Article:

Li, Degui orcid.org/0000-0001-6802-308X, Qian, Junhui and Su, Liangjun (2017) Panel Data Models with Interactive Fixed Effects and Multiple Structural Breaks. Journal of the American Statistical Association. pp. 1804-1819. ISSN: 0162-1459

https://doi.org/10.1080/01621459.2015.1119696

Panel Data Models with Interactive Fixed Effects and Multiple Structural Breaks*

Degui Li[†] Junhui Qian[‡] Liangjun Su[§]

University of York, Shanghai Jiao Tong University, Singapore Management University

October 7, 2015

Abstract

In this paper we consider estimation of common structural breaks in panel data models with interactive fixed effects which are unobservable. We introduce a penalized principal component (PPC) estimation procedure with an adaptive group fused LASSO to detect the multiple structural breaks in the models. Under some mild conditions, we show that with probability approaching one the proposed method can correctly determine the unknown number of breaks and consistently estimate the common break dates. Furthermore, we estimate the regression coefficients through the post-LASSO method and establish the asymptotic distribution theory for the resulting estimators. The developed methodology and theory are applicable to the case of dynamic panel data models. The Monte Carlo simulation results demonstrate that the proposed method works well in finite samples with low false detection probability when there is no structural break and high probability of correctly estimating the break numbers when the structural breaks exist. We finally apply our method to study the environmental Kuznets curve for 74 countries over 40 years and detect two breaks in the data.

*The authors would like to thank the editor, an associate editor and two referees for the helpful and insightful comments which greatly improve an earlier version of the paper.

[†]Department of Mathematics, University of York, Heslington, York, YO10 5DD, United Kingdom. Email: degui.li@york.ac.uk.

[‡]Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200052, China. Email: jhqian@sjtu.edu.cn.

[§]School of Economics, Singapore Management University, Singapore 178903, Singapore. Email: ljsu@smu.edu.sg.

Keywords: Change point; Interactive fixed effects; LASSO; Panel data; Penalized estimation; Principal component analysis.

1 Introduction

As the availability of panel or longitudinal data increases in the last few decades, panel data studies have become increasingly popular among a wide group of statisticians and econometricians. Analysis of panel data sets has various advantages over that of purely time series or cross-sectional data sets. A relatively less exploited advantage of the panel data is that it provides researchers with more flexibility to model cross-sectional dependence over individual units and uncover possible structural changes over time. Structural breaks are, indeed, quite common in many areas such as economics and finance, and may occur for various reasons. For example, the celebrated environmental Kuznets curve may shift as a result of a growing public awareness of environmental issues, a technological breakthrough, or an international coordination and cooperation on environmental protection. If such structural changes are ignored in the modelling, subsequent statistical analyses may lead to incorrect inferences or misleading predictions.

In recent years, there has been a growing literature on the estimation and test of structural breaks in panel data models. Generally speaking, most of the existing literature falls into two categories depending on whether the parameters of interest are allowed to be heterogenous across subjects or not. The first category focuses on homogenous panel data models (c.f., De Watcher and Tzavalis, 2012; and Qian and Su, 2015b) and the second category considers estimation and inference of common breaks in heterogenous panel data models (c.f., Bai, 2010; Kim, 2011; Baltagi et al., 2015). Despite the vast literature on multiple structural breaks in the time series framework (c.f., Csörgö and Horváth, 1997; Bai and Perron, 1998; Qu and Perron, 2007; Harchaoui and Lèvy-Leduc, 2010; Chan et al., 2014; Qian and Su, 2015a), most of the existing work on panel structural breaks focuses on the estimation and inference of a single structural break in panel data models. The only exception is the paper by Qian and Su (2015b) which considers shrinkage estimation of common breaks in panel data models. However, Qian and Su's (2015b) modelling framework does not allow the existence of cross-sectional dependence, which limits the applicability of their techniques as cross-sectional dependence commonly exists in many panel data sets nowadays (such as the panel climate and environmental data).

In this paper, we aim to estimate multiple structural breaks in panel data models with cross-sectional dependence which is described through the unobservable interactive fixed effects.

Such a cross-sectional dependence structure has received increasing interest in the analysis of panel data in recent years; see, for example, Pesaran (2006), Bai (2009), Bai and Li (2014), and Moon and Weidner (2014, 2015). However, to the best of our knowledge, there is virtually no work on estimating multiple structural breaks in panel data models with interactive fixed effects and possible dynamic structure (such as the dynamic autoregressive panel data models). As in Qian and Su (2015b), we apply the shrinkage idea through the adaptive group fused LASSO (AGF-LASSO) to estimate the multiple structural break dates. Nevertheless, the existence of the unobservable interactive fixed effects in our model makes the estimation techniques and the development of the asymptotic theory much more involved than those in Qian and Su (2015b). In Section 2 below, we introduce a novel penalized principal component (PPC) estimation procedure via AGF-LASSO to estimate both the regression coefficients and the factor loadings. Similar to the sparsity result in the high-dimensional variable selection literature (c.f., Fan and Li, 2001, 2006), we establish the consistency for the detection of multiple structural breaks, which indicates that both the number of breaks and the break dates can be consistently estimated. Furthermore, we also estimate the regression coefficients through the post-LASSO method and then establish the asymptotic distribution theory of the resulting estimators, which generalizes the results in Bai (2009) and Moon and Weidner (2014) where there is no structural break. The simulation studies show that the proposed PPC method has a high probability of correctly estimating the number of breaks when the structural breaks exist in panel data models, and a low probability of false detection when there is no structural break. Furthermore, we study the environmental Kuznets curve for 74 countries over 40 years by using our method and find that there exist two structural breaks in the data.

The rest of the paper is organized as follows. Section 2 introduces the model and the PPC estimation method. Section 3 gives the asymptotic properties for the PPC estimator as well as the post-LASSO estimator. Section 4 discusses the determination of the number of the factors and the choice of the tuning parameter in the PPC estimation procedure and reports the Monte Carlo simulation results. Section 5 gives the empirical application of the proposed model and method. Section 6 concludes the paper. Appendices A and B give the assumptions and the proofs of the asymptotic results, respectively. Some technical lemmas as well as their proofs are collected in Appendix C of the supplemental document.

Notation. For an $m \times n$ real matrix $\boldsymbol{A}$ we denote its transpose as $\boldsymbol{A}'$, its Frobenius norm as $\|\boldsymbol{A}\|$ $(\equiv [\mathrm{tr}(\boldsymbol{A}\boldsymbol{A}')]^{1/2})$, its spectral norm as $\|\boldsymbol{A}\|_{\mathrm{sp}}$ $(\equiv [\mu_{\max}(\boldsymbol{A}\boldsymbol{A}')]^{1/2})$, and its Moore-Penrose generalized inverse as $\boldsymbol{A}^+$, where $\mu_{\max}(\cdot)$ denotes the maximum eigenvalue of a square matrix. Let $\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^+\boldsymbol{A}'$ and $\boldsymbol{M}_{\boldsymbol{A}} = \boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{A}}$, where $\boldsymbol{I}_m$ is an $m \times m$ identity matrix. When $\boldsymbol{A}$ is symmetric with $m = n$, we use $\mu_r(\boldsymbol{A})$ to denote its rth largest eigenvalue by counting multiple eigenvalues multiple times, and $\mu_{\max}(\boldsymbol{A})$ and $\mu_{\min}(\boldsymbol{A})$ to denote the largest and smallest eigenvalues of $\boldsymbol{A}$, respectively. Let $\mathsf{vec}(\boldsymbol{A})$ be the vectorization of $\boldsymbol{A}$ and $\mathsf{Tr}(\boldsymbol{A})$ the trace of a square matrix $\boldsymbol{A}$. Let $\mathbf{0}$ denote a null matrix or vector whose size may change from line to line, and $\mathbf{1}\{\cdot\}$ be the usual indicator function. The operator $\stackrel{P}{\to}$ denotes convergence in probability, $\stackrel{D}{\to}$ convergence in distribution, and plim probability limit. We use $(N,T) \to \infty$ to denote that both N and T pass to infinity jointly.

2 Model and estimation

In this section, we first introduce a panel data model with interactive fixed effects and an unknown number of structural breaks, and then propose the PPC estimation method.

2.1 The model

Let Y_{it} be the dependent variable for subject i measured at time t where $i = 1, ..., N$, and $t = 1, ..., T$. We consider the following panel data model with interactive fixed effects

$$Y_{it} = \beta_t' X_{it} + \lambda_i' f_t + \varepsilon_{it}, \quad i = 1, ..., N, \; t = 1, ..., T, \tag{2.1}$$

where X_{it} is a $p \times 1$ vector of explanatory variables, β_t is a $p \times 1$ vector of unknown slope coefficients which may change over time, λ_i and f_t denote an $R_0 \times 1$ vector of unobservable factor loadings and common factors, respectively, both of which may be correlated with X_{it}, and ε_{it} is the idiosyncratic error term. The dimension of the unknown coefficient vector, $p \equiv p_{NT}$, is allowed to be diverging as $(N,T) \to \infty$, and the dimension of the vectors for the factor loadings and common factors, R_0, is a fixed positive integer. Throughout the paper, we denote the true value of a parameter vector with a superscript 0. For instance, β_t^0, λ_i^0 and f_t^0 denote the true values of β_t, λ_i and f_t, respectively. We allow the regression coefficients to vary over the time and model (2.1) thus includes the classical linear panel data models with interactive fixed effects (c.f., Pesaran, 2006; Bai, 2009; Moon and Weidner, 2015) as a special case. As in these papers, we assume that both the cross-sectional size N and the time series length T pass to infinity, which is called as "large dimensional panel" in the literature.

In this paper we assume that the true regression coefficients $\{\beta_1^0, ..., \beta_T^0\}$ exhibit certain *sparse* nature such that the total number of distinct vectors in the set is given by m^0+1, which is unknown but typically much smaller than the time series length T. We allow $m^0 \equiv m_T^0$ to be divergent at an appropriate rate as $T \to \infty$. More specifically, we let

$$\beta_t^0 = \alpha_j^0 \text{ for } t = T_{j-1}^0, ..., T_j^0 - 1 \text{ with } j = 1, ..., m^0+1,$$

where we adopt the convention that $T_0^0 = 1$ and $T_{m^0+1}^0 = T+1$. The indices T_j^0, $j = 1, ..., m^0$, indicate that there are m^0 unobserved break points/dates and the number m^0+1 denotes the total number of regimes. We are interested in estimating the *unknown* number of structural breaks, the *unobservable* break dates, and the regression coefficients in different regimes. Let $\boldsymbol{\beta} = (\beta_1', ..., \beta_T')'$, $\boldsymbol{\alpha}_m = (\alpha_1', ..., \alpha_{m+1}')'$, $\boldsymbol{\Lambda} = (\lambda_1, \lambda_2, ..., \lambda_N)'$, $\boldsymbol{F} = (f_1, f_2, ..., f_T)'$, and $\mathcal{T}_m = (T_1, ..., T_m)$. Throughout the paper, we use m^0, $\boldsymbol{\alpha}_{m^0}^0 = (\alpha_1^{0\prime}, ..., \alpha_{m^0+1}^{0\prime})'$ and $\mathcal{T}_{m^0}^0 = (T_1^0, ..., T_{m^0}^0)$ to denote the true number of structural breaks, the true vector of distinct regression coefficients, and the set of true break dates, respectively.

2.2 PPC estimation

We consider the PPC estimation of the unknown components $(\boldsymbol{\beta}^0, \boldsymbol{\Lambda}^0, \boldsymbol{F}^0)$, the true values of $(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{F})$. Let $Y_t = (Y_{1t}, ..., Y_{Nt})'$ and $X_t = (X_{1t}, ..., X_{Nt})'$. In order to apply the PPC method, we define the objective function through

$$\tilde{Q}_{NT,\gamma}(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{F}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Y_{it} - X_{it}'\beta_t - \lambda_i' f_t\right)^2 + \frac{\gamma}{T}\sum_{t=2}^{T}\dot{w}_t \left\|\beta_t - \beta_{t-1}\right\|, \tag{2.2}$$

which can be written as

$$\frac{1}{NT}\sum_{t=1}^{T}\left(Y_t - X_t\beta_t - \boldsymbol{\Lambda} f_t\right)'\left(Y_t - X_t\beta_t - \boldsymbol{\Lambda} f_t\right) + \frac{\gamma}{T}\sum_{t=2}^{T}\dot{w}_t \left\|\beta_t - \beta_{t-1}\right\|,$$

where $\gamma \equiv \gamma_{NT} > 0$ is a tuning parameter and $\dot{w}_t$ is a data-driven weight defined by

$$\dot{w}_t = \left\|\dot{\beta}_t - \dot{\beta}_{t-1}\right\|^{-\kappa}, \quad t = 2, ..., T, \tag{2.3}$$

$\dot{\beta}_t$, $t = 1, ..., T$, are the preliminary estimates of the regression coefficients β_t, and κ is a user-specified positive constant that usually takes value 2 in the literature. In this paper, the preliminary estimation $\{\dot{\beta}_t, t = 1, ..., T\}$ is constructed to minimize the first term of the objective function in (2.2) by ignoring the penalization device.

By concentrating $\boldsymbol{F}$ out in the first term of the objective function (2.2), we can readily obtain the following objective function

$$\hat{Q}_{NT,\gamma}\left(\boldsymbol{\beta},\boldsymbol{\Lambda}\right)=\hat{Q}_{NT}\left(\boldsymbol{\beta},\boldsymbol{\Lambda}\right)+\frac{\gamma}{T}\sum_{t=2}^{T}\dot{w}_t\left\|\beta_t-\beta_{t-1}\right\|, \tag{2.4}$$

where

$$\hat{Q}_{NT}\left(\boldsymbol{\beta},\Lambda\right)=\frac{1}{NT}\sum_{t=1}^{T}\left(Y_t-X_t\beta_t\right)'\boldsymbol{M}_{\boldsymbol{\Lambda}}\left(Y_t-X_t\beta_t\right).$$

Following Moon and Weidner (2014), we can further concentrate $\boldsymbol{\Lambda}$ out in (2.4) and obtain the objective function

$$\bar{Q}_{NT,\gamma}\left(\boldsymbol{\beta}\right)=\bar{Q}_{NT}\left(\boldsymbol{\beta}\right)+\frac{\gamma}{T}\sum_{t=2}^{T}\dot{w}_t\left\|\beta_t-\beta_{t-1}\right\|, \tag{2.5}$$

where

$$\bar{Q}_{NT}\left(\boldsymbol{\beta}\right)=\frac{1}{N}\sum_{r=R_0+1}^{N}\mu_r\left[\frac{1}{T}\sum_{t=1}^{T}\left(Y_t-X_t\beta_t\right)\left(Y_t-X_t\beta_t\right)'\right]. \tag{2.6}$$

It can be seen that the penalization device in the above objective functions is closely related to the literature on the adaptive LASSO (Zou, 2006), the group LASSO (Yuan and Lin, 2006), and the fused LASSO (Tibshirani et al., 2005; Rinaldo, 2009). The use of the Frobenius norm $\|\cdot\|$ for the vector difference $\beta_t-\beta_{t-1}$ generalizes the fused LASSO to the group fused LASSO; and the use of the weights $\{\dot{w}_t\}$ makes the LASSO procedure adaptive. Therefore, we can call our penalized estimation procedure as an *adaptive group fused LASSO* (AGF-LASSO) procedure.

Following Bai and Ng's (2002) principal component method under the identification restrictions that $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}/N=\boldsymbol{I}_{R_0}$ and $\boldsymbol{F}'\boldsymbol{F}$ is a diagonal matrix, the minimizers to the objective function defined in (2.4), $\hat{\boldsymbol{\beta}}=\left(\hat{\beta}_1',...,\hat{\beta}_T'\right)'$ and $\hat{\boldsymbol{\Lambda}}$ satisfy that

$$\hat{\boldsymbol{\beta}}=\arg\min_{\boldsymbol{\beta}}\hat{Q}_{NT,\gamma}(\boldsymbol{\beta},\hat{\boldsymbol{\Lambda}}), \tag{2.7}$$

and

$$\left[\frac{1}{NT}\sum_{t=1}^{T}\left(Y_t-X_t\hat{\beta}_t\right)\left(Y_t-X_t\hat{\beta}_t\right)'\right]\hat{\boldsymbol{\Lambda}}=\hat{\boldsymbol{\Lambda}}\boldsymbol{V}_{NT} \tag{2.8}$$

where $\boldsymbol{V}_{NT}$ is a diagonal matrix consisting of the R_0 largest eigenvalues of the matrix in the square brackets in (2.8) arranged in descending order. Furthermore, the common factor $\boldsymbol{F}^0$ can be estimated by

$$\hat{\boldsymbol{F}}=(\hat{f}_1,\hat{f}_2,...,\hat{f}_T)' \quad \text{with} \quad \hat{f}_t=N^{-1}\hat{\boldsymbol{\Lambda}}'(Y_t-X_t\hat{\beta}_t). \tag{2.9}$$

An iterative algorithm based on (2.7) and (2.8) can be implemented in practice to estimate $\boldsymbol{\beta}^0$ and $\boldsymbol{\Lambda}^0$. Note that the above calculations are different from those in the existing literature such as Bai (2009) and Lu and Su (2015) by switching the role of $\boldsymbol{\Lambda}$ and $\boldsymbol{F}$, because the regression coefficients are heterogeneous over time.

With the estimated regression coefficients $\hat{\beta}_t$, the set of estimated break dates are given by $\hat{\mathcal{T}}_{\hat{m}} = (\hat{T}_1, ..., \hat{T}_{\hat{m}})$ where $2 \le \hat{T}_1 < ... < \hat{T}_{\hat{m}} \le T$ such that $\|\hat{\beta}_t - \hat{\beta}_{t-1}\| \ne 0$ at $t = \hat{T}_j$ for $j = 1, ..., \hat{m}$. The set $\hat{\mathcal{T}}_{\hat{m}}$ divides the time interval $[1, T]$ into $\hat{m}+1$ regimes such that the parameter estimates remain constant within each regime. Notice that if $\hat{T}_{\hat{m}} = T$, the last break occurs at the end of the sample and the $(\hat{m}+1)$th regime has only one time series observation for each cross-sectional unit. Let $\hat{T}_0 = 1$ and $\hat{T}_{\hat{m}+1} = T+1$. Define $\hat{\alpha}_j = \hat{\alpha}_j(\hat{\mathcal{T}}_{\hat{m}}) = \hat{\beta}_{\hat{T}_{j-1}}$ as the estimate of α_j^0 for $j = 1, ..., \hat{m}+1$. In the sequel, we usually suppress the dependence of $\hat{\alpha}_j$ on $\hat{\mathcal{T}}_{\hat{m}}$ (or the tuning parameter γ) unless necessary. For example, we let $\hat{\boldsymbol{\alpha}}_{\hat{m}} = \left(\hat{\alpha}_1', \hat{\alpha}_2', ..., \hat{\alpha}_{\hat{m}+1}'\right)'$ which denotes $\hat{\boldsymbol{\alpha}}_{\hat{m}}(\hat{\mathcal{T}}_{\hat{m}}) = \left[\hat{\alpha}_1(\hat{\mathcal{T}}_{\hat{m}})', \hat{\alpha}_2(\hat{\mathcal{T}}_{\hat{m}})', ..., \hat{\alpha}_{\hat{m}+1}(\hat{\mathcal{T}}_{\hat{m}})'\right]'$.

3 Asymptotic properties

In this section, we give the large sample theory including the consistency of the proposed PPC estimator and the limiting distribution of the post-LASSO estimator.

3.1 Consistency of the PPC estimator

We start with the consistency result of the PPC estimator $\hat{\boldsymbol{\beta}}$ with preliminary convergence rates.

Theorem 3.1 *Suppose that Assumptions 1 and 2(i)(ii) in Appendix A holds. Then we have (i)* $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|^2/T = O_P\left(p/N + 1/T\right) = O_P\left(\delta_{p,NT}^{-2}\right)$, *and (ii)* $\left\|\hat{\beta}_t - \beta_t^0\right\| = O_P\left(\delta_{p,NT}^{-1}\right)$, *where* $\delta_{p,NT} = \min(\sqrt{N/p}, \sqrt{T})$.

Theorems 3.1 (i) and (ii) establish the *preliminary* mean square and point-wise convergence rates of $\{\hat{\beta}_t\}$, respectively, which is a very general result by allowing the existence of multiple jumps or drops in the regression coefficients. As we allow the regression coefficients vary over time, there is less observational information available for the estimation of each regression coefficient (compared with the model without any structural break). This would in turn affect the estimation accuracy of the factor loading matrix and convergence rates for the parameter estimators. The divergent dimension of the regression coefficients at each time point further slows down the convergence rates. It is easy to find that the total number of the unknown elements in the set $\{\beta_t^0\}$ is pT. Hence, it is not surprising that in Theorem 3.1 we can only obtain the $O_P(\delta_{p,NT}^{-1})$ convergence rate for the PPC estimator $\hat{\beta}_t$, which is much slower than the optimal root-(NT) rate obtained by Bai (2009) and Moon and Weidner (2014) (after bias correction) when there is no change point for the regression coefficients and the dimension of the regression coefficients is fixed.

Recall that $\mathcal{T}_{m^0}^0 = \{T_1^0, ..., T_{m^0}^0\}$ denotes the set of true break dates. Let $\mathcal{T}^c = \{2, ..., T\} \backslash \mathcal{T}_{m^0}^0$. Let $\theta_1^0 = \beta_1^0$, $\hat{\theta}_1 = \hat{\beta}_1$, $\theta_t^0 = \beta_t^0 - \beta_{t-1}^0$ and $\hat{\theta}_t = \hat{\beta}_t - \hat{\beta}_{t-1}$ for $t = 2, ..., T$. The following theorem establishes the detection consistency, which, in some sense, is analogous to the sparsity result in the high-dimensional variable selection literature.

Theorem 3.2 *Suppose that Assumptions 1 and 2 in Appendix A hold. Then*

$$\lim_{(N,T)\to\infty} \mathsf{P}\left(\left\|\hat{\theta}_t\right\| = 0 \textit{ for all } t \in \mathcal{T}^c\right) = 1.$$

Theorem 3.2 shows that with probability approaching one (w.p.a.1), all the zero vectors in $\{\theta_t^0\}$ must be estimated as exactly zero, which is a well-known sparsity result in the high-dimensional variable selection literature (c.f., Fan and Li, 2006). On the other hand, by Theorem 3.1(ii), we know that the estimators of the nonzero vectors in $\{\theta_t^0\}$ are consistent by noting that $\hat{\beta}_t - \hat{\beta}_{t-1}$ consistently estimates $\theta_t^0 = \beta_t^0 - \beta_{t-1}^0$. A combination of Theorems 3.1 and 3.2 implies that the AGF-LASSO penalty has the ability to identify the true regression model with the correct number of structural breaks and the correct break dates, which is stated in the following corollary.

Corollary 3.3 *Suppose that Assumptions 1 and 2 in Appendix A hold. Then (i)*

$$\lim_{(N,T)\to\infty} \mathsf{P}\left(\hat{m} = m^0\right) = 1,$$

and (ii)

$$\lim_{(N,T)\to\infty} \mathsf{P}(\hat{T}_1 = T_1^0, ..., \hat{T}_{m^0} = T_{m^0}^0) = 1.$$

3.2 Post-LASSO estimation

We next introduce the post-LASSO estimation of the regression coefficients, which can improve the convergence rate of the PPC estimation given in Theorem 3.1. For any $p(m+1)$-dimensional vector $\boldsymbol{\alpha}_m = \left(\alpha_1', ..., \alpha_{m+1}'\right)'$ and $\mathcal{T}_m = \{T_1, ..., T_m\}$ with $1 < T_1 < ... < T_m \leq T$, we define the objective function by

$$\begin{aligned} Q_{NT}\left(\boldsymbol{\alpha}_m, \boldsymbol{\Lambda}, \boldsymbol{F}; \mathcal{T}_m\right) &= \frac{1}{NT}\sum_{j=1}^{m+1}\sum_{t=T_{j-1}}^{T_j-1}\sum_{i=1}^{N}\left(Y_{it} - X_{it}'\alpha_j - \lambda_i' f_t\right)^2 \\ &= \frac{1}{NT}\sum_{j=1}^{m+1}\sum_{t=T_{j-1}}^{T_j-1}\left(Y_t - X_t\alpha_j - \boldsymbol{\Lambda} f_t\right)'\left(Y_t - X_t\alpha_j - \boldsymbol{\Lambda} f_t\right). \end{aligned} \tag{3.1}$$

By concentrating $\boldsymbol{F}$ out in the above objective function, we readily obtain the following post-LASSO objective function

$$Q_{NT}\left(\boldsymbol{\alpha}, \boldsymbol{\Lambda}; \mathcal{T}_m\right) = \frac{1}{NT}\sum_{j=1}^{m+1}\sum_{t=T_{j-1}}^{T_j-1}\left(Y_t - X_t\alpha_j\right)'\boldsymbol{M}_{\boldsymbol{\Lambda}}\left(Y_t - X_t\alpha_j\right). \tag{3.2}$$

Let $\tilde{\boldsymbol{\alpha}}_m\left(\mathcal{T}_m\right) = \left[\tilde{\alpha}_1(\mathcal{T}_m)', ..., \tilde{\alpha}_{m+1}(\mathcal{T}_m)'\right]'$ and $\tilde{\boldsymbol{\Lambda}}\left(\mathcal{T}_m\right) = \left[\tilde{\lambda}_1(\mathcal{T}_m), ..., \tilde{\lambda}_N(\mathcal{T}_m)\right]'$ denote the minimizers of the objective function defined in (3.2) for given $\mathcal{T}_m$. By setting $\mathcal{T}_m$ as $\hat{\mathcal{T}}_{\hat{m}} = (\hat{T}_1, ..., \hat{T}_{\hat{m}})$, the set of the estimated break dates constructed in Section 2.2, we obtain the post-LASSO estimators $\tilde{\boldsymbol{\alpha}}_{\hat{m}} \equiv \tilde{\boldsymbol{\alpha}}_{\hat{m}}(\hat{\mathcal{T}}_{\hat{m}})$ and $\tilde{\boldsymbol{\Lambda}} \equiv \tilde{\boldsymbol{\Lambda}}(\hat{\mathcal{T}}_{\hat{m}})$.

We next study the asymptotic distribution of the post-LASSO estimators. Corollary 3.3 above implies that w.p.a.1 $\hat{m} = m^0$ and $\hat{T}_j = T_j^0$ for $j = 1, ..., m^0$. Hence, it follows that $\tilde{\alpha}_{\hat{m}}$ is asymptotically equivalent to the infeasible estimator $\tilde{\boldsymbol{\alpha}}_{m^0}(\mathcal{T}_{m^0})$ which is obtained only if one knows the set $\mathcal{T}_{m^0}^0$ of the true break dates. Let $\tau_j(T) = T_j^0 - T_{j-1}^0$,

$$\begin{aligned} \boldsymbol{B}_{NT}(1) &= \left[B_{NT,1}(1)', ..., B_{NT,m^0+1}(1)'\right]' \text{ and} \\ \boldsymbol{B}_{NT}(2) &= \left[B_{NT,1}(2,1)' - B_{NT,1}(2,2)', ..., B_{NT,m^0+1}(2,1)' - B_{NT,m^0+1}(2,2)'\right]', \end{aligned}$$

where for $j = 1, ..., m^0 + 1$

$$B_{NT,j}(1) = \frac{1}{N^2 T\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_{m^0}}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\tilde{\boldsymbol{\Lambda}}_{m^0}\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}_{m^0}\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0,$$

$$B_{NT,j}(2,1) = \frac{1}{N\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{\Lambda}^0\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+\left(\frac{1}{NT}\sum_{s=1}^{T} f_s^0\varepsilon_s'\varepsilon_t\right),$$

$$B_{NT,j}(2,2) = \frac{1}{N\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{\Lambda}^0\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+\left(\frac{1}{NT}\sum_{s=1}^{T} f_t^0\varepsilon_s'\varepsilon_t^*\right),$$

and $\varepsilon_t^* = \frac{1}{T}\sum_{s=1}^{T}\chi_{st}\varepsilon_s$ with $\chi_{st} = f_s^{0\prime}\big(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\big)^+ f_t^0$, $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_T)$ with $\varepsilon_t = \big(\varepsilon_{1t}, ..., \varepsilon_{Nt}\big)'$. We then define

$$\boldsymbol{B}_{NT} = \boldsymbol{\Omega}_{NT}^+\big[\boldsymbol{B}_{NT}(1) + \boldsymbol{B}_{NT}(2) - \boldsymbol{B}_{NT}(3)\big],$$

where $\boldsymbol{\Omega}_{NT}$ and $\boldsymbol{B}_{NT}(3)$ are defined in Appendix A. Let

$$\mathbf{D}_{NT} = \mathsf{diag}\Big\{\sqrt{N\tau_1(T)}, ..., \sqrt{N\tau_{m^0+1}(T)}\Big\} \otimes \mathbf{I}_p,$$

where $\otimes$ denotes the Kronecker product, and $\mathbf{S}$ be a $k^0 \times p(m^0+1)$ matrix with full row rank and k^0 being a fixed positive integer.

Theorem 3.4 *Suppose that Assumptions 1–3 in Appendix A hold. Then conditional on $\hat{m} = m^0$, we have*

$$\mathbf{S}\mathbf{D}_{NT}\big(\tilde{\boldsymbol{\alpha}}_{\hat{m}} - \boldsymbol{\alpha}^0 + \boldsymbol{B}_{NT}\big) \xrightarrow{D} \mathsf{N}\big(\mathbf{0},\ \mathbf{S}\boldsymbol{\Omega}_0^+\boldsymbol{\Omega}_1\boldsymbol{\Omega}_0^+\mathbf{S}'\big),$$

where $\boldsymbol{\Omega}_0$ and $\boldsymbol{\Omega}_1$ are defined in Assumption 3 in Appendix A.

Despite the use of different notations and proof strategies, $\boldsymbol{\Omega}_{NT}^+\boldsymbol{B}_{NT}(1)$ and $\boldsymbol{\Omega}_{NT}^+\boldsymbol{B}_{NT}(2)$ correspond to the terms $-C$ and $-B$ in Bai (2009) or $-W^{-1}B_3$ and $-W^{-1}B_2$ in Moon and Weidner (2014), respectively. However, these two papers assume that the dimension p is fixed and there is no structural break on the regression coefficients. Hence, our asymptotic distribution theory is derived under a more general framework. Like the term $-W^{-1}B_1$ in Moon and Weidner (2014), $\boldsymbol{\Omega}_{NT}^+\boldsymbol{B}_{NT}(3)$ arises here because we allow the regressor vector X_{it} to contain lagged dependent variable (e.g., $Y_{i,t-1}$) and it is vanishing under Bai's (2009) conditions A-E that include the independence between ε_{it} and $(X_{js}, \lambda_j^0, f_s^0)$ for all i, t, j, s and thus rule out dynamics in the regression equation. As Bai (2009) remarks, in the absence of both serial/cross-sectional correlations and heteroskedasticity and under his Assumption D, all of these three bias terms are asymptotically negligible. In the general case, the bias terms of the post-LASSO estimates can be removed by constructing a bias-corrected estimate. Following Bai (2009) in the case of static panels or Moon and Weidner (2014) in the case of dynamic panels, one can easily construct a bias corrected version of our post-LASSO estimate. We omit the details as the extension is quite straightforward.

Note that the above theorem holds without requiring that N and T diverge to infinity at the same speed and the latter condition was assumed in both Bai (2009) and Moon and Weidner (2014). For the easiness of presentation, we need to assume that $\tau_j(T) = T_j^0 - T_{j-1}^0 \propto T/m^0$ in Assumption 3(ii) in Appendix A, which implies that each regime specific regression coefficient vector α_j^0 can be estimated at the same convergence rate $O_P(\sqrt{pm^0/(NT)})$ after possible bias correction. Apparently, it is possible to weaken this last assumption to $T_j^0 - T_{j-1}^0 \to \infty$ and then we can anticipate that $\tilde{\alpha}_j(\mathcal{T}_m)$'s would have different convergence rates to their true values across different regimes.

4 Practical issues in model estimation and simulation study

In this section we first discuss the determination of the number of factors and the choice of the tuning parameter γ in the PPC estimation procedure, then introduce the algorithm to implement the estimation method, and finally conduct a set of Monte Carlo experiments to evaluate the finite sample performance of the proposed method.

4.1 Determination of the number of factors

In the above analysis we assume that the number of factors R_0 is known. In practice, one has to determine it from data. Here we use R to denote a generic number of factors and assume that it is bounded from above by a finite integer $R_{\max} \geq R_0$. We propose a BIC-type information criterion to determine R_0 before embarking on the AGF-LASSO procedure.

Let $\dot{\beta}_{t,R}$, $\dot{f}_{t,R}$ and $\dot{\lambda}_{i,R}$ denote the PCA estimators (without the penalization device) of β_t, $f_{t,R}$ and $\lambda_{i,R}$ by assuming R factors in the model using the normalization rule: $\mathbf{\Lambda}_R'\mathbf{\Lambda}_R/N = \boldsymbol{I}_R$ and $\boldsymbol{F}_R'\boldsymbol{F}_R$ is a diagonal. Note that we have made the dependence of the parameters and their estimators on R explicitly here. Let $\dot{\boldsymbol{\beta}}_R = \left(\dot{\beta}_{1,R}', ..., \dot{\beta}_{T,R}'\right)'$. Define

$$V(R, \dot{\boldsymbol{\beta}}_R) = \frac{1}{N}\sum_{r=R+1}^{N} \mu_r \left[\frac{1}{T}\sum_{t=1}^{T}\left(Y_t - X_t\dot{\beta}_{t,R}\right)\left(Y_t - X_t\dot{\beta}_{t,R}\right)'\right].$$

Following Bai and Ng (2002), we consider the BIC-type information criterion defined by

$$\mathsf{BIC}(R) = \ln V(R, \dot{\boldsymbol{\beta}}_R) + \rho_1 R, \tag{4.1}$$

where $\rho_1 \equiv \rho_{1,NT}$ is pre-determined which plays the role of $\ln(NT)/(NT)$ in the case of the conventional BIC criterion. Let $\hat{R} = \arg\min_{0 \leq R \leq R_{\max}} \mathsf{BIC}(R)$, which estimates the number of the factors.

Theorem 4.1 *Suppose that Assumptions 1–4 in Appendix A hold. Then*

$$\mathsf{P}\left(\hat{R} = R_0\right) \to 1 \text{ as } (N, T) \to \infty.$$

The above theorem shows that the use of $\mathsf{BIC}(R)$ can consistently estimate R_0. To implement the above information criterion, one needs to choose the penalty coefficient ρ_1. Following Bai and Ng (2002), we can set

$$\rho_1 = \frac{(N+T)\,p}{NT} \ln\left(\frac{NT}{N+T}\right) \quad \text{or} \quad \rho_1 = \frac{(N+T)\,p}{NT} \ln\left(\delta_{NT}^2\right)$$

where $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$ is defined as in Section 3. The penalty coefficient in Bai and Ng (2002) corresponds to $p = 1$ in the above definitions of ρ_1. In our simulations we use the first specification of ρ_1 and search for $\hat{R}$ in the range of $\{1, 2, \ldots, 5\}$ when $R_0 = 2$.

4.2 Choice of the tuning parameter

We now discuss the choice of the tuning parameter γ in the PPC estimation procedure, which is an important issue when the penalized methodology is used in practice. Let

$$\tilde{\boldsymbol{\alpha}}_{\hat{m}_\gamma} = \tilde{\boldsymbol{\alpha}}_{\hat{m}_\gamma}(\hat{\mathcal{T}}_{\hat{m}_\gamma}) = \left[\tilde{\alpha}_1(\hat{\mathcal{T}}_{\hat{m}_\gamma})', ..., \tilde{\alpha}_{\hat{m}_\gamma+1}(\hat{\mathcal{T}}_{\hat{m}_\gamma})'\right]'$$

denote the set of the post-LASSO estimates of the regression coefficients based on the break dates in $\hat{\mathcal{T}}_{\hat{m}_\gamma} = \hat{\mathcal{T}}_{\hat{m}_\gamma}(\gamma)$, where we make the dependence of various estimates on γ explicitly. Let $\tilde{\sigma}^2(\hat{\mathcal{T}}_{\hat{m}_\gamma}) = Q_{NT}\big(\tilde{\boldsymbol{\alpha}}_{\hat{m}_\gamma}, \tilde{\boldsymbol{\Lambda}}, \tilde{\boldsymbol{F}}; \hat{\mathcal{T}}_{\hat{m}_\gamma}\big)$, where $\tilde{\boldsymbol{F}}$ is defined similarly to $\hat{\boldsymbol{F}}$ in (2.9) with $\hat{\boldsymbol{\Lambda}}$ and $\hat{\beta}_t$ replaced by $\tilde{\boldsymbol{\Lambda}}$ and $\tilde{\alpha}_j(\hat{\mathcal{T}}_{\hat{m}_\gamma})$ when $\hat{T}_{j-1} \leq t \leq \hat{T}_j - 1$. We then propose to select the tuning parameter γ by minimizing the following information criterion:

$$\mathsf{IC}(\gamma) = \ln\left[\tilde{\sigma}^2(\hat{\mathcal{T}}_{\hat{m}_\gamma})\right] + \rho_2 p\big(\hat{m}_\gamma + 1\big), \tag{4.2}$$

where $\rho_2 \equiv \rho_{2,NT}$ is pre-determined such that $\rho_2 \to 0$ and $\rho_2 \delta_{NT}^2 \to \infty$. Let $\hat{\gamma} = \text{argmin}_\gamma \, \mathsf{IC}(\gamma)$.

Theorem 4.2 *Suppose that Assumptions 1–2 and 3(ii) and 5 in Appendix A hold. Then*

$$\mathsf{P}\left(\hat{m}_{\hat{\gamma}} = m^0\right) \to 1 \ \text{as} \ (N, T) \to \infty.$$

The above theorem shows that by minimizing $\mathsf{IC}(\gamma)$, we can obtain a data-driven $\hat{\gamma}$ that ensures the correct determination of the number of breaks. When we minimize the objective function in (4.2), we do not restrict γ to satisfy Assumptions 3(i) and (iii) in Appendix A. If these two additional conditions also hold, we know from Corollary 3.3 that $\hat{m}_{\hat{\gamma}} = m^0$ w.p.a.1. But in practice, it is hard to ensure such conditions and Theorem 4.2 becomes handy.

In the following simulation, we choose $\rho_2 = c \log(\min(N,T))/\min(N,T)$, where c is a positive constant. This choice of ρ_2 satisfies the two restrictions specified above. To implement the information criterion in practice, we find an upper bound for the tuning parameter, $\gamma_{\max}$, that would yield zero break in every data generating process (DGP), and a lower bound $\gamma_{\min}$ that would yield many breaks. We then search for the optimal tuning parameter on the 20 evenly-distributed logarithmic grids in the interval $[\gamma_{\min},\ \gamma_{\max}]$. To determine c, we use a data-driven method that is similar to the one in Hallin and Liska (2007). Specifically, given an $N_0 > 0$, we examine subsamples (Y_{it}, X_{it}), $i = 1, ..., N_j$, $t = 1, ..., T$, where $j = 1, \ldots, J$ and $N_0 < N_1 < \cdots < N_J = N$. We examine a range of possible values for c, say $[c_{\min}, c_{\max}]$, where $c_{\min}$ leads to a large number of breaks and $c_{\max}$ leads to zero break for all choices of γ. For each c, we find the number of breaks in each subsample, $\hat{m}_j$, with $j = 1, \ldots, J$. Let $\bar{m}_c = \frac{1}{J}\sum_{j=1}^{J} \hat{m}_j$, we select the smallest $c \in [c_{\min}, c_{\max}]$ that satisfies $S_c = \frac{1}{J}\sum (\hat{m}_j - \bar{m}_c)^2 = 0$ and $\bar{m}_c < T - 1$. Intuitively, the constant c should be chosen such that the estimated number of breaks is constant across the subsamples. In our simulations we set $N_j = N - J + j$ and $J = 3$.

4.3 Implementation of the estimation method

The implementation of the PPC estimation method consists of two steps. In the first step, the preliminary estimation $\dot{\beta}_t$ is obtained along with the estimated number of factors $\hat{R}$. Given a generic number of factors, $\dot{\beta}_t$ is obtained by minimizing the first term of $\tilde{Q}_{NT,\gamma}\big(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{F}\big)$ in (2.2). The minimization problem is solved using an iterative algorithm based on (2.7) and (2.8) with $\hat{Q}_{NT,\gamma}(\boldsymbol{\beta}, \hat{\boldsymbol{\Lambda}})$ replaced by $\hat{Q}_{NT}(\boldsymbol{\beta}, \hat{\boldsymbol{\Lambda}})$, the first term of $\hat{Q}_{NT,\gamma}(\boldsymbol{\beta}, \hat{\boldsymbol{\Lambda}})$. The starting values for the iteration are chosen to be the pooled least squares estimates, assuming that coefficients are time-invariant and that no factor structure exists.

In the second step, given a generic tuning parameter γ, we use the following iterative algorithm to minimize $\tilde{Q}_{NT,\gamma}\big(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{F}\big)$, yielding the set of breaks corresponding to γ. Let $\theta_1 = \beta_1$ and $\theta_t = \beta_t - \beta_{t-1}$, $t = 2, \ldots, T$, and let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_T)'$,

(1) Initialize $\boldsymbol{\theta}^{(0)}$, which implies an initial set of breaks and an initial estimation of parameters in each regime.

