Bootstrap Improved Inference for Factor-Augmented Regressions with CCE

De Vos, Ignace; Stauskas, Ovidijus

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Working Paper 2021:16

Department of Economics
School of Economics and Management

Bootstrap Improved Inference for Factor-Augmented Regressions with CCE

Ignace De Vos
Ovidijus Stauskas

November 2021

Bootstrap Improved Inference for Factor-Augmented Regressions with CCE

Ignace De Vos[1,2] and Ovidijus Stauskas[1]

[1]*Lund University, Department of Economics*
[2]*Ghent University, Department of Economics*

Abstract

The Common Correlated Effects (CCE) methodology is now well established for the analysis of factor-augmented panel models. Yet, it is often neglected that the pooled variant is biased unless the cross-section dimension (N) of the dataset dominates the time series length (T). This is problematic for inference with typical macroeconomic datasets where T often equal or larger than N. Given that an analytical correction is also generally infeasible, the issue remains without a solution. In response, we provide in this paper the theoretical foundation for the 'cross-section' or 'pairs' bootstrap in large N and T panels with $T/N < \infty$. We show that the scheme replicates the distribution of the CCE estimators, under both constant and heterogeneous slopes, such that bias can be eliminated and asymptotically correct inference can ensue even when N does not dominate. Monte Carlo experiments illustrate that the asymptotic properties also translate well to finite samples.

Keywords: bootstrap, pairs bootstrap, factor-augmented panel data models, interactive effects, factors, common correlated effects, bias-correction

1 Introduction

There is an abundance of empirical evidence suggesting that cross-section units in economic panels tend to be contemporaneously correlated and potentially driven by common components. Such cross-section dependence needs to be accounted for in the estimation of the econometric model if estimates and inferences are to be trusted (for details, consult e.g. Andrews, 2005; Sarafidis and Robertson, 2009; Sarafidis and Wansbeek, 2012). One of the leading approaches to model cross-section dependence is by assuming a multi-factor error structure, which was recently also branded an 'interactive effects' structure. The central idea is that cross-section units are simultaneously affected by a finite number of time-varying unobserved common variables, dubbed factors, to which they can respond with unit-specific intensities, which are called loadings. The factors may represent global pandemics, crises, business cycle fluctuations, technological progress, or other global trends and shocks (see in particular Sarafidis and Wansbeek, 2012, for more examples and an overview of cross-section dependence in panel data). Failure to account for these unobserved components results in inconsistent estimates and inferences when they are correlated with the regressors.

One of the leading techniques for estimating panel models with unobserved factors is the Common Correlated Effects (CCE) approach by Pesaran (2006). The method boils down to augmenting the model of interest with the cross-sectional averages (CA) of the observed variables such that asymptotically—as the cross-section dimension $N \to \infty$—the common factor space is eliminated. Both a mean group (CCEMG) and a pooled (CCEP) version are suggested, depending on whether the model slopes are assumed to be heterogeneous (variable) or homogeneous (constant) over cross-sectional units. Thanks to their computational simplicity and robustness (see e.g. Pesaran and Tosetti, 2011; Kapetanios et al., 2011; Westerlund et al., 2019, among others) both CCEMG and CCEP enjoy considerable popularity in practice, as evidenced by numerous applications. Yet, it is often neglected that the pooled CCE variant is biased in large (N, T) samples unless $T/N \to 0$ (see Westerlund and Urbain, 2015; Karabiyik et al., 2017). This result is highly relevant in practice as it implies that inference with standard asymptotic tests will be distorted unless the cross-section dimension of the dataset dominates the time series length. Such dimensions are often encountered in microeconomic applications, but N rarely dominates T in macro panels. In macro settings T is often similar or even larger than N. A few examples of such applications are Özatay et al. (2009); Berger and Heylen (2011); Albanese and Modica (2012); Bertoli and Fernández-Huertas Moraga (2013); Mazzanti and Musolesi (2013); Özmen and Özge Doğanay Yaşar (2016); Stevens and Childs (2017); Eberhardt and Teal (2020). Simulation results in Westerlund and Urbain (2015) confirm that size distortions can be severe for such combinations of N and T. While the obvious solution would be bias correction, this is impeded by the fact that the asymptotic bias depends on whether or not the number of CA exceeds the number of factors, which is unknown, as well as various unobserved matrices without consistent plug-in estimators. Westerlund and Urbain (2013b, 2015) and Karabiyik et al. (2019) have proposed plug-in corrections, but it can only be applied when the unknown number of factors is exactly equal to the number of CA. As this is both highly unlikely in practice and difficult to check, the applicability of the approach is very limited, and the problem remains without a solution.

The bootstrap is an attractive alternative to analytical correction when bias expressions are inestimable. In its essence, if the bootstrap is able to replicate the distribution of an estimator, its bias can be eliminated without explicit knowledge of the functional form.[1] In this paper, we pursue this strategy and establish the theoretical validity of the 'cross-section' (CS) or 'pairs' bootstrap for bias-adjustment and inference with the CCE estimators in large N and T panels. The key advantage of the algorithm is that the number of latent factors or time series properties of the data do not need to be known by the researcher. This is in contrast to many residual-based bootstrap schemes, for example when applied to time series models augmented with factors estimated with the principal components (PC) approach (see e.g. Gonçalves and Perron, 2014; Djogbenou et al., 2015), and it is therefore a considerable advantage in practice. We first derive the generalized asymptotic distribution of the CCE estimators, allowing for both general serial dependence and the possibility that the number of CA exceeds the number of factors, and show that the resampling algorithm replicates the distribution, with all its bias components, under both common and heterogeneous slopes. This generally requires in the common slope setting that the CA of the dependent variable is excluded from the estimation. This restriction is very mild and shown to be without loss of asymptotic efficiency. The heterogeneous slope setting requires no additional restrictions. The key implication is that the cross-section bootstrap enables elimination of the bias of the CCEP estimator and leads to asymptotically correct inferences even when N does not dominate. This is a novel result since the CS-bootstrap was shown for instance by Galvao and Kato (2014); Gonçalves and Kaffo (2015) to not allow correction of the Nickell-type (Nickell, 1981) incidental parameters bias of the fixed effects estimator. The resampling scheme itself was introduced into the panel data literature by Kapetanios (2008) and first studied in a CCEP context by Westerlund et al. (2019) as $N \to \infty$ for fixed T, in which case the analysis is simplified by

[1] See for example Everaert and Pozzi (2007) who use the bootstrap to correct for the Nickell (1981) bias in more general data generating processes than considered by analytical adjustments in e.g. Kiviet (1995); Bun and Carree (2005).

the fact that the CCEP estimator is also unbiased (since $T/N \to 0$).[2] This paper is therefore, to the best of our knowledge, the first to consider the bootstrap for CCE estimators in large N and T panel models with potential heterogeneous slopes, and the first to establish the validity of the cross-section resampling scheme for CCE under joint asymptotics. As such, this article provides the theoretical foundation for the standard errors and bootstrap-t intervals constructed with CS-resampling in e.g. Millo (2019); Juodis et al. (2021). In addition, bootstrap as a bias-correction tool for CCE estimators has not yet been considered. Gonçalves and Perron (2014); Djogbenou et al. (2015) study bootstrap corrections for factor-augmented models, but consider in stead the principal components (PC) estimator of Bai (2009) in a predictive time series model.

The remainder of this paper is structured as follows: the next section introduces the working model and assumptions. Section 3.1 considers the CCEP estimator in the common slope setting and presents its generalized asymptotic distribution as $(N, T) \to \infty$ such that $T/N \to \tau < \infty$. Section 3.1.1 outlines the bootstrap methodology and establishes conditions for its consistency. Section 3.2 presents results under heterogeneous slopes. Section 4 assesses the finite sample validity of our theory with a Monte Carlo experiment, and Section 5 concludes. All proofs are referred to Supplement A, and additional Monte Carlo evidence is provided in Supplement B.

Some notation: we will use $\mathbf{A}^{\dagger}$ to denote the Moore-Penrose pseudo-inverse of the matrix $\mathbf{A}$, $rk(\mathbf{A})$ for its rank, $|\mathbf{A}|$ for the determinant and let $\|\mathbf{A}\| = [tr(\mathbf{A}\mathbf{A}')]^{1/2}$ be the Euclidean (Frobenius) matrix norm. Let furthermore $\boldsymbol{\iota}_a$ be an a-rowed vector of ones and the $vec(.)$, $\otimes$ and $\circ$ operators denote respectively the vectorization operation and the Kronecker- and Hadamard (element-wise) products. Barred variables $\overline{\mathbf{A}}$ denote the cross-section average (CA) over the cross-section specific matrices $\mathbf{A}_i$ as in $\overline{\mathbf{A}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{A}_i$. A starred object $\mathbf{A}_i^*$ stands for an *observed* variable (matrix or scalar) that has been generated in the bootstrap world according to the particular scheme. On the other hand, $\mathbf{A}_{w,i}$ is the *weighted unobserved* primitive of the model. We formalize the bootstrap probability laws similarly to Galvao and Kato (2014). In particular, for any matrix bootstrap sequence $\mathbf{A}_n^*$, which depends on a generic index n, and a deterministic sequence $a_n \in \mathbb{R}_{++}$, we have $\|\mathbf{A}_n^*\| = o_{p^*}(a_n)$ if for every $\epsilon > 0$ and $\delta > 0$, we have $\mathbb{P}(\mathbb{P}^*(a_n^{-1}\|\mathbf{A}_n^*\| > \epsilon) > \delta) \to 0$ as $n \to \infty$, where $\mathbb{P}^*(.)$ is a bootstrap-induced measure. Similarly, $\|\mathbf{A}_n^*\| = O_{p^*}(a_n)$ if for every $\delta > 0$ and $\eta > 0$, there exists a constant $C > 0$, such that

[2] Note that the theory in Kapetanios (2008) does not consider the CCE estimators. This can be seen from model (2.1) and Assumptions 3.2 and 3.4 in that paper, the combination of which is incompatible with the CCE framework. Assumption 3.1 also rules out that the regressors have a factor structure as in Pesaran (2006). In its essence, extrapolations of the provided theory to a CCE context would neglect the impact of factor estimation error with the CA.

$\mathbb{P}(\mathbb{P}^*(a_n^{-1}\|\mathbf{A}_n^*\| > C) > \delta) < \eta$ for all $n \geq 1$. Additionally, $\mathbb{E}^*(\cdot)$, $\mathbb{V}ar^*(\cdot)$ and $\mathbb{C}ov^*(\cdot,\cdot)$ represent, respectively, the expectation, variance and covariance taken with respect to the induced measure $\mathbb{P}^*$, and $\mathbf{A}_n^* = \mathbf{A}^* + o_{p^*}(1)$ means $\|\mathbf{A}_n^* - \mathbf{A}^*\| = o_{p^*}(1)$ for the limiting bootstrap matrix $\mathbf{A}^*$. Lastly, $\to^{p*}$ ($\to^{p}$) and $\xrightarrow{d^*}$ ($\xrightarrow{d}$) represent convergence in probability and distribution with respect to the induced (generic) probability measure.

2 Model and assumptions

Consider the setup in Pesaran (2006) where y_{it} is the scalar dependent variable observed for cross-section $i = 1, \ldots, N$ at time $t = 1, \ldots, T$, and $\mathbf{x}_{i,t}$ is the corresponding $k \times 1$ vector of explanatory variables. The observed $T \times 1$ vector $\mathbf{y}_i = [y_{i,1}, \ldots, y_{i,T}]'$ and $T \times k$ matrix $\mathbf{X}_i = [\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,T}]'$ are generated according to:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \tag{2.1}$$

$$\mathbf{X}_i = \mathbf{F}\boldsymbol{\Gamma}_i + \mathbf{V}_i, \tag{2.2}$$

where $\boldsymbol{\beta}_i$ is a $k \times 1$ vector of unknown coefficients which could be heterogeneous or common over individuals, $\mathbf{F}$ is a $T \times m$ matrix stacking m unobserved factors $\mathbf{f}_t$ over time, with m fixed and finite, and $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Gamma}_i$ are the associated $m \times 1$ vector and $m \times k$ matrix of factor loadings. The $T \times 1$ vector $\boldsymbol{\varepsilon}_i$ and the $T \times k$ matrix $\mathbf{V}_i$ stack the idiosyncratic errors ε_{it} and $\mathbf{v}_{it}$ over time, respectively.

By combining (2.1)-(2.2), the data generating process (DGP) for the $T \times (1+k)$ matrix of observables can be written as an approximate static factor model

$$\mathbf{Z}_i = [\mathbf{y}_i, \mathbf{X}_i] = (\mathbf{F}\widetilde{\mathbf{C}}_i + \widetilde{\mathbf{U}}_i)\mathbf{B}_i = \mathbf{F}\mathbf{C}_i + \mathbf{U}_i, \tag{2.3}$$

where $\widetilde{\mathbf{C}}_i = [\boldsymbol{\gamma}_i, \boldsymbol{\Gamma}_i]$, $\widetilde{\mathbf{U}}_i = [\boldsymbol{\varepsilon}_i, \mathbf{V}_i] = [\widetilde{\mathbf{u}}_{i,1}, \ldots, \widetilde{\mathbf{u}}_{i,T}]'$ and $\widetilde{\mathbf{u}}_{i,t} = [\varepsilon_{i,t}, \mathbf{v}_{i,t}']'$, $\mathbf{B}_i = \mathbf{B} + \widetilde{\mathbf{B}}_i$, with $\mathbf{B} = [[1, \boldsymbol{\beta}']', [\mathbf{0}_{k\times 1}, \mathbf{I}_k]']$ and $\widetilde{\mathbf{B}}_i = [[0, \boldsymbol{v}_i']', \mathbf{0}_{k+1\times k}]$. The $T \times (1+k)$ matrix $\mathbf{U}_i = \widetilde{\mathbf{U}}_i\mathbf{B}_i = [\mathbf{u}_{i1}, \ldots, \mathbf{u}_{i,T}]'$, with $\mathbf{u}_{it} = [\varepsilon_{i,t} + \mathbf{v}_{i,t}'\boldsymbol{\beta}_i, \mathbf{v}_{i,t}']'$ thus combines the idiosyncratic errors and the $m \times (1+k)$ loading matrix $\mathbf{C}_i = \widetilde{\mathbf{C}}_i\mathbf{B}_i$ stipulates the influence of the factors on each column of $\mathbf{Z}_i$.

We make the following assumptions:

Assumption 1 *(Idiosyncratic errors) ε_{it} and $\mathbf{v}_{it}$ are stationary variables, independent across i with $\mathbb{E}(\varepsilon_{i,t}) = 0$, $\mathbb{E}(\mathbf{v}_{i,t}) = \mathbf{0}_{k\times 1}$, $\sigma_i^2 = \mathbb{E}(\varepsilon_{i,t}^2)$, $\boldsymbol{\Sigma}_i = \mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}_{i,t}')$, $\boldsymbol{\Omega}_i = \mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')$, with $\boldsymbol{\Omega}_i, \boldsymbol{\Sigma}_i$ positive definite and $\mathbb{E}(\varepsilon_{i,t}^6) < \infty$, $\mathbb{E}(\|\mathbf{v}_{i,t}\|^6) < \infty$ for all i and t. Additionally, $\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}|E(\varepsilon_{i,t}\varepsilon_{i,s})| = O(1)$*

and $\frac{1}{T^3}\sum_{t=1}^{T}\sum_{q=1}^{T}\sum_{r=1}^{T}\sum_{s=1}^{T}|E(\varepsilon_{i,t}\varepsilon_{i,q}\varepsilon_{i,r}\varepsilon_{i,s})| = O(1)$ *as* $T \to \infty$, $\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 \to \sigma^2 < \infty$ *and* $\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Sigma}_i \to \mathbf{\Sigma} < \infty$ *as* $N \to \infty$, *and we define* $\mathbf{\Sigma}_{\mathbf{u},i} = \mathbb{E}(\mathbf{u}_{i,t}\mathbf{u}_{i,t}') = \mathbf{B}'\mathbf{\Sigma}_{\widetilde{\mathbf{u}},i}\mathbf{B}$ *and* $\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Sigma}_{\mathbf{u},i} \to \mathbf{\Sigma}_{\mathbf{u}} = \mathbf{B}'\mathbf{\Sigma}_{\widetilde{\mathbf{u}}}\mathbf{B}$, *where* $\mathbf{\Sigma}_{\widetilde{\mathbf{u}},i} = [[\sigma_i^2, \mathbf{0}_{1\times k}]', [\mathbf{0}_{k\times 1}, \mathbf{\Sigma}_i]']$ *and* $\mathbf{\Sigma}_{\widetilde{\mathbf{u}}} = [[\sigma^2, \mathbf{0}_{1\times k}]', [\mathbf{0}_{k\times 1}, \mathbf{\Sigma}]']$.

Assumption 2 *(Common factors)* $\mathbf{f}_t$ *is covariance stationary with* $\mathbb{E}(\|\mathbf{f}_t\|^4) < \infty$, *absolute summable autocovariances and* $T^{-1}\mathbf{F}'\mathbf{F} \to \mathbf{\Sigma}_{\mathbf{F}}$ *as* $T \to \infty$, *with* $\mathbf{\Sigma}_{\mathbf{F}}$ *positive definite.*

Assumption 3 *(Factor loadings) The* $\mathbf{C}_i$ *are generated according to*

$$\mathbf{C}_i = \widetilde{\mathbf{C}}_i\mathbf{B}_i = (\widetilde{\mathbf{C}} + \widetilde{\boldsymbol{\eta}}_i)\mathbf{B}_i = \mathbf{C} + \boldsymbol{\eta}_i, \qquad vec(\widetilde{\boldsymbol{\eta}}_i) \sim IID(\mathbf{0}_{m(1+k)}, \mathbf{\Omega}_{\widetilde{\eta}}), \tag{2.4}$$

where $\widetilde{\mathbf{C}} = \mathbb{E}(\widetilde{\mathbf{C}}_i) = [\boldsymbol{\gamma}, \mathbf{\Gamma}]$, $\mathbf{\Omega}_{\widetilde{\eta}} = \mathbb{E}(vec(\widetilde{\boldsymbol{\eta}})vec(\widetilde{\boldsymbol{\eta}})')$ *and* $\mathbf{\Sigma}_{\boldsymbol{\eta}} = \mathbb{E}(\widetilde{\boldsymbol{\eta}}_i \otimes \widetilde{\boldsymbol{\eta}}_i)$ *have finite elements.*

Assumption 4 *(Rank condition)* $rk(\overline{\mathbf{C}}) = m$ *for all* N.

Assumption 5 *(Independence)* $\mathbf{f}_t, \varepsilon_{is}, \mathbf{v}_{jl}, \widetilde{\boldsymbol{\eta}}_n$ *are mutually independent for all* i, j, n, t, s, l.

Assumption 6 *(Slope heterogeneity) The slopes* $\boldsymbol{\beta}_i$ *follow*

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{v}_i, \qquad \boldsymbol{v}_i \sim IID(\mathbf{0}_{k\times 1}, \mathbf{\Omega}_{\boldsymbol{v}})$$

with $\mathbf{\Omega}_{\boldsymbol{v}}$ *a finite nonnegative definite* $k \times k$ *matrix and the* $\boldsymbol{v}_i$ *are independent of* $\mathbf{f}_t, \varepsilon_{is}, \mathbf{v}_{jl}, \widetilde{\boldsymbol{\eta}}_n$ *for all* i, j, n, t, s, l.

