Discussion Paper No. 1037

A ROBUST APPROACH TO HETEROSKEDASTICITY, ERROR SERIAL CORRELATION AND SLOPE HETEROGENEITY FOR LARGE LINEAR PANEL DATA MODELS WITH INTERACTIVE EFFECTS

Guowei Cui
Kazuhiko Hayakawa
Shuichi Nagata
Takashi Yamagata

Revised June 2019
July 2018

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

A robust approach to heteroskedasticity, error serial correlation and slope heterogeneity for large linear panel data models with interactive effects*

Guowei Cui†
Huazhong University of Science and Technology

Kazuhiko Hayakawa‡
Hiroshima University

Shuichi Nagata§
Kwansei Gakuin University

Takashi Yamagata¶
University of York and Osaka University

26 June 2019

Abstract

In this paper, we propose a robust approach against heteroskedasticity, error serial correlation and slope heterogeneity for large linear panel data models. First, we establish the asymptotic validity of the Wald test based on the widely used panel heteroskedasticity and autocorrelation consistent (HAC) variance estimator of the pooled estimator under random coefficient models. Then, we show that a similar result holds with the proposed bias-corrected Bai's (2009) estimator for models with unobserved interactive effects. Our new theoretical result justifies the use of the same slope estimator and the variance estimator, both for slope homogeneous and heterogeneous models. This robust approach can significantly reduce the model selection uncertainty for applied researchers. In addition, we propose a novel test for the correlation and dependence of the random coefficient with covariates. The test is of great importance, since the widely used estimators and/or its variance estimators can become inconsistent when the variation of coefficients depends on covariates, in general. The finite sample evidence supports the usefulness and reliability of our approach.

Key Words: panel data; slope heterogeneity; interactive effects; test for correlated random coefficients

JEL Classification: C12, C13, C23.

*The authors are grateful to Milda Norkutė for helpful comments.

†Department of Economics, Huazhong University of Science and Technology; E-mail address: cuigw610@163.com

‡Graduate School of Social Sciences, Department of Economics, 1-2-1 Kagamiyama, Higashi-Hiroshima City Hiroshima, 739-8525 Japan; E-mail address: kazuhaya@hiroshima-u.ac.jp. This author gratefully acknowledges the financial support by JSPS KAKENHI Grant Numbers 16H03606, 17K03660 and 17KK0070.

§School of Business Administration, Kwansei Gakuin University, 1-155 Ichiban-cho, Uegahara, Nishinomiya, Hyogo 662-8501, Japan; E-mail address: snagata@kwansei.ac.jp.

¶Corresponding author. Department of Economics and Related Studies, University of York, York YO10 5DD, UK; E-mail address: takashi.yamagata@york.ac.uk. This author gratefully acknowledges the financial support by JSPS KAKENHI Grant Numbers 18K01545 and 15H05728.

1 Introduction

The recently increasing availability of panel data sets in which both cross-section dimension N and times series dimension T are large has produced opportunities to develop statistical methods to exploit richer information, while presenting associated technical challenges. In particular, controlling cross-sectional dependence, heterogeneity in parameters and distributions, and serial dependence has been a main focus of the literature.

The celebrated fixed effects model permits intercept to be cross-sectionally heterogeneous whilst slope coefficients are constant across cross-section units and time. Hansen (2007) has shown that, under mild conditions, the heteroskedasticity and autocorrelation consistent (HAC) variance estimator of Arellano (1987), which is originally proposed for a short panel fixed effects estimator, will be asymptotically valid for large panels. Greenaway-McGrevy et al. (2012) propose to use the HAC estimator for the pooled principal component based (PC) estimator for the model with unobserved interactive effects.

The random-coefficient model, in which the slope coefficients are allowed to vary with the cross-sectional units, has attracted great attention in recent years.[1] It can control differences in behaviour across cross-section units which are not captured by the control variables. For such models, the estimate of interest is often the population average of slope coefficients. Interestingly, if the cross-sectional variation of slopes in the random coefficient model is independent of covariates, the fixed effects estimator is consistent to the population average of slope coefficients. A non-parametric variance-covariance estimator for such pooled estimators has been implicitly proposed in Pesaran (2006), in which the population variation of slopes is replaced by its sample counterpart – the variation of the estimates of cross-section specific slopes. The evidence has shown that the variance estimator behaves very well in finite samples.

There are some issues about this variance estimator for our robust approach. First, for the choice between the HAC and this variance estimator, the practitioner would like to know if there is slope heterogeneity or not. Second, some estimation methods, such as Bai's (2009) estimator, do not permit slope heterogeneity models, and making use of statistics involving individual slope estimates might not be asymptotically justified.

In this paper, we propose a robust approach against heteroskedasticy, error serial correlation and slope heterogeneity for large linear panel data models. First, we establish the asymptotic validity of the Wald test based on the panel HAC variance estimator for the pooled estimator under random coefficient models. Then, we show that a similar result holds with the bias-corrected Bai's (2009) estimator for models with interactive effects when the regressor has a factor structure. Our new theoretical result justifies the use of the original Bai's iterative estimator and the variance estimator, both for slope homogeneous and heterogeneous models. This robust approach is expected to substantially reduce the model selection uncertainty for applied researchers.[2]

Another main contribution of this paper is a novel test for the correlation and dependency of the random coefficient on covariates. We extend the test proposed by Wooledridge (2010) by robustifying against (uncorrelated) random coefficients, proposing a Lagrange Multiplier test along with a Wald test, and developing them for the models with unobservable interactive effects. The test is of great importance, since the widely used estimators and/or its variance estimators can become inconsistent when the variation of coefficients is correlated or dependent with covariates, in general.

We have examined the finite sample performance of the estimators, tests of linear restrictions,

[1] See Hsiao and Pesaran (2008) for an excellent survey of random coefficient panel data models.

[2] Galvao and Kato (2014) consider estimation and inference of fixed effects estimation for large panels under misspecification.

and the LM tests for correlated random coefficients. The evidence illustrates the usefulness of our approach. In particular, for the estimation of the models with unobserved interactive effects, the size of the proposed robust Wald test using the bias-corrected Bai's (2009) estimator is very close to the nominal level, under both slope homogeneity and slope heterogeneity, while maintaining satisfactory power. Also, the LM tests for correlated random coefficients have correct size under both slope homogeneity and slope heterogeneity due to pure random coefficients, while exhibiting high power when the random coefficients depend on covariates.

The paper is organised as follows. The robust Wald test is proposed for standard linear panel data models in Section 2, then for the models with unobserved interactive effects in Section 3. A test for correlation of slopes with covariates is proposed in Section 4. The finite sample performance of the proposed bias-corrected estimator, the associated Wald test and the correlation test is investigated using the Monte Carlo method in Section 5. Section 6 contains some concluding remarks. Proofs of the main results in Sections 2 and 3 are contained in Online Appendices A and B, respectively.

Notations: we denote the largest eigenvalues of the $N \times N$ matrix $\mathbf{A} = (a_{ij})$ by $\mu_{\max}(\mathbf{A})$, its trace by $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{N} a_{ii}$, its Frobenius norm by $\|\mathbf{A}\| = \sqrt{\mathrm{tr}(\mathbf{A}'\mathbf{A})}$. The projection matrix on $\mathbf{A}$ is $\mathbf{P_A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ and $\mathbf{M_A} = \mathbf{I} - \mathbf{P_A}$. Δ is a generic positive constant large enough, $\delta_{NT}^2 = \min\{N, T\}$. We use $N, T \to \infty$ to denote that N and T pass to infinity jointly.

2 Benchmark Panel Data Model

Consider a panel data model with cross-sectionally heterogeneous slopes:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it}, \qquad (i = 1, 2, ..., N, t = 1, 2, ..., T) \tag{1}$$

$\mathbf{x}_{it}$ is a $k \times 1$ vector of observed covariates, and ε_{it} is disturbances. The $k \times 1$ slope coefficients are generated as

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}^0 + \boldsymbol{\eta}_i, \tag{2}$$

where $\boldsymbol{\eta}_i$ is independently distributed random vector across i, with $\mathbb{E}(\boldsymbol{\eta}_i) = \mathbf{0}$. When $\boldsymbol{\eta}_i = \mathbf{0}$ for all i, it reduces to the homogeneous slope model. Throughout the paper, our interest is in the estimation and testing of the linear restrictions of $\boldsymbol{\beta}^0$. Now stack the T equations of (1) to form

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \tag{3}$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{iT})'$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})'$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})'$.

Remark 1 *For notational simplicity, we do not include individual and time specific effects in the model. But all the discussion below will hold by replacing* $\{y_{it}, \mathbf{x}_{it}'\}$ *with transformed variables* $\{\ddot{y}_{it}, \ddot{\mathbf{x}}_{it}'\}$*, where* $\ddot{y}_{it}^* = (y_{it} - \bar{y}_i - \bar{y}_t + \bar{y})$ *and* $\ddot{\mathbf{x}}_{it} = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i - \bar{\mathbf{x}}_t + \bar{\mathbf{x}})$ *with* $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$*,* $\bar{y}_t = N^{-1}\sum_{i=1}^{N} y_{it}$*,* $\bar{y} = N^{-1}\sum_{i=1}^{N} \bar{y}_i$*, and* $\bar{\mathbf{x}}_i$*,* $\bar{\mathbf{x}}_t$ *and* $\bar{\mathbf{x}}$ *are defined analogously.*

We can rewrite the equations (2) and (3) as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}^0 + \mathbf{u}_i, \qquad \mathbf{u}_i = \mathbf{X}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i. \tag{4}$$

The pooled estimator of $\boldsymbol{\beta}^0$ is given by

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{y}_i. \tag{5}$$

To analyse the asymptotic properties of $\widehat{\boldsymbol{\beta}}$, we extend the assumptions in Hansen (2007) to accommodate random coefficient models as follows:

Assumption A1: $\{\mathbf{x}_{it}', \varepsilon_{it}\}$ is independent across $i = 1, 2, ..., N$ for all t, a strong mixing sequence in t with α of size $-3s/(s-4)$ for $s > 4$, with $\mathbb{E}\left|\varepsilon_{it}\right|^{4+4\delta} \leq \Delta < \infty$, $\mathbb{E}\left|x_{ith}\right|^{8+8\delta} \leq \Delta < \infty$ for all $i, t, h = 1, 2, ..., k$ and $\mathbb{E}\left(\boldsymbol{\varepsilon}_i|\mathbf{X}_i\right) = \mathbf{0}$; $\left\|\boldsymbol{\beta}^0\right\| \leq \Delta$; $\{\boldsymbol{\eta}_i\}$ is independent across $i = 1, 2, ..., N$ and of $\{\boldsymbol{\varepsilon}_i\}$ for all i, $\mathbb{E}\left|\eta_{ih}\right|^{4+4\delta} \leq \Delta < \infty$ and $\mathbb{E}\left(\boldsymbol{\eta}_i|\mathbf{X}_i\right) = \mathbf{0}$.

Assumption A2: (Identification): $\mathbf{A}_{iT} = T^{-1}\mathbb{E}\left(\mathbf{X}_i'\mathbf{X}_i\right)$ is uniformly positive definite and $\mathbf{A} = \lim_{N,T\to\infty} \mathbf{A}_{NT}$, with $\mathbf{A}_{NT} = N^{-1}\sum_{i=1}^{N} \mathbf{A}_{iT}$, is fixed and positive definite.

Assumption A3: (Variance Matrix 1): $\mathbf{B}_{iT} = T^{-1}\mathbb{E}\left(\mathbf{X}_i'\boldsymbol{\Sigma}_{\varepsilon\varepsilon,i}\mathbf{X}_i\right)$ and $\boldsymbol{\Sigma}_{\varepsilon\varepsilon,i} = \mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'|\mathbf{X}_i\right)$ are uniformly positive definite and $\mathbf{B} = \lim_{N,T\to\infty} \mathbf{B}_{NT}$, with $\mathbf{B}_{NT} = N^{-1}\sum_{i=1}^{N} \mathbf{B}_{iT}$, is fixed and positive definite.

Assumption A4: (Variance Matrix 2): $\mathbf{C}_{iT} = T^{-2}\mathbb{E}\left(\mathbf{X}_i'\mathbf{X}_i\boldsymbol{\Omega}_{\eta\eta,i}\mathbf{X}_i'\mathbf{X}_i\right)$ and $\boldsymbol{\Omega}_{\eta\eta,i} = \mathbb{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i'|\mathbf{X}_i\right)$ are uniformly positive definite and $\mathbf{C} = \lim_{N,T\to\infty} \mathbf{C}_{NT}$, with $\mathbf{C}_{NT} = N^{-1}\sum_{i=1}^{N} \mathbf{C}_{iT}$, is fixed and positive definite.

Assumption A1 allows serial dependence in $\{\mathbf{x}_{it}', \varepsilon_{it}\}$ but assumes independence across i. The random coefficient is independent across i. Both the idiosyncratic errors and random coefficient are assumed to be uncorrelated with $\mathbf{x}_{it}$. Assumption A2 is a fairly standard identification condition. Assumption A3 allows conditional heteroskedasticity across i and t. Assumption A4 permits a conditionally heteroskedastic random coefficient process.

For later use, let us define the sample counterpart of $\mathbf{A}_{NT}$ and $\mathbf{A}_{iT}$ defined in Assumption A2:

$$\bar{\mathbf{A}}_{NT} = \frac{1}{N}\sum_{i=1}^{N} \bar{\mathbf{A}}_{iT}, \qquad \bar{\mathbf{A}}_{iT} = \frac{\mathbf{X}_i'\mathbf{X}_i}{T}. \tag{6}$$

Substituting (3) into (5) gives

$$\begin{aligned} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 &= \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{u}_i \\ &= \bar{\mathbf{A}}_{NT}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N} \mathbf{X}_i'\boldsymbol{\varepsilon}_i + \frac{1}{N}\sum_{i=1}^{N} \bar{\mathbf{A}}_{iT}\boldsymbol{\eta}_i\right). \end{aligned} \tag{7}$$

Let us consider the asymptotic properties of the first term of (7). We state the following theorem, which is proven by Hansen (2007):

Theorem 1 *Consider model (4). Under Assumptions A1-A3, as* $(N,T) \to \infty$,

$$\bar{\mathbf{A}}_{NT}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \mathbf{X}_i'\boldsymbol{\varepsilon}_i \xrightarrow{d} N\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right) \tag{8}$$

where $\bar{\mathbf{A}}_{NT}$, $\mathbf{A}$, *and* $\mathbf{B}$ *are defined in (6), Assumptions A2 and A3, respectively.*

This is a very useful result, since, in the absence of slope heterogeneity $\boldsymbol{\eta}_i$, even when the dimension of $\boldsymbol{\Sigma}_{\varepsilon\varepsilon,i} = \mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'|\mathbf{X}_i\right)$ is unbounded as $T \to \infty$ (but $\mu_{\max}\left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon,i}\right) \leq \Delta$ with serially correlated errors), the theorem tells us that the use of the celebrated heteroskedasticity and autocorrelation consistent (HAC) variance estimator of Arellano (1987) for short panel models will be asymptotically justified for large panels.

The next theorem states the asymptotic properties of the second term of (7).

Theorem 2 *Consider model (4). Under Assumptions A1, A2 and A4, as* $(N,T) \to \infty$,

$$\bar{\mathbf{A}}_{NT}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \bar{\mathbf{A}}_{iT} \boldsymbol{\eta}_i \xrightarrow{d} N\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1}\right) \tag{9}$$

where $\bar{\mathbf{A}}_{NT}$ *and* $\bar{\mathbf{A}}_{iT}$ *are defined in (6),* $\mathbf{A}$ *and* $\mathbf{C}$ *are defined in Assumptions A2 and A4, respectively.*

As discussed in Pesaran (2006) and Reese and Westerlund (2018), the pooled estimator $\widehat{\boldsymbol{\beta}}$ is consistent to the centred value $\boldsymbol{\beta}$ under the random coefficient assumption, and the variation of $\widehat{\boldsymbol{\beta}}$ due to the dispersion of slope coefficients dominates the variation due to the linear function of idiosyncratic errors. The following corollary of these two theorems clarify this point:

Corollary 1 *Consider model (4). Under Assumptions A1-A4, as* $(N,T) \to \infty$,

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1}\right) \tag{10}$$

whilst under slope homogeneity, $\boldsymbol{\eta}_i = \mathbf{0}$ *for all* i,

$$\sqrt{NT}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right), \tag{11}$$

where $\widehat{\boldsymbol{\beta}}$ *is defined by (5),* $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$ *are defined in Assumptions A2, A3 and A4, respectively.*

In view of this, Pesaran (2006) proposes to estimate the variance of $\widehat{\boldsymbol{\beta}}$ under random coefficient assumption by

$$\widetilde{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}} = \frac{1}{N} \bar{\mathbf{A}}_{NT}^{-1} \ddot{\mathbf{C}}_{NT} \bar{\mathbf{A}}_{NT}^{-1}, \tag{12}$$

where

$$\ddot{\mathbf{C}}_{NT} = \frac{1}{N} \sum_{i=1}^{N} \bar{\mathbf{A}}_{iT} \left(\widehat{\boldsymbol{\beta}}_i - \overline{\widehat{\boldsymbol{\beta}}}\right) \left(\widehat{\boldsymbol{\beta}}_i - \overline{\widehat{\boldsymbol{\beta}}}\right)' \bar{\mathbf{A}}_{iT}, \tag{13}$$

$\widehat{\boldsymbol{\beta}}_i = \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1} \mathbf{X}_i'\mathbf{y}_i$ and $\overline{\widehat{\boldsymbol{\beta}}} = N^{-1} \sum_{i=1}^{N} \widehat{\boldsymbol{\beta}}_i$. The idea is to approximate the unobserved slope heterogeneity $\boldsymbol{\eta}_i$ by its sample counterparts, $\widehat{\boldsymbol{\beta}}_i - \overline{\widehat{\boldsymbol{\beta}}}$. The empirical evidence has proven that this estimator works well in finite samples.[3] However, there are some issues with this variance estimator for our robust approach. First, because it is different from the HAC variance estimator assuming slope homogeneity, at the choice the practitioner would like to know if there is slope heterogeneity or not. Second, some estimation methods, such as Bai's (2009) estimator, do not permit slope heterogeneity models and computation of statistics involving individual slope estimates might not be justified[4]. In practice we do not necessarily have a priori information on whether slopes are homogeneous or heterogeneous, which may make the choice of the variance estimator subject to uncertainty.[5]

[3] See experimental results in Pesaran (2006), for example.

[4] In Section 3, we demonstrate that Bai's estimator continues to be consistent even for heterogenous model if the regressor has a factor structure.

[5] Pesaran and Yamagata (2008) and Su and Chen (2013), for example, propose slope homogeneity tests, which can guide such a choice.

We propose a simple robust approach against such a choice. Based on the above discussion, under slope heterogeneity we have

$$\begin{aligned}\frac{1}{NT^2}\sum_{i=1}^{N}\mathbb{E}\left(\mathbf{X}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{X}_i\right) &= \frac{1}{NT^2}\sum_{i=1}^{N}\mathbb{E}\left(\mathbf{X}_i'\mathbf{X}_i\boldsymbol{\Omega}_{\eta\eta,i}\mathbf{X}_i'\mathbf{X}_i\right)+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbb{E}\left(\mathbf{X}_i'\boldsymbol{\Sigma}_{\varepsilon\varepsilon i}\mathbf{X}_i\right)\\ &= \frac{1}{NT^2}\sum_{i=1}^{N}\mathbb{E}\left(\mathbf{X}_i'\mathbf{X}_i\boldsymbol{\Omega}_{\eta\eta,i}\mathbf{X}_i'\mathbf{X}_i\right)+O\left(T^{-1}\right). \qquad (14)\end{aligned}$$

This suggests a new alternative estimator of $\mathbf{C}$:

$$\widehat{\mathbf{C}}_{NT}=\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{C}}_{iT}, \qquad \widehat{\mathbf{C}}_{iT}=\frac{\mathbf{X}_i'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{X}_i}{T^2}, \qquad (15)$$

where $\widehat{\mathbf{u}}_i=\mathbf{y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}}$.

Under homogeneous slopes ($\boldsymbol{\eta}_i=\mathbf{0}$ for all i), $\frac{1}{NT}\sum_{i=1}^{N}\mathbb{E}\left(\mathbf{X}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{X}_i\right)=\frac{1}{NT}\sum_{i=1}^{N}\mathbb{E}\left(\mathbf{X}_i'\boldsymbol{\Sigma}_{\varepsilon\varepsilon i}\mathbf{X}_i\right)$ as $\mathbf{u}_i=\boldsymbol{\varepsilon}_i$, hence, following Hansen (2007), we propose the following estimator of **B:**

$$\widehat{\mathbf{B}}_{NT}=\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{B}}_{iT}, \qquad \widehat{\mathbf{B}}_{iT}=\frac{\mathbf{X}_i'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{X}_i}{T}. \qquad (16)$$

We summarise the asymptotic properties of the estimators $\widehat{\mathbf{C}}_{NT}$ and $\widehat{\mathbf{B}}_{NT}$ in the following proposition:[6]

Proposition 1 *Consider the model (3) and the pooled estimator $\widehat{\boldsymbol{\beta}}$, which is defined by (5). Under Assumptions A1-A4, under slope heterogeneity $\widehat{\mathbf{C}}_{NT}\xrightarrow{p}\mathbf{C}$, whilst under slope homogeneity ($\boldsymbol{\eta}_i=\mathbf{0}$ for all i) $\widehat{\mathbf{B}}_{NT}\xrightarrow{p}\mathbf{B}$, as $(N,T)\to\infty$, where $\widehat{\mathbf{u}}_i=\mathbf{y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}}$, $\widehat{\mathbf{C}}_{NT}$ and $\widehat{\mathbf{B}}_{NT}$ are defined by (15) and (16), and $\mathbf{C}$ and $\mathbf{B}$ are defined in Assumptions A3 and A4.*

This proposition implies that the use of a widely employed HAC variance estimator for short panel data models,

$$\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}=\left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\left[\sum_{i=1}^{N}\mathbf{X}_i'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{X}_i\right]\left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1}, \qquad (17)$$

is asymptotically justified for large panel data models under both slope homogeneity and slope heterogeneity.

When there is strong evidence that coefficients are heterogeneous, an alternative pooled estimator, such as a mean group estimator, may be preferred. In this paper we are more in line with the robust approach, which is widely employed in the literature - avoiding uncertainty in specifying and estimating 'nuisance' parameters for potential efficiency gain. As will be discussed in the next section, this approach turns out to be useful for some popular estimation methods, in particular, estimation of linear panel data models with unobserved interactive effects.

The following theorem formally demonstrate the validity of the Wald test based on the proposed robust variance estimator of $\widehat{\boldsymbol{\beta}}$.

[6] The proof of the consistency of $\widehat{\mathbf{B}}_{NT}$ is given by Hansen (2007).

Theorem 3 *Consider testing q linearly independent restrictions of $\boldsymbol{\beta}^0$, $H_0 : \boldsymbol{\mathcal{R}}\boldsymbol{\beta}^0 = \boldsymbol{r}$ against $H_1 : \boldsymbol{\mathcal{R}}\boldsymbol{\beta}^0 \neq \boldsymbol{r}$, where $\boldsymbol{\mathcal{R}}$ is a $q \times k$ fixed matrix of full row rank. Consider the model (3) and the Wald test statistic*

$$W_{NT} = \left(\boldsymbol{\mathcal{R}}\widehat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left\{\boldsymbol{\mathcal{R}}\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{\mathcal{R}}'\right\}^{-1} \left(\boldsymbol{\mathcal{R}}\widehat{\boldsymbol{\beta}} - \mathbf{r}\right), \tag{18}$$

where $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}$ are defined by (5) and (17), respectively. Suppose that Assumptions A1-A4 hold. Then, under the H_0, for both heterogenous slopes and homogeneous slopes ($\boldsymbol{\eta}_i = \mathbf{0}$ for all i), $W_{NT} \xrightarrow{d} \chi^2_q$, as $(N, T) \to \infty$.

Note that in view of (10), (11), (15) and (16), the rate of convergences of homogeneous and heterogeneous models are different. Such a difference is automatically adjusted in (18).

In this section, we have considered a simple random coefficient panel regression model and showed that valid inference is possible by estimating homogenous panel models even if the true model is homogenous panel model. In the next section, we extend the model to include unobserved factors in the residual, and demonstrate that a robust inference such as Theorem 3 is possible even for such a general model.

3 Panel data models with interactive effects

We consider the following heterogeneous coefficients panel data models

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\boldsymbol{\beta}_i + \boldsymbol{\lambda}_i^{0\prime}\mathbf{f}_t^0 + \varepsilon_{it}, \qquad (i = 1, 2, \ldots, N; t = 1, 2, \ldots, T), &(19)\\ \mathbf{x}_{it} &= \boldsymbol{\Gamma}_i^{0\prime}\boldsymbol{g}_t^0 + \mathbf{v}_{it}, &(20)\end{aligned}$$

where $\mathbf{x}_{it}$ is a $k \times 1$ vector of regressors, $\mathbf{f}_t^0$ and $\boldsymbol{g}_t^0$ denote $r_1 \times 1$ and $r_2 \times 1$ vectors of latent factors, respectively. Correspondingly, their factor loadings are $\boldsymbol{\lambda}_i^0$ and $\boldsymbol{\Gamma}_i^0 = (\gamma_{1i}^0, ..., \gamma_{ki}^0)$. Without loss of generality, we assume that $\mathbf{f}_t$ and $\boldsymbol{g}_t$ are different factors. ε_{it} and $\mathbf{v}_{it}$ are the idiosyncratic disturbance terms.

If we stack the equation (19) and (20) over t, we have

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{F}^0\boldsymbol{\lambda}_i^0 + \boldsymbol{\varepsilon}_i, &(21)\\ \mathbf{X}_i &= \mathbf{G}^0\boldsymbol{\Gamma}_i^0 + \mathbf{V}_i &(22)\end{aligned}$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \cdots, \mathbf{x}_{iT})'$, $\mathbf{F}^0 = (\mathbf{f}_1^0, \cdots, \mathbf{f}_T^0)'$, $\mathbf{G}^0 = (\boldsymbol{g}_1^0, \cdots, \boldsymbol{g}_T^0)'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \cdots, \varepsilon_{iT})'$ and $\mathbf{V}_i = (\mathbf{v}_{i1}, \ldots, \mathbf{v}_{iT})'$. If $\boldsymbol{\beta}_i$ has the form $\boldsymbol{\beta}_i = \boldsymbol{\beta}^0 + \boldsymbol{\eta}_i$ as in the previous section, we can rewrite the heterogeneous model in terms of homogenous model as follows:

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta}^0 + \mathbf{H}^0\boldsymbol{\phi}_i^0 + \mathbf{e}_i, &(23)\\ \mathbf{e}_i &= \mathbf{V}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \end{aligned}$$

where $\mathbf{H}^0 = (\mathbf{G}^0, \mathbf{F}^0)$ and $\boldsymbol{\phi}_i^0 = (\boldsymbol{\eta}_i'\boldsymbol{\Gamma}_i^{0\prime}, \boldsymbol{\lambda}_i^{0\prime})'$ with $r = r_1 + r_2$. For later use, define $\mathbf{u}_i = \mathbf{H}^0\boldsymbol{\phi}_i^0 + \mathbf{e}_i$. Note that this form incorporates the homogeneous panel data models ($\boldsymbol{\eta}_i = \mathbf{0}$), where $\mathbf{H}^0 = \mathbf{F}^0$, $\boldsymbol{\phi}_i^0 = \boldsymbol{\lambda}_i^0$ and $\mathbf{e}_i = \boldsymbol{\varepsilon}_i$. Thus, by setting the definitions of $\mathbf{H}^0$, $\boldsymbol{\phi}_i^0$ and $\mathbf{e}_i$, (23) gives a unified representations for the heterogeneous slope model and homogeneous slope model.

When $\mathbf{H}^0$ is unobserved, it should be replaced with a suitable estimator, and in this case a further careful analysis is required. In particular, using estimated variables will result in some asymptotic biases in the pooled estimator, as discussed in Pesaran (2006), Bai (2009) and Westerlund and Urbain (2015), among others. Here we consider Bai's (2009) estimator.[7] Our theoretical contributions to this strand of literature are: (i) establishing the consistency of a bias-corrected estimator both under homogeneous and heterogeneous slopes; (ii) showing the limit distribution of the Wald test statistic based on the HAC variance estimator both under homogeneous and heterogeneous slopes, and; (iii) proposing a new test for correlation and dependence of the random coefficients with the regressors (in the next section).

Remark 4 *One of the important results in this paper is that if the regressors have a factor structure as in (20), Bai's (2009) estimator continues to be consistent and a valid inference can be conducted even for heterogenous slope models with our approach. Therefore, by assuming a factor structure in the regressor, Bai's (2009) estimator becomes robust to slope heterogeneity. This robustness property is the added value by assuming a factor structure in the regressor.*[8]

We now introduce the Bai's estimator. The least squares objective function is defined as

$$\mathrm{SSR}(\boldsymbol{\beta}, \mathbf{H}, \{\boldsymbol{\phi}_i\}_{i=1}^N) = \frac{1}{NT}\sum_{i=1}^N (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{H}\boldsymbol{\phi}_i)'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{H}\boldsymbol{\phi}_i)$$

subject to the constraints $\mathbf{H}'\mathbf{H}/T = \mathbf{I}_r$ and $\sum_{i=1}^N \boldsymbol{\phi}_i\boldsymbol{\phi}_i'$ being diagonal.

The least squares estimator for $(\boldsymbol{\beta}, \mathbf{H})$ denoted by $\left(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{H}}\right)$ is the solution of the following nonlinear equations:

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= \left(\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{y}_i\right) \\ \left[\frac{1}{NT}\sum_{i=1}^N \left(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}\right)\left(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}\right)'\right]\widehat{\mathbf{H}} &= \widehat{\mathbf{H}}\mathcal{V}_{NT} \end{aligned} \tag{24}$$

where $\mathcal{V}_{NT}$ is a diagonal matrix that contains the r largest eigenvalues of the above matrix in the brackets in decreasing order. Given $\left(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{H}}\right)$, we can estimate $\boldsymbol{\phi}_i$ by

$$\widehat{\boldsymbol{\phi}}_i = \frac{1}{T}\widehat{\mathbf{H}}'(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}).$$

We impose the following assumptions. They are basically similar to Bai (2009).

Assumption B1 (idiosyncratic error in y): (i) ε_{it} distributes independently across i; (ii) $\mathbb{E}\varepsilon_{it} = 0$ and $\mathbb{E}|\varepsilon_{it}|^{8+\delta} \leq \Delta$; (iii) $T^{-1}\sum_{s=1}^T\sum_{t=1}^T \mathbb{E}|\varepsilon_{is}\varepsilon_{it}|^{1+\delta} \leq \Delta$; (iv) $\mathbb{E}|N^{-1/2}\sum_{i=1}^N [\varepsilon_{is}\varepsilon_{it} - \mathbb{E}(\varepsilon_{is}\varepsilon_{it})]|^4 \leq \Delta$ for each (s,t); (v) $N^{-1}T^{-2}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T\sum_{r=1}^T\sum_{w=1}^T |\mathrm{cov}(\varepsilon_{is}\varepsilon_{it}, \varepsilon_{ir}\varepsilon_{iw})| \leq \Delta$; (vi) $\boldsymbol{\Omega}_{\varepsilon,i} = \mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')$ is positive definite and its largest eigenvalue is bounded, uniformly every i and T.

Assumption B2 (idiosyncratic error in x): Let $v_{\ell it}$ be the ℓ-th element of $\mathbf{v}_{it}$ and $\mathbf{v}_{\ell,i} = (v_{\ell i1}, \cdots, v_{\ell iT})'$. Then we assume that (i) $\mathbf{v}_{it}$ is independently distributed across i and group-wise independent from $\{\varepsilon_{js}\}$ for $1 \leq j \leq N$ and $1 \leq s \leq T$; (ii) $\mathbb{E}\mathbf{v}_{it} = 0$ and $\mathbb{E}\|\mathbf{v}_{it}\|^{8+\delta} \leq \Delta$; (iii)

[7] Indeed, the results of this paper hold for the PC estimator due to Westerlund and Urbain (2015). The proof is available upon request from the authors.