(2) Given the initial set of breaks and parameter estimates, calculate factors $\boldsymbol{F}^{(i)}$ using eigenvalue decomposition, where the superscript $^{(i)}$ denotes the i-th iteration.

(3) Given $\boldsymbol{F}^{(i)}$, update $\boldsymbol{\theta}^{(i)}$ (or equivalently $\boldsymbol{\beta}^{(i)}$) that minimizes $\hat{Q}_{NT,\gamma}\left(\boldsymbol{\beta}, \boldsymbol{\Lambda}\right)$ in (2.4). This calculation utilizes a block-coordinate-descent algorithm similar to Qian and Su (2015b).

The updated $\boldsymbol{\theta}^{(i)}$ implies a new set of breaks, and the post-LASSO procedure is used to obtain new estimates of parameters in each regime.

(4) Repeat (2)-(3) until $\|\boldsymbol{\theta}^{(i)} - \boldsymbol{\theta}^{(i-1)}\|$ drops below a pre-determined threshold. Use the post-LASSO procedure to obtain the final estimate of parameters, factors and their loadings.

In the above iterative algorithm, the starting values for the iterations are chosen to be the preliminary estimates of the coefficients obtained in the first step. The post-LASSO procedure minimizes $Q_{NT}\big(\boldsymbol{\alpha}_m, \boldsymbol{\Lambda}, \boldsymbol{F}; \mathcal{T}_m\big)$ in (3.1) with $\mathcal{T}_m$ replaced by the estimated set of break dates in each iteration and with the starting values chosen to be the pooled least squares estimates as in the first step. Finally, we obtain the set of break dates using the tuning parameter that minimizes $\mathsf{IC}(\gamma)$ defined in (4.2).

4.4 Simulation

We consider the following data generating processes:

$$Y_{it} = \beta_{1t} Z_{it} + \beta_{2t} X_{it} + \lambda_i' f_t + \sigma u_{it},\ i = 1, \ldots, N,\ t = 1, \ldots, T,$$

where $f_t = \big[f_t(1), f_t(2)\big]'$ and $\lambda_i = \big[\lambda_i(1), \lambda_i(2)\big]'$ are two-dimensional random vectors, and

- DGP-1 (benchmark): $X_{it} \overset{i.i.d.}{\sim} \mathsf{N}(0,1)$, $Z_{it} = 1$, $\lambda_i \overset{i.i.d.}{\sim} \mathsf{N}(0, \boldsymbol{I}_2)$, $f_t \overset{i.i.d.}{\sim} \mathsf{N}(0, \boldsymbol{I}_2)$, both λ_i and f_t are independent of X_{it}, $u_{it} \overset{i.i.d.}{\sim} \mathsf{N}(0,1)$ and is independent of X_{it}, λ_i and f_t;

- DGP-2 (serial correlation in the common factor and heteroskedasticity in the error): X_{it}, Z_{it}, and λ_i are defined as in DGP-1, each of the two element in f_t is an AR(1) process with unit variance: $f_t(k) = 0.5 f_{t-1}(k) + \epsilon_t(k)$ with $\epsilon_t = \big[\epsilon_t(1), \epsilon_t(2)\big]' \overset{i.i.d.}{\sim} \mathsf{N}(0, 0.75\boldsymbol{I}_2)$, $u_{it} = \big(0.75 + 0.15 x_{it}^2\big)^{1/2} \eta_{it}^*$ with $\eta_{it}^* \overset{i.i.d.}{\sim} \mathsf{N}(0,1)$ and independent of X_{it}, λ_i and f_t;

- DGP-3 (dependent factors and serial correlation in the error): $X_{it} = 0.5\lambda_i' f_t + 0.5(\lambda_i'\iota + f_t'\iota) + \eta_{it}^{\diamond}$ with $\eta_{it}^{\diamond} \overset{i.i.d.}{\sim} \mathsf{N}(0,1)$ and $\iota = (1,1)'$, Z_{it} and λ_i are defined as in DGP-1, f_t is defined as in DGP-2, for each i, u_{it} is an independent ARMA(1,1) process with unit variance such that $u_{it} = 0.5 u_{i,t-1} + \epsilon_{it}^u + 0.5\epsilon_{i,t-1}^u$, where $\epsilon_{it}^u \overset{i.i.d.}{\sim} \mathsf{N}(0, 3/7)$.

- DGP-4 (dynamic panel): $X_{it} = Y_{i,t-1}$, $Z_{it} \overset{i.i.d.}{\sim} \mathsf{N}(0,1)$, λ_i and u_{it} are defined as in DGP-1, f_t is defined as in DGP-2.

In order to evaluate the performance under different noise levels, we select the free parameter σ to be either 0.5 or 1. In DGP-1 with no breaks, $\sigma = 1$ roughly corresponds to a signal-to-noise ratio of 1. We also experiment on different levels of factor loadings λ_i and find that the impact of the magnitude of the factor loadings on the performance of our method is small.

DGP-1 serves as the benchmark case where both the regressor and the idiosyncratic error are sequences of strong white noise. DGP-2 introduces serial correlation in the common factor f_t and conditional heteroskedasticity in the model errors. DGP-3 allows the dependence of both the factor loadings and common factors on the regressor. In addition, DGP-3 introduces serial correlation into the model errors, so the estimated model may be dynamically misspecified. DGP-4 has a dynamic panel AR(1) structure. We experiment on four combinations of dimensions: $(N, T) = (40, 40)$, $(N, T) = (80, 40)$, $(N, T) = (40, 80)$, and $(N, T) = (80, 80)$. The data-driven method to select both the constant c in ρ_2 and the tuning parameter γ is computationally intensive. As a result, we set the number of Monte Carlo replications to be 250, which might be smaller than usual but good enough for our purpose.

For the DGPs 1–3, we set $\beta_{1t} = \beta_{2t} = 1$ for all t when no break exists, $\beta_{1t} = \beta_{2t} = \mathbf{1}\{1 \leq t \leq T/2\}$ when there is one break, and $\beta_{1t} = \beta_{2t} = \mathbf{1}\{1 \leq t \leq \lfloor T/3 \rfloor\} + \mathbf{1}\{T/2 < t \leq T\}$ when there are two breaks. For the DGP-4, we set $\beta_{1t} = \beta_{2t} = 0.5$ for all t when there is no break, $\beta_{1t} = \beta_{2t} = 0.5 \cdot \mathbf{1}\{1 \leq t \leq T/2\}$ when there is only one break, and $\beta_{1t} = \beta_{2t} = 0.5 \cdot \mathbf{1}\{1 \leq t \leq \lfloor T/3 \rfloor\} + 0.5 \cdot \mathbf{1}\{T/2 < t \leq T\}$ when there are two breaks.

We first evaluate the probability of falsely detecting breaks when there is no break in the simulation design. Then we experiment on the DGPs with one or two breaks. We evaluate the probability of correctly detecting the number of breaks and the accuracy of break date estimation when breaks are detected. Tables 1, 2 and 3 report simulation results for the above DGPs. The first panel of Table 1 reports the percentages of falsely detecting breaks when there is no break ($m^0 = 0$). The second and the third panels report the percentages of correctly estimating the number of breaks when the true number of breaks is one and two, respectively. In Table 2, we report the ratio of average Hausdorff distance (HD) between the estimated and true sets of breaks to T, i.e., $100 \cdot \mathrm{HD}(\widehat{\mathcal{T}}_{\widehat{m}}, \mathcal{T}^0_{m^0})/T$, conditional on correct estimation of the number of breaks. Here the average is taken over 250 replications and the HD between two sets A and B is defined as $\mathrm{HD}(A, B) = \max\{\mathcal{D}(A, B), \mathcal{D}(B, A)\}$ with $\mathcal{D}(A, B) \equiv \sup_{b \in B} \inf_{a \in A} |a - b|$. The mean squared or absolute errors of the parameter estimates are roughly proportional to the Hausdorff error of the break-date estimation and hence are not reported. In Table 3 we report the percentages of correctly estimating the number of factors in the Monte Carlo replications.

We summarize the major findings from these tables. (i) When there is no break in the DGPs, the probabilities of falsely detecting breaks decline to zero as either N or T increases. (ii) When there are one or two breaks, the probabilities of correctly estimating the number of breaks converge fairly quickly to 100% or near 100% as both N and T increase. The detection procedure performs slightly better at lower idiosyncratic noise levels ($\sigma = 0.5$) than at higher noise level ($\sigma = 1$). The performance is robust to serial correlation in the common factor, serial correlation and conditional heteroskedasticity in the errors, and the dependence of both the factors and their loadings on the regressor. For the dynamic panel (DGP-4), the procedure performs less satisfactorily. However, this may be due to the fact that the signal-to-noise ratio in this case is roughly 1/3, much less than that in the other three DGPs. (iii) Conditional on the correct estimation of the number of breaks, our procedure estimates the break dates accurately, which can be seen from Table 2 (iv) Finally, Table 3 shows that the BIC-type information criterion specified in (4.1) can accurately determine the number of factors for the interactive fixed effects structure.

5 An empirical application to the environmental Kuznets curve

The environmental Kuznets curve (EKC) has become a standard feature in the environmental policy literature. It hypothesizes that the relationship between income and the emission of chemicals like sulfur dioxide (SO_2) and carbon dioxide (CO_2) or the natural resource usage has an inverted U-shape, which is similar to the relationship between income and inequality in the Kuznets curve hypothesis in economics. In this section we consider the following specification:

$$c_{it} = \beta_{0t} + \beta_{1t} y_{it} + \beta_{2t} y_{it}^2 + \beta_{3t} e_{it} + \lambda_i' f_t + u_{it},$$

where c_{it} represents the logarithm of per capita CO_2 emission for country i in year t, y_{it} represents the logarithm of per capita income (gross domestic product, abbreviated as GDP), e_{it} represents the logarithm of per capita consumption of energy, f_t is a vector of unobservable common factors and λ_i is a vector of factor loadings. Our data-driven BIC criterion determines that the number of factors is five. The controlling of energy consumption in EKC studies was used in the time series regression setting in Ang (2007), and the panel data setting in Apergis and Payne (2009, 2010), Lean and Smyth (2010), Arouri et al. (2012) and Farhani et al. (2014). The panel data

Table 1: The probabilities for falsely detecting breaks when there are none and of correctly detecting the breaks when there are breaks

DGP	σ	N=T=40	N=40,T=80	N=80,T=40	N=T=80
$m^0 = 0$, % of falsely detecting breaks when there are none.					
1	0.5	0	0	0	0
	1	0	0	0	0
2	0.5	0	0	0	0
	1	0.4	0	0	0
3	0.5	2.8	1.2	0.4	0
	1	1.2	0	0.4	0
4	0.5	0	0	0	0
	1	0.4	0	0	0
$m^0 = 1$, % of correctly detecting one break					
1	0.5	100	100	100	100
	1	98.8	99.6	100	100
2	0.5	100	100	100	100
	1	99.6	99.2	100	100
3	0.5	99.2	100	100	100
	1	91.6	98	100	99.6
4	0.5	97.6	99.6	100	100
	1	79.2	76.8	95.2	98
$m^0 = 2$, % of correctly detecting two breaks					
1	0.5	100	100	100	100
	1	99.2	98.4	100	100
2	0.5	99.6	100	100	100
	1	98	99.2	100	100
3	0.5	99.2	100	100	100
	1	87.2	92.4	98	99.2
4	0.5	94	92.8	99.2	100
	1	54.8	58.4	94	94.4

Table 2: Estimation accuracy for the break dates when there is one or two structural breaks

DGP	σ	N=T=40	N=40,T=80	N=80,T=40	N=T=80
			$m^0 = 1$		
1	0.5	0.000	0.000	0.000	0.000
	1.0	0.000	0.005	0.000	0.000
2	0.5	0.000	0.000	0.000	0.000
	1.0	0.005	0.000	0.000	0.000
3	0.5	0.000	0.000	0.000	0.000
	1.0	0.066	0.051	0.000	0.000
4	0.5	0.020	0.030	0.000	0.000
	1.0	0.423	0.540	0.037	0.092
			$m^0 = 2$		
1	0.5	0.000	0.000	0.000	0.000
	1.0	0.010	0.005	0.000	0.000
2	0.5	0.000	0.000	0.000	0.000
	1.0	0.010	0.015	0.000	0.000
3	0.5	0.000	0.000	0.000	0.000
	1.0	0.006	0.011	0.000	0.005
4	0.5	0.027	0.032	0.000	0.000
	1.0	0.246	0.300	0.011	0.053

Note. The table reports $100 \cdot \text{HD}(\widehat{\mathcal{T}}_{\widehat{m}}, \mathcal{T}^0_{m^0})/T$, averaged over 250 replications.

studies in the existing literature, however, assume that the coefficients are constant over time. In our specification, we not only introduce the interactive fixed effects in the panel data models but also allow time-varying coefficients that may capture the instability of the EKC brought by the changing social, political, and economic environment in the past few decades.

We obtain the panel data set from World Bank Development Indicators. The CO_2 emission is measured in metric tones per capita, income is measured using per capita real GDP in constant 2000 USD, and energy consumption is measured with kilogram of oil equivalent per capita. The time frame is selected to be 1971-2010. We exclude OPEC countries, small countries whose populations are less than six million, and other countries with missing observations during the time span. In total, we have $N = 74$ countries and $T = 40$ time points.

The results are summarized in Table 4. The information criterion defined in (4.2) selects a tuning parameter that identifies two breaks ($\hat{m} = 2$) in 1990 and 1992. In the first regime of 1971-1990, the EKC hypothesis is confirmed, as the coefficient on the squared income is significantly negative, implying an inverted U-shape. The elasticities of CO_2 emission per capita with respect to real income per capita in the regime is $(0.198 - 0.02y)$, where y denotes the

Table 3: The probabilities for correctly estimating the number of factors

DGP	σ	N=T=40	N=40,T=80	N=80,T=40	N=T=80
			$m^0 = 0$		
1	0.5	100	100	100	100
	1	98.8	100	100	100
2	0.5	100	100	100	100
	1	100	100	100	100
3	0.5	100	100	100	100
	1	98	100	100	100
4	0.5	100	100	100	100
	1	98.8	100	100	100
			$m^0 = 1$		
1	0.5	100	100	100	100
	1	99.6	100	100	100
2	0.5	100	100	100	100
	1	100	100	100	100
3	0.5	100	100	100	100
	1	97.6	100	100	100
4	0.5	100	100	100	100
	1	98.4	100	100	100
			$m^0 = 2$		
1	0.5	100	100	100	100
	1	99.6	100	100	100
2	0.5	100	100	100	100
	1	100	100	100	100
3	0.5	100	100	100	100
	1	97.6	100	100	100
4	0.5	100	100	100	100
	1	98.8	100	99.6	100

Table 4: A panel data estimation of the EKC for 74 countries from 1971 to 2010

$\hat{m}$	Variables	1971-1989	1990-1991	1992-2010	IC
2	Intercept	$-5.816(0.075)^c$	$-4.641(1.168)^c$	$-6.222(0.187)^c$	-5.948
	y_{it}	$0.198(0.020)^c$	$-0.028(0.239)$	$0.332(0.037)^c$	
	y_{it}^2	$-0.010(0.001)^c$	$-0.004(0.014)$	$-0.017(0.002)^c$	
	e_{it}	$0.847(0.003)^c$	$0.798(0.041)^c$	$0.821(0.005)^c$	
0	Intercept		$-5.841(0.037)^c$		-5.940
	y_{it}		$0.248(0.010)^c$		
	y_{it}^2		$-0.013(0.001)^c$		
	e_{it}		$0.842(0.002)^c$		

Note. Superscript a, b and c denotes significance level at 10%, 5%, and 1%, respectively. Standard errors are given in parentheses.

logarithm of real GDP per capita. The threshold, or the turning points of the EKC, occurs at the per capita income of 19,900 USD (2000). The second regime is a short one, covering only two years, 1990 and 1991. In this regime, the coefficients on both y_{it} and y_{it}^2 are statistically insignificant. The signs of these coefficients do not point to an inverted U-shape. This suggests that, using a short panel or cross-section data set collected in a certain time period, one may reject the EKC hypothesis, while a longer panel data would arrive at the opposite conclusion. In the third regime of 1992-2010, the EKC hypothesis is again confirmed. The elasticities of CO_2 emission per capita with respect to real income per capita in the regime is $(0.332 - 0.034y)$, implying a threshold of 17,400 USD (2000). Comparing with the first regime, we may conclude that the EKC has shifted leftward in the past two decades. The second regime of 1990-1991 may be regarded as a transition period from the first regime to the second regime, which is more environment-friendly. We also report in Table 4 the case of zero break ($\hat{m} = 0$), where coefficients are assumed to be constant. Here the EKC hypothesis is also confirmed, with a threshold at 13,900 USD (2000). Interestingly, the panel data model with constant regression coefficients paints the most optimistic EKC. If we estimate the regression coefficients in the panel data model with two structural breaks detected by the PPC method, however, we see a more cautious picture for the EKC.

6 Conclusions

In this paper, we study the estimation of the panel data models with interactive fixed effects and multiple structural breaks, which substantially generalizes the existing work which either considers the panel models with interactive fixed effects but no structural break (c.f., Bai, 2009), or the panel models with multiple structural breaks but under cross-sectional independence (c.f., Qian and Su, 2015b). We develop a novel PPC estimation procedure with the AGF-LASSO penalty function to consistently estimate both the regression coefficients and the factor loadings. Under some regularity conditions, we show that both the unknown number of structural breaks and the unobservable break dates can be consistently estimated. In order to further improve the convergence rates, we also estimate the regression coefficients (in different regimes) through the post-LASSO method and then establish the asymptotic distribution theory of the resulting estimators. In particular, the developed shrinkage estimation methodology and the asymptotic theory are also applicable to the case of dynamic panel data. We introduce two data-driven methods to determine the number of factors and choose the tuning parameter involved in the PPC estimation procedure, respectively. The simulation studies show that the proposed PPC method has a high probability of correctly estimating the number of breaks when the structural breaks exist in the simulation design, and a low probability of false detection when there is no structural break. We apply our method to study the EKC for 74 countries over 40 years and find two breaks in the panel data.

Appendix

We first give in Appendix A some regularity conditions that are used to derive the asymptotic results. Then we provide some technical lemmas and prove the main theoretical results in Appendix B. The proofs of the technical lemmas are given in Appendix C of the supplemental document.

A Assumptions

We start with the introduction of some notation. Denote

$$\delta_{NT} = \min(\sqrt{N}, \sqrt{T}), \quad \delta_{p,NT} = \min(\sqrt{N/p}, \sqrt{T}),$$
$$\Delta_{NT} = \min_{1\le j\le m^0} \left\|\alpha_{j+1}^0 - \alpha_j^0\right\|, \quad \Delta_{NT}^* = \max_{1\le j\le m^0} \left\|\alpha_{j+1}^0 - \alpha_j^0\right\|.$$

Let $\xi_{ij} = \sum_{t=1}^T \varepsilon_{it}\varepsilon_{jt}$ for $1 \le i, j \le N$, and $\xi_{ts}^* = \sum_{i=1}^N \varepsilon_{it}\varepsilon_{is}$ for $1 \le t, s \le T$. Define

$$\mathbf{\Omega}_{NT} = \mathbf{\Phi}_{NT} - \mathbf{\Phi}_{NT}^*, \quad \mathbf{\Phi}_{NT} = \mathsf{diag}\big(\Phi_1, ..., \Phi_{m^0+1}\big), \quad \mathbf{\Phi}_{NT}^* = \big(\Phi_{jk}^*\big)_{1\le j,k\le m^0+1},$$

where

$$\Phi_j = \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\boldsymbol{\Lambda}^0} X_t, \quad \Phi_{jk}^* = \frac{1}{NT\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} \sum_{s=T_{k-1}^0}^{T_k^0-1} \chi_{st} X_t' \boldsymbol{M}_{\boldsymbol{\Lambda}^0} X_s,$$

$\tau_j(T) = T_j^0 - T_{j-1}^0$ and $\chi_{st} = f_s^{0\prime}\big(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\big)^+ f_t^0$. In order to prove the asymptotic results stated in Sections 3 and 4, we make the following assumptions.

Assumption 1 (i) There exist two positive definite matrices $\boldsymbol{\Sigma}_F$ and $\boldsymbol{\Sigma}_\Lambda$ such that

$$\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0 \xrightarrow{P} \boldsymbol{\Sigma}_F, \quad \frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0 \xrightarrow{P} \boldsymbol{\Sigma}_\Lambda.$$

Furthermore, both the common factors f_t^0 and the factor loadings λ_i^0 have finite 8-th moments.
(ii) Let the regressor X_t satisfy $\max_{1\le t\le T}\|X_t\| = O_P\big(p^{1/2}N^{1/2}\big)$, and

$$c_x \le \inf_{\boldsymbol{\Lambda}} \min_{1\le t\le T} \mu_{\min}\big(N^{-1}X_t'\boldsymbol{M}_{\boldsymbol{\Lambda}}X_t\big) \le \max_{1\le t\le T} \mu_{\max}\big(N^{-1}X_t'X_t\big) \le c_x^*$$

w.p.a.1, where $0 < c_x < c_x^* < \infty$, and $\inf_{\boldsymbol{\Lambda}}$ is taken with respect to $\boldsymbol{\Lambda}$ such that $\frac{1}{N}\boldsymbol{\Lambda}'\boldsymbol{\Lambda} = \boldsymbol{I}_{R_0}$.
(iii) Let $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_T)$. The idiosyncratic error term ε_{it} satisfies $\mathsf{E}[\varepsilon_{it}] = 0$ and $\mathsf{E}[\varepsilon_{it}^8] < c_\varepsilon$ for each i and t and $\|\boldsymbol{\varepsilon}\|_{\mathrm{sp}}$ =max($\sqrt{N}$,$\sqrt{T}$). where c_ε is a bounded positive constant. Furthermore,

$$\max_{1\le s,t\le T} \mathsf{E}\big[\|X_t'\varepsilon_s\|^2\big] = O(pN), \quad \max_{1\le t\le T} \mathsf{E}\big[\|\boldsymbol{\Lambda}^{0\prime}\varepsilon_t\|^2\big] = O(N), \quad \mathsf{E}\big[\|\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\varepsilon}\boldsymbol{F}^0\|^2\big] = O(NT),$$

$$\max_{1\le i,j\le N} \mathsf{E}\Big[\Big\|\sum_{t=1}^T\sum_{s=1}^T \varepsilon_{it}\varepsilon_{js}\gamma_t\gamma_s'\Big\|^2\Big] = O(T^2), \quad \text{and} \quad \max_{1\le i,j\le N} \mathsf{E}\Big[\Big\|\sum_{t=1}^T\sum_{s=1}^T \varepsilon_{it}\varepsilon_{js}\varepsilon_t'\varepsilon_s\Big\|^2\Big] = O(N^2T^2 + T^4),$$

where γ_t can be either 1 or f_t^0.
(iv) Assume that $\max_{1\le i,j\le N}\mathsf{Var}(\xi_{ij}) = \max_{1\le i,j\le N}\mathsf{Var}(\sum_{t=1}^T \varepsilon_{it}\varepsilon_{jt}) = O(T)$, and there exists $\sigma_{ij} > 0$ such that $\big|\mathsf{E}(\xi_{ij})\big| \le \sigma_{ij}T$ and $\sum_{i=1}^N\sum_{j=1}^N \sigma_{ij}^2 = O(N)$. Furthermore, we have

$$\max_{1\le t\le T} \mathsf{E}\Big[\sum_{s=1}^T (\xi_{ts}^*)^2\Big] = O\big(N^2 + NT\big), \quad \mathsf{E}\Big[\Big\|\sum_{t=1}^T\sum_{s=1}^T f_t^{0\prime}\xi_{ts}^* f_s^0\Big\|^2\Big] = O\big(N^2T^2\big).$$

Assumption 2 (i) The tuning parameter γ satisfies that

$$\gamma = o(1), \quad \gamma m^0 \Delta_{NT}^{-\kappa}\delta_{p,NT} = O(1) \quad \text{as } (N,T) \to \infty,$$

where κ is the user-specified positive constant defined in (2.3).
(ii) Let the following restrictions hold:

$$\delta_{p,NT}\Delta_{NT} \to \infty, \quad \Delta_{NT}^* = O(p^{1/2}), \quad pN^{-1/2} + p^{1/2}T^{-1/2} = o(1) \quad \text{as} \quad (N,T) \to \infty.$$

(iii) Let $\gamma\delta_{p,NT}^{\kappa+1} \to \infty$ as $(N,T) \to \infty$.

Assumption 3 (i) There exists a positive definite matrix $\mathbf{\Omega}_0$ such that $\left\|\mathbf{\Omega}_{NT} - \mathbf{\Omega}_0\right\|_{sp} = o_P(1)$.

(ii) There exist $0 < c_\tau \leq c_\tau^* < \infty$ such that

$$\frac{c_\tau T}{m^0} \leq \min_{1\leq j\leq m^0+1} \tau_j(T) \leq \max_{1\leq j\leq m^0+1} \tau_j(T) \leq \frac{c_\tau^* T}{m^0}.$$

(iii) Letting $A_t = \sum_{s=1}^T \mathbf{\Lambda}^{0\prime}\varepsilon_s\varepsilon_s'\varepsilon_t$, $\max_{1\leq t\leq T} \mathsf{E}(A_t^2) = O(N^2(N+T))$.

(iv) Letting $W_{j,NT} = \frac{1}{N\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\mathbf{\Lambda}^0}(\varepsilon_t - \varepsilon_t^*)$ for $j = 1, ..., m^0+1$ and $\boldsymbol{W}_{NT} = (W_{1,NT}', ..., W_{m^0+1,NT}')'$, there exist $\boldsymbol{B}_{NT}(3)$ and $\mathbf{\Omega}_1$ such that

$$\mathbf{S}_*\mathbf{D}_{NT}[\boldsymbol{W}_{NT} - \boldsymbol{B}_{NT}(3)] \xrightarrow{D} \mathsf{N}\big(\mathbf{0},\ \mathbf{S}_*\mathbf{\Omega}_1\mathbf{S}_*'\big),$$

where $\mathbf{D}_{NT}$ is defined in Section 3.2, $\mathbf{S}_*$ is an arbitrary $k^0 \times p(m^0+1)$ matrix with full row rank, and k^0 is a fixed positive integer.

(v) Let $(NT)^{1/2}/\delta_{p,NT}^3 = o(1)$ and $p/\delta_{p,NT} = o(1)$ as $(N,T) \to \infty$.

Assumption 4 As $(N,T) \to \infty$, $\rho_1 \to 0$ and $\delta_{p,NT}^2\rho_1 \to \infty$.

Assumption 5 (i) For any $0 \leq m < m^0$, there exists a positive constant c_β such that

$$\min_{\mathcal{T}_m}\min_{\boldsymbol{\alpha}_m} \frac{m^0}{T\Delta_{NT}^2}\sum_{j=1}^{m+1}\sum_{t=T_{j-1}}^{T_j-1}\left\|\beta_t^0 - \alpha_j\right\|^2 \geq c_\beta,$$

where $\boldsymbol{\alpha}_m$ and $\mathcal{T}_m$ are defined in Section 3.2.

(ii) As $(N,T) \to \infty$, $\frac{m^0}{T\Delta_{NT}^2}(pN^{-1/2} + p^{1/2}T^{-1/2}) = o(1)$.

(iii) As $(N,T) \to \infty$, $m^0p\rho_2 \to 0$ and $\delta_{p,NT}^2 p\rho_2 \to \infty$.

Remark A.1. Assumption 1 imposes some standard moment conditions on X_{it}, f_t^0, λ_i^0 and ε_{it}, which are analogous to those in the existing literature such as Bai and Ng (2002), Bai (2009), Bai and Li (2014), Lu and Su (2015), and Moon and Weidner (2015). As we allow p, the dimension of the regression coefficients, to be divergent, some of our moment conditions might be slightly stronger than those in the literature. Assumptions 1(iii) and (iv) allow weak form of cross-sectional dependence and serial dependence among X_{it}, f_t^0, λ_i^0 and ε_{it}. In particular, unlike Pesaran (2006) and Bai (2009), we do not assume independence between ε_{it} and $(X_{js}, f_s^0, \lambda_j^0)$ for all i, j, t, s, and our theories are thus applicable to the dynamic autoregressive panel data models with interactive fixed effects. Assumption 2 imposes some mild restrictions on the tuning parameter γ and the jump sizes of the regression coefficients, which can be easily justified. For example, assuming that the jump sizes are bounded away from zero and infinity and $N \sim T$, Assumption 2 can be simplified to $\gamma = o(1)$, $\gamma m^0 (N/p)^{1/2} = O(1)$, $p = o(N^{1/2})$ and $\gamma (N/p)^{(\kappa+1)/2} \to \infty$. Assumption 3 imposes some additional conditions for the proof of the asymptotic distribution theory of the post-LASSO estimation, which can be verified under some primitive conditions. For example, if we assume that $\{\varepsilon_{it}, \lambda_i^0\}$ are independent across i and for each i, $\{\varepsilon_{it}\}$ is a martingale difference sequence with respect to the σ-field generated by $(\varepsilon_{i,t-1}, \ldots, \varepsilon_{i1}, f_{t-1}^0, \ldots, f_1^0, \lambda_i^0)$ and $\{\varepsilon_{it}, X_{it}\}$ satisfy some strongly mixing conditions, then the moment condition in Assumption 3(iii) holds. Assumption 4 indicates that ρ_1 has to shrink to zero at an appropriate rate to avoid both over-selection and under-selection of the number of factors. Assumptions 5(i)(ii) impose conditions to avoid the selection of model with fewer breaks than the true number by using an information criterion proposed in Section 4.2. Assumption 5(iii) parallels Assumption 4.

B Proofs of the main asymptotic results

In this appendix, we give the detailed proofs of the asymptotic results in Sections 3 and 4. We start with two technical lemmas whose proofs are provided in Appendix C of the supplemental document.

Lemma B.1 *Suppose that Assumption 1 in Appendix A holds and* $pN^{-1/2} + p^{1/2}T^{-1/2} = o(1)$. *Let* $\dot{\boldsymbol{\beta}} = \left(\dot{\beta}_1', ..., \dot{\beta}_T'\right)'$ *be the preliminary estimates of the regression coefficients which minimize,* $\hat{Q}_{NT}(\boldsymbol{\beta}, \boldsymbol{\Lambda})$*, the first term of the objective function defined in (2.4). Then* $\left\|\dot{\beta}_t - \beta_t^0\right\| = O_P\left(p^{1/2}N^{-1/2} + T^{-1/2}\right) = O_P\left(\delta_{p,NT}^{-1}\right)$ *for any* $t = 1, 2, ..., T$*, where* $\delta_{p,NT}$ *is defined as in Appendix A.*

Lemma B.2 *Suppose that Assumption 1 Appendix A holds and let* $\eta_{NT} = \frac{1}{T}\sum_{t=1}^T \|\hat{\beta}_t - \beta_t^0\|^2$. *Then we have*

(i) $\frac{1}{NT}\sum_{t=1}^T (\hat{\beta}_t - \beta_t^0)' X_t' \boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}} \varepsilon_t = O_P\left(\delta_{p,NT}^{-1}\eta_{NT}^{1/2}\right)$,

(ii) $\sum_{t=1}^T f_t^{0\prime} \boldsymbol{\Lambda}^{0\prime} \boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}} \varepsilon_t = O_P\left(\delta_{NT}^{-2} + \delta_{p,NT}^{-1}\eta_{NT}^{1/2}\right)$, *and*

(iii) $\frac{1}{NT}\sum_{t=1}^T \varepsilon_t' (\boldsymbol{P}_{\hat{\boldsymbol{\Lambda}}} - \boldsymbol{P}_{\boldsymbol{\Lambda}^0}) \varepsilon_t = O_P\left(\delta_{NT}^{-2}\right)$.

We next give the proof of Theorem 3.1 by using the above two lemmas.

Proof of Theorem 3.1. (i) Recall that the penalized estimate of $\boldsymbol{\beta}^0$ is denoted by $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1', ..., \hat{\beta}_T'\right)'$ and the estimated factor loading matrix is denoted by $\hat{\boldsymbol{\Lambda}}$. Note that

$$Y_t - X_t\hat{\beta}_t = X_t(\beta_t^0 - \hat{\beta}_t) + \boldsymbol{\Lambda}^0 f_t^0 + \varepsilon_t. \tag{B.1}$$

Then, by (B.1) and using the fact that $\boldsymbol{M}_{\boldsymbol{\Lambda}^0}\boldsymbol{\Lambda}^0 = \mathbf{0}$, we have

$$
\begin{aligned}
\hat{Q}_{NT,\gamma}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Lambda}}) - \hat{Q}_{NT,\gamma}(\boldsymbol{\beta}^0, \boldsymbol{\Lambda}^0) \;=\;& \frac{1}{T}\sum_{t=1}^{T}\Big[\hat{Q}^{*}_{NT,t}(\beta_t, \boldsymbol{\Lambda}) + \hat{Q}^{\diamond}_{NT,t}(\beta_t, \boldsymbol{\Lambda})\Big] \\
&+ \frac{\gamma}{T}\sum_{t\in\mathcal{T}^0_{m^0}} \dot{w}_t\Big[\|\hat{\beta}_t - \hat{\beta}_{t-1}\| - \|\beta^0_t - \beta^0_{t-1}\|\Big] \\
&+ \frac{\gamma}{T}\sum_{t\in\mathcal{T}^c} \dot{w}_t\Big[\|\hat{\beta}_t - \hat{\beta}_{t-1}\| - \|\beta^0_t - \beta^0_{t-1}\|\Big], \qquad \text{(B.2)}
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{Q}^{*}_{NT,t}(\beta_t, \boldsymbol{\Lambda}) &= \frac{1}{N}\Big[(\hat{\beta}_t - \beta^0_t)'X_t'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_t(\hat{\beta}_t - \beta^0_t) - 2(\hat{\beta}_t - \beta^0_t)'X_t'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0 f^0_t + f^{0\prime}_t\boldsymbol{\Lambda}^{0\prime}\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0 f^0_t\Big], \\
\hat{Q}^{\diamond}_{NT,t}(\beta_t, \boldsymbol{\Lambda}) &= \frac{1}{N}\Big[-2(\hat{\beta}_t - \beta^0_t)'X_t'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\varepsilon_t + 2f^{0\prime}_t\boldsymbol{\Lambda}^{0\prime}\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\varepsilon_t - \varepsilon_t'\boldsymbol{P}_{\hat{\boldsymbol{\Lambda}}}\varepsilon_t + \varepsilon_t'\boldsymbol{P}_{\boldsymbol{\Lambda}^0}\varepsilon_t\Big].
\end{aligned}
$$

As $\beta^0_t - \beta^0_{t-1} = \mathbf{0}$ for $t \in \mathcal{T}^c$, the last term on the right hand side of (B.2) satisfies that

$$
\frac{\gamma}{T}\sum_{t\in\mathcal{T}^c} \dot{w}_t\Big[\|\hat{\beta}_t - \hat{\beta}_{t-1}\| - \|\beta^0_t - \beta^0_{t-1}\|\Big] = \frac{\gamma}{T}\sum_{t\in\mathcal{T}^c} \dot{w}_t\|\hat{\beta}_t - \hat{\beta}_{t-1}\| \geq 0. \qquad \text{(B.3)}
$$

By the triangle inequality, the Cauchy-Schwarz inequality, Lemma B.1 and Assumption 2(ii) in Appendix A, we can prove that

$$
\begin{aligned}
\sum_{t\in\mathcal{T}^0_{m^0}} \dot{w}_t\Big[\|\hat{\beta}_t - \hat{\beta}_{t-1}\| - \|\beta^0_t - \beta^0_{t-1}\|\Big] &\leq O_P(\Delta_{NT}^{-\kappa})\sum_{t\in\mathcal{T}^0_{m^0}}\|\hat{\beta}_t - \beta^0_t\| \\
&\leq O_P(\Delta_{NT}^{-\kappa})(m^0)^{1/2}\left(\sum_{t\in\mathcal{T}^0_{m^0}}\|\hat{\beta}_t - \beta^0_t\|^2\right)^{1/2} \\
&\leq O_P(\Delta_{NT}^{-\kappa})(m^0 T)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{\beta}_t - \beta^0_t\|^2\right)^{1/2}.
\end{aligned}
$$

Note that Assumption 2(i) implies that $\gamma(m^0)^{1/2}T^{-1/2}\Delta_{NT}^{-\kappa} = o(\delta^{-1}_{p,NT})$ where $\delta_{p,NT} = \min(\sqrt{N/p}, \sqrt{T})$. This, together with the above argument, indicates that

$$
\frac{\gamma}{T}\sum_{t\in\mathcal{T}^0_{m^0}} \dot{w}_t\Big[\|\hat{\beta}_t - \hat{\beta}_{t-1}\| - \|\beta^0_t - \beta^0_{t-1}\|\Big] = o_P\left(\delta^{-1}_{p,NT}\eta^{1/2}_{NT}\right). \qquad \text{(B.4)}
$$

By Lemma B.2, we can readily show that

$$
\frac{1}{T}\sum_{t=1}^{T}\hat{Q}^{\diamond}_{NT,t}(\beta_t, \boldsymbol{\Lambda}) = O_P\left(\delta^{-2}_{NT} + \delta^{-1}_{p,NT}\eta^{1/2}_{NT}\right). \qquad \text{(B.5)}
$$

Combining (B.4) and (B.5), we have

$$\hat{Q}_{NT,\gamma}\big(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\Lambda}}\big)-\hat{Q}_{NT,\gamma}\big(\boldsymbol{\beta}^0,\boldsymbol{\Lambda}^0\big)\geq\frac{1}{T}\sum_{t=1}^{T}\hat{Q}^*_{NT,t}(\beta_t,\boldsymbol{\Lambda})+O_P\left(\delta_{NT}^{-2}+\delta_{p,NT}^{-1}\eta_{NT}^{1/2}\right). \tag{B.6}$$

Define the vectors:

$$\hat{\boldsymbol{d}}_\beta=\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0 \text{ and } \hat{\boldsymbol{d}}_\Lambda=\frac{1}{N^{1/2}}\mathsf{vec}(\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0),$$

where $\mathsf{vec}(\cdot)$ denotes the vectorization of a matrix; and define the matrices:

$$\begin{aligned}\hat{\boldsymbol{A}} &= \frac{1}{N}\mathsf{diag}\big(X_1'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_1,...,X_T'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_T\big),\quad \hat{\boldsymbol{B}}=(\boldsymbol{F}^{0\prime}\boldsymbol{F}^0)\otimes\boldsymbol{I}_N, \text{ and}\\ \hat{\boldsymbol{C}} &= \frac{1}{N^{1/2}}\left[f_1^0\otimes\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_1,...,f_T^0\otimes\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_T\right],\end{aligned}$$

where $\otimes$ denotes the Kronecker product. It is easy to verify that

$$\begin{aligned}&\frac{1}{NT}\sum_{t=1}^{T}\big(\hat{\beta}_t-\beta_t^0\big)'X_t'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_t\big(\hat{\beta}_t-\beta_t^0\big)=\frac{1}{T}\hat{\boldsymbol{d}}_\beta'\hat{\boldsymbol{A}}\hat{\boldsymbol{d}}_\beta,\\ &\frac{1}{NT}\sum_{t=1}^{T}\big(\hat{\beta}_t-\beta_t^0\big)'X_t'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0f_t^0=\frac{1}{NT}\sum_{t=1}^{T}\mathsf{Tr}\Big\{\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0f_t^0\big(\hat{\beta}_t-\beta_t^0\big)'X_t'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\Big\}=\frac{1}{T}\hat{\boldsymbol{d}}_\Lambda'\hat{\boldsymbol{C}}\hat{\boldsymbol{d}}_\beta,\\ &\frac{1}{NT}\sum_{t=1}^{T}f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0f_t^0=\frac{1}{NT}\sum_{t=1}^{T}\mathsf{Tr}\Big(\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0f_t^0f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\Big)=\frac{1}{T}\hat{\boldsymbol{d}}_\Lambda'\hat{\boldsymbol{B}}\hat{\boldsymbol{d}}_\Lambda,\end{aligned}$$

where we have used the following facts on matrix calculation: $\mathsf{Tr}\big(\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_3\big)=\mathsf{vec}'\big(\boldsymbol{A}_1\big)\big(\boldsymbol{A}_2\otimes\boldsymbol{I}_k\big)\mathsf{vec}\big(\boldsymbol{A}_3\big)$ and $\mathsf{Tr}\big(\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{A}_4\big)=\mathsf{vec}'\big(\boldsymbol{A}_1\big)\big(\boldsymbol{A}_2\otimes\boldsymbol{A}_4'\big)\mathsf{vec}\big(\boldsymbol{A}_3'\big)$ with k being the size of the column vectors in $\boldsymbol{A}_3$. Using the above notations, we may show that

$$\frac{1}{T}\sum_{t=1}^{T}\hat{Q}^*_{NT,t}(\beta_t,\boldsymbol{\Lambda})=\frac{1}{T}\big(\hat{\boldsymbol{d}}_\beta'\hat{\boldsymbol{A}}\hat{\boldsymbol{d}}_\beta-2\hat{\boldsymbol{d}}_\Lambda'\hat{\boldsymbol{C}}\hat{\boldsymbol{d}}_\beta+\hat{\boldsymbol{d}}_\Lambda'\hat{\boldsymbol{B}}\hat{\boldsymbol{d}}_\Lambda\big)=\frac{1}{T}\big(\hat{\boldsymbol{d}}_\beta'\hat{\boldsymbol{D}}\hat{\boldsymbol{d}}_\beta+\hat{\boldsymbol{d}}_*'\hat{\boldsymbol{B}}\hat{\boldsymbol{d}}_*\big), \tag{B.7}$$

where $\hat{\boldsymbol{D}}=\hat{\boldsymbol{A}}-\hat{\boldsymbol{C}}'\hat{\boldsymbol{B}}^+\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{d}}_*=\hat{\boldsymbol{d}}_\Lambda-\hat{\boldsymbol{B}}^+\hat{\boldsymbol{C}}\hat{\boldsymbol{d}}_\beta$. By Assumption 1(i), we may show that the minimum eigenvalue of $\frac{1}{T}\hat{\boldsymbol{B}}$ is bounded away from zero w.p.a.1, i.e., there exists a positive constant c_1 such that $\mu_{\min}\big(\hat{\boldsymbol{B}}/T\big)>c_1$ for sufficiently large T. We next show that $\mu_{\max}\big(\hat{\boldsymbol{C}}'\hat{\boldsymbol{C}}/T\big)=o_P(1)$. Letting $\nu_{st}=f_s^{0\prime}f_t^0$, it is easy to verify that

$$\hat{\boldsymbol{C}}'\hat{\boldsymbol{C}}=\frac{1}{N}\begin{pmatrix}\nu_{11}X_1'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_1 & \nu_{12}X_1'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_2 & ... & \nu_{1T}X_1'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_T\\ \nu_{21}X_2'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_1 & \nu_{22}X_2'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_2 & ... & \nu_{2T}X_2'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_T\\ \vdots & \vdots & \ddots & \vdots\\ \nu_{T1}X_T'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_1 & \nu_{T2}X_T'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_2 & ... & \nu_{TT}X_T'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_T\end{pmatrix}.$$

Letting

$$\hat{\boldsymbol{C}}_1 = \frac{1}{N}\begin{pmatrix} \nu_{11}X_1'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_1 & \nu_{12}X_1'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_2 & \ldots & \nu_{1T}X_1'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_T \\ \mathbf{0} & \nu_{22}X_2'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_2 & \ldots & \nu_{2T}X_2'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \nu_{TT}X_T'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_T \end{pmatrix}$$

and $\hat{\boldsymbol{C}}_d = \frac{1}{N}\mathsf{diag}\big(\nu_{11}X_1'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_1, ..., \nu_{TT}X_T'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_T\big)$, we have

$$\hat{\boldsymbol{C}}'\hat{\boldsymbol{C}} = \hat{\boldsymbol{C}}_1 + \hat{\boldsymbol{C}}_1' - \hat{\boldsymbol{C}}_d. \tag{B.8}$$

By the fact that the eigenvalues of a block upper/lower triangular matrix are the combined eigenvalues of its diagonal block matrices, Weyl's inequality, and Assumptions 1(i) and (ii), we have

$$\begin{aligned} T^{-1}\mu_{\max}(\hat{\boldsymbol{C}}'\hat{\boldsymbol{C}}) &\leq T^{-1}\{2\mu_{\max}(\hat{\boldsymbol{C}}_1) - \mu_{\min}(\hat{\boldsymbol{C}}_d)\} \\ &\leq 2T^{-1}\max_{1\leq t\leq T}\left\|f_t^0\right\|^2 \mu_{\max}\left(N^{-1}X_t'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}X_t\right) \\ &= O_P(T^{-1})O_P(T^{1/4})O_P(1) = O_P(T^{-3/4}), \end{aligned}$$

where we use the fact that $\max_{1\leq t\leq T}\|f_t^0\|^2 = O_P\left(T^{1/4}\right)$ by Assumption 1(i) and the Markov inequality. On the other hand, we note that the minimum eigenvalue of $\hat{\boldsymbol{A}}$ is positive and bounded away from zero w.p.a.1. Hence, the matrix $\hat{\boldsymbol{D}}$ is asymptotically positive definite as its minimum eigenvalue is positive and bounded away from zero w.p.a.1 by using the above facts. Then, by (B.7) and (B.8), we can readily show that there exist two positive constants c_2 and c_3 such that

$$\frac{c_2}{T}\|\hat{\boldsymbol{d}}_\beta\|^2 + c_3\|\hat{\boldsymbol{d}}_*\|^2 \leq \frac{1}{T}\sum_{t=1}^{T}\hat{Q}_{NT,t}^*(\beta_t, \boldsymbol{\Lambda}), \tag{B.9}$$

which indicates that

$$\frac{c_2}{T}\|\hat{\boldsymbol{d}}_\beta\|^2 + c_3\|\hat{\boldsymbol{d}}_*\|^2 + O_P\big(\delta_{NT}^{-2} + \delta_{p,NT}^{-1}\eta_{NT}^{1/2}\big) \leq \hat{Q}_{NT}\big(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Lambda}}\big) - \hat{Q}_{NT}\big(\boldsymbol{\beta}^0, \boldsymbol{\Lambda}^0\big). \tag{B.10}$$

Multiplying both sides of (B.10) by $\delta_{p,NT}^2$ and noting that $\frac{1}{T}\|\hat{\boldsymbol{d}}_\beta\|^2 = \eta_{NT}$ and $\hat{Q}_{NT}\big(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Lambda}}\big) - \hat{Q}_{NT}\big(\boldsymbol{\beta}^0, \boldsymbol{\Lambda}^0\big) \leq 0$, we readily show that

$$c_2\delta_{p,NT}^2\eta_{NT} + O_P(1) + O_P(1)\cdot\left[\delta_{p,NT}^2\eta_{NT}\right]^{1/2} \leq 0. \tag{B.11}$$

When $\delta_{p,NT}^2\eta_{NT}$ is sufficiently large, the first term on the left hand side of (B.11) would dominate the other two terms, which would lead to a contradiction. Hence, we must have that $\delta_{p,NT}^2\eta_{NT}$

is stochastically bounded, implying that $\eta_{NT} = O_P\left(pN^{-1} + T^{-1}\right)$. This completes the proof of Theorem 3.1(i).