The setting defined by (2.1)-(2.2) and Assumptions 1-6 is similar to that in Pesaran (2006), where the CCE methodology was first proposed, but deviates in the following respects. First, we present the model without fixed effects or other observed factors. This is for ease of exposition only and the main results below follow through in the presence of such elements (see also Remark 1). Second, we require in Ass.1 that sixth moments of the innovations are finite and impose summability conditions of the moments that are similar in spirit to (implied by) a mixing condition. This is stronger than both Pesaran (2006) and Karabiyik et al. (2017), but a consequence of our $T/N \to \tau < \infty$ asymptotics (not considered in Pesaran, 2006) in combination with the presence of serial dependence (which is excluded in Karabiyik et al., 2017). Following Westerlund and Urbain (2013a); Karabiyik et al. (2019), Assumption 3 also generalizes Pesaran (2006) by allowing the factor loadings in the process of the dependent variable γ_i and those in the explanatory variables $\mathbf{\Gamma}_i$ to be correlated for each cross-section unit i. This can be seen from the fact that $\mathbf{\Omega}_{\widetilde{\eta}}$ is not restricted to be a (block) diagonal matrix. We further note that Ass.2-5 are identical to Assumptions 1(iii)-(v) in Karabiyik et al. (2017).

3 CCE estimation in large N and T panels

The central idea behind the CCE methodology is to estimate the unobserved factors in eq.(2.1) with the cross-section averages of the observed data. This idea follows straightforwardly from eq.(2.3), which shows that the CA of the observed data $\overline{\mathbf{Z}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i$ is

$$\overline{\mathbf{Z}} = \mathbf{F}\overline{\mathbf{C}} + \overline{\mathbf{U}} \tag{3.1}$$

which, given the rank condition in Ass.4, can be solved for $\mathbf{F}$ as

$$\mathbf{F} = (\overline{\mathbf{Z}} - \overline{\mathbf{U}})\overline{\mathbf{C}}^{\dagger} \tag{3.2}$$

The key insight from (3.2) is that since $\left\|\overline{\mathbf{U}}\right\| = O_p(N^{-1/2})$ for a fixed T, the CA of the residuals $\overline{\mathbf{U}}$ is negligible and the observed $\overline{\mathbf{Z}}$ therefore asymptotically (as $N \to \infty$) mimics the behavior of the factors (up to a rotation). Factors can thus be estimated as $\widehat{\mathbf{F}} = \overline{\mathbf{Z}}$, and these estimates can in turn be used as additional regressors to control for the unobserved factor space in model (2.1). That is, substituting (3.2) into (2.1) gives the so called factor-augmented model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \overline{\mathbf{Z}}\boldsymbol{\pi}_i + (\boldsymbol{\varepsilon}_i - \overline{\mathbf{U}}\boldsymbol{\pi}_i) \tag{3.3}$$

where the required linear combination $\boldsymbol{\pi}_i = \overline{\mathbf{C}}^{\dagger}\gamma_i$ is estimable by least squares (LS). The Common Correlated Effects estimators are the LS estimators of the parameters in this augmented model and differ depending on whether the slopes $\boldsymbol{\beta}_i$ are assumed to be common or variable over cross-sections. The Common Correlated Effects Pooled (CCEP) estimator is the LS solution in model (3.3) when slopes are homogeneous, $\boldsymbol{\beta}_i = \boldsymbol{\beta}$,

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{y}_i \tag{3.4}$$

where $\mathbf{M}_{\widehat{\mathbf{F}}} = \mathbf{I}_T - \overline{\mathbf{Z}}(\overline{\mathbf{Z}}'\overline{\mathbf{Z}})^{+}\overline{\mathbf{Z}}'$ orthogonalizes the data on the estimated factors. Pesaran (2006) suggests the following non-parametric estimator for its sample variance

$$\widehat{\boldsymbol{\Theta}} = N^{-1}\overline{\mathbf{Q}}^{-1}\widehat{\boldsymbol{\Upsilon}}\overline{\mathbf{Q}}^{-1}, \qquad \widehat{\boldsymbol{\Upsilon}} = \frac{1}{N-1}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{mg})(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{mg})'\widehat{\mathbf{Q}}_i \tag{3.5}$$

where $\widehat{\mathbf{Q}}_i = T^{-1}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i$ and $\overline{\mathbf{Q}} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i$. Although (3.4) is also consistent for $\mathbb{E}(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$ when slopes vary over individuals $\boldsymbol{\beta}_i \neq \boldsymbol{\beta}$, the Mean Group CCE (CCEMG) approach is then a more natural estimator. It is calculated as the average of the individual-specific LS slope estimates $\widehat{\boldsymbol{\beta}}_i$ in (3.3)

$$\widehat{\boldsymbol{\beta}}_{mg} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_i, \qquad \widehat{\boldsymbol{\beta}}_i = \left(\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{y}_i \tag{3.6}$$

and has the following sample variance estimator

$$\widehat{\mathbf{\Omega}}_v = \frac{1}{N(N-1)} \sum_{i=1}^{N} (\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{mg})(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{mg})' \tag{3.7}$$

Note the importance of Ass.4 to obtain (3.3). This rank condition states that the set of CA must span the space of the factors, or in other words that there must be at least as many CA holding linearly independent information about the unobserved factors, as there are factors (m). One implication is that the number of factors m therefore cannot exceed the number of CA. Pesaran (2006) has shown that this rank condition in Ass.4 ensures the validity of the substitution in (3.3) such that the CA consistently estimate the factor space. By consequence, both presented CCE variants are consistent as $N \to \infty$.

Estimating factors with CA is an elegant solution to the unobserved factor problem but not without consequences for the asymptotic properties of the CCE estimators. In practice, the number of factors, m, is typically small and likely to be exceeded by the number of CA (the $k+1$ columns of $\overline{\mathbf{Z}}$). Equation (3.1) implies that $k+1-m$ columns of $\overline{\mathbf{Z}}$ are then asymptotically degenerate so that the pseudo-inverse $(T^{-1}\overline{\mathbf{Z}}'\overline{\mathbf{Z}})^{\dagger}$ that features in $\mathbf{M}_{\widehat{\mathbf{F}}}$ is unbounded as $(N,T) \to \infty$ (see Karabiyik et al., 2017). While this has no consequences for consistency, the asymptotic distribution and how it should be analyzed are significantly affected. Many standard arguments no longer apply. This was first addressed by Karabiyik et al. (2017) for the CCEP estimator under simplified error assumptions with common slopes, but it has to the best of our knowledge not yet been considered in asymptotic analyses of the CCEMG and CCEP estimators when slopes are heterogeneous.[3] Hence, we begin our analysis by deriving generalized asymptotic distributions for the CCE estimators and subsequently motivate and prove consistency of the bootstrap based on those results. Naturally, asymptotic degeneracy ($m < k+1$) in the original sample will, and should, affect the bootstrap distribution as well. We analyze first the common slope setting in the next section, and consider heterogeneous slopes in section 3.2.

3.1 Homogeneous slopes: $\boldsymbol{\beta}_i = \boldsymbol{\beta}$

The CCEP estimator in eq.(3.4) is the natural estimator in model (3.3) when the slopes are common over individuals $\boldsymbol{\beta}_i = \boldsymbol{\beta}$. In the next theorem, we establish its asymptotic distribution as $(N,T) \to \infty$ such that $T/N \to \tau < \infty$ under potential asymptotic degeneracy (due to more CA

[3] Karabiyik et al. (2019) provide some results for their augmented variant of the CCE estimator under assumptions that are close to ours, but crucially impose that the set of averages is successfully pre-selected such that $cols(\overline{\mathbf{Z}}) = m$ and $rk(\overline{\mathbf{Z}}) = m$, where $cols(\overline{\mathbf{Z}})$ is the number of columns of $\overline{\mathbf{Z}}$. This avoids the $m < cols(\overline{\mathbf{Z}})$ problem, but need not hold true in applications. Our analysis does not require this assumption, as it holds for $m \leq cols(\overline{\mathbf{Z}})$.

than factors, $m \leq 1+k$), allowing for both cross-section heteroscedasticity and serial correlation in the residuals.

Theorem 1 *Let Ass.1-5 hold. Then, as $(N,T) \to \infty$ such that $T/N \to \tau < \infty$*

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}) + \sqrt{\tau}\boldsymbol{\Sigma}^{-1}(\mathbf{b} - \mathbf{d})$$

where $\boldsymbol{\Psi} = \text{plim}_{(N,T)\to\infty} \frac{1}{N}\sum_{i=1}^{N}(T^{-1}\mathbf{V}_i'\boldsymbol{\Omega}_i\mathbf{V}_i)$, $\mathbf{b} = \mathbf{b}_1 - \mathbf{b}_2$,

$$\mathbf{b}_1 = \mathbf{q}_{xy}'\boldsymbol{\Sigma}_{\boldsymbol{\eta}}' vec((\mathbf{C}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{C}^{\dagger})$$

$$\mathbf{b}_2 = \boldsymbol{\Gamma}'(\mathbf{C}^{\dagger})'[\sigma^2, \mathbf{0}_{1\times k}]'$$

and $\mathbf{d} = \mathbf{0}_{k\times 1}$ if $m = 1+k$, whereas if $m < 1+k$ then $\mathbf{d} = \mathbf{d}_1 + \mathbf{d}_2$ with

$$\mathbf{d}_1 = \mathbf{q}_{xy}'\boldsymbol{\Sigma}_{\boldsymbol{\eta}}' vec\left((\mathbf{C}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{D}_{-m}\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{C}^{\dagger}\right)$$

$$\mathbf{d}_2 = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\Sigma}_i[\boldsymbol{\beta}, \mathbf{I}_k] - \boldsymbol{\Gamma}'(\mathbf{C}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}})\mathbf{D}_{-m}[\sigma_i^2, \mathbf{0}_{1\times k}]'$$

with $\mathbf{q}_{xy} = (\mathbf{q}_y \otimes \mathbf{q}_x)$, $\mathbf{q}_y = [1, \mathbf{0}_{k\times 1}']'$, $\mathbf{q}_x = [\mathbf{0}_{k\times 1}, \mathbf{I}_k]'$, $\mathbf{D}_{-m} = \mathbf{T}\mathbf{H}_{-m}(\mathbf{H}_{-m}'\mathbf{T}'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m})^{\dagger}\mathbf{H}_{-m}'\mathbf{T}'$. $\mathbf{T}$ is the $(1+k)\times(1+k)$ partitioning matrix such that $\mathbf{C}\mathbf{T} = [\mathbf{C}_m, \mathbf{C}_{-m}]$ with $\mathbf{C}_m$ an $m\times m$ full rank matrix and $\mathbf{C}_{-m}$ is $m\times(k+1-m)$, and $\mathbf{H}_{-m} = [-(\mathbf{C}_m^{-1}\mathbf{C}_{-m})', \mathbf{I}_{k+1-m}]'$.

Theorem 1 generalizes the results in Karabiyik et al. (2017) by allowing both serial correlation and heteroscedasticity in the innovations, and confirms that the asymptotic distribution of the CCEP estimator also features bias terms in this setting unless the cross-section dimension of the dataset dominates the time series length. This is at its root an incidental parameters bias (see e.g. Neyman and Scott, 1948) induced by estimating the factors with the CA at every $t = 1, \ldots, T$ in the dataset, as this accumulates T approximation errors that vanish only with N. If T and N then grow at a similar rate, error accumulates as quickly as it vanishes and a bias will remain in the limit. The larger T is relative to N, the larger this over-accumulation of error, and hence bias, will be, and only when N grows faster than T does error die out sufficiently fast for the distribution to be correctly centered. This is represented by the $T/N \to \tau < \infty$ in Theorem 1, so that $\tau = 0$ is the case where N dominates T, and the $\tau < \infty$ restriction ensures that the accumulation of error is not explosive.

It is not difficult to see that the asymptotic bias can be highly disruptive for inference. This is apparent from the distribution of the t-statistic under the null for the l-th coordinate in $\boldsymbol{\beta}$ as $(N,T) \to \infty$ such that $T/N \to \tau < \infty$

$$t = \frac{\mathbf{q}_l'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sqrt{\mathbf{q}_l'\widehat{\boldsymbol{\Theta}}\mathbf{q}_l}} = \frac{\mathbf{q}_l'\sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sqrt{\mathbf{q}_l' NT\widehat{\boldsymbol{\Theta}}\mathbf{q}_l}} \xrightarrow{d} \mathcal{N}(0,1) + \sqrt{\tau}\frac{\mathbf{q}_l'\boldsymbol{\Sigma}^{-1}(\mathbf{b} - \mathbf{d})}{\sqrt{\mathbf{q}_l'\boldsymbol{\Theta}\mathbf{q}_l}} \tag{3.8}$$

where $\mathbf{q}_l$ a $k \times 1$ vector of zeros with a one as its l-th element. This result follows from Theorem 1 and $NT\widehat{\mathbf{\Theta}} \to^p \mathbf{\Theta} = \mathbf{\Sigma}^{-1}\mathbf{\Psi}\mathbf{\Sigma}^{-1}$ by Theorem 8 of Supplement A. Clearly, unless $\tau = 0$, the bias shifts the center of the distribution away from zero and thereby causes over-rejection of the correct null hypothesis. The t-test will thus tend to be over-sized. The actual severity of the problem depends on the drivers of the asymptotic bias, i.e. the overall noise level in the form of the error and loading variance-covariance matrices $\sigma^2, \mathbf{\Sigma}_\eta, \mathbf{\Sigma}_\mathbf{u}, \mathbf{\Sigma}$, but also on the asymptotic information content in the CA. The latter follows from the presence of $\mathbf{C}^\dagger = \mathbf{C}'(\mathbf{CC}')^{-1}$ in the bias expression. That is, recall from (3.2) that the approximation error when estimating $\mathbf{F}$ with $\overline{\mathbf{Z}}$ is $\overline{\mathbf{U}}\overline{\mathbf{C}}^\dagger$. As alluded to above, the residuals $\overline{\mathbf{U}}$ are the source of the bias in Theorem 1 and they are themselves scaled by $\overline{\mathbf{C}}^\dagger$. Noting that $\overline{\mathbf{C}} \to^p \mathbf{C}$ and $\overline{\mathbf{C}}^\dagger \to^p \mathbf{C}^\dagger$ as $N \to \infty$, we have that when the columns of $\mathbf{C}$ are highly informative about the factors[4], then the determinant $|\mathbf{CC}'|$ is large so that $(\mathbf{CC}')^{-1}$ and hence $\|\mathbf{C}^\dagger\|$ are small. The impact of the estimation error $\overline{\mathbf{U}}$ on the bias, in the form of $\sigma^2, \mathbf{\Sigma}_\eta, \mathbf{\Sigma}_\mathbf{u}, \mathbf{\Sigma}$ in the expressions above, will then be pushed down. Naturally, the converse is also true. As $|\mathbf{CC}'| \to 0$ the asymptotic information content of the CA decreases so that $(\mathbf{CC}')^{-1}$ explodes and $\mathbf{C}^\dagger$ aggravates the impact of the other bias components.[5] The main takeaway is thus highly intuitive: when $\tau > 0$ estimation error from the CA causes bias, and the less informative the set of CA is, the larger the (absolute) asymptotic bias and size distortions will be. If on the other hand N dominates T ($\tau = 0$), error vanishes more quickly than it accumulates and asymptotic information content is not an issue (provided that the rank condition is satisfied). In this case (3.8) is correctly centered and tests based on it are correctly sized. In practice, it is of course unknown whether a set of CA is informative so bias remains a concern.

The obvious solution to the problem would be bias-correction. Yet, the key practical implication of Theorem 1 is that analytical adjustments appear to be infeasible for applications. The specific form of the bias expression depends first of all on whether the number of CA equals ($m = 1 + k$) or surpasses ($m < 1 + k$) the number of factors, with $\mathbf{b}$ featuring in the distribution on both occasions, whereas $\mathbf{d}$ is only present in the latter case.[6] The $\mathbf{d}$ term represents the additional bias caused by the asymptotic singularity of $\overline{\mathbf{Z}}'\overline{\mathbf{Z}}/T$ when $m < 1 + k$. The difficulty for correction is that firstly m is unknown so that the researcher is unaware of whether $\mathbf{d}$ is present or not. Even if m were known (or estimated), the dependence of $\mathbf{d}$ on the unknown rotations $\mathbf{T}$ and $\mathbf{H}_{-m}$ still

[4]For example when $m = 2$ and $\overline{\mathbf{X}}$ loads only on the first factor, while $\overline{\mathbf{y}}$ loads only on the second.

[5]Logically, in the limit, $|\mathbf{CC}'| = 0$, the CA are not informative for the full common factor space. The $\mathbf{C}^\dagger$ matrix is then undefined and the CCE estimators are generally inconsistent. This is excluded under Ass.4.

[6]Note that the distribution in Theorem 1 features one less bias term than Theorem 3 in Karabiyik et al. (2017). Our proof of Theorem 1 in the supplement shows that the random loading assumption (Ass.3 here or Assumption 1.(iv) in Karabiyik et al. (2017)) implies that in their expressions $\mathbf{b}_1 = \mathbf{0}_{k\times 1}$.

hinders an actual correction. In addition, while $\sigma^2, \mathbf{\Sigma}$ and $\mathbf{\Sigma_u}$ have consistent estimators (see e.g. Westerlund and Urbain, 2013b), this is not the case for the loading population mean $\mathbf{C}$. By consequence, $\mathbf{\Gamma}$ and the prominent scaling matrix $\mathbf{C}^\dagger$ remain unknown to the researcher.[7]

Given these difficulties with implementing the bias expression in Theorem 1 for correction we consider in the next section the bootstrap as an alternative. As we will see, this allows us to side-step estimation of any bias components and allows the researcher to also remain agnostic about the number of factors m.

3.1.1 Pairs bootstrap for CCE estimators

The bootstrap is an attractive alternative to analytical bias-correction and inference when bias expressions are inestimable. In its essence, if the bootstrap is able to replicate the distribution of an estimator, its bias can be eliminated without explicit knowledge of the functional form and consistent inferences can be made even when classical asymptotic tests fail.

The objective is thus to replicate with the bootstrap the asymptotic distribution in Theorem 1. To do so, it is paramount that the resampling scheme preserves the correlation and variance structure of the original sample. Cross-section/pairs resampling is an ideal candidate given eq.(2.3) and Ass.1-5. The scheme was first introduced into a panel data context by Kapetanios (2008) and boils down to generating the bootstrap dataset $\mathbf{Z}^* = [\mathbf{Z}_1^{*\prime}, \ldots, \mathbf{Z}_N^{*\prime}]'$ according to

$$\mathbf{Z}_i^* = \mathbf{Z}_{i^*} \qquad \text{for} \qquad i = 1, \ldots, N \tag{3.9}$$

with i^* drawn with replacement from $(1, \ldots, N)$. Put simply, the data for cross-section unit i in the bootstrap sample is generated by taking the data for a random unit i^* of the original sample.[8] This implies given (2.3) that the bootstrap data is

$$\mathbf{Z}_i^* = \mathbf{Z}_{i^*} = \mathbf{F}\mathbf{C}_{i^*} + \mathbf{U}_{i^*}$$

The loading and innovation matrices of the original sample are thus randomly (and implicitly) redistributed over cross-sections while the factors and the time series dimensions remain untouched. This is advantageous in practice as the factors, time series properties, loading means,

[7]Westerlund and Urbain (2013b, 2015) propose the least squares estimator $\widehat{\mathbf{C}}_i = (\overline{\mathbf{Z}}'\overline{\mathbf{Z}})^{-1}\overline{\mathbf{Z}}'\mathbf{Z}_i$ of the individual-specific loading matrix. It can be shown, however, that the mean of this estimator $\widehat{\mathbf{C}} = \frac{1}{N}\sum_{i=1}^N \widehat{\mathbf{C}}_i$ is inconsistent for $\mathbf{C}$.