[8] The model for the regressor specified in (20) can be seen as slightly restrictive comparing to that in Bai (2009), in which no factor structure is imposed. Nonetheless, the process (20) has been widely accepted in the literature including Pesaran (2006), Bai and Li (2014), Westerlund and Urbain (2015), among many others.

$T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}\left\|\mathbf{v}_{is}\mathbf{v}_{it}'\right\|^{1+\delta}\leq\Delta$; (iv) $\mathbb{E}|N^{-1/2}\sum_{i=1}^{N}\left[v_{\ell is}v_{\ell it}-\mathbb{E}(v_{\ell is}v_{\ell it})\right]|^{4}\leq\Delta$ for every ℓ, s and t; In addition, we assume that (v) $N^{-1}T^{-2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{w=1}^{T}|\mathrm{cov}(v_{\ell is}v_{\ell it},v_{\ell ir}v_{\ell iw})|\leq\Delta$; (vi) the largest eigenvalue of $\mathbb{E}(\mathbf{v}_{\ell,i}\mathbf{v}_{\ell,i}')$ is bounded uniformly for every ℓ, i and T.

Assumption B3 (random coefficient): $\boldsymbol{\eta}_i$ is independent across i, and is group-wise independent of ε_{it}, $\mathbf{v}_{it}$, $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Gamma}_i$; $\mathbb{E}\boldsymbol{\eta}_i=\mathbf{0}$ and $\mathbb{E}\|\boldsymbol{\eta}_i\|^4\leq\Delta$; $\mathbb{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}_i')=\bar{\boldsymbol{\Omega}}_{\eta,i}$ is a fixed positive definite matrix uniformly over i.

Assumption B4 (factor components): (i) $\mathbb{E}\|\mathbf{f}_t^0\|^4\leq\Delta$ and $T^{-1}\sum_{t=1}^{T}\mathbf{f}_t^0\mathbf{f}_t^{0\prime}\xrightarrow{p}\boldsymbol{\Sigma}_f$ as $T\to\infty$ which is a fixed positive definite matrix, $\mathbf{f}_t^0$ is group-wise independent from $\mathbf{v}_{it}$ and ε_{it}; (ii) $\mathbb{E}\|\boldsymbol{g}_t^0\|^4\leq\Delta$ and $T^{-1}\sum_{t=1}^{T}\boldsymbol{g}_t^0\boldsymbol{g}_t^{0\prime}\xrightarrow{p}\boldsymbol{\Sigma}_g$ as $T\to\infty$ which is a fixed positive definite matrix, $\boldsymbol{g}_t^0$ is group-wise independent from $\mathbf{v}_{it}$ and ε_{it}; (iii) Let $\mathbf{l}_t=(\mathbf{f}_t^{0\prime},\boldsymbol{g}_t^{0\prime})'$. We assume $T^{-1}\sum_{t=1}^{T}\mathbf{l}_t\mathbf{l}_t'\xrightarrow{p}\boldsymbol{\Sigma}_l$, which is positive definite. (iv) $\mathbb{E}\|\boldsymbol{\Gamma}_i^0\|^4\leq\Delta$, $\mathbb{E}\|\boldsymbol{\lambda}_i^0\|^4\leq\Delta$, $N^{-1}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i^0\boldsymbol{\Gamma}_i^{0\prime}\xrightarrow{p}\boldsymbol{\Sigma}_\Gamma$, $N^{-1}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i^0\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{\Gamma}_i^{0\prime}\xrightarrow{p}\boldsymbol{\Sigma}_{\gamma\eta}$ and $N^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i^0\boldsymbol{\lambda}_i^{0\prime}\xrightarrow{p}\boldsymbol{\Sigma}_\lambda$, where $\boldsymbol{\Sigma}_\Gamma$, $\boldsymbol{\Sigma}_{\gamma\eta}$ and $\boldsymbol{\Sigma}_\lambda$ are positive definite matrices. $\boldsymbol{\Gamma}_i^0$ and $\boldsymbol{\lambda}_i^0$ are group-wise independent from $\mathbf{v}_{it}$ and ε_{it}.

Assumption B5 (identification and Variance Matrices): Define $\mathbf{Z}_i=\mathbf{X}_i-N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\phi}_i^{0\prime}\boldsymbol{\Upsilon}_{\phi^0}^{-1}\boldsymbol{\phi}_\ell^0\mathbf{X}_\ell$ with $\boldsymbol{\Upsilon}_{\phi^0}=N^{-1}\sum_{i=1}^{N}\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i^{0\prime}$. Then, we assume that the four matrices

$$\begin{aligned}\mathbf{A}_0&=\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\mathbb{E}(\mathbf{V}_i'\mathbf{V}_i),\qquad \mathbf{A}_1=\operatorname*{plim}_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}^0}\mathbf{Z}_i,\\ \mathbf{C}_0&=\lim_{N,T\to\infty}\frac{1}{NT^2}\sum_{i=1}^{N}\mathbb{E}(\mathbf{V}_i'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{V}_i),\qquad \mathbf{B}_1=\operatorname*{plim}_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}^0}\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')\mathbf{M}_{\mathbf{H}^0}\mathbf{Z}_i\end{aligned}$$

are fixed and positive definite.

Idiosyncratic errors ε_{it} and $\mathbf{v}_{it}$ are independent groups of each other, independent over i, but allowed to be serially correlated as structured by Assumptions B1 and B2. Assumption B3 implies there are r factors, and the factor loadings $\boldsymbol{\Gamma}_i^0$ and $\boldsymbol{\lambda}_i^0$ have mean zero without loss of generality and are allowed to be correlated with each other. Assumption B4 implies that the random coefficients can be heteroskedastic but should be independent of all other cross-section varying variables.

Under B1-B5, the consistency of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{H}}$ for homogenous slope model can be proved in the same way as Bai (2009). The following proposition is the consistency results for heterogeneous slope model.

Proposition 5 *When the slope in the model is heterogeneous and given Assumptions B1-B5, as $N,T\to\infty$, we have*

$$\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\xrightarrow{p}\mathbf{0}$$

and let $\mathbf{R}=(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)(T^{-1}\mathbf{H}^{0\prime}\widehat{\mathbf{H}})\mathbf{V}_{NT}^{-1}$ where $\boldsymbol{\Phi}=(\boldsymbol{\phi}_1,...,\boldsymbol{\phi}_N)'$. Then $\mathbf{R}$ and $\mathbf{R}^{-1}$ both are $r\times r$ invertible matrices and $O_p(1)$, and

$$\frac{1}{T}\|\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R}\|^2=O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)+O_p(\delta_{NT}^{-2}).$$

Given consistency, we can derive the rate of convergence as follows:

Theorem 6 *Let Assumptions B1-B5 hold. (a)When the model's slopes are heterogeneous, we have*

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = O_p(N^{-1/2}) + O_p(\delta_{NT}^{-2})$$

and

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{V}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{V}_i\boldsymbol{\eta}_i + O_p(\delta_{NT}^{-2}).$$

(b) When the model's slope is homogeneous, we have

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}^0}\mathbf{Z}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i + \frac{1}{N}\boldsymbol{\xi}_{NT} + \frac{1}{T}\boldsymbol{\zeta}_{NT} + O_p(\delta_{NT}^{-3})$$

where the bias terms are given by

$$\boldsymbol{\xi}_{NT} = -\left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{M}_{\mathbf{H}^0}\mathbf{Z}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{Z}_i'\mathbf{H}^0(\mathbf{H}^{0\prime}\mathbf{H}^0)^{-1}\boldsymbol{\Upsilon}_{\phi^0}^{-1}\boldsymbol{\phi}_i^0\mathbb{E}(e_{it}^2),$$

$$\boldsymbol{\zeta}_{NT} = -\left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{M}_{\mathbf{H}^0}\mathbf{Z}_i\right)^{-1}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{X}_i\mathbf{M}_{\mathbf{H}^0}\mathbb{E}(\mathbf{e}_j\mathbf{e}_j')\mathbf{H}^0(\mathbf{H}^{0\prime}\mathbf{H}^0)^{-1}\boldsymbol{\Upsilon}_{\phi^0}^{-1}\boldsymbol{\phi}_i^0.$$

To derive the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$, following Bai (2009), we impose the following assumption:

Assumption B6 (Central Limit Theorem):

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i \xrightarrow{d} N\left(\mathbf{0}, \mathbf{B}_1\right).$$

Corollary 7 *Assume Assumptions B1-B6 hold and $T/N \to \rho \in (0, \Delta]$. (a)When the model's slope is heterogeneous, we have*

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \xrightarrow{d} N(\mathbf{0}, \mathbf{A}_0^{-1}\mathbf{C}_0\mathbf{A}_0^{-1})$$

(b)When the model is homogeneous, we have

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \xrightarrow{d} N(\rho^{1/2}\boldsymbol{\xi}_0 + \rho^{-1/2}\boldsymbol{\zeta}_0, \mathbf{A}_1^{-1}\mathbf{B}_1\mathbf{A}_1^{-1})$$

where $\boldsymbol{\xi}_0$ and $\boldsymbol{\zeta}_0$ are the probability limit of $\boldsymbol{\xi}_{NT}$ and $\boldsymbol{\zeta}_{NT}$, respectively.

From this result, it is found that the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ for homogenous slope model is biased. Therefore, to make a valid inference, we need to remove the bias of the asymptotic distribution. For this, we use the alternative expression given in the following corollary.

Corollary 8 *When the model's slopes are heterogeneous and given Assumptions B1-B5 hold. We have*

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}^0}\mathbf{Z}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}^0}\mathbf{e}_i + O_p(\delta_{NT}^{-2})$$

when the model's slope is homogeneous, we have

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = \left(\frac{1}{NT} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{M}_{\mathbf{H}^0} \mathbf{Z}_i \right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{M}_{\mathbf{H}^0} \mathbf{e}_i + \frac{1}{N} \boldsymbol{\xi}_{NT} + \frac{1}{T} \boldsymbol{\zeta}_{NT} + O_p(\delta_{NT}^{-3}).$$

To derive asymptotically unbiased estimator of $\boldsymbol{\beta}^0$ in the case of homogeneous slope, we consider to estimate the bias terms. Let $\widehat{\mathbf{Z}}_i = \mathbf{X}_i - N^{-1} \sum_{\ell=1}^{N} \widehat{\boldsymbol{\phi}}_i' \widehat{\boldsymbol{\Upsilon}}_\phi^{-1} \widehat{\boldsymbol{\phi}}_\ell \mathbf{X}_\ell$ with $\widehat{\boldsymbol{\Upsilon}}_\phi = N^{-1} \sum_{i=1}^{N} \widehat{\boldsymbol{\phi}}_i \widehat{\boldsymbol{\phi}}_i'$, and $\widehat{\boldsymbol{\Omega}} = \mathrm{diag}(N^{-1} \sum_{j=1}^{N} \widehat{e}_{j1}^2, \cdots, N^{-1} \sum_{j=1}^{N} \widehat{e}_{jT}^2)$. Define

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - \frac{1}{N} \widehat{\boldsymbol{\xi}}_{NT} - \frac{1}{T} \widehat{\boldsymbol{\zeta}}_{NT} \tag{25}$$

where

$$\widehat{\boldsymbol{\xi}}_{NT} = -\left(\frac{1}{NT} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i \mathbf{M}_{\widehat{\mathbf{H}}} \widehat{\mathbf{Z}}_i' \right)^{-1} \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\mathbf{Z}}_i' \widehat{\mathbf{H}} \widehat{\boldsymbol{\Upsilon}}_\phi^{-1} \widehat{\boldsymbol{\phi}}_i \widehat{e}_{it}^2$$

$$\widehat{\boldsymbol{\zeta}}_{NT} = -\left(\frac{1}{NT} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i \mathbf{M}_{\widehat{\mathbf{H}}} \widehat{\mathbf{Z}}_i' \right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{H}}} \widehat{\boldsymbol{\Omega}} \widehat{\mathbf{H}} \widehat{\boldsymbol{\Upsilon}}_\phi^{-1} \widehat{\boldsymbol{\phi}}_i$$

or

$$\widetilde{\boldsymbol{\zeta}}_{NT} = -\left(\frac{1}{NT} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i \mathbf{M}_{\widehat{\mathbf{H}}} \widehat{\mathbf{Z}}_i' \right) \frac{1}{N} \sum_{i=1}^{N} \frac{\widehat{\mathbf{X}_i' \mathbf{M}_{\mathbf{H}} \boldsymbol{\Omega} \mathbf{H}}}{T} \widehat{\boldsymbol{\Upsilon}}_\phi^{-1} \widehat{\boldsymbol{\phi}}_i.$$

with

$$\frac{\widehat{\mathbf{X}_i' \mathbf{M}_{\mathbf{H}} \boldsymbol{\Omega} \mathbf{H}}}{T} = \frac{1}{TN} \sum_{j=1}^{N} \left[\sum_{t=1}^{T} \widehat{u}_{jt}^2 \widehat{\mathbf{x}}_{it} \widehat{\mathbf{h}}_t' + \sum_{s=1}^{S} \sum_{t=s+1}^{T} \left(1 - \frac{s}{S+1}\right) \widehat{u}_{jt} \widehat{u}_{j,t-s} \left(\widehat{\mathbf{x}}_{it} \widehat{\mathbf{h}}_{t-s}' + \widehat{\mathbf{x}}_{i,t-s} \widehat{\mathbf{h}}_t' \right) \right].$$

Note that $\widehat{\boldsymbol{\zeta}}_{NT}$ is a consistent estimator for $\boldsymbol{\zeta}_{NT}$ for serially uncorrelated case while $\widetilde{\boldsymbol{\zeta}}_{NT}$ is a consistent estimator for serially correlated case. In the Monte Carlo simulation in the next section, we use $\widetilde{\boldsymbol{\zeta}}_{NT}$ with $S = \lfloor T^{1/4} \rfloor$.

The stochastic representation of $\widehat{\boldsymbol{\beta}}$ is given by the following theorem.

Theorem 9 *Given Assumptions B1-B5 hold. In addition,* $\mathbb{E}(\varepsilon_{it}^2) = \sigma_{it}^2$ *and* $E(\varepsilon_{it}\varepsilon_{js}) = 0$ *for* $i \neq j$ *or* $t \neq s$. *When the model's slopes are heterogeneous, we have*

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 &= O_p(N^{-1/2}) + O_p(\delta_{NT}^{-2}) \\
&= \left(\frac{1}{NT} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{M}_{\mathbf{H}^0} \mathbf{Z}_i \right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M}_{\mathbf{H}^0} \mathbf{V}_i \boldsymbol{\eta}_i + O_p(\delta_{NT}^{-2}) \\
&= \left(\frac{1}{NT} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{M}_{\mathbf{H}^0} \mathbf{Z}_i \right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{M}_{\mathbf{H}^0} \mathbf{u}_i + O_p(\delta_{NT}^{-2})
\end{aligned}$$

When the model's slope is homogeneous, we have

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 &= \left(\frac{1}{NT} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{M}_{\mathbf{H}^0} \mathbf{Z}_i \right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{M}_{\mathbf{H}^0} \boldsymbol{\varepsilon}_i + O_p(\delta_{NT}^{-3}) \\
&= \left(\frac{1}{NT} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{M}_{\mathbf{H}^0} \mathbf{Z}_i \right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{M}_{\mathbf{H}^0} \mathbf{u}_i + O_p(\delta_{NT}^{-3})
\end{aligned}$$

as $(N, T) \to \infty$ *and* $T/N \to \rho \in (0, \Delta]$.

The asymptotic distribution of the bias-corrected estimator $\widetilde{\boldsymbol{\beta}}$ is given as follows:

Corollary 10 *Assume Assumptions B1-B6 hold and $T/N \to \rho \in (0, \Delta]$. (a)When the model's slope is heterogeneous, we have*

$$\sqrt{N}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \xrightarrow{d} N(\mathbf{0}, \mathbf{A}_0^{-1}\mathbf{C}_0\mathbf{A}_0^{-1}).$$

(b)When the model is homogeneous, we have

$$\sqrt{NT}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \xrightarrow{d} N(\mathbf{0}, \mathbf{A}_1^{-1}\mathbf{B}_1\mathbf{A}_1^{-1}).$$

This result state that the rate of convergence and asymptotic variance of the $\widetilde{\boldsymbol{\beta}}$ are different for homogeneous and heterogeneous models. However, as in the previous section, we can conduct a valid inference without paying attention to that difference. To introduce a robust Wald test, define

$$\widehat{\boldsymbol{\Sigma}}_{\widetilde{\beta}} = \left(\sum_{i=1}^{N} \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\right)^{-1} \left(\sum_{i=1}^{N} \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\right) \left(\sum_{i=1}^{N} \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\right)^{-1} \tag{26}$$

where $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\widetilde{\boldsymbol{\beta}}$ and $\widehat{\mathbf{Z}}_i = \mathbf{X}_i - N^{-1}\sum_{\ell=1}^{N} \widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell$ with $\widehat{\boldsymbol{\Upsilon}}_\phi = N^{-1}\sum_{i=1}^{N} \widehat{\boldsymbol{\phi}}_i\widehat{\boldsymbol{\phi}}_i'$.

Theorem 11 *Consider testing q linearly independent restrictions of $\boldsymbol{\beta}$, $H_0 : \boldsymbol{\mathcal{R}}\boldsymbol{\beta} = \mathbf{r}$ against $H_1 : \boldsymbol{\mathcal{R}}\boldsymbol{\beta} \neq \mathbf{r}$, where $\boldsymbol{\mathcal{R}}$ is a $q \times k$ fixed matrix of full row rank. Consider the model (23) and the Wald statistic*

$$\widetilde{W}_{NT} = (\boldsymbol{\mathcal{R}}\widetilde{\boldsymbol{\beta}} - \mathbf{r})'\left(\boldsymbol{\mathcal{R}}\widehat{\boldsymbol{\Sigma}}_{\widetilde{\beta}}\boldsymbol{\mathcal{R}}'\right)^{-1}(\boldsymbol{\mathcal{R}}\widetilde{\boldsymbol{\beta}} - \mathbf{r}) \tag{27}$$

where $\widetilde{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Sigma}}_{\widetilde{\beta}}$ are defined by (25) and (26), respectively. Suppose that Assumptions B1-B6 hold. Then, under the H_0, for both heterogeneous slopes and homogeneous slopes ($\boldsymbol{\eta}_i = 0$ for all i), $\widetilde{W}_{NT} \xrightarrow{d} \chi_q^2$, as $(N,T) \to \infty$ and $T/N \to \rho \in (0, \Delta]$.

Remark 12 *Our approach is also robust against mixtures of homogeneous and heterogeneous slopes.*[9] *To see this, consider the mode without common components and the case in which the k slopes are partitioned in such a way that $k = k_1 + k_2$, without loss of generality, where $\boldsymbol{\beta}_i = \left(\boldsymbol{\beta}_{1i}', \boldsymbol{\beta}_2^{0\prime}\right)'$, $\boldsymbol{\beta}_{1i} = \boldsymbol{\beta}_1^0 + \boldsymbol{\eta}_{1i}$, $\mathbb{E}\left(\boldsymbol{\eta}_{1i}\right) = \mathbf{0}$ and $Var\left(\boldsymbol{\eta}_{1i}\right) = \boldsymbol{\Omega}_{1i}$, with $\boldsymbol{\beta}^0 = \left(\boldsymbol{\beta}_1^{0\prime}, \boldsymbol{\beta}_2^{0\prime}\right)'$.The expansion of the pooled estimator $\widehat{\boldsymbol{\beta}}$ gives*

$$\begin{aligned}
\sqrt{N}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) &= \mathbf{A}_{NT}^{-1}\frac{\sqrt{N}}{NT}\begin{bmatrix} \sum_{i=1}^{N}\mathbf{X}_{1i}'\mathbf{u}_i \\ \sum_{i=1}^{N}\mathbf{X}_{2i}'\mathbf{u}_i \end{bmatrix} \\
&= \mathbf{A}_{NT}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{u}_i = \mathbf{A}_{NT}^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\mathbf{X}_i'\mathbf{X}_{1i}}{T}\right)\boldsymbol{\eta}_{1i} + O_p\left(1/\sqrt{T}\right) \\
&\xrightarrow{d} N\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1}\right)
\end{aligned}$$

where

$$\mathbf{C} = \operatorname*{plim}_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{X}_i'\mathbf{X}_{1i}}{T}\right)\boldsymbol{\eta}_{1i}\boldsymbol{\eta}_{1i}'\left(\frac{\mathbf{X}_{1i}'\mathbf{X}_i}{T}\right).$$

Observe that under assumptions we have made, $\mathbf{C}$ is positive definite. Also note that the convergence rate of ($\widehat{\boldsymbol{\beta}}_1$ and) $\widehat{\boldsymbol{\beta}}_2$ is $\sqrt{N}$, as the variation of $\widehat{\boldsymbol{\beta}}$ is dominated by $\boldsymbol{\eta}_{1i}$.

[9] We do not consider cross-sectional and/or time-series structural breaks in $\boldsymbol{\beta}_i$ which is beyond the scope of this paper.

Now consider a special case in which $\mathbf{X}_{1i}'\mathbf{X}_{2i} = \mathbf{0}$*. Define a scaling diagonal matrix of order* k *as*

$$\mathbf{D} = \left[\begin{array}{cc} \sqrt{N}\mathbf{I}_{k_1} & \mathbf{0} \\ \mathbf{0} & \sqrt{NT}\mathbf{I}_{k_2} \end{array}\right], \tag{28}$$

so that

$$\mathbf{D}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right)'\mathbf{D} = \left(\mathbf{D}\bar{\mathbf{A}}_{NT}^{-1}\mathbf{D}^{-1}\right)\left(\mathbf{D}\sum_{i=1}^{N}\frac{\mathbf{X}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{X}_i}{T^2N^2}\mathbf{D}\right)\left(\mathbf{D}^{-1}\bar{\mathbf{A}}_{NT}^{-1}\mathbf{D}\right). \tag{29}$$

It is easily seen that $\mathbf{D}\bar{\mathbf{A}}_{NT}^{-1}\mathbf{D}^{-1} = \bar{\mathbf{A}}_{NT}^{-1}$ *since* $\bar{\mathbf{A}}_{NT}^{-1}$ *is block diagonal. Recalling that* $\mathbf{u}_i = \mathbf{X}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i$, $\boldsymbol{\eta}_i = (\boldsymbol{\eta}_{1i}', \mathbf{0}')'$ *and* $E(\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_i') = \mathbf{0}$, *the probability limit of the middle term is*

$$\operatorname*{plim}_{N,T\to\infty}\sum_{i=1}^{N}\mathbf{D}\frac{\mathbf{X}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{X}_i}{T^2N^2}\mathbf{D} = \left(\begin{array}{cc} \mathbf{C}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{22} \end{array}\right), \tag{30}$$

where

$$\mathbf{C}_{11} = \operatorname*{plim}_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{X}_{1i}'\mathbf{X}_{1i}}{T}\boldsymbol{\eta}_{1i}\boldsymbol{\eta}_{1i}'\frac{\mathbf{X}_{1i}'\mathbf{X}_{1i}}{T}\right), \tag{31}$$

$$\mathbf{B}_{22} = \operatorname*{plim}_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{X}_{2i}'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{X}_{2i}}{T}. \tag{32}$$

Therefore, the asymptotic normality of $\widehat{\boldsymbol{\beta}}$*, the consistency of the HAC estimator and the asymptotic validity of Wald test hold with mixtures of homogeneous and heterogeneous slopes.*

4 Wald and LM tests for Correlation of Random Coefficients with Covariates

As discussed earlier, the proposed robust approach works for random coefficients. If it is fixed cross-sectionally varying coefficients or correlated random coefficients with $\mathbf{X}_i$, the approach may not work. To see this, consider the model (4) but without factor components. We have $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = \left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\left[\mathbf{X}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i\right]$. If $\mathbb{E}(\boldsymbol{\eta}_i|\mathbf{X}_i) \neq \mathbf{0}$, $\mathbb{E}\left[\mathbf{X}_i'\mathbf{X}_iE(\boldsymbol{\eta}_i|\mathbf{X}_i)\right]$ is not necessarily zero, and in general it renders $\widehat{\boldsymbol{\beta}}$ biased.

In view of this, we propose novel tests for correlation or dependence of random coefficients with covariates, substantially extending the test proposed by Wooledridge (2010; Ch11.7.4). The main distinctions of our tests from Wooldridge's are: (i) we consider the test for large panels whilst he considers for short panels; (ii) our tests are robust against (uncorrelated) random coefficients;[10] (iii) we propose a Lagrange Multiplier test along with a Wald test; (iv) ours permit $\mathbb{E}(\boldsymbol{\eta}_i|\mathbf{X}_i)$ to be a non-linear function of $\mathbf{X}_i$.

More generally, suppose that the random part of the coefficients is modeled as

$$\boldsymbol{\eta}_i = \mathbf{h}(\mathbf{X}_i) - \boldsymbol{\mu}_{\mathbf{h}} + \boldsymbol{\zeta}_i \tag{33}$$

with $\mathbb{E}[\mathbf{h}(\mathbf{X}_i)] = \boldsymbol{\mu}_{\mathbf{h}}$ and $\mathbb{E}(\boldsymbol{\zeta}_i|\mathbf{X}_i) = \mathbf{0}$, where various forms of function of $\mathbf{X}_i$ can be entertained. For the testing purpose, we consider $\mathbf{h}(\mathbf{X}_i) = \boldsymbol{\Xi}_i\boldsymbol{\delta}$ with

$$\boldsymbol{\Xi}_i = \left(\bar{\mathbf{x}}_i^{(1)}, \bar{\mathbf{x}}_i^{(2)}, ..., \bar{\mathbf{x}}_i^{(g)}\right), \tag{34}$$

[10] Wooldridge (2010;p.386) points out that the drawback of his test is that it cannot detect heterogeneity in $\boldsymbol{\beta}_i$ that is uncorrelated with $\bar{\mathbf{x}}_i$. In our robustified test, this becomes the desirable property.

where $\bar{\mathbf{x}}_i^{(g)} = \left(\bar{x}_{i1}^{(g)}, \bar{x}_{i2}^{(g)}, ..., \bar{x}_{ik}^{(g)}\right)'$, $\bar{x}_{ih}^{(g)} = T^{-1}\sum_{t=1}^{T} x_{ith}^{g}$.[11] Note that x_{ith} is the (t,h) element of $\mathbf{X}_i$.[12] Consider an augmented regression

$$\mathbf{y}_i = \mathbf{W}_i\boldsymbol{\theta} + \boldsymbol{\epsilon}_i, \tag{35}$$

where $\mathbf{W}_i = [\mathbf{X}_i, \mathbf{L}_i]$ with

$$\mathbf{L}_i = \mathbf{X}_i\left(\boldsymbol{\Xi}_i - \bar{\boldsymbol{\Xi}}\right), \tag{36}$$

$\bar{\boldsymbol{\Xi}} = N^{-1}\sum_{i=1}^{N}\boldsymbol{\Xi}_i$, $\boldsymbol{\theta} = \left(\boldsymbol{\beta}', \boldsymbol{\delta}'\right)'$, and the associated unrestricted estimator $\widehat{\boldsymbol{\theta}} = \left(\widehat{\widehat{\boldsymbol{\beta}}}', \widehat{\boldsymbol{\delta}}\right)' = \left(\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{W}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{y}_i$. Under the null hypothesis of $H_0 : \boldsymbol{\delta} = \mathbf{0}$ and Assumptions A1-A4, for homogeneous or heterogeneous slopes, Theorem 3 establishes that

$$W_{CRC}^{(g)} = \widehat{\boldsymbol{\delta}}'\widehat{\boldsymbol{\Sigma}}_{\delta\delta}^{-1}\widehat{\boldsymbol{\delta}} \xrightarrow{d} \chi_g^2 \tag{37}$$

as $(N,T) \to \infty$, where $\widehat{\boldsymbol{\Sigma}}_{\delta\delta}$ is defined as the bottom right partition of $\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\theta}}} = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{\beta\beta} & \widehat{\boldsymbol{\Sigma}}_{\beta\delta} \\ \widehat{\boldsymbol{\Sigma}}_{\delta\beta} & \widehat{\boldsymbol{\Sigma}}_{\delta\delta} \end{pmatrix} = \left(\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{W}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{W}_i'\widehat{\boldsymbol{\epsilon}}_i\widehat{\boldsymbol{\epsilon}}_i'\mathbf{W}_i\right)\left(\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{W}_i\right)^{-1}$, $\widehat{\boldsymbol{\epsilon}}_i = \mathbf{y}_i - \mathbf{W}_i\widehat{\boldsymbol{\theta}}$.

For the model with unobserved factors, the test statistic is computed based on $\mathbf{y}_i$ and $\widehat{\mathbf{W}}_i = [\widehat{\mathbf{Z}}_i, \widehat{\mathbf{L}}_i]$ with $\widehat{\mathbf{L}}_i = \widehat{\mathbf{Z}}_i\left(\widehat{\boldsymbol{\Xi}}_i - \widehat{\bar{\boldsymbol{\Xi}}}\right)$ and $\widehat{\boldsymbol{\Xi}}_i = \left(\widehat{\mathbf{x}}_i^{(1)}, \widehat{\mathbf{x}}_i^{(2)}, ..., \widehat{\mathbf{x}}_i^{(g)}\right)$, $\widehat{\mathbf{x}}_i^{(g)} = \left(\widehat{x}_{i1}^{(g)}, \widehat{x}_{i2}^{(g)}, ..., \widehat{x}_{ik}^{(g)}\right)'$, $\widehat{x}_{ih}^{(g)} = T^{-1}\sum_{t=1}^{T}\widehat{x}_{ith}^{g}$, $\widehat{\mathbf{X}}_i = \mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i = \left(\widehat{x}_{ith}^{g}\right)$ and the bias-corrected Bai's estimator $\widetilde{\boldsymbol{\theta}}_{Bai} = \left(\widetilde{\widetilde{\boldsymbol{\beta}}}'_{Bai}, \widetilde{\boldsymbol{\delta}}_{Bai}\right)'$, which is discussed in Section 3.

We also consider the Lagrange Multiplier (LM) or Score test of the correlated random coefficient. One of the advantages of employing the LM test is that, unlike the Wald test, computation of the LM test only requires the estimation results of the null model. The LM test statistic for the model without factors is defined as

$$LM_{CRC}^{(g)} = \left(\sum_{i=1}^{N}\mathbf{L}_i'\widehat{\mathbf{u}}_i\right)'\left(\sum_{i=1}^{N}\mathbf{K}_i'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{K}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{L}_i'\widehat{\mathbf{u}}_i\right) \tag{38}$$

where $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}$ with $\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{y}_i$ and

$$\mathbf{K}_i' = \mathbf{L}_i' - \left(\sum_{i=1}^{N}\mathbf{L}_i'\mathbf{X}_i\right)\left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\mathbf{X}_i'. \tag{39}$$

For the Bai's estimator, the LM test statistic is given by

$$LM_{CRC_{Bai}}^{(g)} = \left(\sum_{i=1}^{N}\widehat{\mathbf{L}}_i'\widetilde{\mathbf{u}}_i\right)'\left(\sum_{i=1}^{N}\widehat{\mathbf{K}}_i'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\widehat{\mathbf{K}}_i\right)^{-1}\left(\sum_{i=1}^{N}\widehat{\mathbf{L}}_i'\widetilde{\mathbf{u}}_i\right) \tag{40}$$

where $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_{Bai}$, $\widetilde{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\widetilde{\boldsymbol{\beta}}_{Bai}$ with $\widetilde{\boldsymbol{\beta}}_{Bai}$ being the bias corrected estimator, and

$$\widehat{\mathbf{K}}_i' = \widehat{\mathbf{L}}_i' - \left(\sum_{i=1}^{N}\widehat{\mathbf{L}}_i'\widehat{\mathbf{Z}}_i\right)\left(\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\widehat{\mathbf{Z}}_i\right)^{-1}\widehat{\mathbf{Z}}_i'. \tag{41}$$

[11] Cross product terms, such as $T^{-1}\sum_{t=1}^{T} x_{ith}^{(g)}x_{itj}^{(f)}$ for $h \neq j$, could be included in $\boldsymbol{\Xi}_i$.