(ii) The proof for the point-wise convergence result is similar to the proof of Theorem 3.2(ii) in Qian and Su (2015b), where the condition $\gamma m^0 \Delta_{NT}^{-\kappa} \delta_{p,NT} = O(1)$ in Assumption 2(i) is used to handle the penalty term. We omit the details to save space.

We have thus completed the proof of Theorem 3.1. ■

Proof of Theorem 3.2. To prove the sparsity, it is equivalent to showing

$$\mathsf{P}\left(\left\|\hat{\theta}_t\right\| \neq 0 \text{ for some } t \in \mathcal{T}^c\right) \to 0 \tag{B.12}$$

as $(N,T) \to \infty$. We consider two cases: (i) $2 \leq t \leq T-1$ and $t \in \mathcal{T}^c$; and (ii) $t = T$ and $t \in \mathcal{T}^c$. Recall that $\delta_{p,NT} = \min(p^{-1/2}N^{1/2}, T^{1/2})$.

For case (i), there would be two possible circumstances: (i.1) $t+1 = T_j^0 \in \mathcal{T}_{m^0}^0$ for some $j = 1, ..., m^0$; and (i.2) $t+1 \in \mathcal{T}^c$. We invoke subdifferential calculus (e.g., Bersekas, 1995, Appendix B.5) to obtain the following Karush-Kuhn-Tucker condition with respect to β_t to the objective function in (2.4):

$$\delta_{p,NT}\left[\frac{-2}{N}X_t' \boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\left(Y_t - X_t\hat{\beta}_t\right) + \gamma\dot{w}_t\frac{\hat{\beta}_t - \hat{\beta}_{t-1}}{\left\|\hat{\beta}_t - \hat{\beta}_{t-1}\right\|} - \gamma\dot{w}_{t+1}\frac{\hat{\beta}_{t+1} - \hat{\beta}_t}{\left\|\hat{\beta}_{t+1} - \hat{\beta}_t\right\|}\right] = \mathbf{0}, \tag{B.13}$$

where for any $p \times 1$ vector a with $\|a\| = 0$, $a/\|a\|$ is defined as an arbitrary $p \times 1$ vector with Frobenius norm smaller than or equal to 1. Let $U_{NT,1} = \frac{1}{N}X_t'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\left(Y_t - X_t\hat{\beta}_t\right)$, $U_{NT,2} = \gamma\dot{w}_t\frac{\hat{\beta}_t - \hat{\beta}_{t-1}}{\left\|\hat{\beta}_t - \hat{\beta}_{t-1}\right\|}$ and $U_{NT,3} = \gamma\dot{w}_{t+1}\frac{\hat{\beta}_{t+1} - \hat{\beta}_t}{\left\|\hat{\beta}_{t+1} - \hat{\beta}_t\right\|}$. Following the proof of Theorem 3.1 and using Lemma B.2, we may show that

$$\delta_{p,NT}\|U_{NT,1}\| = O_P(1). \tag{B.14}$$

If circumstance (i.1) holds, by Lemma B.1 and Assumption 2(ii), we have

$$\dot{w}_{t+1} = \|\dot{\beta}_{t+1} - \dot{\beta}_t\|^{-\kappa} \leq \left[\min_{1 \leq j \leq m^0}\left\|\alpha_{j+1}^0 - \alpha_j^0\right\| + O_P\left(\delta_{p,NT}^{-1}\right)\right]^{-\kappa} = O_P(\Delta_{NT}^{-\kappa}), \tag{B.15}$$

which together with Assumption 2(i), indicates that

$$\delta_{p,NT}\|U_{NT,3}\| = O_P(\gamma\delta_{p,NT}\Delta_{NT}^{-\kappa}) = O_P(1). \tag{B.16}$$

However, for case (i) with $2 \leq t \leq T-1$ and $t \in \mathcal{T}^c$, by Lemma B.1, we may show that w.p.a.1

$$\dot{w}_t = \|\dot{\beta}_t - \dot{\beta}_{t-1}\|^{-\kappa} \geq C\delta_{p,NT}^{\kappa}, \tag{B.17}$$

for some positive constant C. Hence, it is not difficult to see that when $\hat{\theta}_t \neq \mathbf{0}$,

$$\delta_{p,NT}\|U_{NT,2}\| \geq C\gamma\delta_{p,NT}^{\kappa+1} \to \infty \tag{B.18}$$

by using Assumption 2(iii). By (B.14), (B.16) and (B.18), the equation (B.13) cannot hold as $(N,T) \to \infty$. Hence, $\hat{\theta}_t$ can only take the value of $\mathbf{0}$ at which $||\hat{\theta}_t||$ is not differentiable. Furthermore, as an implication of the above result, if $t = T_j^0 - 1 \in \mathcal{T}^c$ for some $j = 1, ..., m^0$, then we have

$$\delta_{p,NT}\gamma\dot{w}_t\frac{\hat{\beta}_t - \hat{\beta}_{t-1}}{\|\hat{\beta}_t - \hat{\beta}_{t-1}\|} = \delta_{p,NT}\gamma\dot{w}_{T_j^0-1}\frac{\hat{\beta}_{T_j^0-1} - \hat{\beta}_{T_j^0-2}}{\|\hat{\beta}_{T_j^0-1} - \hat{\beta}_{T_j^0-2}\|} = O_P(1). \tag{B.19}$$

We next prove (B.12) for circumstance (*i.2*). Following the above argument, we can show that when $t = T_j^0 - 2$ and $\hat{\theta}_{T_j^0-2} \neq \mathbf{0}$,

$$\frac{\delta_{p,NT}}{N}X_t'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\left(Y_t - X_t\hat{\beta}_t\right) = O_P(1), \quad \delta_{p,NT}\gamma\dot{w}_t\frac{\hat{\beta}_t - \hat{\beta}_{t-1}}{\|\hat{\beta}_t - \hat{\beta}_{t-1}\|} \to \infty, \tag{B.20}$$

which, together with (B.19), implies that (B.13) cannot hold as $(N,T) \to \infty$. Hence, $\hat{\theta}_{T_j^0-2}$ can only be $\mathbf{0}$. Deducting in this way until we reach $t = T_{j-1}^0 + 1 \in \mathcal{T}^c$, we can complete the proof of sparsity for case (i).

For case (ii), note that the consequence of the Karush-Kuhn-Tucker condition with respect to β_T leads to

$$\delta_{p,NT}\left[\frac{1}{N}X_T'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\left(Y_T - X_T\hat{\beta}_T\right) + \gamma\dot{w}_T\frac{\hat{\beta}_T - \hat{\beta}_{T-1}}{\|\hat{\beta}_T - \hat{\beta}_{T-1}\|}\right] = \mathbf{0}. \tag{B.21}$$

As there is only one penalty term in (B.21), the proof is much simpler than that for case (i). Hence, we omit the details here.

We have completed the proof of Theorem 3.2. ■

Proof of Corollary 3.3. By Theorem 3.2, as $(N,T) \to \infty$, no time point in $\mathcal{T}^c$ can be identified as the break time, which implies that $\hat{m} \leq m^0$. On the other hand, by Theorem 3.1, for any $t \in \mathcal{T}_{m^0}^0$,

$$\left\|\hat{\theta}_t\right\| = \left\|\hat{\beta}_t - \hat{\beta}_{t-1}\right\| = \|\beta_t^0 - \beta_{t-1}^0\| + O_P(\delta_{p,NT}^{-1}) = \|\theta_t^0\| + O_P(\delta_{p,NT}^{-1}),$$

which indicates that $\|\theta_t^0\| = O_P(\delta_{p,NT}^{-1})$ if $\hat{\theta}_t = \mathbf{0}$ (i.e., $t \in \mathcal{T}_{m^0}^0$ is not identified as a break point). However, the conclusion $\|\theta_t^0\| = O_P\left(\delta_{p,NT}^{-1}\right)$ would violate the condition $\delta_{p,NT}\Delta_{NT} \to \infty$ which is assumed in Assumption 2(ii). Hence, each time point in $\mathcal{T}^0_{m^0}$ must be identified as the break time, which implies that $\hat{m} = m^0$ w.p.a.1 and thus both the results (i) and (ii) are proved. ■

To prove the asymptotic distribution theory for the post-LASSO estimator in Theorem 3.4, we need to use the following lemma whose proof is given in Appendix C of the supplemental document. Let $\tilde{\mathbf{\Lambda}}_{m^0} \equiv \tilde{\mathbf{\Lambda}}(\mathcal{T}^0_{m^0})$ be the infeasible estimator of the factor loadings in the post-LASSO estimation procedure, $\tilde{\boldsymbol{H}} = \left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)\left(\frac{1}{N}\mathbf{\Lambda}^{0\prime}\,\tilde{\mathbf{\Lambda}}_{m^0}\right)\tilde{\boldsymbol{V}}^+_{NT}$, and $\tilde{\boldsymbol{\alpha}}_{m^0} \equiv \tilde{\boldsymbol{\alpha}}_{m^0}(\mathcal{T}^0_{m^0})$, where $\tilde{\boldsymbol{V}}_{NT}$ will be defined later in (B.25).

Lemma B.3 *Suppose that the conditions in Theorem 3.4 hold. Then,*

(i) for each $j = 1, ..., m^0 + 1$, we have

$$\left\|\frac{1}{N\tau_j(T)}\sum_{t=T^0_{j-1}}^{T^0_j-1} X_t'\left(\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}_{m^0}} - \boldsymbol{M}_{\mathbf{\Lambda}^0}\right)\varepsilon_t + B_{NT,j}(2,1)\right\| = O_P\left(\delta^{-1}_{NT}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m_0} - \boldsymbol{\alpha}^0\| + \delta^{-3}_{p,NT}\right),$$

where $\tau_j(T)$ and $B_{NT,j}(2,1)$ are defined as in Theorem 3.4;

(ii) for each $j = 1, ..., m^0 + 1$, we have

$$\Bigg\|\frac{1}{N\tau_j(T)}\sum_{t=T^0_{j-1}}^{T^0_j-1} X_t'\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}_{m^0}}\left(\mathbf{\Lambda}^0 - \tilde{\mathbf{\Lambda}}_{m^0}\tilde{\boldsymbol{H}}^+\right)f^0_t + \frac{1}{N\tau_j(T)}\sum_{t=T^0_{j-1}}^{T^0_j-1} X_t'\boldsymbol{M}_{\mathbf{\Lambda}^0}\varepsilon^*_t + B_{NT,j}(1)$$
$$-B_{NT,j}(2,2) + \left(\Phi^*_{j1}, ..., \Phi^*_{j,m^0+1}\right)\left(\tilde{\boldsymbol{\alpha}}_{m^0} - \boldsymbol{\alpha}^0\right)\Bigg\| = O_P\left(p\delta^{-1}_{p,NT}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m_0} - \boldsymbol{\alpha}^0\| + \delta^{-3}_{p,NT}\right),$$

where $\varepsilon^*_t = \frac{1}{T}\sum_{s=1}^T \chi_{st}\varepsilon_s$, Φ^*_{jk}, $1 \le j, k \le m^0 + 1$, are defined at the beginning of Appendix A, and $B_{NT,j}(1)$ and $B_{NT,j}(2,2)$, $j = 1, ..., m^0 + 1$, are defined as in Theorem 3.4.

We are now ready to prove Theorem 3.4.

Proof of Theorem 3.4. Let $\mathcal{G}_\mathcal{T} = \left\{\hat{T}_j = T^0_j \text{ for } j = 1, ..., m^0\right\}$. By Corollary 3.3, we readily have

$$\begin{aligned}
&\mathsf{P}\left\{\mathbf{SD}_{NT}\left(\tilde{\boldsymbol{\alpha}}_{\hat{m}} - \boldsymbol{\alpha}^0\right) \in \mathcal{C}\middle|\hat{m} = m^0\right\} \\
=\ &\mathsf{P}\left\{\mathbf{SD}_{NT}\left(\tilde{\boldsymbol{\alpha}}_{\hat{m}} - \boldsymbol{\alpha}^0\right) \in \mathcal{C}, \mathcal{G}_\mathcal{T}\middle|\hat{m} = m^0\right\} + \mathsf{P}\left\{\mathbf{SD}_{NT}\left(\tilde{\boldsymbol{\alpha}}_{\hat{m}} - \boldsymbol{\alpha}^0\right) \in \mathcal{C}, \mathcal{G}^c_\mathcal{T}\middle|\hat{m} = m^0\right\} \\
=\ &\mathsf{P}\left\{\mathbf{SD}_{NT}\left(\tilde{\boldsymbol{\alpha}}_{m^0} - \boldsymbol{\alpha}^0\right) \in \mathcal{C}\right\} + o(1), \qquad \text{(B.22)}
\end{aligned}$$

where $\mathcal{C} \subset \mathbb{R}^{k^0}$, $\mathcal{G}^c_\mathcal{T}$ is the complement of $\mathcal{G}_\mathcal{T}$ and $\tilde{\boldsymbol{\alpha}}_{m^0} = \tilde{\boldsymbol{\alpha}}_{m^0}(\mathcal{T}_{m^0})$ is the infeasible estimate of $\boldsymbol{\alpha}^0$. Hence, throughout the proof, we can replace $\hat{m}$ and $\hat{T}_j$ $(j = 1, ..., \hat{m})$ by m^0 and T^0_j, respectively, which would not affect the asymptotic distribution of the post-LASSO estimator.

Letting $m = m^0$ and $T_j = T_j^0$ in the objective function (3.1), we have

$$Q_{NT}\big(\boldsymbol{\alpha}_{m^0}, \boldsymbol{\Lambda}, \boldsymbol{F}; \mathcal{T}_{m^0}^0\big) = \frac{1}{NT}\sum_{j=1}^{m^0+1}\sum_{t=T_{j-1}^0}^{T_j^0-1}\big(Y_t - X_t\alpha_j - \boldsymbol{\Lambda}f_t\big)'\big(Y_t - X_t\alpha_j - \boldsymbol{\Lambda}f_t\big),$$

and

$$\min_{\boldsymbol{F}} Q_{NT}\big(\boldsymbol{\alpha}_{m^0}, \boldsymbol{\Lambda}, \boldsymbol{F}; \mathcal{T}_{m^0}^0\big) = \frac{1}{NT}\sum_{j=1}^{m^0+1}\sum_{t=T_{j-1}^0}^{T_j^0-1}\big(Y_t - X_t\alpha_j\big)'\boldsymbol{M}_{\boldsymbol{\Lambda}}\big(Y_t - X_t\alpha_j\big). \qquad \text{(B.23)}$$

Recall that $\tilde{\boldsymbol{\Lambda}}_{m^0} = \tilde{\boldsymbol{\Lambda}}(\mathcal{T}_{m^0}^0)$ which is defined as in Lemma B.3. Let

$$\begin{aligned}
\tilde{\boldsymbol{\Phi}}_{NT}(\tilde{\boldsymbol{\Lambda}}_{m^0}) &= \mathsf{diag}\left\{\tilde{\Phi}_1(\tilde{\boldsymbol{\Lambda}}_{m^0}), ..., \tilde{\Phi}_{m^0+1}(\tilde{\boldsymbol{\Lambda}}_{m^0})\right\} \quad \text{and} \\
\tilde{\boldsymbol{\Xi}}_{NT}(\tilde{\boldsymbol{\Lambda}}_{m^0}) &= \left[\tilde{\Xi}_1(\tilde{\boldsymbol{\Lambda}}_{m^0})', ..., \tilde{\Xi}_{m^0+1}(\tilde{\boldsymbol{\Lambda}}_{m^0})'\right]',
\end{aligned}$$

where $\tilde{\Phi}_j(\tilde{\boldsymbol{\Lambda}}_{m^0}) = \frac{1}{N\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_{m^0}}X_t$, and $\tilde{\Xi}_j(\tilde{\boldsymbol{\Lambda}}_{m^0}) = \frac{1}{N\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_{m^0}}Y_t$ for $j = 1, ..., m^0 + 1$. Then, the solution $\big(\tilde{\boldsymbol{\alpha}}_{m^0},\ \tilde{\boldsymbol{\Lambda}}_{m^0}\big)$ to the minimization of the objective function in (B.23) satisfies

$$\tilde{\boldsymbol{\alpha}}_{m^0} = \tilde{\boldsymbol{\Phi}}_{NT}^{+}(\tilde{\boldsymbol{\Lambda}}_{m^0})\tilde{\boldsymbol{\Xi}}_{NT}(\tilde{\boldsymbol{\Lambda}}_{m^0}) \ \text{ with } \ \tilde{\alpha}_{m^0j} = \tilde{\Phi}_j^{+}(\tilde{\boldsymbol{\Lambda}}_{m^0})\tilde{\Xi}_j(\tilde{\boldsymbol{\Lambda}}_{m^0}), \qquad \text{(B.24)}$$

and

$$\left[\frac{1}{NT}\sum_{j=1}^{m^0+1}\sum_{t=T_{j-1}}^{T_j-1}\big(Y_t - X_t\tilde{\alpha}_{m^0j}\big)\big(Y_t - X_t\tilde{\alpha}_{m^0j}\big)'\right]\tilde{\boldsymbol{\Lambda}}_{m^0} = \tilde{\boldsymbol{\Lambda}}_{m^0}\tilde{\boldsymbol{V}}_{NT}, \qquad \text{(B.25)}$$

where $\tilde{\alpha}_{m^0j}$ is the j-th p-dimensional element of $\tilde{\boldsymbol{\alpha}}_{m^0}$ and $\tilde{\boldsymbol{V}}_{NT}$ is a diagonal matrix consisting of the R_0 largest eigenvalues of the above matrix in the square brackets in (B.25) arranged in descending order.

To simplify the notation, we further let $\tilde{\boldsymbol{\Lambda}} \equiv \tilde{\boldsymbol{\Lambda}}_{m^0}$ in the remaining proof when no confusion can arise. For $j = 1, ..., m^0 + 1$, using the expression that $Y_t = X_t\alpha_j^0 + \boldsymbol{\Lambda}^0 f_t^0 + \varepsilon_t$ for $t \in$

$[T_{j-1}^0, T_j^0 - 1]$ and the fact that $\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}\tilde{\boldsymbol{\Lambda}} = \mathbf{0}$, we have

$$\begin{aligned}
\tilde{\Xi}_j(\tilde{\boldsymbol{\Lambda}}) &= \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \big(X_t \alpha_j^0 + \boldsymbol{\Lambda}^0 f_t^0 + \varepsilon_t \big) \\
&= \Big[\frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} X_t \Big] \alpha_j^0 + \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \big(\boldsymbol{\Lambda}^0 - \tilde{\boldsymbol{\Lambda}} \tilde{\boldsymbol{H}}^+ \big) f_t^0 \\
&\quad + \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \varepsilon_t.
\end{aligned}$$

Plugging the above expression into the formula of $\tilde{\alpha}_{m^0 j}$ in (B.24) yields

$$\tilde{\Phi}_j(\tilde{\boldsymbol{\Lambda}}) \big(\tilde{\alpha}_{m^0 j} - \alpha_j^0 \big) = \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \big(\boldsymbol{\Lambda}^0 - \tilde{\boldsymbol{\Lambda}} \tilde{\boldsymbol{H}}^+ \big) f_t^0 + \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \varepsilon_t. \quad \text{(B.26)}$$

We first consider the second term on the right hand side of (B.26). By Lemma B.3(i),

$$\begin{aligned}
&\Big\| \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \varepsilon_t - \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\boldsymbol{\Lambda}^0} \varepsilon_t + B_{NT,j}(2,1) \Big\| \\
&= O_P\Big(\delta_{NT}^{-1} (m^0)^{-1/2} \|\tilde{\boldsymbol{\alpha}}_{m_0} - \boldsymbol{\alpha}^0\| + \delta_{p,NT}^{-3} \Big) \qquad \text{(B.27)}
\end{aligned}$$

for each $j = 1, ..., m^0 + 1$. On the other hand, for the first term on the right hand side of (B.26), by Lemma B.3(ii), we have

$$\begin{aligned}
&\Big\| \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_{m^0}} \big(\boldsymbol{\Lambda}^0 - \tilde{\boldsymbol{\Lambda}}_{m^0} \tilde{\boldsymbol{H}}^+ \big) f_t^0 + \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\boldsymbol{\Lambda}^0} \varepsilon_t^* + B_{NT,j}(1) - B_{NT,j}(2,2) \\
&+ \big(\Phi_{j1}^*, ..., \Phi_{j,m^0+1}^* \big) \big(\tilde{\boldsymbol{\alpha}}_{m^0} - \boldsymbol{\alpha}^0 \big) \Big\| = O_P\Big(\delta_{NT}^{-1} (m^0)^{-1/2} \|\tilde{\boldsymbol{\alpha}}_{m_0} - \boldsymbol{\alpha}^0\| + \delta_{p,NT}^{-3} \Big), \qquad \text{(B.28)}
\end{aligned}$$

Recall that $\boldsymbol{\Omega}_{NT} = \boldsymbol{\Phi}_{NT} - \boldsymbol{\Phi}_{NT}^*$ with $\boldsymbol{\Phi}_{NT}$ and $\boldsymbol{\Phi}_{NT}^*$ defined at the beginning of Appendix A. Then, using the definitions of $\boldsymbol{B}_{NT}(1)$ and $\boldsymbol{B}_{NT}(2)$ in Section 3.2, the definition of $\boldsymbol{W}_{NT}$ in Assumption 3(iv), the condition $(NT)^{1/2} = o(\delta_{p,NT}^3)$ in Assumption 3(v) as well as (B.26)–(B.28), we have

$$\big\| \mathbf{SD}_{NT} \big[\boldsymbol{\Omega}_{NT} (\tilde{\boldsymbol{\alpha}}_{m^0} - \boldsymbol{\alpha}^0) + \boldsymbol{B}_{NT}(1) + \boldsymbol{B}_{NT}(2) - \boldsymbol{W}_{NT} \big] \big\| = o_P(1). \qquad \text{(B.29)}$$

Furthermore, by Assumptions 3(i)(iv) and noting that $\mathbf{\Omega}_0$ is positive definite, we have

$$\mathbf{SD}_{NT}\left[\tilde{\boldsymbol{\alpha}}_{m^0} - \boldsymbol{\alpha}^0 + \boldsymbol{B}_{NT}\right] \xrightarrow{D} \mathsf{N}\left(\mathbf{0},\ \mathbf{S}\mathbf{\Omega}_0^+\mathbf{\Omega}_1\mathbf{\Omega}_0^+\mathbf{S}'\right),$$

where $\boldsymbol{B}_{NT} = \mathbf{\Omega}_{NT}^+\left[\boldsymbol{B}_{NT}(1) + \boldsymbol{B}_{NT}(2) - \boldsymbol{B}_{NT}(3)\right]$. We have thus completed the proof of Theorem 3.4. ■

To prove Theorem 4.1 in Section 4.1, we need the following lemma whose proof is given in Appendix C of the supplemental document.

Lemma B.4 *Suppose that the conditions in Theorem 4.1 hold. Then*

(i) there exists a $c_R > 0$ *such that plim inf*$_{(N,T)\to\infty}\left[V(R,\dot{\boldsymbol{\beta}}_R) - V(R_0,\dot{\boldsymbol{\beta}}_{R_0})\right] \geq c_R$ *for each* R *with* $1 \leq R < R_0$,

(ii) $V(R,\dot{\boldsymbol{\beta}}_R) - V(R_0,\dot{\boldsymbol{\beta}}_{R_0}) = O_P\left(\delta_{p,NT}^{-2}\right)$ *for each* R *with* $R \geq R_0$.

Proof of Theorem 4.1. The proof is analogous to that of Corollary 1 in Bai and Ng (2002). For notational simplicity, let $V(R) = V(R,\dot{\boldsymbol{\beta}}_R)$ for all R. Note that

$$\mathsf{BIC}(R) - \mathsf{BIC}(R_0) = \ln\left[V(R)/V(R_0)\right] + (R - R_0)\,\rho_1.$$

We discuss the following two cases: (a) $R < R_0$, and (b) $R_0 < R \leq R_{\max}$.

For case (a), by Lemma B.4(i), $V(R)/V(R_0) > 1+\epsilon_0$ and thus $\ln\left[V(R)/V(R_0)\right] \geq \epsilon_0/2$ for some $\epsilon_0 > 0$ w.p.a.1. This, in conjunction with the fact that $(R - R_0)\,\rho_1 \to 0$ under Assumption 4, implies that $\mathsf{BIC}(R) - \mathsf{BIC}(R_0) \geq \epsilon_0/4$ w.p.a.1. It follows that

$$\mathsf{P}\left(\mathsf{BIC}(R) - \mathsf{BIC}(R_0) > 0\right) \to 1$$

as $(N,T) \to \infty$ for any $R < R_0$.

For case (b), we apply Lemma B.4(ii) and Assumption 4 to obtain

$$\begin{aligned}\mathsf{P}\left(\mathsf{BIC}(R) - \mathsf{BIC}(R_0) > 0\right) &= \mathsf{P}\left(\ln\left[V(R)/V(R_0)\right] + (R - R_0)\,\rho_1 > 0\right) \\ &= \mathsf{P}\left(O_P(1) + (R - R_0)\,\rho_1\delta_{p,NT}^2 > 0\right) \to 1\end{aligned}$$

as $(N,T) \to \infty$ for any $R_0 < R \leq R_{\max}$. Consequently, the minimizer of $\mathsf{BIC}(R)$ can only be achieved at $R = R_0$ w.p.a.1. That is, $\mathsf{P}(\hat{R} = R_0) \to 1$ for any $R \in [1, R_{\max}]$ as $(N,T) \to \infty$. ■

Let $\mathbb{T}_m$ consist of $\mathcal{T}_m = \{T_1, ..., T_m\}$ such that $2 \leq T_1 < ... < T_m \leq T$, $T_0 = 1$ and $T_{m+1} = T + 1$; and let $\bar{\mathbb{T}}_m$ consist of $\mathcal{T}_m = \{T_1, ..., T_m\}$ such that $\mathcal{T}_{m^0} \subset \mathcal{T}_m$, $2 \leq T_1 < ... < T_m \leq T$ for $m^0 < m \leq m_{\max}$. To prove Theorem 4.2, we need the following two useful lemmas.

Lemma B.5 *Suppose that the conditions in Theorem 4.2 hold. Then there exists a positive constant c_m such that*

$$\min_{0\leq m<m^0}\inf_{\mathcal{T}_m\in\mathbb{T}_m}\frac{m^0}{T\Delta_{NT}^2}\left[\tilde{\sigma}^2(\mathcal{T}_m)-\tilde{\sigma}^2(\mathcal{T}_{m^0}^0)\right]\geq c_m+o_P(1).$$

Lemma B.6 *Suppose that the conditions in Theorem 4.2 hold. Then we have*

$$\max_{m^0<m\leq m_{\max}}\sup_{\mathcal{T}_m\in\bar{\mathbb{T}}_m}\delta_{p,NT}^2\left|\tilde{\sigma}^2(\mathcal{T}_m)-\tilde{\sigma}^2(\mathcal{T}_{m^0}^0)\right|=O_P(1).$$

Proof of Theorem 4.2. Denote $\Gamma=[0,\gamma_{\max}]$, a bounded interval in $\mathbb{R}^+$, which is divided into three subsets Γ_0, Γ_- and Γ_+ as follows

$$\Gamma_0=\left\{\gamma\in\Gamma:\hat{m}_\gamma=m^0\right\},\ \Gamma_-=\left\{\gamma\in\Gamma:\hat{m}_\gamma<m^0\right\},\ \text{and}\ \Gamma_+=\left\{\gamma\in\Gamma:\hat{m}_\gamma>m^0\right\}.$$

Clearly, Γ_0, Γ_- and Γ_+ denote the three subsets of Γ in which the correct-, under- and over-number of breaks are selected by the AGF-LASSO procedure, respectively. Recall that $\tilde{\boldsymbol{\alpha}}_{\hat{m}_\gamma}=(\tilde{\alpha}_1(\hat{\mathcal{T}}_{\hat{m}_\gamma})', ..., \tilde{\alpha}_{\hat{m}_\gamma+1}(\hat{\mathcal{T}}_{\hat{m}_\gamma})')'$ and $\tilde{\mathbf{\Lambda}}(\hat{\mathcal{T}}_{\hat{m}_\gamma})$ denote the post-LASSO estimators of the regression coefficients and factor loadings based on the break dates in $\hat{\mathcal{T}}_{\hat{m}_\gamma}=\hat{\mathcal{T}}_{\hat{m}_\gamma}(\gamma)=(\hat{T}_1(\gamma),...,\hat{T}_{\hat{m}_\gamma}(\gamma))$, where we make the dependence of various estimates on γ explicit. Recall that $\tilde{\sigma}^2(\hat{\mathcal{T}}_{\hat{m}_\gamma})=Q_{NT}(\tilde{\alpha}_{\hat{m}_\gamma},\tilde{\mathbf{\Lambda}}(\hat{\mathcal{T}}_{\hat{m}_\gamma});\hat{\mathcal{T}}_{\hat{m}_\gamma})$. Let $\gamma^0\equiv\gamma_{NT}^0$ denote an element in Γ_0 that also satisfies the conditions on γ in Assumptions 2(i) and (iii), and let $\hat{T}_j(\gamma^0)$ be the AGF-LASSO estimate of the true break date T_j^0 using the tuning parameter γ^0. For any $\gamma^0\in\Gamma_0$, we have $\hat{m}_{\gamma^0}=m^0$ w.p.a.1, and by Corollary 3.3,

$$\lim_{(N,T)\to\infty}\mathsf{P}\left(\hat{T}_j(\gamma^0)=T_j^0,\quad j=1,...,m^0\right)=1.$$

It follows that w.p.a.1 $\tilde{\sigma}^2(\hat{\mathcal{T}}_{\hat{m}_{\gamma^0}})=\tilde{\sigma}^2(\mathcal{T}_{m^0}^0)$. By the proof of Lemma B.5 in Appendix C of the supplemental document,

$$\begin{aligned}\tilde{\sigma}^2(\mathcal{T}_{m^0}^0) &= \frac{1}{NT}\sum_{j=1}^{m^0+1}\sum_{t=T_{j-1}^0+1}^{T_j^0-1}\left[Y_t-X_t\tilde{\alpha}_j(\mathcal{T}_{m^0})\right]'\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}}\left[Y_t-X_t\tilde{\alpha}_j(\mathcal{T}_{m^0})\right]\\ &= \frac{1}{NT}\sum_{t=1}^{T}\varepsilon_t'\varepsilon_t+O_P(\delta_{p,NT}^{-2})\xrightarrow{P}\sigma_0^2,\end{aligned}$$

where $\sigma_0^2\equiv\lim_{(N,T)\to\infty}\frac{1}{NT}\sum_{t=1}^T\mathsf{E}\left[\varepsilon_t'\varepsilon_t\right]$. Thus $\tilde{\sigma}^2(\mathcal{T}_{m^0}^0)\xrightarrow{P}\sigma_0^2$ and $\mathsf{IC}\left(\gamma^0\right)=\ln(\tilde{\sigma}^2(\mathcal{T}_{m^0}^0))+\rho_2p(m^0+1)\xrightarrow{P}\ln(\sigma_0^2)$ as $\rho_2p\left(m^0+1\right)=o(1)$ by Assumption 5(iii). We next consider the cases of under- and over-fitted models separately.

Case 1 (Under-fitted model with $\hat{m}_\gamma < m^0$): By Lemma B.5 and Assumption 5(iii),

$$\mathsf{P}\left(\inf_{\gamma\in\Gamma_-}\mathsf{IC}\left(\gamma\right)>\mathsf{IC}\left(\gamma^0\right)\right)=\mathsf{P}\left(\inf_{\gamma\in\Gamma_-}\frac{m^0}{T\Delta_{NT}^2}\left[\ln\left(\tilde{\sigma}^2(\tilde{\mathcal{T}}_{\hat{m}_\gamma})/\tilde{\sigma}^2(\mathcal{T}_{m^0}^0)\right)+\rho_2 p\left(\hat{m}_\gamma-m^0\right)\right]>0\right)$$
$$\geq \mathsf{P}\left(c/2+o_P\left(1\right)>0\right)\to 1,$$

where c is a positive constant.

Case 2 (Over-fitted model with $\hat{m}_\gamma > m^0$): For given $\mathcal{T}_m=\{T_1,...,T_m\}\in\mathbb{T}_m$, we let $\bar{\mathcal{T}}_{m^*+m^0}=\left\{\bar{T}_1,\bar{T}_2,...,\bar{T}_{m^*+m^0}\right\}$ denote the union of $\mathcal{T}_m$ and $\mathcal{T}_{m^0}^0$ with elements ordered in non-descending order: $2\leq\bar{T}_1<\bar{T}_2<\cdots<\bar{T}_{m^*+m^0}\leq T$ for some $m^*\in\{0,1,...,m\}$. Let

$$\left(\tilde{\boldsymbol{\alpha}}_m(\mathcal{T}_m),\tilde{\boldsymbol{\Lambda}}\left(\mathcal{T}_m\right)\right)=\arg\min_{(\boldsymbol{\alpha}_m,\boldsymbol{\Lambda})}Q_{NT}\left(\boldsymbol{\alpha}_m,\boldsymbol{\Lambda};\mathcal{T}_m\right)$$

subject to $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}/N=\boldsymbol{I}_{R^0}$. Let $\tilde{\sigma}^2(\mathcal{T}_m)\equiv Q_{NT}(\tilde{\boldsymbol{\alpha}}_m(\mathcal{T}_m),\tilde{\boldsymbol{\Lambda}}\left(\mathcal{T}_m\right);\mathcal{T}_m)$ and let $\tilde{\sigma}^2(\bar{\mathcal{T}}_{m^*+m^0})$ be analogously defined. In view of the fact that $\tilde{\sigma}^2(\bar{\mathcal{T}}_{m^*+m^0})\leq\tilde{\sigma}^2(\mathcal{T}_m)$ for all $\mathcal{T}_m\in\mathbb{T}_m$,

$$\delta_{p,NT}^2\left[\tilde{\sigma}^2(\bar{\mathcal{T}}_{m^*+m^0})-\bar{\sigma}_{NT}^2\right]=O_P(1)$$

uniformly in $\mathcal{T}_m\in\mathbb{T}_m$ by Lemma B.6, and $\delta_{p,NT}^2 p\rho_2\to\infty$ by Assumption 5(iii), we have

$$\begin{aligned}
&\mathsf{P}\left(\inf_{\gamma\in\Gamma_+}\mathsf{IC}(\gamma)>\mathsf{IC}(\gamma^0)\right)\\
\geq\ &\mathsf{P}\left(\min_{m^0<m\leq m_{\max}}\inf_{\mathcal{T}_m\in\mathbb{T}_m}\left\{\delta_{p,NT}^2\left[\ln\left(\tilde{\sigma}^2(\mathcal{T}_m)/\tilde{\sigma}^2(\mathcal{T}_{m^0})\right)\right]+\delta_{p,NT}^2\rho_2 p\left(m-m^0\right)\right\}>0\right)\\
\geq\ &\mathsf{P}\left(\min_{m^0<m\leq m_{\max}}\inf_{\mathcal{T}_m\in\mathbb{T}_m}\left\{\delta_{p,NT}^2\left[\ln\left(\tilde{\sigma}^2(\bar{\mathcal{T}}_{m^*+m^0})/\tilde{\sigma}^2(\mathcal{T}_{m^0})\right)\right]+\delta_{p,NT}^2\rho_2 p\left(m-m^0\right)\right\}>0\right)\\
\to\ &1.
\end{aligned}$$

We have completed the proof of Theorem 4.2 ■

References

Ang, J.B. (2007), "CO_2 emissions, energy consumption,and output in France," *Energy Policy,* 35, 4772-4778.