[8]Note that the presentation of the algorithm is kept simple here for ease of exposition. The interested reader is referred to section 3.1 of the supplement for a formal exposition of its properties. There, the bootstrap dataset is generated as $\mathbf{Z}^* = \mathbf{W}_T\mathbf{Z}$, with $\mathbf{W}_T = (\mathbf{w} \otimes \mathbf{I}_T)$, $\mathbf{w} = [\mathbf{w}_1', \ldots, \mathbf{w}_N']'$ and $\mathbf{w}_i$ is a $1 \times N$ Boolean selection vector drawn for a multinomial distribution with one trial and $k = N$ events, each with probability N^{-1}.

and the within unit variance-covariances of the loadings and errors $(\boldsymbol{\Sigma}_{\boldsymbol{\eta}}, \sigma_i^2, \boldsymbol{\Sigma}_i, \boldsymbol{\Sigma}_{\mathbf{u},i}, \boldsymbol{\Omega}_i)$ are replicated while the researcher remains agnostic about these components and even the number of factors m that are at play. Note that the aforementioned are all key components of the asymptotic distribution in Theorem 1. The ability to keep m unknown in the resampling process is particularly attractive because it can be difficult to estimate, and miss-specification in model-based resampling schemes would significantly distort the bootstrap distribution. Gonçalves and Perron (2014, 2020) for instance employ a residual bootstrap scheme in a PC context and therefore implicitly assume that the number of factors is known or correctly selected for their asymptotic theory to follow through. With residual-based schemes one also needs to be careful with how residuals are resampled, as the Wild Bootstrap scheme in Gonçalves and Perron (2014) for instance does not reproduce error serial correlation (Djogbenou et al., 2015). In stead, none of these considerations are required with (3.9), which makes the algorithm both easy and broadly applicable. The key assumptions are in stead that $N \to \infty$ and that loadings $\mathbf{C}_i$ and innovations $\mathbf{U}_i$ are cross-sectionally independent as in Ass.1 and 3. It is important to note that this does not imply that the observed data $(\mathbf{y}_i, \mathbf{X}_i)$ is cross-sectionally independent, but rather that the cross-section dependence stems from $\mathbf{F}$ only, and that the data is therefore cross-sectionally independent *conditional* on the sigma algebra $\mathcal{F} = \sigma\{\mathbf{F}\}$ (see also Andrews, 2005). Violation of this assumption would require the use of blocked resampling variants (see e.g. Lahiri, 2003), but this is beyond the scope of this paper. As a final note, observe that it is crucial for our purposes to regenerate also $\mathbf{X}_i^*$ in the bootstrap sample, as the bias in Theorem 1 is induced by error from estimating factors with the CA, of which $\overline{\mathbf{X}}$ is an integral part. Conditioning on (fixing) the regressors as in classical residual bootstrap methods would therefore not replicate bias. Gonçalves and Perron (2014) make a similar observation in the PC context. Similarly, fixing the CA over bootstrap iterations as in Westerlund et al. (2019) (who consider a fixed T setting) will not replicate bias when $(N, T) \to \infty$.

We have thus far argued that the resampling algorithm retains the key features of the original dataset under Ass.1-5. This however does not guarantee that the asymptotic distribution of the CCE estimators will also be replicated. Conditions for this are investigated next.

REMARK 1 *Note that while our assumptions and DGP exclude cross-section fixed effects, this is for ease of exposition only. It can be shown that the distribution of the CCE estimators is invariant to these effects so long as a* $T \times 1$ *row of constants* $\boldsymbol{\iota}_T = [1, \ldots, 1]'$ *is added to the matrix of CA, that is,* $\overline{\mathbf{H}} = [\boldsymbol{\iota}_T, \overline{\mathbf{Z}}]$*, such that* $\mathbf{M}_{\widehat{\mathbf{F}}} = \mathbf{I}_T - \overline{\mathbf{H}}(\overline{\mathbf{H}}'\overline{\mathbf{H}})^{\dagger}\overline{\mathbf{H}}'$ *(see for instance Lemma 1 in De Vos and Everaert, 2021, in a dynamic setting). This implies that the CS-resampling scheme and the conditions for its consistency presented* *below apply directly to the fixed effects setting, provided a vector of ones is added to the CA, regardless of whether those fixed effects are cross-sectionally dependent or not.*

3.1.2 Asymptotic analysis

Recall that $\mathbf{Z}_i^* = [\mathbf{y}_i^*, \mathbf{X}_i^*]$, let $\overline{\mathbf{Z}}^* = \frac{1}{N}\sum_{i=1}^N \mathbf{Z}_i^*$ be the CA of the observables generated from (3.9) in the bootstrap sample and let $\mathbf{M}_{\widehat{\mathbf{F}}^*} = \mathbf{I}_T - \overline{\mathbf{Z}}^*((\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*)^\dagger(\overline{\mathbf{Z}}^*)'$ be the corresponding orthogonalization matrix. The CCEP estimator in the bootstrap dataset is then

$$\widehat{\boldsymbol{\beta}}^* = \left(\sum_{i=1}^N \mathbf{X}_i^{*\prime}\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{X}_i^*\right)^{-1}\sum_{i=1}^N \mathbf{X}_i^{*\prime}\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{y}_i^* = \left(\sum_{i=1}^N s_i\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{X}_i\right)^{-1}\sum_{i=1}^N s_i\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{y}_i \tag{3.10}$$

where s_i denotes the frequency with which cross-section i was resampled. Theorem 2 establishes the asymptotic distribution of (3.10) in the bootstrap world as $(N, T) \to \infty$ such that $T/N \to \tau < \infty$.

Theorem 2 *Let Ass.1-5 hold. Then, as $(N, T) \to \infty$ such that $T/N \to \tau < \infty$*

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \xrightarrow{d^*} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}) + \sqrt{\tau}\boldsymbol{\Sigma}^{-1}(\mathbf{b} - \mathbf{d} - \mathbf{d}^+)$$

where $\mathbf{d}^+ = (1/2)\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \boldsymbol{\Sigma}_i[\boldsymbol{\beta}, \mathbf{I}_k]\mathbf{D}_{-m}[\sigma_i^2, \mathbf{0}_{1\times k}]'$ and $\mathbf{b}$, $\mathbf{d}$, $\boldsymbol{\Psi}$ and $\mathbf{D}_{-m}$ are defined in Theorem 1, with $\mathbf{d} = \mathbf{d}^+ = \mathbf{0}_{k\times 1}$ if $m = 1 + k$. If either $\tau = 0$ or $m = 1 + k$, then

$$\sup_{x\in\mathbb{R}^{k\times 1}}\left|\mathbb{P}^*[\sqrt{NT}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \le x] - \mathbb{P}[\sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \le x]\right| \longrightarrow^p 0,$$

where inequalities should be interpreted coordinate-wise.

This result reveals that the bootstrap replicates the distribution of the CCEP estimator in Theorem 1, but that it also generates an additional bias $\mathbf{d}^+$ when and $m < 1 + k$. The new term is effectively the exacerbation of $\mathbf{d}$ in the bootstrap distribution, and it implies that the bootstrap is only consistent when either $\tau = 0$ (no bias), or when the number of CA equals the number of factors $m = 1 + k$ (as $\mathbf{d}$ is then absent and cannot be magnified). This is the last statement of the theorem.

As is, the conclusions from Theorem 2 are relatively disappointing. Bootstrap bias-corrections, confidence intervals and tests, to be discussed in further detail below, are only applicable to $\widehat{\boldsymbol{\beta}}$ when either $\tau = 0$ or $m = 1 + k$. The obvious practical issue, as with analytical corrections, is that the researcher is typically unaware of whether $m < 1 + k$ or $m = 1 + k$ so that it is unclear whether bias will remain and conclusions can be trusted. Yet, close inspection reveals that $\mathbf{d}^+$

is a consequence of the dependence generated by estimating the factors with $\overline{\mathbf{y}}^*$ (in combination with $m < 1 + k$). Hence, the solution is to exclude $(\overline{\mathbf{y}}, \overline{\mathbf{y}}^*)$ from the employed set of CA in both the original and the bootstrap world and use $\overline{\mathbf{X}}$ ($\overline{\mathbf{X}}^*$) in stead of $\overline{\mathbf{Z}}$ ($\overline{\mathbf{Z}}^*$) to estimate the unknown factors. The implied reduction of information requires that Ass.4 is sharpened to Ass.7 below to ensure that $\overline{\mathbf{X}}$ still consistently estimates the factor space.

Assumption 7 *(Rank condition) $rk(\overline{\boldsymbol{\Gamma}}) = m$ for all N.*

Consider then the CCEP estimator that excludes $\overline{\mathbf{y}}$ from the estimation in the original sample

$$\widehat{\boldsymbol{\beta}}_{\mathbf{x}} = \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}_{\mathbf{x}}} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}_{\mathbf{x}}} \mathbf{y}_i, \qquad \text{with} \qquad \mathbf{M}_{\widehat{\mathbf{F}}_{\mathbf{x}}} = \mathbf{I}_T - \overline{\mathbf{X}}(\overline{\mathbf{X}}'\overline{\mathbf{X}})^{\dagger}\overline{\mathbf{X}}' \tag{3.11}$$

Corollary 1 presents its asymptotic distribution. An important observation is that the asymptotic variance of $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ is identical to that of $\widehat{\boldsymbol{\beta}}$, which means that no asymptotic efficiency is lost by the exclusion of $\overline{\mathbf{y}}$ under Ass.7. In addition, while the bias remains inestimable for analytical correction purposes, it is significantly less involved than that in Theorem 1.

Corollary 1 *Let Assumptions 1-3, 5 and 7 hold. Then, as $(N, T) \to \infty$ such that $T/N \to \tau < \infty$*

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}) + \sqrt{\tau}\boldsymbol{\Sigma}^{-1}\mathbf{g}$$

where

$$\mathbf{g} = \mathbf{q}_{xy}' \boldsymbol{\Sigma}_{\boldsymbol{\eta}}' vec((\boldsymbol{\Gamma}^{\dagger})'\boldsymbol{\Sigma}(\mathbf{I}_k - \mathbf{D}_{\mathbf{x},-m}\boldsymbol{\Sigma})\boldsymbol{\Gamma}^{\dagger})$$

and $\mathbf{D}_{\mathbf{x},-m} = \mathbf{T}_{\mathbf{x}}\mathbf{H}_{\mathbf{x},-m}(\mathbf{H}_{\mathbf{x},-m}'\mathbf{T}_{\mathbf{x}}'\boldsymbol{\Sigma}\mathbf{T}_{\mathbf{x}}\mathbf{H}_{\mathbf{x},-m})^{\dagger}\mathbf{H}_{\mathbf{x},-m}'\mathbf{T}_{\mathbf{x}}'$, with $\mathbf{T}_{\mathbf{x}}$ the $k \times k$ partitioning matrix such that $\boldsymbol{\Gamma}\mathbf{T}_{\mathbf{x}} = [\boldsymbol{\Gamma}_m, \boldsymbol{\Gamma}_{-m}]$ with $\boldsymbol{\Gamma}_m$ an $m \times m$ full rank matrix and $\boldsymbol{\Gamma}_{-m}$ is $m \times (k-m)$, and $\mathbf{H}_{\mathbf{x},-m} = [-(\boldsymbol{\Gamma}_m^{-1}\boldsymbol{\Gamma}_{-m})', \mathbf{I}_{k-m}]'$.

The reduced number of terms compared to Theorem 1 does not, however, guarantee that the asymptotic bias of $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ is also smaller than that of $\widehat{\boldsymbol{\beta}}$. The relative size depends on how much information about the factors is lost by excluding $\overline{\mathbf{y}}$, as reflected by the presence of $\boldsymbol{\Gamma}^{\dagger}$ rather than $\mathbf{C}^{\dagger}$ of Theorem 1. That is, if the population means $\boldsymbol{\Gamma}$ are sufficiently less informative than $\mathbf{C} = [\gamma + \boldsymbol{\Gamma}\boldsymbol{\beta}, \boldsymbol{\Gamma}]$, asymptotic bias may yet be aggravated due to $\left\|\boldsymbol{\Gamma}^{\dagger}\right\| > \left\|\mathbf{C}^{\dagger}\right\|$. Our Monte Carlo experiments of Section 4 confirm that the bias of $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ tends to be larger than that of $\widehat{\boldsymbol{\beta}}$.
Hence, $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ may be more biased than $\widehat{\boldsymbol{\beta}}$, but we nevertheless posit that $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ is better suited for correction because the bootstrap distribution of

$$\widehat{\boldsymbol{\beta}}_{\mathbf{x}}^* = \left(\sum_{i=1}^{N} \mathbf{X}_i^{*\prime} \mathbf{M}_{\widehat{\mathbf{F}}_{\mathbf{x}}^*} \mathbf{X}_i^*\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i^{*\prime} \mathbf{M}_{\widehat{\mathbf{F}}_{\mathbf{x}}^*} \mathbf{y}_i^* \quad \text{with} \quad \mathbf{M}_{\widehat{\mathbf{F}}_{\mathbf{x}}^*} = \mathbf{I}_T - \overline{\mathbf{X}}^*((\overline{\mathbf{X}}^*)'\overline{\mathbf{X}}^*)^{\dagger}(\overline{\mathbf{X}}^*)' \tag{3.12}$$

should not magnify bias when $m < k$ (note that $\overline{\mathbf{X}}$ has k in stead of $k+1$ columns). This is confirmed by Corollary 2.

Corollary 2 *Under Ass.1-3, 5 and 7, as $(N,T) \to \infty$ such that $T/N \to \tau < \infty$,*

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{\mathbf{x}}^{*} - \widehat{\boldsymbol{\beta}}_{\mathbf{x}}) \xrightarrow{d^*} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}) + \sqrt{\tau}\boldsymbol{\Sigma}^{-1}\mathbf{g}$$

with $\mathbf{g}$ defined in Corollary 1, and

$$\sup_{x\in\mathbb{R}^{k\times 1}} \left| \mathbb{P}^*[\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{\mathbf{x}}^{*} - \widehat{\boldsymbol{\beta}}_{\mathbf{x}}) \leq x] - \mathbb{P}[\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \boldsymbol{\beta}) \leq x] \right| \longrightarrow^{p} 0,$$

where the inequalities are interpreted coordinate-wise.

REMARK 2 *Note that Assumption 7 and Corollary 1 and 2 are perfectly compatible with the usual DGPs considered in the CCE literature. That is, the same $\mathbf{F}$ enters equations for $\mathbf{X}_i$ and $\mathbf{y}_i$. Provided this, we only need to increase informativeness of $\boldsymbol{\Gamma}_i$ to compensate for dependence reduction to make the popular pairs bootstrap consistent in the CCE context when more CAs than needed are used as proxies for the factor space. This is the main intuition behind Assumption 7. Interestingly, slightly more general DGPs resembling those of quasi-maximum likelihood (QMLE) literature can be compatible with this sharpened requirement. Following, Juodis et al. (2021), we can have a $T \times m_y$ matrix $\mathbf{F}_{\mathbf{y}}$ which represents the factors that enter equation of $\mathbf{y}_i$. Let $\mathbf{X}_i$ load on $T \times m_x$ matrix $\mathbf{F}_{\mathbf{X}}$. If $\mathbf{F}_{\mathbf{y}} \subseteq \mathbf{F}_{\mathbf{X}}$, then $m = m_x$, and under Assumption 7 the results presented above still go through, because $\mathbf{X}_i$ is sufficiently informative about the factor space. On the other hand, if $\mathbf{F}_{\mathbf{X}} \subset \mathbf{F}_{\mathbf{y}}$, the proposed solution could result in a loss of information that cannot be compensated without further restrictions on DGP. One possible restriction is to let the loadings of the factors entering $\mathbf{y}_i$ only, say $\gamma_{i,-\mathbf{X}} \in \mathbb{R}^{(m_y-m_x)\times 1}$, be uncorrelated with the other loadings. Similar loading properties are considered in, for example, Pesaran (2006), Chudik et al. (2011) and Kapetanios et al. (2011).*

The above corollaries establish the consistency of the bootstrap for the distribution of $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ for general $m \leq k$, and hence validates the construction of bootstrap confidence intervals for this estimator in a general setting. The next theorem confirms that the obtained bootstrap mean can also be used for explicit bias-correction of $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ (and leads to over-correction for $\widehat{\boldsymbol{\beta}}$).

Theorem 3 *Under Ass.1-5 strengthened with $\mathbb{E}(\|\mathbf{v}_{i,t}\|^8) < \infty$ we have as $(N,T) \to \infty$ such that $T/N \to \tau < \infty$ that $\mathbf{A}^* \longrightarrow^{p^*} \sqrt{\tau}\mathbf{A}$, where $\mathbf{A}^* = \mathbb{E}^*(\sqrt{NT}(\widehat{\boldsymbol{\beta}}^{*} - \widehat{\boldsymbol{\beta}}))$ and $\mathbf{A} = \boldsymbol{\Sigma}^{-1}(\mathbf{b} - \mathbf{d} - \mathbf{d}^{+})$. If also Ass.7 holds, then $\mathbf{A}_{\mathbf{x}}^* \longrightarrow^{p^*} \sqrt{\tau}\mathbf{A}_{\mathbf{x}}$, with $\mathbf{A}_{\mathbf{x}}^* = \mathbb{E}^*(\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{\mathbf{x}}^{*} - \widehat{\boldsymbol{\beta}}_{\mathbf{x}}))$ and $\mathbf{A}_{\mathbf{x}} = \boldsymbol{\Sigma}^{-1}\mathbf{g}$.*

Consider then the following bootstrap estimator for the bias of $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$

$$\widehat{\mathbf{b}}_{\mathbf{x}}^{*} = \overline{\boldsymbol{\beta}}_{\mathbf{x}}^{*} - \widehat{\boldsymbol{\beta}}_{\mathbf{x}} \tag{3.13}$$

where $\overline{\boldsymbol{\beta}}^*_{\mathbf{x}} = \frac{1}{B}\sum_{b=1}^{B}\widehat{\boldsymbol{\beta}}^*_{\mathbf{x},b}$ and $\widehat{\boldsymbol{\beta}}^*_{\mathbf{x},b}$ is the CCEP estimator in (3.12) applied to bootstrap sample $b = 1, \ldots, B$. Theorem 3 establishes that $\widehat{\mathbf{b}}^*_{\mathbf{x}}$ is consistent for the bias of $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ derived in Corollary 1, and by consequence, that the bootstrap adjusted estimator

$$\widehat{\boldsymbol{\beta}}_{\mathbf{x},c} = \widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \widehat{\mathbf{b}}^*_{\mathbf{x}}$$

is asymptotically unbiased as $(N, T, B) \to \infty$ such that $\tau < \infty$ in the general $m \leq k$ setting. It is important to reiterate the relative ease with which this bias correction is achieved: it does not require an analytical formula or knowledge of m, only the nonparametric resampling scheme in (3.9) and calculation of the CCEP estimates.

