_{\lambda\lambda'}^{-1}]_i\right\|\left\|[T\bar{L}_{\lambda\lambda'}^{-1}]_i\right\|\left\|[T\hat{\bar{L}}_{\lambda\lambda'}]_i - [T\bar{L}_{\lambda\lambda'}]_i\right\|L_T\sqrt{\sum_{t=1}^{T}\left\|\hat{f}_t\right\|^2\sum_{s=1}^{T}\left\|f_s^0\right\|^2} \\
=\; & \frac{N}{T\sqrt{NT}}O_p(1)O_p(1)O_p(\frac{1}{c_{NT}})L_T T = \frac{NL_T}{\sqrt{NT}}O_p(\frac{1}{c_{NT}}),
\end{aligned}$$

where we uses $\max_i\left\|[T\hat{\bar{L}}_{\lambda\lambda'}]_i - [T\bar{L}_{\lambda\lambda'}]_i\right\| = \max_i\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_\phi(d_{it})\hat{f}_t\hat{f}_t' - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_\phi(d_{it})f_t^0 f_t^{0\prime}\right\| = O_p(\frac{1}{c_{NT}})$ for the first equality.

Step (3):

$$\begin{aligned}
&\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\frac{1}{T}\Gamma(\frac{s-t}{L_T})[d_{it}v_{it}d_{is}x_{isk} - \mathbb{E}_\phi(d_{it}v_{it}d_{is}x_{isk})]f_t^{0\prime}[T\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0\right\| \\
\leq\; & \frac{1}{T\sqrt{NT}}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T\wedge(t+L_T)}\left\|f_t^0\right\|\left\|f_s^0\right\|\left\|\sum_{i=1}^{N}[d_{it}v_{it}d_{is}x_{isk} - \mathbb{E}_\phi(d_{it}v_{it}d_{is}x_{isk})][T\bar{L}_{\lambda\lambda'}^{-1}]_i\right\| \\
\leq\; & \frac{M\sqrt{TL_T}}{T\sqrt{NT}}\sqrt{\sum_{t=1}^{T-1}\sum_{s=t+1}^{T\wedge(t+L_T)}\left\|\sum_{i=1}^{N}[d_{it}v_{it}d_{is}x_{isk} - \mathbb{E}_\phi(d_{it}v_{it}d_{is}x_{isk})][T\bar{L}_{\lambda\lambda'}^{-1}]_i\right\|^2} \\
=\; & \frac{M\sqrt{TL_T}}{T\sqrt{NT}}O_p(\sqrt{NTL_T}) = O_p(\frac{1}{\sqrt{T}}L_T) = o_p(1), \qquad \text{(G.10)}
\end{aligned}$$

where the first equality follows from Assumption 9(i).

Step (4):

$$\begin{aligned}
&\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{T}\Gamma(\frac{s-t}{L_T})\mathbb{E}_\phi(d_{it}v_{it}d_{is}x_{isk})f_t^{0\prime}[T\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0 - b_{2k}\right| \\
=\; & \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=t+L_T+1}^{T}\frac{1}{T}\mathbb{E}_\phi(d_{it}v_{it}d_{is}x_{isk})f_t^{0\prime}[T\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{T\sqrt{NT}} \max_{i,t,s} \left\| f_t^{0\prime} [T\bar{L}_{\lambda\lambda'}^{-1}]_i f_s^0 \right\| \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=t+L_T+1}^{T} \mathbb{E}_\phi(d_{it} v_{it} d_{is} x_{isk}) \\
&\leq \frac{1}{T\sqrt{NT}} MN \sum_{t=1}^{T} \sum_{s=t+L_T+1}^{T} M \left| s-t \right|^{c_2} \leq \frac{M^2 N}{\sqrt{NT}} \sum_{s-t=L_T+1}^{\infty} \left| s-t \right|^{c_2} \\
&= o(\frac{N}{\sqrt{NT}}) \text{ as } L_T \to \infty. \qquad \text{(G.11)}
\end{aligned}
$$

Expressions (G.7), (G.9), (G.10) and (G.11) prove the consistency of $\hat{b}_{2k}$ when $\frac{N}{T} \to \varepsilon$, $\frac{L_T}{T^{\frac{1}{2} - \frac{4}{\varrho \wedge \zeta}}} \to 0$ and $L_T \to \infty$.

The consistency of $\hat{b}_{3k}, ..., \hat{b}_{6k}$ and $\hat{\Omega}_x$ when there is serial dependence (under Assumption 9) can be proved similarly to $\hat{b}_{2k}$. The main difference is that the rate in Lemma G.1 is $\frac{1}{\sqrt{c_{NT}}}$ rather than $\frac{1}{c_{NT}}$, consequently we need $\frac{L_T}{T^{\frac{1}{4} - \frac{4}{\varrho \wedge \zeta}}} \to 0$. Note that these conditions on L_T is sufficient but not necessary. ■

H Details for the Alternating Direction Method of Multipliers Algorithm

ADMM is a powerful algorithm for solving optimization problems with structured objectives and constraints, like the NNR problem. First, it splits complex optimization problems into simpler subproblems and is capable to handle non-smooth objective functions, which makes it particularly effective for solving NNR, as the optimization can be separated into a smooth component and a proximal operator for the nuclear norm, leveraging SVT. Second, the tuning parameter set up is easy for ADMM, which can achieve robust convergence behavior. Third, ADMM has well-established convergence properties. Compared to common optimization approaches like gradient descent which is not suitable for NNR due to the non-differentiability, ADMM provides a balance of flexibility, scalability, and computational efficiency that makes it a preferred choice for our problem.