Apergis, N., Payne, J.E. (2009), "CO_2 emissions, energy usage,and output in Central America," *Energy Policy,* 37, 3282-3286.

Apergis, N., Payne, J.E. (2010), "The emissions, energy consumption,and growth nexus: evidence from the common wealth of independent states," *Energy Policy,* 38, 650-655.

Arouri, M.H., Youssef, A.B., M'Henni, H., Rault, C., 2012, "Energy consumption, economic growth and CO_2 emissions in Middle East and North African countries," *Energy Policy,* 45, 342-349.

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

Bai, J. (2010), "Common breaks in means and variances for panel data," *Journal of Econometrics,* 157, 78-92.

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

Bai, J., Ng, S. (2002), "Determining the number of factors in approximate factor models," *Econometrica,* 70, 191–221.

Bai, J., Perron, P. (1998), "Estimating and testing liner models with multiple structural changes," *Econometrica,* 66, 47-78.

Baltagi, B. H., Feng, Q., Kao, C. (2015), "Estimation of heterogeneous panels with structural breaks," *Journal of Econometrics*, forthcoming.

Chan, N. H., Yau, C. Y., Zhang, R. (2014), "Group LASSO for structural break time series," *Journal of American Statistical Association,* 109, 590-599.

Csörgö, M., Horváth, L. (1997), *Limit Theorems in Change-Point Analysis.* Wiley Series in Probability and Statistics.

De Watcher, S., Tzavalis, E. (2012), "Detection of structural breaks in linear dynamic panel data models," *Computational Statistics and Data Analysis,* 56, 3020-3034.

Fan, J., Li, R. (2001), "Variable selection via nonconcave penalized likelihood and it oracle properties," *Journal of American Statistical Association,* 96, 1348-1360.

Fan, J., Li, R. (2006), "Statistical challenges with high dimensionality: feature selection in knowledge discovery," In: Sanz-Sole, M., Soria, J., Varona, J. L., Verdera, J. (eds.) *Proceedings of the International Congress of Mathematicians,* Vol. III, European Mathematical Society, Zurich, pp. 595-622.

Farhani, S., Mrizak, S., Chaibi, A., Rault, C. (2014)), "The environmental Kuznets curve and sustainability: A panel data analysis," *Energy Policy* 71, 189-198.

Hallin, M., Liska, R. (2007), "Determining the number of factors in the general dynamic factor model," *Journal of the American Statistical Association,* 102, 603-617.

Harchaoui, Z., Lévy-Leduc, C. (2010), "Multiple change-point estimation with a total variation penalty," *Journal of the American Statistical Association,* 105, 1480-1493.

Kim, D. (2011), "Estimating a common deterministic time trend break in large panels with cross sectional dependence," *Journal of Econometrics,* 164, 310-330.

Lean, H.H., Smyth, R. (2010), "CO_2 emissions, electricity consumption and output in ASEAN," *Applied Energy* 87,1858-1864.

Lu, X., Su, L. (2015), "Shrinkage estimation of dynamic panel data models with interactive fixed effects," *Working paper*, Singapore Management University.

Moon, H. R. Weidner, M. (2014), "Dynamic linear panel data regression models with interactive fixed effects," *Working paper*, University of Southern California.

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

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

Qian, J., Su, L. (2015a), "Shrinkage estimation of regression models with multiple structural changes," *Econometric Theory*, forthcoming.

Qian, J., Su, L. (2015b), "Shrinkage estimation of common breaks in panel data models via adaptive group fused Lasso," *Working paper*, Singapore Management University.

Qu, Z., Perron, P. (2007), "Estimating and testing structural changes in multiple regressions," *Econometrica,* 75, 459-502.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K. (2005), "Sparsity and smoothness via the fused Lasso," *Journal of the Royal Statistical Society, Series B,* 67, 91-108.

Rinaldo, A. (2009), "Properties and refinement of the fused Lasso," *Annals of Statistics,* 37, 2922-2952.

Yuan, M., Lin, Y. (2006), "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society, Series B,* 68, 49-67.

Zou, H. (2006), "The adaptive Lasso and its oracle properties," *Journal of the American Statistical Association,* 101, 1418-1429.

Supplementary Material for

"Panel Data Models with Interactive Fixed Effects and Multiple Structural Breaks"

This supplemental document provides the proofs of all the technical lemmas in Appendix B of the main document.

C Proofs of the technical lemmas

In this appendix we give the detailed proofs of the technical lemmas used in Appendix B. Before proving Lemma B.1 on the convergence rates of $\dot{\beta}_t$, we give some preliminary results. Let $\boldsymbol{b} = (b_1', b_2', ..., b_T')'$ where b_t is a p-dimensional column vector and let C be a positive constant whose value may change from line to line. Recall that $\delta_{NT} = \min(\sqrt{N}, \sqrt{T})$.

Lemma C.1 *Suppose that Assumption 1 in Appendix A holds. Then we have*

(i) $\sup_{\boldsymbol{b}} \sup_{\boldsymbol{\Lambda}} \left| \frac{1}{NT} \sum_{t=1}^T b_t' X_t' \boldsymbol{M}_{\boldsymbol{\Lambda}} \varepsilon_t \right| = O_P(pN^{-1/2} + p^{1/2} T^{-1/2})$,

(ii) $\sup_{\boldsymbol{\Lambda}} \left| \frac{1}{NT} \sum_{t=1}^T f_t^{0\prime} \boldsymbol{\Lambda}^{0\prime} \boldsymbol{M}_{\boldsymbol{\Lambda}} \varepsilon_t \right| = O_P(\delta_{NT}^{-1})$,

(iii) $\sup_{\boldsymbol{\Lambda}} \left| \frac{1}{NT} \sum_{t=1}^T \varepsilon_t' \boldsymbol{P}_{\boldsymbol{\Lambda}} \varepsilon_t \right| = O_P(\delta_{NT}^{-2})$,

(iv) $\frac{1}{NT} \sum_{t=1}^T \varepsilon_t' \boldsymbol{P}_{\boldsymbol{\Lambda}^0} \varepsilon_t = O_P(N^{-1})$,

where $\sup_{\boldsymbol{b}}$ *is taken with respect to* $\boldsymbol{b}$ *such that* $\|\boldsymbol{b}\| \le C(pT)^{1/2}$ *and* $\sup_{\boldsymbol{\Lambda}}$ *is taken with respect to* $\boldsymbol{\Lambda}$ *such that* $\frac{1}{N}\boldsymbol{\Lambda}'\boldsymbol{\Lambda} = \boldsymbol{I}_{R_0}$.

Proof of Lemma C.1. **(i)** Note that $\frac{1}{NT} \sum_{t=1}^T b_t' X_t' \boldsymbol{M}_{\boldsymbol{\Lambda}} \varepsilon_t = \frac{1}{NT} \sum_{t=1}^T b_t' X_t' \varepsilon_t - \frac{1}{N^2 T} \sum_{t=1}^T b_t' X_t' \boldsymbol{\Lambda}\boldsymbol{\Lambda}' \varepsilon_t$ if $\frac{1}{N}\boldsymbol{\Lambda}'\boldsymbol{\Lambda} = \boldsymbol{I}_{R_0}$. By Assumption 1(iii) and the Cauchy-Schwarz inequality, we have

$$\big|\sum_{t=1}^T b_t' X_t' \varepsilon_t\big| = \big(\sum_{t=1}^T \|b_t\|^2\big)^{1/2} \cdot \big(\sum_{t=1}^T \|X_t'\varepsilon_t\|^2\big)^{1/2} = O_P\big(pTN^{1/2}\big) \tag{C.1}$$

for $\|\boldsymbol{b}\|^2 = \sum_{t=1}^T \|b_t\|^2 \le CpT$. On the other hand, by some elementary calculations, we have

$$\begin{aligned} \big|\sum_{t=1}^T b_t' X_t' \boldsymbol{\Lambda}\boldsymbol{\Lambda}' \varepsilon_t\big| &\le \sum_{t=1}^T \big|b_t' X_t' \boldsymbol{\Lambda}\boldsymbol{\Lambda}' \varepsilon_t\big| \le \max_{1\le t\le T} \|X_t'\boldsymbol{\Lambda}\| \sum_{t=1}^T \|b_t\| \|\boldsymbol{\Lambda}'\varepsilon_t\| \\ &\le \max_{1\le t\le T} \|X_t'\boldsymbol{\Lambda}\| \Big(\sum_{t=1}^T \|b_t\|^2\Big)^{1/2} \Big(\sum_{t=1}^T \|\boldsymbol{\Lambda}'\varepsilon_t\|^2\Big)^{1/2}. \end{aligned}$$

By the restriction on $\mathbf{\Lambda}$ and Assumption 1(ii), we have

$$\max_{1\le t\le T}\left\|X_t'\mathbf{\Lambda}\right\|^2=\max_{1\le t\le T}\operatorname{tr}\left(\mathbf{\Lambda}'X_tX_t'\mathbf{\Lambda}\right)\le\max_{1\le t\le T}\mu_{\max}\left(X_t'X_t\right)\|\mathbf{\Lambda}\|^2=O_P(N^2). \tag{C.2}$$

On the other hand, using $\frac{1}{N}\mathbf{\Lambda}'\mathbf{\Lambda}=\boldsymbol{I}_{R_0}$ and Assumption 1(iii), we have

$$\begin{aligned}\sum_{t=1}^T||\mathbf{\Lambda}'\varepsilon_t||^2 &= \sum_{t=1}^T\mathsf{Tr}(\mathbf{\Lambda}'\varepsilon_t\varepsilon_t'\mathbf{\Lambda})=\mathsf{Tr}(\mathbf{\Lambda}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{\Lambda})\\ &\le N\|\boldsymbol{\varepsilon}\|_{\text{sp}}^2\,\mathsf{Tr}(\mathbf{\Lambda}'\mathbf{\Lambda}/N)=NR_0\|\boldsymbol{\varepsilon}\|_{\text{sp}}^2=O_P\left(N(N+T)\right).\end{aligned} \tag{C.3}$$

It follows that

$$\Big|\sum_{t=1}^T b_t'X_t'\mathbf{\Lambda}\mathbf{\Lambda}'\varepsilon_t\Big|=O_P\big(p^{1/2}(N^2T^{1/2}+N^{3/2}T)\big), \tag{C.4}$$

as $\|\boldsymbol{b}\|\le C(pT)^{1/2}$. Then, by (C.1) and (C.4), we can complete the proof of (i).

(ii) By the definition of $\boldsymbol{M}_{\mathbf{\Lambda}}$ and noting that $\frac{1}{N}\mathbf{\Lambda}'\mathbf{\Lambda}=\boldsymbol{I}_{R_0}$, we have

$$\frac{1}{NT}\sum_{t=1}^T f_t^{0\prime}\mathbf{\Lambda}^{0\prime}\boldsymbol{M}_{\mathbf{\Lambda}}\varepsilon_t=\frac{1}{NT}\sum_{t=1}^T f_t^{0\prime}\mathbf{\Lambda}^{0\prime}\varepsilon_t-\frac{1}{N^2T}\sum_{t=1}^T f_t^{0\prime}\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}\mathbf{\Lambda}'\varepsilon_t.$$

By Assumptions 1(i) and (iii), we readily have

$$\Big|\sum_{t=1}^T f_t^{0\prime}\mathbf{\Lambda}^{0\prime}\varepsilon_t\Big|=\Big(\sum_{t=1}^T\|f_t^{0\prime}\|^2\Big)^{1/2}\cdot\Big(\sum_{t=1}^T\|\mathbf{\Lambda}^{0\prime}\varepsilon_t\|^2\Big)^{1/2}=O_P(\sqrt{N}T). \tag{C.5}$$

On the other hand, as in the proof of (C.4) above we can show

$$\Big|\sum_{t=1}^T f_t^{0\prime}\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}\mathbf{\Lambda}'\varepsilon_t\Big|=O_P(N^2T^{1/2}+N^{3/2}T). \tag{C.6}$$

We then complete the proof of (ii) by using (C.5) and (C.6).

(iii) As $\frac{1}{N}\mathbf{\Lambda}'\mathbf{\Lambda}=\boldsymbol{I}_{R_0}$, we have $\frac{1}{NT}\sum_{t=1}^T\varepsilon_t'\boldsymbol{P}_{\mathbf{\Lambda}}\varepsilon_t=\frac{1}{N^2T}\sum_{t=1}^T\varepsilon_t'\mathbf{\Lambda}\mathbf{\Lambda}'\varepsilon_t$, which together with (C.3), completes the proof of (iii).

(iv) Using Assumption 1(iii) and the fact $\frac{1}{N}\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\xrightarrow{P}\mathbf{\Sigma}_{\mathbf{\Lambda}}$ under Assumption 1(i), we have

$$\begin{aligned}\Big|\frac{1}{NT}\sum_{t=1}^T\varepsilon_t'\boldsymbol{P}_{\mathbf{\Lambda}^0}\varepsilon_t\Big| &\le \frac{1}{N}\Big\|\Big(\frac{1}{N}\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\Big)^+\Big\|\cdot\frac{1}{NT}\sum_{t=1}^T\left\|\mathbf{\Lambda}^{0\prime}\varepsilon_t\right\|^2\\ &= O_P(N^{-1})\cdot O_P(1)\cdot O_P(1)=O_P(N^{-1}),\end{aligned} \tag{C.7}$$

which completes the proof of (iv).

We has thus completed the proof of Lemma C.1. ■

Lemma C.2 *Suppose that Assumption 1 in Appendix A holds and* $pN^{-1/2}+p^{1/2}T^{-1/2}=o(1)$. *Let* $\dot{\boldsymbol{\beta}}=(\dot{\beta}_1',...,\dot{\beta}_T')'$ *and* $\dot{\boldsymbol{\Lambda}}=(\dot{\lambda}_1',...,\dot{\lambda}_N')'$ *be the preliminary estimates of* $\boldsymbol{\beta}^0$ *and* $\boldsymbol{\Lambda}^0$ *which minimize* $\hat{Q}_{NT}(\boldsymbol{\beta},\boldsymbol{\Lambda})$, *the first term of the objective function defined in (2.4). Then*

$$\frac{1}{T}\sum_{t=1}^{T}\|\dot{\beta}_t-\beta_t^0\|^2=O_P\left(pN^{-1/2}+p^{1/2}T^{-1/2}\right)=o_P(1).$$

Proof of Lemma C.2. The proof of this lemma is similar to that of Theorem 3.1 in Appendix B of the main document. Notice that

$$\hat{Q}_{NT}(\boldsymbol{\beta},\boldsymbol{\Lambda})=\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{N}(Y_t-X_t\beta_t)'\boldsymbol{M}_{\boldsymbol{\Lambda}}(Y_t-X_t\beta_t)\right]\equiv\frac{1}{T}\sum_{t=1}^{T}\hat{Q}_{NT,t}(\beta_t,\boldsymbol{\Lambda}) \tag{C.8}$$

and

$$Y_t-X_t\dot{\beta}_t=X_t(\beta_t^0-\dot{\beta}_t)+\boldsymbol{\Lambda}^0f_t^0+\varepsilon_t. \tag{C.9}$$

Then, by (C.8) and (C.9) and using the fact that $\boldsymbol{M}_{\boldsymbol{\Lambda}^0}\boldsymbol{\Lambda}^0=\boldsymbol{0}$, we have

$$\begin{aligned}
& Q_{NT}(\dot{\boldsymbol{\beta}},\dot{\boldsymbol{\Lambda}})-Q_{NT}(\boldsymbol{\beta}^0,\boldsymbol{\Lambda}^0)\\
=\ & \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\Big[(Y_t-X_t\dot{\beta}_t)'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}(Y_t-X_t\dot{\beta}_t)-(Y_t-X_t\beta_t^0)'\boldsymbol{M}_{\boldsymbol{\Lambda}^0}(Y_t-X_t\beta_t^0)\Big]\\
=\ & \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\Big[(\dot{\beta}_t-\beta_t^0)'X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}X_t(\dot{\beta}_t-\beta_t^0)-2(\dot{\beta}_t-\beta_t^0)'X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0f_t^0+f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0f_t^0\Big]\\
& +\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\Big[-2(\dot{\beta}_t-\beta_t^0)'X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\varepsilon_t+2f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\varepsilon_t-\varepsilon_t'\boldsymbol{P}_{\dot{\boldsymbol{\Lambda}}}\varepsilon_t+\varepsilon_t'\boldsymbol{P}_{\boldsymbol{\Lambda}^0}\varepsilon_t\Big].
\end{aligned} \tag{C.10}$$

By Lemma C.1 above, we can prove that

$$\frac{1}{NT}\sum_{t=1}^{T}\Big[-2(\dot{\beta}_t-\beta_t^0)'X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\varepsilon_t+2f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\varepsilon_t-\varepsilon_t'\boldsymbol{P}_{\dot{\boldsymbol{\Lambda}}}\varepsilon_t+\varepsilon_t'\boldsymbol{P}_{\boldsymbol{\Lambda}^0}\varepsilon_t\Big]=O_P(pN^{-1/2}+p^{1/2}T^{-1/2}). \tag{C.11}$$

Let $\dot{\boldsymbol{d}}_\beta=\dot{\boldsymbol{\beta}}-\boldsymbol{\beta}^0$ and $\dot{\boldsymbol{d}}_\Lambda=\frac{1}{N^{1/2}}\mathsf{vec}(\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0)$ where $\mathsf{vec}(\cdot)$ denotes the vectorization of a matrix. Define

$$\begin{aligned}
\dot{\boldsymbol{A}} &= \frac{1}{N}\mathsf{diag}(X_1'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}X_1,...,X_T'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}X_T),\quad \dot{\boldsymbol{B}}=(\boldsymbol{F}^{0\prime}\boldsymbol{F}^0)\otimes\boldsymbol{I}_N,\ \text{and}\\
\dot{\boldsymbol{C}} &= \frac{1}{N^{1/2}}\left[f_1^0\otimes\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}X_1,...,f_T^0\otimes\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}X_T\right],
\end{aligned}$$

where $\otimes$ denotes the Kronecker product. It is easy to verify that

$$\frac{1}{NT}\sum_{t=1}^{T}\left(\dot\beta_t-\beta_t^0\right)'X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}X_t\left(\dot\beta_t-\beta_t^0\right)=\frac{1}{T}\dot{\boldsymbol{d}}_\beta'\dot{\boldsymbol{A}}\dot{\boldsymbol{d}}_\beta,$$

$$\frac{1}{NT}\sum_{t=1}^{T}\left(\dot\beta_t-\beta_t^0\right)'X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0f_t^0=\frac{1}{NT}\sum_{t=1}^{T}\mathsf{Tr}\left\{\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0f_t^0\left(\dot\beta_t-\beta_t^0\right)'X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\right\}=\frac{1}{T}\dot{\boldsymbol{d}}_\Lambda'\dot{\boldsymbol{C}}\dot{\boldsymbol{d}}_\beta,$$

and

$$\frac{1}{NT}\sum_{t=1}^{T}f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0f_t^0=\frac{1}{NT}\sum_{t=1}^{T}\mathsf{Tr}\left(\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0f_t^0f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\right)=\frac{1}{T}\dot{\boldsymbol{d}}_\Lambda'\dot{\boldsymbol{B}}\dot{\boldsymbol{d}}_\Lambda,$$

where we have used the following fact on matrix calculation that $\mathsf{Tr}\left(\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_3\right)=\mathsf{vec}'\left(\boldsymbol{A}_1\right)\left(\boldsymbol{A}_2\otimes\boldsymbol{I}_k\right)\mathsf{vec}\left(\boldsymbol{A}_3\right)$ and that $\mathsf{Tr}\left(\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{A}_4\right)=\mathsf{vec}'\left(\boldsymbol{A}_1\right)\left(\boldsymbol{A}_2\otimes\boldsymbol{A}_4'\right)\mathsf{vec}\left(\boldsymbol{A}_3'\right)$ with k being the size of the column vectors in $\boldsymbol{A}_3$ (in the first equation). With the above notations, we may show that

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\left[\left(\dot\beta_t-\beta_t^0\right)'X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}X_t\left(\dot\beta_t-\beta_t^0\right)-2\left(\dot\beta_t-\beta_t^0\right)'X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0f_t^0+f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0f_t^0\right]\\=&\ \frac{1}{T}\left(\dot{\boldsymbol{d}}_\beta'\dot{\boldsymbol{A}}\dot{\boldsymbol{d}}_\beta-2\dot{\boldsymbol{d}}_\Lambda'\dot{\boldsymbol{C}}\dot{\boldsymbol{d}}_\beta+\dot{\boldsymbol{d}}_\Lambda'\dot{\boldsymbol{B}}\dot{\boldsymbol{d}}_\Lambda\right)=\frac{1}{T}\left(\dot{\boldsymbol{d}}_\beta'\dot{\boldsymbol{D}}\dot{\boldsymbol{d}}_\beta+\dot{\boldsymbol{d}}_*'\dot{\boldsymbol{B}}\dot{\boldsymbol{d}}_*\right),\end{aligned}$$

where $\dot{\boldsymbol{D}}=\dot{\boldsymbol{A}}-\dot{\boldsymbol{C}}'\dot{\boldsymbol{B}}^+\dot{\boldsymbol{C}}$ and $\dot{\boldsymbol{d}}_*=\dot{\boldsymbol{d}}_\Lambda-\dot{\boldsymbol{B}}^+\dot{\boldsymbol{C}}\dot{\boldsymbol{d}}_\beta$. By Assumption 1(i), we may show that the minimum eigenvalue of $\frac{1}{T}\dot{\boldsymbol{B}}$ is bounded away from zero w.p.a.1, i.e., there exists a positive constant c_4 such that $\mu_{\min}\left(\dot{\boldsymbol{B}}/T\right)>c_4$ w.p.a.1. Using a decomposition similar to (B.8) in Appendix B, we can readily show that $\mu_{\max}\left(\dot{\boldsymbol{C}}'\dot{\boldsymbol{C}}/T\right)=o_P(1)$. By Assumption 1(ii), we can also show that the minimum eigenvalue of $\dot{\boldsymbol{A}}$ is bounded away from zero w.p.a.1, i.e., there exists a positive constant c_x (defined in Assumption 1(ii)) such that $\mu_{\min}\left(\dot{\boldsymbol{A}}\right)>c_x$ w.p.a.1. Hence, we have proved that the matrix $\dot{\boldsymbol{D}}$ is asymptotically positive definite as its minimum eigenvalue is positive and bounded away from zero w.p.a.1.

Note that

$$\frac{1}{T}\left(\dot{\boldsymbol{d}}_\beta'\dot{\boldsymbol{D}}\dot{\boldsymbol{d}}_\beta+\dot{\boldsymbol{d}}_*'\dot{\boldsymbol{B}}\dot{\boldsymbol{d}}_*\right)+O_P\left(pN^{-1/2}+p^{1/2}T^{-1/2}\right)\le Q_{NT}\left(\dot{\boldsymbol{\beta}},\dot{\boldsymbol{\Lambda}}\right)-Q_{NT}\left(\boldsymbol{\beta}^0,\boldsymbol{\Lambda}^0\right)\le 0,\quad\text{(C.12)}$$

$\dot{\boldsymbol{d}}_*'\dot{\boldsymbol{B}}\dot{\boldsymbol{d}}_*$ is asymptotically nonnegative, and $\dot{\boldsymbol{d}}_\beta'\dot{\boldsymbol{D}}\dot{\boldsymbol{d}}_\beta\ge c_5\|\dot{\boldsymbol{d}}_\beta\|^2$ where c_5 is a positive constant. It follows that $\frac{1}{T}\|\dot{\boldsymbol{d}}_\beta\|^2=\frac{1}{T}\sum_{t=1}^{T}\|\dot\beta_t-\beta_t^0\|^2=O_P(pN^{-1/2}+p^{1/2}T^{-1/2})=o_P(1)$, completing the proof of Lemma C.2. ■

Lemma C.3 *Suppose that Assumption 1 in Appendix A holds and $pN^{-1/2}+p^{1/2}T^{-1/2}=o(1)$. Let $\dot{\boldsymbol{H}}\equiv\dot{\boldsymbol{H}}_{NT}=\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\dot{\boldsymbol{\Lambda}}\right)\dot{\boldsymbol{V}}_{NT}^+$, where $\dot{\boldsymbol{V}}_{NT}$ is analogously defined as $\boldsymbol{V}_{NT}$ in (2.7) with $\hat\beta_t$ replaced by $\dot\beta_t$. Denote $\dot\eta_{NT}=\frac{1}{T}\sum_{t=1}^{T}\|\dot\beta_t-\beta_t^0\|^2$. Then we have*

(i) $\frac{1}{N}\left\|\dot{\mathbf{\Lambda}}-\mathbf{\Lambda}^0\dot{\boldsymbol{H}}\right\|^2=O_P\left(\delta_{NT}^{-2}+\dot{\eta}_{NT}\right)$,
(ii) $\frac{1}{N}\left(\dot{\mathbf{\Lambda}}-\mathbf{\Lambda}^0\dot{\boldsymbol{H}}\right)'\mathbf{\Lambda}^0\dot{\boldsymbol{H}}=O_P\left(\delta_{NT}^{-2}+\dot{\eta}_{NT}^{1/2}\right)$,
(iii) $\frac{1}{N}\left(\dot{\mathbf{\Lambda}}-\mathbf{\Lambda}^0\dot{\boldsymbol{H}}\right)'\dot{\mathbf{\Lambda}}=O_P\left(\delta_{NT}^{-2}+\dot{\eta}_{NT}^{1/2}\right)$,
(iv) $\frac{1}{N}\left(\dot{\mathbf{\Lambda}}'\dot{\mathbf{\Lambda}}-\dot{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\dot{\boldsymbol{H}}\right)=O_P\left(\delta_{NT}^{-2}+\dot{\eta}_{NT}^{1/2}\right)$,
(v) $\left\|\boldsymbol{P}_{\dot{\mathbf{\Lambda}}}-\boldsymbol{P}_{\mathbf{\Lambda}^0\dot{\boldsymbol{H}}}\right\|=O_P\left(\delta_{NT}^{-2}+\dot{\eta}_{NT}^{1/2}\right)$,
(vi) $\frac{1}{NT}\sum_{s=1}^T(\dot{\mathbf{\Lambda}}-\mathbf{\Lambda}^0\dot{\boldsymbol{H}})'\varepsilon_s\gamma_s'=O_P\left(\delta_{NT}^{-2}+\dot{\eta}_{NT}^{1/2}\right)$ *with* $\gamma_s=1$ *or* f_s^0, *and*
(vii) $\frac{1}{NT}\sum_{s=1}^T||(\dot{\mathbf{\Lambda}}-\mathbf{\Lambda}^0\dot{\boldsymbol{H}})'\varepsilon_s||^2=O_P\left((1+NT^{-1})(\delta_{NT}^{-2}+\dot{\eta}_{NT})\right)$.

Proof of Lemma C.3. **(i)** By (2.7) and (C.9) and letting $d_t=\dot{\beta}_t-\beta_t^0$, we have

$$\begin{aligned}
&\dot{\mathbf{\Lambda}}\dot{\boldsymbol{V}}_{NT}-\mathbf{\Lambda}^0\dot{\boldsymbol{H}}\dot{\boldsymbol{V}}_{NT}\\
=&\Big[\frac{1}{NT}\sum_{t=1}^T\left(Y_t-X_t\dot{\beta}_t\right)\left(Y_t-X_t\dot{\beta}_t\right)\Big]\dot{\mathbf{\Lambda}}-\mathbf{\Lambda}^0\dot{\boldsymbol{H}}\dot{\boldsymbol{V}}_{NT}\\
=&\Big\{\frac{1}{NT}\sum_{t=1}^T\left[-X_td_t+\mathbf{\Lambda}^0f_t^0+\varepsilon_t\right]\left[-X_td_t+\mathbf{\Lambda}^0f_t^0+\varepsilon_t\right]'\Big\}\dot{\mathbf{\Lambda}}-\mathbf{\Lambda}^0\dot{\boldsymbol{H}}\dot{\boldsymbol{V}}_{NT}\\
=&\frac{1}{NT}\sum_{t=1}^TX_td_td_t'X_t'\dot{\mathbf{\Lambda}}-\frac{1}{NT}\sum_{t=1}^TX_td_tf_t^{0\prime}\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}-\frac{1}{NT}\sum_{t=1}^TX_td_t\varepsilon_t'\dot{\mathbf{\Lambda}}-\frac{1}{NT}\sum_{t=1}^T\mathbf{\Lambda}^0f_t^0d_t'X_t'\dot{\mathbf{\Lambda}}\\
&+\frac{1}{NT}\sum_{t=1}^T\mathbf{\Lambda}^0f_t^0\varepsilon_t'\dot{\mathbf{\Lambda}}-\frac{1}{NT}\sum_{t=1}^T\varepsilon_td_t'X_t'\dot{\mathbf{\Lambda}}+\frac{1}{NT}\sum_{t=1}^T\varepsilon_tf_t^{0\prime}\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}+\frac{1}{NT}\sum_{t=1}^T\varepsilon_t\varepsilon_t'\dot{\mathbf{\Lambda}}\\
\equiv&\sum_{j=1}^8\dot{u}_{NT,j}.
\end{aligned}\tag{C.13}$$

Noting that $\mathsf{Tr}\,(AB)\leq\mathsf{Tr}\,(A)\,\mathsf{Tr}\,(B)$ for conformable positive semidefinite matrices A and B, $\|\dot{\mathbf{\Lambda}}\|=O_P(N^{1/2})$ and $\max_{1\leq t\leq T}\mu_{\max}^2\left(X_t'X_t/N\right)=O_P\left(1\right)$ by Assumption 1(ii), we have

$$\begin{aligned}
\|\dot{u}_{NT,1}\|^2&=\frac{1}{N^2T^2}\sum_{t=1}^T\sum_{s=1}^T\mathsf{Tr}(X_td_td_t'X_t'\dot{\mathbf{\Lambda}}\dot{\mathbf{\Lambda}}'X_sd_sd_s'X_s')\\
&\leq\left\|\dot{\mathbf{\Lambda}}\right\|^2\left\{\frac{1}{NT}\sum_{t=1}^T\mathsf{Tr}(X_td_td_t'X_t')\right\}^2=\left\|\dot{\mathbf{\Lambda}}\right\|^2\left\{\frac{1}{NT}\sum_{t=1}^Td_t'X_t'X_td_t\right\}^2\\
&\leq\left\|\dot{\mathbf{\Lambda}}\right\|^2\left[\max_{1\leq t\leq T}\mu_{\max}^2\left(X_t'X_t/N\right)\right]\left\{\frac{1}{T}\sum_{t=1}^T\|d_t\|^2\right\}^2=O_P(N\dot{\eta}_{NT}^2).
\end{aligned}\tag{C.14}$$

Noting that $\mathsf{Tr}\,(AB) \leq \mathsf{Tr}\,(AA')^{1/2}\,\mathsf{Tr}\,(BB')^{1/2}$ for conformable matrices A and B, we have

$$\begin{aligned}
\|\dot{u}_{NT,2}\|^2 &= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathsf{Tr}(X_t d_t f_t^{0\prime}\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}\dot{\mathbf{\Lambda}}'\mathbf{\Lambda}^0 f_s^0 d_s' X_s') \\
&\leq \left\|\dot{\mathbf{\Lambda}}\right\|^2 \mu_{\max}(\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0)\frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathsf{Tr}(X_t d_t f_t^{0\prime} f_s^0 d_s' X_s') \\
&\leq \frac{1}{N}\left\|\dot{\mathbf{\Lambda}}\right\|^2 \mu_{\max}(\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0/N)\left(\frac{1}{T}\sum_{t=1}^{T}\left\{\mathsf{Tr}(f_t^0 d_t' X_t' X_t d_t f_t^{0\prime})\right\}^{1/2}\right)^2 \\
&\leq \left\|\dot{\mathbf{\Lambda}}\right\|^2 \mu_{\max}(\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0/N)\left[\max_{1\leq t\leq T}\mu_{\max}\left(X_t'X_t/N\right)\right]\left(\frac{1}{T}\sum_{t=1}^{T}\|d_t\|\left\|f_t^0\right\|\right)^2 \\
&= O_P\left(N\right)O_P\left(1\right)O_P\left(1\right)\frac{1}{T}\sum_{t=1}^{T}\|d_t\|^2\frac{1}{T}\sum_{t=1}^{T}\left\|f_t^0\right\|^2 = O_P\left(N\dot{\eta}_{NT}\right), \qquad \text{(C.15)}
\end{aligned}$$

and analogously

$$\|\dot{u}_{NT,4}\|^2 = O_P\left(N\left(\frac{1}{T}\sum_{t=1}^{T}\|\dot{\beta}_t - \beta_t^0\|^2\right)\right) = O_P\left(N\dot{\eta}_{NT}\right). \qquad \text{(C.16)}$$

Noting that $\sum_{t=1}^{T}\|\varepsilon_t\|^2 = O_P(NT)$ by Assumption 1(iii) and $\max_{1\leq t\leq T}\mu_{\max}\left(X_t'X_t/N\right) = O_P(1)$ by Assumption 1(ii), we can show that

$$\begin{aligned}
\|\dot{u}_{NT,3}\|^2 &= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathsf{Tr}(X_t d_t \varepsilon_t'\dot{\mathbf{\Lambda}}\dot{\mathbf{\Lambda}}'\varepsilon_s d_s' X_s') \leq \left\|\dot{\mathbf{\Lambda}}\right\|^2\frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathsf{Tr}(X_t d_t \varepsilon_t'\varepsilon_s d_s' X_s') \\
&\leq \left\|\dot{\mathbf{\Lambda}}\right\|^2\left\{\frac{1}{NT}\sum_{t=1}^{T}\left\{\mathsf{Tr}(\varepsilon_t d_t' X_t' X_t d_t \varepsilon_t')\right\}^{1/2}\right\}^2 \\
&\leq \frac{1}{N}\left\|\dot{\mathbf{\Lambda}}\right\|^2\left[\max_{1\leq t\leq T}\mu_{\max}\left(X_t'X_t/N\right)\right]\left\{\frac{1}{T}\sum_{t=1}^{T}\|\varepsilon_t\|\,\|d_t\|\right\}^2 \\
&\leq O_P\left(1\right)\frac{1}{T}\sum_{t=1}^{T}\|\varepsilon_t\|^2\,\frac{1}{T}\sum_{t=1}^{T}\|d_t\|^2 = O_P\left(N\dot{\eta}_{NT}\right) \qquad \text{(C.17)}
\end{aligned}$$

and analogously

$$\|\dot{u}_{NT,6}\|^2 = O_P\left(\frac{N}{T}\sum_{t=1}^{T}\|\dot{\beta}_t - \beta_t^0\|^2\right) = O_P\left(N\dot{\eta}_{NT}\right). \qquad \text{(C.18)}$$

The analysis of the remaining three terms is similar to the proof of Theorem 1 in Bai and Ng (2002) by switching the roles of f_t and λ_i. For $\dot{u}_{NT,5}$, using the fact that $\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0 = O_P(N)$,

$\|\dot{\mathbf{\Lambda}}\| = O_P(N^{1/2})$ and Assumptions 1(iii) and (iv), we can prove that

$$
\begin{aligned}
\|\dot{u}_{NT,5}\|^2 &= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathsf{Tr}\Big(\mathbf{\Lambda}^0 f_t^0\varepsilon_t'\dot{\mathbf{\Lambda}}\dot{\mathbf{\Lambda}}'\varepsilon_s f_s^{0\prime}\mathbf{\Lambda}^{0\prime}\Big) = \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathsf{Tr}\Big(f_t^0\varepsilon_t'\dot{\mathbf{\Lambda}}\dot{\mathbf{\Lambda}}'\varepsilon_s f_s^{0\prime}\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\Big) \\
&= O_P\left(\frac{1}{NT^2}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{k=1}^{N}\varepsilon_{it}\varepsilon_{ks}\dot{\lambda}_i'\dot{\lambda}_k f_t^0 f_s^{0\prime}\right\|\right) \\
&= O_P\left(\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left|\dot{\lambda}_i'\dot{\lambda}_k\right|\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\varepsilon_{it}\varepsilon_{ks}f_t^0 f_s^{0\prime}\right\|\right) \\
&= O_P\left(\frac{1}{NT^2}\left(\sum_{i=1}^{N}\sum_{k=1}^{N}\|\dot{\lambda}_i\|^2\|\dot{\lambda}_k\|^2\right)^{1/2}\left(\sum_{i=1}^{N}\sum_{k=1}^{N}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\varepsilon_{it}\varepsilon_{ks}f_t^0 f_s^{0\prime}\right\|^2\right)^{1/2}\right) \\
&= O_P\left(\frac{1}{T^2}\left(\sum_{i=1}^{N}\sum_{k=1}^{N}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\varepsilon_{it}\varepsilon_{ks}f_t^0 f_s^{0\prime}\right\|^2\right)^{1/2}\right) = O_P(N/T), \qquad \text{(C.19)}
\end{aligned}
$$

and

$$
\begin{aligned}
\|\dot{u}_{NT,7}\|^2 &= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathsf{Tr}\left(\varepsilon_t f_t^{0\prime}\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}\dot{\mathbf{\Lambda}}'\mathbf{\Lambda}^0 f_s^0\varepsilon_s'\right) = \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathsf{Tr}\left(\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}\dot{\mathbf{\Lambda}}'\mathbf{\Lambda}^0 f_s^0\varepsilon_s'\varepsilon_t f_t^{0\prime}\right) \\
&= O_P\left(\frac{1}{T^2}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}f_s^0\varepsilon_s'\varepsilon_t f_t^{0\prime}\right\|\right) = O_P(N/T). \qquad \text{(C.20)}
\end{aligned}
$$

By the assumption that $\max_{1\le i,j\le N}\mathsf{E}\Big[\big\|\sum_{t=1}^{T}\sum_{s=1}^{T}\varepsilon_{it}\varepsilon_{js}\varepsilon_t'\varepsilon_s\big\|^2\Big] = O(N^2T^2+T^2)$ in Assumption 1(iii), we can similarly prove

$$\|\dot{u}_{NT,8}\|^2 = O_P(N/T). \qquad \text{(C.21)}$$

By (C.13)–(C.21), we can prove that

$$\frac{1}{N}\big\|\dot{\mathbf{\Lambda}}\dot{\boldsymbol{V}}_{NT} - \mathbf{\Lambda}^0\dot{\boldsymbol{H}}\dot{\boldsymbol{V}}_{NT}\big\|^2 = O_P(\delta_{NT}^{-2} + \dot{\eta}_{NT}). \qquad \text{(C.22)}$$

Premultiplying (C.13) by $\dot{\mathbf{\Lambda}}'$, and using the identification restriction on $\dot{\mathbf{\Lambda}}$: $\frac{1}{N}\dot{\mathbf{\Lambda}}'\dot{\mathbf{\Lambda}} = \boldsymbol{I}_{R_0}$, (C.22) and Lemma C.2, we may show that

$$\dot{\boldsymbol{V}}_{NT} - \left(\frac{1}{N}\dot{\mathbf{\Lambda}}'\mathbf{\Lambda}^0\right)\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)\left(\frac{1}{N}\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}\right) = o_P(1). \qquad \text{(C.23)}$$

Furthermore, applying (C.12) in the proof of Lemma C.2 and noting that the matrix $\dot{\boldsymbol{B}}$ is positive definite, we can show that

$$\frac{1}{N}\mathbf{\Lambda}^{0\prime}\boldsymbol{M}_{\dot{\mathbf{\Lambda}}}\mathbf{\Lambda}^0 = \frac{1}{N}\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0 - \left(\frac{1}{N}\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}\right)\left(\frac{1}{N}\dot{\mathbf{\Lambda}}'\mathbf{\Lambda}^0\right) = o_P(1),$$

which together with Assumption 1(i), implies that $\frac{1}{N}\dot{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}^0$ is asymptotically invertible and thus $\dot{\boldsymbol{V}}_{NT}$ is also asymptotically invertible. We can then complete the proof of (i) by using this fact and (C.22).