[12] For the model with fixed effects, the test variable $\boldsymbol{\Xi}_i$ should not be based on within-transformed $\mathbf{X}_i$, otherwise $\bar{\mathbf{x}}_i^{(1)} = \mathbf{0}$ for all i.

By the standard discussion of asymptotic equivalence of the LM and Wald tests, it is readily established that under the null hypothesis $LM_{CRC} \xrightarrow{d} \chi_g^2$ as $(N,T) \to \infty$, and $LM_{CRC_{Bai}} \xrightarrow{d} \chi_g^2$ as $(N,T) \to \infty$ such that $N/T \to c \in (0, \Delta]$. It may be sufficient to consider $g = 2$ to approximate the function $\mathbf{g}(\mathbf{X}_i)$ for our testing purpose.

When the test is rejected in favour of alternatives, it is preferable to employ estimators which are consistent when variation of $\boldsymbol{\beta}_i$ is dependent on covariates. For the estimation of the models with observed factors, the mean group estimator proposed by Chamberlain (1982) and Pesaran and Smith (1995) would be possible choices.

5 Monte Carlo Experiments

In this section we investigate the finite sample performance of our robust approach against slope heterogeneity, error serial correlation and heteroskedasticity. We consider the performance of the following estimators: (two-way) fixed effects estimator $\widehat{\boldsymbol{\beta}}_{FE}$, which is the pooled ordinary least square (OLS) estimator of within-transformed and cross-sectionally demeaned variables; Bai's (2009) iterative PC estimator, both bias-non-corrected $\widehat{\boldsymbol{\beta}}_{Bai}$ defined by (24) and the bias-corrected estimator $\widetilde{\boldsymbol{\beta}}_{Bai}$ defined by (25). For simplicity, in all the experiments, we assume that the number of factors r is known.[13]

In particular, we examine bias and root mean square errors (RMSE) of the estimators, and empirical size and power of the (Wald) test for linear restrictions of slope coefficients, as well as the performance of the LM test for correlation and dependence of slope coefficients with covariates.[14]

5.1 Design

Consider the following data generating process:

$$y_{it} = \sum_{h=1}^{k} x_{ith}\beta_{ih} + \sum_{\ell=1}^{r} f_{t\ell}\lambda_{i\ell} + \sigma_{\varepsilon,it}\varepsilon_{it}, \; i = 1, 2, ..., N; t = 1, 2, ..., T \tag{42}$$

where $\lambda_{i\ell} \sim iidN(0,1)$, $f_{t\ell} = \rho_f f_{t-1,\ell} + \sqrt{1-\rho_f^2}\nu_{t\ell}$, $\nu_{t\ell} \sim iidN(0,1)$ with $f_{0,\ell} \sim iidN(0,1)$ for $\ell = 1, .., r$, $\varepsilon_{it} = \rho_\varepsilon \varepsilon_{it-1} + \sqrt{1-\rho_\varepsilon^2}\xi_{it}$, $\xi_{it} \sim iidN(0,1)$ with $\varepsilon_{i0} \sim iidN(0,1)$, and

$$\sigma_{\varepsilon,it} = (\kappa_{\varepsilon,i}\kappa_{\varepsilon,t})^{1/2}, \; \kappa_{\varepsilon,i} \sim iidU(0.5, 1.5) \text{ and } \kappa_{\varepsilon,t} = 0.5 + t/T. \tag{43}$$

The regressors x_{ith}, $h = 1, 2, .., k$, are generated as

$$x_{ith} = \sum_{\ell=1}^{r} f_{t\ell}\gamma_{ih\ell} + \varrho\sigma_{\mathrm{v},it}\mathrm{v}_{ith}, \tag{44}$$

where $\mathrm{v}_{ith} = \rho_{\mathrm{v}}\mathrm{v}_{it-1,h} + \sqrt{1-\rho_{\mathrm{v}}^2}\varpi_{it,h}$. We consider two types of distribution for $\varpi_{it,h}$: (i) $\varpi_{it,h} = \left(\varpi_{it,h}^* - c\right)/\sqrt{2c}$, $\varpi_{it,h}^* \sim iid\chi_c^2$ and $\mathrm{v}_{i0,h} = \left(\mathrm{v}_{i0,h}^* - c\right)/\sqrt{2c}$, $\mathrm{v}_{i0,h}^* \sim iid\chi_c^2$ with $c = 6$, and (ii) $\varpi_{it,h} \sim iidN(0,1)$ with $\mathrm{v}_{i0,h} \sim iidN(0,1)$. The factor loadings in x_{ith} are generated as

$$\gamma_{ih\ell} = 0.7\lambda_{i\ell} + \left(1 - 0.7^2\right)^{1/2}\varphi_{ih\ell}, \tag{45}$$

[13] The Pesaran's (2006) CCE estimator is not considered in our experiments, since, to our knowledge, feasible analytical bias correction for the pooled estimator under slope homogeneity is not available.

[14] The finite sample performance of the Wald version of the correlated random effects test is much worse than the LM test version. Therefore, its summary results are not reported.

$\varphi_{ih\ell} \sim iidN(0,1)$ for $h = 1,..,k$ and $\ell = 1,..,r$, so that they are correlated with factor loadings in y_{it}.

$$\sigma_{\mathrm{v},it} = (\kappa_{\mathrm{v},i}\kappa_{\mathrm{v},t})^{1/2}, \ \kappa_{\mathrm{v},i} \sim iidU(0.5, 1.5) \text{ and } \kappa_{\mathrm{v},t} = 0.5 + t/T, \tag{46}$$

and $\phi^2 = \{2, 3\}$. Finally we have

$$\beta_{ih} = \beta_h + \sigma_\eta \left(\sqrt{1 - \rho_{x\eta}^2}\eta_{ih} + \rho_{x\eta}w_{ih} \right), \tag{47}$$

$\eta_{ih} \sim iidN(0,1)$ for $h = 1,..,k$, and

$$w_{ih} = \frac{1}{\sqrt{q}} \sum_{p=1}^{q} \frac{z_{ih,p} - \bar{z}_{h,p}}{s_{zh,p}}, \tag{48}$$

where $\bar{z}_{h,p} = N^{-1}\sum_{i=1}^{N} z_{ih,p}$, $s_{zh,p}^2 = (N-1)^{-1}\sum_{i=1}^{N}(z_{ih,p} - \bar{z}_{h,p})^2$. We consider $z_{ih,p} = T^{-1}\sum_{t=1}^{T}(x_{ith}^*)^p$.

We set $k = 2$ (two regressors) for all the experiments. We consider two sets of design: the model without factors ($r = 0$) to examine the fixed effects estimator where $\sum_{\ell=1}^{r} f_{t\ell}\lambda_{i\ell}$ is removed from (42), and the model with two factors ($r = 2$) to examine the Bai's estimator. As recommended in Remark 1, before the estimation the data is all within transformed and cross-sectionally demeaned, to make the results invariant to the inclusion of (additive) individual effects and time effects. For parameter values, we set $(\beta_1, \beta_2) = (1, 3)$.

To look into the bias and RMSE of the estimators, and the size and power of the test of linear restrictions for the estimators, we consider the following sets of designs:

(A) homogeneous slopes ($\sigma_\eta = 0$ in (47));

(B) heterogeneous slopes ($\sigma_\eta = 0.2$ in (47)).

In order to see the effects of dependence of β_i with the regressors upon the bias of the estimators and the associated tests, we set $\rho_{x\eta} = 0.5$ in (47). To investigate the effects of the symmetry of the distribution upon the performance of the estimators and the tests, we consider two types of distribution of disturbances in x_{ith} :

(C) $\left(\varpi_{it,h}^* - 6\right)/\sqrt{12}$, $\varpi_{it,h}^* \sim iid\chi_6^2$, with $\rho_{x\eta} = 0.5$

(D) $\varpi_{it,h} \sim iidN(0,1)$, with $\rho_{x\eta} = 0.5$.

For designs (C) and (D), we consider two types of dependence of β_{ih} upon regressors: β_{ih} is a linear function of the following cross-sectionally standardised values: (i) $T^{-1}\sum_{t=1}^{T} x_{ith}$ (i.e., $q = 1$ and $p = 1$ in (48)) and (ii) $T^{-1}\sum_{t=1}^{T} x_{ith}^2$ (i.e., $q = 1$ and $p = 2$ in (48)).

Finally, the size and the power of the LM tests with degrees $g = 1, 2$, are examined as the set **(E)**. The empirical size is obtained using designs (A) and (B), and the empirical power is computed by designs (C) and (D).

We consider all the combinations of $N = 50, 100, 200$ and $T = 25, 50, 100, 200$. Throughout the experiments, we set $\rho_f = 0.5$, $\rho_\varepsilon = 0.5$ and $\rho_\mathrm{v} = 0.5$. To save space, we report the results with $\varrho^2 = 2$ only.[15] All the tests are conducted at the five per cent significance level. All the experimental results are based on 2,000 replications.

5.2 Results

Table 1 summarises the performance of the Fixed Effect estimator for the model of $(\beta_1, \beta_2) = (1, 3)$, with time-series and cross-section heteroskedastic, serially correlated errors in the absence of interactive effects. Panel A reports the bias, the root mean square error (RMSE) of estimates

[15] The results with $\varrho^2 = 3$ are qualitatively very similar to those with $\varrho^2 = 2$, which are available upon request from the authors.

of β_1, and the size of the Wald test for $H_0 : \beta_1 = 1$ and the power for $H_0 : \beta_1 = 0.95$, under homogeneous slopes, and Panel B under heterogeneous random slopes. The results for β_2 are qualitatively similar and not reported. As predicted by the theory, the Wald test based on the HAC variance estimator has correct size both under slope homogeneity and heterogeneity. Panels C&D report the bias of the estimates and the size of the Wald test for $H_0 : \beta_1 = 1$, to see the effects of dependence between random coefficients and regressors. In Panel C the regressors are generated by asymmetric disturbances and in Panel D, they are drawn from symmetric distribution. In Panel C, when $\boldsymbol{\eta}_i$ depends on $\sum_{t=1}^{T} x_{ith}$, the fixed effects estimator exhibits systematic bias, but in Panel D, it does not. This is because when the third moment of x_{ith} is zero, by construction $\mathbb{E}\left[\mathbf{X}_i'\mathbf{X}_i\boldsymbol{\eta}_i\right] = \mathbf{0}$ which makes the estimator unbiased. However, as can be seen in Panel D, the size of the test declines systematically as sample size rises, which suggests that the HAC variance estimates will not be consistent. When $\boldsymbol{\eta}_i$ is a linear function of $\sum_{t=1}^{T} x_{ith}^2$, regardless of the shape of the distribution of regressors, it exhibits serious bias in estimates (see Panels C&D). Therefore, it is of great importance to statistically check the evidence of dependence of $\boldsymbol{\beta}_i$ with regressors. The performance of the proposed LM test for correlation and dependence of random coefficients with regressors is summarised in Panel E. As can be seen, it has correct size with slope homogeneity and random coefficients, and the LM test with $g = 2$ has high power against both types of dependence of $\boldsymbol{\beta}_i$, $\sum_{t=1}^{T} x_{ith}$ and $\sum_{t=1}^{T} x_{ith}^2$, whilst the LM test with $g = 1$ lacks power when $\boldsymbol{\beta}_i$ depends on $\sum_{t=1}^{T} x_{ith}^2$ only. Therefore, it is recommended to employ $g = 2$ in practice.

Let us turn our attention to the estimation of the models with unobservable interactive effects. The relevant results are reported in Table 2. Table2 contains Panels A-E, which correspond to the panels in Table 1. To illustrate the effectiveness of the bias-correction, we report the results both for bias-non-corrected and bias-corrected estimators.

Consider Panel A of Table 1, which deals with the slope homogeneous case. First look at the bias of the estimators. Non-bias-corrected estimator ($\widehat{\beta}_{Bai}$) has very little bias and the magnitude of correction is very small. As reported in Bai (2009), the bias-corrected estimator ($\widetilde{\beta}_{Bai}$) has very small bias and it becomes smaller as N and/or T rise. In terms of RMSE, $\widehat{\beta}_{Bai}$ and $\widetilde{\beta}_{Bai}$ are very similar for all the combinations of (N, T). The size of the Wald test based on $\widehat{\beta}_{Bai}$ and $\widetilde{\beta}_{Bai}$ is close to nominal level for the sample sizes which we consider.

Now let us turn our attention to the random coefficient model, the results of which are summarised in Panel B, Table 2. The magnitude of the bias of the estimators under slope heterogeneity is larger than under slope homogeneity, especially with small N and T, but it gets smaller as N and T increase. As in the homogeneous slope case, the bias of both $\widehat{\beta}_{Bai}$ and $\widetilde{\beta}_{Bai}$ is relatively small. The properties of the results reported in Panels C, D and E are very similar to those commented earlier on the corresponding panels in Table 1.

6 Concluding Remarks

In this paper, we have proposed a robust approach against heteroskedasticity, error serial correlation and slope heterogeneity for large linear panel data models. First, we have established the asymptotic validity of the Wald test based on the panel HAC variance estimator of the pooled estimator under random coefficient models. Then, we have shown that a similar result holds with the bias-corrected Bai's estimator for models with unobserved interactive effects. Our new theoretical result has justified the use of the same slope estimator and the variance estimator, both for slope homogeneous and heterogeneous models. This robust approach can significantly reduce the model selection uncertainty for applied researchers.

In addition, we have proposed a novel test for correlation and dependence of the random coefficient with covariates. The test is of great importance, since the widely used estimators and/or its variance estimators can become inconsistent when the variation of coefficients depends on covariates, in general.

We have examined the finite sample performance of the estimators, tests of linear restrictions, and the LM tests for correlated random coefficients. The evidence illustrates the usefulness of our approach. In particular, for the estimation of the models with unobserved interactive effects, the size of the proposed robust Wald test using the bias-corrected Bai's (2009) estimator is very close to the nominal level, under both slope homogeneity and slope heterogeneity, while maintaining satisfactory power. Also, the LM tests for correlated random coefficients have correct size under both slope homogeneity and slope heterogeneity due to pure random coefficients, while exhibiting high power when the random coefficients depend on covariates. In view of these finite sample performance, the proposed robust approach based on Bai's (2009) estimator is useful in practice.

Recently estimation of panel models with a group structure has been gaining great interest in the literature. Su et al. (2016) propose a method for identifying and estimating latent group structures using so called C-Lasso. Consider:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}_g + u_{it}, \quad u_{it} = \mathbf{f}_t'\boldsymbol{\gamma}_i + \varepsilon_{it},$$

$g = 1, ..., G$. For individual $i \in g$, the slope is given by $\boldsymbol{\beta}_g$. It may be possible to assume heterogeneous slope, mean of which has group structure: $\boldsymbol{\beta}_{i\in g} = \boldsymbol{\beta}_g + \eta_{i\in g}$.

As emphasised in the paper, when the test of correlated random coefficient rejects the null in favour of alternatives, it is preferable to employ estimators which are consistent when variation of slopes is dependent on covariates. The mean group estimator proposed by Chamberlain (1982), Pesaran and Smith (1995) and Pesaran (2006) would be possible choices, however, to our knowledge, no satisfactory inferential methods have been proposed in the literature. Thus, developing such methods will be an important future research theme.

Table 1: Summary results of Fixed Effects estimator for the model with $\{\beta_1, \beta_2\} = \{1, 3\}$, heteroskedastic and serially correlated errors

Panel A: Homogeneous Slopes, $\beta_{ih} = \beta_h$ for all i, $h = 1, 2$												
for β_1	Bias ($\times$100)			RMSE ($\times$100)			Size			Power		
T, N	50	100	200	50	100	200	50	100	200	50	100	200
$\widehat{\beta}_{FE}$												
25	-0.148	-0.115	-0.005	2.572	1.848	1.276	5.5	5.9	4.9	47.4	76.1	96.9
50	-0.077	-0.040	0.008	1.853	1.305	0.920	6.0	5.4	5.2	76.0	96.6	100.0
100	-0.061	-0.015	-0.001	1.372	0.954	0.674	5.9	5.7	5.4	95.5	100.0	100.0
200	-0.017	0.004	0.007	0.955	0.677	0.479	5.4	6.1	5.3	100.0	100.0	100.0
Panel B: Heterogeneous Slopes, $\beta_{ih} = \beta_h + \eta_{ih}$ with $\eta_{ih} \sim iidN\,(0, 0.04)$ for all i, $h = 1, 2$												
for β_1	Bias ($\times$100)			RMSE ($\times$100)			Size			Power		
T, N	50	100	200	50	100	200	50	100	200	50	100	200
$\widehat{\beta}_{FE}$												
25	-0.038	-0.064	0.045	4.122	2.966	2.146	5.6	5.5	5.4	23.0	39.3	66.6
50	-0.015	-0.010	0.057	3.679	2.592	1.869	6.6	5.4	5.1	30.2	50.5	78.6
100	-0.031	0.009	0.039	3.327	2.328	1.661	5.7	4.6	4.8	33.0	57.3	85.2
200	0.025	0.037	0.050	3.129	2.194	1.562	6.1	5.5	4.9	36.5	63.2	90.0

Notes for Panels A and B: Data is generated as $y_{it}^* = x_{it,1}^*\beta_{i1} + x_{it,2}^*\beta_{i2} + \sigma_{\varepsilon,it}\varepsilon_{it}$, $i = 1, ..., N$, $t = 1, ..., T$, $\varepsilon_{it} = \rho_\varepsilon \varepsilon_{it-1} + \sqrt{1-\rho_\varepsilon^2}\xi_{it}$, $\xi_{it} \sim iidN(0,1)$ with $\varepsilon_{i0} \sim iidN(0,1)$, $\sigma_{\varepsilon,it} = (\kappa_{\varepsilon,i}\kappa_{\varepsilon,t})^{1/2}$, $\kappa_{\varepsilon,i} \sim iidU\,(0.5, 1.5)$ and $\kappa_{\varepsilon,t} = 0.5 + t/T$; $x_{it,h}^* = \varrho\sigma_{\mathrm{v},it}\mathrm{v}_{it,h}$, where $\mathrm{v}_{it,h} = \rho_{\mathrm{v}}\mathrm{v}_{it-1,h} + \sqrt{1-\rho_{\mathrm{v}}^2}\varpi_{it,h}$, $\varpi_{it,h} \sim iid\left(\chi_6^2 - 6\right)/\sqrt{12}$ with $\mathrm{v}_{i0,h} \sim iid\left(\chi_6^2 - 6\right)/\sqrt{12}$, $\sigma_{\mathrm{v},it} = (\kappa_{\mathrm{v},i}\kappa_{\mathrm{v},t})^{1/2}$, $\kappa_{\mathrm{v},i} \sim iidU\,(0.5, 1.5)$ and $\kappa_{\mathrm{v},t} = 0.5 + t/T$. We set $\rho_\varepsilon = \rho_{\mathrm{v}} = 0.5$ and $\varrho^2 = 2$. $\widehat{\beta}_{FE}$ is the pooled regression of within-transformed and cross-sectionally demeaned variables. The size is rejection frequency of the proposed Wald test (defined by (18)) for $H_0 : \beta_1 = 1$ and the power for $H_0 : \beta_1 = 0.95$, based on the 5% level test. All results are based on 2000 replications.

Table 1 continued

Panel C: Correlated Heterogeneous Slopes, $\rho_{x\eta} = 0.5$, x^*_{ith} generated using χ^2_6												
	(i) β_{ih} is function of $\sum_t x^*_{ith}$						(ii) β_{ih} is function of $\sum_t (x^*_{ith})^2$					
for β_1	Bias ($\times 100$)			Size			Bias ($\times 100$)			Size		
T, N	50	100	200	50	100	200	50	100	200	50	100	200
$\widehat{\beta}_{FE}$												
25	0.068	0.019	0.147	5.2	5.0	4.4	1.145	1.158	1.273	6.0	6.9	8.5
50	0.107	0.088	0.164	5.7	4.6	3.3	1.208	1.247	1.315	7.8	6.8	9.4
100	0.075	0.099	0.153	5.5	3.6	3.8	1.236	1.317	1.350	7.7	7.6	9.9
200	0.151	0.145	0.188	5.2	4.3	3.2	1.398	1.450	1.473	7.4	8.9	12.7
Panel D: Correlated Heterogeneous Slopes, $\rho_{x\eta} = 0.5$, x^*_{ith} generated using $N(0,1)$												
	(i) β_{ih} is function of $\sum_t x^*_{ith}$						(ii) β_{ih} is function of $\sum_t (x^*_{ith})^2$					
for β_1	Bias ($\times 100$)			Size			Bias ($\times 100$)			Size		
T, N	50	100	200	50	100	200	50	100	200	50	100	200
$\widehat{\beta}_{FE}$												
25	0.046	0.022	0.031	6.6	5.2	4.7	1.213	1.185	1.205	7.6	7.4	7.6
50	-0.045	-0.008	-0.016	6.4	4.6	3.5	1.134	1.168	1.183	6.8	6.2	8.5
100	-0.033	-0.021	-0.015	5.7	4.7	2.9	1.192	1.194	1.222	6.1	7.6	8.7
200	-0.062	-0.049	-0.043	6.0	4.6	3.3	1.235	1.236	1.266	7.1	7.4	9.8

Notes for Panels C and D: The data generating process (DGP) is the same as that for Panel B, except $\beta_{ih} = \beta_h + \sigma_\eta \left(\sqrt{1-\rho^2_{x\eta}}\eta_{ih} + \rho_{x\eta} w_{ih} \right), \eta_{ih} \sim iidN(0,1)$ for $h = 1, 2$, $w_{ih} = \frac{z_{ih,p} - \bar{z}_{h,p}}{s_{zh,p}}$, where $\bar{z}_{h,p} = N^{-1} \sum_{i=1}^N z_{ih,p}$, $s^2_{zh,p} = (N-1)^{-1} \sum_{i=1}^N (z_{ih,p} - \bar{z}_{h,p})^2$, $z_{ih,p} = T^{-1} \sum_{t=1}^T \left(x^*_{it,h} \right)^p$, $p = 1, 2$. The DGP for Panel D is identical to of Panel C, except that $\varpi_{it,h} \sim iidN(0,1)$.

Table 1 continued

Panel E: Size and Power of the LM test of Correlated Random Coefficient																								
	(A) Size			(B) Size			(C) Power, x_{ith} generated using χ^2_6						(D) Power, x_{ith} generated using $N(0,1)$					
	Slope Homo			Slope Hetero			$\sum_t x^*_{ith}$			$\sum_t (x^*_{ith})^2$			$\sum_t x^*_{ith}$			$\sum_t (x^*_{ith})^2$		
T,N	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200
$LM^{(1)}_{CRC}$																		
25	5.1	5.4	4.9	4.9	5.3	5.1	87.1	99.7	100.0	5.1	4.5	4.9	89.2	99.5	100.0	4.8	5.4	5.3
50	5.5	5.6	5.2	4.4	5.2	5.7	96.7	100.0	100.0	5.2	4.4	5.4	96.8	100.0	100.0	4.5	4.4	5.9
100	5.7	4.9	4.9	4.6	4.9	5.7	99.0	100.0	100.0	4.8	4.7	5.8	99.3	100.0	100.0	4.9	4.6	5.8
200	4.2	4.4	5.4	4.5	5.1	4.9	99.7	100.0	100.0	5.0	4.9	5.6	99.8	100.0	100.0	5.7	4.8	5.8
$LM^{(2)}_{CRC}$																		
25	4.1	3.9	4.6	3.7	4.2	4.6	75.6	98.6	100.0	37.4	76.2	98.5	76.9	99.0	100.0	43.3	83.4	98.9
50	3.6	5.5	4.8	3.6	4.4	4.1	90.0	99.8	100.0	57.2	92.6	99.9	91.3	99.9	100.0	61.6	94.5	99.8
100	4.0	4.1	4.3	3.3	4.1	4.9	95.9	100.0	100.0	72.3	96.7	99.9	97.4	100.0	100.0	73.2	97.2	100.0
200	3.4	3.8	4.5	4.1	4.1	5.0	98.4	100.0	100.0	82.8	98.6	100.0	98.9	100.0	100.0	82.6	98.9	100.0

Notes for Panel E: The results of column blocks (A), (B), (C) and (D) are obtained using the same DGPs as Panels A, B, C and D, respectively. $LM^{(g)}_{CRC}$ is the proposed LM tests of correlated random effects defined by (38). The test statistics are referred to the 95% quantile of χ^2_g distribution, $g = 1, 2$. All results are based on 2000 replications.

Table 2: Summary results of Bai's estimators for the model with $\{\beta_1, \beta_2\} = \{1, 3\}$, interactive effects, heteroskedastic and serially correlated errors

Panel A: Homogeneous Slopes, $\beta_{ih} = \beta_h$ for all i, $h = 1, 2$												
for β_1	Bias ($\times$100)			RMSE ($\times$100)			Size			Power		
T, N	50	100	200	50	100	200	50	100	200	50	100	200
$\widehat{\beta}_{Bai}$												
25	0.048	0.015	0.091	2.717	1.932	1.336	6.9	6.5	5.5	50.5	76.2	97.1
50	0.013	0.018	0.048	1.919	1.341	0.942	6.9	6.4	5.6	76.5	96.6	100.0
100	-0.021	0.018	0.021	1.413	0.972	0.681	7.4	6.4	5.6	95.1	99.9	100.0
200	0.007	0.024	0.021	0.989	0.688	0.485	7.3	6.2	5.5	100.0	100.0	100.0
$\widetilde{\beta}_{Bai}$												
25	0.038	0.006	0.082	2.712	1.931	1.334	6.9	6.4	5.5	50.2	76.1	97.1
50	0.003	0.009	0.040	1.918	1.341	0.942	6.8	6.2	5.6	76.3	96.7	100.0
100	-0.031	0.010	0.015	1.415	0.973	0.682	7.3	6.6	5.5	95.1	99.9	100.0
200	-0.008	0.014	0.014	0.996	0.689	0.485	7.1	6.1	5.4	99.9	100.0	100.0
Panel B: Heterogeneous Slopes, $\beta_{ih} = \beta_h + \eta_{ih}$ with $\eta_{ih} \sim iidN\,(0, 0.04)$ for all i, $h = 1, 2$												
for β_1	Bias ($\times$100)			RMSE ($\times$100)			Size			Power		
T, N	50	100	200	50	100	200	50	100	200	50	100	200
$\widehat{\beta}_{Bai}$												
25	0.002	-0.009	0.108	4.228	3.049	2.190	7.8	6.3	6.2	26.2	41.8	67.3
50	-0.095	-0.038	0.052	3.741	2.627	1.882	8.5	6.7	6.1	32.9	52.8	79.0
100	-0.170	-0.050	0.013	3.364	2.344	1.668	7.7	5.8	5.2	36.5	58.8	85.7
200	-0.134	-0.037	0.017	3.152	2.198	1.564	8.1	6.5	5.3	39.8	64.4	90.3
$\widetilde{\beta}_{Bai}$												
25	-0.016	-0.024	0.096	4.227	3.049	2.189	7.7	6.3	6.2	25.9	41.6	66.9
50	-0.116	-0.056	0.038	3.742	2.629	1.882	8.4	6.9	5.9	32.8	52.6	78.8
100	-0.196	-0.070	-0.001	3.365	2.345	1.668	7.6	5.7	5.3	36.2	58.6	85.4
200	-0.181	-0.071	-0.004	3.153	2.200	1.565	7.7	6.4	5.3	39.4	63.3	90.0

Notes for Panel A: Data is generated as $y^*_{it} = \sum_{h=1}^{2} x^*_{ith}\beta_{ih} + \sum_{\ell=1}^{2} f_{t\ell}\lambda_{i\ell} + \sigma_{\varepsilon,it}\varepsilon_{it}$, $i = 1, 2, ..., N; t = 1, 2, ..., T$, where $\lambda_{i\ell} \sim iidN(0, 1)$, $f_{t\ell} = \rho_f f_{t-1,\ell} + \sqrt{1-\rho_f^2}\nu_{t\ell}$, $\nu_{t\ell} \sim iidN(0, 1)$ with $f_{0,\ell} \sim iidN\,(0, 1)$ for $\ell = 1, .., r$, $\varepsilon_{it} = \rho_\varepsilon \varepsilon_{it-1} + \sqrt{1-\rho_\varepsilon^2}\xi_{it}$, $\xi_{it} \sim iidN(0, 1)$ with $\varepsilon_{i0} \sim iidN\,(0, 1)$, and $\sigma_{\varepsilon,it} = \left(\kappa_{\varepsilon,i}\kappa_{\varepsilon,t}\right)^{1/2}$, $\kappa_{\varepsilon,i} \sim iidU\,(0.5, 1.5)$ and $\kappa_{\varepsilon,t} = 0.5 + t/T$; $x^*_{ith} = \sum_{\ell=1}^{r} f_{t\ell}\gamma_{ih\ell} + \varrho\sigma_{\mathrm{v},it}\mathrm{v}_{ith}$,where $\mathrm{v}_{ith} = \rho_{\mathrm{v}}\mathrm{v}_{it-1,h} + \sqrt{1-\rho_{\mathrm{v}}^2}\varpi_{it,h}$, $\varpi_{it,h} \sim iid\left(\chi^2_6 - 6\right)/\sqrt{12}$ with $\mathrm{v}_{i0,h} \sim iid\left(\chi^2_6 - 6\right)/\sqrt{12}$, $\gamma_{ih\ell} = 0.7\lambda_{i\ell} + \left(1 - 0.7^2\right)^{1/2}\varphi_{ih\ell}$, $\varphi_{ih\ell} \sim iidN(0, 1)$, $\sigma_{\mathrm{v},it} = \left(\kappa_{\mathrm{v},i}\kappa_{\mathrm{v},t}\right)^{1/2}$, $\kappa_{\mathrm{v},i} \sim iidU\,(0.5, 1.5)$ and $\kappa_{\mathrm{v},t} = 0.5 + t/T$, $\varrho^2 = 2$. $\widehat{\beta}_{Bai}$ is non-bias-corrected and $\widetilde{\beta}_{Bai}$ is bias-corrected estimator proposed by Bai (2009). The size is rejection frequency of the proposed Wald test (defined by (18)) for $H_0 : \beta_1 = 1$ and the power for $H_0 : \beta_1 = 0.95$, based on the 5% level test. All results are based on 2000 replications. Notes for Panel B: See notes to Panel A.