Corollary 2 justifies inferences on $\boldsymbol{\beta}$ with the 'basic' $100(1-\alpha)\%$ confidence interval

$$CI(\alpha, \widehat{\boldsymbol{\beta}}^*_{\mathbf{x}}) = [2\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \boldsymbol{\theta}^*_{(1-\alpha/2)}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}}), 2\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \boldsymbol{\theta}^*_{\alpha/2}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}})] \tag{3.14}$$

with $\boldsymbol{\theta}^*_{\alpha}(\cdot)$ the empirical α-quantile of the obtained bootstrap distribution for the statistic inside the brackets. The implicit bias adjustment means that the interval attains the coordinate-wise nominal coverage $\mathbb{P}[\boldsymbol{\beta} \in CI(\alpha, \widehat{\boldsymbol{\beta}}^*_{\mathbf{x}})] = 1 - \alpha$ as $(N, T, B) \to \infty$ provided $m \leq k$ and $\tau < \infty$.[9] To establish also the asymptotic validity of bootstrap$-t$ intervals, define $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}$ and let $\widehat{\boldsymbol{\Theta}}^*$ be the bootstrap world equivalent of the variance estimator in (3.5)

$$\widehat{\boldsymbol{\Theta}}^* = N^{-1}\overline{\mathbf{Q}}^{*-1}\widehat{\boldsymbol{\Psi}}^*\overline{\mathbf{Q}}^{*-1}, \qquad \widehat{\boldsymbol{\Psi}}^* = \frac{1}{N-1}\sum_{i=1}^{N}\widehat{\mathbf{Q}}^*_i(\widehat{\boldsymbol{\beta}}^*_i - \widehat{\boldsymbol{\beta}}^*_{mg})(\widehat{\boldsymbol{\beta}}^*_i - \widehat{\boldsymbol{\beta}}^*_{mg})'\widehat{\mathbf{Q}}^*_i \tag{3.15}$$

where $\overline{\mathbf{Q}}^* = \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}^*_i$ and

$$\widehat{\mathbf{Q}}^*_i = T^{-1}\mathbf{X}^{*\prime}_i\mathbf{M}_{\widehat{\mathbf{F}}^*_{\mathbf{x}}}\mathbf{X}^*_i, \qquad \widehat{\boldsymbol{\beta}}^*_i = \widehat{\mathbf{Q}}^{*-1}_i T^{-1}\mathbf{X}^{*\prime}_i\mathbf{M}_{\widehat{\mathbf{F}}^*_{\mathbf{x}}}\mathbf{y}^*_i, \qquad \widehat{\boldsymbol{\beta}}^*_{mg} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}^*_i$$

with the obvious substitution of the projection matrices with $\mathbf{M}_{\widehat{\mathbf{F}}^*}$ when $(\overline{\mathbf{y}}, \overline{\mathbf{y}}^*)$ are included in the matrix of CA. Theorem 8 in the supplement shows that $NT\widehat{\boldsymbol{\Theta}} \to^p \boldsymbol{\Theta}$ and $NT\widehat{\boldsymbol{\Theta}}^* \to^{p^*} \boldsymbol{\Theta}$ as $(N, T) \to \infty$, irrespective of whether the number of employed CA exceeds or equals m or if $(\overline{\mathbf{y}}, \overline{\mathbf{y}}^*)$ are also employed to estimate factors. It then follows from Corollary 1, for the $l-$th coordinate of $\boldsymbol{\beta}$ in the original sample

$$t = \frac{\mathbf{q}'_l(\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \boldsymbol{\beta})}{\sqrt{\mathbf{q}'_l\widehat{\boldsymbol{\Theta}}\mathbf{q}_l}} \xrightarrow{d} \mathcal{N}(0,1) + \sqrt{\tau}\frac{\mathbf{q}'_l\boldsymbol{\Sigma}^{-1}\mathbf{g}}{\sqrt{\mathbf{q}'_l\boldsymbol{\Theta}\mathbf{q}_l}} \tag{3.16}$$

[9]Note that the absence of an implicit bias-adjustment implies that the classical percentile interval $CI_p(\alpha, \widehat{\boldsymbol{\beta}}^*_{\mathbf{x}}) = [\boldsymbol{\theta}^*_{\alpha/2}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}}), \boldsymbol{\theta}^*_{(1-\alpha/2)}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}})]$ will require in stead $\tau = 0$ to attain nominal asymptotic coverage. See e.g. DiCiccio and Efron (1996) for several more advanced percentile bootstrap methods.

and for the studentized bootstrap statistic, by Corollary 2,

$$t^* = \frac{\mathbf{q}_l'(\widehat{\boldsymbol{\beta}}_{\mathbf{x}}^* - \widehat{\boldsymbol{\beta}}_{\mathbf{x}})}{\sqrt{\mathbf{q}_l'\widehat{\boldsymbol{\Theta}}^*\mathbf{q}_l}} \xrightarrow{d^*} \mathcal{N}(0,1) + \sqrt{\tau}\frac{\mathbf{q}_l'\boldsymbol{\Sigma}^{-1}\mathbf{g}}{\sqrt{\mathbf{q}_l'\boldsymbol{\Theta}\mathbf{q}_l}} \tag{3.17}$$

as $(N,T) \to \infty$ such that $\tau < \infty$. It is apparent from (3.16) that the asymptotic bias also shifts the center of the distribution of the t-statistic based on $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ and causes size distortions for the classical t-test unless $\tau = 0$. Yet, given that $t \sim t^*$ by (3.16)-(3.17) under the null, asymptotically correct size is achieved with the bootstrap as $(N,T,B) \to \infty$ such that $\tau < \infty$ by rejecting the null hypothesis when $t > \boldsymbol{\theta}^*_{(1-\alpha/2)}(t^*)$ or $t < \boldsymbol{\theta}^*_{(\alpha/2)}(t^*)$. This is equivalent to rejecting when $\boldsymbol{\beta}_0$ of the null hypothesis falls outside the 'bootstrap-t' interval constructed with the roots (t,t^*) (see also van Giersbergen and Kiviet, 2002):

$$CI_t(\alpha, \widehat{\boldsymbol{\beta}}_{\mathbf{x}}^*) = [\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - diag(\widehat{\boldsymbol{\Theta}}^{1/2}) \circ \boldsymbol{\theta}^*_{(1-\alpha/2)}(t^*), \widehat{\boldsymbol{\beta}}_{\mathbf{x}} - diag(\widehat{\boldsymbol{\Theta}}^{1/2}) \circ \boldsymbol{\theta}^*_{(\alpha/2)}(t^*)] \tag{3.18}$$

Bootstrap-t intervals date back to Efron (1979, 1981) and are widely advocated in econometrics because the studentized roots in (3.16)-(3.17) lead to improved finite sample behavior, compared to say (3.14), if they are also (asymptotically) pivotal (see e.g. Diciccio and Romano, 1988; Beran, 1987; Hall, 1988; Hall et al., 1996). The bias terms in (3.16)-(3.17) imply, however, that the roots (t,t^*) are not pivotal in our setting unless $\tau = 0$, but we can re-pivot them with the bias estimate in (3.13). Let (t_c, t_c^*) be these corrected roots. We then have by Corollaries 1,2 and Theorems 3,8

$$t_c = \frac{\mathbf{q}_l'(\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \widehat{\mathbf{b}}_{\mathbf{x}}^* - \boldsymbol{\beta})}{\sqrt{\mathbf{q}_l'\widehat{\boldsymbol{\Theta}}\mathbf{q}_l}} \xrightarrow{d} \mathcal{N}(0,1), \qquad t_c^* = \frac{\mathbf{q}_l'(\widehat{\boldsymbol{\beta}}_{\mathbf{x}}^* - \widehat{\mathbf{b}}_{\mathbf{x}}^* - \widehat{\boldsymbol{\beta}}_{\mathbf{x}})}{\sqrt{\mathbf{q}_l'\widehat{\boldsymbol{\Theta}}^*\mathbf{q}_l}} \xrightarrow{d^*} \mathcal{N}(0,1)$$

as $(N,T,B) \to \infty$ such that $\tau < \infty$. That is, the corrected roots (t_c, t_c^*) are asymptotically pivotal and lead to the adjusted bootstrap-t_c confidence interval

$$CI_{t,c}(\alpha, \widehat{\boldsymbol{\beta}}_{\mathbf{x}}^*) = [\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \widehat{\mathbf{b}}_{\mathbf{x}}^* - diag(\widehat{\boldsymbol{\Theta}}^{1/2}) \circ \boldsymbol{\theta}^*_{(1-\alpha/2)}(t_c^*), \widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \widehat{\mathbf{b}}_{\mathbf{x}}^* - diag(\widehat{\boldsymbol{\Theta}}^{1/2}) \circ \boldsymbol{\theta}^*_{(\alpha/2)}(t_c^*)] \tag{3.19}$$

Hypothesis tests based on (3.19) are equivalent to rejecting the null hypothesis when $t_c < \boldsymbol{\theta}^*_{(\alpha/2)}(t_c^*)$ or $t_c > \boldsymbol{\theta}^*_{(1-\alpha/2)}(t_c^*)$. Since the confidence interval and hypothesis tests are now based on asymptotically pivotal roots, classical theory (e.g. Hall, 1988; Diciccio and Romano, 1988) suggests that they will have better finite sample properties than (3.18). This is confirmed by the Monte Carlo experiments of Section 4.

To sum up, we have shown with the analysis above that the cross-section bootstrap allows elimination of the asymptotic bias of the CCEP estimator and enables asymptotically unbiased inferences in panels where $0 \leq \tau < \infty$. This is a considerable generalization compared to standard normal theory intervals and tests, which were shown to require the highly stringent $\tau = 0$ restriction. We have argued that the procedures are most generally effective when applied to $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ compared to $\widehat{\boldsymbol{\beta}}$, as bootstrap consistency for the former does not depend on whether the number of CA exceeds or equals the number of factors, which is unknown. Bias adjustment and interval construction can in other words proceed for $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ without this knowledge, while $\widehat{\boldsymbol{\beta}}$ requires in stead verification that $m = 1 + k$.[10] The only cost associated with $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ compared to $\widehat{\boldsymbol{\beta}}$ is that the maximum number of factors that can be allowed is reduced from $k + 1$ to k, but there is otherwise no asymptotic efficiency loss.

3.2 Heterogeneous slopes: $\boldsymbol{\beta}_i \neq \boldsymbol{\beta}$

Consider next the case where the slope coefficients $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N)$ in the model are heterogeneous over individuals and characterized by Ass.6. We take that the researcher is interested in the population mean $\mathbb{E}(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$ in

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{F}\gamma_i + \boldsymbol{\varepsilon}_i, \tag{3.20}$$

The CCEP estimator has the following asymptotic distribution in this model:

Theorem 4 *Under Ass.1-6, with* $\mathbb{E}(\|\mathbf{v}_{i,t}\|^8) < \infty$ *and* $\mathbb{E}(\|\boldsymbol{v}_i\|^6) < \infty$, *as* $(N, T) \to \infty$

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}_h\boldsymbol{\Sigma}^{-1})$$

where $\boldsymbol{\Psi}_h = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \boldsymbol{\Sigma}_i \boldsymbol{\Omega}_v \boldsymbol{\Sigma}_i$.

This is, to the best of our knowledge, the first derivation of the asymptotic distribution of the CCEP estimator under heterogeneous slopes which also allows $m < 1 + k$ settings. This result is thus a generalization of that in Pesaran (2006), and reveals that with the potential for $m < 1 + k$ comes the additional requirement that 6th and 8th moments of $\boldsymbol{v}_i$ and $\mathbf{v}_{it}$, respectively, are bounded. Restrictions on the relative expansion rate of T and N are then not required, but these moment bounds can also be dispensed with by imposing in stead that $\sqrt{N}/T \to 0$. The main conclusion remains that the estimator is now asymptotically unbiased because the convergence rate is reduced to $\sqrt{N}$ compared to the $\sqrt{NT}$ rate in the homogeneous slope setting. This is because the heterogeneity in $\boldsymbol{\beta}_i$ is now the slowest decaying error component. The distribution of $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ is identical to that in Theorem 4 under Ass.7.

The cross-section resampling scheme of section 3.1.1 can also be applied to the CCE estimators when slopes are heterogeneous. The only new requirement is that the $\boldsymbol{\beta}_i$ are independent over

[10] Note that should $m = 1 + k$ or $\tau = 0$ hold true, then (3.14), (3.18) and (3.19) can be applied to $\widehat{\boldsymbol{\beta}}$.

cross-sections as in Ass.6. In this case, the absence of asymptotic bias in Theorem 4 implies that bootstrap consistency no longer hinges on the exclusion of $(\bar{\mathbf{y}}, \bar{\mathbf{y}}^*)$ when $m < 1 + k$. This is formalized in the next theorem.

Theorem 5 *Under the conditions of Theorem 4*

$$\sup_{x \in \mathbb{R}^{k \times 1}} \left| \mathbb{P}^*[\sqrt{N}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \leq x] - \mathbb{P}[\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq x] \right| \longrightarrow^p 0,$$

where inequalities are to be interpreted coordinate-wise.

The result is identical for $(\widehat{\boldsymbol{\beta}}_{\mathbf{x}}^*, \widehat{\boldsymbol{\beta}}_{\mathbf{x}})$ under Ass.7. Theorem 5 establishes the validity of percentile and basic bootstrap intervals as $(N, T, B) \to \infty$ without further restrictions on N and T. In addition, by the consistency of both $N\widehat{\boldsymbol{\Theta}}$ and $N\widehat{\boldsymbol{\Theta}}^*$ in (3.15) for the asymptotic variance $\boldsymbol{\Theta}_h = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}_h\boldsymbol{\Sigma}^{-1}$ of Theorem 4 (which is established in Theorem 8 of the supplement) the bootstrap-t intervals presented in (3.18) and (3.19) also give asymptotically correct coverage when slopes are heterogeneous. This applies irrespective of whether $(\bar{\mathbf{y}}, \bar{\mathbf{y}}^*)$ are employed in the estimation, provided a rank condition holds (Ass.4 or 7). In other words, also

$$CI_t(\alpha, \widehat{\boldsymbol{\beta}}^*) = [\widehat{\boldsymbol{\beta}} - diag(\widehat{\boldsymbol{\Theta}}^{1/2}) \circ \boldsymbol{\theta}^*_{(1-\alpha/2)}(t^*), \widehat{\boldsymbol{\beta}} - diag(\widehat{\boldsymbol{\Theta}}^{1/2}) \circ \boldsymbol{\theta}^*_{(\alpha/2)}(t^*)] \tag{3.21}$$

has the desired coverage in this setting, where it should be clear that $t^* = \mathbf{q}_l'(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})/\sqrt{\mathbf{q}_l'\widehat{\boldsymbol{\Theta}}^*\mathbf{q}_l}$. Note that the absence of bias now means that the roots (t, t^*) are asymptotically pivotal so that the interval can be motivated by the usual arguments. Bias-adjustment as in

$$CI_{t,c}(\alpha, \widehat{\boldsymbol{\beta}}^*) = [\widehat{\boldsymbol{\beta}} - \widehat{\mathbf{b}}^* - diag(\widehat{\boldsymbol{\Theta}}^{1/2}) \circ \boldsymbol{\theta}^*_{(1-\alpha/2)}(t_c^*), \widehat{\boldsymbol{\beta}} - \widehat{\mathbf{b}}^* - diag(\widehat{\boldsymbol{\Theta}}^{1/2}) \circ \boldsymbol{\theta}^*_{(\alpha/2)}(t_c^*)] \tag{3.22}$$

with $t_c^* = \mathbf{q}_l'(\widehat{\boldsymbol{\beta}}^* - \widehat{\mathbf{b}}^* - \widehat{\boldsymbol{\beta}})/\mathbf{q}_l'(\widehat{\boldsymbol{\Theta}}^*)^{1/2}\mathbf{q}_l$ is thus not strictly necessary, but it is asymptotically innocuous due to $\widehat{\mathbf{b}}^* \to^{p^*} \mathbf{0}_{k\times 1}$ as $(N, T, B) \to \infty$, and may yet improve finite sample behavior as the CCEP estimator is 'only' asymptotically unbiased (and not in small samples).

The main conclusion from the analysis so far is that the resampling scheme in (3.9) and the resulting bias adjustments and bootstrap confidence intervals of section 3.1 also lead to asymptotically unbiased inferences on $\mathbb{E}(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$ when slopes are heterogeneous. The researcher can in other words safely apply the exact same bootstrap procedures indiscriminately and remain agnostic about whether slopes vary over individuals in practice. Notwithstanding, efficiency gains could be achieved by applying in stead of CCEP the Mean Group CCE (CCEMG) estimator defined in (3.6). Given that slope heterogeneity is also here the slowest decaying source of error, the CCEMG estimator converges at the slower $\sqrt{N}$ rate as well and is asymptotically unbiased in the general $m \leq 1 + k$ setting.

Theorem 6 *Under Ass.1-6 as $(N, T) \to \infty$*

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{mg} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Omega}_{\boldsymbol{v}})$$

This is the first result for the CCEMG estimator that accounts for the potential $m < 1 + k$ problem. Nevertheless, the asymptotic distribution is identical to that in Pesaran (2006).

The bootstrap can also in this setting lead to refinements by for instance reducing finite sample bias (also $\widehat{\boldsymbol{\beta}}_{mg}$ is only asymptotically unbiased). In addition, if the distribution of the slopes is non-normal and skewed, bootstrap percentiles and critical values can be substantially more accurate than the normal approximation in finite samples. The next theorem establishes bootstrap consistency. Results are identical when $(\overline{\mathbf{y}}, \overline{\mathbf{y}}^*)$ is excluded.

Theorem 7 *Under the conditions of Theorem 6, as $(N, T) \to \infty$*

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^*_{mg} - \widehat{\boldsymbol{\beta}}_{mg}) \xrightarrow{d^*} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Omega}_{\boldsymbol{v}}),$$

and in addition

$$\sup_{x\in\mathbb{R}^{k\times 1}} \left| \mathbb{P}^*[\sqrt{N}(\widehat{\boldsymbol{\beta}}^*_{mg} - \widehat{\boldsymbol{\beta}}_{mg}) \leq x] - \mathbb{P}[\sqrt{N}(\widehat{\boldsymbol{\beta}}_{mg} - \boldsymbol{\beta}) \leq x] \right| \longrightarrow^p 0,$$

where inequalities are to be interpreted coordinate-wise.

Theorem 8 in the supplement establishes that the bootstrap world equivalent of (3.7)

$$\widehat{\boldsymbol{\Omega}}^*_{\boldsymbol{v}} = \frac{1}{N(N-1)} \sum_{i=1}^{N} (\widehat{\boldsymbol{\beta}}^*_i - \widehat{\boldsymbol{\beta}}^*_{mg})(\widehat{\boldsymbol{\beta}}^*_i - \widehat{\boldsymbol{\beta}}^*_{mg})'$$

also consistently estimates the variance in Theorem 6, $N\widehat{\boldsymbol{\Omega}}^*_{\boldsymbol{v}} \to^{p^*} \boldsymbol{\Omega}_{\boldsymbol{v}}$ as $(N, T) \to \infty$. It thus follows in combination with Theorem 7 that the percentile, basic and (corrected) bootstrap-t intervals apply equally to the CCEMG estimator. That is, $CI(\alpha, \widehat{\boldsymbol{\beta}}^*_{mg})$, $CI_t(\alpha, \widehat{\boldsymbol{\beta}}^*_{mg})$ and $CI_{t,c}(\alpha, \widehat{\boldsymbol{\beta}}^*_{mg})$, with obvious adaptations to the CCEMG estimator, have asymptotically correct coverage.

4 Monte Carlo Simulation

In this section we assess the finite properties for the algorithms presented above.

4.1 Design

Data for $\mathbf{y}_i$ and $\mathbf{X}_i$ are generated according to eqs.(2.1)-(2.2), assuming $m = 2$ unobserved factors and $k = 3$ regressors. This corresponds to the likely practical setting where there are multiple factors but their number is exceeded by the number of regressors. Slope coefficients are

$$\boldsymbol{\beta}_i = \beta \boldsymbol{\iota}_{k\times 1} + \boldsymbol{v}_i, \qquad \textit{with} \qquad v_{i,l} \sim (\chi^2_1 - 1)\sqrt{\sigma^2_v/2} \qquad \text{for } l = 1, \ldots, k$$

where $v_{i,l}$ denotes the l-th row of $\boldsymbol{v}_i$, so that $\sigma^2_v \in (0, 5)$ considers respectively the common and variable slopes setting. We vary the slope population mean as $\beta \in (1, 5)$ to also determine its impact on bias. Time varying unobservables in (2.1)-(2.2) follow

$$\begin{aligned}
\mathbf{f}_t &= \theta \mathbf{f}_{t-1} + \sqrt{1-\theta^2}\boldsymbol{\nu}^f_t, \qquad & \boldsymbol{\nu}^f_t &\sim \mathcal{N}(\mathbf{0}_{m\times 1}, \mathbf{I}_m/m) \\
\varepsilon_{i,t} &= \rho \varepsilon_{i,t-1} + \sqrt{1-\rho^2}\boldsymbol{\nu}^\varepsilon_{i,t}, \qquad & \boldsymbol{\nu}^\varepsilon_{i,t} &\sim \mathcal{N}(0, \sigma^2_i) \\
\mathbf{v}_{i,t} &= \rho_x \mathbf{v}_{i,t-1} + \sqrt{1-\rho_x^2}\boldsymbol{\nu}^x_{i,t}, \qquad & \boldsymbol{\nu}^x_{i,t} &\sim \mathcal{N}(\mathbf{0}_{k\times 1}, \sigma^2_{x,i}\mathbf{I}_k)
\end{aligned}$$

where each variable is initiated at 0 and the first 50 periods are discarded as a burn-in to neutralize the initial conditions. In accordance with the high serial correlation that is typically encountered in practice we set $\rho = \rho_x = \theta = 0.8$ for all experiments. To further illustrate the robustness to heteroscedasticity, variances in the processes are drawn from $\sigma^2_i = \sigma^2 + (\omega_{1,i} - 1)$ and $\sigma^2_{x,i} = \sigma^2_x + (\omega_{2,i} - 1)$ respectively, where $\omega_{1,i} \sim \chi^2_1$ and $\omega_{2,i} \sim \chi^2_1$. We set $\sigma^2_x = 2$ for all experiments to ensure, given $k = 3$, a minimal signal to noise level.