(ii) Observe that by (C.13)

$$\frac{1}{N}(\dot{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}})'\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}} = \frac{1}{N}\sum_{j=1}^{8}\dot{\boldsymbol{V}}_{NT}^{+}\dot{u}_{NT,j}'\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}} \equiv \frac{1}{N}\sum_{j=1}^{8}\dot{u}_{NT,j}^{*}. \tag{C.24}$$

By Assumption 1(i) and (C.14), we can readily prove

$$\frac{1}{N}\|\dot{u}_{NT,1}^{*}\| \leq \left(\frac{1}{N^{1/2}}\|\dot{u}_{NT,1}\|\right)\cdot\|\dot{\boldsymbol{V}}_{NT}^{+}\|\cdot\left(\frac{1}{N^{1/2}}\|\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\|\right) = O_P\left(\dot{\eta}_{NT}\right). \tag{C.25}$$

Analogously, by (C.15) and (C.16), we can prove that

$$\frac{1}{N}\|\dot{u}_{NT,2}^{*}\| = O_P\left(\dot{\eta}_{NT}^{1/2}\right) \quad\text{and}\quad \frac{1}{N}\|\dot{u}_{NT,4}^{*}\| = O_P\left(\dot{\eta}_{NT}^{1/2}\right). \tag{C.26}$$

For $\dot{u}_{NT,3}^{*}$, by the definition of $\dot{u}_{NT,3}$, we have

$$\begin{aligned}
-\dot{u}_{NT,3}^{*} &= -\dot{\boldsymbol{V}}_{NT}^{+}\dot{u}_{NT,3}'\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}} = \frac{1}{NT}\dot{\boldsymbol{V}}_{NT}^{+}\sum_{t=1}^{T}\dot{\boldsymbol{\Lambda}}'\varepsilon_t d_t' X_t'\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}} \\
&= \frac{1}{NT}\dot{\boldsymbol{V}}_{NT}^{+}\sum_{t=1}^{T}\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\varepsilon_t d_t' X_t'\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}} + \frac{1}{NT}\dot{\boldsymbol{V}}_{NT}^{+}\sum_{t=1}^{T}(\dot{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}})'\varepsilon_t d_t' X_t'\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}} \\
&\equiv \dot{u}_{NT,3a}^{*} + \dot{u}_{NT,3b}^{*}.
\end{aligned} \tag{C.27}$$

By the Cauchy-Schwarz inequality and Assumptions 1(ii) and (iii), we have

$$\|\dot{u}_{NT,3a}^{*}\| \leq \frac{C}{T}\sum_{t=1}^{T}\|\boldsymbol{\Lambda}^{0\prime}\varepsilon_t d_t'\| \leq C\left(\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{\Lambda}^{0\prime}\varepsilon_t\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\|d_t\|^2\right)^{1/2} = O_P\left((N\dot{\eta}_{NT})^{1/2}\right). \tag{C.28}$$

Similarly, with the help of Lemma C.3(i), we can also prove that

$$\|\dot{u}_{NT,3b}^{*}\| = O_P\left(N\dot{\eta}_{NT} + N\delta_{NT}^{-1}\dot{\eta}_{NT}^{1/2}\right). \tag{C.29}$$

By (C.27)–(C.29), we have

$$\frac{1}{N}\|\dot{u}_{NT,3}^{*}\| = O_P\left(\dot{\eta}_{NT} + \delta_{NT}^{-1}\dot{\eta}_{NT}^{1/2}\right). \tag{C.30}$$

Similarly, we can also show that

$$\frac{1}{N}\|\dot{u}_{NT,6}^{*}\| = O_P\left(\dot{\eta}_{NT} + \delta_{NT}^{-1}\dot{\eta}_{NT}^{1/2}\right). \tag{C.31}$$

For $\dot{u}^*_{NT,5}$, by the definition of $\dot{u}_{NT,5}$, we have

$$\begin{aligned}\dot{u}^*_{NT,5} &= \frac{1}{NT}\dot{\boldsymbol{V}}^+_{NT}\sum_{t=1}^T \dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\varepsilon_t f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}} + \frac{1}{NT}\dot{\boldsymbol{V}}^+_{NT}\sum_{t=1}^T \left(\dot{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\right)'\varepsilon_t f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}} \\ &\equiv \dot{u}^*_{NT,5a} + \dot{u}^*_{NT,5b}. \end{aligned} \tag{C.32}$$

By Assumptions 1(i) and (iii), we have

$$\|\dot{u}^*_{NT,5a}\| \leq C\frac{1}{T}\|\sum_{t=1}^T \boldsymbol{\Lambda}^{0\prime}\varepsilon_t f_t^{0\prime}\| = O_P\Big(\frac{1}{T}\|\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\varepsilon}\boldsymbol{F}^0\|\Big) = O_P\left(N^{1/2}T^{-1/2}\right). \tag{C.33}$$

Using Lemma C.3(i), we can also prove that

$$\|\dot{u}^*_{NT,5b}\| = O_P\left(N\dot{\eta}_{NT} + N\delta_{NT}^{-2}\right). \tag{C.34}$$

By (C.32)–(C.34), we have

$$\frac{1}{N}\|\dot{u}^*_{NT,5}\| = O_P\left(\dot{\eta}_{NT} + \delta_{NT}^{-2}\right). \tag{C.35}$$

Noting that $\dot{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}^0 = O_P(N)$ and using the assumption $\mathsf{E}\big[\big\|\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\varepsilon}\boldsymbol{F}^0\big\|^2\big] = O(NT)$ in Assumption 1(iii), we can also show that

$$\frac{1}{N}\|\dot{u}^*_{NT,7}\| = O_P\left(\dot{\eta}_{NT} + \delta_{NT}^{-2}\right) \quad \text{and} \quad \frac{1}{N}\|\dot{u}^*_{NT,8}\| = O_P\left(\dot{\eta}_{NT} + \delta_{NT}^{-2}\right). \tag{C.36}$$

By (C.24)–(C.26), (C.30), (C.31), (C.35) and (C.36), we can complete the proof of (ii).

(iii) and **(iv)** The proofs of (iii) and (iv) can be completed by using the results in Lemmas C.3(i) and (ii).

(v) Note that

$$\boldsymbol{P}_{\dot{\boldsymbol{\Lambda}}} - \boldsymbol{P}_{\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}} = \dot{\boldsymbol{\Lambda}}\big(\dot{\boldsymbol{\Lambda}}'\dot{\boldsymbol{\Lambda}}\big)^+\dot{\boldsymbol{\Lambda}}' - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)^+\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime} \equiv \sum_{j=1}^7 \dot{v}_{NT,j}, \tag{C.37}$$

where

$$\begin{aligned}
\dot{v}_{NT,1} &= \big(\dot{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)\big(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)^+\big(\dot{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)', \\
\dot{v}_{NT,2} &= \big(\dot{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)\big(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)^+\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}, \\
\dot{v}_{NT,3} &= \big(\dot{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)\big[\big(\dot{\boldsymbol{\Lambda}}'\dot{\boldsymbol{\Lambda}}\big)^+ - \big(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)^+\big]\big(\dot{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)', \\
\dot{v}_{NT,4} &= \big(\dot{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)\big[\big(\dot{\boldsymbol{\Lambda}}'\dot{\boldsymbol{\Lambda}}\big)^+ - \big(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)^+\big]\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}, \\
\dot{v}_{NT,5} &= \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)^+\big(\dot{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)', \\
\dot{v}_{NT,6} &= \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big[\big(\dot{\boldsymbol{\Lambda}}'\dot{\boldsymbol{\Lambda}}\big)^+ - \big(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)^+\big]\big(\dot{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)', \\
\dot{v}_{NT,7} &= \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big[\big(\dot{\boldsymbol{\Lambda}}'\dot{\boldsymbol{\Lambda}}\big)^+ - \big(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)^+\big]\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}.
\end{aligned}$$

Using the results in Lemmas C.3(i) and (iv), we can prove (v).

(**vi**) The proof is analogous to that of part (ii) and thus omitted.

(**vii**) By Assumption 1(iii) and part (i),

$$\begin{aligned}\frac{1}{NT}\sum_{s=1}^{T}\|(\dot{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}})'\varepsilon_s\|^2 &= \frac{1}{NT}\mathsf{Tr}\Big((\dot{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}})'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'(\dot{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}})\Big)\\ &\leq \frac{1}{T}\|\boldsymbol{\varepsilon}\|_{\text{sp}}^2\cdot\frac{1}{N}\mathsf{Tr}\big((\dot{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}})'(\dot{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}})\big)\\ &= O_P((1+NT^{-1})(\delta_{NT}^{-2}+\dot{\eta}_{NT})).\end{aligned}$$

We have thus completed the proof of Lemma C.3. ■

With the above three lemmas, we are ready to give the proof of Lemma B.1.

Proof of Lemma B.1. Let $\hat{Q}_{NT,t}(\beta_t,\boldsymbol{\Lambda})$ be defined as in (C.8), $\dot{\boldsymbol{\beta}}$ and $\dot{\boldsymbol{\Lambda}}$ be defined in Lemma C.2, and $\dot{\boldsymbol{H}}$ be defined in Lemma C.3. Note that

$$Y_t - X_t\dot{\beta}_t = X_t(\beta_t^0-\dot{\beta}_t)+\dot{\boldsymbol{\Lambda}}\dot{\boldsymbol{H}}^+f_t^0+\big(\boldsymbol{\Lambda}^0-\dot{\boldsymbol{\Lambda}}\dot{\boldsymbol{H}}^+\big)f_t^0+\varepsilon_t. \tag{C.38}$$

The preliminary estimate $\dot{\beta}_t$ which minimizes $\hat{Q}_{NT,t}(\beta_t,\boldsymbol{\Lambda})$ (with respect to β_t) satisfies that

$$\Big(\frac{1}{N}X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}X_t\Big)(\dot{\beta}_t-\beta_t^0)=\frac{1}{N}X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\varepsilon_t+\frac{1}{N}X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\big(\boldsymbol{\Lambda}^0-\dot{\boldsymbol{\Lambda}}\dot{\boldsymbol{H}}^+\big)f_t^0, \tag{C.39}$$

as $\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\dot{\boldsymbol{\Lambda}}=\mathbf{0}$, where $\mathbf{0}$ is a null matrix or vector whose size may change from line to line.

We first consider the term $\frac{1}{N}X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\varepsilon_t$. Notice that

$$\frac{1}{N}X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}\varepsilon_t=\frac{1}{N}X_t'\boldsymbol{M}_{\boldsymbol{\Lambda}^0}\varepsilon_t+\frac{1}{N}X_t'\big(\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}}-\boldsymbol{M}_{\boldsymbol{\Lambda}^0}\big)\varepsilon_t. \tag{C.40}$$

By the definition of $\boldsymbol{M}_{\boldsymbol{\Lambda}^0}$, we have

$$\frac{1}{N}X_t'\boldsymbol{M}_{\boldsymbol{\Lambda}^0}\varepsilon_t=\frac{1}{N}X_t'\varepsilon_t-\frac{1}{N}X_t'\boldsymbol{\Lambda}^0(\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0)^+\boldsymbol{\Lambda}^{0\prime}\varepsilon_t. \tag{C.41}$$

By Assumption 1(iii), we can show that for each $1\leq t\leq T$

$$\frac{1}{N}\|X_t'\varepsilon_t\|=O_P\left(p^{1/2}N^{-1/2}\right). \tag{C.42}$$

By Assumptions 1(i)–(iii), we can show that for each $1\leq t\leq T$

$$\|X_t'\boldsymbol{\Lambda}^0\|=O_P(N),\quad \|\boldsymbol{\Lambda}^{0\prime}\varepsilon_t\|=O_P(N^{1/2})\;\text{ and }\;\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right)^+\xrightarrow{P}\boldsymbol{\Sigma}_\Lambda^+,$$

which imply that

$$\frac{1}{N}\|X_t'\boldsymbol{\Lambda}^0(\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0)^+\boldsymbol{\Lambda}^{0\prime}\varepsilon_t\|=O_P\left(N^{-1/2}\right). \tag{C.43}$$

Thus, by (C.41)–(C.43), we have

$$\frac{1}{N}\|X_t'\boldsymbol{M}_{\boldsymbol{\Lambda}^0}\varepsilon_t\| = O_P\left(p^{1/2}N^{-1/2}\right). \tag{C.44}$$

To derive the order of $X_t'\big(\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}} - \boldsymbol{M}_{\boldsymbol{\Lambda}^0}\big)\varepsilon_t$, we need to investigate the term $\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}} - \boldsymbol{M}_{\boldsymbol{\Lambda}^0}$. By (C.37), we have

$$-(\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}} - \boldsymbol{M}_{\boldsymbol{\Lambda}^0}) = \dot{\boldsymbol{\Lambda}}\big(\dot{\boldsymbol{\Lambda}}'\dot{\boldsymbol{\Lambda}}\big)^+\dot{\boldsymbol{\Lambda}}' - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)^+\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime} = \sum_{j=1}^{7}\dot{v}_{NT,j}. \tag{C.45}$$

We next show that

$$\frac{1}{N}\Big\|X_t'\big(\sum_{j=1}^{7}\dot{v}_{NT,j}\big)\varepsilon_t\Big\| = O_P\big(\delta_{NT}^{-1}\big). \tag{C.46}$$

To save the space, we only consider the case of $j=5$. Other cases can be studied similarly. For $X_t'\dot{v}_{NT,5}\varepsilon_t$, note that

$$\begin{aligned}
\dot{v}_{NT,5} &= \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)^+\big(\dot{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)', \\
&= \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)^+\dot{\boldsymbol{V}}_{NT}^+\big(\dot{\boldsymbol{\Lambda}}\dot{\boldsymbol{V}}_{NT} - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\dot{\boldsymbol{V}}_{NT}\big)', \\
&= \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)^+\dot{\boldsymbol{V}}_{NT}^+\big(\sum_{j=1}^{8}\dot{u}_{NT,j}\big)',
\end{aligned} \tag{C.47}$$

where $\dot{u}_{NT,j}$, $j=1,...,8$, are defined in the proof of Lemma C.3(i) above. By the fact that both $\dot{\boldsymbol{H}}$ and $\dot{\boldsymbol{V}}_{NT}$ are asymptotically invertible and similar to the proof of Lemma C.3(i), we readily prove that

$$\frac{1}{N}\left\|X_t'\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\left(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\right)^+\dot{\boldsymbol{V}}_{NT}^+\left(\sum_{j=1}^{5}\dot{u}_{NT,j} + \dot{u}_{NT,8}\right)'\varepsilon_t\right\| = O_P\left(\delta_{NT}^{-2} + \delta_{NT}^{-1}\dot{\eta}_{NT}^{1/2}\right). \tag{C.48}$$

Meanwhile, by Assumptions 1(i)(ii) and noting that

$$\max_{1\le t\le T}\mathsf{E}\big[\sum_{s=1}^{T}\big|\varepsilon_s'\varepsilon_t\big|^2\big] = \max_{1\le t\le T}\mathsf{E}\big[\sum_{s=1}^{T}(\xi_{st}^*)^2\big] = O(N^2+NT)$$

by Assumption 1(iv), we can prove that

$$\begin{aligned}
&\frac{1}{N}\left\|X_t'\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)^+\dot{\boldsymbol{V}}_{NT}^+\dot{u}_{NT,6}'\varepsilon_t\right\| \\
=\;& \frac{1}{N}\left\|X_t'\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big(\dot{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}\big)^+\dot{\boldsymbol{V}}_{NT}^+\left(\frac{1}{NT}\sum_{s=1}^{T}\varepsilon_s d_s' X_s'\dot{\boldsymbol{\Lambda}}\right)'\varepsilon_t\right\| \\
=\;& O_P\left(\frac{1}{N^2T}\left\|\sum_{s=1}^{T}\dot{\boldsymbol{\Lambda}}'X_s d_s\varepsilon_s'\varepsilon_t\right\|\right)
\end{aligned}$$

and

$$\begin{aligned}\frac{1}{N^2T}\left\|\sum_{s=1}^{T}\dot{\mathbf{\Lambda}}'X_sd_s\varepsilon_s'\varepsilon_t\right\| &\leq N^{-1/2}\left(\frac{1}{N^2T}\sum_{s=1}^{T}\left\|\dot{\mathbf{\Lambda}}'X_sd_s\right\|^2\right)^{1/2}\cdot\left(\frac{1}{NT}\sum_{s=1}^{T}\left\|\varepsilon_s'\varepsilon_t\right\|^2\right)^{1/2}\\ &= O_P\left(\delta_{NT}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\|d_s\|^2\right)^{1/2}\right),\end{aligned}$$

which together with Lemma C.2, indicate that

$$\frac{1}{N}\left\|X_t'\mathbf{\Lambda}^0\dot{\boldsymbol{H}}(\dot{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\dot{\boldsymbol{H}})^{+}\dot{\boldsymbol{V}}_{NT}^{+}\dot{u}_{NT,6}'\varepsilon_t\right\| = O_P\left(\delta_{NT}^{-1}\dot{\eta}_{NT}^{1/2}\right). \tag{C.49}$$

Similarly, we can also show that

$$\begin{aligned}&\frac{1}{N}\left\|X_t'\mathbf{\Lambda}^0\dot{\boldsymbol{H}}(\dot{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\dot{\boldsymbol{H}})^{+}\dot{\boldsymbol{V}}_{NT}^{+}\dot{u}_{NT,7}'\varepsilon_t\right\|\\ =\;&\frac{1}{N}\left\|X_t'\mathbf{\Lambda}^0\dot{\boldsymbol{H}}(\dot{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\dot{\boldsymbol{H}})^{+}\dot{\boldsymbol{V}}_{NT}^{+}\left(\frac{1}{NT}\sum_{s=1}^{T}\varepsilon_sf_s^{0\prime}\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}\right)'\varepsilon_t\right\| = O_P(1)\frac{1}{NT}\left\|\sum_{s=1}^{T}f_s^0\varepsilon_s'\varepsilon_t\right\|\\ =\;&O_P(N^{-1/2})\left(\frac{1}{T}\sum_{s=1}^{T}\left\|f_s^0\right\|^2\right)^{1/2}\cdot\left(\frac{1}{NT}\sum_{s=1}^{T}\left\|\varepsilon_s'\varepsilon_t\right\|^2\right)^{1/2} = O_P\left(\delta_{NT}^{-1}\right).\end{aligned} \tag{C.50}$$

Then, by (C.48)–(C.50) and using the fact that $\dot{\eta}_{NT} = o_P(1)$ in Lemma C.2, we can readily prove that

$$\frac{1}{N}\left\|X_t'\dot{v}_{NT,5}\varepsilon_t\right\| = O_P\left(\delta_{NT}^{-1}\right). \tag{C.51}$$

Then we complete the proof of (C.46), which implies that

$$\frac{1}{N}\left\|X_t'(\boldsymbol{M}_{\dot{\mathbf{\Lambda}}} - \boldsymbol{M}_{\mathbf{\Lambda}^0})\varepsilon_t\right\| = O_P\left(\delta_{NT}^{-1}\right). \tag{C.52}$$

We next consider the term $\frac{1}{N}X_t'\boldsymbol{M}_{\dot{\mathbf{\Lambda}}}(\mathbf{\Lambda}^0 - \dot{\mathbf{\Lambda}}\dot{\boldsymbol{H}}^{+})f_t^0$. Note that

$$\frac{1}{N}X_t'\boldsymbol{M}_{\dot{\mathbf{\Lambda}}}(\mathbf{\Lambda}^0-\dot{\mathbf{\Lambda}}\dot{\boldsymbol{H}}^{+})f_t^0 = \frac{1}{N}X_t'\boldsymbol{M}_{\mathbf{\Lambda}^0\dot{\boldsymbol{H}}}(\mathbf{\Lambda}^0-\dot{\mathbf{\Lambda}}\dot{\boldsymbol{H}}^{+})f_t^0+\frac{1}{N}X_t'(\boldsymbol{M}_{\dot{\mathbf{\Lambda}}}-\boldsymbol{M}_{\mathbf{\Lambda}^0\dot{\boldsymbol{H}}})(\mathbf{\Lambda}^0-\dot{\mathbf{\Lambda}}\dot{\boldsymbol{H}}^{+})f_t^0. \tag{C.53}$$

Applying Lemmas C.3(i) and (v), we can find that $\frac{1}{N}X_t'\boldsymbol{M}_{\mathbf{\Lambda}^0\dot{\boldsymbol{H}}}(\mathbf{\Lambda}^0 - \dot{\mathbf{\Lambda}}\dot{\boldsymbol{H}}^{+})f_t^0$ is the leading term, which will be the major focus in the following proof. Note that

$$\mathbf{\Lambda}^0 - \dot{\mathbf{\Lambda}}\dot{\boldsymbol{H}}^{+} = (\mathbf{\Lambda}^0\dot{\boldsymbol{H}}\dot{\boldsymbol{V}}_{NT} - \dot{\mathbf{\Lambda}}\dot{\boldsymbol{V}}_{NT})\dot{\boldsymbol{V}}_{NT}^{+}\dot{\boldsymbol{H}}^{+}.$$

We can apply the decomposition (C.13) for $\mathbf{\Lambda}^0\dot{\boldsymbol{H}}\dot{\boldsymbol{V}}_{NT}-\dot{\mathbf{\Lambda}}\dot{\boldsymbol{V}}_{NT}$, use the fact that $\boldsymbol{M}_{\mathbf{\Lambda}^0\dot{\boldsymbol{H}}}\mathbf{\Lambda}^0\dot{\boldsymbol{H}} = \mathbf{0}$ and both $\dot{\boldsymbol{H}}$ and $\dot{\boldsymbol{V}}_{NT}$ are asymptotically invertible, and then obtain

$$\frac{1}{N}X_t'\boldsymbol{M}_{\mathbf{\Lambda}^0\dot{\boldsymbol{H}}}(\mathbf{\Lambda}^0 - \dot{\mathbf{\Lambda}}\dot{\boldsymbol{H}}^{+})f_t^0 = -\frac{1}{N}X_t'\boldsymbol{M}_{\mathbf{\Lambda}^0\dot{\boldsymbol{H}}}\left(\sum_{j=1}^{3}\dot{u}_{NT,j}+\sum_{j=6}^{8}\dot{u}_{NT,j}\right)\dot{\boldsymbol{V}}_{NT}^{+}\dot{\boldsymbol{H}}^{+}f_t^0. \tag{C.54}$$

Similar to the proof of Lemma C.3(i) and using the decomposition $\dot{\mathbf{\Lambda}} = (\dot{\mathbf{\Lambda}} - \mathbf{\Lambda}^0 \dot{\boldsymbol{H}}) + \mathbf{\Lambda}^0 \dot{\boldsymbol{H}}$, we may prove that

$$\frac{1}{N}\left\| X_t' \boldsymbol{M}_{\mathbf{\Lambda}^0 \dot{\boldsymbol{H}}} \left(\dot{u}_{NT,1} + \dot{u}_{NT,3} + \sum_{j=6}^{8} \dot{u}_{NT,j} \right) \dot{\boldsymbol{V}}_{NT}^{+} \dot{\boldsymbol{H}}^{+} f_t^0 \right\| = O_P\big(\delta_{NT}^{-1} + \dot{\eta}_{NT}\big). \tag{C.55}$$

Meanwhile, letting $\chi_{st} = f_s^{0\prime}\big(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\big)^{+} f_t^0$, we may also obtain

$$\begin{aligned} -\frac{1}{N} X_t' \boldsymbol{M}_{\mathbf{\Lambda}^0 \dot{\boldsymbol{H}}} \dot{u}_{NT,2} \dot{\boldsymbol{V}}_{NT}^{+} \dot{\boldsymbol{H}}^{+} f_t^0 &= \frac{1}{N^2 T} \sum_{s=1}^{T} X_t' \boldsymbol{M}_{\mathbf{\Lambda}^0 \dot{\boldsymbol{H}}} X_s d_s f_s^{0\prime} \mathbf{\Lambda}^{0\prime} \dot{\mathbf{\Lambda}} \dot{\boldsymbol{V}}_{NT}^{+} \dot{\boldsymbol{H}}^{+} f_t^0 \\ &= \frac{1}{NT} \sum_{s=1}^{T} X_t' \boldsymbol{M}_{\mathbf{\Lambda}^0 \dot{\boldsymbol{H}}} X_s \chi_{st} d_s. \end{aligned} \tag{C.56}$$

Note that

$$\frac{1}{N} X_t' \boldsymbol{M}_{\dot{\mathbf{\Lambda}}} X_t (\dot{\beta}_t - \beta_t^0) \overset{P}{\sim} \frac{1}{N} X_t' \boldsymbol{M}_{\mathbf{\Lambda}^0 \dot{\boldsymbol{H}}} X_t d_t, \tag{C.57}$$

where $a \overset{P}{\sim} b$ denotes $a = b(1 + o_P(1))$. By (C.39), (C.44), and (C.52)–(C.57), we have

$$\left\| \frac{1}{N} X_t' \boldsymbol{M}_{\mathbf{\Lambda}^0 \dot{\boldsymbol{H}}} X_t d_t - \frac{1}{NT} \sum_{s=1}^{T} X_t' \boldsymbol{M}_{\mathbf{\Lambda}^0 \dot{\boldsymbol{H}}} X_s \chi_{st} d_s \right\| = O_P\left(p^{1/2} N^{-1/2} + T^{-1/2} + \dot{\eta}_{NT} \right). \tag{C.58}$$

Let $\boldsymbol{L}_{NT} = \mathsf{diag}\left\{ \frac{1}{N} X_1' \boldsymbol{M}_{\mathbf{\Lambda}^0 \dot{\boldsymbol{H}}} X_1, ..., \frac{1}{N} X_T' \boldsymbol{M}_{\mathbf{\Lambda}^0 \dot{\boldsymbol{H}}} X_T \right\}$ and $\boldsymbol{L}_{NT,*}$ be the $T \times T$ block matrix with the (t, s) block being $\frac{1}{NT} X_t' \boldsymbol{M}_{\mathbf{\Lambda}^0 \dot{\boldsymbol{H}}} X_s \chi_{st}$. By (C.58), we may show that

$$\big(\boldsymbol{L}_{NT} - \boldsymbol{L}_{NT,*}\big) \dot{\boldsymbol{d}}_\beta = \boldsymbol{R}_{NT}, \tag{C.59}$$

where $\dot{\boldsymbol{d}}_\beta$ is defined in the proof of Lemma C.2, $\boldsymbol{R}_{NT} = (R_1', \ldots, R_T')'$ with

$$\|R_t\| = O_P\left(p^{1/2} N^{-1/2} + T^{-1/2} + \dot{\eta}_{NT} \right) \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T} \|R_t\|^2 = O_P\left(pN^{-1} + T^{-1} + \dot{\eta}_{NT}^2 \right).$$

Using the arguments as used in the proofs of Theorem 3.1 and Lemma C.2, we can prove that $\boldsymbol{L}_{NT} - \boldsymbol{L}_{NT,*}$ is asymptotically positive definite with the smallest eigenvalue bounded away from zero. Hence, (C.59) indicates that

$$\frac{1}{T}\|\dot{\boldsymbol{d}}_\beta\|^2 = \frac{1}{T}\sum_{t=1}^{T} \|\dot{\beta}_t - \beta_t^0\|^2 = O_P\left(pN^{-1} + T^{-1} + \dot{\eta}_{NT}^2 \right), \tag{C.60}$$

which, in conjunction with the definition of $\dot{\eta}_{NT}$ in the statement of Lemma C.3, implies that $\frac{1}{T}\|\dot{\boldsymbol{d}}_\beta\|^2 = O_P\left(pN^{-1} + T^{-1} \right)$, and strengthens the consistency result in Lemma C.2. By the fact that the matrix $\frac{1}{N}X_t'\boldsymbol{M}_{\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}}X_t$ is positive definite as well as (C.58) and (C.60), we can prove that

$$\left\|\dot{\beta}_t - \beta_t^0\right\| = O_P\left(p^{1/2}N^{-1/2} + T^{-1/2}\right) = O_P\left(\delta_{p,NT}^{-1}\right)$$

for each t, completing the proof of Lemma B.1 in Appendix B. ■

Proof of Lemma B.2. **(i)** Using the argument in the proof of Lemma C.2 (with some modifications), we may prove that $\eta_{NT} = o_P(1)$. Then, following the proofs of (C.44) and (C.52) above, we can readily show that

$$\frac{1}{N^2T}\sum_{t=1}^{T}\left\|X_t'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\varepsilon_t\right\|^2 = O_P\big(pN^{-1} + T^{-1}\big). \tag{C.61}$$

Furthermore, by the Cauchy-Schwarz inequality, we have

$$\frac{1}{NT}\sum_{t=1}^{T}(\hat{\beta}_t - \beta_t^0)'X_t'\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\varepsilon_t = O_P\big(p^{1/2}\delta_{NT}^{-1}\big)\cdot\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{\beta}_t - \beta_t^0\right\|^2\right)^{1/2} = O_P\left(\delta_{p,NT}^{-1}\eta_{NT}^{1/2}\right). \tag{C.62}$$

(ii) As $\boldsymbol{\Lambda}^{0\prime}\boldsymbol{M}_{\boldsymbol{\Lambda}^0} = \boldsymbol{0}$, we have $\sum_{t=1}^{T} f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}}\varepsilon_t = \sum_{t=1}^{T} f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\big(\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}} - \boldsymbol{M}_{\boldsymbol{\Lambda}^0}\big)\varepsilon_t$. Similar to the decomposition in (C.37), we have

$$\boldsymbol{P}_{\hat{\boldsymbol{\Lambda}}} - \boldsymbol{P}_{\boldsymbol{\Lambda}^0\boldsymbol{H}} = \hat{\boldsymbol{\Lambda}}\big(\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}}\big)^{+}\hat{\boldsymbol{\Lambda}}' - \boldsymbol{\Lambda}^0\boldsymbol{H}\big(\boldsymbol{H}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\boldsymbol{H}\big)^{+}\boldsymbol{H}'\boldsymbol{\Lambda}^{0\prime} \equiv \sum_{j=1}^{7} v_{NT,j}, \tag{C.63}$$

where $\boldsymbol{H} \equiv \boldsymbol{H}_{NT} = \big(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\big)\big(\frac{1}{T}\boldsymbol{\Lambda}^{0\prime}\hat{\boldsymbol{\Lambda}}\big)\boldsymbol{V}_{NT}^{+}$, $\boldsymbol{V}_{NT}$ is defined in (2.7), and $v_{NT,j}$, $j = 1, ..., 7$, are analogously defined as $\dot{v}_{NT,j}$ in the proof of Lemma C.3(v) with $\dot{\boldsymbol{\Lambda}}$ and $\dot{\boldsymbol{H}}$ replaced by $\hat{\boldsymbol{\Lambda}}$ and $\boldsymbol{H}$, respectively. We only need to show that

$$\left|\sum_{t=1}^{T} f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\big(\boldsymbol{M}_{\hat{\boldsymbol{\Lambda}}} - \boldsymbol{M}_{\boldsymbol{\Lambda}^0}\big)\varepsilon_t\right| = \left|\sum_{t=1}^{T} f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\big(\sum_{j=1}^{7} v_{NT,j}\big)\varepsilon_t\right| = O_P\left(\delta_{NT}^{-2} + \delta_{p,NT}^{-1}\eta_{NT}^{1/2}\right). \tag{C.64}$$

When $(\dot{\boldsymbol{\Lambda}}, \dot{\boldsymbol{H}})$ is replaced by $(\hat{\boldsymbol{\Lambda}}, \boldsymbol{H})$, it is easy to verify that the convergence results in Lemma C.3 still hold with $\dot{\eta}_{NT}$ replaced by η_{NT}. By Assumption 1(iii),

$$\left\|\sum_{t=1}^{T}\boldsymbol{\Lambda}^{0\prime}\varepsilon_t f_t^0\right\| = O_P(\sqrt{NT}), \tag{C.65}$$

which together with Lemma C.3 (with some modifications to allow the replacement of $\dot{\eta}_{NT}$, $\dot{\boldsymbol{\Lambda}}$, and $\dot{\boldsymbol{H}}$ by η_{NT}, $\hat{\boldsymbol{\Lambda}}$, and $\boldsymbol{H}$, respectively) indicates that

$$\frac{1}{NT}\left\|\sum_{t=1}^{T} f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\left(v_{NT,2} + v_{NT,4} + v_{NT,7}\right)\varepsilon_t\right\| = O_P\big((NT)^{-1/2}(\delta_{NT}^{-2} + \eta_{NT}^{1/2})\big). \tag{C.66}$$

On the other hand, note that

$$\left\|\sum_{t=1}^{T}\big(\hat{\mathbf{\Lambda}}\boldsymbol{V}_{NT}-\mathbf{\Lambda}^0\boldsymbol{H}\boldsymbol{V}_{NT}\big)'\varepsilon_t f_t^0\right\|=\left\|\sum_{t=1}^{T}\left(\sum_{j=1}^{8}u_{NT,j}\right)'\varepsilon_t f_t^0\right\|, \tag{C.67}$$

where $u_{NT,j}$, $j=1,...,8$, are defined similarly to $\dot{u}_{NT,j}$ in the proof of Lemma C.3 (i) with $\dot{\beta}_t$ and $\dot{\mathbf{\Lambda}}$ replaced by $\hat{\beta}_t$ and $\hat{\mathbf{\Lambda}}$, respectively. Let $\hat{d}_s=\hat{\beta}_s-\beta_s^0$. Then, by the definition of $u_{NT,j}$ and using Assumptions 1(i)–(iii), we can prove that

$$\begin{aligned}\left\|\sum_{t=1}^{T}u_{NT,1}'\varepsilon_t f_t^0\right\| &= \frac{1}{NT}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\mathbf{\Lambda}}'X_s\hat{d}_s\hat{d}_s'X_s'\varepsilon_t f_t^0\right\|\\ &= O_P\big(T^{-1}\big)\cdot\sum_{t=1}^{T}\|f_t^0\|\sum_{s=1}^{T}\|\hat{d}_s\|^2\|X_s'\varepsilon_t\|\\ &= O_P\left(N^{1/2}Tp^{1/2}\eta_{NT}\right),\end{aligned} \tag{C.68}$$

and

$$\begin{aligned}\left\|\sum_{t=1}^{T}u_{NT,2}'\varepsilon_t f_t^0\right\| &= \frac{1}{NT}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\mathbf{\Lambda}}'\mathbf{\Lambda}^0 f_s^0\hat{d}_s'X_s'\varepsilon_t f_t^0\right\|\\ &= O_P(T^{-1})\sum_{t=1}^{T}\|f_t^0\|\sum_{s=1}^{T}\|\hat{d}_s\|\|f_s^0\|\|X_s'\varepsilon_t\|\\ &= O_P\big(N^{1/2}T(p\eta_{NT})^{1/2}\big).\end{aligned} \tag{C.69}$$

By analogous arguments, we can also show that

$$\left\|\sum_{t=1}^{T}u_{NT,4}'\varepsilon_t f_t^0\right\|=O_P\big(N^{1/2}T\eta_{NT}^{1/2}\big). \tag{C.70}$$

On the other hand, using Lemma C.3 we can show that

$$\begin{aligned}
\left\|\sum_{t=1}^{T} u'_{NT,3}\varepsilon_t f_t^0\right\| &= \frac{1}{NT}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\mathbf{\Lambda}}'\varepsilon_s \hat{d}'_s X'_s \varepsilon_t f_t^0\right\| \\
&= \frac{1}{NT}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{H}'\mathbf{\Lambda}'_0\varepsilon_s \hat{d}'_s X'_s \varepsilon_t f_t^0\right\| + \frac{1}{NT}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}_0\boldsymbol{H}\right)'\varepsilon_s \hat{d}'_s X'_s \varepsilon_t f_t^0\right\| \\
&\le \|\boldsymbol{H}\|\left(\frac{1}{NT}\sum_{s=1}^{T}\|\mathbf{\Lambda}'_0\varepsilon_s\|^2\right)^{1/2}\left(\frac{1}{NT}\sum_{s=1}^{T}\left\|\hat{d}'_s\sum_{s=1}^{T}X'_s\varepsilon_t f_t^0\right\|^2\right)^{1/2} \\
&\quad + \left(\frac{1}{NT}\sum_{s=1}^{T}||(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}_0\boldsymbol{H})'\varepsilon_s||^2\right)^{1/2}\left(\frac{1}{NT}\sum_{s=1}^{T}\left\|\hat{d}'_s\sum_{t=1}^{T}X'_s\varepsilon_t f_t^0\right\|^2\right)^{1/2} \\
&= O_P\left(T(p\eta_{NT})^{1/2}\right) + O_P\left((1+N^{1/2}T^{-1/2})(\delta_{NT}^{-1}+\eta_{NT}^{1/2})T(p\eta_{NT})^{1/2}\right) \\
&= O_P\left((1+N^{1/2}T^{-1/2}\delta_{NT}^{-1}+N^{1/2}T^{-1/2}\eta_{NT}^{1/2})T(p\eta_{NT})^{1/2}\right), \qquad \text{(C.71)}
\end{aligned}$$

and analogously

$$\begin{aligned}
\left\|\sum_{t=1}^{T} u'_{NT,5}\varepsilon_t f_t^0\right\| &= \frac{1}{NT}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\mathbf{\Lambda}}'\varepsilon_s f_s^{0\prime}\mathbf{\Lambda}^{0\prime}\varepsilon_t f_t^0\right\| \\
&= \frac{1}{NT}\left\|\sum_{s=1}^{T}\sum_{t=1}^{T}\boldsymbol{H}'\mathbf{\Lambda}'_0\varepsilon_s f_s^{0\prime}\mathbf{\Lambda}^{0\prime}\varepsilon_t f_t^0\right\| + \frac{1}{NT}\left\|\sum_{s=1}^{T}\sum_{t=1}^{T}\left(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}_0\boldsymbol{H}\right)'\varepsilon_s f_s^{0\prime}\mathbf{\Lambda}^{0\prime}\varepsilon_t f_t^0\right\| \\
&\le \|\boldsymbol{H}\|\frac{1}{NT}\|\mathbf{\Lambda}^{0\prime}\boldsymbol{\varepsilon}\boldsymbol{F}^0\|^2 + \frac{1}{NT}\left\|\sum_{s=1}^{T}\left(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}_0\boldsymbol{H}\right)'\varepsilon_s f_s^{0\prime}\right\|\left\|\mathbf{\Lambda}^{0\prime}\boldsymbol{\varepsilon}\boldsymbol{F}^0\right\| \\
&= O_P(1) + O_P\left(N^{1/2}T^{1/2}(\delta_{NT}^{-2}+\eta_{NT}^{1/2})\right). \qquad \text{(C.72)}
\end{aligned}$$

Using the fact that under Assumptions 1(i) and (iv)

$$\sum_{s=1}^{T}\left\|\sum_{t=1}^{T}\varepsilon'_s\varepsilon_t f_t^0\right\|^2 \le \left(\sum_{s=1}^{T}\sum_{t_1=1}^{T}\|\varepsilon'_s\varepsilon_{t_1}\|^2\right)\left(\sum_{t_2=1}^{T}\|f_{t_2}^0\|^2\right) = O_P\left(T^2N(N+T)\right), \qquad \text{(C.73)}$$

we have

$$\begin{aligned}
\left\|\sum_{t=1}^{T} u'_{NT,6}\varepsilon_t f_t^0\right\| &= \frac{1}{NT}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\mathbf{\Lambda}}'X_s\hat{d}_s\varepsilon'_s\varepsilon_t f_t^0\right\| \\
&\le \frac{1}{NT}\max_{1\le s\le T}\|\hat{\mathbf{\Lambda}}'X_s\|\cdot\left(\sum_{s=1}^{T}\|\hat{d}_s\|^2\right)^{1/2}\cdot\left(\sum_{s=1}^{T}\left\|\sum_{t=1}^{T}\varepsilon'_s\varepsilon_t f_t^0\right\|^2\right)^{1/2} \\
&= O_P(T^{-1})\cdot O_P\left(T^{1/2}\eta_{NT}^{1/2}\right)\cdot O_P\left(TN^{1/2}(N^{1/2}+T^{1/2})\right) \\
&= O_P\left(\eta_{NT}^{1/2}(NT^{1/2}+N^{1/2}T)\right). \qquad \text{(C.74)}
\end{aligned}$$

Notice that

$$\begin{aligned}\left\|\sum_{t=1}^{T} u_{NT,8}'\varepsilon_t f_t^0\right\| &= \frac{1}{NT}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\mathbf{\Lambda}}'\varepsilon_s\varepsilon_s'\varepsilon_t f_t^0\right\| \\ &\leq \frac{1}{NT}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{H}'\mathbf{\Lambda}_0'\varepsilon_s\varepsilon_s'\varepsilon_t f_t^0\right\| + \frac{1}{NT}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{\mathbf{\Lambda}}-\mathbf{\Lambda}_0\boldsymbol{H}\right)'\varepsilon_s\varepsilon_s'\varepsilon_t f_t^0\right\|.\end{aligned}$$

For the first term on the right hand side, by the Cauchy-Schwarz inequality and Assumption 1(iii) and (C.73) we may show that

$$\begin{aligned}\frac{1}{NT}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{H}'\mathbf{\Lambda}_0'\varepsilon_s\varepsilon_s'\varepsilon_t f_t^0\right\| &\leq \frac{1}{NT}\|\boldsymbol{H}\|\cdot\left(\sum_{s=1}^{T}\|\mathbf{\Lambda}_0'\varepsilon_s\|^2\right)^{1/2}\cdot\left(\sum_{s=1}^{T}\|\sum_{t=1}^{T}\varepsilon_s'\varepsilon_t f_t^0\|^2\right)^{1/2} \\ &= O_P\big((NT)^{-1/2}\big)O_P\big(TN^{1/2}(N^{1/2}+T^{1/2})\big) = O_P\big((NT)^{1/2}+T\big).\end{aligned}$$

For the second term on the right hand side, by Lemma C.3(vii) (with $\dot{\eta}_{NT}$, $\dot{\mathbf{\Lambda}}$, and $\dot{\boldsymbol{H}}$ replaced by η_{NT}, $\hat{\mathbf{\Lambda}}$, and $\boldsymbol{H}$, respectively), we have

$$\begin{aligned}\frac{1}{NT}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{\mathbf{\Lambda}}-\mathbf{\Lambda}_0\boldsymbol{H}\right)'\varepsilon_s\varepsilon_s'\varepsilon_t f_t^0\right\| &\leq \left(\frac{1}{NT}\sum_{s=1}^{T}\|\left(\hat{\mathbf{\Lambda}}-\mathbf{\Lambda}_0\boldsymbol{H}\right)'\varepsilon_s\|^2\right)^{1/2}\cdot\left(\frac{1}{NT}\sum_{s=1}^{T}\|\sum_{t=1}^{T}\varepsilon_s'\varepsilon_t f_t^0\|^2\right)^{1/2} \\ &= O_P\left((1+N^{1/2}T^{-1/2})(\delta_{NT}^{-1}+\eta_{NT}^{1/2})\right)O_P\left(T+N^{1/2}T^{1/2}\right) \\ &= O_P\left((T+N)(\delta_{NT}^{-1}+\eta_{NT}^{1/2})\right).\end{aligned}$$

It follows that

$$\left\|\sum_{t=1}^{T} u_{NT,8}'\varepsilon_t f_t^0\right\| = O_P\left((NT)^{1/2}+T+N\eta_{NT}^{1/2}\right). \tag{C.75}$$

Finally, noting that $\left|\sum_{s=1}^{T}\sum_{t=1}^{T} f_s^{0\prime}\varepsilon_s'\varepsilon_t f_t^0\right| = O_P(NT)$ by Assumption 1(iv), we can also show that

$$\left\|\sum_{t=1}^{T} u_{NT,7}'\varepsilon_t f_t^0\right\| = O_P(N). \tag{C.76}$$

By (C.67)–(C.76), we have

$$\frac{1}{NT}\left\|\sum_{t=1}^{T}\left(\hat{\mathbf{\Lambda}}\boldsymbol{V}_{NT}-\mathbf{\Lambda}^0\boldsymbol{H}\boldsymbol{V}_{NT}\right)'\varepsilon_t f_t^0\right\| = O_P\left(\delta_{NT}^{-2}+\delta_{p,NT}^{-1}\eta_{NT}^{1/2}\right). \tag{C.77}$$

With this, we readily prove that

$$\frac{1}{NT}\left\|\sum_{t=1}^{T} f_t^{0\prime}\mathbf{\Lambda}^{0\prime}(v_{NT,1}+v_{NT,3}+v_{NT,5}+v_{NT,6})\varepsilon_t\right\| = O_P\left(\delta_{NT}^{-2}+\delta_{p,NT}^{-1}\eta_{NT}^{1/2}\right), \tag{C.78}$$

which together with (C.66), leads to (C.64). Hence, we complete the proof of (ii).