Table 2 continued

Panel C: Correlated Heterogeneous Slopes, $\rho_{x\eta}=0.5$, x^*_{ith} generated using χ^2_6												
	(i) β_{ih} is function of $\sum_t x^*_{ith}$						(ii) β_{ih} is function of $\sum_t (x^*_{ith})^2$					
for β_1	Bias ($\times$100)			Size, $H_0:\beta_1=1$			Bias ($\times$100)			Size, $H_0:\beta_1=1$		
T,N	50	100	200	50	100	200	50	100	200	50	100	200
$\widehat{\beta}_{Bai}$												
25	0.439	0.456	0.576	7.5	6.1	5.4	0.591	0.724	0.918	6.7	6.1	6.9
50	0.364	0.456	0.546	8.6	5.9	5.1	0.556	0.754	0.911	8.4	6.5	7.1
100	0.352	0.497	0.571	7.4	5.4	5.4	0.526	0.789	0.918	7.8	6.7	7.0
200	0.476	0.594	0.660	7.3	6.2	5.2	0.650	0.881	1.008	7.4	7.6	8.4
$\widetilde{\beta}_{Bai}$												
25	0.424	0.442	0.563	7.5	6.2	5.5	0.580	0.712	0.907	6.7	6.2	6.8
50	0.347	0.439	0.533	8.6	5.9	5.1	0.540	0.739	0.898	8.4	6.4	6.9
100	0.332	0.479	0.558	7.4	5.5	5.4	0.508	0.767	0.904	7.9	6.7	7.0
200	0.444	0.566	0.641	7.1	5.9	5.0	0.619	0.843	0.978	7.7	7.5	8.0
Panel D: Correlated Heterogeneous Slopes, $\rho_{x\eta}=0.5$, x^*_{ith} generated using $N(0,1)$												
	(i) β_{ih} is function of $\sum_t x^*_{ith}$						(ii) β_{ih} is function of $\sum_t (x^*_{ith})^2$					
for β_1	Bias ($\times$100)			Size, $H_0:\beta_1=1$			Bias ($\times$100)			Size, $H_0:\beta_1=1$		
T,N	50	100	200	50	100	200	50	100	200	50	100	200
$\widehat{\beta}_{Bai}$												
25	-0.006	-0.016	0.083	8.6	5.9	4.3	0.605	0.736	0.873	9.2	6.5	6.5
50	-0.174	-0.090	-0.014	7.7	6.2	4.3	0.460	0.678	0.804	8.2	6.8	6.2
100	-0.177	-0.114	-0.033	7.5	5.1	3.6	0.497	0.698	0.818	8.2	6.5	6.4
200	-0.227	-0.159	-0.079	7.4	5.5	4.0	0.525	0.709	0.831	8.3	6.4	6.6
$\widetilde{\beta}_{Bai}$												
25	-0.022	-0.030	0.071	8.5	6.0	4.2	0.592	0.723	0.862	9.2	6.4	6.5
50	-0.191	-0.104	-0.026	7.9	6.2	4.3	0.443	0.665	0.793	8.1	6.8	6.2
100	-0.200	-0.131	-0.046	7.4	5.3	3.5	0.474	0.675	0.804	8.1	6.3	6.3
200	-0.266	-0.186	-0.097	7.3	5.5	4.1	0.486	0.669	0.800	8.4	6.1	6.6

Notes for Panel C: The data generating process (DGP) is the same as Panel B, except $\beta_{ih}=\beta_h+\sigma_\eta\left(\sqrt{1-\rho^2_{x\eta}}\eta_{ih}+\rho_{x\eta}w_{ih}\right), \eta_{ih}\sim iid\left(\chi^2_6-6\right)/\sqrt{12}$ for $h=1,2$, $w_{ih}=\frac{z_{ih,p}-\bar{z}_{h,p}}{s_{zh,p}}$, where $\bar{z}_{h,p}=N^{-1}\sum_{i=1}^N z_{ih,p}$, $s^2_{zh,p}=(N-1)^{-1}\sum_{i=1}^N\left(z_{ih,p}-\bar{z}_{h,p}\right)^2$, $z_{ih,p}=T^{-1}\sum_{t=1}^T\left(x^*_{it,h}\right)^p$, $p=1,2$. See notes to Panel C. The DGP for Panel D is identical to of Panel C, except that $\varpi_{it,h}\sim iidN(0,1)$.

Table 2 continued

Panel E: Size and Power of the LM test of Correlated Random Coefficient																								
	(A) Size			(B) Size			(C) Power, x_{ith} generated using χ^2_6						(D) Power, x_{ith} generated using $N(0,1)$					
	Slope Homo			Slope Hetero			$\sum_t x^*_{ith}$			$\sum_t (x^*_{ith})^2$			$\sum_t x^*_{ith}$			$\sum_t (x^*_{ith})^2$		
T,N	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200
$LM^{(1)}_{CRC}$																		
25	5.0	5.6	5.8	5.0	4.3	4.3	22.7	37.7	54.7	6.0	5.4	5.9	22.8	40.3	56.0	5.1	5.7	4.6
50	5.8	5.2	5.8	5.6	5.1	5.0	40.4	59.6	75.9	5.9	5.2	6.1	40.0	61.5	76.7	5.2	5.2	5.3
100	5.8	4.8	5.1	4.8	4.4	4.8	71.7	89.5	96.1	5.1	5.4	6.1	71.6	89.7	96.2	5.2	5.7	5.3
200	3.7	4.5	4.7	5.3	5.3	5.3	99.6	100.0	100.0	5.8	7.1	10.2	99.8	100.0	100.0	4.5	5.0	4.8
$LM^{(2)}_{CRC}$																		
25	3.8	4.7	5.3	4.0	4.2	5.1	17.0	32.7	49.5	10.7	20.6	38.5	16.2	32.6	51.0	12.3	24.6	47.2
50	4.0	5.0	5.8	4.8	4.8	4.3	31.7	55.0	72.2	13.6	25.7	48.7	30.3	53.3	72.9	15.6	30.6	55.3
100	4.5	4.8	4.6	4.5	4.6	5.4	59.7	85.2	95.1	14.8	32.2	58.9	60.6	86.3	95.1	17.3	35.4	63.6
200	3.9	5.0	4.7	6.2	4.6	4.6	98.3	100.0	100.0	17.5	40.3	74.5	98.6	100.0	100.0	17.2	38.5	70.4

Notes for Panel E: The results of column blocks (A), (B), (C) and (D) are obtained using the same DGPs as Panels A, B, C and D, respectively. The $LM^{(g)}_{CRC_{Bai}}$ is the proposed LM test of correlated random effects defined by (40), based on bias-corrected Bai (2009) estimator (defined by (25)). The test statistics are referred to the 95% quantile of χ^2_g distribution, $g=1,2$. All results are based on 2000 replications.

References

Arellano, M. (1987). Computing robust standard errors for within-group estimators. *Oxford Bulletin of Economics and Statistics*, 49, 431–434.

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

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

Bai, J. (2009b). Supplement to "Panel data models with interactive effects": Technical details and proofs. Econometrica Supplemental Material.

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

Bai, J. and Liao, Y. (2017). Inferences in panel data with interactive effects using large covariance matrices. *Journal of Econometrics*, 200(1), 59–78.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1), 18–29.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, 18, 5–46.

Doukhan, P. (1994). Mixing: properties and examples. Springer-Verlag, New York.

Galvao, A.F. and Kato, K. Estimation and inference for linear panel data models under misspecification when both N and T are large *Journal of Business and Economic Statistics* 32, 285–309

Greenaway-McGrevy, R., Han, C., Sul, D. (2012). Asymptotic distribution of factor augmented estimators for panel regression. *Journal of Econometrics*, 169, 48–53.

Hsiao, C., Pesaran, M. H. (2008). Random coefficient models. In: Matyas, L., Sevestre, P., eds. The Econometrics of Panel Data. Vol. 46. Berlin: Springer Berlin Heidelberg.

Hansen, C.B. (2007). Asymptotic properties of a robust variance matrix estimator for panel data with T is large. *Journal of Econometrics*, 141, 597-620.

Kapetanios, G., Pesaran, M. H. (2005). Alternative approaches to estimation and inference in large multifactor panels: Small sample results with an application to modelling of asset returns. CESifoWorking Paper Series 1416.

Norkutė, M., Sarafidis, V., Yamagata, T. (2018). Instrumental variable estimation of dynamic linear panel data models with defactored regressors and a multifactor error structure. *ISER Discussion Paper No. 1019.*

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

Pesaran, M. H., Smith, Ron (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68, 79–113.

Pesaran, M.H., Yamagata, T. (2008). Testing slope homogeneity in large panels. *Journal of Econometrics*, 142, 50–93.

Reese, S. and Westerlund J. (2018). Estimation of factor-augmented panel regressions with weakly influential factors. *Econometric Reviews*, 37, 401–465.

Su, L., Chen, Q. (2013). Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory*, 29, 1079–1135.

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

Wooldridge, J. (2010). Econometric analysis of cross section and panel data. MIT press.

Supplementary Appendix

for

"A robust approach to heteroskedasticity, error serial correlation and slope heterogeneity for large linear panel data models with interactive effects"

by G. Cui, K. Hayakawa, S. Nagata and T. Yamagata

In what follows, we repeatedly use Cauchy-Schwarz inequality, triangular inequality, Minkowski in equality, Holder's inequality, and other well-established results: for conformable matrices $\mathbf{ABC}$, $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A}) \text{vec}(\mathbf{B})$, $E \|\mathbf{A} \otimes \mathbf{B}\|^s \leq \left(E \|\mathbf{A}\|^{2s} E \|\mathbf{B}\|^{2s}\right)^{1/2}$, for square matrices, $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \mu_{\max} \|\mathbf{B}\|$.

Appendix A: Lemmas and Proofs for the Results in Section 2

We rely on the law of large numbers and central limit theorem results, which are stated in Lemmas A.1 and A.2, which are given and proved in Hansen (2007). The results which are stated as Lemmas A.3-A.6 are discussed and proven in Hansen (2007), but replicated here for convenience. The proof of main results, which are readily proven based on the lemmas, are given in A.2. We provide proofs of LemmaA.8 in A.3.[16]

A.1: Lemmas for Section 2

Lemma A.1 *Suppose $\{W_{i,T}\}$ are independent across $i = 1, 2, ..., N$ for all T with $\mathbb{E}(W_{i,T}) = \mu_{i,T}$ and $\mathbb{E}|W_{i,T}|^{1+\delta} < \Delta < \infty$ for some $\delta > 0$ and all i, T. Then $N^{-1} \sum_{i=1}^{N} \left(W_{i,T} - \mu_{i,T}\right) \xrightarrow{p} 0$ as $(N, T) \xrightarrow{j} \infty$.*

Lemma A.2 *Suppose $\{\mathbf{w}_{i,T}\}$, $h \times 1$ random vectors, are independent across $i = 1, 2, ..., N$ for all T with $\mathbb{E}(\mathbf{w}_{i,T}) = \mathbf{0}$, $\mathbb{E}\left(\mathbf{w}_{i,T}\mathbf{w}_{i,T}'\right) = \mathbf{\Sigma}_{i,T}$ and $\mathbb{E}\|\mathbf{w}_{i,T}\|^{2+\delta} < \Delta < \infty$ for some $\delta > 0$ and all i, T. Assume $\mathbf{\Sigma} = \lim_{N,T\to\infty} N^{-1} \sum_{i=1}^{N} \mathbf{\Sigma}_{i,T}$ is positive definite and the smallest eigenvalue of $\mathbf{\Sigma}$ is strictly positive. Then, $N^{-1/2} \sum_{i=1}^{N} \mathbf{w}_{i,T} \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma})$ as $(N, T) \xrightarrow{j} \infty$.*

Lemma A.3 *Let $\{w_t\}$ be a strong mixing sequence with $E(w_t) = 0$, $E|w_t|^{s+\delta} < \Delta \leq \infty$ and mixing coefficient $\alpha(m)$ of size $(1-c)r/(r-c)$ where $c \in 2\mathbb{N}$, $s \leq c < r$. Then, there is a constant C depending only on s and $\alpha(m)$ such that $\mathbb{E}\left|\sum_{t=1}^{T} w_t\right|^s \leq C\ D(s, \delta, T)$, where $D(s, \delta, T)$ is as defined in Doukhan (1994) and satisfying $D(s, \delta, T) = O(T)$ for $s \leq 2$ and $D(s, \delta, T) = O(T^{s/2})$ for $s > 2$.*

Lemma A.4 *Under Assumptions A1 and A2, $\bar{\mathbf{A}}_{NT} - \mathbf{A} \xrightarrow{p} \mathbf{0}$ and $\bar{\mathbf{A}}_{NT}^{-1} - \mathbf{A}^{-1} \xrightarrow{p} \mathbf{0}$ as $(N, T) \to \infty$, where $\bar{\mathbf{A}}_{NT}$ and $\mathbf{A}$ are defined by (6) and in Assumption A2, respectively.*

Lemma A.5 *Under Assumptions A1-A3, $\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{X}_i'\boldsymbol{\varepsilon}_i \xrightarrow{d} N(\mathbf{0}, \mathbf{B})$, where $\mathbf{B}$ is defined in Assumption A3.*

Lemma A.6 *Under Assumptions A1-A3, $N^{-1} \sum_{i=1}^{N} \widehat{\mathbf{B}}_{i,T} - \mathbf{B} \xrightarrow{p} \mathbf{0}$ as $(N, T) \to \infty$, where $\widehat{\mathbf{B}}_{i,T} = T^{-1}\mathbf{X}_i'\widehat{\boldsymbol{\varepsilon}}_i\widehat{\boldsymbol{\varepsilon}}_i'\mathbf{X}_i$ with $\widehat{\boldsymbol{\varepsilon}}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}$ with $\boldsymbol{\eta}_i = \mathbf{0}$ for all i, and $\mathbf{B}$ is defined in Assumption A3.*

Lemma A.7 *Under Assumptions A1-A4, $\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{X}_i\boldsymbol{\eta}_i \xrightarrow{d} N(\mathbf{0}, \mathbf{C})$, where $\mathbf{C}$ is defined in Assumption A4.*

[16] Proof of other Lemmas in this subsection is provided in Appendix C.1 for convenience.

Lemma A.8 *Under Assumptions A1-A4,* $N^{-1}\sum_{i=1}^{N}\widehat{\mathbf{C}}_{i,T}-\mathbf{C}\xrightarrow{p}\mathbf{0}$ *as* $(N,T)\rightarrow\infty$*, where* $\widehat{\mathbf{C}}_{i,T}=T^{-2}\mathbf{X}_i'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{X}_i$ *with* $\widehat{\mathbf{u}}_i=\mathbf{y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}}$ *and* $\mathbf{C}$ *is defined in Assumption A4.*

Proof of Lemma A.4. $\mathbb{E}\left\|\bar{\mathbf{A}}_{iT}\right\|^{1+\delta}=\mathbb{E}\left\|T^{-1}\mathbf{X}_i'\mathbf{X}_i\right\|^{1+\delta}\leq\mathbb{E}\left\|T^{-1/2}\mathbf{X}_i\right\|^{2+2\delta}\leq\mathbb{E}\left|tr\left(T^{-1}\mathbf{X}_i'\mathbf{X}_i\right)\right|^{1+\delta}$ $=T^{-(1+\delta)}\left[\left(\mathbb{E}\left|\sum_{h=1}^{k}\sum_{t=1}^{T}x_{ith}^2\right|^{1+\delta}\right)^{\frac{1}{1+\delta}}\right]^{1+\delta}\leq T^{-(1+\delta)}\left[\sum_{h=1}^{k}\sum_{t=1}^{T}\left(\mathbb{E}\left|x_{ith}^2\right|^{1+\delta}\right)^{\frac{1}{1+\delta}}\right]^{1+\delta}\leq k^{1+\delta}\Delta<\infty$ using Holder's and Minkowski's inequality and Assumption A2, then applying Lemma A.1 gives $\bar{\mathbf{A}}_{NT}-\mathbf{A}\xrightarrow{p}\mathbf{0}$. Applying continuous mapping theorem yields $\bar{\mathbf{A}}_{NT}^{-1}-\mathbf{A}^{-1}\xrightarrow{p}\mathbf{0}$. ■

Proof of Lemma A.5. We have

$$\begin{aligned}\mathbb{E}\left\|T^{-1/2}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\right\|^{2+2\delta}&\leq\mathbb{E}\left|\sum_{h=1}^{k}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_{ith}\varepsilon_{it}\right|^2\right|^{1+\delta}\leq\left[\sum_{h=1}^{k}\left(\mathbb{E}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_{ith}\varepsilon_{it}\right|^{2+2\delta}\right)^{\frac{1}{1+\delta}}\right]^{1+\delta}\quad\text{(A.1)}\\&\leq k^{1+\delta}\left(T^{-\frac{2+2\delta}{2}}C\ D\left(s,\delta,T\right)\right)\leq\Delta<\infty,\end{aligned}$$

where the third inequality follows, because, by Assumption A1, $\mathbb{E}\left(x_{ith}\varepsilon_{it}\right)=0$, $\mathbb{E}\left|x_{ith}\varepsilon_{it}\right|^{s+\delta}\leq\mathbb{E}\left|x_{ith}\right|^{2s+2\delta}E\left|\varepsilon_{it}\right|^{2s+2\delta}\leq 2\Delta^{2s+2\delta}<\infty$ for $s>2$ and all $h=1,...,k$, and using Lemma A.3 $\mathbb{E}\left|\sum_{t=1}^{T}x_{ith}\varepsilon_{it}\right|^{2+2\delta}=C\ D\left(s,\delta,T\right)=O\left(T^{\frac{2+2\delta}{2}}\right)$. Therefore $\mathbb{E}\left\|T^{-1/2}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\right\|^{2+2\delta}\leq\Delta$ and together with Assumption A3, applying Lemma A.2 the result follows. ■

Proof of Lemma A.6. We write

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{B}}_{i,T}&=\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{X}_i-\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\boldsymbol{\varepsilon}_i'\mathbf{X}_i\quad\text{(A.2)}\\&\quad-\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\mathbf{X}_i'\mathbf{X}_i+\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\mathbf{X}_i'\mathbf{X}_i\\&=\mathbf{D}_1-\mathbf{D}_2-\mathbf{D}_3+\mathbf{D}_4.\quad\text{(A.3)}\end{aligned}$$

First

$$T^{-1}\text{vec}\left(\mathbf{D}_3\right)=\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\left(\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\otimes\frac{\mathbf{X}_i'\boldsymbol{\varepsilon}_i}{\sqrt{T}}\right)\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right).\quad\text{(A.4)}$$

$\mathbb{E}\left(\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\otimes\frac{\mathbf{X}_i'\boldsymbol{\varepsilon}_i}{\sqrt{T}}\right)=\mathbf{0}$ by Assumptions A1 and A2. Noting

$$\begin{aligned}\mathbb{E}\left\|T^{-1}\mathbf{X}_i'\mathbf{X}_i\right\|^{2+2\delta}&\leq\mathbb{E}\left\|T^{-1/2}\mathbf{X}_i\right\|^{4+4\delta}=\mathbb{E}\left|tr\left(T^{-1}\mathbf{X}_i'\mathbf{X}_i\right)\right|^{2+2\delta}=T^{-(1+\delta)}\left[\left(\mathbb{E}\left|\sum_{h=1}^{k}\sum_{t=1}^{T}x_{ith}^2\right|^{2+2\delta}\right)^{\frac{1}{2+2\delta}}\right]^{2+2\delta}\\&\leq T^{-(1+\delta)}\left[\sum_{h=1}^{k}\sum_{t=1}^{T}\left(\mathbb{E}\left|x_{ith}^4\right|^{1+\delta}\right)^{\frac{1}{2+2\delta}}\right]^{2+2\delta}\leq k^{2+2\delta}\Delta<\infty\quad\text{(A.5)}\end{aligned}$$

and $\mathbb{E}\left\|\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\otimes\frac{\mathbf{X}_i'\boldsymbol{\varepsilon}_i}{\sqrt{T}}\right\|^{1+\delta}\leq\left(\mathbb{E}\left\|\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\right\|^{2+2\delta}\mathbb{E}\left\|\frac{\mathbf{X}_i'\boldsymbol{\varepsilon}_i}{\sqrt{T}}\right\|^{2+2\delta}\right)^{1/2}\leq\Delta$ and by Lemma A.1 $N^{-1}\sum_{i=1}^{N}\left(\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\otimes\frac{\mathbf{X}_i'\boldsymbol{\varepsilon}_i}{\sqrt{T}}\right)=o_p\left(1\right)$ and together with $\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}=O_p\left(1/\sqrt{NT}\right)$, $\text{vec}\left(\mathbf{D}_3\right)=o_p\left(N^{-1/2}\right)$. In a similar manner, $\text{vec}\left(\mathbf{D}_2\right)=o_p\left(N^{-1/2}\right)$. $\left\|\mathbf{D}_4\right\|\leq\left(T/N\right)\sum_{i=1}^{N}\left\|T^{-1}\mathbf{X}_i'\mathbf{X}_i\right\|^2\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|^2=O_p\left(N^{-1}\right)$. $\mathbb{E}\left\|\mathbf{D}_1\right\|^{1+\delta}=\mathbb{E}\left\|T^{-1}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{X}_i\right\|^{1+\delta}\leq\mathbb{E}\left\|T^{-1/2}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\right\|^{2+2\delta}=O\left(1\right)$ by (A.1), and we apply Lemma A.1 to conclude $p\lim_{N,T\rightarrow\infty}N^{-1}\sum_{i=1}^{N}\left(\widehat{\mathbf{B}}_{i,T}-\mathbf{B}_{i,T}\right)=\mathbf{0}$ as required. ■

Proof of Lemma A.7. First $\mathbb{E}\left(T^{-1}\mathbf{X}_i'\mathbf{X}_i\boldsymbol{\eta}_i\right)=\mathbf{0}$ and $Var\left(T^{-1}\mathbf{X}_i'\mathbf{X}_i\boldsymbol{\eta}_i\right)=\mathbf{C}_{iT}$.

$$\begin{aligned}\mathbb{E}\left\|T^{-1}\mathbf{X}_i'\mathbf{X}_i\boldsymbol{\eta}_i\right\|^{2+2\delta} &\leq T^{-(2+2\delta)}\mathbb{E}\left|\sum_{h=1}^k\left|\sum_{\ell=1}^k\sum_{t=1}^T x_{ith}x_{it\ell}\eta_{i\ell}\right|^2\right|^{1+\delta}\\ &\leq T^{-(2+2\delta)}\left|\sum_{h=1}^k\left(\mathbb{E}\left|\sum_{\ell=1}^k\sum_{t=1}^T x_{ith}x_{it\ell}\eta_{i\ell}\right|^{2+2\delta}\right)^{\frac{1}{1+\delta}}\right|^{1+\delta}\end{aligned}$$

but as

$$\mathbb{E}\left|\sum_{\ell=1}^k\sum_{t=1}^T x_{ith}x_{it\ell}\eta_{i\ell}\right|^{2+2\delta}\leq\left[\sum_{\ell=1}^k\sum_{t=1}^T\left(\mathbb{E}\left|x_{ith}x_{it\ell}\eta_{i\ell}\right|^{2+2\delta}\right)^{\frac{1}{2+2\delta}}\right]^{2+2\delta}$$

$$\begin{aligned}\mathbb{E}\left|x_{ith}x_{it\ell}\eta_{i\ell}\right|^{2+2\delta} &\leq \left(\mathbb{E}\left|x_{ith}x_{it\ell}\right|^{4+4\delta}\mathbb{E}\left|\eta_{i\ell}\right|^{4+4\delta}\right)^{1/2}\\ &\leq \left(\mathbb{E}\left|x_{ith}\right|^{8+8\delta}\mathbb{E}\left|x_{it\ell}\right|^{8+8\delta}\right)^{1/4}\left(\mathbb{E}\left|\eta_{i\ell}\right|^{4+4\delta}\right)^{1/2}\leq\Delta\end{aligned}$$

we have

$$\begin{aligned}\mathbb{E}\left\|T^{-1}\mathbf{X}_i'\mathbf{X}_i\boldsymbol{\eta}_i\right\|^{2+2\delta} &\leq T^{-(2+2\delta)}\left|\sum_{h=1}^k\left(\left[\sum_{\ell=1}^k\sum_{t=1}^T\left(\mathbb{E}\left|x_{ith}x_{it\ell}\eta_{i\ell}\right|^{2+2\delta}\right)^{\frac{1}{2+2\delta}}\right]^{2+2\delta}\right)^{\frac{1}{1+\delta}}\right|^{1+\delta}\\ &\leq T^{-(2+2\delta)}\left|\sum_{h=1}^k\left(\left[\sum_{\ell=1}^k\sum_{t=1}^T(\Delta)^{\frac{1}{2+2\delta}}\right]^{2+2\delta}\right)^{\frac{1}{1+\delta}}\right|^{1+\delta}\\ &\leq T^{-(2+2\delta)}k^{1+\delta}\Delta\left[kT\right]^{2+2\delta}=O(1) \qquad \text{(A.6)}\end{aligned}$$

Applying Lemma A.2 the required result follows. ■

Proof of Lemma A.8. We write

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N\widehat{\mathbf{C}}_{i,T} &= \frac{1}{NT^2}\sum_{i=1}^N\mathbf{X}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{X}_i-\frac{1}{NT^2}\sum_{i=1}^N\mathbf{X}_i'\mathbf{X}_i\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\mathbf{u}_i'\mathbf{X}_i\\ &\quad-\frac{1}{NT^2}\sum_{i=1}^N\mathbf{X}_i'\mathbf{u}_i\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\mathbf{X}_i'\mathbf{X}_i+\frac{1}{NT^2}\sum_{i=1}^N\mathbf{X}_i'\mathbf{X}_i\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\mathbf{X}_i'\mathbf{X}_i\\ &= \mathbf{E}_1-\mathbf{E}_2-\mathbf{E}_3+\mathbf{E}_4.\end{aligned}$$

Recall $\mathbf{u}_i=\mathbf{X}_i\boldsymbol{\eta}_i+\boldsymbol{\varepsilon}_i$. First

$$\begin{aligned}\mathbf{E}_3 &= \frac{1}{NT^2}\sum_{i=1}^N\mathbf{X}_i'\mathbf{X}_i\boldsymbol{\eta}_i\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\mathbf{X}_i'\mathbf{X}_i+\frac{1}{NT^2}\sum_{i=1}^N\mathbf{X}_i'\boldsymbol{\varepsilon}_i\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\mathbf{X}_i'\mathbf{X}_i \qquad \text{(A.7)}\\ &= \mathbf{E}_{31}+\mathbf{E}_{32},\text{ say.} \qquad \text{(A.8)}\\ \mathrm{vec}\left(\mathbf{E}_{31}\right) &= \frac{1}{N}\sum_{i=1}^N\left(\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\otimes\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\boldsymbol{\eta}_i\right)\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \qquad \text{(A.9)}\end{aligned}$$

but $\mathbb{E}\left\|\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\otimes\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\boldsymbol{\eta}_i\right\|^{1+\delta}\leq\left(\mathbb{E}\left\|\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\right\|^{2+2\delta}\mathbb{E}\left\|\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\boldsymbol{\eta}_i\right\|^{2+2\delta}\right)^{1/2}\leq\Delta$ by (A.1) and (A.6). As $\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)=O_p\left(N^{-1/2}\right)$, $\mathbf{E}_{31}=O_p\left(N^{-1/2}\right)$. Similarly $\mathrm{vec}\left(\mathbf{E}_{32}\right)=N^{-1}T^{-1/2}\sum_{i=1}^N\left(\frac{\mathbf{X}_i'\mathbf{X}_i}{T}\otimes\frac{\mathbf{X}_i'\boldsymbol{\varepsilon}_i}{\sqrt{T}}\right)\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)=$

$O_p\left(N^{-1/2}T^{-1/2}\right)$, thus, $\mathbf{E}_3 = O_p\left(N^{-1/2}\right) + O_p\left(N^{-1/2}T^{-1/2}\right)$. It is easily seen that $\mathbf{E}_2 = O_p\left(N^{-1/2}\right) + O_p\left(N^{-1/2}T^{-1/2}\right)$. $\|\mathbf{E}_4\| \leq N^{-1}\sum_{i=1}^N \left\|T^{-1}\mathbf{X}_i'\mathbf{X}_i\right\|^2 \left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^2 = O_p\left(N^{-1}\right)$. Finally,

$$
\begin{aligned}
\mathbf{E}_1 &= \frac{1}{NT^2}\sum_{i=1}^N \mathbf{X}_i'\left(\mathbf{X}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i\right)\left(\mathbf{X}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i\right)'\mathbf{X}_i \\
&= \frac{1}{NT^2}\sum_{i=1}^N \mathbf{X}_i'\mathbf{X}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{X}_i'\mathbf{X}_i + \frac{1}{NT^2}\sum_{i=1}^N \mathbf{X}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{X}_i + \frac{1}{NT^2}\sum_{i=1}^N \mathbf{X}_i'\mathbf{X}_i\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_i'\mathbf{X}_i + \frac{1}{NT^2}\sum_{i=1}^N \mathbf{X}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\eta}_i'\mathbf{X}_i'\mathbf{X}_i \\
&= \mathbf{G}_1 + \mathbf{G}_2 + \mathbf{G}_3 + \mathbf{G}_4, \text{ say.}
\end{aligned}
$$

Since, $\mathbb{E}\left\|T^{-3/2}\mathbf{X}_i'\mathbf{X}_i\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_i'\mathbf{X}_i\right\|^{1+\delta} \leq \mathbb{E}\left(\left\|T^{-1}\mathbf{X}_i'\mathbf{X}_i\boldsymbol{\eta}_i\right\|\left\|T^{-1/2}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\right\|\right)^{1+\delta} \leq \mathbb{E}\left\|T^{-1}\mathbf{X}_i'\mathbf{X}_i\boldsymbol{\eta}_i\right\|^{2+2\delta}\mathbb{E}\left\|T^{-1/2}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\right\|^{2+2\delta} \leq \Delta$ by by (A.1) and (A.6), $\mathbf{G}_3 = O_p\left(T^{-1/2}\right)$. By a similar derivation, it is easily seen that $\mathbf{G}_4 = O_p\left(T^{-1/2}\right)$. By (A.1), $\mathbf{G}_2 = O_p\left(T^{-1}\right)$. Finally by (A.6), $\mathbf{G}_1 - \mathbf{C} \to_p \mathbf{0}$, and the required result follows. ■

A.2: Proofs of Main Results in Section 2

Proof of Theorem 1. Applying Lemmas A.4 and A.5, the result immediately follows. ■
Proof of Theorem 2. Applying Lemmas A.4 and A.7, the result immediately follows. ■
Proof of Proposition 1. Applying Lemmas A.8 and A.6, the result immediately follows. ■
Proof of Theorem 3. In the case of both slope homogeneity and slope heterogeneity, using Theorems 1&2, together with Proposition 1, $\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}^{-1/2}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{I}_k\right)$ as $(N,T) \to \infty$. It is straightforward to impose the linear restriction $H_0 : \mathcal{R}\boldsymbol{\beta} = \boldsymbol{r}$ and show that under the null, $\left[\mathcal{R}\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}\mathcal{R}'\right]^{-1/2}\left(\mathcal{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{I}_q\right)$ which implies that $\left(\mathcal{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}\right)'\left[\mathcal{R}\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}\mathcal{R}'\right]^{-1}\left(\mathcal{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}\right) \xrightarrow{d} \chi_q^2$ as $(N,T) \to \infty$. This completes the proof. ■

Appendix B: Lemmas and Proofs for the Results in Section 3

Lemma B.1 *For slope heterogeneous models, when Assumptions B1 to B4 hold, we have*

$$
\begin{aligned}
&(a) \quad \sup_{\mathbf{H}\in\boldsymbol{\mathcal{H}}}\left\|\frac{1}{NT}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{\mathbf{H}}(\mathbf{V}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i)\right\| = o_p(1)\,, \\
&(b) \quad \sup_{\mathbf{H}\in\boldsymbol{\mathcal{H}}}\left\|\frac{1}{NT}\sum_{i=1}^N \boldsymbol{\phi}_i^{0\prime}\mathbf{H}^{0\prime}\mathbf{M}_{\mathbf{H}}(\mathbf{V}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i)\right\| = o_p(1)\,, \\
&(c) \quad \sup_{\mathbf{H}\in\boldsymbol{\mathcal{H}}}\left\|\frac{1}{NT}\sum_{i=1}^N (\mathbf{V}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i)'\mathbf{P}_{\mathbf{H}}(\mathbf{V}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i)\right\| = o_p(1)\,.
\end{aligned}
$$

where $\boldsymbol{\mathcal{H}} = \{\mathbf{H} : \mathbf{H}'\mathbf{H}/T = \mathbf{I}_r\}$.