Our optimization problem is equivalent to the following one:

$$
\begin{aligned}
&\min_{\beta,\Theta,Z_\Theta} \frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it}(y_{it} - \Theta_{it} - x_{it}'\beta)^2 + \nu_{NT} \left\| Z_\Theta \right\|_* + \frac{\rho}{2} \left\| Z_\Theta - \Theta \right\|_F^2, \\
&\text{s.t. } Z_\Theta - \Theta = 0,
\end{aligned}
$$

where $\rho > 0$ is the penalty parameter. The corresponding augmented Lagrangian (in scaled form) is

$$
\mathscr{L}(\beta, \Theta, Z_\Theta, U_\Theta) = \frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it}(y_{it} - \Theta_{it} - x_{it}'\beta)^2 + \nu_{NT} \left\| Z_\Theta \right\|_* + \frac{\rho}{2} \left\| Z_\Theta - \Theta + U_\Theta \right\|_F^2 - \frac{\rho}{2} \left\| U_\Theta \right\|_F^2.
$$

The detailed optimization algorithm is as below.

Algorithm 3

1. *Set up the initial value:*

$$
\Theta_{it}^{[0]} = (d_{it} + \rho)^{-1} d_{it} y_{it}, \quad Z_\Theta^{[0]} = \Theta^{[0]}, \quad \beta^{[0]} = 0, \quad U_\Theta^{[0]} = 0.
$$

2. Given $\beta^{[k]}$, $Z_{\Theta}^{[k]}$, and $U_{\Theta}^{[k]}$, we notice that

$$\Theta^{[k+1]} = \min_{\Theta} \frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it}(y_{it} - \Theta_{it} - x_{it}'\beta^{[k]})^2 + \frac{\rho}{2} \left\| Z_{\Theta}^{[k]} - \Theta + U_{\Theta}^{[k]} \right\|_F^2,$$

which produces $\Theta_{it}^{[k+1]} = [d_{it}(y_{it} - x_{it}'\beta) + \rho(Z_{\Theta,it} + U_{\Theta,it})] (d_{it} + \rho)^{-1}$.

3. Given $\Theta^{[k+1]}$, we notice that

$$Z_{\Theta}^{[k+1]} = \min_{Z_{\Theta}} \nu_{NT} \left\| Z_{\Theta} \right\|_* + \frac{\rho}{2} \left\| Z_{\Theta} - \Theta^{[k+1]} + U_{\Theta}^{[k]} \right\|_F^2.$$

By the SVD, we have $\Theta^{[k+1]} - U_{\Theta}^{[k]} = P^{[k+1]} D^{[k+1]} Q^{[k+1]\prime}$. Let $D_{\nu_{NT}}^{[k+1]} = \max(D^{[k+1]} - \frac{\nu_{NT}}{\rho}, 0)$, then we update $Z_{\Theta}^{[k+1]} = P^{[k+1]} D_{\nu_{NT}}^{[k+1]} Q^{[k+1]\prime}$. Moreover, we update

$$\beta^{[k+1]} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} d_{it} x_{it} x_{it}' \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it} x_{it} \left(y_{it} - \Theta_{it}^{[k+1]} \right), \text{ and}$$
$$U_{\Theta}^{[k+1]} = Z_{\Theta}^{[k+1]} - \Theta^{[k+1]} + U_{\Theta}^{[k]}.$$

4. Iterate Steps 2-3 until convergence and the final result produce the initial estimates $\hat{\beta}^{(0)}$ and $\tilde{\Theta}^{(0)}$.

In practice, to choose the tuning parameter ρ, we can simply fix it as $\rho = 1$ or $\rho = 10$. Otherwise, we can do the adjustment in each iteration. In our simulation, we start with the default choice $\rho = 1$ and adopt the common residual balancing strategy to update rho in each iteration; see Section 3.4 in Boyd et al. (2011). While a fixed ρ can be used for simpler problems, residual balancing adaptively maintains balance between residuals for faster and more stable convergence. Specifically, define $PR = \left\| Z_{\Theta}^{[k+1]} - \Theta^{[k+1]} \right\|_F$ and $DR = \left\| Z_{\Theta}^{[k+1]} - Z_{\Theta}^{[k]} \right\|_F$. Then we adjust $\rho^{[k+1]}$ as

$$\rho^{[k+1]} = \begin{cases} \rho^{[k]} \tau_{incr}, & \text{if } PR > \mu DR, \\ \rho^{[k]} / \tau_{decr}, & \text{if } DR > \mu PR, \\ \rho^{[k]}, & \text{otherwise.} \end{cases}$$

If the primal residual (PR) is much larger than the dual residual (DR), it indicates that the constraint $\Theta = Z_\theta$ is poorly enforced, and we increase ρ. If the dual residual is much larger than the primal residual, it indicates instability in the dual updates, and we decrease ρ. Otherwise, ρ remains unchanged. τ_{incr}, τ_{decr}, and μ are all adjustment parameters. In the simulation, we set $\tau_{incr} = \tau_{decr} = 2$, and $\mu = 10$.