(iii) This follows from Lemmas C.1(iii) and (iv). ■

Before proving Lemma B.3 in Appendix B, we need to introduce two technical lemmas. The first lemma is similar to Lemma C.3 with the preliminary estimates replaced by the post-LASSO estimates. Let $\tilde{\mathbf{\Lambda}}_{m^0} = \tilde{\mathbf{\Lambda}}(\mathcal{T}^0_{m^0})$ be the infeasible estimate of the factor loadings in the post-LASSO estimation procedure, $\tilde{\boldsymbol{H}} = \big(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\big)\big(\frac{1}{N}\mathbf{\Lambda}^{0\prime}\ \tilde{\mathbf{\Lambda}}_{m^0}\big)\tilde{\boldsymbol{V}}^+_{NT}$ with $\tilde{\boldsymbol{V}}_{NT}$ defined in the proof of Theorem 3.4 in Appendix B, and $\tilde{\eta}_{NT} = \frac{1}{m^0}\sum_{j=1}^{m^0+1}\|\tilde{\alpha}_{m^0j} - \alpha^0_j\|^2$, where $\tilde{\alpha}_{m^0j}$ is the j-th p-dimensional element of the infeasible estimate $\tilde{\boldsymbol{\alpha}}_{m^0} = \tilde{\boldsymbol{\alpha}}_{m^0}(\mathcal{T}^0_{m^0})$.

Lemma C.4 *Suppose that the conditions in Theorem 3.4 hold. Then we have*

(i) $\frac{1}{N}\big\|\tilde{\mathbf{\Lambda}}_{m^0} - \mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big\|^2 = O_P\big(\delta^{-2}_{NT} + \tilde{\eta}_{NT}\big)$,

(ii) $\frac{1}{N}\big(\tilde{\mathbf{\Lambda}}_{m^0} - \mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)'\mathbf{\Lambda}^0\tilde{\boldsymbol{H}} = O_P\big(\delta^{-2}_{NT} + \tilde{\eta}^{1/2}_{NT}\big)$,

(iii) $\frac{1}{N}\big(\tilde{\mathbf{\Lambda}}_{m^0} - \mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)'\tilde{\mathbf{\Lambda}}_{m^0} = O_P\big(\delta^{-2}_{NT} + \tilde{\eta}^{1/2}_{NT}\big)$,

(iv) $\frac{1}{N}\big(\tilde{\mathbf{\Lambda}}'_{m^0}\tilde{\mathbf{\Lambda}}_{m^0} - \tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big) = O_P\big(\delta^{-2}_{NT} + \tilde{\eta}^{1/2}_{NT}\big)$,

(v) $\big\|\boldsymbol{P}_{\tilde{\mathbf{\Lambda}}_{m^0}} - \boldsymbol{P}_{\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}}\big\| = O_P\big(\delta^{-1}_{NT} + \tilde{\eta}^{1/2}_{NT}\big)$,

(vi) $\frac{1}{NT}\sum_{s=1}^T(\tilde{\mathbf{\Lambda}}_{m^0} - \mathbf{\Lambda}^0\tilde{\boldsymbol{H}})'\varepsilon_s\gamma'_s = O_P\big(\delta^{-2}_{NT} + \tilde{\eta}^{1/2}_{NT}\big)$ *with* $\gamma_s = 1$ *or* f^0_s, *and*

(vii) $\frac{1}{NT}\sum_{s=1}^T||(\tilde{\mathbf{\Lambda}}_{m^0} - \mathbf{\Lambda}^0\tilde{\boldsymbol{H}})'\varepsilon_s||^2 = O_P\big((1 + NT^{-1})(\delta^{-2}_{NT} + \tilde{\eta}_{NT})\big)$.

Proof of Lemma C.4. The proof is analogous to that of Lemma C.3. Hence, we only sketch it. For notational simplicity, we let $\tilde{\boldsymbol{V}} \equiv \tilde{\boldsymbol{V}}_{NT}$, and $\tilde{\eta}_j = \tilde{\alpha}_{m^0j} - \alpha^0_j$, $j = 1, ..., m^0 + 1$. By (B.25) in the proof of Theorem 3.4, we have

$$
\begin{aligned}
&\tilde{\mathbf{\Lambda}}_{m^0}\tilde{\boldsymbol{V}} - \mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\tilde{\boldsymbol{V}}\\
= &\left[\frac{1}{NT}\sum_{j=1}^{m^0+1}\sum_{t=T^0_{j-1}}^{T^0_j-1}\big(Y_t - X_t\tilde{\alpha}_{m^0j}\big)\big(Y_t - X_t\tilde{\alpha}_{m^0j}\big)'\right]\tilde{\mathbf{\Lambda}}_{m^0} - \mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\tilde{\boldsymbol{V}}\\
= &\left[\frac{1}{NT}\sum_{j=1}^{m^0+1}\sum_{t=T^0_{j-1}}^{T^0_j-1}\big(-X_t\tilde{\eta}_j + \mathbf{\Lambda}^0 f^0_t + \varepsilon_t\big)\big(-X_t\tilde{\eta}_j + \mathbf{\Lambda}^0 f^0_t + \varepsilon_t\big)'\right]\tilde{\mathbf{\Lambda}}_{m^0} - \mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\tilde{\boldsymbol{V}}\\
= &\frac{1}{NT}\sum_{j=1}^{m^0+1}\sum_{t=T^0_{j-1}}^{T^0_j-1}X_t\tilde{\eta}_j\tilde{\eta}'_jX'_t\tilde{\mathbf{\Lambda}}_{m^0} - \frac{1}{NT}\sum_{j=1}^{m^0+1}\sum_{t=T^0_{j-1}}^{T^0_j-1}X_t\tilde{\eta}_jf^{0\prime}_t\mathbf{\Lambda}^{0\prime}\tilde{\mathbf{\Lambda}}_{m^0} - \frac{1}{NT}\sum_{j=1}^{m^0+1}\sum_{t=T^0_{j-1}}^{T^0_j-1}X_t\tilde{\eta}_j\varepsilon'_t\tilde{\mathbf{\Lambda}}_{m^0}
\end{aligned}
$$

$$
\begin{aligned}
& -\frac{1}{NT}\sum_{j=1}^{m^0+1}\sum_{t=T_{j-1}^0}^{T_j^0-1}\mathbf{\Lambda}^0 f_t^0\tilde{\eta}_j' X_t'\tilde{\mathbf{\Lambda}}_{m^0}+\frac{1}{NT}\sum_{t=1}^{T}\mathbf{\Lambda}^0 f_t^0\varepsilon_t'\tilde{\mathbf{\Lambda}}_{m^0}-\frac{1}{NT}\sum_{j=1}^{m^0+1}\sum_{t=T_{j-1}^0}^{T_j^0-1}\varepsilon_t\tilde{\eta}_j' X_t'\tilde{\mathbf{\Lambda}}_{m^0} \\
& +\frac{1}{NT}\sum_{t=1}^{T}\varepsilon_t f_t^{0\prime}\mathbf{\Lambda}^{0\prime}\tilde{\mathbf{\Lambda}}_{m^0}+\frac{1}{NT}\sum_{t=1}^{T}\varepsilon_t\varepsilon_t'\tilde{\mathbf{\Lambda}}_{m^0} \\
\equiv & \sum_{j=1}^{8}\tilde{u}_{NT,j}. \qquad \text{(C.79)}
\end{aligned}
$$

Then following the proof of Lemma C.3 with $\dot{\mathbf{\Lambda}}$ and d_t replaced by $\tilde{\mathbf{\Lambda}}_{m^0}$ and $\tilde{\eta}_j$, respectively, and using Assumption 3(ii), we can readily prove Lemma C.4(i). Note that

$$
\frac{1}{N}\big(\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)'\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}=\frac{1}{N}\sum_{j=1}^{8}\tilde{\boldsymbol{V}}^{+}\tilde{u}_{NT,j}'\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\equiv\frac{1}{N}\sum_{j=1}^{8}\tilde{u}_{NT,j}^{*}. \qquad \text{(C.80)}
$$

Then following the proof of Lemma C.3(ii) and using Lemma C.4(i), we readily prove Lemma C.4(ii). The results in (iii) and (iv) can be proved by combining Lemmas C.4(i) and (ii). Similar to (C.37), we have the following decomposition:

$$
\boldsymbol{P}_{\tilde{\mathbf{\Lambda}}_{m^0}}-\boldsymbol{P}_{\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}}=\tilde{\mathbf{\Lambda}}_{m^0}\big(\tilde{\mathbf{\Lambda}}_{m^0}'\tilde{\mathbf{\Lambda}}_{m^0}\big)^{+}\tilde{\mathbf{\Lambda}}_{m^0}'-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big(\tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)^{+}\tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\equiv\sum_{j=1}^{7}\tilde{v}_{NT,j}, \qquad \text{(C.81)}
$$

where

$$
\begin{aligned}
\tilde{v}_{NT,1} &= \big(\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)\big(\tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)^{+}\big(\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)', \\
\tilde{v}_{NT,2} &= \big(\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)\big(\tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)^{+}\tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}, \\
\tilde{v}_{NT,3} &= \big(\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)\big[\big(\tilde{\mathbf{\Lambda}}_{m^0}'\tilde{\mathbf{\Lambda}}_{m^0}\big)^{+}-\big(\tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)^{+}\big]\big(\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)', \\
\tilde{v}_{NT,4} &= \big(\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)\big[\big(\tilde{\mathbf{\Lambda}}_{m^0}'\tilde{\mathbf{\Lambda}}_{m^0}\big)^{+}-\big(\tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)^{+}\big]\tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}, \\
\tilde{v}_{NT,5} &= \mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big(\tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)^{+}\big(\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)', \\
\tilde{v}_{NT,6} &= \mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big[\big(\tilde{\mathbf{\Lambda}}_{m^0}'\tilde{\mathbf{\Lambda}}_{m^0}\big)^{+}-\big(\tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)^{+}\big]\big(\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)', \\
\tilde{v}_{NT,7} &= \mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big[\big(\tilde{\mathbf{\Lambda}}_{m^0}'\tilde{\mathbf{\Lambda}}_{m^0}\big)^{+}-\big(\tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\big)^{+}\big]\tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}.
\end{aligned}
$$

By (C.81) and Lemmas C.4(i) and (iv), we can prove (v). The proofs of (vi) and (vii) parallel to those of Lemmas C.3(vi) and (vii). We have thus completed the proof of Lemma C.4. ■

Lemma C.5 *Suppose that the conditions in Theorem 3.4 hold. Then we have*
(i) $\tilde{\eta}_{NT}=\frac{1}{m^0}\sum_{j=1}^{m^0+1}\|\tilde{\alpha}_{m^0 j}-\alpha_j^0\|^2=O_P\big(\delta_{p,NT}^{-2}\big)$,

(ii) $\frac{1}{N}(\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}})'\varepsilon_t = \tilde{\boldsymbol{H}}'\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+\left(\frac{1}{NT}\sum_{s=1}^T f_s^0\varepsilon_s'\varepsilon_t\right)+O_P\left(\delta_{NT}^{-1}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\|\right)$ $+O_P\left(\delta_{p,NT}^{-3}\right)$ *for* $t=1,...,T$,

(iii) $\frac{1}{N\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1}X_t'\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\left(\tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\right)^+\left(\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\right)'\varepsilon_t-\sum_{t=T_{j-1}^0}^{T_j^0-1}X_t'\mathbf{\Lambda}^0\left(\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\right)^+\right.$ $\left.\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+\left(\frac{1}{NT}\sum_{s=1}^T f_s^0\varepsilon_s'\varepsilon_t\right)\right\| = O_P\left(\delta_{NT}^{-1}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\|\right)+O_P\left(\delta_{p,NT}^{-3}\right)$ *for* $j=1,...,m^0+1$,

(iv) $\frac{1}{NT}\sum_{t=1}^T\left\|(\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}})'\varepsilon_t f_t^0\right\| = O_P\left(\delta_{p,NT}^{-2}\right)$.

Proof of Lemma C.5. As the proof of the convergence rates for $\tilde{\boldsymbol{\alpha}}_{m^0}$ in (i) is similar to the proof of Lemma B.1, we omit the details. Furthermore, the results in (iii) and (iv) can be easily proved by using (ii). Hence we only focus on the proof of the result in (ii).

Note that for any $t=1,...,T$,

$$\frac{1}{N}(\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}})'\varepsilon_t = \frac{1}{N}\tilde{\boldsymbol{V}}^+(\tilde{\mathbf{\Lambda}}_{m^0}\tilde{\boldsymbol{V}}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\tilde{\boldsymbol{V}})'\varepsilon_t = \frac{1}{N}\tilde{\boldsymbol{V}}^+(\sum_{k=1}^8\tilde{u}_{NT,k})'\varepsilon_t \tag{C.82}$$

by using (C.79) in the proof of Lemma C.4. By Lemma C.5(i), Assumptions 1(ii), (iii) and 3(ii), and the Jensen inequality, we have

$$\begin{aligned}
\frac{1}{N}\left\|\tilde{\boldsymbol{V}}^+\tilde{u}_{NT,1}'\varepsilon_t\right\| &= \frac{1}{N^2T}\left\|\tilde{\boldsymbol{V}}^+\sum_{k=1}^{m^0+1}\sum_{s=T_{k-1}^0}^{T_k^0-1}\tilde{\mathbf{\Lambda}}_{m^0}'X_s\tilde{\eta}_k\tilde{\eta}_k'X_s'\varepsilon_t\right\| \\
&= O_P(N^{-2}T^{-1})\left\|\tilde{\mathbf{\Lambda}}_{m^0}\right\|\max_{1\le s\le T}\mu_{\max}^{1/2}(X_s'X_s)\cdot\sum_{k=1}^{m^0+1}\|\tilde{\eta}_k\|^2\sum_{s=T_{k-1}^0}^{T_k^0-1}\|X_s'\varepsilon_t\| \\
&= O_P\left(p^{1/2}N^{-1/2}\tilde{\eta}_{NT}\right) = O_P\left(\delta_{p,NT}^{-3}\right).
\end{aligned} \tag{C.83}$$

By Lemmas C.4(i) and C.5(i) and Assumptions 1(iii), (iv) and 3(ii), we can show that

$$\begin{aligned}
&\frac{1}{N}\|\tilde{\boldsymbol{V}}^+\tilde{u}_{NT,3}'\varepsilon_t\| \\
=\ & \frac{1}{N^2T}\left\|\tilde{\boldsymbol{V}}^+\left[\sum_{k=1}^{m^0+1}\sum_{s=T_{k-1}^0}^{T_k^0-1}\tilde{\boldsymbol{H}}'\mathbf{\Lambda}^{0\prime}\varepsilon_s\tilde{\eta}_k'X_s'\varepsilon_t+\sum_{k=1}^{m^0+1}\sum_{s=T_{k-1}^0}^{T_k^0-1}(\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}})'\varepsilon_s\tilde{\eta}_k'X_s'\varepsilon_t\right]\right\| \\
=\ & O_P(N^{-2}T^{-1})\left[\sum_{k=1}^{m^0+1}\|\tilde{\eta}_k\|\sum_{s=T_{k-1}^0}^{T_k^0-1}\|\mathbf{\Lambda}^{0\prime}\varepsilon_s\|\|X_s'\varepsilon_t\|+\|\tilde{\mathbf{\Lambda}}_{m^0}-\mathbf{\Lambda}^0\tilde{\boldsymbol{H}}\|\sum_{k=1}^{m^0+1}\|\tilde{\eta}_k\|\sum_{s=T_{k-1}^0}^{T_k^0-1}\|\varepsilon_s\|\|X_s'\varepsilon_t\|\right] \\
=\ & O_P\left(N^{-1}(p\tilde{\eta}_{NT})^{1/2}\right)+O_P\left(N^{-1/2}(\delta_{NT}^{-1}+\tilde{\eta}_{NT}^{1/2})(p\tilde{\eta}_{NT})^{1/2}\right) \\
=\ & O_P\left(N^{-1}(p\tilde{\eta}_{NT})^{1/2}+\delta_{p,NT}^{-3}\right).
\end{aligned} \tag{C.84}$$

By Assumptions 1(i), (iii) and 3(ii), and Lemma C.5(i), we have

$$\begin{aligned}
\frac{1}{N}\tilde{\boldsymbol{V}}^{+}\tilde{u}_{NT,4}'\varepsilon_t &= \frac{1}{N^2T}\tilde{\boldsymbol{V}}^{+}\left(\sum_{k=1}^{m^0+1}\sum_{s=T_{k-1}^0}^{T_k^0-1}\tilde{\boldsymbol{\Lambda}}_{m^0}'X_s\tilde{\eta}_k f_s^{0\prime}\boldsymbol{\Lambda}^{0\prime}\right)\varepsilon_t \\
&= O_P\big(N^{-2}T^{-1}\big)\cdot\sum_{k=1}^{m^0+1}\|\tilde{\eta}_k\|\left(\sum_{s=T_{k-1}^0}^{T_k^0-1}\|\tilde{\boldsymbol{\Lambda}}_{m^0}'X_s\|\|f_s^0\|\|\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}\|\right) \\
&= O_P\left(\delta_{NT}^{-1}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\|\right). \qquad (C.85)
\end{aligned}$$

Analogously, we can show that

$$\frac{1}{N}\tilde{\boldsymbol{V}}^{+}\tilde{u}_{NT,2}'\varepsilon_t = O_P\left(\delta_{NT}^{-1}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\|\right). \qquad (C.86)$$

By Assumptions 1(iii) and (iv), we can prove that

$$\begin{aligned}
\frac{1}{N}\tilde{\boldsymbol{V}}^{+}\tilde{u}_{NT,5}'\varepsilon_t &= \frac{1}{N^2T}\tilde{\boldsymbol{V}}^{+}\left(\sum_{s=1}^{T}\boldsymbol{\Lambda}^0 f_s^0\varepsilon_s'\tilde{\boldsymbol{\Lambda}}_{m^0}\right)'\varepsilon_t \\
&= \frac{1}{N^2T}\tilde{\boldsymbol{V}}^{+}\sum_{s=1}^{T}\left(\tilde{\boldsymbol{\Lambda}}_{m^0}-\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}}\right)'\varepsilon_s f_s^0\boldsymbol{\Lambda}^{0\prime}\varepsilon_t + \frac{1}{N^2T}\tilde{\boldsymbol{V}}^{+}\left(\sum_{s=1}^{T}\tilde{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\varepsilon_s f_s^{0\prime}\boldsymbol{\Lambda}^{0\prime}\varepsilon_t\right) \\
&= \frac{1}{N^2T}\tilde{\boldsymbol{V}}^{+}\sum_{s=1}^{T}\left(\tilde{\boldsymbol{\Lambda}}_{m^0}-\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}}\right)'\varepsilon_s f_s^0\boldsymbol{\Lambda}^{0\prime}\varepsilon_t + O_P\left(\frac{1}{N^2T}\|\sum_{s=1}^{T}\boldsymbol{\Lambda}^{0\prime}\varepsilon_s f_s^0\|\|\boldsymbol{\Lambda}^{0\prime}\varepsilon_t\|\right) \\
&= \frac{1}{N^2T}\tilde{\boldsymbol{V}}^{+}\sum_{s=1}^{T}\left(\tilde{\boldsymbol{\Lambda}}_{m^0}-\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}}\right)'\varepsilon_s f_s^0\boldsymbol{\Lambda}^{0\prime}\varepsilon_t + O_P\left(\delta_{NT}^{-3}\right). \qquad (C.87)
\end{aligned}$$

By Assumptions 1(ii), (iv) and Lemma C.5(i), we have

$$\begin{aligned}
\frac{1}{N}\tilde{\boldsymbol{V}}^{+}\tilde{u}_{NT,6}'\varepsilon_t &= \frac{1}{N^2T}\tilde{\boldsymbol{V}}^{+}\left(\sum_{k=1}^{m^0+1}\sum_{s=T_{k-1}^0}^{T_k^0-1}\tilde{\boldsymbol{\Lambda}}_{m^0}'X_s\tilde{\eta}_k\varepsilon_s'\right)\varepsilon_t \\
&= O_P\big(N^{-2}T^{-1}\big)\cdot\sum_{k=1}^{m^0+1}\|\tilde{\eta}_k\|\left[\sum_{s=T_{k-1}^0}^{T_k^0-1}\|\tilde{\boldsymbol{\Lambda}}_{m^0}'X_s\|\left\|\sum_{i=1}^{N}\varepsilon_{it}\varepsilon_{is}\right\|\right] \\
&= O_P\left(\delta_{NT}^{-1}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\|\right) \qquad (C.88)
\end{aligned}$$

By the definition of $\tilde{\boldsymbol{H}}$ and noting that $\tilde{\boldsymbol{V}}_{NT}^{+}$ is diagonal, we have

$$\frac{1}{N}\tilde{\boldsymbol{V}}^{+}\tilde{u}_{NT,7}'\varepsilon_t = \left(\frac{1}{N}\tilde{\boldsymbol{V}}^{+}\tilde{\boldsymbol{\Lambda}}_{m^0}'\boldsymbol{\Lambda}^0\right)\left[\frac{1}{NT}\sum_{s=1}^{T}f_s^0\varepsilon_s'\varepsilon_t\right] = \tilde{\boldsymbol{H}}'\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^{+}\left[\frac{1}{NT}\sum_{s=1}^{T}f_s^0\varepsilon_s'\varepsilon_t\right]. \qquad (C.89)$$

By the definition of $\tilde{u}_{NT,8}$ and Assumption 3(iii),

$$\begin{aligned}\frac{1}{N}\tilde{\boldsymbol{V}}^{+}\tilde{u}'_{NT,8}\varepsilon_t &= \frac{1}{N^2T}\tilde{\boldsymbol{V}}^{+}\sum_{s=1}^{T}\left(\tilde{\boldsymbol{\Lambda}}_{m^0}-\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}}\right)'\varepsilon_s\varepsilon_s'\varepsilon_t+\frac{1}{N^2T}\tilde{\boldsymbol{V}}^{+}\tilde{\boldsymbol{H}}'\sum_{s=1}^{T}\boldsymbol{\Lambda}^{0\prime}\varepsilon_s\varepsilon_s'\varepsilon_t \\ &= \frac{1}{N^2T}\tilde{\boldsymbol{V}}^{+}\sum_{s=1}^{T}\left(\tilde{\boldsymbol{\Lambda}}_{m^0}-\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}}\right)'\varepsilon_s\varepsilon_s'\varepsilon_t+O_P\left(\delta_{NT}^{-3}\right). \qquad (C.90)\end{aligned}$$

Combining the results in (C.82)–(C.90) yields

$$\begin{aligned}\frac{1}{N}(\tilde{\boldsymbol{\Lambda}}_{m^0}-\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}})'\varepsilon_t &= \tilde{\boldsymbol{H}}'\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^{+}\frac{1}{NT}\sum_{s=1}^{T}f_s^0\varepsilon_s'\varepsilon_t+\frac{1}{N^2T}\tilde{\boldsymbol{V}}^{+}\sum_{s=1}^{T}\left(\tilde{\boldsymbol{\Lambda}}_{m^0}-\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}}\right)'\varepsilon_s f_s^0\boldsymbol{\Lambda}^{0\prime}\varepsilon_t \\ &\quad +\frac{1}{N^2T}\tilde{\boldsymbol{V}}^{+}\sum_{s=1}^{T}\left(\tilde{\boldsymbol{\Lambda}}_{m^0}-\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}}\right)'\varepsilon_s\varepsilon_s'\varepsilon_t+O_P\left(\delta_{p,NT}^{-3}\right) \\ &\quad +O_P\left(\delta_{NT}^{-1}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\|\right). \qquad (C.91)\end{aligned}$$

By Assumptions 1(i) and (iv), the first term on the right hand side of (C.91) is $O_P\left(\delta_{NT}^{-2}\right)$; by Assumptions 1(iii) and Lemmas C.4(vi) and C.5(i) we can show the second term is $O_P\left(\delta_{p,NT}^{-1}\delta_{NT}^{-1}\right)$; by Assumptions 1(iii) and (iv) and Lemma C.4(vii) and , we can show the third and fourth terms are $O_P\left(\delta_{p,NT}^{-1}\delta_{NT}^{-1}\right)$. It follows that

$$\frac{1}{N}\left(\tilde{\boldsymbol{\Lambda}}_{m^0}-\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}}\right)'\varepsilon_t=O_P\left(\delta_{p,NT}^{-1}\delta_{NT}^{-1}\right). \qquad (C.92)$$

By (C.92) and following the above arguments, we can further show that the second and third terms on the right hand side of (C.91) is $O_P\left(\delta_{p,NT}^{-3}\right)$. This completes the proof of Lemma C.5(ii). ■

Proof of Lemma B.3. For notional simplicity, we let $\tilde{\boldsymbol{\Lambda}}=\tilde{\boldsymbol{\Lambda}}_{m^0}$ throughout this proof.

(i) Noting that

$$-(\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}-\boldsymbol{M}_{\boldsymbol{\Lambda}^0})=\tilde{\boldsymbol{\Lambda}}(\tilde{\boldsymbol{\Lambda}}'\tilde{\boldsymbol{\Lambda}})^{+}\tilde{\boldsymbol{\Lambda}}'-\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}}(\tilde{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}})^{+}\tilde{\boldsymbol{H}}'\boldsymbol{\Lambda}^{0\prime}=\sum_{k=1}^{7}\tilde{v}_{NT,k} \qquad (C.93)$$

and by using the decomposition (C.81), we have

$$\frac{1}{N\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1}X_t'(\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}-\boldsymbol{M}_{\boldsymbol{\Lambda}^0})\varepsilon_t=-\frac{1}{N\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1}X_t'\left(\sum_{k=1}^{7}\tilde{v}_{NT,k}\right)\varepsilon_t. \qquad (C.94)$$

By (C.94), Lemmas C.4(i), (iv) and C.5(iii), we can prove that for any $j = 1, ..., m^0 + 1$,

$$
\begin{aligned}
& \left\| \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'(\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} - \boldsymbol{M}_{\boldsymbol{\Lambda}^0})\varepsilon_t + B_{NT,j}(2,1) \right\| \\
\leq & \left\| \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\Big(\sum_{j=1,\neq 5}^{7} \tilde{v}_{NT,j}\Big)\varepsilon_t \right\| + \left\| \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\tilde{v}_{NT,5}\varepsilon_t - B_{NT,j}(2,1) \right\| \\
= & \left\| \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\tilde{v}_{NT,5}\varepsilon_t - B_{NT,j}(2,1) \right\| + O_P\left(\delta_{NT}^{-1}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m_0} - \boldsymbol{\alpha}^0\| + \delta_{p,NT}^{-3} \right) \\
= & \; O_P\left(\delta_{NT}^{-1}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m_0} - \boldsymbol{\alpha}^0\| + \delta_{p,NT}^{-3} \right) \qquad \text{(C.95)}
\end{aligned}
$$

which completes the proof of Lemma B.3(i).

(ii) Noting that for any $j = 1, ..., m^0 + 1$,

$$
\frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}(\boldsymbol{\Lambda}^0 - \tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{H}}^+)f_t^0 = \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}(\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}}\tilde{\boldsymbol{V}} - \tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{V}})\tilde{\boldsymbol{V}}^+\tilde{\boldsymbol{H}}^+ f_t^0,
$$

and $\tilde{\boldsymbol{V}}^+\tilde{\boldsymbol{H}}^+ = \big(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\big)^+\big(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\big)^+$, by the decomposition (C.79), we have

$$
\begin{aligned}
& \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}(\boldsymbol{\Lambda}^0 - \tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{H}}^+)f_t^0 \\
= & -\frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \left(\sum_{l=1}^{8} \tilde{u}_{NT,l} \right) \left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}} \right)^+ \left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0 \right)^+ f_t^0. \qquad \text{(C.96)}
\end{aligned}
$$

We next analyze each term on the right hand side of the equation (C.96).

For $l=1$, by the definition of $\tilde{u}_{NT,1}$, Assumptions 1(i)(ii), and Lemma C.5(i), we have

$$\begin{aligned}
&\frac{1}{N\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}\tilde{u}_{NT,1}\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
=\;& \frac{1}{N\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}\left(\frac{1}{NT}\sum_{k=1}^{m^0+1}\sum_{s=T_{k-1}^0}^{T_k^0-1} X_s\tilde{\eta}_k\tilde{\eta}_k' X_s'\tilde{\boldsymbol{\Lambda}}\right)\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
=\;& O_P\left(\frac{1}{N^2T}\sum_{k=1}^{m^0+1}\|\tilde{\eta}_k\|^2\cdot\frac{1}{\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1}\sum_{s=T_{k-1}^0}^{T_k^0-1}\left\|X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}X_s\right\|\left\|X_s'\tilde{\boldsymbol{\Lambda}}\right\|\|f_t^0\|\right) \\
=\;& O_P\left(p\tilde{\eta}_{NT}\right) = O_P\left(p\delta_{p,NT}^{-1}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\|\right). \qquad \text{(C.97)}
\end{aligned}$$

For $l=2$, by the definition of $\tilde{u}_{NT,2}$, we have

$$\begin{aligned}
&\frac{1}{N\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}\tilde{u}_{NT,2}\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0 \\
=\;& -\frac{1}{NT\tau_j(T)}\sum_{k=1}^{m^0+1}\sum_{t=T_{j-1}^0}^{T_j^0-1}\sum_{s=T_{k-1}^0}^{T_k^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}X_s\tilde{\eta}_k f_s^{0\prime}\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0 \\
=\;& -\sum_{k=1}^{m^0+1}\left(\frac{1}{NT\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1}\sum_{s=T_{k-1}^0}^{T_k^0-1}\chi_{st}X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}X_s\right)\tilde{\eta}_k \\
=\;& -\left[\tilde{\Phi}_{j1}^*(\tilde{\boldsymbol{\Lambda}}),\ldots,\tilde{\Phi}_{j,m^0+1}^*(\tilde{\boldsymbol{\Lambda}})\right]\left(\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\right), \qquad \text{(C.98)}
\end{aligned}$$

where $\chi_{st}=f_s^{0\prime}\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0$ and $\tilde{\Phi}_{jk}^*(\tilde{\boldsymbol{\Lambda}})=\frac{1}{NT\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1}\sum_{s=T_{k-1}^0}^{T_k^0-1}\chi_{st}X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}X_s$. By Lemmas C.4(v) and C.5(i), we may show that

$$\left\|\tilde{\Phi}_{jk}^*(\tilde{\boldsymbol{\Lambda}})-\Phi_{jk}^*\right\| = O_P\left(p\delta_{p,NT}^{-1}(m^0)^{-1}\right),\quad 1\le j,k\le m^0+1, \qquad \text{(C.99)}$$

where $\Phi_{jk}^*=\frac{1}{NT\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1}\sum_{s=T_{k-1}^0}^{T_k^0-1}\chi_{st}X_t'\boldsymbol{M}_{\boldsymbol{\Lambda}^0}X_s$. Hence, by (C.98), (C.99) and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\left\|\frac{1}{N\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}\tilde{u}_{NT,2}\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}_{m^0}\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0+\left(\Phi_{j1}^*,\ldots,\Phi_{j,m^0+1}^*\right)\left(\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\right)\right\| \\
=\;& \left\|\left[\tilde{\Phi}_{j1}^*(\tilde{\boldsymbol{\Lambda}}),\ldots,\tilde{\Phi}_{j,m^0+1}^*(\tilde{\boldsymbol{\Lambda}})\right]\left(\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\right)-\left(\Phi_{j1}^*,\ldots,\Phi_{j,m^0+1}^*\right)\left(\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\right)\right\| \\
=\;& O_P\left(p\delta_{p,NT}^{-1}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\|\right). \qquad \text{(C.100)}
\end{aligned}$$

For $l=3$, by the definition of $\tilde{u}_{NT,3}$, Assumptions 1 and 3(ii), as well as (C.92), we have

$$\begin{aligned}
& \frac{1}{N\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}\tilde{u}_{NT,3}\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
= & \frac{1}{N\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}\left(\frac{1}{NT}\sum_{k=1}^{m^0+1}\sum_{s=T_{k-1}^0}^{T_k^0-1} X_s\tilde{\eta}_k\varepsilon_s'\tilde{\boldsymbol{\Lambda}}\right)\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
= & \frac{O_P(1)}{N^2T\tau_j(T)}\sum_{k=1}^{m^0+1}\|\tilde{\eta}_k\|\left[\sum_{t=T_{j-1}^0}^{T_j^0-1}\sum_{s=T_{k-1}^0}^{T_k^0-1}\|X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}X_s\|\left(\|\varepsilon_s'\boldsymbol{\Lambda}^0\|+\|\varepsilon_s'(\tilde{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}})\|\right)\|f_t^0\|\right] \\
= & O_P\left(p\delta_{NT}^{-1}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\|\right). \qquad \text{(C.101)}
\end{aligned}$$

To study the next two terms, we can apply the arguments used in the proof of Lemma C.3(ii) and show that $\frac{1}{N}||X_t'(\boldsymbol{\Lambda}^0-\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{H}}^+)||=O_P(p^{1/2}\delta_{NT}^{-2}+\tilde{\eta}_{NT}^{1/2})$. This, in conjunction with Lemma C.4(iii), implies that

$$\frac{1}{N}\|X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}(\boldsymbol{\Lambda}^0-\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{H}}^+)\|=O_P\left(p^{1/2}\delta_{NT}^{-2}+\tilde{\eta}_{NT}^{1/2}\right) \qquad \text{(C.102)}$$

and similarly for $j=1,\cdots,m^0+1$,

$$\frac{1}{N\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1}\|X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}(\boldsymbol{\Lambda}^0-\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{H}}^+)\|\,\|f_t^0\|=O_P\left(p^{1/2}\delta_{NT}^{-2}+\tilde{\eta}_{NT}^{1/2}\right). \qquad \text{(C.103)}$$

For $l=4$, by the definition of $\tilde{u}_{NT,4}$, (C.103), and Lemma C.5(i) and noting that $\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}\tilde{\boldsymbol{\Lambda}}=\boldsymbol{0}$,

$$\begin{aligned}
& \frac{1}{N\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}\tilde{u}_{NT,4}\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
= & \frac{1}{N\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}(\boldsymbol{\Lambda}^0-\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{H}}^+)\left(\frac{1}{NT}\sum_{k=1}^{m^0+1}\sum_{s=T_{k-1}^0}^{T_k^0-1} f_s^0\tilde{\eta}_k'X_s'\tilde{\boldsymbol{\Lambda}}\right)\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
= & O_P\left(\frac{1}{N^2T\tau_j(T)}\sum_{k=1}^{m^0+1}\|\tilde{\eta}_k\|\sum_{t=T_{j-1}^0}^{T_j^0-1}\sum_{s=T_{k-1}^0}^{T_k^0-1}\|X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}(\boldsymbol{\Lambda}^0-\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{H}}^+)\|\|X_s'\tilde{\boldsymbol{\Lambda}}\|\|f_s^0\|\|f_t^0\|\right) \\
= & O_P\left(\delta_{p,NT}^{-1}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\|\right). \qquad \text{(C.104)}
\end{aligned}$$

For $l=5$, by the definition of $\tilde{u}_{NT,5}$, Assumptions 1(i)(iii), (C.103), and Lemma C.5(iv), we have

$$\begin{aligned}
&\frac{1}{N\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \tilde{u}_{NT,5}\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
= &\frac{1}{N\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \left(\frac{1}{NT}\sum_{s=1}^{T}\boldsymbol{\Lambda}^0 f_s^0\varepsilon_s'\tilde{\boldsymbol{\Lambda}}\right)\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
\leq &\frac{1}{N^2T\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}(\boldsymbol{\Lambda}^0-\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{H}}^+)\left(\sum_{s=1}^{T} f_s^0\varepsilon_s'\boldsymbol{\Lambda}^0\right)\tilde{\boldsymbol{H}}\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
&+\frac{1}{N^2T\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}(\boldsymbol{\Lambda}^0-\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{H}}^+)\left[\sum_{s=1}^{T} f_s^0\varepsilon_s'(\tilde{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}})\right]\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+\left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
= &O_P\left(\frac{1}{N^2T\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1}\|X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}(\boldsymbol{\Lambda}^0-\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{H}}^+)\|\|\sum_{s=1}^{T} f_s^0\varepsilon_s'\boldsymbol{\Lambda}^0\|\|f_t^0\|\right) \\
&+O_P\left(\frac{1}{N^2T\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1}\sum_{s=1}^{T}\|X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}(\boldsymbol{\Lambda}^0-\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{H}}^+)\|\|\varepsilon_s'(\tilde{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}^0\tilde{\boldsymbol{H}})f_s^0\|\|f_t^0\|\right) \\
= &O_P\left(\delta_{p,NT}^{-3}+\delta_{p,NT}^{-2}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\|\right). \qquad \text{(C.105)}
\end{aligned}$$

For $l=6$, by the definition of $\tilde{u}_{NT,6}$ and Assumptions 1(i)-(iii), 2(ii) and 3(ii), we have

$$\begin{aligned}
& \frac{1}{N\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \tilde{u}_{NT,6} \left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+ \left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
= & \frac{1}{N\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \left(\frac{1}{NT}\sum_{k=1}^{m^0+1}\sum_{s=T_{k-1}^0}^{T_k^0-1} \varepsilon_s \tilde{\eta}_k' X_s' \tilde{\boldsymbol{\Lambda}}\right) \left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+ \left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
\leq & \frac{1}{N\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} \left(\frac{1}{NT}\sum_{k=1}^{m^0+1}\sum_{s=T_{k-1}^0}^{T_k^0-1} X_t'\varepsilon_s \tilde{\eta}_k' X_s' \tilde{\boldsymbol{\Lambda}}\right) \left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+ \left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
& + \frac{1}{N\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \left(\frac{1}{NT}\sum_{k=1}^{m^0+1}\sum_{s=T_{k-1}^0}^{T_k^0-1} (\boldsymbol{P}_{\tilde{\boldsymbol{\Lambda}}}-\boldsymbol{P}_{\boldsymbol{\Lambda}^0})\varepsilon_s \tilde{\eta}_k' X_s' \tilde{\boldsymbol{\Lambda}}\right) \left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+ \left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
& + \frac{1}{N\tau_j(T)}\left\|\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{\Lambda}^0(\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0)^+ \left(\frac{1}{NT}\sum_{k=1}^{m^0+1}\sum_{s=T_{k-1}^0}^{T_k^0-1} \boldsymbol{\Lambda}^{0\prime}\varepsilon_s \tilde{\eta}_k' X_s' \tilde{\boldsymbol{\Lambda}}\right) \left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+ \left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0\right\| \\
= & O_P\left(p^{1/2}\delta_{p,NT}^{-1}(m^0)^{-1/2}\|\tilde{\boldsymbol{\alpha}}_{m^0}-\boldsymbol{\alpha}^0\|\right). \qquad \text{(C.106)}
\end{aligned}$$

For $l=7$, by the definitions of $\tilde{u}_{NT,7}$ and χ_{st}, we have

$$\begin{aligned}
& \frac{1}{N\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \tilde{u}_{NT,7} \left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+ \left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0 \\
= & \frac{1}{N\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \left(\frac{1}{NT}\sum_{s=1}^{T}\varepsilon_s f_s^{0\prime}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right) \left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+ \left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0 \\
= & \frac{1}{NT\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1}\sum_{s=1}^{T}\chi_{st}X_t'\boldsymbol{M}_{\boldsymbol{\Lambda}^0}\varepsilon_s + \frac{1}{NT\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1}\sum_{s=1}^{T}\chi_{st}X_t'(\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}-\boldsymbol{M}_{\boldsymbol{\Lambda}^0})\varepsilon_s \\
= & \frac{1}{N\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1} X_t'\boldsymbol{M}_{\boldsymbol{\Lambda}^0}\varepsilon_t^* + \frac{1}{NT\tau_j(T)}\sum_{t=T_{j-1}^0}^{T_j^0-1}\sum_{s=1}^{T}\chi_{st}X_t'(\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}-\boldsymbol{M}_{\boldsymbol{\Lambda}^0})\varepsilon_s, \qquad \text{(C.107)}
\end{aligned}$$

where $\varepsilon_t^* = \frac{1}{T}\sum_{s=1}^T \chi_{st}\varepsilon_s$. On the other hand, following the proof of Lemma B.3(i) and (C.95) in particular, we may show that

$$\begin{aligned} &\left\| \frac{1}{NT\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} \sum_{s=1}^{T} \chi_{st} X_t' (\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} - \boldsymbol{M}_{\boldsymbol{\Lambda}^0}) \varepsilon_s + B_{NT,j}(2,2) \right\| \\ = \;& O_P\left(\delta_{NT}^{-1} (m^0)^{-1/2} \|\tilde{\boldsymbol{\alpha}}_{m_0} - \boldsymbol{\alpha}^0\| + \delta_{p,NT}^{-3} \right). \end{aligned} \tag{C.108}$$

Hence, it follows that

$$\begin{aligned} &\left\| \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \tilde{u}_{NT,7} (\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}})^+ (\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0)^+ f_t^0 - \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\boldsymbol{\Lambda}^0} \varepsilon_t^* + B_{NT,j}(2,2) \right\| \\ = \;& O_P\left(\delta_{NT}^{-1} (m^0)^{-1/2} \|\tilde{\boldsymbol{\alpha}}_{m_0} - \boldsymbol{\alpha}^0\| + \delta_{p,NT}^{-3} \right). \end{aligned} \tag{C.109}$$

For $l = 8$, by the definition of $\tilde{u}_{NT,8}$, we have

$$\begin{aligned} &\frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \tilde{u}_{NT,8} \left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+ \left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0 \\ = \;& \frac{1}{N\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \left(\frac{1}{NT}\sum_{s=1}^T \varepsilon_s\varepsilon_s'\tilde{\boldsymbol{\Lambda}}\right) \left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+ \left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0 \\ = \;& \frac{1}{N^2T\tau_j(T)} \sum_{t=T_{j-1}^0}^{T_j^0-1} X_t' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\tilde{\boldsymbol{\Lambda}} \left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\tilde{\boldsymbol{\Lambda}}\right)^+ \left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)^+ f_t^0 \equiv B_{NT,j}(1). \end{aligned} \tag{C.110}$$

By (C.96), (C.97), (C.100), (C.101), (C.104)–(C.106), (C.109) and (C.110), we can complete the proof of Lemma B.3(ii).