Proof of Lemma B.1. Consider (a). By Bai (2009a), we have $\sup_{\mathbf{H}\in\boldsymbol{\mathcal{H}}}\left\|\frac{1}{NT}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\boldsymbol{\varepsilon}_i\right\| = o_p(1)$. Note that $\boldsymbol{\eta}_i$ is independent from $\mathbf{X}_i'\mathbf{V}_i$ by Assumption B3, then $\left\|\frac{1}{NT}\sum_{i=1}^N \mathbf{X}_i'\mathbf{V}_i\boldsymbol{\eta}_i\right\| = o_p(1)$. Then it is sufficient to show that $\sup_{\mathbf{H}\in\boldsymbol{\mathcal{H}}}\left\|\frac{1}{NT}\sum_{i=1}^N \mathbf{X}_i'\mathbf{P}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\right\| = o_p(1)$. Using $\mathbf{P}_{\mathbf{H}} = T^{-1}\mathbf{H}\mathbf{H}'$, we have

$$
\begin{aligned}
\left\|\frac{1}{NT}\sum_{i=1}^N \mathbf{X}_i'\mathbf{P}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\right\| &= \left\|\frac{1}{N}\sum_{i=1}^N (T^{-1}\mathbf{X}_i'\mathbf{H})(T^{-1}\mathbf{H}'\mathbf{V}_i\boldsymbol{\eta}_i)\right\| \\
&\leq \|T^{-1/2}\mathbf{H}\| \cdot \frac{1}{N}\sum_{i=1}^N \left\|T^{-1/2}\mathbf{X}_i\right\|\left\|\frac{1}{T}\sum_{t=1}^T \mathbf{h}_t\mathbf{v}_{it}'\right\|\|\boldsymbol{\eta}_i\|.
\end{aligned}
$$

Since $\|T^{-1/2}\mathbf{H}\| = \sqrt{r_1 + r_2}$, the above term is bounded in norm by

$$\sqrt{\left(\frac{r}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|^2\|\boldsymbol{\eta}_i\|^2\right)\cdot\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{h}_t\mathbf{v}_{it}'\right\|^2\right)}. \tag{B.1}$$

Note that $\mathbf{X}_i$ is independent of $\boldsymbol{\eta}_i$, we can take the expectation of the first term of (B.1) to show that it is $O_p(1)$ easily. The second term of (B.1) is equal to

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbf{h}_t'\mathbf{h}_s\mathbb{E}(\mathbf{v}_{is}'\mathbf{v}_{it}) + \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbf{h}_t'\mathbf{h}_s(\mathbf{v}_{is}'\mathbf{v}_{it} - \mathbb{E}(\mathbf{v}_{is}'\mathbf{v}_{it})).$$

The first expression is bounded in norm by

$$\begin{aligned}\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\mathbf{h}_t\|\|\mathbf{h}_s\|\left|\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(\mathbf{v}_{is}'\mathbf{v}_{it})\right| \leq& \sqrt{\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\mathbf{h}_t\|^2\|\mathbf{h}_s\|^2\cdot\sum_{s=1}^{T}\sum_{t=1}^{T}\left|\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(\mathbf{v}_{is}'\mathbf{v}_{it})\right|^2} \\ \leq& r\frac{1}{\sqrt{T}}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\tau_{st}^2} \leq \frac{C}{\sqrt{T}}\end{aligned}$$

with $\left|N^{-1}\sum_{i=1}^{N}\mathbb{E}(\mathbf{v}_{is}'\mathbf{v}_{it})\right| \leq \tau_{st}$ by Assumption B2(i) and (iii). The second expression is bounded in norm by

$$\frac{1}{\sqrt{N}}\cdot\sqrt{\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\mathbf{h}_t\|^2\|\mathbf{h}_s\|^2\cdot\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{v}_{is}'\mathbf{v}_{it} - \mathbb{E}(\mathbf{v}_{is}'\mathbf{v}_{it}))\right|^2} = O_p(N^{-1/2})$$

by Assumption B2(iv). With the above two expressions, the second term of (B.1) is $O_p(N^{-1/2}) + O_p(T^{-1/2})$ uniformly over $\boldsymbol{\mathcal{H}}$, which implies that $\sup_{\mathbf{H}\in\boldsymbol{\mathcal{H}}}\left\|\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{P}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\right\| = o_p(1)$, so we have (a). The proofs for (b) and (c) are similar to that of (a), so are omitted. This completes the proof. ■

Proof of Proposition 5.

For the slope heterogeneous model, without loss of generality, we assume that $\mathbf{f}_t$ and $\boldsymbol{g}_t$ are different factors. Then $\mathbf{H}^0 = (\mathbf{G}^0, \mathbf{F}^0)$ and $\boldsymbol{\phi}_i^0 = (\boldsymbol{\eta}_i'\boldsymbol{\Gamma}_i^{0\prime}, \boldsymbol{\lambda}_i^{0\prime})'$. Given $\boldsymbol{\beta}$ and $\mathbf{H}$, we can concentrate out $\{\boldsymbol{\phi}_i\}_{i=1}^N$, and derive the following concentrated objective function

$$\mathrm{S}_{NT}(\boldsymbol{\beta}, \mathbf{H}) = \frac{1}{NT}\sum_{i=1}^{N}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\mathbf{M}_{\mathbf{H}}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}).$$

Define

$$\begin{aligned}\widetilde{\mathrm{S}}_{NT}(\boldsymbol{\beta}, \mathbf{H}) =& \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{\beta}^0 - \boldsymbol{\beta})'\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) + \frac{2}{NT}\sum_{i=1}^{N}(\boldsymbol{\beta}^0 - \boldsymbol{\beta})'\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{H}^0\boldsymbol{\phi}_i^0 \\ &+ \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\phi}_i^{0\prime}\mathbf{H}^{0\prime}\mathbf{M}_{\mathbf{H}}\mathbf{H}^0\boldsymbol{\phi}_i^0.\end{aligned}$$

Then

$$\begin{aligned}\mathrm{S}_{NT}(\boldsymbol{\beta}, \mathbf{H}) =& \widetilde{\mathrm{S}}_{NT}(\boldsymbol{\beta}, \mathbf{H}) + \frac{2}{NT}\sum_{i=1}^{N}(\boldsymbol{\beta}^0 - \boldsymbol{\beta})'\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}(\mathbf{V}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i) \\ &+ \frac{2}{NT}\sum_{i=1}^{N}\boldsymbol{\phi}_i^{0\prime}\mathbf{H}^{0\prime}\mathbf{M}_{\mathbf{H}}(\mathbf{V}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i) + \frac{1}{NT}\sum_{i=1}^{N}(\mathbf{V}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i)'(\mathbf{P}_{\mathbf{H}} - \mathbf{P}_{\mathbf{H}^0})(\mathbf{V}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i).\end{aligned}$$

With Lemma B.1, we can follow the argument in the proof of Proposition 1 in Bai (2009a) to show that $\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0 \xrightarrow{p} \mathbf{0}$ and $\|\mathbf{P}_{\widehat{\mathbf{H}}}-\mathbf{P}_{\mathbf{H}^0}\| \xrightarrow{p} 0$. In addition, we have that $\mathbf{H}^{0\prime}\widehat{\mathbf{H}}/T$ is invertible and $(\mathbf{H}^{0\prime}\widehat{\mathbf{H}}/T)^{-1} = O_p(1)$.

With the definition of $\widehat{\mathbf{H}}$, we have

$$\frac{1}{NT}\sum_{i=1}^{N}(\mathbf{y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}})(\mathbf{y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}})'\widehat{\mathbf{H}} = \widehat{\mathbf{H}}\mathcal{V}_{NT}$$

where $\mathbf{y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}} = \mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})+\mathbf{H}^0\boldsymbol{\phi}_i^0+\mathbf{V}_i\boldsymbol{\eta}_i+\boldsymbol{\varepsilon}_i$. Then, we have

$$\begin{aligned}
& \frac{1}{\sqrt{T}}\widehat{\mathbf{H}}\mathcal{V}_{NT}-\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i^{0\prime}\mathbf{H}^{0\prime}\widehat{\mathbf{H}} \\
= \;& \frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_i'\widehat{\mathbf{H}}+\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})\boldsymbol{\phi}_i^{0\prime}\mathbf{H}^{0\prime}\widehat{\mathbf{H}} \\
& +\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'\widehat{\mathbf{H}}+\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\mathbf{H}^0\boldsymbol{\phi}_i^0(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_i'\widehat{\mathbf{H}} \\
& +\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_i'\widehat{\mathbf{H}}+\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\varepsilon}_i'\widehat{\mathbf{H}}+\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i^{0\prime}\mathbf{H}^{0\prime}\widehat{\mathbf{H}} \\
& +\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\widehat{\mathbf{H}}+\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})\boldsymbol{\eta}_i'\mathbf{V}_i'\widehat{\mathbf{H}}+\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\mathbf{V}_i\boldsymbol{\eta}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_i'\widehat{\mathbf{H}} \\
& +\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\eta}_i'\mathbf{V}_i'\widehat{\mathbf{H}}+\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\phi}_i^{0\prime}\mathbf{H}^{0\prime}\widehat{\mathbf{H}}+\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_i'\widehat{\mathbf{H}} \\
& +\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\widehat{\mathbf{H}}+\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\widehat{\mathbf{H}} \\
= \;& \mathbb{A}_1+\cdots+\mathbb{A}_{15}\,.
\end{aligned} \tag{B.2}$$

Following the argument in the proof of Proposition A.1 in Bai (2009a), the first five terms are $O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)$. The sixth to the eighth term $O_p(\delta_{NT}^{-1})$ with $\delta_{NT}=\min[\sqrt{N},\sqrt{T}]$. Analogously, the ninth to tenth terms are $O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)$. The eleventh to fourteenth terms are $O_p(N^{-1/2})$. The last term can be decomposed as follows:

$$\mathbb{A}_{15} = \frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\mathbb{E}(\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i')\widehat{\mathbf{H}}+\frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\left(\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'-\mathbb{E}(\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i')\right)\widehat{\mathbf{H}}. \tag{B.3}$$

Note that

$$\frac{1}{N}\sum_{i=1}^{N}\left|\mathbb{E}\left(\mathbf{v}_{is}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{v}_{it}\right)\right| = \frac{1}{N}\sum_{i=1}^{N}\left|\mathrm{tr}\left(\mathbb{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}_i')\mathbb{E}(\mathbf{v}_{it}\mathbf{v}_{is}')\right)\right| \le \Delta\frac{1}{N}\sum_{i=1}^{N}\|\mathbb{E}(\mathbf{v}_{it}\mathbf{v}_{is}')\|$$

by Assumption B2(i). The first expression of (B.3) is bounded in norm by

$$\frac{1}{\sqrt{T}}\cdot\sqrt{\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\left|\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(\mathbf{v}_{is}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{v}_{it}\right)\right|^2}\cdot\|T^{-1/2}\widehat{\mathbf{H}}\| \le \Delta\frac{1}{\sqrt{T}}\cdot\sqrt{\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\mathbb{E}(\mathbf{v}_{it}\mathbf{v}_{is}')\|} = O_p(T^{-1/2})$$

with Assumption B2(iii). The second expression (B.3) is bounded in norm by

$$\frac{1}{\sqrt{N}}\cdot\sqrt{\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{v}_{is}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{v}_{it}-\mathbb{E}\left(\mathbf{v}_{is}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{v}_{it}\right))\right|^2}\cdot\|T^{-1/2}\widehat{\mathbf{H}}\| = O_p(N^{-1/2}).$$

Thus, $\mathbb{A}_{15}$ is $O_p(\delta_{NT}^{-1})$. Collecting the above terms, we have

$$\frac{1}{\sqrt{T}}\widehat{\mathbf{H}}\mathcal{V}_{NT} - \frac{1}{NT\sqrt{T}}\sum_{i=1}^{N}\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i^{0\prime}\mathbf{H}^{0\prime}\widehat{\mathbf{H}} = O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-1}).$$

Since $N^{-1}\sum_{i=1}^{N}\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i^{0\prime}$ and $T^{-1}\mathbf{H}^{0\prime}\widehat{\mathbf{H}}$ both are invertible and $O_p(1)$, then

$$\frac{1}{\sqrt{T}}\widehat{\mathbf{H}}\mathcal{V}_{NT}\left(\frac{1}{T}\mathbf{H}^{0\prime}\widehat{\mathbf{H}}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i^{0\prime}\right)^{-1} - \frac{1}{\sqrt{T}}\mathbf{H}^0 = O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-1}).$$

The rest of the proof exactly follows Proposition A.1 in Bai (2009a) with changes in notation. Below we summarize the results as follows:

1. $\mathcal{V}_{NT}$ is invertible and $\mathcal{V}_{NT}\xrightarrow{p}\mathcal{V}$, where $\mathcal{V}$ $(r\times r)$ is a diagonal matrix consisting of the eigenvalues of $\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}\boldsymbol{\Sigma}_{\mathbf{H}}$.
2. Let $\mathbf{R} = (\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)(T^{-1}\mathbf{H}^{0\prime}\widehat{\mathbf{H}})\mathcal{V}_{NT}^{-1}$. Then $\mathbf{R}$ and $\mathbf{R}^{-1}$ both are $r\times r$ invertible matrices and $O_p(1)$, and
$$\frac{1}{T}\|\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R}\|^2 = O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2) + O_p(\delta_{NT}^{-2})$$
This completes the proof. ■

In all remaining proofs, $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^0$ are used interchangeably, and so are $\mathbf{H}$ and $\mathbf{H}^0$.

Lemma B.2 *Under Assumptions B1 to B4, when the slopes are heterogeneous, we have*

(a) $T^{-1}\mathbf{H}^{0\prime}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R}) = O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-2})$,

(b) $T^{-1}\widehat{\mathbf{H}}'(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R}) = O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-2})$,

(c) $\mathbf{R}\mathbf{R}' - (\mathbf{H}^{0\prime}\mathbf{H}^0/T)^{-1} = O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-2})$,

(d) $T^{-1}\mathbf{V}_k'(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R}) = O_p(T^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(N^{-1}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-2})$ *for each* $k=1,...,N$

(e) $T^{-1}\boldsymbol{\varepsilon}_k'(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R}) = O_p(T^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(N^{-1}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-2})$ *for each* $k=1,...,N$

(f) $\mathbf{M}_{\widehat{\mathbf{H}}} - \mathbf{M}_{\mathbf{H}^0} = O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-1})$,

(g) $\displaystyle\frac{1}{NT}\sum_{k=1}^{N}\mathbf{V}_k'(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R}) = O_p((NT)^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(N^{-1}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(N^{-1/2}\delta_{NT}^{-2}) + O_p(N^{-1})$,

(h) $\displaystyle\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{\varepsilon}_k'(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R}) = O_p((NT)^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(N^{-1}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(N^{-1/2}\delta_{NT}^{-2}) + O_p(N^{-1})$.

Proof of Lemma B.2. Without loss of generality, we assume that $\mathbf{f}_t$ and $\boldsymbol{g}_t$ are different factors. Then $\mathbf{H}^0 = (\mathbf{G}^0,\mathbf{F}^0)$ and $\boldsymbol{\phi}_i^0 = (\boldsymbol{\eta}_i'\boldsymbol{\Gamma}_i^{0\prime},\boldsymbol{\lambda}_i^{0\prime})'$. Consider (a). With (B.2), we have

$$T^{-1}\mathbf{H}^{0\prime}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R}) = T^{-1/2}\mathbf{H}^{0\prime}\mathbb{A}_1\mathcal{V}_{NT}^{-1} + \cdots + T^{-1/2}\mathbf{H}^{0\prime}\mathbb{A}_{15}\mathcal{V}_{NT}^{-1}$$

we can follow the proof in Bai (2009a) to show that the terms $T^{-1/2}\mathbf{H}^{0\prime}\mathbb{A}_1\mathbf{V}_{NT}^{-1}$ to $T^{-1/2}\mathbf{H}^{0\prime}\mathbb{A}_8\mathbf{V}_{NT}^{-1}$ is $O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-2})$. Since $\mathcal{V}_{NT}^{-1} = O_p(1)$, we omit it in the following proof. Analogously, we can prove that the terms $T^{-1/2}\mathbf{H}^{0\prime}\mathbb{A}_9\mathcal{V}_{NT}^{-1}$ and $T^{-1/2}\mathbf{H}^{0\prime}\mathbb{A}_{10}\mathcal{V}_{NT}^{-1}$ are both $O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)$. $T^{-1/2}\mathbf{H}^{0\prime}\mathbb{A}_{11}$ is bounded in norm by

$$\begin{aligned}
&\frac{1}{\sqrt{NT}}\cdot\|T^{-1/2}\mathbf{H}^0\|^2\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\phi}_i^0\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{H}^0\right\|\|\mathbf{R}\| \\
&+\frac{1}{\sqrt{N}}\cdot\|T^{-1/2}\mathbf{H}^0\|^2\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\phi}_i^0\boldsymbol{\eta}_i'\mathbf{V}_i'^{-1/2}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})\right\| \\
=&O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-1/2})\left[O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-1})\right] = O_p(N^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-2})
\end{aligned}$$

by Proposition 5 and $\mathbf{R}=O_p(1)$. Similarly, we can show that $T^{-1/2}\mathbf{H}^{0\prime}\mathbb{A}_{12}$ is $O_p\left(1/\sqrt{NT}\right)$. $T^{-1/2}\mathbf{H}^{0\prime}\mathbb{A}_{13}$ is bounded in norm by

$$\frac{1}{\sqrt{NT}}\left\|\frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\mathbf{H}^{0\prime}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_i'\right\|\left\|T^{-1/2}\widehat{\mathbf{H}}\right\|=O_p\left(1/\sqrt{NT}\right).$$

Similarly, we can show that $T^{-1/2}\mathbf{H}^{0\prime}\mathbb{A}_{14}$ is also $O_p(1/\sqrt{NT})$. By Proposition 5 and $\mathbf{R}=O_p(1)$, the last term $T^{-1/2}\mathbf{H}^{0\prime}\mathbb{A}_{15}$ is bounded in norm by

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{H}^{0\prime}\mathbf{V}_i\boldsymbol{\eta}_i\|^2\|\mathbf{R}\|+\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{H}^{0\prime}\mathbf{V}_i\boldsymbol{\eta}_i\|\cdot T^{-1/2}\|\boldsymbol{\eta}_i'\mathbf{V}_i'^{-1}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})\|\\
\leq&O_p(T^{-1})+\sqrt{\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{H}^{0\prime}\mathbf{V}_i\boldsymbol{\eta}_i\|^2\cdot\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\boldsymbol{\eta}_i'\mathbf{v}_{it}\|^2}\|T^{-1}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})\|\\
=&O_p(T^{-1})+O_p(T^{-1/2})\left[O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-1})\right]=O_p(T^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-2}).
\end{aligned}$$

Thus,

$$\frac{1}{T}\mathbf{H}^{0\prime}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})=O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-2}).$$

Consider (b), which is given by

$$\frac{1}{T}\widehat{\mathbf{H}}'(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})=\frac{1}{T}\mathbf{R}'\mathbf{H}^{0\prime}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})+\frac{1}{T}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})'(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R}).$$

The first term is $O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-2})$ by (a), the second is bounded in norm by $T^{-1}\|\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R}\|^2=O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)+O_p(\delta_{NT}^{-2})$ by Proposition 5. Then we have (b).

With (a) and (b), we can follow the argument in the proof of Lemma A.7 of Bai (2009a) to derive (c).

Consider (d). With (B.2), we have

$$\begin{aligned}
&\frac{1}{T}\mathbf{V}_k'(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})\\
=&\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_i'\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})\boldsymbol{\phi}_i^{0\prime}\mathbf{H}^{0\prime}\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}\\
&+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\mathbf{H}^0\boldsymbol{\phi}_i^0(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_i'\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}\\
&+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\boldsymbol{\varepsilon}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_i'\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\varepsilon}_i'\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}\\
&+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i^{0\prime}\mathbf{H}^{0\prime}\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}\\
&+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})\boldsymbol{\eta}_i'\mathbf{V}_i'\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\mathbf{V}_i\boldsymbol{\eta}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_i'\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}\\
&+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\eta}_i'\mathbf{V}_i'\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\phi}_i^{0\prime}\mathbf{H}^{0\prime}\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_i'\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}\\
&+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\boldsymbol{\varepsilon}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\widehat{\mathbf{H}}\mathcal{V}_{NT}^{-1}\\
=&\mathbb{B}_1+\mathbb{B}_2+\cdots+\mathbb{B}_{15}.
\end{aligned}$$

Hereafter, we ignore $\mathcal{V}_{NT}^{-1}$, which is $O_p(1)$. With $\mathbf{X}_i = \mathbf{H}^0\mathbf{\Gamma}_i^0 + \mathbf{V}_i$, $\mathbb{B}_1$ is equal to

$$\mathbb{B}_1 = \frac{1}{NT^2}\sum_{i=1}^N \mathbf{V}_k'\mathbf{H}^0\mathbf{\Gamma}_i^0(\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}})(\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}})'\mathbf{X}_i'\widehat{\mathbf{H}} + \frac{1}{NT^2}\sum_{i=1}^N \mathbf{V}_k'\mathbf{V}_i(\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}})(\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}})'\mathbf{X}_i'\widehat{\mathbf{H}}.$$

The first term is bounded in norm by

$$\frac{1}{\sqrt{T}} \cdot \|T^{-1/2}\mathbf{V}_k'\mathbf{H}^0\|\|T^{-1/2}\widehat{\mathbf{H}}\| \cdot \|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|^2 \cdot \frac{1}{N}\sum_{i=1}^N \|\mathbf{\Gamma}_i^0\|\|T^{-1/2}\mathbf{X}_i\| = O_p(T^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|^2)$$

while the second term is bounded in norm by

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N \|T^{-1}\mathbb{E}(\mathbf{V}_k'\mathbf{V}_i)\|\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|^2\|T^{-1/2}\widehat{\mathbf{H}}\| \\
&+ \frac{1}{\sqrt{T}}\frac{1}{N}\sum_{i=1}^N \|T^{-1/2}\left(\mathbf{V}_k'\mathbf{V}_i - \mathbb{E}(\mathbf{V}_k'\mathbf{V}_i)\right)\|\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|^2\|T^{-1/2}\widehat{\mathbf{H}}\| \\
=&O_p(N^{-1}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|^2) + O_p(T^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|^2)
\end{aligned}$$

with Assumption B2(i). Then, $\mathbb{B}_1$ is $O_p(N^{-1}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|^2) + O_p(T^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|^2)$. Analogously, we can show that $\mathbb{B}_2$, $\mathbb{B}_3$, $\mathbb{B}_9$ and $\mathbb{B}_{10}$ both are $O_p(N^{-1}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(T^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|)$. We can also show that $\mathbb{B}_4$ and $\mathbb{B}_5$ both are $O_p(T^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|)$. $\mathbb{B}_6$ is bounded in norm by

$$\frac{1}{\sqrt{NT}} \cdot \|T^{-1/2}\mathbf{V}_k'\mathbf{H}^0\| \left\| \frac{1}{\sqrt{NT}}\sum_{i=1}^N \boldsymbol{\phi}_i^0\boldsymbol{\varepsilon}_i'^{-1/2}\widehat{\mathbf{H}} \right\| = O_p(N^{-1/2}T^{-1/2}),$$

$\mathbb{B}_7$ is bounded in norm by

$$\frac{1}{\sqrt{NT}} \cdot \left\| \frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{V}_k'\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i^{0\prime} \right\| \|T^{-1/2}\mathbf{H}^0\|\|T^{-1/2}\widehat{\mathbf{H}}\| = O_p(N^{-1/2}T^{-1/2}),$$

by Assumptions B4(iv). $\mathbb{B}_8$ is bounded in norm by

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^N \|T^{-1/2}\mathbf{V}_k'\boldsymbol{\varepsilon}_i\|\|T^{-1/2}\boldsymbol{\varepsilon}_i'\mathbf{H}^0\|\|\mathbf{R}\| + \frac{1}{\sqrt{T}}\frac{1}{N}\sum_{i=1}^N \|T^{-1/2}\mathbf{V}_k'\boldsymbol{\varepsilon}_i\|\|T^{-1/2}\boldsymbol{\varepsilon}_i\|\|T^{-1/2}(\widehat{\mathbf{H}} - \mathbf{H}^0\mathbf{R})\| \\
=&O_p(T^{-1}) + O_p(T^{-1/2})\left[O_p(\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-1})\right] = O_p(T^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-2}),
\end{aligned}$$

by Proposition 5. $\mathbb{B}_{11}$ is bounded in norm by

$$\begin{aligned}
&\frac{1}{\sqrt{NT}} \cdot \|T^{-1/2}\mathbf{V}_k'\mathbf{H}^0\| \left\| \frac{1}{\sqrt{NT}}\sum_{i=1}^N \boldsymbol{\phi}_i^0\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{H}^0 \right\| \|\mathbf{R}\| \\
&+ \frac{1}{\sqrt{NT}} \cdot \|T^{-1/2}\mathbf{V}_k'\mathbf{H}^0\| \left\| \frac{1}{\sqrt{NT}}\sum_{i=1}^N \boldsymbol{\phi}_i^0\boldsymbol{\eta}_i'\mathbf{V}_i' \right\| \left\| T^{-1/2}\left(\widehat{\mathbf{H}} - \mathbf{H}^0\mathbf{R}\right) \right\| \\
=&O_p(N^{-1/2}T^{-1}) + O_p(N^{-1/2}T^{-1/2})\left[O_p(\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-1})\right] = O_p(N^{-1/2}T^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-2}).
\end{aligned}$$

$\mathbb{B}_{12}$ is equal to (ignoring $\mathbf{H}^{0\prime}\widehat{\mathbf{H}}/T$ since it is $O_p(1)$)

$$\frac{1}{NT}\sum_{i=1}^N \mathbb{E}(\mathbf{V}_k'\mathbf{V}_i)\boldsymbol{\eta}_i\boldsymbol{\phi}_i^{0\prime} + \frac{1}{NT}\sum_{i=1}^N \left[\mathbf{V}_k'\mathbf{V}_i - \mathbb{E}(\mathbf{V}_k'\mathbf{V}_i)\right]\boldsymbol{\eta}_i\boldsymbol{\phi}_i^{0\prime}$$

in which the first term to be $O_p(1/N)$, the second terms can be shown to be $O_p(1/\sqrt{NT})$, easily. Then $\mathbb{B}_{12} = O_p(1/\sqrt{NT}) + O_p(1/N)$. Similar to the argument in the proof $\mathbb{B}_{12}$, we can show that $\mathbb{B}_{13} = O_p(1/\sqrt{NT}) + O_p(1/N)$.

$\mathbb{B}_{14}$ is bounded in norm by

$$
\begin{aligned}
\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\boldsymbol{\varepsilon}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\widehat{\mathbf{H}} =&\frac{1}{NT}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{V}_k'\boldsymbol{\varepsilon}_i\|\|\boldsymbol{\eta}_i\|\|T^{-1/2}\mathbf{V}_i'\mathbf{H}^0\|\|\mathbf{R}\| \\
&+\frac{1}{\sqrt{T}}\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{V}_k'\boldsymbol{\varepsilon}_i\|\|\boldsymbol{\eta}_i\|\|T^{-1/2}\mathbf{V}_i\|\|T^{-1/2}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})\| \\
=&O_p(T^{-1})+O_p(T^{-1/2})\left[O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-1})\right]=O_p(T^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-2}).
\end{aligned}
$$

$\mathbb{B}_{15}$ is equal to

$$
\mathbb{B}_{15}=\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{H}^0\mathbf{R}+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_k'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})\,. \tag{B.4}
$$

The first term of (B.4) is bounded in norm by (ignoring $\mathbf{R}$)

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\frac{1}{N}\sum_{i=1}^{N}\|T^{-1}\mathbb{E}(\mathbf{V}_k'\mathbf{V}_i)\|\|\boldsymbol{\eta}_i\|^2\|T^{-1/2}\mathbf{V}_i'\mathbf{H}^0\| \\
&+\frac{1}{NT}\sum_{i=1}^{N}\|T^{-1/2}\left(\mathbf{V}_k'\mathbf{V}_i-\mathbb{E}(\mathbf{V}_k'\mathbf{V}_i)\right)\|\|\boldsymbol{\eta}_i\|^2\|T^{-1/2}\mathbf{V}_i'\mathbf{H}^0\|=O_p(N^{-1}T^{-1/2})+O_p(T^{-1})\,,
\end{aligned}
$$

and the second term is bounded in norm by

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\|T^{-1}\mathbb{E}(\mathbf{V}_k'\mathbf{V}_i)\|\|\boldsymbol{\eta}_i\|^2\|T^{-1/2}\mathbf{V}_i\|\cdot\|T^{-1/2}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})\| \\
&+\frac{1}{\sqrt{T}}\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\left(\mathbf{V}_k'\mathbf{V}_i-\mathbb{E}(\mathbf{V}_k'\mathbf{V}_i)\right)\|\|\boldsymbol{\eta}_i\|^2\|T^{-1/2}\mathbf{V}_i\|\cdot\|T^{-1/2}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})\| \\
=&\left[O_p(N^{-1})+O_p(T^{-1/2})\right]\left[O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-1})\right]\,.
\end{aligned}
$$

Thus, $\mathbb{B}_{15}$ is $O_p(T^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(N^{-1}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-2})$. Collecting the above terms, we have

$$
T^{-1}\mathbf{V}_k'(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})=O_p(T^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(N^{-1}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-2})\,.
$$

The claim (e) can be proved by following the argument in the proof of (d), then details are omitted. For (f), we decompose the left hand side term as

$$
\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}^0}=-\frac{1}{T}\widehat{\mathbf{H}}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})'^{-1}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})\mathbf{R}'\mathbf{H}^{0\prime}-\frac{1}{T}\mathbf{H}^0\left(\mathbf{R}\mathbf{R}'^{-1}\mathbf{H}^{0\prime}\mathbf{H}^0)^{-1}\right)\mathbf{H}^{0\prime}
$$

then it will be bounded in norm by

$$
\begin{aligned}
&\|T^{-1/2}\widehat{\mathbf{H}}\|\|T^{-1/2}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})\|+\|\mathbf{R}\|\|T^{-1/2}\mathbf{H}^0\|\|T^{-1/2}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})\|+\|T^{-1/2}\mathbf{H}^0\|^2\|\mathbf{R}\mathbf{R}'^{-1}\mathbf{H}^{0\prime}\mathbf{H}^0)^{-1}\| \\
=&O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-1})
\end{aligned}
$$

with (a), (b), (c) and the facts that $\|T^{-1/2}\widehat{\mathbf{H}}\|^2=r_1+r_2$ and $\mathbb{E}\|T^{-1/2}\mathbf{H}^0\|^2\le\Delta$ by Assumption B4(i) and (ii). Thus, we complete the proof. (g) and (h) are derived from (e) and (f), respectively.■

Proofs of Theorem 6 and Corollary 8.