As discussed below Theorem 1, an important driver of the asymptotic bias of the CCEP estimator is the extent to which the chosen set of cross-section averages are (asymptotically) informative about the unobserved factors. We measure this with the determinant $d = |\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'|$ in our experiments and control it by choosing an upper bound d^u and generating the entries in $\widetilde{\mathbf{C}}$ independently from $\mathcal{U}[0, 2]$ such that $d^u - 0.1 \leq d \leq d^u$. The obtained population mean $\widetilde{\mathbf{C}}$ that adheres to this restriction is then fixed over Monte Carlo replications and sample sizes. We take $d = 10$ as our baseline scenario with good information content, and study the impact of a less informative setting by lowering d to 5.[11] To avoid that the majority of the information stems only from $\overline{\mathbf{y}}$ we impose that $d_x = |\boldsymbol{\Gamma}\boldsymbol{\Gamma}'|$ adheres to the same bounds as d.[12] Given $\widetilde{\mathbf{C}}$, the cross-section-specific loadings are generated as $\widetilde{\mathbf{C}}_i = \widetilde{\mathbf{C}} + \widetilde{\boldsymbol{\eta}}_i \boldsymbol{\iota}'_{1+k}$, with $\widetilde{\boldsymbol{\eta}}_i \sim \mathcal{N}(\mathbf{0}_{m\times 1}, \sigma^2_\eta \mathbf{I}_m)$. This implies that loadings

[11] These numbers are based on the (simulated) distribution of the determinant of 2×4 matrices with elements drawn from $\mathcal{U}[0, 2]$, which ranges roughly from 0 to 40 (with a long right tail) with $\mathbb{E}(d) \approx 9.2$.

[12] Note that this does not in any way inhibit the information content in $\overline{\mathbf{y}}$ because we restrict $\widetilde{\mathbf{C}} = [\gamma, \boldsymbol{\Gamma}]$ rather than $\mathbf{C} = [\gamma + \boldsymbol{\Gamma}\boldsymbol{\beta}, \boldsymbol{\Gamma}]$, so that given that in all our experiments $\boldsymbol{\beta} \neq \mathbf{0}$ the information content in $\overline{\mathbf{X}}$ also feeds into $\overline{\mathbf{y}}$ and $\mathbf{C}$ will tend to be more informative than $\boldsymbol{\Gamma}$, as is likely the case in practice.

in $\mathbf{Z}_i$ are perfectly correlated within individuals, and the covariance ($\mathbf{\Sigma}_\eta$ in Theorem 1) scales up one-to-one with $\sigma_\eta^2 = (1, 5)$.

Our experiments can be summarized as follows. We take $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (10, 1, 1, 1, 0)$ as our baseline scenario. It considers the homogeneous slope setting with more CA than factors and bias components at a standard level: CA are reasonably informative and none of the variances and covariances are excessive (relative to the others). We use this to assess the properties of CCEP and the bootstrap in a 'regular' or relatively forgiving setting, and subsequently perform stress tests by boosting one of the other DGP components to its more extreme 5 setting (while the other components stay on their standard setting). In each case we generate datasets where N and T take the values $(25, 50, 100, 200, 500, 1000)$, such that we cover both micro (where N is larger than T) and macro (T similar or larger than N) panels. We generate 2000 datasets for each combination of N and T and report bias, root mean squared error (*rmse*) and empirical size for tests performed at the $\alpha = 0.05$ level. In each Monte Carlo iteration the bootstrap-adjusted CCE estimate and confidence intervals are calculated from $B = 2000$ bootstrap samples generated with the cross-section resampling algorithm. We report results both with and without exclusion of $\bar{\mathbf{y}}$ from the matrix of CA, with the former denoted as CCEP and CCEMG and the latter by a x subscript as in CCEP_x and CCEMG_x. Their respective pairs bootstrap corrections are denoted 'pairs' and 'pairs$_x$'. The CCE estimators with an x subscript therefore employ $k = 3 > m$ cross-section averages, while the CCE estimators without this subscript employ $1 + k = 4 > m$. Note that we could in principle equalize the number of CA for the CCE and CCE_x approaches (i.e. estimate both versions with 3 CA) but choose in stead the current setting to correspond to empirical practice where one typically employs all the available CA in the estimation, and excluding $\bar{\mathbf{y}}$ as we propose therefore entails a loss of information, and one less CA than one would typically use. The Monte Carlo results represent this fact and are therefore more relevant for practice.

The next section discusses results with homogeneous slopes in the baseline setting, and all the corresponding stress tests are discussed in section 4.2.1. Section 4.3 presents results for heterogeneous slopes.

4.2 Results: Homogeneous slopes

Consider first the baseline scenario where the CA have reasonable asymptotic information content and none of the respective bias components are excessive. Table 1 focuses on the estimation results and presents (100×) the bias and root mean squared error (*rmse*) of the CCEP and CCEP$_x$ estimators as well as their respective bootstrap corrections 'pairs' and 'pairs$_x$'. Note that in this and subsequent tables we have $\tau = 1$ along the diagonal, whereas respectively $\tau > 1$ above the diagonal and $\tau < 1$ below the diagonal. Clearly, the results confirm the theory of section 3.1 as the CCEP estimators are generally biased when T is large compared to N. This holds true for both CCEP and its variant CCEP$_x$ which excludes the CA of the dependent variable. As alluded to below Corollary 1, the information loss incurred by excluding $\overline{\mathbf{y}}$ leads to larger biases for CCEP$_x$ compared to CCEP, but the *rmse* at the $N = T = 1000$ intersection confirms that this exclusion does not lead to efficiency losses in large samples.

Table 1: Estimation results: Baseline setting, fixed slopes

		bias × 100						*rmse* × 100					
	(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
CCEP	25	3.05	2.67	2.74	2.81	2.70	2.75	6.35	5.09	4.15	3.60	3.09	2.99
	50	1.56	1.50	1.42	1.39	1.37	1.41	4.49	3.28	2.60	2.09	1.69	1.58
	100	0.87	0.72	0.77	0.75	0.75	0.78	2.93	2.19	1.76	1.30	1.05	0.92
	200	0.52	0.38	0.36	0.37	0.38	0.37	2.06	1.52	1.14	0.84	0.61	0.51
	500	0.14	0.11	0.18	0.15	0.15	0.16	1.25	0.94	0.69	0.50	0.34	0.26
	1000	0.05	0.06	0.05	0.07	0.07	0.07	0.88	0.66	0.48	0.34	0.23	0.17
CCEP$_x$	25	3.17	2.73	2.80	2.88	2.80	2.85	6.61	5.27	4.22	3.69	3.21	3.09
	50	1.56	1.53	1.44	1.43	1.41	1.44	4.58	3.36	2.63	2.13	1.73	1.61
	100	0.85	0.72	0.78	0.76	0.76	0.79	2.99	2.17	1.76	1.32	1.05	0.93
	200	0.52	0.39	0.37	0.37	0.38	0.38	2.09	1.50	1.15	0.84	0.61	0.51
	500	0.12	0.12	0.17	0.15	0.16	0.16	1.25	0.94	0.69	0.50	0.34	0.26
	1000	0.07	0.07	0.05	0.07	0.07	0.07	0.88	0.66	0.48	0.34	0.23	0.17
pairs	25	1.22	0.85	0.87	0.95	0.76	0.84	7.06	5.17	3.67	2.66	1.75	1.42
	50	0.33	0.32	0.21	0.17	0.15	0.19	5.11	3.33	2.41	1.68	1.05	0.75
	100	0.24	0.04	0.08	0.07	0.07	0.10	3.33	2.36	1.72	1.13	0.75	0.51
	200	0.20	0.01	-0.01	0.00	0.02	0.01	2.36	1.64	1.15	0.79	0.49	0.36
	500	-0.01	-0.05	0.04	-0.01	0.00	0.01	1.44	1.04	0.71	0.50	0.31	0.21
	1000	-0.04	-0.02	-0.02	-0.01	0.00	-0.01	1.02	0.72	0.51	0.35	0.22	0.15
pairs$_x$	25	1.08	0.64	0.65	0.74	0.59	0.64	6.58	4.89	3.38	2.45	1.63	1.24
	50	0.28	0.27	0.15	0.15	0.10	0.14	4.80	3.19	2.28	1.61	1.01	0.72
	100	0.19	0.03	0.08	0.06	0.06	0.09	3.14	2.18	1.63	1.10	0.73	0.50
	200	0.20	0.02	0.01	0.01	0.01	0.01	2.24	1.53	1.12	0.77	0.49	0.35
	500	-0.03	-0.04	0.03	-0.01	0.01	0.01	1.35	0.98	0.69	0.49	0.30	0.21
	1000	-0.01	0.00	-0.02	-0.01	0.00	0.00	0.95	0.69	0.49	0.34	0.22	0.15

Notes: The baseline DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (10, 1, 1, 1, 0)$, with $m = 2$ factors and $k = 3$ regressors. CCEP and CCEP$_x$ denote respectively the CCEP estimator with and without $\overline{\mathbf{y}}$ included in the matrix of CA. 'Pairs' and 'pairs$_x$' correspond to their respective bootstrap-corrected estimates obtained from 2000 bootstrap replications with the pairs (cross-section) resampling algorithm.

The table also reveals that the respective bootstrap corrected estimators 'pairs' and 'pairs$_x$' lead to substantial bias reductions. The corrections are clearly effective even in the smallest samples but appear to gain particular traction when $N > 50$ and $T > 25$, in which case the remaining distortions can be more than 10 times smaller than those for the uncorrected CCEP estimators. Both corrected estimators are essentially unbiased for the mentioned combinations of T and N. The bootstrap corrections do lead to increased variance in smaller samples, but the asymptotic variance is unaffected and it is clear that this increase is generally compensated for by the bias reductions in a mean square error sense. Theorem 2 and Corollary 2 also predicted that the bootstrap correction for CCEP_x in this $m < 1 + k$ setting would be more effective than that of CCEP, as the bootstrap distribution of the latter generates an additional distortion. Indeed, even in this relatively low bias setting it is clear from Table 1 that although the bias for CCEP_x is larger than that of CCEP, the pairs_x correction is more effective for removing bias than the pairs correction, for which more bias remains across the board. This experiment also illustrates that the additional bias $\mathbf{d}^+$ generated by the bootstrap when applied to CCEP, which we recall makes it technically inconsistent in $m < 1 + k$ settings, is generally quite small and potentially even negligible at least in the baseline setting.

Table 2: Empirical size: Baseline setting, fixed slopes

CCEP

(N,T)	*t*-test 25	50	100	200	500	1000	*basic* 25	50	100	200	500	1000
25	0.10	0.13	0.18	0.30	0.47	0.69	0.07	0.07	0.05	0.05	0.03	0.01
50	0.08	0.10	0.11	0.17	0.33	0.50	0.09	0.07	0.05	0.04	0.03	0.01
100	0.07	0.07	0.10	0.11	0.24	0.37	0.06	0.06	0.07	0.05	0.05	0.03
200	0.07	0.05	0.07	0.09	0.14	0.19	0.07	0.06	0.06	0.06	0.05	0.05
500	0.06	0.06	0.05	0.07	0.08	0.10	0.07	0.07	0.05	0.06	0.06	0.04
1000	0.05	0.05	0.06	0.06	0.07	0.08	0.07	0.06	0.06	0.05	0.05	0.05

(N,T)	Bootstrap$-t$ 25	50	100	200	500	1000	Bootstrap$-t_c$ 25	50	100	200	500	1000
25	0.11	0.09	0.08	0.08	0.09	0.12	0.11	0.10	0.08	0.08	0.07	0.06
50	0.12	0.09	0.08	0.07	0.08	0.07	0.12	0.09	0.07	0.07	0.07	0.04
100	0.10	0.07	0.09	0.06	0.07	0.06	0.10	0.07	0.09	0.07	0.07	0.05
200	0.09	0.08	0.07	0.07	0.06	0.07	0.09	0.08	0.07	0.07	0.06	0.07
500	0.08	0.08	0.06	0.06	0.06	0.05	0.08	0.08	0.06	0.06	0.06	0.05
1000	0.07	0.06	0.07	0.06	0.06	0.06	0.07	0.06	0.06	0.06	0.06	0.06

CCEP_x

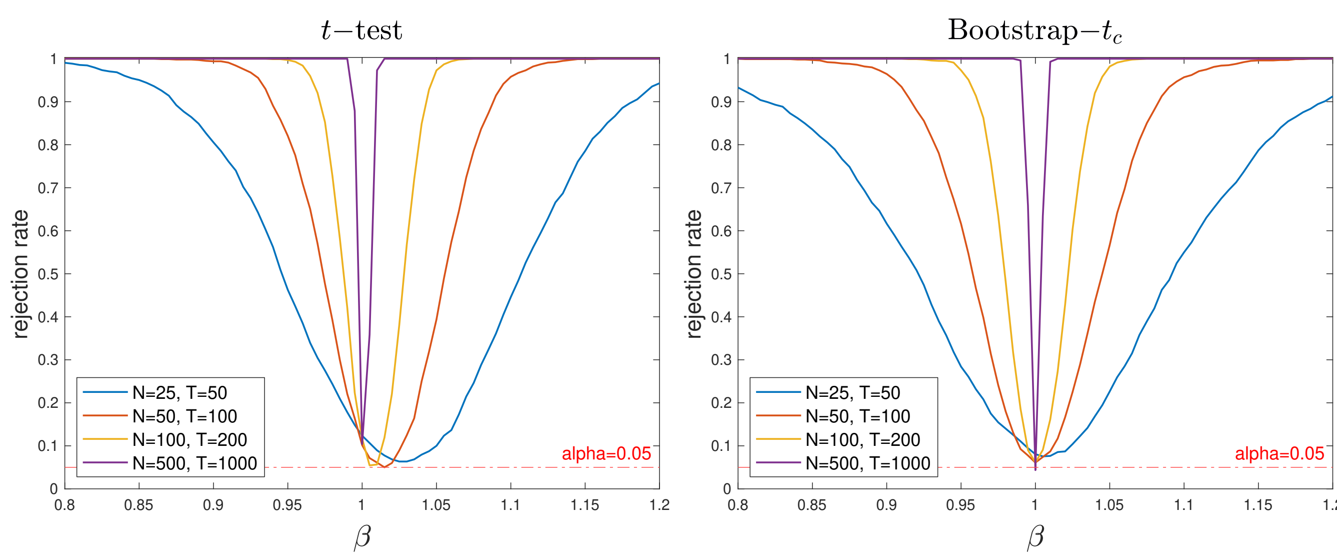
(N,T)	*t*-test 25	50	100	200	500	1000	*basic* 25	50	100	200	500	1000
25	0.09	0.12	0.17	0.28	0.47	0.68	0.04	0.05	0.03	0.03	0.01	0.00
50	0.08	0.09	0.10	0.17	0.32	0.51	0.07	0.04	0.03	0.03	0.01	0.01
100	0.07	0.06	0.10	0.12	0.23	0.37	0.05	0.05	0.05	0.04	0.04	0.03
200	0.07	0.05	0.08	0.08	0.13	0.19	0.06	0.05	0.05	0.05	0.05	0.04
500	0.05	0.06	0.05	0.06	0.08	0.10	0.05	0.06	0.05	0.05	0.05	0.04
1000	0.05	0.05	0.05	0.06	0.07	0.08	0.06	0.05	0.05	0.05	0.06	0.06

(N,T)	Bootstrap$-t$ 25	50	100	200	500	1000	Bootstrap$-t_c$ 25	50	100	200	500	1000
25	0.10	0.08	0.07	0.07	0.07	0.09	0.09	0.08	0.07	0.07	0.06	0.03
50	0.12	0.08	0.06	0.07	0.06	0.06	0.12	0.08	0.06	0.06	0.05	0.03
100	0.09	0.06	0.08	0.06	0.06	0.06	0.09	0.06	0.08	0.06	0.06	0.05
200	0.08	0.07	0.06	0.06	0.06	0.06	0.08	0.07	0.06	0.07	0.06	0.06
500	0.06	0.07	0.06	0.06	0.06	0.04	0.06	0.07	0.06	0.06	0.06	0.04
1000	0.06	0.05	0.06	0.06	0.07	0.06	0.06	0.05	0.06	0.06	0.06	0.06

Notes: The baseline DGP is $(d_u, \beta, \sigma^2, \sigma^2_\eta, \sigma^2_v) = (10, 1, 1, 1, 0)$, with $m = 2$ factors and $k = 3$ regressors. CCEP and CCEP_x denote respectively the CCEP estimator with and without $\bar{\mathbf{y}}$ included in the matrix of CA. '*t*-test' resports the empirical size for a *t*-test at the $\alpha = 0.05$ significance level. '*basic*' reports empirical size for tests based on the basic ('empirical percentile') bootstrap interval, and bootstrap$-t$ and bootstrap$-t_c$ are respectively empirical size for the plain and corrected bootstrap$-t$ interval. All bootstrap tests are based on $B = 2000$ replications with the pairs (cross-section) resampling algorithm.

The biases in Table 1 are not terribly large but nevertheless warrant serious concern for hypothesis testing. This is because the deviations $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ are scaled up by a factor $\sqrt{NT}$ in the numerator of the *t*-statistic under the null (recall e.g.(3.8)), so that even the modest numbers in the table above can still lead to large location shifts away from zero in the distribution of the *t*-statistic, and hence lead to size distortions. The actual impact on testing can be discerned from Table 2, where we report the empirical size of the conventional (asymptotic) *t*-test and the 'basic', bootstrap$-t$ and corrected bootstrap$-t_c$ confidence intervals, each based on 2000 bootstrap replications. As predicted by theory, the asymptotic t-test performs poorly as it only attains the nominal 5% size when N dominates the time series length, or in other words in the lower left quadrants of the tables. This is true for both the CCEP and CCEP$_x$ estimator. Otherwise, size distortions become quite severe the larger is T relative to N. The objective of the bootstrap was to alleviate the $T/N \to 0$ restriction needed for inference with the t-test, which the table confirms to be quite successful. The basic percentile interval for instance achieves large improvements and has an empirical size that is very close to the nominal level for nearly all combinations of N and T. Only when T is excessive compared to N, for instance when $N = 25$, $T = 1000$, we find that the test is undersized. This is in line with Corollary 2 and the overall requirement that $\tau < \infty$, or in other words that T should not dominate, as this leads to an over-accumulation of error. The test is correctly sized for any other combination of N and T, which makes it is clear that the bootstrap has significantly relaxed the $T/N \to 0$ restriction to allow unbiased inferences even when $T/N \not\to 0$. The corrected bootstrap-t_c interval too achieves tremendous improvements over the classical t-test and shows that re-pivoting of the roots has made the test more robust than the basic percentile interval to large T settings, as witnessed by the relatively small size distortions even in the upper-right quadrant where T dominates. The main cost is that the bootstrap$-t_c$ is somewhat more sensitive than the basic interval in very small samples, i.e. $N \leq 50, T \leq 50$, in which case there are some remaining size distortions.