We have thus completed the proof of Lemma B.3. ■

Let

$$\dot{\boldsymbol{\Lambda}}_R = \left(\dot{\lambda}_{1,R}, ..., \dot{\lambda}_{N,R}\right)' \quad \text{and} \quad \breve{\boldsymbol{\Lambda}}_R = \frac{1}{NT}\sum_{t=1}^T (Y_t - X_t\dot{\beta}_{t,R})(Y_t - X_t\dot{\beta}_{t,R})'\dot{\boldsymbol{\Lambda}}_R = \left(\breve{\lambda}_{1,R}, ..., \breve{\lambda}_{N,R}\right)'.$$

In order to prove Lemma B.4 in Appendix B, we first need to prove the following technical lemma.

Lemma C.6 *Suppose that Assumptions 1 and 2 in Appendix A hold and $R > R_0$. Define the $R_0 \times R$ matrix $\dot{\boldsymbol{H}}_R \equiv \left(\frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\right)\left(\frac{1}{N}\boldsymbol{\Lambda}^{0\prime}\dot{\boldsymbol{\Lambda}}_R\right)$ with the Moore-Penrose generalized inverse $\dot{\boldsymbol{H}}_R^+ =$*

$\begin{bmatrix} \dot{\boldsymbol{H}}_R^+(1) \\ \dot{\boldsymbol{H}}_R^+(2) \end{bmatrix}$, *where* $\dot{\boldsymbol{H}}_R^+(1)$ *and* $\dot{\boldsymbol{H}}_R^+(2)$ *are* $R_0 \times R_0$ *and* $(R - R_0) \times R_0$ *matrices, respectively. Let* $\dot{\boldsymbol{V}}_{NT,R}$ *denote an* $R \times R$ *diagonal matrix consisting of the* R *largest eigenvalues of the* $N \times N$ *matrix* $\frac{1}{NT}\sum_{t=1}^T (Y_t - X_t\dot{\beta}_{t,R})(Y_t - X_t\dot{\beta}_{t,R})'$ *where the eigenvalues are in decreasing order along the main diagonal line. Write* $\dot{\boldsymbol{\Lambda}}_R = \left[\dot{\boldsymbol{\Lambda}}_R(1), \dot{\boldsymbol{\Lambda}}_R(2)\right]$ *and* $\dot{\boldsymbol{H}}_R = \left[\dot{\boldsymbol{H}}_R(1), \dot{\boldsymbol{H}}_R(2)\right]$, *where* $\dot{\boldsymbol{\Lambda}}_R(1)$, $\dot{\boldsymbol{\Lambda}}_R(2)$, $\dot{\boldsymbol{H}}_R(1)$, *and* $\dot{\boldsymbol{H}}_R(2)$ *are* $N \times R_0$, $N \times (R - R_0)$, $R_0 \times R_0$, *and* $R_0 \times (R - R_0)$ *matrices, respectively. Furthermore, write* $\dot{\boldsymbol{V}}_{NT,R} = \mathsf{diag}\left\{\dot{\boldsymbol{V}}_{NT,R}(1), \dot{\boldsymbol{V}}_{NT,R}(2)\right\}$, *where* $\dot{\boldsymbol{V}}_{NT,R}(1)$ *denotes the upper-left* $R_0 \times R_0$ *submatrix of* $\dot{\boldsymbol{V}}_{NT,R}$. *Then we have*

(i) $\frac{1}{N}\left\|\breve{\boldsymbol{\Lambda}}_R - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R\right\|^2 = O_P\left(\delta_{p,NT}^{-2}\right)$,

(ii) $\frac{1}{N}\left\|\breve{\boldsymbol{\Lambda}}_R'\breve{\boldsymbol{\Lambda}}_R - \dot{\boldsymbol{H}}_R'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R\right\| = O_P\left(\delta_{p,NT}^{-1}\right)$,

(iii) $\frac{1}{N}\left\|\dot{\boldsymbol{\Lambda}}_R(1) - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R(1)\dot{\boldsymbol{V}}_{NT,R}^+(1)\right\|^2 = O_P\left(\delta_{p,NT}^{-2}\right)\Big)$ *and* $\left\|\dot{\boldsymbol{H}}_R(2)\right\|^2 = O_P\left(\delta_{p,NT}^{-2}\right)$,

(iv) $\left\|\dot{\boldsymbol{H}}_R^+(1)\right\| = O_P(1)$ *and* $\left\|\dot{\boldsymbol{H}}_R^+(2)\right\| = O_P\left(\delta_{p,NT}^{-1}\right)$.

Proof of Lemma C.6. **(i)** When $R > R_0$, we can follow the proof of Lemma C.2 and show that

$$\dot{\eta}_R \equiv \frac{1}{T}\sum_{t=1}^T \|\dot{\beta}_{t,R} - \beta_t^0\|^2 = o_P(1).$$

Next, using $Y_t - X_t\dot{\beta}_{t,R} = \boldsymbol{\Lambda}^0 f_t^0 + \varepsilon_t + X_t(\beta_t^0 - \dot{\beta}_{t,R})$ and $\dot{d}_{t,R} = \dot{\beta}_{t,R} - \beta_t^0$, we have

$$\begin{aligned}
\breve{\boldsymbol{\Lambda}}_R - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R &= \frac{1}{NT}\sum_{t=1}^T (Y_t - X_t\dot{\beta}_{t,R})(Y_t - X_t\dot{\beta}_{t,R})'\dot{\boldsymbol{\Lambda}}_R - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R \\
&= \frac{1}{NT}\sum_{t=1}^T \left[-X_t\dot{d}_{t,R} + \boldsymbol{\Lambda}^0 f_t^0 + \varepsilon_t\right]\left[-X_t\dot{d}_{t,R} + \boldsymbol{\Lambda}^0 f_t^0 + \varepsilon_t\right]'\dot{\boldsymbol{\Lambda}}_R - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R \\
&= \frac{1}{NT}\sum_{t=1}^T X_t\dot{d}_{t,R}\dot{d}_{t,R}'X_t'\dot{\boldsymbol{\Lambda}}_R - \frac{1}{NT}\sum_{t=1}^T X_t d_{t,R} f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\dot{\boldsymbol{\Lambda}}_R - \frac{1}{NT}\sum_{t=1}^T X_t\dot{d}_{t,R}\varepsilon_t'\dot{\boldsymbol{\Lambda}}_R \\
&\quad -\frac{1}{NT}\sum_{t=1}^T \boldsymbol{\Lambda}^0 f_t^0\dot{d}_{t,R}'X_t'\dot{\boldsymbol{\Lambda}}_R + \frac{1}{NT}\sum_{t=1}^T \boldsymbol{\Lambda}^0 f_t^0\varepsilon_t'\dot{\boldsymbol{\Lambda}}_R - \frac{1}{NT}\sum_{t=1}^T \varepsilon_t\dot{d}_{t,R}'X_t'\dot{\boldsymbol{\Lambda}}_R \\
&\quad +\frac{1}{NT}\sum_{t=1}^T \varepsilon_t f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\dot{\boldsymbol{\Lambda}}_R + \frac{1}{NT}\sum_{t=1}^T \varepsilon_t\varepsilon_t'\dot{\boldsymbol{\Lambda}}_R \\
&\equiv \sum_{j=1}^8 \dot{u}_{R,j}. \qquad \text{(C.111)}
\end{aligned}$$

Following the proof of Lemma C.3(i), we can readily show that $\frac{1}{N}||\dot{u}_{R,j}||^2 = O_P\left(\delta_{NT}^{-2} + \dot{\eta}_R\right)$. Then we readily have $\frac{1}{N}||\breve{\boldsymbol{\Lambda}}_R - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R||^2 = O_P\left(\delta_{NT}^{-2} + \dot{\eta}_R\right)$. With this, we can apply the arguments used in the proof of Theorem 3.1 to show that $\dot{\eta}_R = O_P\left(\delta_{p,NT}^{-2}\right)$. Then we may complete the proof of (i).

(ii) Noting that

$$\begin{aligned} & \frac{1}{N}\breve{\mathbf{\Lambda}}_R'\breve{\mathbf{\Lambda}}_R - \frac{1}{N}\dot{\boldsymbol{H}}_R'\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R \\ = & \frac{1}{N}(\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R)'(\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R) + \frac{1}{N}(\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R)'\mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R + \frac{1}{N}\dot{\boldsymbol{H}}_R'\mathbf{\Lambda}^{0\prime}(\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R), \end{aligned}$$

the convergence result (ii) follows from the triangle and Cauchy-Schwarz inequalities, Lemma C.6(i), and the fact that $||\mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R||^2 = O_P(N)$.

(iii) Let $\dot{\boldsymbol{V}}_R$ and $\dot{\boldsymbol{V}}_R(1)$ denote the probability limits of $\dot{\boldsymbol{V}}_{NT,R}$ and $\dot{\boldsymbol{V}}_{NT,R}(1)$, respectively, as $(N,T) \to \infty$. Recall that $\dot{\boldsymbol{H}}_R = \frac{1}{NT}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}_R$ and $\frac{1}{N}\dot{\mathbf{\Lambda}}_R'\dot{\mathbf{\Lambda}}_R = \boldsymbol{I}_R$. As the application of PCA method, we have the identity

$$\frac{1}{NT}\sum_{t=1}^{T}(Y_t - X_t\dot{\beta}_{t,R})(Y_t - X_t\dot{\beta}_{t,R})'\dot{\mathbf{\Lambda}}_R = \dot{\mathbf{\Lambda}}_R\dot{\boldsymbol{V}}_{NT,R}.$$

Pre-multiplying both sides of the above equation by $\dot{\mathbf{\Lambda}}_R'/N$ and using the normalization $\frac{1}{N}\dot{\mathbf{\Lambda}}_R'\dot{\mathbf{\Lambda}}_R = \boldsymbol{I}_R$ yields

$$\frac{1}{N^2T}\dot{\mathbf{\Lambda}}_R'\left[\sum_{t=1}^{T}(Y_t - X_t\dot{\beta}_{t,R})(Y_t - X_t\dot{\beta}_{t,R})'\right]\dot{\mathbf{\Lambda}}_R = \dot{\boldsymbol{V}}_{NT,R},$$

which together with $Y_t - X_t\dot{\beta}_{t,R} = X_t(\beta_t^0 - \dot{\beta}_{t,R}) + \mathbf{\Lambda}^0 f_t^0 + \varepsilon_t$, yields

$$\frac{1}{N^2T}\dot{\mathbf{\Lambda}}_R'\mathbf{\Lambda}^0\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}_R + d_{NT,R} = \dot{\boldsymbol{V}}_{NT,R},$$

where

$$\begin{aligned} d_{NT,R} = & \frac{1}{N^2T}\dot{\mathbf{\Lambda}}_R'\sum_{t=1}^{T}\Big[X_t(\beta_t^0 - \dot{\beta}_{t,R})(\beta_t^0 - \dot{\beta}_{t,R})'X_t' + \varepsilon_t\varepsilon_t' + X_t(\beta_t^0 - \dot{\beta}_{t,R})f_t^{0\prime}\mathbf{\Lambda}^{0\prime} \\ & +\mathbf{\Lambda}^0 f_t^0(\beta_t^0 - \dot{\beta}_{t,R})'X_t' + X_t(\beta_t^0 - \dot{\beta}_{t,R})\varepsilon_t' + \varepsilon_t(\beta_t^0 - \dot{\beta}_{t,R})'X_t' \\ & +\mathbf{\Lambda}^0 f_t^0\varepsilon_t' + \varepsilon_t f_t^{0\prime}\mathbf{\Lambda}^{0\prime}\Big]\dot{\mathbf{\Lambda}}_R \\ \equiv & \sum_{j=1}^{8} d_{R,j}. \end{aligned}$$

Following the proof of Lemma C.3, it is easy to show that $\|d_{NT,R}\| = O_P\left(\delta_{p,NT}^{-1}\right)$ by proving that $d_{R,j}$, $j = 1, 2, ..., 8$, are either $O_P(\delta_{p,NT}^{-1})$ or of smaller order. For example,

$$\begin{aligned}
\|d_{R,1}\| &= \frac{1}{N^2T}\left\|\dot{\mathbf{\Lambda}}_R'\left[\sum_{t=1}^T X_t(\beta_t^0 - \dot{\beta}_{t,R})(\beta_t^0 - \dot{\beta}_{t,R})'X_t'\right]\dot{\mathbf{\Lambda}}_R\right\| \\
&\leq \frac{1}{N}\left\|\dot{\mathbf{\Lambda}}_R\right\|^2 \mu_{\max}\left(X_t'X_t/N\right)\frac{1}{T}\sum_{t=1}^T\left\|\beta_t^0 - \dot{\beta}_{t,R}\right\|^2 = O_P\left(\delta_{p,NT}^{-2}\right), \\
\|d_{R,2}\| &= \frac{1}{N^2T}\left\|\dot{\mathbf{\Lambda}}_R'\left[\sum_{t=1}^T \varepsilon_t\varepsilon_t'\right]\dot{\mathbf{\Lambda}}_R\right\| \leq \frac{1}{NT}\|\boldsymbol{\varepsilon}\|_{\text{sp}}^2\frac{1}{N}\left\|\dot{\mathbf{\Lambda}}_R\right\|^2 = O_P\left(\delta_{NT}^{-2}\right),
\end{aligned}$$

and

$$\begin{aligned}
\|d_{R,3}\| &= \frac{1}{N^2T}\left\|\dot{\mathbf{\Lambda}}_R'\sum_{t=1}^T X_t(\beta_t^0 - \dot{\beta}_{t,R})f_t^{0\prime}\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}_R\right\| \\
&\leq \frac{1}{N}\left\|\dot{\mathbf{\Lambda}}_R\right\|^2\frac{1}{N^{1/2}}\left\|\mathbf{\Lambda}^0\right\|\mu_{\max}^{1/2}\left(X_t'X_t/N\right)\left(\frac{1}{T}\sum_{t=1}^T\left\|\beta_t^0 - \dot{\beta}_{t,R}\right\|^2\right)^{1/2}\frac{1}{T^{1/2}}\left\|\mathbf{F}^0\right\| \\
&\leq O_P\left(\dot{\eta}_R^{1/2}\right) = O_P\left(\delta_{p,NT}^{-1}\right).
\end{aligned}$$

Then

$$\frac{1}{N^2T}\dot{\mathbf{\Lambda}}_R'\mathbf{\Lambda}^0\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}_R = \dot{\boldsymbol{V}}_{NT,R} - d_{NT,R} \xrightarrow{P} \dot{\boldsymbol{V}}_R. \tag{C.112}$$

Observe that $\frac{1}{N^2T}\dot{\mathbf{\Lambda}}_R'\mathbf{\Lambda}^0\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}_R$ has rank R_0 at most in both finite and large samples. Let $\mathbf{\Delta}_{NT}(l) = \frac{1}{N}\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}_R(l)$ for $l = 1, 2$ and $\hat{\mathbf{\Sigma}}_F = \frac{1}{T}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0$. Then

$$\frac{1}{N^2T}\dot{\mathbf{\Lambda}}_R'\mathbf{\Lambda}^0\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}_R = \begin{bmatrix} \mathbf{\Delta}_{NT}'(1)\hat{\mathbf{\Sigma}}_F\mathbf{\Delta}_{NT}(1) & \mathbf{\Delta}_{NT}'(1)\hat{\mathbf{\Sigma}}_F\mathbf{\Delta}_{NT}(2) \\ \mathbf{\Delta}_{NT}'(2)\hat{\mathbf{\Sigma}}_F\mathbf{\Delta}_{NT}(1) & \mathbf{\Delta}_{NT}'(2)\hat{\mathbf{\Sigma}}_F\mathbf{\Delta}_{NT}(2) \end{bmatrix}.$$

Note that $\hat{\mathbf{\Sigma}}_F = \mathbf{\Sigma}_F + o_P(1)$ by Assumption 1(i). Following the proof of Lemma A.3(ii) in Bai (2003), we can show that $\text{plim}_{(N,T)\to\infty}\mathbf{\Delta}_{NT}'(1)\hat{\mathbf{\Sigma}}_F\mathbf{\Delta}_{NT}(1) = \dot{\boldsymbol{V}}_R(1)$ which has full rank R_0. This ensures that $\frac{1}{N^2T}\dot{\mathbf{\Lambda}}_R'\mathbf{\Lambda}^0\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}_R$ has rank R_0 in large samples and $\mathbf{\Delta}_{NT}'(2)\hat{\mathbf{\Sigma}}_F\mathbf{\Delta}_{NT}(2) \xrightarrow{P} \mathbf{0}$. Then $\mathbf{\Delta}_{NT}'(1)\hat{\mathbf{\Sigma}}_F\mathbf{\Delta}_{NT}(2) \xrightarrow{P} \mathbf{0}$ by the Cauchy-Schwarz inequality. By the asymptotic nonsingularity of $\hat{\mathbf{\Sigma}}_F$, this also implies that $\mathbf{\Delta}_{NT}(2) = o_P(1)$ and $\mathbf{\Delta}_{NT}(1) \xrightarrow{P} \mathbf{\Delta}(1)$ for some $R_0 \times R_0$ nonsingular matrix $\mathbf{\Delta}(1)$. Consequently, we have

$$\dot{\boldsymbol{H}}_R(1) = \frac{1}{NT}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}_R(1) \xrightarrow{P} \mathbf{\Sigma}_F\mathbf{\Delta}(1)$$

and

$$\dot{\boldsymbol{H}}_R(2) = \frac{1}{NT}\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\mathbf{\Lambda}^{0\prime}\dot{\mathbf{\Lambda}}_R(2) = o_P(1).$$

Then $\dot{\boldsymbol{H}}_R(1)$ is asymptotically nonsingular and $\dot{\boldsymbol{H}}_R$ has rank R_0.

By the definition $\breve{\boldsymbol{\Lambda}}_R = \frac{1}{NT}\sum_{t=1}^{T}(Y_t - X_t\dot{\beta}_{t,R})(Y_t - X_t\dot{\beta}_{t,R})'\dot{\boldsymbol{\Lambda}}_R$ and the identity $\frac{1}{NT}\sum_{t=1}^{T}(Y_t - X_t\dot{\beta}_{t,R})(Y_t - X_t\dot{\beta}_{t,R})'\dot{\boldsymbol{\Lambda}}_R = \dot{\boldsymbol{\Lambda}}_R\dot{\boldsymbol{V}}_{NT,R}$ from the PCA, we have

$$\begin{aligned}
\frac{1}{N}\left\|\breve{\boldsymbol{\Lambda}}_R - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R\right\|^2 &= \frac{1}{N}\left\|\dot{\boldsymbol{\Lambda}}_R\dot{\boldsymbol{V}}_{NT,R} - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R\right\|^2 \\
&= \frac{1}{N}\left\|\dot{\boldsymbol{\Lambda}}_R\dot{\boldsymbol{V}}_{NT,R}(1) - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R(1)\right\|^2 + \frac{1}{N}\left\|\dot{\boldsymbol{\Lambda}}_R\dot{\boldsymbol{V}}_{NT,R}(2) - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R(2)\right\|^2.
\end{aligned}$$

Lemma C.6(i) implies that $\frac{1}{N}\left\|\dot{\boldsymbol{\Lambda}}_R\dot{\boldsymbol{V}}_{NT,R}(l) - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R(l)\right\|^2 = O_P\left(\delta_{p,NT}^{-2}\right)$ for $l = 1, 2$. Since $\dot{\boldsymbol{V}}_R(1)$ is nonsingular, it follows that

$$\frac{1}{N}\left\|\dot{\boldsymbol{\Lambda}}_R - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R(1)\dot{\boldsymbol{V}}_{NT,R}^{+}(1)\right\|^2 = O_P\left(\delta_{p,NT}^{-2}\right)$$

and

$$\left\|\dot{\boldsymbol{V}}_{NT,R}^{+}(1)\right\| \leq \left\|\dot{\boldsymbol{V}}_{R}^{+}(1)\right\| + \left\|\dot{\boldsymbol{V}}_{NT,R}^{+}(1) - \dot{\boldsymbol{V}}_{R}^{+}(1)\right\| = O_P(1).$$

In addition,

$$\begin{aligned}
\frac{1}{N}\left\|\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R(2)\right\|^2 &\leq \frac{2}{N}\left\|\dot{\boldsymbol{\Lambda}}_R\dot{\boldsymbol{V}}_{NT,R}(2) - \boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R(2)\right\|^2 + \frac{2}{N}\left\|\dot{\boldsymbol{\Lambda}}_R\dot{\boldsymbol{V}}_{NT,R}(2)\right\|^2 \\
&= O_P\left(\delta_{p,NT}^{-2}\right) + O_P\left(\delta_{p,NT}^{-2}\right) = O_P\left(\delta_{p,NT}^{-2}\right),
\end{aligned}$$

because

$$\frac{1}{N}\left\|\dot{\boldsymbol{\Lambda}}_R\dot{\boldsymbol{V}}_{NT,R}(2)\right\|^2 \leq \left[\mu_{\max}^2\left(\dot{\boldsymbol{V}}_{NT,R}(2)\right)\right]\left\|\dot{\boldsymbol{\Lambda}}_R\right\|^2/N = R\left[\mu_{\max}^2\left(\dot{\boldsymbol{V}}_{NT,R}(2)\right)\right]$$

and

$$\mu_{\max}\left(\dot{\boldsymbol{V}}_{NT,R}(2)\right) \leq \mu_{R_0+1}\left(\dot{\boldsymbol{\Lambda}}_R'\boldsymbol{\Lambda}^0\boldsymbol{F}^{0\prime}\boldsymbol{F}^0\boldsymbol{\Lambda}^{0\prime}\dot{\boldsymbol{\Lambda}}_R/(N^2T)\right) + \|d_{NT,R}\| = \|d_{NT,R}\| = O_P\left(\delta_{p,NT}^{-1}\right),$$

where $\mu_{R_0+1}(\cdot)$ denotes the (R_0+1)-th largest eigenvalue of the square matrix in the parentheses. In view of the fact that

$$\frac{1}{N}\left\|\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R(2)\right\|^2 = \frac{1}{N}\mathsf{Tr}\left(\dot{\boldsymbol{H}}_R(2)\dot{\boldsymbol{H}}_R(2)'\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0\right) \geq \mu_{\min}\left(\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0/N\right)\left\|\dot{\boldsymbol{H}}_R(2)\right\|^2,$$

we have

$$\left\|\dot{\boldsymbol{H}}_R(2)\right\|^2 \leq \left[\mu_{\min}\left(\boldsymbol{\Lambda}^{0\prime}\boldsymbol{\Lambda}^0/N\right)\right]^{-1}\frac{1}{N}\left\|\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R(2)\right\|^2 = O_P\left(\delta_{p,NT}^{-2}\right).$$

(iv) Since $\dot{\boldsymbol{H}}_R$ is right invertible asymptotically, by Proposition 6.1.5 in Bernstein (2005, p.225), the $R \times R_0$ generalized inverse $\dot{\boldsymbol{H}}_R^+$ of $\dot{\boldsymbol{H}}_R$ is given by

$$\dot{\boldsymbol{H}}_R^+ = \dot{\boldsymbol{H}}_R' \left[\dot{\boldsymbol{H}}_R \dot{\boldsymbol{H}}_R'\right]^{-1} = \left[\begin{array}{c} \dot{\boldsymbol{H}}_R'(1) \left(\dot{\boldsymbol{H}}_R \dot{\boldsymbol{H}}_R'\right)^{-1} \\ \dot{\boldsymbol{H}}_R'(2) \left(\dot{\boldsymbol{H}}_R \dot{\boldsymbol{H}}_R'\right)^{-1} \end{array} \right] = \left[\begin{array}{c} \dot{\boldsymbol{H}}_R^+(1) \\ \dot{\boldsymbol{H}}_R^+(2) \end{array} \right].$$

Then by Lemma C.6(iii)

$$\begin{aligned} \left\|\dot{\boldsymbol{H}}_R^+(1)\right\| &\leq \left\|\dot{\boldsymbol{H}}_R(1)\right\| \left\|\left(\dot{\boldsymbol{H}}_R \dot{\boldsymbol{H}}_R'\right)^{-1}\right\| = O_P(1), \text{ and} \\ \left\|\dot{\boldsymbol{H}}_R^+(2)\right\| &\leq \left\|\dot{\boldsymbol{H}}_R(2)\right\| \left\|\left(\dot{\boldsymbol{H}}_R \dot{\boldsymbol{H}}_R'\right)^{-1}\right\| = O_P\left(\delta_{p,NT}^{-1}\right). \end{aligned}$$

We have thus completed the proof of Lemma C.6. ■

Proof of Lemma B.4. **(i)** The proof is similar to that of Lemma C.2. Notice that

$$\hat{Q}_{NT}(\boldsymbol{\beta}, \boldsymbol{\Lambda}_R) = \frac{1}{NT}\sum_{t=1}^{T}(Y_t - X_t\beta_t)' \boldsymbol{M}_{\boldsymbol{\Lambda}_R}(Y_t - X_t\beta_t).$$

Using $Y_t - X_t\dot{\beta}_{t,R} = X_t(\beta_t^0 - \dot{\beta}_{t,R}) + \boldsymbol{\Lambda}^0 f_t^0 + \varepsilon_t$, we have

$$\begin{aligned} 0 &\geq \hat{Q}_{NT}(\dot{\boldsymbol{\beta}}_R, \dot{\boldsymbol{\Lambda}}_R) - \hat{Q}_{NT}(\boldsymbol{\beta}^0, \dot{\boldsymbol{\Lambda}}_R) \\ &= \frac{1}{NT}\sum_{t=1}^{T}\left[(Y_t - X_t\dot{\beta}_{t,R})' \boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R}(Y_t - X_t\dot{\beta}_{t,R}) - (Y_t - X_t\beta_t^0)' \boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R}(Y_t - X_t\beta_t^0)\right] \\ &= \frac{1}{NT}\sum_{t=1}^{T}\left[(\dot{\beta}_{t,R} - \beta_t^0)' X_t' \boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R} X_t(\dot{\beta}_{t,R} - \beta_t^0) - 2(\dot{\beta}_{t,R} - \beta_t^0)' X_t' \boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R} \boldsymbol{\Lambda}^0 f_t^0\right] \\ &\quad -\frac{2}{NT}\sum_{t=1}^{T}(\dot{\beta}_{t,R} - \beta_t^0)' X_t' \boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R}\varepsilon_t. \end{aligned}$$

By Lemma C.1(i) (with R_0 and $\boldsymbol{\Lambda}$ being replaced by R and $\boldsymbol{\Lambda}_R$), we can prove that

$$\frac{1}{NT}\sum_{t=1}^{T}(\dot{\beta}_{t,R} - \beta_t^0)' X_t' \boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R}\varepsilon_t = O_P\left(p^{1/2}\delta_{p,NT}^{-1}\right).$$

Let $\dot{\boldsymbol{d}}_{\beta,R} = \dot{\boldsymbol{\beta}}_R - \boldsymbol{\beta}^0$ and $\dot{\boldsymbol{d}}_{\Lambda,R} = \frac{1}{N^{1/2}}\mathsf{vec}(\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R}\boldsymbol{\Lambda}^0)$. Define

$$\dot{\boldsymbol{A}}_R = \frac{1}{N}\mathsf{diag}(X_1' \boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R} X_1, ..., X_T' \boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R} X_T) \text{ and } \dot{\boldsymbol{C}}_R = \frac{1}{N^{1/2}}[f_1^0 \otimes \boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R} X_1, ..., f_T^0 \otimes \boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R} X_T].$$

Then

$$\frac{1}{NT}\sum_{t=1}^{T}\left[(\dot{\beta}_{t,R}-\beta_t^0)'X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R}X_t(\dot{\beta}_{t,R}-\beta_t^0)-2(\dot{\beta}_{t,R}-\beta_t^0)'X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R}\boldsymbol{\Lambda}^0 f_t^0\right]$$
$$= \frac{1}{T}\dot{\boldsymbol{d}}_{\beta,R}'\dot{\boldsymbol{A}}_R\dot{\boldsymbol{d}}_{\beta,R}-\frac{2}{T}\dot{\boldsymbol{d}}_{\Lambda,R}'\dot{\boldsymbol{C}}_R\dot{\boldsymbol{d}}_{\beta,R}.$$

It follows that

$$\frac{1}{T}\dot{\boldsymbol{d}}_{\beta,R}'\dot{\boldsymbol{A}}_R\dot{\boldsymbol{d}}_{\beta,R}-\frac{2}{T}\dot{\boldsymbol{d}}_{\Lambda,R}'\dot{\boldsymbol{C}}_R\dot{\boldsymbol{d}}_{\beta,R}+O_P\left(p^{1/2}\delta_{p,NT}^{-1}\right)\leq 0.$$

This, in junction with the fact that

$$\begin{aligned}\left|\dot{\boldsymbol{d}}_{\Lambda,R}'\dot{\boldsymbol{C}}_R\dot{\boldsymbol{d}}_{\beta,R}\right| &\leq \left[\dot{\boldsymbol{d}}_{\Lambda,R}'\dot{\boldsymbol{d}}_{\Lambda,R}\right]^{1/2}\left[\dot{\boldsymbol{d}}_{\beta,R}'\dot{\boldsymbol{C}}_R'\dot{\boldsymbol{C}}_R\dot{\boldsymbol{d}}_{\beta,R}\right]^{1/2}\\ &\leq \left\|\dot{\boldsymbol{d}}_{\Lambda,R}\right\|\left\|\dot{\boldsymbol{d}}_{\beta,R}\right\|\left[\mu_{\max}^{1/2}(\dot{\boldsymbol{C}}_R'\dot{\boldsymbol{C}}_R)\right],\end{aligned}$$

implies that

$$\frac{1}{T}\dot{\boldsymbol{d}}_{\beta,R}'\dot{\boldsymbol{A}}_R\dot{\boldsymbol{d}}_{\beta,R}-\frac{2}{T^{1/2}}\left\|\dot{\boldsymbol{d}}_{\Lambda,R}\right\|\left\|\dot{\boldsymbol{d}}_{\beta,R}\right\|\left[\mu_{\max}^{1/2}\left(\dot{\boldsymbol{C}}_R'\dot{\boldsymbol{C}}_R/T\right)\right]+O_P(p^{1/2}\delta_{p,NT}^{-1})\leq 0.$$

Using a decomposition similar to (B.8) in Appendix B, we can readily show that $\mu_{\max}(\dot{\boldsymbol{C}}_R'\dot{\boldsymbol{C}}_R/T)$ $= o_P(1)$. By Assumption 1(ii), $\mu_{\min}(\dot{\boldsymbol{A}}_R)>c_x$ w.p.a.1. and $||\dot{\boldsymbol{d}}_{\Lambda,R}||=O_P(1)$. It follows that

$$\frac{1}{T}\|\dot{\boldsymbol{d}}_{\beta,R}\|^2=\frac{1}{T}\sum_{t=1}^{T}\|\dot{\beta}_{t,R}-\beta_t^0\|^2=o_P(1).$$

Note that

$$V(R,\dot{\boldsymbol{\beta}}_R)=\min_{\boldsymbol{\beta},\boldsymbol{\Lambda}_R}\hat{Q}_{NT}(\boldsymbol{\beta},\boldsymbol{\Lambda}_R)$$

subject to $\boldsymbol{\Lambda}_R'\boldsymbol{\Lambda}_R/N=\boldsymbol{I}_R$. Let $s_r(\boldsymbol{\beta})=\mu_r\left[\sum_{t=1}^{T}(Y_t-X_t\beta_t)(Y_t-X_t\beta_t)'/T\right]$. For any $R<R_0$, we make the following decomposition:

$$V(R,\boldsymbol{\beta})=\frac{1}{N}\sum_{r=R_0+1}^{N}s_r(\boldsymbol{\beta})+\frac{1}{N}\sum_{r=R+1}^{R_0}s_r(\boldsymbol{\beta})\equiv S_1(\boldsymbol{\beta})+S_{2R}(\boldsymbol{\beta}).$$

Noting that $S_1(\dot{\boldsymbol{\beta}}_R)\geq S_1(\dot{\boldsymbol{\beta}}_{R_0})=V(R_0,\dot{\boldsymbol{\beta}}_{R_0})$, we have

$$V(R,\dot{\boldsymbol{\beta}}_R)-V(R_0,\dot{\boldsymbol{\beta}}_{R_0})=\left[S_1(\dot{\boldsymbol{\beta}}_R)-S_1(\dot{\boldsymbol{\beta}}_{R_0})\right]+S_{2R}(\dot{\boldsymbol{\beta}}_R)\geq S_{2R}(\dot{\boldsymbol{\beta}}_R).$$

Let $s_r^0 = \mu_r\left(\frac{1}{T}\sum_{t=1}^T\left[\mathbf{\Lambda}^0 f_t^0 f_t^{0\prime}\mathbf{\Lambda}^{0\prime} + \varepsilon_t\varepsilon_t' + X_t(\beta_t^0 - \dot{\beta}_{t,R})(\beta_t^0 - \dot{\beta}_{t,R})'X_t'\right]\right)$. Notice that

$$
\begin{aligned}
& \frac{1}{N}\left|s_r(\dot{\boldsymbol{\beta}}_R) - s_r^0\right| \\
\leq\ & \frac{1}{NT}\Big\|\sum_{t=1}^T\Big\{(\mathbf{\Lambda}^0 f_t^0\varepsilon_t' + \varepsilon_t f_t^{0\prime}\mathbf{\Lambda}^{0\prime}) + [\mathbf{\Lambda}^0 f_t^0(\beta_t^0 - \dot{\beta}_{t,R})'X_t' + X_t(\beta_t^0 - \dot{\beta}_{t,R})f_t^{0\prime}\mathbf{\Lambda}^{0\prime}] \\
& +[\varepsilon_t(\beta_t^0 - \dot{\beta}_{t,R})'X_t' + X_t(\beta_t^0 - \dot{\beta}_{t,R})\varepsilon_t']\Big\}\Big\|_{\mathrm{sp}} \\
\leq\ & \frac{2}{NT}\left\|\sum_{t=1}^T\mathbf{\Lambda}^0 f_t^0\varepsilon_t'\right\|_{\mathrm{sp}} + \frac{2}{NT}\left\|\sum_{t=1}^T\mathbf{\Lambda}^0 f_t^0(\beta_t^0 - \dot{\beta}_{t,R})'X_t'\right\|_{\mathrm{sp}} + \frac{2}{NT}\left\|\sum_{t=1}^T\varepsilon_t(\beta_t^0 - \dot{\beta}_{t,R})'X_t'\right\|_{\mathrm{sp}}.
\end{aligned}
$$

Under Assumptions 1-2 and using the fact that $\frac{1}{T}\|\dot{\boldsymbol{d}}_{\beta,R}\|^2 = o_P(1)$, we can readily show that the second and third terms in the last expression are $o_P(1)$. The first term is $O_P((NT)^{-1/2})$ by Assumption 1(iii). It follows that

$$
\begin{aligned}
S_{2R}(\dot{\boldsymbol{\beta}}_R) &\geq \frac{1}{N}\sum_{r=R+1}^{R_0} s_r^0 + o_P(1) \\
&\geq \frac{1}{NT}\sum_{r=R+1}^{R_0}\mu_r\left(\mathbf{\Lambda}^0\mathbf{F}^{0\prime}\mathbf{F}^0\mathbf{\Lambda}^{0\prime}\right) + o_P(1) \\
&\geq (R_0 - R)\,\mu_{\min}(\mathbf{F}^{0\prime}\mathbf{F}^0/T)\mu_{\min}(\mathbf{\Lambda}^{0\prime}\mathbf{\Lambda}^0/N) + o_P(1) \\
&= (R_0 - R)\,\mu_{\min}(\mathbf{\Sigma}_F)\mu_{\min}(\mathbf{\Sigma}_\Lambda) + o_P(1),
\end{aligned}
$$

where the second inequality follows from Weyl's inequality. In sum, we have

$$
\operatorname{plim}\inf_{(N,T)\to\infty} V(R,\dot{\boldsymbol{\beta}}_R) - V(R_0,\dot{\boldsymbol{\beta}}_{R_0}) \geq c_R, \quad c_R = (R_0 - R)\,\mu_{\min}(\mathbf{\Sigma}_F)\mu_{\min}(\mathbf{\Sigma}_\Lambda)/2,
$$

completing the proof of Lemma B.4(i).