Without loss of generality, we assume that the factors $\mathbf{F}$ are different from $\mathbf{G}$. Since the slopes are heterogeneous, $\mathbf{H}^0 = (\mathbf{G}^0, \mathbf{F}^0)$ and $\boldsymbol{\phi}_i^0 = (\boldsymbol{\eta}_i'\boldsymbol{\Gamma}_i^{0\prime}, \boldsymbol{\lambda}_i^{0\prime})'$. By the definition of $\widehat{\boldsymbol{\beta}}$, we have

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} &= \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i \\
&= \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0 + \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{V}_i\boldsymbol{\eta}_i \\
&\quad + \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i
\end{aligned}$$

which implies that

$$\left(\frac{1}{NT}\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i\right)^{-1} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \frac{1}{NT}\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0 + \frac{1}{NT}\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i + \frac{1}{NT}\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{V}_i\boldsymbol{\eta}_i. \tag{B.5}$$

Consider the first term of (B.5). With (B.2), we have

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0 &= \frac{1}{NT}\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}(\mathbf{H}^0 - \widehat{\mathbf{H}}\mathbf{R}^{-1})\boldsymbol{\phi}_i^0 \\
&= -\frac{1}{\sqrt{T}}\frac{1}{N}\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}(\mathbb{A}_1 + \cdots + \mathbb{A}_{15})(\mathbf{H}^{0\prime}\widehat{\mathbf{H}}/T)^{-1}(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\boldsymbol{\phi}_i^0 \\
&\equiv \mathbb{F}_1 + \mathbb{F}_2 + \cdots + \mathbb{F}_{15}.
\end{aligned}$$

$\mathbb{F}_1$ can be shown to be $O_p(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2)$ easily. $\mathbb{F}_2$ is equal to

$$\begin{aligned}
\mathbb{F}_2 &= -\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}(\mathbf{H}^0 - \widehat{\mathbf{H}}\mathbf{R}^{-1})\boldsymbol{\Gamma}_k^0\boldsymbol{\phi}_k^{0\prime}(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\boldsymbol{\phi}_i^0 \cdot (\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}) \\
&\quad -\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N} \mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{V}_k\boldsymbol{\phi}_k^{0\prime}(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\boldsymbol{\phi}_i^0 \cdot (\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}})
\end{aligned}$$

which is bounded in norm by

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N} \|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\| \cdot \frac{1}{N}\sum_{k=1}^{N} \|\boldsymbol{\Gamma}_k^0\|\|\boldsymbol{\phi}_k^0\| \cdot \|T^{-1/2}(\mathbf{H}^0 - \widehat{\mathbf{H}}\mathbf{R}^{-1})\|\|(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\| \cdot \|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\| \\
&+ \frac{1}{N^2}\frac{1}{\sqrt{T}}\sum_{i=1}^{N} \|T^{-1/2}\mathbf{X}_i\| \left\|\sum_{k=1}^{N} \boldsymbol{\phi}_i^{0\prime}(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\boldsymbol{\phi}_k^0\mathbf{V}_k\right\| \cdot \|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\| \\
=&O_p(\delta_{NT}^{-1}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|^2) \\
&+ \frac{1}{N^2}\frac{1}{\sqrt{T}}\sum_{i=1}^{N} \|T^{-1/2}\mathbf{X}_i\| \sqrt{\sum_{s=1}^{T}\left\|\sum_{k=1}^{N} \boldsymbol{\phi}_i^{0\prime}(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\boldsymbol{\phi}_k^0\mathbf{v}_{ks}'\right\|^2} \cdot \|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\| \\
=&O_p(\delta_{NT}^{-1}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|^2) \\
&+ \frac{1}{N}\sum_{i=1}^{N} \|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\| \cdot \|(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\| \sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^{N} \boldsymbol{\phi}_k^0\mathbf{v}_{ks}'\right\|^2} \cdot N^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\| \\
=&O_p(\delta_{NT}^{-1}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|^2)
\end{aligned}$$

by Proposition 5. $\mathbb{F}_3$ is bounded in norm by

$$
\begin{aligned}
&\frac{1}{NT\sqrt{T}}\sum_{k=1}^{N}\|\mathbf{X}_k\|\|\boldsymbol{\varepsilon}_k'\widehat{\mathbf{H}}\|\cdot\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|\cdot\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|\cdot\|(\mathbf{H}^{0\prime}\widehat{\mathbf{H}}/T)^{-1}\|\|(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\|\\
=&\frac{1}{N}\frac{1}{T\sqrt{T}}\sum_{k=1}^{N}\|\mathbf{X}_k\|\|\boldsymbol{\varepsilon}_k'\widehat{\mathbf{H}}\|\cdot O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)\\
=&\frac{1}{N}\frac{1}{\sqrt{T}}\sum_{k=1}^{N}\|T^{-1/2}\mathbf{X}_k\|\|T^{-1/2}\boldsymbol{\varepsilon}_k'\mathbf{H}^0\|\|\mathbf{R}\|\cdot O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)\\
&+\frac{1}{N}\sum_{k=1}^{N}\|T^{-1/2}\mathbf{X}_k\|\|T^{-1/2}\boldsymbol{\varepsilon}_k\|\cdot\|T^{-1/2}(\widehat{\mathbf{H}}-\mathbf{H}^0\mathbf{R})\|\cdot O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)\\
=&O_p(\delta_{NT}^{-1}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2).
\end{aligned}
$$

Analogously, $\mathbb{F}_9$ is proved to be $O_p(\delta_{NT}^{-1}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)$. $\mathbb{F}_4$ is equal to

$$
\mathbb{F}_4=-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})\boldsymbol{\phi}_k^0(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_k'\widehat{\mathbf{H}}(\mathbf{H}^{0\prime}\widehat{\mathbf{H}}/T)^{-1}(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\boldsymbol{\phi}_i^0
$$

which is bounded in norm by

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}(\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|)^2\cdot\|T^{-1/2}\widehat{\mathbf{H}}\|\|T^{-1/2}(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})\|\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|\cdot\|(\mathbf{H}^{0\prime}\widehat{\mathbf{H}}/T)^{-1}\|\|(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\|\\
=&\left[O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-1})\right]\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|=O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)+O_p(\delta_{NT}^{-1}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)
\end{aligned}
$$

with $\|\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i\|\leq\|\mathbf{X}_i\|$, $T^{-1/2}\|\widehat{\mathbf{H}}\|=\sqrt{r_1+r_2}$ and Proposition 5. $\mathbb{F}_5$ is equal to

$$
\mathbb{F}_5=-\frac{1}{N^2}\frac{1}{T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_k(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_k'\widehat{\mathbf{H}}(\mathbf{H}^{0\prime}\widehat{\mathbf{H}}/T)^{-1}(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\boldsymbol{\phi}_i^0
$$

which is bounded in norm by

$$
\begin{aligned}
&\frac{1}{NT}\left\|\sum_{k=1}^{N}\boldsymbol{\varepsilon}_k(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_k'\right\|\cdot\|T^{-1/2}\widehat{\mathbf{H}}\|\|(\mathbf{H}^{0\prime}\widehat{\mathbf{H}}/T)^{-1}\|\|(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\|\cdot\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|\\
=&\frac{1}{NT}\left\|\sum_{k=1}^{N}\boldsymbol{\varepsilon}_k(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_k'\right\|\cdot O_p(1)\leq\frac{1}{\sqrt{N}}\cdot\sqrt{\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\boldsymbol{\varepsilon}_{ks}\mathbf{x}_{kt}'\right\|^2}\cdot\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|\cdot O_p(1)\\
=&O_p(N^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|).
\end{aligned}
$$

$\mathbb{F}_6$ is equal to

$$
\mathbb{F}_6=-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})\boldsymbol{\phi}_k^0\boldsymbol{\varepsilon}_k'\widehat{\mathbf{H}}(\mathbf{H}^{0\prime}\widehat{\mathbf{H}}/T)^{-1}(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\boldsymbol{\phi}_i^0
$$

which is bounded in norm by

$$
\begin{aligned}
&\|T^{-1/2}(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})\|\left\|\frac{1}{NT}\sum_{k=1}^N\boldsymbol{\phi}_k^0\boldsymbol{\varepsilon}_k'\widehat{\mathbf{H}}\right\|\cdot O_p(1)\\
\le&\|T^{-1/2}(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})\|\left\|\frac{1}{NT}\sum_{k=1}^N\boldsymbol{\phi}_k^0\boldsymbol{\varepsilon}_k'\mathbf{H}^0\right\|\|\mathbf{R}\|\cdot O_p(1)\\
&+\|T^{-1/2}(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})\|^2\cdot\frac{1}{\sqrt{T}}\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^N\boldsymbol{\phi}_k^0\boldsymbol{\varepsilon}_k'\right\|\cdot O_p(N^{-1/2})\\
=&O_p(N^{-1/2}T^{-1/2}\delta_{NT}^{-1})+O_p(N^{-1/2}T^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(T^{-1/2}\delta_{NT}^{-2})+O_p(T^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)\\
=&O_p(T^{-1/2}\delta_{NT}^{-2})+O_p(N^{-1/2}T^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(T^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)
\end{aligned}
$$

with Proposition 5. Since $\mathbf{H}^0=(\mathbf{G}^0,\mathbf{F}^0)$, $\mathbf{F}^0=\mathbf{H}^0\mathbf{S}$, where $\mathbf{S}=(\mathbf{0}_{r_1\times r_2},\mathbf{I}_{r_1})'$. $\mathbb{F}_7$ is equal to

$$
\begin{aligned}
&-\frac{1}{N^2T}\sum_{i=1}^N\sum_{k=1}^N\boldsymbol{\Gamma}_i^{0\prime}\mathbf{S}'(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_k\boldsymbol{\phi}_k^{0\prime}(N^{-1}\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\phi}_i^0-\frac{1}{N^2T}\sum_{i=1}^N\sum_{k=1}^N\mathbf{V}_i'\boldsymbol{\varepsilon}_k\boldsymbol{\phi}_k^{0\prime}(N^{-1}\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\phi}_i^0\\
&+\frac{1}{N^2T^2}\sum_{i=1}^N\sum_{k=1}^N\mathbf{V}_i'\widehat{\mathbf{H}}\widehat{\mathbf{H}}'\boldsymbol{\varepsilon}_k\boldsymbol{\phi}_k^{0\prime}(N^{-1}\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\phi}_i^0
\end{aligned}
$$

which is bounded in norm by (ignoring $(N^{-1}\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}$)

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N\|\boldsymbol{\Gamma}_i^0\|\|\boldsymbol{\phi}_i^0\|\|N^{-1/2}T^{-1/2}(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})\|\cdot\frac{1}{\sqrt{T}}\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^N\boldsymbol{\varepsilon}_k\boldsymbol{\phi}_k^{0\prime}\right\|\\
&+\frac{1}{\sqrt{NT}}\cdot\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{\sqrt{NT}}\sum_{k=1}^N\mathbf{V}_i'\boldsymbol{\varepsilon}_k\boldsymbol{\phi}_k^{0\prime}\right\|\|\boldsymbol{\phi}_i^0\|+\left\|\frac{1}{NT}\sum_{i=1}^N\mathbf{V}_i'\widehat{\mathbf{H}}\right\|\left\|\frac{1}{NT}\sum_{k=1}^N\widehat{\mathbf{H}}'\boldsymbol{\varepsilon}_k\boldsymbol{\phi}_k^{0\prime}\right\|\|\boldsymbol{\phi}_i^0\|\\
=&O_p(N^{-1/2}\delta_{NT}^{-1})+O_p(N^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|).
\end{aligned}
$$

Analogously, we can prove that $\mathbb{F}_8$ is equal to $O_p(\delta_{NT}^{-2})+O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)$. $\mathbb{F}_{10}$ is equal to

$$
\mathbb{F}_{10}=-\frac{1}{N^2T^2}\sum_{i=1}^N\sum_{k=1}^N\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{V}_k\boldsymbol{\eta}_k(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_k'\widehat{\mathbf{H}}(\mathbf{H}^{0\prime}\widehat{\mathbf{H}}/T)^{-1}(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\boldsymbol{\phi}_i^0
$$

which is bounded in norm by

$$
\begin{aligned}
&\frac{1}{NT}\left\|\sum_{k=1}^N\mathbf{V}_k\boldsymbol{\eta}_k(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_k'\right\|\cdot\|T^{-1/2}\widehat{\mathbf{H}}\|\|(\mathbf{H}^{0\prime}\widehat{\mathbf{H}}/T)^{-1}\|\|(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\|\cdot\frac{1}{N}\sum_{i=1}^N\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|\\
=&\frac{1}{NT}\left\|\sum_{k=1}^N\mathbf{V}_k\boldsymbol{\eta}_k(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{X}_k'\right\|\cdot O_p(1)\\
\le&\frac{1}{N\sqrt{T}}\left\|\sum_{k=1}^N\mathbf{V}_k\boldsymbol{\eta}_k(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\boldsymbol{\Gamma}_k^{0\prime}\right\|\|T^{-1/2}\mathbf{H}^0\|\cdot O_p(1)+\frac{1}{NT}\left\|\sum_{k=1}^N\mathbf{V}_k\boldsymbol{\eta}_k(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})'\mathbf{V}_k'\right\|\cdot O_p(1)\\
\le&\frac{1}{\sqrt{N}}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|\cdot\sqrt{\frac{1}{T}\sum_{s=1}^T\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^N\boldsymbol{\Gamma}_k^{0\prime}\mathbf{v}_{ks}'\boldsymbol{\eta}_k\right\|^2}\cdot O_p(1)\\
&+\frac{1}{\sqrt{N}}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|\cdot\sqrt{\frac{1}{T^2}\sum_{s=1}^T\sum_{t=1}^T\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^N\boldsymbol{\eta}_k'\mathbf{v}_{ks}\mathbf{v}_{kt}'\right\|}\cdot O_p(1)=O_p(N^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)
\end{aligned}
$$

by Assumption B3. Following the argument in the proof of $\mathbb{F}_6$, we can prove that $\mathbb{F}_{11}$ is equal to

$$
-\frac{1}{N^2T^2}\sum_{i=1}^N\sum_{k=1}^N\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_k^0\boldsymbol{\eta}_k'\mathbf{V}_k'\widehat{\mathbf{H}}(\mathbf{H}^{0\prime}\widehat{\mathbf{H}}/T)^{-1}(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\boldsymbol{\phi}_i^0=O_p(\delta_{NT}^{-2})+O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2).
$$

$\mathbb{F}_{12}$ is equal to

$$
\begin{aligned}
&-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N}\mathbf{\Gamma}_i^{0\prime}\mathbf{S}'(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{V}_k\boldsymbol{\eta}_k\boldsymbol{\phi}_k^{0\prime}(\mathbf{\Phi}'\mathbf{\Phi}/N)^{-1}\boldsymbol{\phi}_i^0\\
&-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N}\mathbb{E}(\mathbf{V}_i'\mathbf{V}_k)\boldsymbol{\eta}_k\boldsymbol{\phi}_k^{0\prime}(\mathbf{\Phi}'\mathbf{\Phi}/N)^{-1}\boldsymbol{\phi}_i^0+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\mathbf{V}_i'\mathbf{V}_k-\mathbb{E}(\mathbf{V}_i'\mathbf{V}_k)\right)\boldsymbol{\eta}_k\boldsymbol{\phi}_k^{0\prime}(\mathbf{\Phi}'\mathbf{\Phi}/N)^{-1}\boldsymbol{\phi}_i^0\\
&+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\mathbf{V}_i'\widehat{\mathbf{H}}\widehat{\mathbf{H}}'\mathbf{V}_k\boldsymbol{\eta}_k\boldsymbol{\phi}_k^{0\prime}(\mathbf{\Phi}'\mathbf{\Phi}/N)^{-1}\boldsymbol{\phi}_i^0.
\end{aligned}
$$

It is easy to show that the first expression is $O_p(N^{-1/2}\delta_{NT}^{-1})+O_p(N^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)$, the second expression is $O_p(N^{-1})$, the third term is $O_p(N^{-1/2}T^{-1/2})$, the forth term is $O_p(\delta_{NT}^{-2})+O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)$. $\mathbb{F}_{13}$ is bounded in norm by

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|\cdot\|(\mathbf{F}^{0\prime}\widehat{\mathbf{F}}/T)^{-1}\|\|(\mathbf{\Phi}'\mathbf{\Phi}/N)^{-1}\|\cdot\frac{1}{N}\frac{1}{T\sqrt{T}}\left\|\sum_{k=1}^{N}\mathbf{V}_k\boldsymbol{\eta}_k\boldsymbol{\varepsilon}_k'\widehat{\mathbf{F}}\right\|\\
=&\frac{1}{\sqrt{NT}}\cdot\frac{1}{\sqrt{T}}\left\|\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\mathbf{V}_k\boldsymbol{\eta}_k\boldsymbol{\varepsilon}_k'\widehat{\mathbf{F}}\right\|\cdot O_p(1)\\
\le&\frac{1}{\sqrt{NT}}\cdot\frac{1}{\sqrt{T}}\left\|\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\mathbf{V}_k\boldsymbol{\eta}_k\boldsymbol{\varepsilon}_k'\mathbf{F}^0\right\|\|\mathbf{R}\|\cdot O_p(1)+\frac{1}{\sqrt{N}}\cdot\frac{1}{T}\left\|\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\mathbf{V}_k\boldsymbol{\eta}_k\boldsymbol{\varepsilon}_k'\right\|\left\|T^{-1/2}\left(\widehat{\mathbf{F}}-\mathbf{F}^0\right)\right\|\cdot O_p(1)\\
=&O_p(N^{-1/2}\delta_{NT}^{-1})+O_p(N^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|).
\end{aligned}
$$

Similar to the argument in the proof of $\mathbb{F}_{13}$, we can prove that $\mathbb{F}_{14}$ is equal to $O_p(N^{-1/2}\delta_{NT}^{-1})+O_p(N^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)$. $\mathbb{F}_{15}$ is equal to

$$
\mathbb{F}_{15}=-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{V}_k\boldsymbol{\eta}_k\boldsymbol{\eta}_k'\mathbf{V}_k'\widehat{\mathbf{H}}(\mathbf{H}^{0\prime}\widehat{\mathbf{H}}/T)^{-1}(\mathbf{\Phi}'\mathbf{\Phi}/N)^{-1}\boldsymbol{\phi}_i^0
$$

which can be proved to be $O_p(\delta_{NT}^{-2})+O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)$ by following the argument in the proof of $\mathbb{F}_{12}$. Collecting the above terms, we can show that the first term of (B.5) can be written as

$$
\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0=O_p(\delta_{NT}^{-2})+O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)+O_p(\delta_{NT}^{-1}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|). \tag{B.6}
$$

Consider the second term of (B.5). By arranging the terms, we have

$$
\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i=&\frac{1}{NT}\sum_{i=1}^{N}\mathbf{\Gamma}_i^{0\prime}\mathbf{S}'(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i+\frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\\
=&\frac{1}{NT}\sum_{i=1}^{N}\mathbf{\Gamma}_i^{0\prime}\mathbf{S}'(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\boldsymbol{\varepsilon}_i-\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{\Gamma}_i^{0\prime}\mathbf{S}'(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\widehat{\mathbf{H}}\mathbf{R}'\mathbf{H}^{0\prime}\boldsymbol{\varepsilon}_i\\
&+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{\Gamma}_i^{0\prime}\mathbf{S}'(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\widehat{\mathbf{H}}(\widehat{\mathbf{H}}-\mathbf{H}\mathbf{R})'\boldsymbol{\varepsilon}_i+\frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_i-\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{V}_i'\widehat{\mathbf{H}}\widehat{\mathbf{H}}'\boldsymbol{\varepsilon}_i.
\end{aligned}
$$

Then, using $\operatorname{vec}(\mathbf{ABC})=(\mathbf{C}'\otimes\mathbf{A})\operatorname{vec}(\mathbf{B})$ for any comfortable matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, the first expression can be written as

$$
\frac{1}{NT}\sum_{i=1}^{N}\operatorname{vec}\left(\mathbf{\Gamma}_i^{0\prime}\mathbf{S}'(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\boldsymbol{\varepsilon}_i\right)=\frac{1}{NT}\sum_{i=1}^{N}\left[\boldsymbol{\varepsilon}_i'\otimes(\mathbf{\Gamma}_i^{0\prime}\mathbf{S}')\right]\operatorname{vec}\left[(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\right]
$$

which is bounded in norm by

$$
\begin{aligned}
&\frac{1}{\sqrt{N}} \cdot \frac{1}{\sqrt{T}} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\varepsilon}_i' \otimes (\boldsymbol{\Gamma}_i^{0\prime} \mathbf{S}') \right\| \cdot \left\| \frac{1}{\sqrt{T}} \mathrm{vec}\left[(\mathbf{H}^0 - \widehat{\mathbf{H}}\mathbf{R}^{-1})' \right] \right\| \\
=&\frac{1}{\sqrt{N}} \cdot \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \varepsilon_{it} \boldsymbol{\Gamma}_i^{0\prime} \mathbf{S}' \right\|^2} \cdot \| T^{-1/2} (\mathbf{H}^0 - \widehat{\mathbf{H}}\mathbf{R}^{-1}) \| \\
=&O_p(N^{-1/2} \|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(N^{-1/2} \delta_{NT}^{-1}) \,.
\end{aligned}
$$

Analogously, we can prove that the third expression is $o_p(N^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + o_p(N^{-1/2}\delta_{NT}^{-1})$. It is easy to show that the second expression is $O_p(T^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(T^{-1/2}\delta_{NT}^{-2})$, the fourth expression is $O_p(N^{-1/2}T^{-1/2})$. The last expression is further decomposed into

$$
\begin{aligned}
&\frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i' \mathbf{H}\mathbf{R}\mathbf{R}'\mathbf{H}' \boldsymbol{\varepsilon}_i + \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i' (\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R}) \mathbf{R}'\mathbf{H}' \boldsymbol{\varepsilon}_i \\
&+ \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i' \mathbf{H}\mathbf{R} (\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})' \boldsymbol{\varepsilon}_i + \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i' (\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})(\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})' \boldsymbol{\varepsilon}_i
\end{aligned}
$$

which is equal to

$$
\begin{aligned}
&\frac{1}{NT} \sum_{i=1}^{N} \left[\boldsymbol{\varepsilon}_i' \otimes \left(T^{-1/2} \mathbf{V}_i' \mathbf{H} \right) \right] \cdot \mathrm{vec}\left(\mathbf{R}\mathbf{R}' T^{-1/2} \mathbf{H}' \right) + \frac{1}{NT} \sum_{i=1}^{N} \left[(T^{-1/2} \boldsymbol{\varepsilon}_i' \mathbf{H}) \otimes \mathbf{V}_i' \right] \cdot \mathrm{vec}\left(T^{-1/2} (\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R}) \mathbf{R}' \right) \\
&+ \frac{1}{NT} \sum_{i=1}^{N} \left[\boldsymbol{\varepsilon}_i' \otimes \left(T^{-1/2} \mathbf{V}_i' \mathbf{H} \right) \right] \cdot \mathrm{vec}\left(T^{-1/2} \mathbf{R} (\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})' \right) + \frac{1}{NT^2} \sum_{i=1}^{N} [\boldsymbol{\varepsilon}_i \otimes \mathbf{V}_i'] \cdot \mathrm{vec}\left((\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})(\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})' \right)
\end{aligned}
$$

then following the argument in the proof of the first expression, we can show that it is $O_p(N^{-1/2}T^{-1/2}) + O_p(N^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|^2) + O_p(N^{-1/2}\delta_{NT}^{-2})$. Then, we have

$$
\frac{1}{NT} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{H}}} \boldsymbol{\varepsilon}_i = O_p(\delta_{NT}^{-1} \|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(N^{-1/2}\delta_{NT}^{-1}) + O_p(T^{-1/2}\delta_{NT}^{-2}) \tag{B.7}
$$

Consider the third term of (B.5). Since $\mathbf{M}_{\widehat{\mathbf{H}}} - \mathbf{M}_{\mathbf{H}} = -T^{-1}(\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})\mathbf{R}'\mathbf{H}'^{-1}\mathbf{H}\mathbf{R}(\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})'^{-1}(\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})(\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})'^{-1}\mathbf{H}(\mathbf{R}\mathbf{R}' - (\mathbf{H}'\mathbf{H}/T)^{-1})\mathbf{H}'$, we have

$$
\begin{aligned}
&\frac{1}{NT} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{H}}} \mathbf{V}_i \boldsymbol{\eta}_i - \frac{1}{NT} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M}_{\mathbf{H}} \mathbf{V}_i \boldsymbol{\eta}_i \\
=& - \frac{1}{NT^2} \sum_{i=1}^{N} \boldsymbol{\Gamma}_i^{0\prime} \mathbf{F}^{0\prime} (\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R}) \mathbf{R}'\mathbf{H}' \mathbf{V}_i \boldsymbol{\eta}_i - \frac{1}{NT^2} \sum_{i=1}^{N} \boldsymbol{\Gamma}_i^{0\prime} \mathbf{F}^{0\prime} \mathbf{H}\mathbf{R} (\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})' \mathbf{V}_i \boldsymbol{\eta}_i \\
&- \frac{1}{NT^2} \sum_{i=1}^{N} \boldsymbol{\Gamma}_i^{0\prime} \mathbf{F}^{0\prime} (\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})(\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})' \mathbf{V}_i \boldsymbol{\eta}_i - \frac{1}{NT^2} \sum_{i=1}^{N} \boldsymbol{\Gamma}_i^{0\prime} \mathbf{F}^{0\prime} \mathbf{H} (\mathbf{R}\mathbf{R}' - (\mathbf{H}'\mathbf{H}/T)^{-1}) \mathbf{H}' \mathbf{V}_i \boldsymbol{\eta}_i \\
&- \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i' (\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R}) \mathbf{R}'\mathbf{H}' \mathbf{V}_i \boldsymbol{\eta}_i - \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i' \mathbf{H}\mathbf{R} (\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})' \mathbf{V}_i \boldsymbol{\eta}_i \\
&- \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i' (\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})(\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})' \mathbf{V}_i \boldsymbol{\eta}_i - \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i' \mathbf{H} (\mathbf{R}\mathbf{R}' - (\mathbf{H}'\mathbf{H}/T)^{-1}) \mathbf{H}' \mathbf{V}_i \boldsymbol{\eta}_i .
\end{aligned}
$$

Following the argument in the proof of the first expression of the second term, we can show that the first expression is $O_p(N^{-1/2}T^{-1/2}) \cdot T^{-1}\|\mathbf{H}^{0\prime}(\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R})\|$, which is $O_p(N^{-1/2}T^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(N^{-1/2}T^{-1/2}\delta_{NT}^{-2})$ by Lemma B.2(a); The second expression is $O_p(N^{-1/2}) \cdot T^{-1/2}\|\widehat{\mathbf{H}} - \mathbf{H}\mathbf{R}\|$, which is $O_p(N^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(N^{-1/2}\delta_{NT}^{-1})$ by Proposition 5; the third term is $O_p(N^{-1/2}) \cdot T^{-1}\|\mathbf{H}^{0\prime}(\widehat{\mathbf{H}} - \mathbf{HR})\| \cdot T^{-1/2}\|\widehat{\mathbf{H}} - \mathbf{HR}\|$, which is dominated by the second expression; the forth term is $O_p(N^{-1/2}T^{-1/2}) \cdot \|\mathbf{RR}' - (\mathbf{H}'\mathbf{H}/T)^{-1}\|$, which is $O_p(N^{-1/2}T^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(N^{-1/2}T^{-1/2}\delta_{NT}^{-2})$ by Lemma B.2(c); the fifth term and the sixth term both are $O_p(N^{-1/2}T^{-1/2}) \cdot T^{-1/2}\|\widehat{\mathbf{H}} - \mathbf{HR}\|$, which are dominated by the second term, thus omitted; the seventh term is $O_p(N^{-1/2}) \cdot T^{-1}\|\widehat{\mathbf{H}} - \mathbf{HR}\|^2$, which is also dominated by the second term; the eighth term is $O_p(N^{-1/2}T^{-1}) \cdot \|\mathbf{RR}' - (\mathbf{H}'\mathbf{H}/T)^{-1}\|$, which is dominated by the forth term. Then, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{V}_i\boldsymbol{\eta}_i = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i + O_p(N^{-1/2}\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(N^{-1/2}\delta_{NT}^{-1}). \tag{B.8}$$

In addition, with Lemma B.2 (f), we can derive that

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'(\mathbf{M}_{\widehat{\mathbf{H}}} - \mathbf{M}_{\mathbf{H}})\mathbf{X}_i = O_p(\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\|) + O_p(\delta_{NT}^{-1}) = o_p(1). \tag{B.9}$$

With equations (B.6) to (B.9), we can rewrite (B.5) as follows

$$\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i\right) + o_p(1)\right](\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i + O_p(\delta_{NT}^{-2})$$

which implies that

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i + O_p(\delta_{NT}^{-2}).$$

Then we have $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_p(N^{-1/2}) + O_p(\delta_{NT}^{-2})$. It is easy to show that

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i &= \frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{V}_i + O_p(T^{-1}),\\
\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i &= \frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{V}_i\boldsymbol{\eta}_i + +O_p(T^{-1}).
\end{aligned}$$

Thus, for the slope heterogeneous model, we can derive the following expression as given in Theorem 6

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{V}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{V}_i\boldsymbol{\eta}_i + O_p(\delta_{NT}^{-2}).$$

Next, we derive an asymptotic representation in slope heterogeneous case as given in Corollary 8. Since $\boldsymbol{\Upsilon}_\phi = N^{-1}\sum_{i=1}^{N}\boldsymbol{\phi}_i\boldsymbol{\phi}_i'$ and $\mathbf{Z}_i = \mathbf{X}_i - N^{-1}\sum_{j=1}^{N}\mathbf{X}_j\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_j$, we have

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i - \frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i\\
=&\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{X}_j'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_j = \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{V}_j'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_j\\
=&\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}(\mathbf{V}_j'\mathbf{V}_i)\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_j + \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbf{V}_j'\mathbf{V}_i - \mathbb{E}(\mathbf{V}_j'\mathbf{V}_i))\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_j\\
&-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{V}_j'\mathbf{H}(\mathbf{H}'\mathbf{H}/T)^{-1}\mathbf{H}'\mathbf{V}_i\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_j.
\end{aligned}$$

The first term is $N^{-2}T^{-1}\sum_{i=1}^N \mathbb{E}(\mathbf{V}_i'\mathbf{V}_i)\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_i$ by Assumption B2(i), which is bounded in norm by $\Delta N^{-2}\sum_{i=1}^N \|\boldsymbol{\phi}_i\|^2\|\boldsymbol{\Upsilon}_\phi^{-1}\| = O_p(N^{-1})$ by Assumption B2(iv). The second term is $O_p(N^{-1}T^{-1/2})$. The third term is bounded by $T^{-1}\cdot(N^{-1}\sum_{i=1}^N \|T^{-1/2}\mathbf{H}'\mathbf{V}_i\|\|\boldsymbol{\phi}_i\|)^2\|\boldsymbol{\Upsilon}_\phi^{-1}\|\|(\mathbf{H}'\mathbf{H}/T)^{-1}\| = O_p(T^{-1})$. With the above three terms, we have

$$\frac{1}{NT}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{M_H}\mathbf{Z}_i - \frac{1}{NT}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M_H}\mathbf{X}_i = O_p(N^{-1}) + O_p(T^{-1})$$

which implies that

$$\left(\frac{1}{NT}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{M_H}\mathbf{Z}_i\right)^{-1} - \left(\frac{1}{NT}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M_H}\mathbf{X}_i\right)^{-1} = O_p(N^{-1}) + O_p(T^{-1}).$$

With this equation and the fact that $(NT)^{-1}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M_H}\mathbf{V}_i\boldsymbol{\eta}_i = O_p(N^{-1/2})$, we can derive that

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\frac{1}{NT}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{M_H}\mathbf{Z}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M_H}\mathbf{V}_i\boldsymbol{\eta}_i + O_p(\delta_{NT}^{-2}). \tag{B.10}$$

Furthermore, we can show that

$$\begin{aligned}
&\frac{1}{N^2T}\sum_{i=1}^N\sum_{\ell=1}^N \boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell\mathbf{X}_\ell'\mathbf{M_H}\mathbf{V}_i\boldsymbol{\eta}_i\\
=&\mathrm{vec}\left(\frac{1}{N^2T}\sum_{i=1}^N\sum_{\ell=1}^N \boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell\mathbf{V}_\ell'\mathbf{V}_i\boldsymbol{\eta}_i\right) - \frac{1}{N^2T}\sum_{i=1}^N\sum_{\ell=1}^N \boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell\mathbf{V}_\ell'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{V}_i\boldsymbol{\eta}_i\\
=&\frac{1}{N}\cdot\frac{1}{\sqrt{NT}}\sum_{i=1}^N\left[(\boldsymbol{\eta}_i'\mathbf{V}_i')\otimes\boldsymbol{\phi}_i'\right]\cdot\mathrm{vec}\left[\boldsymbol{\Upsilon}_\phi^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{\ell=1}^N \boldsymbol{\phi}_\ell\mathbf{V}_\ell'\right)\right]\\
&-\frac{1}{\sqrt{NT}}\cdot\frac{1}{N}\sum_{i=1}^N \boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{\ell=1}^N \boldsymbol{\phi}_\ell\mathbf{V}_\ell'\mathbf{H}\right)(\mathbf{H}'\mathbf{H}/T)^{-1}(T^{-1/2}\mathbf{H}'\mathbf{V}_i)\boldsymbol{\eta}_i\\
=&O_p(N^{-1}) + O_p(N^{-1/2}T^{-1}).
\end{aligned}$$

It is easy to show that $N^{-2}T^{-1}\sum_{i=1}^N\sum_{\ell=1}^N \boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell\mathbf{X}_\ell'\mathbf{M_H}\boldsymbol{\varepsilon}_i = O_p(N^{-1/2}T^{-1/2})$ and $N^{-1}T^{-1}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M_H}\boldsymbol{\varepsilon}_i = O_p(N^{-1/2}T^{-1/2})$. With (B.10) and the above facts, we can derive that

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\frac{1}{NT}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{M_H}\mathbf{Z}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{M_H}\mathbf{e}_i + O_p(\delta_{NT}^{-2})$$

as $\mathbf{Z}_i = \mathbf{X}_i - N^{-1}\sum_{\ell=1}^N \boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell\mathbf{X}_\ell$.