Figure 1: **Power functions: $T/N = 2$, baseline setting**



Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (10, 1, 1, 5, 0)$ with $m = 2$ and $k = 3$. The vertical axis presents rejection rates of the respective tests for the hypothesized value of β on the horizontal axis. Both the left and right panel are based on the CCEP$_x$ estimator (with $\bar{\mathbf{y}}$ excluded).

To assess power, consider the rejection functions for combinations of N and T such that $T/N = 2$ plotted in Fig.1. The plot displays rejection rates for the t-test and bootstrap$-t_c$ based on the CCEP$_x$ estimator for the hypothesized values of β on the horizontal axis. The correct null hypothesis is $\beta = 1$. It is clear that the rejection curves for the $t-$test (left panel) are generally not correctly centered around the true β, but are in stead shifted to the right of it as a result of the bias. Only in the largest sample size considered (the purple line) is the curve is correctly centered at $\beta = 1$, but nominal size $\alpha = 0.05$ (dotted red line) is then still not attained. The bootstrap$-t_c$ interval is shown in the right panel and can be seen to have resolved these rightward shifts and size distortions for most of the considered combinations of N and T, without incurring significant reductions in power. That is, the main effect appears to have been a re-centering of rejection curves around the true parameter, as required.

4.2.1 Stress-tests

The results so far have shown large improvements of the bootstrap over the asymptotic t-test in our baseline setting. Next, we explore a number of more extreme scenarios to gauge their impact on the CCEP estimator and to stress-test the inference procedures. We find that the most challenging scenario is when the variance of the loadings is relatively large, i.e. when σ_η^2 is boosted to 5. This scales up the $\boldsymbol{\Sigma}_\eta$ matrix in Theorem 1 and Corollary 1 with a factor of 5 and it can be seen from the estimation results in Table 3 that the boosted noise level leads to an increase in bias by a factor of at least 3 compared to Table 1.

Table 3: Estimation results: Boosted loading variance ($\sigma_\eta^2 = 5$), fixed slopes

	(N,T)	*bias* × 100						*rmse* × 100					
		25	50	100	200	500	1000	25	50	100	200	500	1000
CCEP	25	10.04	9.66	9.89	10.06	10.00	10.07	12.04	11.03	10.68	10.58	10.32	10.34
	50	5.97	6.02	6.03	6.01	6.04	6.06	7.67	6.94	6.61	6.36	6.24	6.21
	100	3.44	3.35	3.39	3.41	3.39	3.43	4.72	4.13	3.85	3.63	3.52	3.51
	200	1.93	1.79	1.78	1.78	1.80	1.79	2.89	2.40	2.13	1.97	1.88	1.84
	500	0.72	0.70	0.78	0.74	0.75	0.75	1.47	1.20	1.04	0.89	0.81	0.78
	1000	0.35	0.36	0.36	0.37	0.37	0.37	0.96	0.76	0.61	0.50	0.43	0.40
CCEP_x	25	11.42	11.13	11.32	11.48	11.50	11.57	13.62	12.73	12.33	12.25	12.12	12.12
	50	6.48	6.59	6.53	6.55	6.60	6.60	8.19	7.57	7.19	6.96	6.86	6.81
	100	3.54	3.46	3.53	3.54	3.53	3.58	4.84	4.23	3.99	3.78	3.67	3.66
	200	1.96	1.82	1.83	1.83	1.83	1.82	2.95	2.41	2.19	2.01	1.92	1.87
	500	0.71	0.72	0.78	0.74	0.76	0.76	1.47	1.21	1.04	0.90	0.82	0.79
	1000	0.37	0.37	0.36	0.37	0.37	0.37	0.97	0.77	0.61	0.50	0.43	0.40
pairs	25	6.24	5.82	5.99	6.19	5.98	6.07	10.79	8.91	7.78	7.36	6.72	6.67
	50	2.39	2.48	2.44	2.39	2.38	2.41	6.61	4.83	4.00	3.33	2.90	2.75
	100	0.96	0.80	0.80	0.83	0.78	0.83	4.08	2.95	2.17	1.57	1.21	1.06
	200	0.44	0.24	0.21	0.21	0.23	0.21	2.61	1.83	1.29	0.90	0.59	0.46
	500	0.03	-0.02	0.08	0.03	0.04	0.04	1.51	1.09	0.74	0.52	0.32	0.22
	1000	-0.02	-0.01	-0.01	0.00	0.00	0.00	1.05	0.73	0.52	0.35	0.22	0.16
pairs_x	25	6.24	6.01	6.12	6.30	6.14	6.19	11.09	9.49	8.40	8.07	7.55	7.48
	50	1.93	2.06	1.94	1.93	1.89	1.91	6.20	4.66	3.70	2.97	2.49	2.32
	100	0.68	0.53	0.55	0.58	0.52	0.58	3.80	2.69	2.01	1.43	1.02	0.84
	200	0.37	0.16	0.17	0.15	0.15	0.14	2.50	1.69	1.25	0.86	0.56	0.42
	500	-0.01	-0.01	0.06	0.02	0.03	0.03	1.41	1.04	0.72	0.51	0.32	0.22
	1000	0.00	0.00	-0.01	0.00	0.00	0.00	0.98	0.71	0.50	0.34	0.22	0.16

Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (10, 1, 1, 5, 0)$, with $m = 2$ factors and $k = 3$ regressors. CCEP and CCEP_x denote respectively the CCEP estimator with and without $\bar{\mathbf{y}}$ included in the matrix of CA. 'Pairs' and 'pairs_x' correspond to their respective bootstrap-corrected estimates obtained from 2000 bootstrap replications with the pairs (cross-section) resampling algorithm.

Interestingly, many conclusions from the baseline scenario extend to this more extreme setting, save that they are now more explicit because there is more bias to correct for. That is, while biases are again larger for CCEP_x than for CCEP, less bias remains for the bootstrap correction of the former (pairs_x) than for the latter (pairs). This is exactly as predicted by our theory. Fortunately, the remaining bias is for both corrections but a fraction of that for the original estimator, in particular when $N > 25$. The increased noise levels in other words made the situation more challenging for the bootstrap, but the algorithms remain highly effective given sufficient N. This also leads to substantial improvements for nearly all combinations of N and T in a mean square error sense.

Table 4: Empirical size: Boosted loading variance ($\sigma_\eta^2 = 5$), fixed slopes

	CCEP											
	t-test						*basic*					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.44	0.56	0.78	0.94	0.99	1.00	0.19	0.20	0.24	0.36	0.43	0.50
50	0.34	0.52	0.72	0.89	0.99	1.00	0.09	0.09	0.09	0.11	0.11	0.12
100	0.26	0.35	0.58	0.82	0.98	1.00	0.06	0.05	0.05	0.03	0.02	0.02
200	0.19	0.23	0.39	0.64	0.94	0.99	0.05	0.04	0.03	0.02	0.01	0.01
500	0.10	0.14	0.23	0.34	0.69	0.93	0.06	0.05	0.04	0.04	0.03	0.02
1000	0.07	0.08	0.12	0.20	0.41	0.69	0.05	0.04	0.05	0.04	0.05	0.04
	Bootstrap$-t$						Bootstrap$-t_c$					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.23	0.24	0.29	0.43	0.57	0.67	0.22	0.24	0.29	0.43	0.55	0.64
50	0.13	0.13	0.17	0.25	0.42	0.60	0.13	0.12	0.15	0.20	0.27	0.33
100	0.09	0.07	0.09	0.10	0.22	0.46	0.09	0.07	0.09	0.06	0.09	0.11
200	0.07	0.06	0.05	0.05	0.08	0.16	0.07	0.06	0.05	0.04	0.04	0.04
500	0.07	0.06	0.05	0.05	0.04	0.04	0.08	0.06	0.05	0.05	0.04	0.03
1000	0.06	0.05	0.06	0.04	0.05	0.05	0.06	0.05	0.06	0.04	0.05	0.04

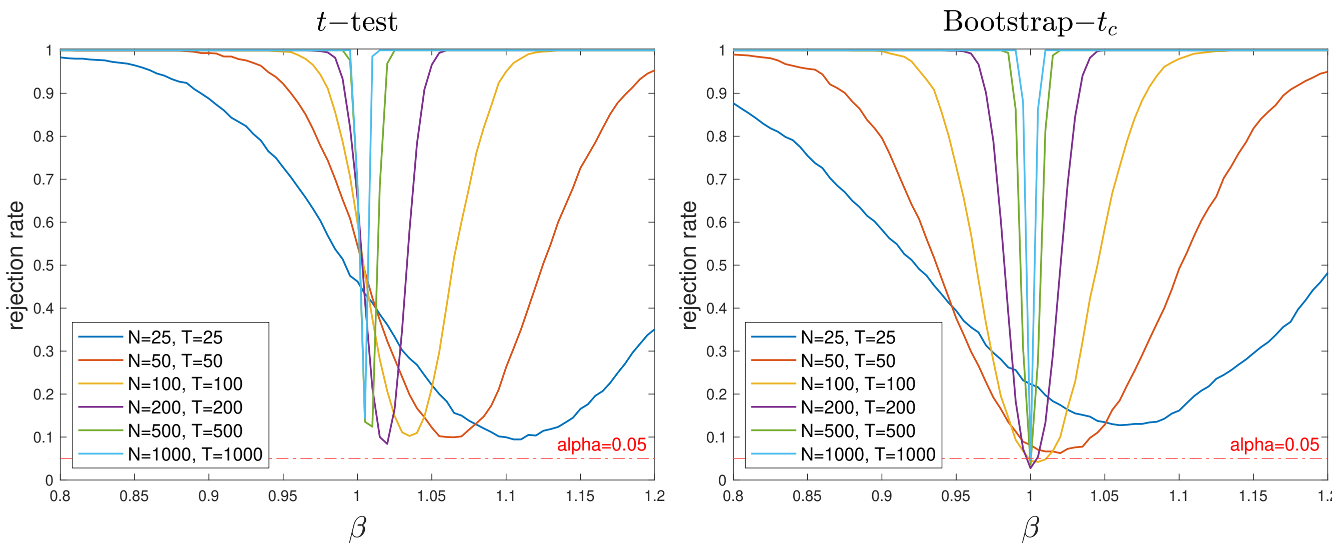
	CCEP_x											
	t-test						*basic*					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.46	0.61	0.82	0.94	0.99	1.00	0.15	0.16	0.19	0.22	0.19	0.20
50	0.37	0.55	0.74	0.91	0.99	1.00	0.06	0.05	0.03	0.02	0.01	0.00
100	0.26	0.38	0.60	0.84	0.98	1.00	0.04	0.03	0.02	0.01	0.00	0.00
200	0.19	0.24	0.41	0.66	0.94	0.99	0.04	0.03	0.03	0.02	0.01	0.01
500	0.09	0.14	0.21	0.35	0.69	0.93	0.03	0.05	0.04	0.04	0.03	0.02
1000	0.07	0.09	0.12	0.19	0.41	0.69	0.05	0.04	0.04	0.04	0.04	0.04
	Bootstrap$-t$						Bootstrap$-t_c$					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.23	0.24	0.30	0.38	0.42	0.48	0.22	0.23	0.27	0.33	0.35	0.37
50	0.10	0.09	0.09	0.12	0.20	0.30	0.10	0.08	0.06	0.07	0.06	0.06
100	0.07	0.04	0.06	0.05	0.11	0.24	0.07	0.05	0.05	0.03	0.03	0.02
200	0.06	0.04	0.06	0.03	0.06	0.10	0.06	0.05	0.05	0.03	0.02	0.02
500	0.04	0.06	0.04	0.05	0.04	0.03	0.04	0.06	0.05	0.05	0.04	0.02
1000	0.05	0.04	0.05	0.05	0.04	0.05	0.05	0.04	0.05	0.05	0.04	0.05

Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (10, 1, 1, 5, 0)$, with $m = 2$ factors and $k = 3$ regressors. CCEP and CCEP_x denote respectively the CCEP estimator with and without $\bar{\mathbf{y}}$ included in the matrix of CA. '*t*-test' resports the empirical size for a *t*-test at the $\alpha = 0.05$ significance level. '*basic*' reports empirical size for tests based on the basic ('empirical percentile') bootstrap interval, and bootstrap$-t$ and bootstrap$-t_c$ are respectively empirical size for the plain and corrected bootstrap$-t$ interval. All bootstrap tests are based on $B = 2000$ replications with the pairs (cross-section) resampling algorithm.

While bias itself is now also a more direct concern for the point estimates, the main worry remains its impact on hypothesis testing due to the implied scaling up by $\sqrt{NT}$. Indeed, Table 4 discloses very large size distortions for the asymptotic *t*-test on all combinations of N and T, much more so than in Table 2. Even the setting where the cross-section dimension dominates the time series length $N = 1000$, $T = 25$, does not seem to suffice to obtain a correctly sized *t*-test, and it suggests that N should preferably be even larger in this experiment. The benefits of the bootstrap tests are now even more apparent and they even appear to be quintessential for correct inferences in this high noise situation. The basic and bootstrap-t tests perform orders of magnitude better than the t-test on all sample sizes, again displaying the relaxation of the $T/N \to 0$ restriction, but the most robust approach is clearly the corrected bootstrap-t_c interval when applied to the CCEP$_x$ estimator (lower panel). The fact that the bootstrap is then generally first order consistent and that the confidence interval (and test) is here based on an asymptotically pivotal statistic results in close to nominal test sizes on all $N > 25$ settings. This is a tremendous improvement over not only the classical t-test, but also compared to the intervals based on non-pivotal roots (basic and bootstrap$-t$). As expected, re-pivoting appears to have translated to faster convergence and increased robustness to settings where T is relatively large compared to N. This makes the bootstrap-t_c attractive on all settings, but its comparative advantages are biggest when T is not small. Note that while the theoretical inconsistency of the bootstrap when applied to CCEP (upper panel) does not seem to have too much of a negative impact on the large sample behavior of the bootstrap tests, the CCEP$_x$ variants (lower panel) perform markedly better in small samples thanks to their consistency. This confirms our theory and strengthens our preference for the CCEP$_x$ estimator in practice.

While it is good to see as in Table 4 that the test size is even in this challenging scenario well controlled at the nominal $\alpha = 0.05$ level by the bootstrap$-t_c$, this is ideally not at the expense of power. This is visualized in Figure 2, which plots the power functions of both the t-test and bootstrap$-t_c$ test based on the CCEP$_x$ estimator, now for combinations of N and T such that $T/N = 1$. Note that this is more forgiving for the $t-$test than $T/N = 2$.

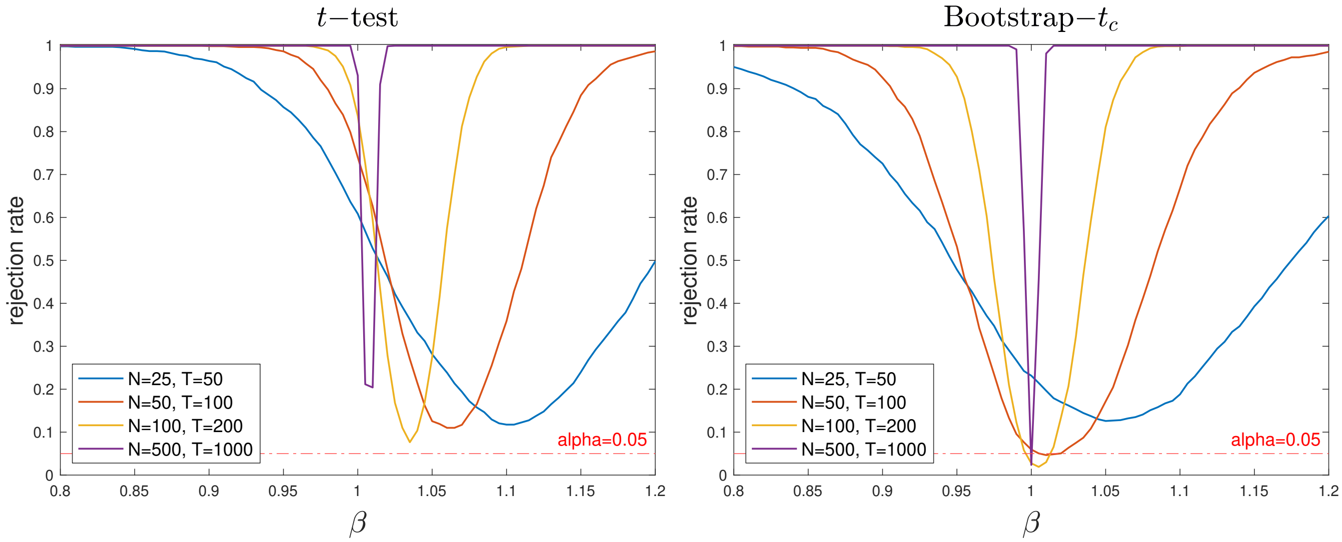
Figure 2: **Power functions:** $T/N = 1$**, boosted loading variance (**$\sigma_\eta^2 = 5$**)**



Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (10, 1, 1, 5, 0)$ with $m = 2$ and $k = 3$. The vertical axis presents rejection rates of the respective tests for the hypothesized value of β on the horizontal axis. Both the left and right panel are based on the CCEP$_x$ estimator (with $\overline{\mathbf{y}}$ excluded).

It is again very apparent that the power function of the t-test is highly distorted for all $T/N = 1$ combinations, with the 'dip' for most of the curves now located much farther to the right of the true $\beta = 1$ null hypothesis. The obvious conclusion is that the t-test cannot be trusted for inferences on β. In contrast, the power function for the bootstrap-t_c test in the right panel has a more regular form with the majority of the curves still enveloping the true $\beta = 1$ value, where the nominal $\alpha = 0.05$ size is also attained. Yet, the situation is clearly not perfect. The plot shows that with the increased noise level also comes a requirement for larger N to adequately correct the distortions. The smallest $N = 25$ setting (dark blue), for instance, is clearly not sufficient to deal with the boosted loading variance. This is not unexpected given the relatively extreme scenario and the fact that cross-section resampling algorithm requires at its core that $N \to \infty$. Fortunately, increasing the number of cross-sections to $N = 50$ appears to resolve the remaining rightward shift and the $N > 25$ curves indeed quickly tighten around $\beta = 1$, as required. As the $T/N = 2$ setting is also fairly common in macroeconomics we display it in Figure 3. Clearly, the distortions are now even larger for the $t-$test compared to the $T/N = 1$ setting in Fig.2 whereas the power curves for the bootstrap in the right panel are much less affected by the increase of T/N, which solidifies that $T/N \to 0$ is no longer required for inference. This is a major advantage in practice.