(ii) Recall that $V(R,\dot{\boldsymbol{\beta}}_R) = \min_{\boldsymbol{\beta},\mathbf{\Lambda}_R}\hat{Q}_{NT}(\boldsymbol{\beta},\mathbf{\Lambda}_R)$ subject to $\mathbf{\Lambda}_R'\mathbf{\Lambda}_R/N = \boldsymbol{I}_R$. Noting that $V(R,\dot{\boldsymbol{\beta}}_R) = \hat{Q}_{NT}(\dot{\boldsymbol{\beta}}_R,\dot{\mathbf{\Lambda}}_R)$, by the triangle inequality, we have

$$
\begin{aligned}
& \left|V(R,\dot{\boldsymbol{\beta}}_R) - V(R_0,\dot{\boldsymbol{\beta}}_{R_0})\right| \\
\leq\ & \left|\hat{Q}_{NT}(\dot{\boldsymbol{\beta}}_R,\dot{\mathbf{\Lambda}} - R) - \hat{Q}_{NT}(\boldsymbol{\beta}^0,\mathbf{\Lambda}^0)\right| + \left|\hat{Q}_{NT}(\dot{\boldsymbol{\beta}}_{R_0},\dot{\mathbf{\Lambda}}_{R_0}) - \hat{Q}_{NT}(\boldsymbol{\beta}^0,\mathbf{\Lambda}^0)\right| \\
\leq\ & 2\max_{R_0\leq R\leq R_{\max}}\left|\hat{Q}_{NT}(\dot{\boldsymbol{\beta}}_R,\dot{\mathbf{\Lambda}}_R) - \hat{Q}_{NT}(\boldsymbol{\beta}^0,\mathbf{\Lambda}^0)\right|.
\end{aligned}
$$

It suffices to show that $\hat{Q}_{NT}(\dot{\boldsymbol{\beta}}_R,\dot{\mathbf{\Lambda}}_R) - \hat{Q}_{NT}(\boldsymbol{\beta}_0,\mathbf{\Lambda}_0) = O_P\left(\delta_{p,NT}^{-2}\right)$ for each $R \in [R_0, R_{\max}]$. Let $\dot{\boldsymbol{H}}_R^+$ denote the Moore-Penrose generalized inverse of $\dot{\boldsymbol{H}}_R$ such that $\dot{\boldsymbol{H}}_R\dot{\boldsymbol{H}}_R^+ = \boldsymbol{I}_{R_0}$; see, for example, the proof of Lemma C.6(iv). Noting that $Y_t - X_t\beta_t^0 = \mathbf{\Lambda}^0 f_t^0 + \varepsilon_t$ and $\boldsymbol{M}_{\mathbf{\Lambda}^0}\mathbf{\Lambda}^0 = \mathbf{0}$, we may show that

$$\hat{Q}_{NT}(\boldsymbol{\beta}^0, \mathbf{\Lambda}^0) = \frac{1}{NT}\sum_{t=1}^{T}(Y_t - X_t\beta_t^0)'\boldsymbol{M}_{\mathbf{\Lambda}^0}(Y_t - X_t\beta_t^0) = \frac{1}{NT}\sum_{t=1}^{T}\varepsilon_t'\boldsymbol{M}_{\mathbf{\Lambda}^0}\varepsilon_t.$$

Let $\breve{\varepsilon}_t = \varepsilon_t - (\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R)\dot{\boldsymbol{H}}_R^+ f_t^0$. Noting that

$$\begin{aligned}
Y_t - X_t\dot{\beta}_{t,R} &= (X_t\beta_t^0 + \mathbf{\Lambda}^0 f_t^0 + \varepsilon_t) - X_t\dot{\beta}_{t,R} \\
&= X_t(\beta_t^0 - \dot{\beta}_{t,R}) + \breve{\mathbf{\Lambda}}_R\dot{\boldsymbol{H}}_R^+ f_t^0 + \varepsilon_t + (\mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R - \breve{\mathbf{\Lambda}}_R)\dot{\boldsymbol{H}}_R^+ f_t^0 \\
&= X_t(\beta_t^0 - \dot{\beta}_{t,R}) + \breve{\mathbf{\Lambda}}_R\dot{\boldsymbol{H}}_R^+ f_t^0 + \breve{\varepsilon}_t
\end{aligned}$$

and $\boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R}\breve{\mathbf{\Lambda}}_R = \boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R}\left(\dot{\mathbf{\Lambda}}_R\dot{\boldsymbol{V}}_{NT,R}\right) = \mathbf{0}$, we have

$$\begin{aligned}
\hat{Q}_{NT}(\dot{\boldsymbol{\beta}}_R, \dot{\mathbf{\Lambda}}_R) &= \frac{1}{NT}\sum_{t=1}^{T}(Y_t - X_t\dot{\beta}_{t,R})'\boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R}(Y_t - X_t\dot{\beta}_{t,R}) \\
&= \frac{1}{NT}\sum_{t=1}^{T}\left[X_t(\beta_t^0 - \dot{\beta}_{t,R}) + \breve{\varepsilon}_t\right]'\boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R}\left[X_t(\beta_t^0 - \dot{\beta}_{t,R}) + \breve{\varepsilon}_t\right] \\
&= \frac{1}{NT}\sum_{t=1}^{T}\breve{\varepsilon}_t'\boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R}\breve{\varepsilon}_t + \frac{1}{NT}\sum_{t=1}^{T}(\dot{\beta}_{t,R} - \beta_t^0)'X_t'\boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R}X_t(\dot{\beta}_{t,R} - \beta_t^0) \\
&\quad -\frac{2}{NT}\sum_{t=1}^{T}\breve{\varepsilon}_t'\boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R}X_t(\dot{\beta}_{t,R} - \beta_t^0) \\
&\equiv I_1 + I_2 - 2I_3.
\end{aligned}$$

We next prove Lemma B.4(ii) by only showing that

$$I_1 - \hat{Q}_{NT}(\boldsymbol{\beta}^0, \mathbf{\Lambda}^0) = O_P\left(\delta_{p,NT}^{-2}\right),$$

and

$$I_2 = O_P\left(\delta_{p,NT}^{-2}\right), \quad I_3 = O_P\left(\delta_{p,NT}^{-2}\right).$$

First, using $\breve{\varepsilon}_t = \varepsilon_t - (\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R)\dot{\boldsymbol{H}}_R^+ f_t^0$, we make the following decomposition:

$$\begin{aligned}
I_1 &= \frac{1}{NT}\sum_{t=1}^{T}[\varepsilon_t - (\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R)\dot{\boldsymbol{H}}_R^+ f_t^0]'\boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R}[\varepsilon_t - (\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R)\dot{\boldsymbol{H}}_R^+ f_t^0] \\
&= \frac{1}{NT}\sum_{r=1}^{T}\varepsilon_t'\boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R}\varepsilon_t - \frac{2}{NT}\sum_{r=1}^{T}f_t^{0\prime}\dot{\boldsymbol{H}}_R^{+\prime}(\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R)'\boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R}\varepsilon_t \\
&\quad +\frac{1}{NT}\sum_{r=1}^{T}f_t^{0\prime}\dot{\boldsymbol{H}}_R^{+\prime}(\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R)'\boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R}(\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0\dot{\boldsymbol{H}}_R)\dot{\boldsymbol{H}}_R^+ f_t^0 \\
&\equiv I_{1,1} - 2I_{1,2} + I_{1,3}.
\end{aligned}$$

Using the arguments as in the proof of Lemmas C.1(iii)(iv), we can show that

$$I_{1,1}-\hat{Q}_{NT}(\boldsymbol{\beta}^0,\boldsymbol{\Lambda}^0)=\frac{1}{NT}\sum_{t=1}^{T}\varepsilon_t'(\boldsymbol{P}_{\boldsymbol{\Lambda}^0}-\boldsymbol{P}_{\dot{\boldsymbol{\Lambda}}^R})\varepsilon_t=O_P\left(\delta_{NT}^{-2}\right)=O_P\left(\delta_{p,NT}^{-2}\right).$$

For $I_{1,2}$, we have

$$\begin{aligned}I_{1,2} &= \frac{1}{NT}\sum_{t=1}^{T}f_t^{0\prime}\dot{\boldsymbol{H}}_R^{+\prime}\left(\breve{\boldsymbol{\Lambda}}_R-\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R\right)'\varepsilon_t-\frac{1}{NT}\sum_{t=1}^{T}f_t^{0\prime}\dot{\boldsymbol{H}}_R^{+\prime}(\breve{\boldsymbol{\Lambda}}_R-\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R)'\boldsymbol{P}_{\dot{\boldsymbol{\Lambda}}_R}\varepsilon_t\\ &\equiv I_{1,2a}-I_{1,2b}.\end{aligned}$$

Using the decomposition in (C.111) and Lemma C.6(i), we can readily show that $I_{1,12a}=O_P\left(\delta_{p,NT}^{-2}\right)$. By the Cauchy-Schwarz inequality, the fact that $\boldsymbol{P}_{\dot{\boldsymbol{\Lambda}}_R}$ is a projection matrix, and Lemma C.1(iii),

$$\begin{aligned}|I_{1,2b}| &\leq \left[\frac{1}{NT}\sum_{t=1}^{T}\left\|(\breve{\boldsymbol{\Lambda}}_R-\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R)\dot{\boldsymbol{H}}_R^{+}f_t^0\right\|^2\right]^{1/2}\left[\frac{1}{NT}\sum_{t=1}^{T}\varepsilon_t'\boldsymbol{P}_{\dot{\boldsymbol{\Lambda}}_R}\varepsilon_t\right]^{1/2}\\ &= O_P\left(\delta_{p,NT}^{-1}\right)\cdot O_P\left(\delta_{NT}^{-1}\right)=O_P\left(\delta_{p,NT}^{-2}\right),\end{aligned}$$

where the following result which can be proved by Lemma C.6 has also been used:

$$\begin{aligned}\frac{1}{NT}\sum_{t=1}^{T}\left\|(\breve{\boldsymbol{\Lambda}}_R-\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R)\dot{\boldsymbol{H}}_R^{+}f_t^0\right\|^2 &\leq \frac{1}{N}\left\|\breve{\boldsymbol{\Lambda}}_R-\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R\right\|^2\left\|\dot{\boldsymbol{H}}_R^{+}\right\|^2\frac{1}{T}\sum_{t=1}^{T}\left\|f_t^0\right\|^2\\ &= O_P\left(\delta_{p,NT}^{-2}\right).\end{aligned}\qquad\text{(C.113)}$$

Thus we have $I_{1,2}=O_P\left(\delta_{p,NT}^{-2}\right)$. Similarly, using the fact that $\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R}$ is a projection matrix and by (C.113),

$$I_{1,3}\leq\frac{1}{NT}\sum_{r=1}^{T}\left\|(\breve{\boldsymbol{\Lambda}}_R-\boldsymbol{\Lambda}^0\dot{\boldsymbol{H}}_R)\dot{\boldsymbol{H}}_R^{+}f_t^0\right\|^2=O_P\left(\delta_{p,NT}^{-2}\right).$$

As a consequence, we may complete the proof of $I_1-\hat{Q}_{NT}(\boldsymbol{\beta}^0,\boldsymbol{\Lambda}^0)=O_P(\delta_{p,NT}^{-2})$ for each $R\in[R_0,R_{\max}]$.

Next, by Assumption 1(ii) and the fact that $\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R}$ is a projection matrix and that $\dot{\eta}_R=\frac{1}{T}\sum_{t=1}^{T}||\dot{\beta}_{t,R}-\beta_t^0||^2=O_P\left(\delta_{p,NT}^{-2}\right)$, we have

$$I_2\leq\frac{1}{NT}\sum_{t=1}^{T}\left\|(\dot{\beta}_{t,R}-\beta_t^0)'X_t'\boldsymbol{M}_{\dot{\boldsymbol{\Lambda}}_R}X_t(\dot{\beta}_{t,R}-\beta_t^0)\right\|\leq\max_{1\leq t\leq T}\mu_{\max}\left(X_t'X_t/N\right)\dot{\eta}_R=O_P\left(\delta_{p,NT}^{-2}\right).$$

To study I_3, we apply $\breve{\varepsilon}_t = \varepsilon_t - (\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0 \dot{\boldsymbol{H}}_R)\dot{\boldsymbol{H}}_R^+ f_t^0$ and $\boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R} = \boldsymbol{I}_N - \boldsymbol{P}_{\dot{\mathbf{\Lambda}}_R}$ and make the following decomposition:

$$\begin{aligned} I_3 &= \frac{1}{NT}\sum_{t=1}^T \breve{\varepsilon}_t' \boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R} X_t(\dot{\beta}_{t,R} - \beta_t^0) \\ &= \frac{1}{NT}\sum_{t=1}^T \varepsilon_t' X_t(\dot{\beta}_{t,R} - \beta_t^0) - \frac{1}{NT}\sum_{t=1}^T \varepsilon_t' \boldsymbol{P}_{\dot{\mathbf{\Lambda}}_R} X_t(\dot{\beta}_{t,R} - \beta_t^0) \\ &\quad - \frac{1}{NT}\sum_{t=1}^T f_t^{0\prime} \dot{\boldsymbol{H}}_R^{+\prime} (\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0 \dot{\boldsymbol{H}}_R)' \boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R} X_t(\dot{\beta}_{t,R} - \beta_t^0) \\ &\equiv I_{3,1} - I_{3,2} - I_{3,3}. \end{aligned}$$

By the Cauchy-Schwarz inequality, Assumptions 1(ii)-(iii), the fact that

$$\dot{\eta}_R = \frac{1}{T}\sum_{t=1}^T \left\| \dot{\beta}_{t,R} - \beta_t^0 \right\|^2 = O_P\left(\delta_{p,NT}^{-2}\right), \quad \frac{1}{NT}\sum_{t=1}^T \varepsilon_t' \boldsymbol{P}_{\dot{\mathbf{\Lambda}}_R} \varepsilon_t = O_P\left(\delta_{NT}^{-2}\right), \quad \mu_{\max}(\boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R}) = 1,$$

and Lemma C.6(i), we have

$$\begin{aligned} |I_{3,1}| &\leq \left[\frac{1}{N^2 T}\sum_{t=1}^T \varepsilon_t' X_t X_t' \varepsilon_t\right]^{1/2} \dot{\eta}_R^{1/2} = O_P(p^{1/2} N^{-1/2}) O_P(\delta_{p,NT}^{-1}) = O_P\left(\delta_{p,NT}^{-2}\right), \\ |I_{3,2}| &\leq \left[\frac{1}{NT}\sum_{t=1}^T \varepsilon_t' \boldsymbol{P}_{\dot{\mathbf{\Lambda}}_R} \varepsilon_t\right]^{1/2} \left[\frac{1}{NT}\sum_{t=1}^T (\dot{\beta}_{t,R} - \beta_t^0)' X_t' X_t (\dot{\beta}_{t,R} - \beta_t^0)\right]^{1/2} \\ &\leq O_P\left(\delta_{NT}^{-1}\right) \mu_{\max}\left(X_t' X_t / N\right)^{1/2} \dot{\eta}_R^{1/2} = O_P\left(\delta_{p,NT}^{-2}\right), \end{aligned}$$

and

$$\begin{aligned} |I_{3,3}| &\leq \left[\frac{1}{NT}\sum_{t=1}^T f_t^{0\prime} \dot{\boldsymbol{H}}_R^{+\prime} (\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0 \dot{\boldsymbol{H}}_R)' \boldsymbol{M}_{\dot{\mathbf{\Lambda}}_R} (\breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0 \dot{\boldsymbol{H}}_R) \dot{\boldsymbol{H}}_R^+ f_t^0\right]^{1/2} \\ &\quad \times \left[\frac{1}{NT}\sum_{t=1}^T (\dot{\beta}_{t,R} - \beta_t^0)' X_t' X_t (\dot{\beta}_{t,R} - \beta_t^0)\right]^{1/2} \\ &\leq \frac{1}{N^{1/2}} \left\| \breve{\mathbf{\Lambda}}_R - \mathbf{\Lambda}^0 \dot{\boldsymbol{H}}_R \right\| \left\| \dot{\boldsymbol{H}}_R^+ \right\| \left[\frac{1}{T}\sum_{t=1}^T \left\| f_t^0 \right\|^2\right]^{1/2} \mu_{\max}^{1/2}\left(X_t' X_t / N\right) \dot{\eta}_R^{1/2} \\ &= O_P(\delta_{p,NT}^{-1}) O_P(1) O_P(\delta_{p,NT}^{-1}) = O_P\left(\delta_{p,NT}^{-2}\right). \end{aligned}$$

Hence $I_3 = O_P\left(\delta_{p,NT}^{-2}\right)$. In sum, we have shown that $\hat{Q}_{NT}(\dot{\boldsymbol{\beta}}_R, \dot{\mathbf{\Lambda}}_R) - \hat{Q}_{NT}(\boldsymbol{\beta}_0, \mathbf{\Lambda}_0) = O_P\left(\delta_{p,NT}^{-2}\right)$ for each $R \in [R_0, R_{\max}]$, completing the proof of Lemma B.4(ii). ■

Proof of Lemma B.5. Let

$$D_{NT}\left(\boldsymbol{\alpha}_m, \boldsymbol{\Lambda}; \mathcal{T}_m\right) = \frac{1}{NT} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}}^{T_j-1} \left[(Y_t - X_t\alpha_j)' \boldsymbol{M}_{\boldsymbol{\Lambda}} \left(Y_t - X_t\alpha_j\right) - \varepsilon_t'\varepsilon_t\right]$$

and $\bar{\sigma}_{NT}^2 = \frac{1}{NT}\sum_{t=1}^T \varepsilon_t'\varepsilon_t$. Note that

$$\left(\tilde{\boldsymbol{\alpha}}_m(\mathcal{T}_m), \tilde{\boldsymbol{\Lambda}}(\mathcal{T}_m)\right) = \arg\min_{(\boldsymbol{\alpha}_m, \boldsymbol{\Lambda})} D_{NT}\left(\boldsymbol{\alpha}_m, \boldsymbol{\Lambda}; \mathcal{T}_m\right),$$

and

$$\tilde{\sigma}^2(\mathcal{T}_m) - \tilde{\sigma}^2(\mathcal{T}_{m^0}^0) = \left[\tilde{\sigma}^2(\mathcal{T}_m) - \bar{\sigma}_{NT}^2\right] - \left[\tilde{\sigma}^2(\mathcal{T}_{m^0}^0) - \bar{\sigma}_{NT}^2\right]$$

with $\tilde{\sigma}^2(\mathcal{T}_m) - \bar{\sigma}_{NT}^2 = D_{NT}(\tilde{\boldsymbol{\alpha}}_m(\mathcal{T}_m), \tilde{\boldsymbol{\Lambda}}(\mathcal{T}_m); \mathcal{T}_m)$. We prove the lemma by showing that (i)

$$\frac{m^0}{T\Delta_{NT}^2}\left[\tilde{\sigma}^2(\mathcal{T}_{m^0}^0) - \bar{\sigma}_{NT}^2\right] = o_P(1); \tag{C.114}$$

and (ii)

$$\frac{m^0}{T\Delta_{NT}^2}(\tilde{\sigma}^2(\mathcal{T}_m) - \bar{\sigma}_{NT}^2) \geq c + o_P(1) \ \text{ w.p.a.1 \ for \ some } \ c > 0. \tag{C.115}$$

We first show (C.114) in (i). We make the following decomposition:

$$\begin{aligned}
\tilde{\sigma}^2_{\mathcal{T}_{m^0}^0} &= \frac{1}{NT} \sum_{j=1}^{m^0+1} \sum_{t=T_{j-1}^0}^{T_j^0-1} [Y_t - X_t\tilde{\alpha}_j]' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} [Y_t - X_t\tilde{\alpha}_j] \\
&= \frac{1}{NT} \sum_{j=1}^{m^0+1} \sum_{t=T_{j-1}^0}^{T_j^0-1} \left[X_t(\alpha_j^0 - \tilde{\alpha}_j) + \boldsymbol{\Lambda}^0 f_t^0 + \varepsilon_t\right]' \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} \left[X_t(\alpha_j^0 - \tilde{\alpha}_j) + \boldsymbol{\Lambda}^0 f_t^0 + \varepsilon_t\right] \\
&= \frac{1}{NT} \sum_{j=1}^{m^0+1} \sum_{t=T_{j-1}^0}^{T_j^0-1} \left[\varepsilon_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}\varepsilon_t + f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0 f_t^0 + (\alpha_j^0 - \tilde{\alpha}_j)'X_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}X_t(\alpha_j^0 - \tilde{\alpha}_j)\right. \\
&\quad \left. + 2\varepsilon_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}X_t(\alpha_j^0 - \tilde{\alpha}_j) + 2\varepsilon_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}\boldsymbol{\Lambda}^0 f_t^0 + 2f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}X_t(\alpha_j^0 - \tilde{\alpha}_j)\right] \\
&\equiv d_{1NT} + d_{2NT} + d_{3NT} + 2d_{4NT} + 2d_{5NT} + 2d_{6NT},
\end{aligned}$$

where we suppress the dependence of $\tilde{\alpha}_j = \tilde{\alpha}_j(\mathcal{T}_{m^0}^0)$ and $\tilde{\boldsymbol{\Lambda}} = \tilde{\boldsymbol{\Lambda}}(\mathcal{T}_{m^0}^0)$ on $\mathcal{T}_{m^0}^0$ for notational simplicity. By Lemma C.1(iii),

$$d_{1NT} = \frac{1}{NT}\sum_{t=1}^T \varepsilon_t'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}\varepsilon_t = \frac{1}{NT}\sum_{t=1}^T \varepsilon_t'\varepsilon_t + O_P\left(\delta_{NT}^{-2}\right) = \bar{\sigma}_{NT}^2 + O_P\left(\delta_{NT}^{-2}\right).$$

Using the preliminary results in Lemmas C.4 and C.5(i) and Theorem 3.4, we may show that $d_{lNT} = O_P(\delta_{p,NT}^{-2})$ for $l = 3, 4, 6$. Using $\boldsymbol{M}_{\boldsymbol{\Lambda}^0}\boldsymbol{\Lambda}^0 = 0$ and (C.79), and decomposing $\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} - \boldsymbol{M}_{\boldsymbol{\Lambda}^0} = -(\boldsymbol{P}_{\tilde{\boldsymbol{\Lambda}}} - \boldsymbol{P}_{\boldsymbol{\Lambda}^0})$ as in (C.81), we can readily show that

$$\begin{aligned} d_{2NT} &= \frac{1}{NT}\sum_{t=1}^{T} f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}\left(\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} - \boldsymbol{M}_{\boldsymbol{\Lambda}^0}\right)\boldsymbol{\Lambda}^0 f_t^0 = O_P\left(\delta_{p,NT}^{-2}\right), \text{ and} \\ d_{5NT} &= \frac{1}{NT}\sum_{t=1}^{T} \varepsilon_t'\left(\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}} - \boldsymbol{M}_{\boldsymbol{\Lambda}^0}\right)\boldsymbol{\Lambda}^0 f_t^0 = O_P\left(\delta_{p,NT}^{-2}\right). \end{aligned}$$

It follows that

$$\tilde{\sigma}^2(\mathcal{T}_{m^0}^0) - \bar{\sigma}_{NT}^2 = O_P\left(\delta_{p,NT}^{-2}\right), \tag{C.116}$$

which, together with Assumption 2(ii), leads to (C.114).

We now show (C.115) in (ii). We consider three cases: (a) $m^0 = 1$, (b) $m^0 = 2$, and (c) $3 < m^0 \le m_{\max}$. For case (a) of $m^0 = 1$, if $n < m^0$, we have $m = 0$ and $\mathcal{T}_m = \mathcal{T}_0 = \varnothing$. The true model contains one structural break:

$$Y_t = \begin{cases} X_t\alpha_1^0 + \boldsymbol{\Lambda}^0 f_t^0 + \varepsilon_t & \text{if } 1 \le t \le T_1^0 - 1, \\ X_t\alpha_2^0 + \boldsymbol{\Lambda}^0 f_t^0 + \varepsilon_t & \text{if } T_1^0 \le t \le T; \end{cases}$$

while the working model that ignores the structural break in the regression coefficient is

$$Y_t = X_t\alpha + \boldsymbol{\Lambda}^0 f_t^0 + e_t,\ 1 \le t \le T,$$

where e_t is the error term. Note that $\tilde{\sigma}^2(\mathcal{T}_0) = \frac{1}{NT}\sum_{t=1}^{T}(Y_t - X_t\tilde{\alpha})'\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}(Y_t - X_t\tilde{\alpha})$, where

$$(\tilde{\alpha}, \tilde{\boldsymbol{\Lambda}}) = \arg\min_{\alpha,\boldsymbol{\Lambda}} \frac{1}{NT}\sum_{t=1}^{T}(Y_t - X_t\alpha)'\boldsymbol{M}_{\boldsymbol{\Lambda}}(Y_t - X_t\alpha)$$

subject to $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}/N = \boldsymbol{I}_{R_0}$, and we suppress the dependence of $\tilde{\alpha}$ and $\tilde{\boldsymbol{\Lambda}}$ on $\mathcal{T}_0$. Using $Y_t - X_t\alpha = X_t(\beta_t^0 - \alpha) + \boldsymbol{\Lambda}^0 f_t^0 + \varepsilon_t$ and Lemmas C.1(i)(ii), we can readily show that

$$\begin{aligned} &\frac{1}{NT}\sum_{t=1}^{T}(Y_t - X_t\alpha)'\boldsymbol{M}_{\boldsymbol{\Lambda}}(Y_t - X_t\alpha) \\ =& \frac{1}{NT}\sum_{t=1}^{T}\left[X_t(\beta_t^0 - \alpha) + \boldsymbol{\Lambda}^0 f_t^0 + \varepsilon_t\right]'\boldsymbol{M}_{\boldsymbol{\Lambda}}\left[X_t(\beta_t^0 - \alpha) + \boldsymbol{\Lambda}^0 f_t^0 + \varepsilon_t\right] \\ =& \frac{1}{NT}\sum_{t=1}^{T}\left[X_t(\beta_t^0 - \alpha) + \boldsymbol{\Lambda}^0 f_t^0\right]'\boldsymbol{M}_{\boldsymbol{\Lambda}}\left[X_t(\beta_t^0 - \alpha) + \boldsymbol{\Lambda}^0 f_t^0\right] + \frac{1}{NT}\sum_{t=1}^{T}\varepsilon_t'\varepsilon_t + O_P(p^{1/2}\delta_{p,NT}^{-1}) \end{aligned}$$

uniformly in α and $\mathbf{\Lambda}$ such that $\mathbf{\Lambda}'\mathbf{\Lambda}/N = \boldsymbol{I}_{R_0}$ and $\|\alpha\| \leq Cp^{1/2}$. It follows that

$$\begin{aligned}
\tilde{\sigma}^2(\mathcal{T}_0) &= \frac{1}{NT}\sum_{t=1}^{T}\tilde{Y}_t'\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}}\tilde{Y}_t + \bar{\sigma}_{NT}^2 + O_P(p^{1/2}\delta_{p,NT}^{-1}) \\
&\geq \min_{\mathbf{\Lambda}:\ \mathbf{\Lambda}'\mathbf{\Lambda}/N=\boldsymbol{I}_{R_0}} \frac{1}{NT}\sum_{t=1}^{T}\tilde{Y}_t'\boldsymbol{M}_{\mathbf{\Lambda}}\tilde{Y}_t + \bar{\sigma}_{NT}^2 + O_P(p^{1/2}\delta_{p,NT}^{-1}) \\
&= \frac{1}{NT}\sum_{r=R_0+1}^{N}\mu_r\left[\sum_{t=1}^{T}\tilde{Y}_t\tilde{Y}_t'\right] + \bar{\sigma}_{NT}^2 + O_P(p^{1/2}\delta_{p,NT}^{-1}) \\
&\geq \frac{1}{NT}\sum_{r=R_0+1}^{N}\mu_r\left[\sum_{t=1}^{T}X_t(\beta_t^0-\tilde{\alpha})(\beta_t^0-\tilde{\alpha})'X_t'\right] + \bar{\sigma}_{NT}^2 + O_P(p^{1/2}\delta_{p,NT}^{-1}) \\
&= \frac{1}{NT}\min_{\mathbf{\Lambda}:\ \mathbf{\Lambda}'\mathbf{\Lambda}/N=\boldsymbol{I}_{R_0}}\left[\sum_{t=1}^{T}(\beta_t^0-\tilde{\alpha})'X_t'\boldsymbol{M}_{\mathbf{\Lambda}}X_t(\beta_t^0-\tilde{\alpha})\right] + \bar{\sigma}_{NT}^2 + O_P(p^{1/2}\delta_{p,NT}^{-1}) \\
&\geq c_x\cdot\frac{1}{T}\sum_{t=1}^{T}\left\|\beta_t^0-\tilde{\alpha}\right\|^2 + \bar{\sigma}_{NT}^2 + O_P(p^{1/2}\delta_{p,NT}^{-1}),
\end{aligned}$$

where $\tilde{Y}_t = X_t(\beta_t^0 - \tilde{\alpha}) + \mathbf{\Lambda}^0 f_t^0$, the second and third inequalities follow from Weyl's inequality and Assumption 1(ii), respectively. Consequently, we have by Assumptions 5(i)-(ii)

$$\frac{m^0}{T\Delta_{NT}^2}\left[\tilde{\sigma}^2(\mathcal{T}_0) - \bar{\sigma}_{NT}^2\right] \geq c_x c_\beta + o_P(1),$$

where c_β is defined in Assumption 5(i). We have completed the proof of (C.115) for case (a).

In cases (b)-(c), it suffices to consider the case where $m = m^0 - 1$ (If $m < m^0 - 1$, one can always augment the set $\mathcal{T}_m$ by $m^0 - 1 - m$ true break points which are not inside $\mathcal{T}_m$ to make $D_{NT}(\tilde{\boldsymbol{\alpha}}_m(\mathcal{T}_m), \tilde{\mathbf{\Lambda}}(\mathcal{T}_m); \mathcal{T}_m)$ smaller). For the case (b) with $m = 1$, we consider three subcases: (b.1) $2 \leq T_1 \leq T_1^0$, (b.2) $T_1^0 < T_1 \leq T_2^0$, and (b.3) $T_2^0 < T_1 \leq T$. In the subcase (b.1), $[1, T_1 - 1]$ does not contain a break point while $[T_1, T]$ contains two true break points T_1^0 and T_2^0. Observe that

$$\begin{aligned}
D_{NT}(\tilde{\boldsymbol{\alpha}}_1(\mathcal{T}_1), \tilde{\mathbf{\Lambda}}(\mathcal{T}_1); \mathcal{T}_1) &= \frac{1}{NT}\sum_{t=1}^{T_1-1}\left\{[Y_t - X_t\tilde{\alpha}_1(\mathcal{T}_1)]'\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}(\mathcal{T}_1)}[Y_t - X_t\tilde{\alpha}_1(\mathcal{T}_1)] - \varepsilon_t'\varepsilon_t\right\} \\
&\quad + \frac{1}{NT}\sum_{t=T_1}^{T}\left\{[Y_t - X_t\tilde{\alpha}_2(\mathcal{T}_1)]'\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}(\mathcal{T}_1)}[Y_t - X_t\tilde{\alpha}_2(\mathcal{T}_1)] - \varepsilon_t'\varepsilon_t\right\} \\
&\equiv D_{NT,1} + D_{NT,2}.
\end{aligned}$$

Noting that the interval $[1, T_1 - 1]$ does not contain a break point, using the arguments as used in the study of case (a), we can readily show that

$$D_{NT,1} \geq \frac{c_x}{T} \sum_{t=1}^{T_1-1} \left\| \alpha_1^0 - \tilde{\alpha}_1(\mathcal{T}_1) \right\|^2 + O_P(p^{1/2} \delta_{p,NT}^{-1}).$$

Similarly, we can show that

$$D_{NT,2} \geq \frac{c_x}{T} \sum_{t=T_1}^{T} \left\| \beta_t^0 - \tilde{\alpha}_2(\mathcal{T}_1) \right\|^2 + O_P(p^{1/2} \delta_{p,NT}^{-1}).$$

Then by Assumptions 5(i)(ii)

$$\begin{aligned}
& \frac{m^0}{T\Delta_{NT}^2} D_{NT}(\tilde{\boldsymbol{\alpha}}_1(\mathcal{T}_1), \tilde{\boldsymbol{\Lambda}}(\mathcal{T}_1); \mathcal{T}_1) \\
\geq\ & \frac{m^0}{T\Delta_{NT}^2} \left\{ \frac{c_x}{T} \sum_{t=1}^{T_1-1} \left\| \alpha_1^0 - \tilde{\alpha}_1(\mathcal{T}_1) \right\|^2 + \frac{c_x}{T} \sum_{t=T_1}^{T} \left\| \beta_t^0 - \tilde{\alpha}_2(\mathcal{T}_1) \right\|^2 + O_P(p^{1/2} \delta_{p,NT}^{-1}) \right\} \\
\geq\ & c_x \min_{\alpha_1, \alpha_2} \frac{m^0}{T\Delta_{NT}^2} \sum_{j=1}^{2} \sum_{t=T_{j-1}}^{T_j-1} \left\| \beta_t^0 - \alpha_j \right\|^2 + o_P(1) \\
\geq\ & c_x c_\beta + o_P(1).
\end{aligned}$$

In the subcase (b.2), both $[2, T_1 - 1]$ and $[T_1, T]$ contain a break. As in subcase (b.1), we can show that

$$\begin{aligned}
& \frac{m^0}{T\Delta_{NT}^2} D_{NT}(\tilde{\boldsymbol{\alpha}}_1(\mathcal{T}_1), \tilde{\boldsymbol{\Lambda}}(\mathcal{T}_1); \mathcal{T}_1) \\
\geq\ & \frac{m^0}{T\Delta_{NT}^2} \left\{ \frac{c_x}{T} \sum_{t=1}^{T_1-1} \left\| \beta_t^0 - \tilde{\alpha}_1(\mathcal{T}_1) \right\|^2 + \frac{c_x}{T} \sum_{t=T_1}^{T} \left\| \beta_t^0 - \tilde{\alpha}_2(\mathcal{T}_1) \right\|^2 + O_P(pN^{-1/2} + p^{1/2} T^{-1/2}) \right\} \\
\geq\ & c_x \min_{\alpha_1, \alpha_2} \frac{m^0}{T\Delta_{NT}^2} \sum_{j=1}^{2} \sum_{t=T_{j-1}}^{T_j-1} \left\| \beta_t^0 - \alpha_j \right\|^2 \geq c_x c_\beta + o_P(1).
\end{aligned}$$

The proof for the subcase (b.3) is analogous to that for the subcase (b.1). Hence, the conclusion (C.115) follows in the subcase (b). Case (c) can be studied analogously. This completes the proof of the lemma. ■

Proof of Lemma B.6. For $\mathcal{T}_m \in \bar{\mathbb{T}}_m$ with $m^0 < m \leq m_{\max}$, we recall that

$$\begin{aligned}\tilde{\sigma}^2(\mathcal{T}_m) &= Q_{NT}(\tilde{\boldsymbol{\alpha}}_m(\mathcal{T}_m), \tilde{\mathbf{\Lambda}}(\mathcal{T}_m); \mathcal{T}_m) \\ &= \min_{\boldsymbol{\alpha}_m, \mathbf{\Lambda}} \frac{1}{NT} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}}^{T_j-1} (Y_t - X_t\alpha_j)' \boldsymbol{M}_{\mathbf{\Lambda}} (Y_t - X_t\alpha_j) \\ &= \min_{\boldsymbol{\alpha}_m} \frac{1}{NT} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}}^{T_j-1} (Y_t - X_t\alpha_j)' \boldsymbol{M}_{\tilde{\mathbf{\Lambda}}(\mathcal{T}_m)} (Y_t - X_t\alpha_j),\end{aligned}$$

and $\bar{\sigma}^2_{NT} = \frac{1}{NT}\sum_{t=1}^T \varepsilon_t'\varepsilon_t$. In view of the fact that

$$\tilde{\sigma}^2(\mathcal{T}^0_{m^0}) \geq \tilde{\sigma}^2(\mathcal{T}_m) \text{ and } \tilde{\sigma}^2(\mathcal{T}^0_{m^0}) = \bar{\sigma}^2_{NT} + O_P(\delta^{-2}_{p,NT})$$

by (C.116), we have

$$0 \leq \tilde{\sigma}^2(\mathcal{T}^0_{m^0}) - \tilde{\sigma}^2(\mathcal{T}_m) = \bar{\sigma}^2_{NT} - \tilde{\sigma}^2(\mathcal{T}_m) + O_P(\delta^{-2}_{p,NT}) = \sum_{j=1}^{m+1} J_{NT,j} + O_P(\delta^{-2}_{p,NT}), \tag{C.117}$$

where $J_{NT,j} \equiv -\inf_\alpha S_j(\alpha)$, $S_j(\alpha) = \frac{1}{NT}\sum_{t=T_{j-1}}^{T_j-1}\left[(Y_t - X_t\alpha)' \boldsymbol{M}_{\tilde{\mathbf{\Lambda}}(\mathcal{T}_m)}(Y_t - X_t\alpha) - \varepsilon_t'\varepsilon_t\right]$ and $[T_{j-1}, T_j - 1]$ does not contain any break point for $j = 1, ..., m+1$. Let $\alpha^0_{j,m} = \beta^0_{T_{j-1}}$ and $\tilde{\alpha}_{j,m} = \tilde{\alpha}_j(\mathcal{T}_m) = \arg\min_\alpha S_j(\alpha) = \left(\sum_{t=T_{j-1}}^{T_j-1} X_t'\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}(\mathcal{T}_m)}X_t\right)^{-1}\sum_{t=T_{j-1}}^{T_j-1} X_t'\,\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}(\mathcal{T}_m)}Y_t$ for $j = 1, ..., m+1$. As in the proofs of Lemma C.4(i) and Theorems 3.1 and 3.4, we can show that $\frac{1}{N}||\tilde{\mathbf{\Lambda}}(\mathcal{T}_m) - \mathbf{\Lambda}^0||^2 = O_P(\delta^{-2}_{p,NT})$ and $||\tilde{\alpha}_{j,m} - \alpha^0_{j,m}|| = O_P(\delta^{-1}_{p,NT})$. Then using $Y_t - X_t\tilde{\alpha}_{j,m} = \varepsilon_t + \mathbf{\Lambda}^0 f^0_t + X_t(\alpha^0_{j,m} - \tilde{\alpha}_{j,m})$, we have

$$\begin{aligned}S_j(\tilde{\alpha}_{j,m}) &= \frac{1}{NT}\sum_{t=T_{j-1}}^{T_j-1}\left[(Y_t - X_t\tilde{\alpha}_{j,m})'\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}(\mathcal{T}_m)}(Y_t - X_t\tilde{\alpha}_{j,m}) - \varepsilon_t'\varepsilon_t\right] \\ &= \frac{1}{NT}\sum_{t=T_{j-1}}^{T_j-1}\left\{\left[\varepsilon_t + \mathbf{\Lambda}^0 f^0_t + X_t(\alpha^0_{j,m} - \tilde{\alpha}_{j,m})\right]'\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}(\mathcal{T}_m)}\left[\varepsilon_t + \mathbf{\Lambda}^0 f^0_t + X_t(\alpha^0_{j,m} - \tilde{\alpha}_{j,m})\right] - \varepsilon_t'\varepsilon_t\right\} \\ &= \frac{-1}{NT}\sum_{t=T_{j-1}}^{T_j-1}\varepsilon_t'\boldsymbol{P}_{\tilde{\mathbf{\Lambda}}(\mathcal{T}_m)}\varepsilon_t + \frac{1}{NT}\sum_{t=T_{j-1}}^{T_j-1} f^{0\prime}_t\mathbf{\Lambda}^{0\prime}\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}(\mathcal{T}_m)}\mathbf{\Lambda}^0 f^0_t \\ &\quad + \frac{1}{NT}\sum_{t=T_{j-1}}^{T_j-1}\left(\alpha^0_{j,m} - \tilde{\alpha}_{j,m}\right)' X_t'\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}(\mathcal{T}_m)}X_t\left(\alpha^0_{j,m} - \tilde{\alpha}_{j,m}\right) + \frac{2}{NT}\sum_{t=T_{j-1}}^{T_j-1}\varepsilon_t'\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}(\mathcal{T}_m)}\mathbf{\Lambda}^0 f^0_t \\ &\quad + \frac{2}{NT}\sum_{t=T_{j-1}}^{T_j-1}\varepsilon_t'\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}(\mathcal{T}_m)}X_t\left(\alpha^0_{j,m} - \tilde{\alpha}_{j,m}\right) + \frac{2}{NT}\sum_{t=T_{j-1}}^{T_j-1} f^{0\prime}_t\mathbf{\Lambda}^{0\prime}\boldsymbol{M}_{\tilde{\mathbf{\Lambda}}(\mathcal{T}_m)}X_t\left(\alpha^0_{j,m} - \tilde{\alpha}_{j,m}\right) \\ &\equiv S_{j,1} + S_{j,2} + S_{j,3} + 2S_{j,4} + 2S_{j,5} + 2S_{j,6}.\end{aligned}$$

By Lemma C.1(iii),

$$\sum_{j=1}^{m+1} S_{j,1} = \frac{-1}{NT}\sum_{t=1}^{T} \varepsilon_t' \boldsymbol{P}_{\tilde{\boldsymbol{\Lambda}}(\mathcal{T}_m)} \varepsilon_t = O_P\left(\delta_{NT}^{-2}\right).$$

In addition, we can show that

$$\begin{aligned}
\sum_{j=1}^{m+1} S_{j,2} &= \frac{1}{NT}\sum_{t=1}^{T} f_t^{0\prime}\boldsymbol{\Lambda}^{0\prime}(\boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}(\mathcal{T}_m)} - \boldsymbol{M}_{\boldsymbol{\Lambda}^0})\boldsymbol{\Lambda}^0 f_t^0 = O_P\left(\delta_{p,NT}^{-2}\right), \\
\sum_{j=1}^{m+1} S_{j,3} &\leq \frac{1}{T}\sum_{j=1}^{m+1} \left\|\alpha_{j,m}^0 - \tilde{\alpha}_{j,m}\right\|^2 \sum_{t=T_{j-1}}^{T_j-1} \mu_{\max}\left(X_t' X_t / N\right) = O_P\left(\delta_{p,NT}^{-2}\right),
\end{aligned}$$

and similarly $\sum_{j=1}^{m+1} S_{j,l} = O_P\left(\delta_{p,NT}^{-2}\right)$ for $l = 4, 5, 6$. Then by (C.117), $\tilde{\sigma}^2(\mathcal{T}_m) - \bar{\sigma}_{NT}^2 = O_P\left(\delta_{p,NT}^{-2}\right)$ for all $m \in \left\{m^0+1, ..., m_{\max}\right\}$ and $\mathcal{T}_m = \{T_1, ..., T_m\}$, which completes the proof of Lemma B.6. ■

References

Bai, J. (2003), "Inferential theory for factor models with large dimensions," *Econometrica,* 71, 135-171.

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

Bai, J., Ng, S. (2002), "Determining the number of factors in approximate factor models," *Econometrica,* 70, 191-221.

Bernstein, D. S. (2005), Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory. Princeton University Press, Princeton.