Lastly, we consider the case in which the panel's slope is homogeneous. Then $\boldsymbol{\eta}_i = 0$, which implies that $\mathbf{H}^0 = \mathbf{F}^0$, $\boldsymbol{\phi}_i^0 = \boldsymbol{\lambda}_i^0$ and $\mathbf{e}_i = \boldsymbol{\varepsilon}_i$. By Proposition A.3 in Bai (2009a), $\widehat{\boldsymbol{\beta}}$ has the following asymptotic representation

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} =& \left(\frac{1}{NT}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{M_H}\mathbf{Z}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{M_H}\mathbf{e}_i\\
&-\left(\frac{1}{NT}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{M_H}\mathbf{Z}_i\right)^{-1}\frac{1}{N^2T}\sum_{i=1}^N\sum_{t=1}^T \mathbf{Z}_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_j\mathbb{E}e_{it}^2\\
&-\left(\frac{1}{NT}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{M_H}\mathbf{Z}_i\right)^{-1}\frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N \mathbf{X}_i'\mathbf{M_H}\mathbb{E}(\mathbf{e}_j\mathbf{e}_j')\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_i + O_p(\delta_{NT}^{-3})\\
=&\left(\frac{1}{NT}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{M_H}\mathbf{Z}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{M_H}\mathbf{e}_i + \frac{1}{N}\boldsymbol{\xi}_{NT} + \frac{1}{T}\boldsymbol{\zeta}_{NT} + O_p(\delta_{NT}^{-3}).
\end{aligned}$$

Thus, we complete the proof. ■

Lemma B.3 *Under Assumptions B1 to B4, when the slopes are heterogeneous, we have*

$$(a)\quad N^{-1}\|\widehat{\mathbf{\Phi}}-\mathbf{\Phi}\mathbf{R}^{-1\prime}\|^2=\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\|^2=O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)+O_p(\delta_{NT}^{-2})\,,$$

$$(b)\quad N^{-1}(\widehat{\mathbf{\Phi}}-\mathbf{\Phi}\mathbf{R}^{-1\prime})'\mathbf{\Phi}=\frac{1}{N}\sum_{i=1}^{N}(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0)\boldsymbol{\phi}_i^{0\prime}=O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-2})\,,$$

$$(c)\quad \widehat{\mathbf{\Phi}}'\widehat{\mathbf{\Phi}}/N-\mathbf{R}^{-1}(\mathbf{\Phi}'\mathbf{\Phi}/N)\mathbf{R}^{-1\prime}=O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-2})\,,$$

$$(d)\quad (\widehat{\mathbf{\Phi}}'\widehat{\mathbf{\Phi}}/N)^{-1}-\mathbf{R}'(\mathbf{\Phi}'\mathbf{\Phi}/N)^{-1}\mathbf{R}=O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-2})\,,$$

$$(e)\quad \frac{1}{N}\sum_{i=1}^{N}\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\|=O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-1})\,,$$

$$(f)\quad \frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\|=O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-1})\,,$$

Proof of Lemma B.3. Since $\mathbf{y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}}=\mathbf{H}^0\boldsymbol{\phi}_i^0+\mathbf{V}_i\boldsymbol{\eta}_i+\boldsymbol{\varepsilon}_i+\mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})$, we have

$$\begin{aligned}\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0=&T^{-1}\widehat{\mathbf{H}}'(\mathbf{y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}})-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\\=&T^{-1}\widehat{\mathbf{H}}'\mathbf{H}^0\boldsymbol{\phi}_i^0-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0+T^{-1}\widehat{\mathbf{H}}'\mathbf{V}_i\boldsymbol{\eta}_i+T^{-1}\widehat{\mathbf{H}}'\boldsymbol{\varepsilon}_i+T^{-1}\widehat{\mathbf{H}}'\mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})\\=&T^{-1}\widehat{\mathbf{H}}'(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})\boldsymbol{\phi}_i^0+T^{-1}\widehat{\mathbf{H}}'\mathbf{V}_i\boldsymbol{\eta}_i+T^{-1}\widehat{\mathbf{H}}'\boldsymbol{\varepsilon}_i+T^{-1}\widehat{\mathbf{H}}'\mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})\end{aligned}$$

with $\mathbf{H}^0=(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})+\widehat{\mathbf{H}}\mathbf{R}^{-1}$. For (a), we have,

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\|^2\leq&4\|T^{-1}\widehat{\mathbf{H}}'(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})\|^2\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{\phi}_i^0\|^2+\frac{4}{N}\sum_{i=1}^{N}\|T^{-1}\widehat{\mathbf{H}}'\mathbf{V}_i\boldsymbol{\eta}_i\|^2\\&+\frac{4}{N}\sum_{i=1}^{N}\|T^{-1}\widehat{\mathbf{H}}'\boldsymbol{\varepsilon}_i\|^2+\frac{4}{N}\sum_{i=1}^{N}\|T^{-1}\widehat{\mathbf{H}}'\mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})\|^2.\end{aligned}$$

Hereafter we omit the scale 4. With Lemma B.2(b), the first term is $O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)+O_p(\delta_{NT}^{-4})$. The second term is bounded in norm by

$$\begin{aligned}&\frac{1}{T}\|\mathbf{R}\|^2\cdot\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{H}'\mathbf{V}_i\boldsymbol{\eta}_i\|^2+\|T^{-1/2}(\widehat{\mathbf{H}}-\mathbf{H}\mathbf{R})\|^2\cdot\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{V}_i\boldsymbol{\eta}_i\|^2\\=&O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)+O_p(\delta_{NT}^{-2}).\end{aligned}$$

Following the proof of the second term, we can prove that the third term is $O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)+O_p(\delta_{NT}^{-2})$. It is easy to show that the forth term is $O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)$. Thus,

$$\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\|^2=O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|^2)+O_p(\delta_{NT}^{-2}).$$

Consider (b), we have

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0)\boldsymbol{\phi}_i^{0\prime}=&\frac{1}{T}\widehat{\mathbf{H}}'(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})\cdot\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i^{0\prime}+\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{H}}'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\phi}_i^{0\prime}\\&+\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{H}}'\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i^{0\prime}+\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{H}}'\mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})\boldsymbol{\phi}_i^{0\prime}.\end{aligned}$$

The first term is $O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-2})$ by Lemma B.2(b). By decomposition $\widehat{\mathbf{H}}=(\widehat{\mathbf{H}}-\mathbf{HR})+\mathbf{HR}$, we can derive that the second and the third terms is $O_p(N^{-1/2}T^{-1/2})+O_p(N^{-1/2}\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(N^{-1/2}\delta_{NT}^{-1})$. It is easy to show that the forth term is $O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)$. With the above four terms, we can derive that

$$\frac{1}{N}\sum_{i=1}^{N}(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0)\boldsymbol{\phi}_i^{0\prime}=O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-2}).$$

By adding and substracting terms, (a) and (b), we have (c) and (d). For (e), we have,

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\| \le &\|T^{-1}\widehat{\mathbf{H}}'(\mathbf{H}^0-\widehat{\mathbf{H}}\mathbf{R}^{-1})\|\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{\phi}_i^0\|+\frac{1}{N}\sum_{i=1}^{N}\|T^{-1}\widehat{\mathbf{H}}'\mathbf{V}_i\boldsymbol{\eta}_i\| \\ &+\frac{1}{N}\sum_{i=1}^{N}\|T^{-1}\widehat{\mathbf{H}}'\boldsymbol{\varepsilon}_i\|+\frac{1}{N}\sum_{i=1}^{N}\|T^{-1}\widehat{\mathbf{H}}'\mathbf{X}_i(\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}})\|.\end{aligned}$$

The first term is $O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-2})$ by Lemma B.2(b). The second term is bounded in norm by

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}\|T^{-1}\mathbf{R}'\mathbf{H}'\mathbf{V}_i\boldsymbol{\eta}_i\|+\frac{1}{N}\sum_{i=1}^{N}\|T^{-1}(\widehat{\mathbf{H}}-\mathbf{HR})'\mathbf{V}_i\boldsymbol{\eta}_i\| \\ \le &T^{-1/2}\|\mathbf{R}\|\cdot\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{H}'\mathbf{V}_i\boldsymbol{\eta}_i\|+\left\|T^{-1/2}(\widehat{\mathbf{H}}-\mathbf{HR})'\right\|\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{V}_i\boldsymbol{\eta}_i\| \\ =&O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-1})\end{aligned}$$

by Proposition 5. Similarly, we can derive that the third term is $O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-1})$. It is easy to show that the fourth term is $O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)$. Then

$$\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\|=O_p(\|\boldsymbol{\beta}^0-\widehat{\boldsymbol{\beta}}\|)+O_p(\delta_{NT}^{-1}).$$

Analogously, we can show that (f). Thus, we complete the proof. ■

Proof of Theorem 9.

For the slope homogeneous case, we refer the proof to Bai (2009a). Then, it is sufficient to prove the theorem in the slopes heterogeneous case. Consider the heterogeneous slope models. We decompose $N^{-1}T^{-1}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i$ into the following five terms, that is

$$\begin{aligned}&\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i-\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i \\ =&\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\mathbf{X}_i-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell\mathbf{X}_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\mathbf{X}_\ell \\ &-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0)'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_\ell \\ &-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\boldsymbol{\phi}_i'\mathbf{R}^{-1\prime}\left(\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}-\mathbf{R}'(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\mathbf{R}\right)\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_\ell \\ &-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\boldsymbol{\phi}_i'(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\mathbf{R}(\widehat{\boldsymbol{\phi}}_\ell-\mathbf{R}^{-1}\boldsymbol{\phi}_\ell)\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_\ell.\end{aligned}$$

The first term is bounded in norm by

$$\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|^2\cdot\|\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}}\| = O_p(\delta_{NT}^{-1}).$$

Analogously, we can show that the second term is $O_p(\delta_{NT}^{-1})$. Note that $\|\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_\ell\| \leq \|\mathbf{X}_\ell\|$ and $\|\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\| = O_p(1)$, which is implied by Lemma B.3(d). Then the third term is bounded in norm by

$$\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\|\|T^{-1/2}\mathbf{X}_i\|\cdot\frac{1}{N}\sum_{\ell=1}^{N}\|T^{-1/2}\mathbf{X}_\ell\|\|\widehat{\boldsymbol{\phi}}_\ell\|\cdot\|\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\| = O_p(\delta_{NT}^{-1})$$

with Lemma B.3(f). Analogously, we can show that the fifth term is $O_p(\delta_{NT}^{-1})$. With Lemma B.3(d), we can derive that the forth term is $O_p(\delta_{NT}^{-1})$. Combining the above terms, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i-\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i = O_p(\delta_{NT}^{-1}).$$

With the above equation, we can show that

$$\left(\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\right)^{-1}-\left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i\right)^{-1} = O_p(\delta_{NT}^{-1}). \tag{B.11}$$

Thus, to investigate the stochastic orders of $\widehat{\boldsymbol{\xi}}_{NT}$ and $\widehat{\boldsymbol{\zeta}}_{NT}$ are $O_p(1)$, it is sufficient to focus on the stochastic orders of two terms $-N^{-1}T^{-2}\sum_{i=1}^N\sum_{t=1}^T\widehat{\mathbf{Z}}_i'\widehat{\mathbf{H}}\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_i\widehat{e}_{it}^2$ and $-N^{-1}T^{-1}\sum_{i=1}^N\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\boldsymbol{\Omega}}\widehat{\mathbf{H}}\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_i$. Specifically, we have

$$\begin{aligned}
\widehat{\boldsymbol{\xi}}_{NT}^\dagger =& -\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\mathbf{Z}}_i'\widehat{\mathbf{H}}\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_i\widehat{e}_{it}^2\\
=& -\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\widehat{\mathbf{H}}\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_i\widehat{e}_{it}^2+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\sum_{t=1}^{T}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell'\widehat{\mathbf{H}}\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_i\widehat{e}_{it}^2\\
=&\widehat{\boldsymbol{\xi}}_{1NT}^\dagger+\widehat{\boldsymbol{\xi}}_{2NT}^\dagger
\end{aligned}$$

and

$$\begin{aligned}
\widehat{\boldsymbol{\zeta}}_{NT}^\dagger =& -\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\boldsymbol{\Omega}}\widehat{\mathbf{H}}\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_i = -\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\widehat{\boldsymbol{\Omega}}\widehat{\mathbf{H}}\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_i+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{X}_i'\widehat{\mathbf{H}}\widehat{\mathbf{H}}'\widehat{\boldsymbol{\Omega}}\widehat{\mathbf{H}}\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_i\\
=& -\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\widehat{\mathbf{h}}_t'\widehat{e}_{jt}^2\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_i+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\widehat{\mathbf{H}}\widehat{\mathbf{h}}_t\widehat{\mathbf{h}}_t'\widehat{e}_{jt}^2\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_i\\
=&\widehat{\boldsymbol{\zeta}}_{1NT}^\dagger+\widehat{\boldsymbol{\zeta}}_{2NT}^\dagger.
\end{aligned}$$

For the term $\widehat{\boldsymbol{\xi}}^{\dagger}_{1NT}$, which has the following decomposition (ignoring the sign)

$$
\begin{aligned}
&\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\widehat{\mathbf{H}}\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_i\widehat{e}_{it}^2-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_i\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2\\
=&\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'(\widehat{\mathbf{H}}-\mathbf{H}\mathbf{R})\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_i\widehat{e}_{it}^2+\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\mathbf{H}\mathbf{R}(\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}-\mathbf{R}'\boldsymbol{\Upsilon}_{\phi}^{-1}\mathbf{R})\widehat{\boldsymbol{\phi}}_i\widehat{e}_{it}^2\\
&+\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\mathbf{H}\mathbf{R}\mathbf{R}'\boldsymbol{\Upsilon}_{\phi}^{-1}\mathbf{R}(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0)\widehat{e}_{it}^2\\
&+\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\mathbf{H}\mathbf{R}\mathbf{R}'\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_i^0\left[\widehat{e}_{it}^2-\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2\right]\\
&+\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\mathbf{H}\left[\mathbf{R}\mathbf{R}'-(\mathbf{H}'\mathbf{H}/T)^{-1}\right]\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_i^0\left[\widehat{e}_{it}^2-\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2\right]\\
\equiv&\mathbb{G}_1+\mathbb{G}_2+\mathbb{G}_3+\mathbb{G}_4+\mathbb{G}_5.
\end{aligned}
$$

Using that $\widehat{e}_{it}=\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it}-\mathbf{x}_{it}'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0)-(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0)'\widehat{\mathbf{h}}_t-\boldsymbol{\phi}_i^{0\prime}\mathbf{R}^{-1\prime}(\widehat{\mathbf{h}}_t-\mathbf{R}'\mathbf{h}_t^0)$, we have

$$
\begin{aligned}
&\left\|\frac{1}{T}\sum_{t=1}^{T}\widehat{e}_{it}^2-\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2\right\|\\
\leq&\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{x}_{it}'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0)-(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0)'\widehat{\mathbf{h}}_t-\boldsymbol{\phi}_i^{0\prime}\mathbf{R}^{-1\prime}(\widehat{\mathbf{h}}_t-\mathbf{R}'\mathbf{h}_t^0)\right]^2\right\|\\
&+2\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{x}_{it}'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0)-(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0)'\widehat{\mathbf{h}}_t-\boldsymbol{\phi}_i^{0\prime}\mathbf{R}^{-1\prime}(\widehat{\mathbf{h}}_t-\mathbf{R}'\mathbf{h}_t^0)\right](\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})\right\|\\
&+\left\|\frac{1}{T}\sum_{t=1}^{T}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2-\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2\right\|\\
\leq&\frac{3}{T}\sum_{t=1}^{T}\|\mathbf{x}_{it}\|^2\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|^2+\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\|^2\cdot\frac{3}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t\|^2\\
&+\|\boldsymbol{\phi}_i^0\|^2\|\mathbf{R}^{-1}\|^2\cdot\frac{3}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t-\mathbf{R}'\mathbf{h}_t^0\|^2+\frac{2}{T}\sum_{t=1}^{T}\|\mathbf{x}_{it}\|\|\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it}\|\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|\\
&+\frac{2}{T}\left\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\right\|\left\|\sum_{t=1}^{T}\mathbf{h}_t^0(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})\right\|\cdot\|\mathbf{R}\|\\
&+\frac{2}{T}\|\widehat{\boldsymbol{\phi}}_i\|\|\boldsymbol{\eta}_i\|\cdot\left\|\sum_{t=1}^{T}(\widehat{\mathbf{h}}_t-\mathbf{R}'\mathbf{h}_t^0)\mathbf{v}_{it}'^{-1}\right\|\|\widehat{\boldsymbol{\phi}}_i\|\left\|\sum_{t=1}^{T}(\widehat{\mathbf{h}}_t-\mathbf{R}'\mathbf{h}_t^0)\varepsilon_{it}\right\|\\
&+\left\|\frac{1}{T}\sum_{t=1}^{T}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2-\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2\right\|.
\end{aligned}\tag{B.12}
$$

Consider the term $\mathbb{G}_4$, which is bounded in norm by

$$
\begin{aligned}
&\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\|\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\widehat{e}_{it}^2-\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2\right]\right\|\cdot\|\mathbf{R}\|^2\|\boldsymbol{\Upsilon}_{\phi}^{-1}\|\|T^{-1/2}\mathbf{H}\|\\
=&\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\widehat{e}_{it}^2-\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2\right]\right\|\cdot O_p(1).
\end{aligned}
$$

By plugging (B.12) into the above equation, $\mathbb{G}_4$ is further bounded in norm by

$$\frac{3}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|^3\|\boldsymbol{\phi}_i^0\|\cdot\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|^2+\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\|^2\cdot\frac{3}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t\|^2$$
$$+\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|^3\|\mathbf{R}^{-1}\|^2\cdot\frac{3}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t-\mathbf{R}'\mathbf{h}_t^0\|^2$$
$$+\frac{2}{NT}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|\sum_{t=1}^{T}\|\mathbf{x}_{it}\|\|\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it}\|\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|$$
$$+\frac{2}{N\sqrt{T}}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\|\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{h}_t^0(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})\right\|\cdot\|\mathbf{R}\|$$
$$+\frac{2}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|\|\widehat{\boldsymbol{\phi}}_i\|\|\boldsymbol{\eta}_i\|\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{v}_{it}\|^2}\cdot\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t-\mathbf{R}'\mathbf{h}_t^0\|^2}$$
$$+\frac{2}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|\|\widehat{\boldsymbol{\phi}}_i\|\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\varepsilon_{it}\|^2}\cdot\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t-\mathbf{R}'\mathbf{h}_t^0\|^2}$$
$$+\frac{1}{N}\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2\right\|.$$

The first term is $O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|^2)$, the third term is $O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|^2)+O_p(\delta_{NT}^{-2})$ by Proposition 5, the fourth term is $O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)$, the eighth term is $O_p(T^{-1/2})$. Following the proof of Lemma B.3 (a) and (e), we show that

$$\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\|^2=O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|^2)+O_p(\delta_{NT}^{-2}),$$
$$\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i^0\|\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0\|\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{h}_t^0(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})\right\|=O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1}).$$

Then the second term is $O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|^2)+O_p(\delta_{NT}^{-2})$, the fifth term is $O_p(T^{-1/2}\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(T^{-1/2}\delta_{NT}^{-1})$. It is easy to show that the sixth term and the seventh term both are $O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1})$. With the above terms, $\mathbb{G}_4=O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1})$. With Lemma B.2(c), we can show that $\mathbb{G}_5=O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|^2)+O_p(\delta_{NT}^{-1}\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-3})$. The term $\mathbb{G}_3$ is decomposed into

$$\mathbb{G}_3=\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\mathbf{H}\mathbf{R}\mathbf{R}'\boldsymbol{\Upsilon}_\phi^{-1}\mathbf{R}(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0)\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2$$
$$+\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\mathbf{H}\mathbf{R}\mathbf{R}'\boldsymbol{\Upsilon}_\phi^{-1}\mathbf{R}(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i^0)\left[\widehat{e}_{it}^2-\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2\right].$$

The first term can be shown to be $O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1})$, by following the argument in the proof of Lemma B.3(e). The second term will be proved to be $O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|^2)+O_p(\delta_{NT}^{-1}\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-2})$, by following the argument in the proof of $\mathbb{G}_4$. $\mathbb{G}_2$ is decomposed into

$$\mathbb{G}_2=\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\mathbf{H}\mathbf{R}(\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}-\mathbf{R}'\boldsymbol{\Upsilon}_\phi^{-1}\mathbf{R})\widehat{\boldsymbol{\phi}}_i\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2$$
$$+\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\mathbf{H}\mathbf{R}(\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}-\mathbf{R}'\boldsymbol{\Upsilon}_\phi^{-1}\mathbf{R})\widehat{\boldsymbol{\phi}}_i\left[\widehat{e}_{it}^2-\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2\right].$$

The first term is bounded in norm by

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|T^{-1/2}\mathbf{X}_i\|\|\widehat{\boldsymbol{\phi}}_i\|\|\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2\|\cdot\|T^{-1/2}\mathbf{H}\|\|\mathbf{R}\|\|\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}-\mathbf{R}'\boldsymbol{\Upsilon}_{\phi}^{-1}\mathbf{R}\|=O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-2})$$

with Lemma B.3(c). Following the argument in the proof of $\mathbb{G}_4$, the second term is shown to be $o_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+o_p(\delta_{NT}^{-2})$. Then $\mathbb{G}_2=O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-2})$. Similarly, we can show that $\mathbb{G}_1=O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1})$. Thus, we derive that

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\widehat{\mathbf{H}}\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_i\widehat{e}_{it}^2-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_i\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2=O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1}).$$

Analogously, we can prove that

$$\begin{aligned}&\frac{1}{N^2\sqrt{N}}\frac{1}{T\sqrt{T}}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\sum_{t=1}^{T}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell'\widehat{\mathbf{H}}\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_i\widehat{e}_{it}^2\\&-\frac{1}{N^2\sqrt{N}}\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\sum_{t=1}^{T}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_i\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it}^2)\\=&O_p(N^{-1/2}T^{1/2}\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(N^{-1/2}T^{1/2}\delta_{NT}^{-1}).\end{aligned}$$

Following the argument in the proof Lemma A.11 of Bai (2009), we can derive that

$$\begin{aligned}&\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\sum_{t=1}^{T}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_i\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it}^2)\\&-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\sum_{t=1}^{T}\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_\ell\mathbf{X}_\ell'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_i\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it}^2)\\=&O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1}).\end{aligned}$$

Thus, we can also derive that

$$\begin{aligned}&\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\sum_{t=1}^{T}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell'\widehat{\mathbf{F}}\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_i\widehat{e}_{it}^2\\&-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\sum_{t=1}^{T}\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_\ell\mathbf{X}_\ell'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_i\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it}^2)\\=&O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1}).\end{aligned}$$

With the above all terms, we have

$$\begin{aligned}&\widehat{\boldsymbol{\xi}}_{1NT}^{\dagger}+\widehat{\boldsymbol{\xi}}_{2NT}^{\dagger}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_i\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})^2\\&-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\sum_{t=1}^{T}\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_\ell\mathbf{X}_\ell'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_i\mathbb{E}(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it}^2)\\=&O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1})=O_p(\delta_{NT}^{-1}).\end{aligned}$$

Combining the above facts, the bias term $\widehat{\boldsymbol{\xi}}_{NT}=O_p(1)$.

Next, we consider $\widehat{\boldsymbol{\zeta}}_{NT}$ where there is no serial correlation. First, note that $\widehat{\boldsymbol{\zeta}}_{1NT}^{\dagger}$ has the following decomposition (ignoring the sign)

$$
\begin{aligned}
&\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\widehat{\mathbf{h}}_t'\widehat{e}_{jt}^2\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_i-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{h}_t'\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2(\mathbf{H}'\mathbf{H}/T)^{-1}\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_i\\
=&\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\widehat{\mathbf{h}}_t'(\widehat{e}_{jt}^2-\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2)\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_i\\
&+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}(\widehat{\mathbf{h}}_t-\mathbf{R}'\mathbf{h}_t)'\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_i\\
&+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{h}_t'\mathbf{R}\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2(\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}-\mathbf{R}'\boldsymbol{\Upsilon}_\phi^{-1}\mathbf{R})\widehat{\boldsymbol{\phi}}_i\\
&+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{h}_t'\mathbf{R}\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2\mathbf{R}'\boldsymbol{\Upsilon}_\phi^{-1}\mathbf{R}(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i)\\
&+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{h}_t'\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2(\mathbf{R}\mathbf{R}'-(\mathbf{H}'\mathbf{H}/T)^{-1})\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_i\\
=&\mathbb{H}_1+\cdots+\mathbb{H}_5.
\end{aligned}
$$

Consider $\mathbb{H}_3$, which is bounded in norm by

$$
\begin{aligned}
&\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\|\mathbf{h}_t\|\left(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_{it}\|^2\right)^{1/2}\|\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_i+\varepsilon_{jt})^2\|\cdot\sqrt{\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\boldsymbol{\phi}}_i\|^2\|\mathbf{R}\|}\left\|\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}-\mathbf{R}'\boldsymbol{\Upsilon}_\phi^{-1}\mathbf{R}\right\|\\
=&O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-2})
\end{aligned}
$$

with Lemma B.3(c), then $\mathbb{H}_3=O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-2})$. Analogously, we can show that $\mathbb{H}_5=O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-2})$. Next, note that, analogous to (B.12), we have

$$
\begin{aligned}
&\left\|\frac{1}{N}\sum_{j=1}^{N}\widehat{e}_{jt}^2-\frac{1}{N}\sum_{j=1}^{T}\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2\right\|\\
\leq&\frac{3}{N}\sum_{j=1}^{N}\|\mathbf{x}_{jt}\|^2\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|^2+\frac{3}{N}\sum_{j=1}^{N}\|\widehat{\boldsymbol{\phi}}_j-\mathbf{R}^{-1}\boldsymbol{\phi}_j^0\|^2\cdot\|\widehat{\mathbf{h}}_t\|^2\\
&+\frac{3}{N}\sum_{j=1}^{N}\|\boldsymbol{\phi}_j^0\|^2\|\mathbf{R}^{-1}\|^2\cdot\|\widehat{\mathbf{h}}_t-\mathbf{R}'\mathbf{h}_t^0\|^2+\frac{2}{N}\sum_{j=1}^{N}\|\mathbf{x}_{jt}\|\|\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt}\|\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|\\
&+\frac{2}{N}\|\widehat{\mathbf{h}}_t-\mathbf{R}'\mathbf{h}_t^0\|\left\|\sum_{j=1}^{N}\boldsymbol{\phi}_j^0(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})\right\|\cdot\|\mathbf{R}^{-1}\|\\
&+\frac{2}{N}\|\widehat{\mathbf{h}}_t\|\cdot\left\|\sum_{j=1}^{N}(\widehat{\boldsymbol{\phi}}_j-\mathbf{R}^{-1}\boldsymbol{\phi}_j^0)\mathbf{v}_{jt}'\boldsymbol{\eta}_j\right\|+\frac{2}{N}\|\widehat{\mathbf{h}}_t\|\left\|\sum_{j=1}^{N}(\widehat{\boldsymbol{\phi}}_j-\mathbf{R}^{-1}\boldsymbol{\phi}_j^0)\varepsilon_{jt}\right\|\\
&+\left\|\frac{1}{N}\sum_{j=1}^{N}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2-\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2\right\|.
\end{aligned}
\tag{B.13}
$$

Consider $\mathbb{H}_1$, which is bounded in norm by

$$\begin{aligned}
&\frac{1}{N^2T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t\|\left\|\sum_{j=1}^{N}(\widehat{e}_{jt}^2-\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2)\right\|\sum_{i=1}^{N}\|\mathbf{x}_{it}\|\|\widehat{\boldsymbol{\phi}}_i\|\cdot\|\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\| \\
\leq&\frac{1}{NT}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t\|\sqrt{\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_{it}\|^2}\left\|\sum_{j=1}^{N}(\widehat{e}_{jt}^2-\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2)\right\|\cdot\left\|\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\right\|\sqrt{\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\boldsymbol{\phi}}_i\|^2} \\
=&\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t\|\sqrt{\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_{it}\|^2}\left\|\frac{1}{N}\sum_{j=1}^{N}(\widehat{e}_{jt}^2-\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2)\right\|\cdot O_p(1)
\end{aligned}$$

with (B.13), it is further bounded in norm by

$$\begin{aligned}
&\frac{3}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t\|\left(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_{it}\|^2\right)^{3/2}\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|^2 \\
&+\frac{3}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t\|^3\left(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_{it}\|^2\right)^{1/2}\cdot\frac{1}{N}\sum_{j=1}^{N}\|\widehat{\boldsymbol{\phi}}_j-\mathbf{R}^{-1}\boldsymbol{\phi}_j^0\|^2 \\
&+\frac{3}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t-\mathbf{R}'\mathbf{h}_t^0\|^2\|\widehat{\mathbf{h}}_t\|\left(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_{it}\|^2\right)^{1/2}\cdot\frac{1}{N}\sum_{j=1}^{N}\|\boldsymbol{\phi}_j^0\|^2\|\mathbf{R}^{-1}\|^2 \\
&+\frac{2}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t\|\left(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_{it}\|^2\right)^{1/2}\frac{1}{N}\sum_{j=1}^{N}\|\mathbf{x}_{jt}\|\|\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt}\|\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\| \\
&+\frac{2}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t\|\|\widehat{\mathbf{h}}_t-\mathbf{R}'\mathbf{h}_t^0\|\left(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_{it}\|^2\right)^{1/2}\left\|\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\phi}_j^0(\mathbf{v}_{it}'\boldsymbol{\eta}_i+\varepsilon_{it})\right\|\cdot\|\mathbf{R}^{-1}\| \\
&+\frac{2}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t\|^2\left(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_{it}\|^2\right)^{1/2}\sqrt{\frac{1}{N}\sum_{j=1}^{N}\|\mathbf{v}_{jt}\|^2\|\boldsymbol{\eta}_j\|^2}\cdot\sqrt{\frac{1}{N}\sum_{j=1}^{N}\|\widehat{\boldsymbol{\phi}}_j-\mathbf{R}^{-1}\boldsymbol{\phi}_j^0\|^2} \\
&+\frac{2}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t\|^2\left(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_{it}\|^2\right)^{1/2}\sqrt{\frac{1}{N}\sum_{j=1}^{N}\|\varepsilon_{jt}\|^2}\cdot\sqrt{\frac{1}{N}\sum_{j=1}^{N}\|\widehat{\boldsymbol{\phi}}_j-\mathbf{R}^{-1}\boldsymbol{\phi}_j^0\|^2} \\
&+\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{h}}_t\|\left(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_{it}\|^2\right)^{1/2}\left\|\frac{1}{N}\sum_{j=1}^{N}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2-\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2\right\|.
\end{aligned}\tag{B.14}$$

Following the arguments in the proof of $\mathbb{G}_4$, we can show the terms in the above equation is $O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1})$, then $\mathbb{H}_1=O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1})$. Analogous to the arguments in the proof of $\mathbb{G}_1$, $\mathbb{H}_2=O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1})$. Following the argument in the proof $\mathbb{G}_3$, $\mathbb{H}_4=O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1})$. Collecting the above terms, we can derive that

$$\widehat{\boldsymbol{\zeta}}_{1NT}^{\dagger}-\left(-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{h}_t'\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2(\mathbf{H}'\mathbf{H}/T)^{-1}\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_i\right)=O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1}).$$

Analogously, we can derive that

$$\begin{aligned}
&\widehat{\boldsymbol{\zeta}}_{2NT}^{\dagger}-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\mathbf{H}(\mathbf{H}'\mathbf{H}/T)^{-1}\mathbf{h}_t\mathbf{h}_t'\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_i \\
=&O_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\|)+O_p(\delta_{NT}^{-1}).
\end{aligned}$$

Note that it is easy to show that $N^{-2}T^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{h}_t'\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2(\mathbf{H}'\mathbf{H}/T)^{-1}\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_i = O_p(1)$ and $N^{-2}T^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\mathbf{H}(\mathbf{H}'\mathbf{H}/T)^{-1}\mathbf{h}_t\mathbf{h}_t'\mathbb{E}(\mathbf{v}_{jt}'\boldsymbol{\eta}_j+\varepsilon_{jt})^2\boldsymbol{\Upsilon}_{\phi}^{-1}\boldsymbol{\phi}_i = O_p(1)$. Collecting the above facts, we obtain that

$$\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\widehat{\mathbf{h}}_t'\widehat{e}_{jt}^2\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_i - \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbf{X}_i'\widehat{\mathbf{H}}\widehat{\mathbf{h}}_t\widehat{\mathbf{h}}_t'\widehat{e}_{jt}^2\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_i = O_p(1)$$

which implies that $\widehat{\boldsymbol{\zeta}}_{NT}^{\dagger} = O_p(1)$. Similarly, we can show that $\widetilde{\boldsymbol{\zeta}}_{NT}$ for the case with serially correlated errors is $O_p(1)$.