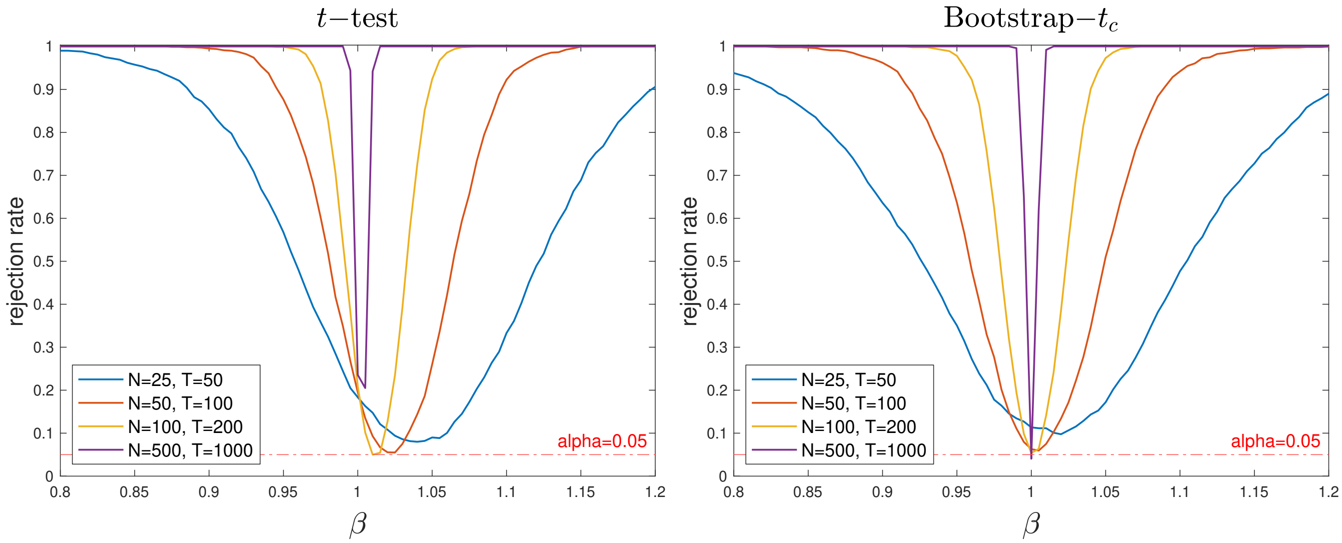
Figure 3: **Power functions:** $T/N = 2$**, boosted loading variance (**$\sigma_\eta^2 = 5$**)**



Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (10, 1, 1, 5, 0)$ with $m = 2$ and $k = 3$. The vertical axis presents rejection rates of the respective tests for the hypothesized value of β on the horizontal axis. Both the left and right panel are based on the CCEP$_x$ estimator (with $\bar{\mathbf{y}}$ excluded).

In contrast to the loading variance, boosting σ^2 or β has a relatively small effect on performance. Conclusions are largely the same as in the baseline setting discussed above so we will not report the results here and in stead provide the tables in Supplement B. A more interesting scenario is when we reduce the asymptotic information content of the cross-section averages (i.e. we reduce the bounds of the determinant of the loading matrix d_u from 10 to 5). The impact on the estimation results is shown in Table 5 and it is largely as predicted below Theorem 1: even though the noise levels (loading and error variances) are identical to those in the baseline scenario, the bias in this scenario is larger because the error components are scaled up as a result of the less informative cross-section averages. The benefit of the bootstrap correction is that this information content does not need to be known or estimated, and the bootstrap corrections clearly remain highly effective at removing bias. Naturally, the increased bias compared to the baseline setting also leads to larger size distortions for the t-test in Table 6 compared to Table 2. Fortunately, it can also be seen that with the exception of the very small $N = 25$ setting, the performance of the bootstrap procedures is largely unaffected by the information drop, with the CCEP$_x$ bootstrap-t_c test once again displaying the most robustness to even dominant T. Again, performance for the bootstrap applied to CCEP$_x$ (lower panel) is better than when applied to CCEP (upper panel).

Figure 4: **Power functions:** $T/N = 2$**, reduced information (**$d_u = 5$**)**



Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (5, 1, 1, 1, 0)$ with $m = 2$ and $k = 3$. The vertical axis presents rejection rates of the respective tests for the hypothesized value of β on the horizontal axis. Both the left and right panel are based on the CCEP$_x$ estimator (with $\bar{\mathbf{y}}$ excluded).

The power function for $T/N = 2$ in Figure 4 also allows comparison to earlier situations. This scenario is clearly somewhat less disruptive for the $t-$test compared to the boosted loading variance scenario, but a correct size is still not attained, not even in large samples, and the majority of the curves are still considerably shifted to the right of the true parameter. The bootstrap curves in the right panel are much better behaved, with the relatively minor distortions limited to the smallest $N = 25$ case.

In conclusion, we have confirmed with our fixed slope experiments that the properties of the $t-$test are indeed highly sensitive to the $T/N \to 0$ restriction. The higher the noise in the dataset (or CA), the larger N needs to be compared to T. In contrast, the results confirm that the bootstrap does not require $T/N \to 0$, as evidenced by the near nominal size and adequate power on nearly all combinations of N and T. The bootstrap is in addition substantially more robust to the challenging settings we have considered. This is particularly true for the bootstrap-t_c procedure applied to CCEP$_x$, provided that N is not too small.

Table 5: Estimation results: reduced asymptotic information content ($d_u = 5$), fixed slopes

		bias × 100						*rmse* × 100					
	(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
CCEP	25	4.35	3.81	3.95	4.05	4.01	4.04	7.26	5.92	5.15	4.70	4.34	4.26
	50	2.29	2.26	2.17	2.17	2.17	2.20	4.92	3.73	3.11	2.70	2.41	2.34
	100	1.35	1.17	1.21	1.18	1.20	1.22	3.15	2.36	2.02	1.60	1.41	1.32
	200	0.72	0.61	0.60	0.60	0.62	0.61	2.14	1.62	1.24	0.97	0.79	0.70
	500	0.24	0.22	0.27	0.25	0.25	0.26	1.28	0.97	0.72	0.54	0.39	0.33
	1000	0.10	0.11	0.10	0.12	0.12	0.12	0.88	0.67	0.49	0.35	0.25	0.19
CCEP_x	25	4.62	4.03	4.17	4.26	4.26	4.28	7.64	6.23	5.33	4.91	4.60	4.50
	50	2.36	2.34	2.24	2.27	2.27	2.29	5.02	3.83	3.19	2.80	2.51	2.42
	100	1.34	1.18	1.24	1.21	1.22	1.25	3.24	2.36	2.04	1.63	1.43	1.35
	200	0.73	0.63	0.62	0.61	0.63	0.62	2.14	1.61	1.25	0.98	0.79	0.71
	500	0.22	0.23	0.27	0.25	0.26	0.26	1.27	0.97	0.72	0.54	0.40	0.33
	1000	0.12	0.12	0.10	0.12	0.12	0.12	0.89	0.68	0.48	0.35	0.25	0.20
pairs	25	2.32	1.68	1.78	1.87	1.70	1.76	7.73	5.68	4.22	3.24	2.43	2.19
	50	0.67	0.67	0.53	0.53	0.49	0.54	5.38	3.46	2.53	1.80	1.20	0.94
	100	0.43	0.17	0.20	0.17	0.17	0.20	3.41	2.32	1.77	1.16	0.78	0.55
	200	0.18	0.05	0.03	0.03	0.05	0.04	2.39	1.68	1.17	0.80	0.50	0.36
	500	0.00	-0.03	0.03	0.00	0.01	0.01	1.48	1.06	0.71	0.50	0.31	0.21
	1000	-0.03	-0.02	-0.03	0.00	-0.01	0.00	1.02	0.73	0.50	0.34	0.22	0.15
pairs_x	25	2.18	1.47	1.55	1.61	1.47	1.51	7.33	5.44	3.87	2.97	2.21	1.94
	50	0.62	0.57	0.43	0.46	0.41	0.44	5.03	3.31	2.39	1.72	1.13	0.86
	100	0.35	0.13	0.17	0.15	0.14	0.17	3.28	2.18	1.70	1.13	0.76	0.53
	200	0.17	0.05	0.05	0.04	0.04	0.03	2.21	1.58	1.12	0.78	0.50	0.36
	500	-0.04	-0.02	0.02	0.00	0.01	0.01	1.37	0.99	0.68	0.49	0.30	0.21
	1000	0.00	0.00	-0.03	-0.01	0.00	0.00	0.96	0.70	0.49	0.34	0.22	0.15

Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma^2_\eta, \sigma^2_v) = (5, 1, 1, 1, 0)$, with $m = 2$ factors and $k = 3$ regressors. CCEP and CCEP_x denote respectively the CCEP estimator with and without $\bar{\mathbf{y}}$ included in the matrix of CA. 'Pairs' and 'pairs_x' correspond to their respective bootstrap-corrected estimates obtained from 2000 bootstrap replications with the pairs (cross-section) resampling algorithm.

Table 6: Empirical size: reduced information content ($d_u = 5$), fixed slopes

	CCEP											
	t-test						*basic*					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.15	0.19	0.28	0.45	0.70	0.87	0.11	0.09	0.07	0.07	0.04	0.02
50	0.12	0.13	0.20	0.30	0.58	0.81	0.10	0.05	0.05	0.04	0.02	0.01
100	0.09	0.09	0.15	0.21	0.42	0.71	0.06	0.05	0.07	0.04	0.04	0.02
200	0.08	0.08	0.11	0.13	0.27	0.45	0.07	0.06	0.05	0.05	0.05	0.04
500	0.06	0.07	0.06	0.09	0.12	0.23	0.07	0.07	0.06	0.05	0.05	0.04
1000	0.04	0.05	0.06	0.07	0.09	0.13	0.06	0.06	0.06	0.05	0.06	0.05
	Bootstrap$-t$						Bootstrap$-t_c$					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.13	0.12	0.11	0.15	0.16	0.26	0.13	0.12	0.11	0.14	0.11	0.16
50	0.12	0.08	0.08	0.09	0.10	0.15	0.12	0.08	0.08	0.07	0.07	0.07
100	0.08	0.06	0.09	0.06	0.08	0.08	0.08	0.06	0.08	0.06	0.07	0.05
200	0.09	0.07	0.07	0.06	0.07	0.07	0.09	0.07	0.06	0.06	0.07	0.06
500	0.08	0.08	0.06	0.06	0.05	0.05	0.08	0.08	0.06	0.06	0.06	0.05
1000	0.07	0.07	0.06	0.05	0.06	0.06	0.07	0.07	0.06	0.05	0.06	0.06
	$CCEP_x$											
	t-test						*basic*					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.14	0.18	0.29	0.45	0.71	0.88	0.07	0.07	0.05	0.05	0.02	0.01
50	0.11	0.14	0.20	0.31	0.58	0.80	0.07	0.04	0.03	0.03	0.02	0.00
100	0.08	0.09	0.16	0.21	0.43	0.71	0.06	0.03	0.06	0.03	0.04	0.01
200	0.07	0.07	0.10	0.14	0.27	0.46	0.06	0.04	0.05	0.04	0.04	0.04
500	0.05	0.07	0.06	0.08	0.13	0.24	0.05	0.06	0.05	0.05	0.04	0.04
1000	0.04	0.06	0.06	0.07	0.09	0.14	0.05	0.05	0.05	0.05	0.05	0.05
	Bootstrap$-t$						Bootstrap$-t_c$					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.12	0.12	0.11	0.13	0.15	0.24	0.12	0.11	0.10	0.11	0.08	0.09
50	0.09	0.07	0.07	0.08	0.10	0.15	0.09	0.07	0.06	0.06	0.05	0.04
100	0.08	0.05	0.07	0.06	0.07	0.08	0.09	0.05	0.07	0.05	0.07	0.05
200	0.07	0.06	0.06	0.06	0.06	0.06	0.07	0.05	0.06	0.06	0.06	0.06
500	0.06	0.07	0.06	0.05	0.05	0.04	0.06	0.07	0.06	0.05	0.05	0.04
1000	0.05	0.06	0.06	0.05	0.06	0.06	0.05	0.06	0.06	0.05	0.06	0.06

Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (5, 1, 1, 1, 0)$, with $m = 2$ factors and $k = 3$ regressors. CCEP and $CCEP_x$ denote respectively the CCEP estimator with and without $\overline{\mathbf{y}}$ included in the matrix of CA. '*t*-test' resports the empirical size for a *t*-test at the $\alpha = 0.05$ significance level. '*basic*' reports empirical size for tests based on the basic ('empirical percentile') bootstrap interval, and bootstrap$-t$ and bootstrap$-t_c$ are respectively empirical size for the plain and corrected bootstrap$-t$ interval. All bootstrap tests are based on $B = 2000$ replications with the pairs (cross-section) resampling algorithm.

4.3 Results: Heterogeneous slopes

In this section we set $\sigma_v^2 = 5$ and present results for when the slope coefficients in the model are heterogeneous. We find that the pairs bootstrap also in this setting leads to finite sample bias reductions for both the CCEP and CCEMG estimator. These results are reported in Table A-3 of Supplement B. Regarding inference, we find that the results are qualitatively the same with or without the exclusion of $\overline{\mathbf{y}}$. Hence, to save space we only report results for the latter, given also our preference for it in the (fixed slope) theory section. The full results are presented in Tables B-3 and B-4 of Supplement B. Table 7 summarizes empirical size with the CCEP_x estimator in the top panel and the bottom panel contains results for the CCEMG_x estimator.

Table 7: Empirical size: Heterogeneous slopes ($\sigma_v^2 = 5$)

CCEP_x

	t-test						*basic*					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.10	0.08	0.09	0.11	0.11	0.11	0.19	0.18	0.16	0.18	0.19	0.19
50	0.08	0.09	0.08	0.08	0.09	0.09	0.13	0.14	0.12	0.13	0.13	0.14
100	0.07	0.06	0.08	0.09	0.08	0.08	0.10	0.10	0.12	0.11	0.12	0.12
200	0.06	0.07	0.08	0.07	0.06	0.06	0.08	0.10	0.10	0.09	0.07	0.08
500	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.07
1000	0.05	0.05	0.04	0.04	0.06	0.06	0.06	0.06	0.04	0.05	0.06	0.06

	Bootstrap$-t$						Bootstrap$-t_c$					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.18	0.13	0.11	0.11	0.10	0.09	0.16	0.12	0.10	0.11	0.09	0.09
50	0.12	0.12	0.10	0.09	0.09	0.08	0.12	0.11	0.09	0.09	0.08	0.07
100	0.10	0.08	0.09	0.08	0.07	0.07	0.10	0.08	0.09	0.08	0.07	0.06
200	0.09	0.08	0.09	0.07	0.07	0.06	0.09	0.08	0.09	0.07	0.07	0.06
500	0.07	0.06	0.05	0.06	0.06	0.06	0.07	0.06	0.05	0.06	0.06	0.06
1000	0.07	0.06	0.04	0.05	0.07	0.06	0.07	0.06	0.04	0.05	0.07	0.06

CCEMG_x

	t-test						*basic*					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.09	0.09	0.10	0.09	0.09	0.11	0.12	0.12	0.13	0.13	0.13	0.13
50	0.07	0.08	0.06	0.07	0.08	0.07	0.09	0.10	0.08	0.09	0.09	0.09
100	0.07	0.07	0.06	0.07	0.07	0.06	0.08	0.08	0.07	0.09	0.07	0.08
200	0.05	0.06	0.05	0.06	0.06	0.06	0.06	0.07	0.06	0.08	0.06	0.06
500	0.05	0.07	0.05	0.06	0.06	0.05	0.06	0.06	0.06	0.06	0.06	0.05
1000	0.04	0.04	0.05	0.05	0.05	0.04	0.04	0.05	0.05	0.05	0.06	0.05

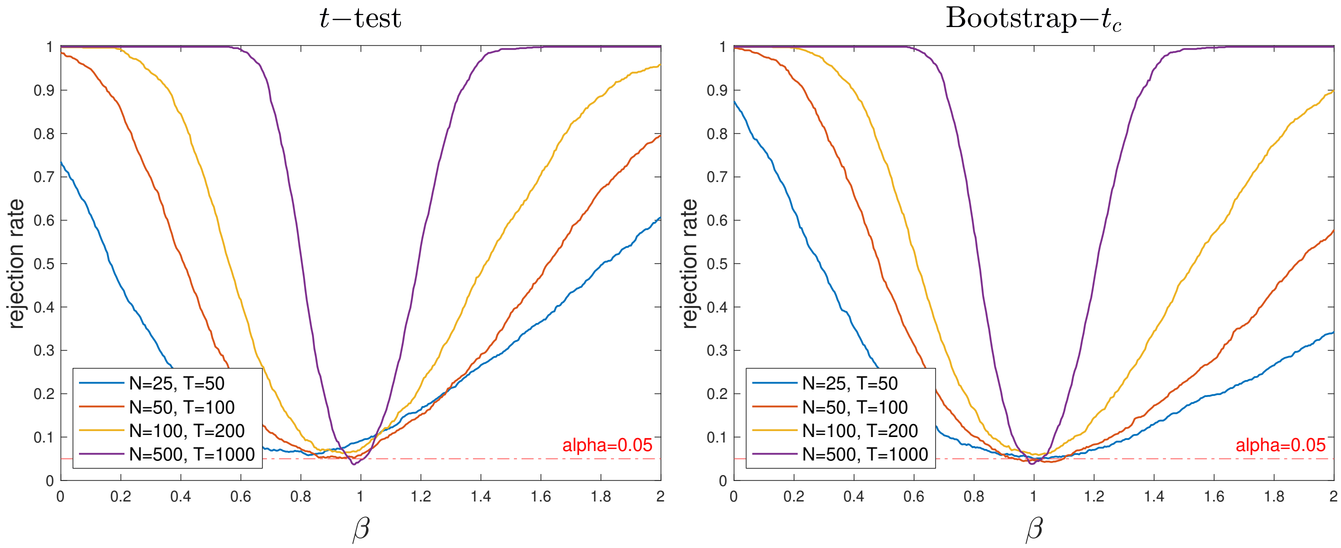
	Bootstrap$-t$						Bootstrap$-t_c$					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.06	0.05	0.06	0.07	0.05	0.05	0.06	0.05	0.06	0.06	0.05	0.05
50	0.06	0.06	0.05	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.05	0.05
100	0.05	0.06	0.05	0.06	0.05	0.05	0.05	0.06	0.04	0.06	0.05	0.05
200	0.05	0.06	0.04	0.05	0.05	0.05	0.05	0.06	0.04	0.05	0.05	0.04
500	0.05	0.06	0.05	0.05	0.05	0.04	0.05	0.06	0.05	0.05	0.05	0.04
1000	0.04	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.05

Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (10, 1, 1, 1, 5)$, with $m = 2$ factors and $k = 3$ regressors. CCEP_x and CCEMG_x denote respectively the CCEP and CCEMG estimators with $\bar{\mathbf{y}}$ excluded from the matrix of CA. '*t*-test' resports the empirical size for a *t*-test at the $\alpha = 0.05$ significance level. '*basic*' reports empirical size for tests based on the basic ('empirical percentile') bootstrap interval, and bootstrap$-t$ and bootstrap$-t_c$ are respectively empirical size for the plain and corrected bootstrap$-t$ interval. All bootstrap tests are based on $B = 2000$ replications with the pairs (cross-section) resampling algorithm.

As predicted by Theorems 4 and 7, the *t*-test performs relatively well in the heterogeneous slope setting. This is because both the CCEP_x and CCEMG_x estimators are asymptotically unbiased. The bootstrap is therefore not as quintessential as in the fixed slope setting. For the CCEP_x estimator (top panel), the basic percentile interval performs noticeably worse in small N samples. The bootstrap-t_c interval, on the other hand, offers performance gains over the t-test when T is relatively large but otherwise largely mimics its behavior, with the t-test even performing slightly better in very T small samples. Hence, for the CCEP$_x$ estimator with heterogeneous slopes, the relative strength of the bootstrap-t_c is more pronounced in larger samples, and gains are generally not as large as when slopes are fixed. The main advantage is in stead that the bootstrap$-t_c$ applies directly to both the fixed and heterogeneous slope setting, while the results of the previous section clearly indicate that the t-test should not be attempted when slopes may be homogeneous. There is in other words, in contrast to the $t-$test, little risk associated with the bootstrap$-t_c$ on either setting so that one does not need to know whether slopes are heterogeneous to decide whether inferences can be trusted.

While the bootstrap$-t_c$ is more valuable in larger samples when applied to CCEP$_x$, the converse seems to be true for the CCEMG$_x$ estimator in the bottom panel. The basic bootstrap again does not offer performance gains, but the bootstrap-t and bootstrap-t_c tests do improve significantly over the t-test when N is very small, over the entire range of T. We find that the latter are correctly sized for all combinations of N and T, whereas the t-test is oversized when $N \leq 100$. This reflects an improved finite sample approximation of the null distribution by the bootstrap. That is, given that slope heterogeneity is drawn from the χ^2_1 distribution, the null distribution of the t-statistic is likely to be asymmetric and heavy tailed for small N so that the normal approximation employed by the t-test is not very accurate. The bootstrap intervals do not impose a distribution, or symmetry, and therefore achieve a more accurate test size. Indeed, we find that (unreported) symmetric versions of the (corrected) bootstrap-t test perform markedly worse than the unrestricted ones for this reason. The power functions for the CCEMG$_x$ estimator when $T/N = 2$ in Fig.5 are also informative for the situation.