Note that, for slope heterogeneous models, by Theorem 6 and Corollary 8, we have

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta} &= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i + O_p(\delta_{NT}^{-2}),\\
&= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{e}_i + O_p(\delta_{NT}^{-2})
\end{aligned}$$

combining the facts that $\widehat{\boldsymbol{\xi}}_{NT}^{\dagger} = O_p(1)$ and $\widehat{\boldsymbol{\zeta}}_{NT}^{\dagger} = O_p(1)$, we have

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta} =& \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta} - \frac{1}{N}\widehat{\boldsymbol{\xi}}_{NT} - \frac{1}{T}\widehat{\boldsymbol{\zeta}}_{NT}\\
=& \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i + O_p(\delta_{NT}^{-2})\\
=& \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{u}_i + O_p(\delta_{NT}^{-2})
\end{aligned}$$

which implies that $\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta} = O_p(N^{-1/2}) + O_p(\delta_{NT}^{-2})$.

For the slope homogeneous case, by Bai (2009a), we have

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta} &= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\boldsymbol{\varepsilon}_i + O_p(\delta_{NT}^{-3})\\
&= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{u}_i + O_p(\delta_{NT}^{-3})
\end{aligned}$$

Thus, we complete the proof. ■

Proof of Theorem 11

For the heterogeneous slope models, with (B.11), we have

$$\left(\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\right)^{-1} - \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i\right)^{-1} = O_p(\delta_{NT}^{-1})$$

which implies that

$$\left(\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\right)^{-1} - \mathbf{A}_0^{-1} = o_p(1) \tag{B.15}$$

For the case with homogeneous slopes, by proof of Proposition 2 in Bai (2009b), we can also derive that

$$\left(\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\right)^{-1} - \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i\right)^{-1} = O_p(\delta_{NT}^{-1})$$

which implies that

$$\left(\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\right)^{-1}-\mathbf{A}_1^{-1}=o_p(1). \tag{B.16}$$

Using the above facts, we just need to focus on the term $\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i$. For the homogeneous slope case, $\widehat{\mathbf{u}}_i=\mathbf{H}^0\boldsymbol{\phi}_i^0+\boldsymbol{\varepsilon}_i+\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})$, we have

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i-\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\\
=&\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i+\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})'\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\\
&+\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})'\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i+\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\\
&+\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})'\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i+\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\\
&+\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i+\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\\
=&\mathbb{I}_1+\mathbb{I}_2+\cdots+\mathbb{I}_8.
\end{aligned}$$

Consider the term $\mathbb{I}_1$, which is decomposed into

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\\
=&\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i\\
&-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell\\
&+\frac{1}{N^3T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\sum_{j=1}^{N}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_{\phi}^{-1}\widehat{\boldsymbol{\phi}}_j\mathbf{X}_j\\
=&\mathbb{I}_{1.1}+\mathbb{I}_{1.2}+\mathbb{I}_{1.3}+\mathbb{I}_{1.4}.
\end{aligned}$$

Consider $\mathbb{I}_{1.1}$, it is further decomposed into

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i+\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i\\
&+\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\mathbf{X}_i+\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\mathbf{X}_i.
\end{aligned}$$

The second term is bounded in norm by

$$\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\boldsymbol{\varepsilon}_i\|\|T^{-1/2}\mathbf{X}_i\|^3\cdot T\|\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}}\|\|\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}}\|=O_p(N^{-1/2}T^{1/2}\delta_{NT}^{-1}).$$

Similarly, we can prove that the third term is $O_p(N^{-1/2}T^{1/2}\delta_{NT}^{-1})$ and the fourth term is $O_p(N^{-1/2}T^{1/2}\delta_{NT}^{-2})$. The first term is

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'\mathbf{X}_i-\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i(\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{X}_i$$

which is bounded in norm by

$$\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|^2\|T^{-1/2}\boldsymbol{\varepsilon}_i'\mathbf{X}_i\|\cdot T^{1/2}\|\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}}\|$$
$$+\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|^3\|T^{-1/2}\boldsymbol{\varepsilon}_i'\mathbf{H}\|\cdot\|(\mathbf{H}'\mathbf{H}/T)^{-1}\|\|\mathbf{H}\|\|\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}}\|=O_p(N^{-1/2}).$$

Then we can show that $\mathbb{I}_{1.1}=O_p(N^{-1/2}T^{1/2}\delta_{NT}^{-1})$. Analogously, we can prove that $\mathbb{I}_{1.2}=O_p(N^{-1/2}T^{1/2}\delta_{NT}^{-1})$, $\mathbb{I}_{1.3}=O_p(N^{-1/2}T^{1/2}\delta_{NT}^{-1})$ and $\mathbb{I}_{1.4}=O_p(N^{-1/2}T^{1/2}\delta_{NT}^{-1})$. Thus, $\mathbb{I}_1=O_p(N^{-1/2}T^{1/2}\delta_{NT}^{-1})$. Following the argument in the proof of $\mathbb{I}_1$, we can show that $\mathbb{I}_3=O_p(N^{-1/2}T^{1/2}\delta_{NT}^{-1})$. Consider the term $\mathbb{I}_2$, which is bounded in norm by

$$\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\widehat{\mathbf{Z}}_i\|^2\|T^{-1/2}\mathbf{X}_i\|^2\cdot T\|\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}}\|^2=O_p(N^{-1}).$$

Note that $\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}=\mathbf{M}_{\widehat{\mathbf{H}}}(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})$, then $\mathbb{I}_4$ is bounded in norm by

$$\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\widehat{\mathbf{Z}}_i\|^2\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i\|\cdot T\|T^{-1/2}(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})\|\|\boldsymbol{\beta}^0-\bar{\boldsymbol{\beta}}\|=O_p(N^{-1/2}T^{1/2}\delta_{NT}^{-1}).$$

Analogously, we can show that $\mathbb{I}_5=O_p(N^{-1/2}T^{1/2}\delta_{NT}^{-1})$. Consider the term $\mathbb{I}_6$, which is

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i\\
&-\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell\\
&+\frac{1}{N^3T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\ell=1}^{N}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_j\mathbf{X}_j\\
=&\mathbb{I}_{6.1}+\mathbb{I}_{6.2}+\mathbb{I}_{6.3}+\mathbb{I}_{6.4}.
\end{aligned}$$

For the term $\mathbb{I}_{6.1}$, it is

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\mathbf{X}_i-\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\widehat{\mathbf{H}}\widehat{\mathbf{H}}'\mathbf{X}_i\\
&-\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}\boldsymbol{\phi}_i'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\mathbf{X}_i+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}\boldsymbol{\phi}_i'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\widehat{\mathbf{H}}\widehat{\mathbf{H}}'\mathbf{X}_i\\
&+\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\mathbf{X}_i-\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{X}_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\widehat{\mathbf{H}}\widehat{\mathbf{H}}'\mathbf{X}_i.
\end{aligned}$$

The first term is bounded in norm by

$$\sqrt{T}\cdot\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i'\boldsymbol{\varepsilon}_i\|\|\boldsymbol{\phi}_i\|\|T^{-1}(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\mathbf{X}_i\|=O_p(T^{1/2}\delta_{NT}^{-2}).$$

Analogously, we can show that both the second term, the third term and the fourth term are $O_p(T^{1/2}\delta_{NT}^{-2})$. The fifth term is bounded in norm by

$$\frac{T}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|T^{-1/2}\boldsymbol{\varepsilon}_i\|\|\boldsymbol{\phi}_i\|\|T^{-1}(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\mathbf{X}_i\|\cdot\|\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}}\|=O_p(T\delta_{NT}^{-3}).$$

Analogously, we can show that the sixth term is $O_p(T\delta_{NT}^{-3})$. Collecting the above six terms, we can show that $\mathbb{I}_{6.1} = O_p(T\delta_{NT}^{-3})$. Following the argument in the proof of $\mathbb{I}_{6.1}$, we can prove that $\mathbb{I}_{6.2} = O_p(T\delta_{NT}^{-3})$, $\mathbb{I}_{6.3} = O_p(T\delta_{NT}^{-3})$ and $\mathbb{I}_{6.4} = O_p(T\delta_{NT}^{-3})$. Thus, we derive that $\mathbb{I}_6 = O_p(T\delta_{NT}^{-3})$. Consider $\mathbb{I}_7$, which is decomposed into

$$
\begin{aligned}
\mathbb{I}_7 =& \frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i - \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_\ell\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell \\
&- \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i \\
&+ \frac{1}{N^3T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\ell=1}^{N}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_j'\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_j \\
=& \mathbb{I}_{7.1} + \mathbb{I}_{7.2} + \mathbb{I}_{7.3} + \mathbb{I}_{7.4}.
\end{aligned}
$$

For $\mathbb{I}_{7.1}$, it can be further decomposed into

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\mathbf{X}_i \\
&- \frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{X}_i'\widehat{\mathbf{H}}\widehat{\mathbf{H}}'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\mathbf{X}_i \\
&- \frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{X}_i'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\widehat{\mathbf{H}}\widehat{\mathbf{H}}'\mathbf{X}_i \\
&+ \frac{1}{NT^3}\sum_{i=1}^{N}\mathbf{X}_i'\widehat{\mathbf{H}}\widehat{\mathbf{H}}'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})'\widehat{\mathbf{H}}\widehat{\mathbf{H}}'\mathbf{X}_i.
\end{aligned}
$$

The first term is bounded in norm by

$$
\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|\boldsymbol{\phi}_i\|^2\|T^{-1}\mathbf{X}_i'(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})\|\cdot T\|T^{-1/2}(\mathbf{H}-\widehat{\mathbf{H}}\mathbf{R}^{-1})\| = O_p(T\delta_{NT}^{-3}).
$$

Analogously, we can prove that the second term and the third term both are $O_p(T\delta_{NT}^{-3})$, the fourth term is $O_p(T\delta_{NT}^{-4})$. Collecting the above four terms, we can derive that $\mathbb{I}_{7.1} = O_p(T\delta_{NT}^{-3})$. Similarly, we can prove that $\mathbb{I}_{7.2} = O_p(T\delta_{NT}^{-3})$, $\mathbb{I}_{7.3} = O_p(T\delta_{NT}^{-3})$ and $\mathbb{I}_{7.4} = O_p(T\delta_{NT}^{-3})$. Consequently, we have $\mathbb{I}_7 = O_p(T\delta_{NT}^{-3})$. Following the argument in the proof of $\mathbb{I}_6$, we can prove that $\mathbb{I}_8 = O_p(T\delta_{NT}^{-3})$.

Combining $\mathbb{I}_1$ to $\mathbb{I}_8$, we can derive that

$$
\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i = \frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i + O_p(T\delta_{NT}^{-3}).
$$

Next, we consider the term $(NT)^{-1}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i$, which is

$$
\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i =& \frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i - \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_\ell\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell \\
&- \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i \\
&+ \frac{1}{N^3T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\ell=1}^{N}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_j\mathbf{X}_j'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_\ell\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell \\
=& \mathbb{I}_{9.1} + \mathbb{I}_{9.2} + \mathbb{I}_{9.3} + \mathbb{I}_{9.4}.
\end{aligned}
$$

For $\mathbb{I}_{9.1}$, we have

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\varepsilon_i\varepsilon_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i-\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\varepsilon_i\varepsilon_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i\\
=&\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\varepsilon_i\varepsilon_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i+\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\varepsilon_i\varepsilon_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\mathbf{X}_i\\
&+\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\varepsilon_i\varepsilon_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\mathbf{X}_i\\
=&\mathbb{I}_{9.1.1}+\mathbb{I}_{9.1.2}+\mathbb{I}_{9.1.3}.
\end{aligned}
$$

Since $\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}}=-T^{-1}(\widehat{\mathbf{H}}-\mathbf{HR})\mathbf{R}'\mathbf{H}'^{-1}\mathbf{HR}(\widehat{\mathbf{H}}-\mathbf{HR})'^{-1}(\widehat{\mathbf{H}}-\mathbf{HR})(\widehat{\mathbf{H}}-\mathbf{HR})'^{-1}\mathbf{H}(\mathbf{RR}'-\big(T^{-1}\mathbf{H}'\mathbf{H}\big)^{-1})\mathbf{H}'$, $\mathbb{I}_{9.1.1}$ is

$$
\begin{aligned}
&-\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{X}_i'(\widehat{\mathbf{H}}-\mathbf{HR})\mathbf{R}'\mathbf{H}'\varepsilon_i\varepsilon_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i-\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{HR}(\widehat{\mathbf{H}}-\mathbf{HR})'\varepsilon_i\varepsilon_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i\\
&-\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{X}_i'(\widehat{\mathbf{H}}-\mathbf{HR})(\widehat{\mathbf{H}}-\mathbf{HR})'\varepsilon_i\varepsilon_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i\\
&-\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{H}(\mathbf{RR}'-\big(T^{-1}\mathbf{H}'\mathbf{H}\big)^{-1})\mathbf{H}'\varepsilon_i\varepsilon_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i.
\end{aligned}
$$

The first term is bounded in norm by

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|T^{-1/2}\mathbf{H}'\varepsilon_i\|\|T^{-1/2}\varepsilon_i'\mathbf{X}_i\|\cdot\|T^{-1/2}(\widehat{\mathbf{H}}-\mathbf{HR})\|\|\mathbf{R}\|\\
&+\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|^2\|T^{-1/2}\mathbf{H}'\varepsilon_i\|^2\cdot\|T^{-1/2}(\widehat{\mathbf{H}}-\mathbf{HR})\|\|\mathbf{R}\|\|(T^{-1}\mathbf{H}'\mathbf{H})^{-1}\|\|T^{-1/2}\mathbf{H}\|=O_p(\delta_{NT}^{-1}).
\end{aligned}
$$

The second term is bounded in norm by

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|\|T^{-1}(\widehat{\mathbf{H}}-\mathbf{HR})'\varepsilon_i\|\|T^{-1/2}\varepsilon_i'\mathbf{X}_i\|\cdot T^{1/2}\|T^{-1/2}\mathbf{H}\|\|\mathbf{R}\|\\
&+\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{X}_i\|^2\|T^{-1/2}\mathbf{H}'\varepsilon_i\|\|T^{-1}(\widehat{\mathbf{H}}-\mathbf{HR})'\varepsilon_i\|\cdot T^{1/2}\|\mathbf{R}\|\|(T^{-1}\mathbf{H}'\mathbf{H})^{-1}\|\|T^{-1/2}\mathbf{H}\|^2\\
=&O_p(T^{1/2}\delta_{NT}^{-2}).
\end{aligned}
$$

Analogously, we can prove that the third term is $O_p(T^{1/2}\delta_{NT}^{-2})$ and the fourth term is $O_p(\delta_{NT}^{-2})$. Thus, $\mathbb{I}_{9.1.1}=O_p(T^{1/2}\delta_{NT}^{-2})$. $\mathbb{I}_{9.1.2}$ is the transpose of $\mathbb{I}_{9.1.1}$, then $\mathbb{I}_{9.1.2}=O_p(T^{1/2}\delta_{NT}^{-2})$. Following the argument in the proof of $\mathbb{I}_{9.1.1}$, $\mathbb{I}_{9.1.3}=O_p(T\delta_{NT}^{-3})$. Thus,

$$
\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\varepsilon_i\varepsilon_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i-\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\varepsilon_i\varepsilon_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i=O_p(T\delta_{NT}^{-3}).
$$

Similarly, we can derive that

$$\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_\ell\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell - \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_\ell\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell = O_p(T\delta_{NT}^{-3})$$

$$\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i - \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell\mathbf{X}_\ell'\mathbf{M}_{\mathbf{H}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i = O_p(T\delta_{NT}^{-3})$$

$$\frac{1}{N^3T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\ell=1}^{N}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_j\mathbf{X}_j'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_\ell\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell$$
$$- \frac{1}{N^3T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\ell=1}^{N}\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_j\mathbf{X}_j'\mathbf{M}_{\mathbf{H}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_\ell\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell = O_p(T\delta_{NT}^{-3}).$$

Thus, we can derive that

$$\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i - \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i = O_p(T\delta_{NT}^{-3}).$$

Also, since

$$\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i = \frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i + O_p(T\delta_{NT}^{-3})$$
$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i - \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i = O_p(N^{-1/2})$$

and noting that $\mathbf{Z}_i = \widetilde{\mathbf{X}}_i$, we have, for the homogeneous slope model,

$$\frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i - \mathbf{B}_1 = o_p(1). \tag{B.17}$$

For the heterogeneous slopes case, using $\widehat{\mathbf{u}}_i = \mathbf{H}^0\boldsymbol{\phi}_i^0 + \mathbf{V}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i + \mathbf{X}_i(\boldsymbol{\beta}^0 - \bar{\boldsymbol{\beta}})$, we have

$$
\begin{aligned}
&\frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i - \frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i \\
=&\frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i(\boldsymbol{\beta}^0 - \bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i + \frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i(\boldsymbol{\beta}^0 - \bar{\boldsymbol{\beta}})(\boldsymbol{\beta}^0 - \bar{\boldsymbol{\beta}})'\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i \\
&+ \frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i(\boldsymbol{\beta}^0 - \bar{\boldsymbol{\beta}})'\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i + \frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i(\boldsymbol{\beta}^0 - \bar{\boldsymbol{\beta}})\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i \\
&+ \frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i(\boldsymbol{\beta}^0 - \bar{\boldsymbol{\beta}})\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i + \frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{V}_i\boldsymbol{\eta}_i(\boldsymbol{\beta}^0 - \bar{\boldsymbol{\beta}})'\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i \\
&+ \frac{2}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0(\boldsymbol{\beta}^0 - \bar{\boldsymbol{\beta}})'\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i + \frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i \\
&+ \frac{2}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i + \frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i \\
&+ \frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\phi}_i'\mathbf{H}'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i + \frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H}^0\boldsymbol{\phi}_i^0\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i \\
&+ \frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i + \frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i \\
&+ \frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i \\
=&\mathbb{J}_1 + \mathbb{J}_2 + \cdots + \mathbb{J}_{15}.
\end{aligned}
$$

The term $\mathbb{J}_1$ is bounded in norm by

$$
\frac{1}{N}\sum_{i=1}^N \|T^{-1/2}\widehat{\mathbf{Z}}_i\|^2\|T^{-1/2}\mathbf{X}_i\|\|T^{-1/2}\boldsymbol{\varepsilon}_i\| \cdot \|\boldsymbol{\beta}^0 - \bar{\boldsymbol{\beta}}\| = O_p(N^{-1/2}).
$$

Following the argument in the proof of the term $\mathbb{J}_1$, we can prove that $\mathbb{J}_2 = O_p(N^{-1})$, the terms $\mathbb{J}_3$ to $\mathbb{J}_7$ both are $O_p(N^{-1/2})$. Note that $\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{H} = \mathbf{M}_{\widehat{\mathbf{H}}}(\mathbf{H} - \widehat{\mathbf{H}}\mathbf{R}^{-1})$, $\mathbb{J}_8$ is bounded in norm by

$$
\begin{aligned}
&\left\|\frac{1}{NT^2}\sum_{i=1}^N \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\boldsymbol{\varepsilon}_i\boldsymbol{\phi}_i'(\mathbf{H} - \widehat{\mathbf{H}}\mathbf{R}^{-1})'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i\right\| \\
\leq&\frac{1}{NT^2}\sum_{i=1}^N \|T^{-1/2}\widehat{\mathbf{Z}}_i\|^2\|T^{-1/2}\boldsymbol{\varepsilon}_i\|\|\boldsymbol{\phi}_i\| \cdot \|T^{-1/2}(\mathbf{H} - \widehat{\mathbf{H}}\mathbf{R}^{-1})\| = O_p(\delta_{NT}^{-1}).
\end{aligned}
$$

Analogously, we can derive that $\mathbb{J}_9 = O_p(\delta_{NT}^{-2})$, the terms $\mathbb{J}_{10}$ to $\mathbb{J}_{12}$ both are $O_p(\delta_{NT}^{-1})$. Consider $\mathbb{J}_{13}$, which is

$$
\begin{aligned}
&\frac{1}{NT^2}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\mathbf{H}}\widehat{\mathbf{Z}}_i+\frac{1}{NT^2}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\widehat{\mathbf{Z}}_i\\
&+\frac{1}{NT^2}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_i'(\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}})\widehat{\mathbf{Z}}_i+\frac{1}{NT^2}\sum_{i=1}^{N}(\widehat{\mathbf{Z}}_i-\mathbf{Z}_i)'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i\\
&+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\mathbf{H}}(\widehat{\mathbf{Z}}_i-\mathbf{Z}_i)+\frac{1}{NT^2}\sum_{i=1}^{N}(\widehat{\mathbf{Z}}_i-\mathbf{Z}_i)'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\mathbf{H}}(\widehat{\mathbf{Z}}_i-\mathbf{Z}_i)\\
&+\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\mathbf{H}}\mathbf{Z}_i\\
=&\mathbb{J}_{13.1}+\cdots+\mathbb{J}_{13.7}.
\end{aligned}
$$

The term $\mathbb{J}_{13.1}$ is bounded in norm by

$$
\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\widehat{\mathbf{Z}}_i\|^2\|T^{-1/2}\mathbf{V}_i\|\|\boldsymbol{\eta}_i\|\|T^{-1/2}\boldsymbol{\varepsilon}_i\|\cdot\|\mathbf{M}_{\widehat{\mathbf{H}}}-\mathbf{M}_{\mathbf{H}}\|=O_p(\delta_{NT}^{-1})
$$

Similarly, we can derive that $\mathbb{J}_{13.2}$ and $\mathbb{J}_{13.3}$ both are $O_p(\delta_{NT}^{-1})$, and $\mathbb{J}_{13.4}$ is $O_p(T^{-1/2})$. Note that $\widehat{\mathbf{Z}}_i-\mathbf{Z}_i=N^{-1}\sum_{\ell=1}^{N}(\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell-\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell)\mathbf{X}_\ell$, we can derive $\mathbb{J}_{13.5}$ and $\mathbb{J}_{13.6}$ both are $O_p(T^{-1/2})$. It is easy to show that $\mathbb{J}_{13.7}$ is $O_p(T^{-1/2})$. Thus, we have $\mathbb{J}_{13}=O_p(\delta_{NT}^{-1})$. Following the argument in the proof of $\mathbb{J}_{13}$, we can derive that $\mathbb{J}_{14}=O_p(\delta_{NT}^{-1})$ and $\mathbb{J}_{15}=O_p(\delta_{NT}^{-1})$. Collecting the above terms, we have

$$
\frac{1}{NT^2}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i=\frac{1}{NT^2}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i+O_p(\delta_{NT}^{-1}).
$$

It is easy to show that

$$
\frac{1}{NT^2}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i=\frac{1}{NT^2}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\mathbf{H}}\widehat{\mathbf{Z}}_i+O_p(\delta_{NT}^{-1}).
$$

To proceed, we investigate $N^{-1}T^{-2}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\mathbf{H}}\widehat{\mathbf{Z}}_i$, which is decomposed into

$$
\begin{aligned}
\mathbb{J}_{16}\equiv&\frac{1}{NT^2}\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\mathbf{H}}\widehat{\mathbf{Z}}_i\\
=&\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_\ell\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell\\
&-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell\mathbf{X}_\ell'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\mathbf{H}}\widehat{\mathbf{Z}}_i\\
&+\frac{1}{N^3T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_j'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_\ell\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_j\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell\\
=&\mathbb{J}_{16.1}+\mathbb{J}_{16.2}+\mathbb{J}_{16.3}+\mathbb{J}_{16.4}.
\end{aligned}
$$

Consider $\mathbb{J}_{16.2}$ (ignoring the sign), we have

$$
\begin{aligned}
&\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M_H}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M_H}\mathbf{X}_\ell\widehat{\boldsymbol{\phi}}_i'\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}\widehat{\boldsymbol{\phi}}_\ell \\
=&\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M_H}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M_H}\mathbf{X}_\ell(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i)'\mathbf{R}'(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\boldsymbol{\phi}_\ell \\
&+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M_H}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M_H}\mathbf{X}_\ell\boldsymbol{\phi}_i'(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\mathbf{R}(\boldsymbol{\phi}_\ell-\mathbf{R}^{-1}\boldsymbol{\phi}_\ell) \\
&+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M_H}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M_H}\mathbf{X}_\ell\boldsymbol{\phi}_i'\mathbf{R}^{-1\prime}(\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}-\mathbf{R}'(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\mathbf{R})\mathbf{R}^{-1}\boldsymbol{\phi}_\ell \\
&+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M_H}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M_H}\mathbf{X}_\ell\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell \\
&+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M_H}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M_H}\mathbf{X}_\ell(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i)'\mathbf{R}'(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\mathbf{R}(\widehat{\boldsymbol{\phi}}_\ell-\mathbf{R}^{-1}\boldsymbol{\phi}_\ell) \\
&+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M_H}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M_H}\mathbf{X}_\ell\boldsymbol{\phi}_i'\mathbf{R}^{-1\prime}(\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}-\mathbf{R}'(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\mathbf{R})(\widehat{\boldsymbol{\phi}}_\ell-\mathbf{R}^{-1}\boldsymbol{\phi}_\ell) \\
&+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M_H}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M_H}\mathbf{X}_\ell(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i)'(\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}-\mathbf{R}'(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\mathbf{R})\mathbf{R}^{-1}\boldsymbol{\phi}_\ell \\
&+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{X}_i'\mathbf{M_H}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M_H}\mathbf{X}_\ell(\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i)'\left(\widehat{\boldsymbol{\Upsilon}}_\phi^{-1}-\mathbf{R}'(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\mathbf{R}\right)(\widehat{\boldsymbol{\phi}}_\ell-\mathbf{R}^{-1}\boldsymbol{\phi}_\ell).
\end{aligned}
$$

The first term is bounded in norm by

$$
\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\mathbf{V}_i\|^3\|\boldsymbol{\eta}_i\|^2\left\|\widehat{\boldsymbol{\phi}}_i-\mathbf{R}^{-1}\boldsymbol{\phi}_i\right\|\cdot\frac{1}{N}\sum_{\ell=1}^{N}\|T^{-1/2}\mathbf{V}_\ell\|\|\boldsymbol{\phi}_\ell\|\cdot\|\mathbf{R}\|\|(\boldsymbol{\Phi}'\boldsymbol{\Phi}/N)^{-1}\|=O_p(\delta_{NT}^{-1}).
$$

Analogously, we can prove that the remaining terms except the fourth term is $O_p(\delta_{NT}^{-1})$. For the fourth term, it is

$$
\begin{aligned}
&\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{V}_i'\mathbf{M_H}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M_H}\mathbf{V}_\ell\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell \\
=&\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{V}_i'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{V}_\ell\boldsymbol{\phi}_\ell'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_i-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{V}_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{V}_\ell\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell \\
&-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{V}_i'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{V}_\ell\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell \\
&+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{V}_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{V}_\ell\boldsymbol{\phi}_i'\boldsymbol{\Upsilon}_\phi^{-1}\boldsymbol{\phi}_\ell
\end{aligned}
$$

which is bounded in norm by

$$
\begin{aligned}
&\frac{1}{\sqrt{N}} \cdot \frac{1}{N} \sum_{i=1}^{N} \|T^{-1/2}\mathbf{V}_i\|^3 \|\boldsymbol{\eta}_i\|^2 \|\boldsymbol{\phi}_i\| \cdot \left\| \frac{1}{\sqrt{NT}} \sum_{\ell=1}^{N} \mathbf{V}_\ell \boldsymbol{\phi}_\ell' \right\| \|\boldsymbol{\Upsilon}_\phi^{-1}\| \\
&+ \frac{1}{T} \cdot \frac{1}{N} \sum_{i=1}^{N} \|T^{-1/2}\mathbf{V}_i'\mathbf{H}\|^2 \|\boldsymbol{\eta}_i\|^2 \|T^{-1/2}\mathbf{V}_i\| \|\boldsymbol{\phi}_i\| \cdot \frac{1}{N} \sum_{\ell=1}^{N} \|T^{-1/2}\mathbf{V}_\ell\| \|\boldsymbol{\phi}_\ell\| \cdot \|(\mathbf{H}'\mathbf{H}/T)^{-1}\| \|\boldsymbol{\Upsilon}_\phi^{-1}\| \\
&+ \frac{1}{\sqrt{NT}} \cdot \frac{1}{N} \sum_{i=1}^{N} \|T^{-1/2}\mathbf{V}_i'\mathbf{H}\| \|T^{-1/2}\mathbf{V}_i\|^2 \|\boldsymbol{\eta}_i\|^2 \|\boldsymbol{\phi}_i\| \cdot \left\| \frac{1}{\sqrt{NT}} \sum_{\ell=1}^{N} \mathbf{H}'\mathbf{V}_\ell \boldsymbol{\phi}_\ell \right\| \cdot \|\boldsymbol{\Upsilon}_\phi^{-1}\| \|(\mathbf{H}'\mathbf{H}/T)^{-1}\| \\
&+ \frac{1}{\sqrt{NT^2}} \cdot \frac{1}{N} \sum_{i=1}^{N} \|T^{-1/2}\mathbf{V}_i'\mathbf{H}\|^3 \|\boldsymbol{\eta}_i\|^2 \|\boldsymbol{\phi}_i\| \cdot \left\| \frac{1}{\sqrt{NT}} \sum_{\ell=1}^{N} \mathbf{H}'\mathbf{V}_\ell \boldsymbol{\phi}_\ell \right\| \cdot \|\boldsymbol{\Upsilon}_\phi^{-1}\| \|(\mathbf{H}'\mathbf{H}/T)^{-1}\|^2 \\
=&O_p(N^{-1/2}) + O_p(T^{-1}).
\end{aligned}
$$

Collecting the above terms, we can derive that $\mathbb{J}_{16.2} = O_p(\delta_{NT}^{-1})$. Following the argument in the proof of $\mathbb{J}_{16.2}$, we can derive that $\mathbb{J}_{16.3} = O_p(\delta_{NT}^{-1})$ and $\mathbb{J}_{16.4} = O_p(\delta_{NT}^{-1})$. Consider the term $\mathbb{J}_{16.1}$, which is

$$
\begin{aligned}
&\frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i = \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i \\
=&\frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{V}_i - \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{V}_i \\
&- \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{V}_i + \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{V}_i.
\end{aligned}
$$

It is easy to show that the last three terms both are $O_p(T^{-1})$. Then

$$
\mathbb{J}_{16.1} = \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\mathbf{H}}\mathbf{X}_i = \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{V}_i + O_p(T^{-1})
$$

which implies that

$$
\frac{1}{NT^2} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i'\mathbf{M}_{\mathbf{H}}\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{M}_{\mathbf{H}}\widehat{\mathbf{Z}}_i = \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{V}_i + O_p(\delta_{NT}^{-1}).
$$

Collecting the terms $\mathbb{J}_1$ to $\mathbb{J}_{16}$, we can derive that

$$
\frac{1}{NT^2} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i = \frac{1}{NT^2} \sum_{i=1}^{N} \mathbf{V}_i'\mathbf{V}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{V}_i'\mathbf{V}_i + O_p(\delta_{NT}^{-1})
$$

which implies that

$$
\frac{1}{NT^2} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\widehat{\mathbf{Z}}_i - \mathbf{C}_0 = o_p(1). \tag{B.18}
$$

Combining the equations (B.15), (B.16), (B.18) and (B.17), we can derive the theorem easily. Thus, we complete the proof. ■