Figure 5: **Power functions:** $T/N = 2$**, Heterogeneous slopes, CCEMG**$_x$



Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (10, 1, 1, 1, 5)$ with $m = 2$ and $k = 3$. The vertical axis presents rejection rates of the respective tests for the hypothesized value of β on the horizontal axis. Both the left and right panel are based on the CCEMG$_x$ estimator (with $\bar{\mathbf{y}}$ excluded).

Fig.5 shows that the small N rejection curves have been shifted to the left of $\beta = 1$ for the t-test but that the bootstrap has ensured that their minimum once again coincides with the correct $\beta = 1$ hypothesis at the $\alpha = 0.05$ intersection. Note that while this correction appears to have somewhat flattened the right hand side of the small N curves (blue and red) compared to those of the $t-$test, which implies a loss of power against $\beta > 1$ alternatives, this is compensated to some extent by the increase in power against $\beta < 1$ hypotheses caused by the re-centering.

We have also run experiments where the slope heterogeneity is drawn from the normal distribution. In this case there are no noticeable performance gains of the bootstrap as in Table 7. This is because the leading term in the distribution of the CCEMG estimator then has an *exact* normal distribution, i.e. for all sample sizes, in which case numerical approximations like the bootstrap will not lead to improvements. Assuming normality of the slopes is of course a strong assumption in practice and the experiment shows that the bootstrap offers robustness and performance gains in case the slope distribution is less well behaved.

5 Conclusion

We propose in this paper the cross-section or pairs bootstrap to improve inference with the CCE estimators in large N and T panels where the cross-section dimension need not dominate the time series length. In datasets of these dimensions standard asymptotic inference with the CCEP estimator is distorted by bias terms for which analytical corrections are not generally feasible.

We show in this paper that the cross-section bootstrap enables the elimination of this bias, and asymptotically correct inference, even when N does not dominate, so long as $0 \leq T/N < \infty$. This result holds true most generally when the cross-section average of the dependent variable is excluded from the estimation, but otherwise requires the number of factors to equal the number of cross-section averages. In the former setting, the number of factors or the general time series properties of the original dataset do not need to be known, which makes the algorithm both extremely simple and generally applicable. We show in addition that the bootstrap is also consistent for the distribution of the CCE estimators when slopes are heterogeneous, and leads to improved inference in this setting as well. In other words, the exact same bootstrap algorithm and inference procedures achieve asymptotically correct inferences on the population mean of the slopes, irrespective of whether those slopes are heterogeneous or not. This is a considerable advantage in practice. Monte Carlo simulations illustrate that these asymptotic properties also translate well to finite samples.

6 Acknowledgments

The authors thank Joakim Westerlund, Simon Reese, Arturas Juodis, and the participants of the Lund University, Department of Economics econometrics seminar for helpful comments. The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the Flemish Supercomputer Center, funded by Ghent University; the Hercules Foundation; and the Economy, Science, and Innovation Department of the Flemish Government.

References

Albanese, G. and Modica, S. (2012). Government size, the role of commitments*. *Oxford Bulletin of Economics and Statistics*, 74(4):532–546.

Andrews, D. W. K. (2005). Cross-section regression with common shocks. *Econometrica*, 73(5):1551–1585.

Bai, J. (2009). Panel Data Models with Interactive Fixed Effects. *Econometrica*, 77(4):1229–1279.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika*, 74(3):457–468.

Berger, T. and Heylen, F. (2011). Differences in hours worked in the oecd: Institutions or fiscal policies? *Journal of Money, Credit and Banking*, 43(7):1333–1369.

Bertoli, S. and Fernández-Huertas Moraga, J. (2013). Multilateral resistance to migration. *Journal of Development Economics*, 102:79 – 100. Migration and Development.

Bun, M. and Carree, M. (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business and Economic Statistics*, 23(2):200–210.

Chudik, A., Pesaran, M., and Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal*, 14(1):C45–C90.

De Vos, I. and Everaert, G. (2021). Bias-corrected common correlated effects pooled estimation in dynamic panels. *Journal of Business & Economic Statistics*, 39(1):294–306.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3):189–212.

Diciccio, T. J. and Romano, J. P. (1988). A review of bootstrap confidence intervals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(3):338–354.

Djogbenou, A., Gonçalves, S., and Perron, B. (2015). Bootstrap inference in regressions with estimated factors and serial correlation. *Journal of Time Series Analysis*, 36(3):481–502.

Eberhardt, M. and Teal, F. (2020). The magnitude of the task ahead: Macro implications of heterogeneous technology. *Review of Income and Wealth*, 66(2):334–360.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Efron, B. (1981). Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics*, 9(2):139–158.

Everaert, G. and Pozzi, L. (2007). Bootstrap-based bias correction for dynamic panels. *Journal of Economic Dynamics and Control*, 31(4):1160–1184.

Galvao, A. F. and Kato, K. (2014). Estimation and inference for linear panel data models under misspecification when both n and t are large. *Journal of Business & Economic Statistics*, 32(2):285–309.

Gonçalves, S. and Kaffo, M. (2015). Bootstrap inference for linear dynamic panel data models with individual fixed effects. *Journal of Econometrics*, 186(2):407–426. High Dimensional Problems in Econometrics.

Gonçalves, S. and Perron, B. (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics*, 182(1):156 – 173. Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions.

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

Hall, A. R., Rudebusch, G. D., and Wilcox, D. W. (1996). Judging instrument relevance in instrumental variables estimation. *International Economic review*, 37(2):283–298.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *The Annals of Statistics*, 16(3):927–953.

Juodis, A., Karabiyik, H., and Westerlund, J. (2021). On the robustness of the pooled cce estimator. *Journal of Econometrics*, 220(2):325–348. Annals Issue: Celebrating 40 Years of Panel Data Analysis: Past, Present and Future.

Kapetanios, G. (2008). A bootstrap procedure for panel data sets with many cross-sectional units. *Econometrics Journal*, 11(2):377–395.

Kapetanios, G., Pesaran, M., and Yamagata, T. (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics*, 160(2):326–348.

Karabiyik, H., Reese, S., and Westerlund, J. (2017). On the role of the rank condition in CCE estimation of factor-augmented panel regressions. *Journal of Econometrics*, 197(1):60 – 64.

Karabiyik, H., Urbain, J.-P., and Westerlund, J. (2019). CCE estimation of factor-augmented regression models with more factors than observables. *Journal of Applied Econometrics*, 34(2):268–284.

Kiviet, J. (1995). On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models. *Journal of Econometrics*, 68:53–78.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. Springer Series in Statistics. Springer.

Mazzanti, M. and Musolesi, A. (2013). The heterogeneity of carbon kuznets curves for advanced countries: comparing homogeneous, heterogeneous and shrinkage/bayesian estimators. *Applied Economics*, 45(27):3827–3842.

Millo, G. (2019). Private returns to R&D in the presence of spillovers, revisited. *Journal of Applied Econometrics*, 34(1):155–159.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, 49(6):1417–1426.

Pesaran, M. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica*, 74(4):967–1012.

Pesaran, M. and Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161(2):182–202.

Sarafidis, V. and Robertson, D. (2009). On the Impact of Error Cross-Sectional Dependence in Short Dynamic Panel Estimation. *Econometrics Journal*, 12(1):62–81.

Sarafidis, V. and Wansbeek, T. (2012). Cross-Sectional Dependence in Panel Data Analysis. *Econometric Reviews*, 31(5):483–531.

Stevens, J. and Childs, J. (2017). Re-examining the economic determinants of alcohol consumption in canada: controlling for the presence of common correlated effects. *Applied Economics Letters*, 24(16):1177–1180.

van Giersbergen, N. P. and Kiviet, J. F. (2002). How to implement the bootstrap in static or stable dynamic regression models: test statistic versus confidence region approach. *Journal of Econometrics*, 108(1):133 – 156.

Westerlund, J., Petrova, Y., and Norkute, M. (2019). CCE in fixed-T panels. *Journal of Applied Econometrics*, 34(5):746–761.

Westerlund, J. and Urbain, J. (2013a). On the estimation and inference in factor-augmented panel regressions with correlated loadings. *Economics Letters*, 119(3):247–250.

Westerlund, J. and Urbain, J.-P. (2013b). On the implementation and use of factor-augmented regressions in panel data. *Journal of Asian Economics*, 28:3 – 11. ACAES-FEG Special Issue: Econometrics of Financial Markets in Asia.

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

Özatay, F., Özmen, E., and Şahinbeyoğlu, G. (2009). Emerging market sovereign spreads, global financial conditions and u.s. macroeconomic news. *Economic Modelling*, 26(2):526 – 531.

Özmen, E. and Özge Doğanay Yaşar (2016). Emerging market sovereign bond spreads, credit ratings and global financial crisis. *Economic Modelling*, 59:93 – 101.

SUPPLEMENT A: MATHEMATICAL PROOFS FOR "BOOTSTRAP IMPROVED INFERENCE FOR FACTOR-AUGMENTED REGRESSIONS WITH CCE"

Ignace De Vos[1,2] and Ovidijus Stauskas[1]

[1]*Lund University, Department of Economics*
[2]*Ghent University, Department of Economics*

This supplement can be divided into three parts. The first section introduces the stacked notation and restates the working assumptions. Section 2 gathers proofs in the original sample, with preliminary results given in 2.1, and main results are derived in 2.2 for homogeneous slopes and in section 2.3 for heterogeneous slopes. Section 3 contains proofs in the bootstrap world, with basic properties for the bootstrap resampling operator presented in 3.1 and preliminary results derived in section 3.2. The main analysis under homogeneous slopes is presented in section 3.3 and slope heterogeneity is considered in 3.4.

1 Notation and assumptions

Let the stacked matrices of the main variables in model (1)-(3) of the main text be

$$\underset{(TN\times 1)}{\mathbf{y}} = [\mathbf{y}_1', \mathbf{y}_2', \dots, \mathbf{y}_N']', \quad \underset{(TN\times k)}{\mathbf{X}} = [\mathbf{X}_1', \mathbf{X}_2', \dots, \mathbf{X}_N']', \quad \underset{(TN\times 1+k)}{\mathbf{Z}} = [\mathbf{Z}_1', \dots, \mathbf{Z}_N']' = [\mathbf{y}, \mathbf{X}]$$
$$\underset{(TN\times 1)}{\boldsymbol{\varepsilon}} = [\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2', \dots, \boldsymbol{\varepsilon}_N']', \quad \underset{(TN\times k)}{\mathbf{V}} = [\mathbf{V}_1', \mathbf{V}_2', \dots, \mathbf{V}_N']', \quad \underset{(TN\times 1+k)}{\mathbf{U}} = [\mathbf{U}_1', \dots, \mathbf{U}_N']' = [\boldsymbol{\varepsilon}, \mathbf{V}]\mathbf{B}$$

so that we have the following data generating processes (DGP) for the stacked observed matrices

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \underline{\mathbf{F}}\,\underline{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon} \tag{1.1}$$
$$\mathbf{X} = \underline{\mathbf{F}}\,\underline{\boldsymbol{\Gamma}} + \mathbf{V} \tag{1.2}$$
$$\mathbf{Z} = \underline{\mathbf{F}}\,\underline{\mathbf{C}} + \mathbf{U} \tag{1.3}$$

where the remaining unobservables are defined as

$$\underset{(TN\times mN)}{\underline{\mathbf{F}}} = (\mathbf{I}_N \otimes \mathbf{F})$$
$$\underset{(Nm\times k)}{\underline{\mathbf{\Gamma}}} = [\mathbf{\Gamma}_1', \mathbf{\Gamma}_2', \ldots, \mathbf{\Gamma}_N']'$$
$$\underset{(Nm\times 1)}{\underline{\gamma}} = [\gamma_1', \gamma_2', \ldots, \gamma_N']'$$
$$\underset{(Nm\times 1+k)}{\boldsymbol{\eta}} = [\boldsymbol{\eta}_1', \boldsymbol{\eta}_2', \ldots, \boldsymbol{\eta}_N']'$$
$$\underset{(Nm\times 1+k)}{\underline{\mathbf{C}}} = [\underline{\gamma}, \underline{\mathbf{\Gamma}}]\mathbf{B} = (\boldsymbol{\iota}_N \otimes \mathbf{C}) + \boldsymbol{\eta}$$

The cross-section average operation for general 'stacks' of N cross-section specific $l-$rowed matrices is

$$\mathbf{A}_l = N^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{I}_l), \tag{1.4}$$

where we note that for matrices repeated over individuals, such as $\underline{\mathbf{F}}$, this averaging operator is commutative

$$\mathbf{A}_T\underline{\mathbf{F}} = N^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{I}_T)(\mathbf{I}_N \otimes \mathbf{F}) = N^{-1}(\boldsymbol{\iota}_N'\mathbf{I}_N \otimes \mathbf{F}) = \mathbf{F}N^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{I}_m) = \mathbf{F}\mathbf{A}_m \tag{1.5}$$

We work under (a subset of) the following assumptions:

Assumption 1 *(Idiosyncratic errors) $\varepsilon_{i,t}$ and $\mathbf{v}_{i,t}$ are stationary and independent across i with absolute summable autocovariances, $\mathbb{E}(\varepsilon_{i,t}) = 0$, $\mathbb{E}(\mathbf{v}_{i,t}) = \mathbf{0}_{k\times 1}$, $\sigma_i^2 = \mathbb{E}(\varepsilon_{i,t}^2)$, $\mathbf{\Sigma}_i = \mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}_{i,t}')$, $\mathbf{\Omega}_i = \mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')$, with $\mathbf{\Omega}_i, \mathbf{\Sigma}_i$ positive definite and $\mathbb{E}(\varepsilon_{i,t}^6) < \infty$, $\mathbb{E}(\|\mathbf{v}_{i,t}\|^6) < \infty$ for all i and t. Additionally, $\frac{1}{N}\sum_{i=1}^N \sigma_i^2 \to \sigma^2 < \infty$ and $\frac{1}{N}\sum_{i=1}^N \mathbf{\Sigma}_i \to \mathbf{\Sigma} < \infty$ as $N \to \infty$, and we define $\mathbf{\Sigma}_{\mathbf{u},i} = \mathbb{E}(\mathbf{u}_{i,t}\mathbf{u}_{i,t}') = \mathbf{B}'\mathbb{E}(\widetilde{\mathbf{u}}_{i,t}\widetilde{\mathbf{u}}_{i,t}')\mathbf{B} = \mathbf{B}'\mathbf{\Sigma}_{\widetilde{\mathbf{u}},i}\mathbf{B}$ and $\frac{1}{N}\sum_{i=1}^N \mathbf{\Sigma}_{\mathbf{u},i} \to \mathbf{B}'\mathbf{\Sigma}_{\widetilde{\mathbf{u}}}\mathbf{B} = \mathbf{\Sigma}_{\mathbf{u}}$, where $\mathbf{\Sigma}_{\widetilde{\mathbf{u}},i} = [[\sigma_i^2, \mathbf{0}_{1\times k}]', [\mathbf{0}_{k\times 1}, \mathbf{\Sigma}_i]']$ and $\mathbf{\Sigma}_{\widetilde{\mathbf{u}}} = [[\sigma^2, \mathbf{0}_{1\times k}]', [\mathbf{0}_{k\times 1}, \mathbf{\Sigma}]']$.*

Assumption 2 *(Common factors) $\mathbf{f}_t$ is covariance stationary with $\mathbb{E}(\|\mathbf{f}_t\|^4) < \infty$, absolute summable autocovariances and $T^{-1}\mathbf{F}'\mathbf{F} \to \mathbf{\Sigma}_{\mathbf{F}}$ as $T \to \infty$, with $\mathbf{\Sigma}_{\mathbf{F}}$ positive definite.*

Assumption 3 *(Factor loadings) The $\mathbf{C}_i$ are generated according to*

$$\mathbf{C}_i = \widetilde{\mathbf{C}}_i\mathbf{B}_i = (\widetilde{\mathbf{C}} + \widetilde{\boldsymbol{\eta}}_i)\mathbf{B}_i = \mathbf{C} + \boldsymbol{\eta}_i, \qquad vec(\widetilde{\boldsymbol{\eta}}_i) \sim IID(\mathbf{0}_{m(1+k)}, \mathbf{\Omega}_{\widetilde{\eta}}), \tag{1.6}$$

where $\widetilde{\mathbf{C}} = \mathbb{E}(\widetilde{\mathbf{C}}_i) = [\gamma, \mathbf{\Gamma}]$, $\mathbf{\Omega}_{\widetilde{\eta}} = \mathbb{E}(vec(\widetilde{\boldsymbol{\eta}})vec(\widetilde{\boldsymbol{\eta}})')$ and $\left\|\widetilde{\mathbf{C}}\right\|, \|\mathbf{\Omega}_{\widetilde{\eta}}\| < \infty$. We also define $\mathbf{\Sigma}_{\boldsymbol{\eta}} = \mathbb{E}(\widetilde{\boldsymbol{\eta}}_i \otimes \widetilde{\boldsymbol{\eta}}_i)$, which is a restructuring of $\mathbf{\Omega}_{\widetilde{\eta}}$

Assumption 4 *(Rank condition) $rk(\overline{\mathbf{C}}) = m$ for all N.*

Assumption 5 *(Independence) $\mathbf{f}_t, \varepsilon_{is}, \mathbf{v}_{jl}, \widetilde{\boldsymbol{\eta}}_n$ are mutually independent for all i, j, n, t, s, l.*

Assumption 6 *(Slope heterogeneity) The heterogeneous slope coefficients follow*

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{v}_i, \qquad \boldsymbol{v}_i \sim IID(\mathbf{0}_{k\times 1}, \boldsymbol{\Omega}_{\boldsymbol{v}})$$

with $\boldsymbol{\Omega}_{\boldsymbol{v}}$ *a finite nonnegative definite* $k \times k$ *matrix and the* $\boldsymbol{v}_i$ *are independent of* $\mathbf{f}_t, \varepsilon_{is}, \mathbf{v}_{jl}, \widetilde{\boldsymbol{\eta}}_n$ *for all* i, j, n, t, s, l.

Assumption 7 *(Rank condition)* $rk(\overline{\boldsymbol{\Gamma}}) = m$ *for all* N.

Some additional notation: In this supplement we use $\mathbf{A}^{\dagger}$ to denote the Moore-Penrose pseudo-inverse of the matrix $\mathbf{A}$, $rk(\mathbf{A})$ for its rank, $|\mathbf{A}|$ for the determinant and let $\|\mathbf{A}\| = [tr(\mathbf{A}\mathbf{A}')]^{1/2}$ be the Euclidean (Frobenius) matrix norm. Let furthermore $\boldsymbol{\iota}_a$ be an a-rowed vector of ones and the $vec(.)$, $\otimes$ and $\circ$ operators denote respectively the vectorization operation and the Kronecker- and Hadamard (element-wise) products. Barred variables $\overline{\mathbf{A}}$ denote the cross-section average (CA) over the cross-section specific matrices $\mathbf{A}_i$ as in $\overline{\mathbf{A}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{A}_i$. A starred object $\mathbf{A}_i^*$ stands for an *observed* variable (matrix or scalar) that has been generated in the bootstrap world according to the particular scheme. On the other hand, $\mathbf{A}_{w,i}$ is the *weighted unobserved* primitive of the model. We formalize the bootstrap probability laws similarly to Galvao and Kato (2014). In particular, for any matrix bootstrap sequence $\mathbf{A}_n^*$, which depends on a generic index n, and a deterministic sequence $a_n \in \mathbb{R}_{++}$, we have $\|\mathbf{A}_n^*\| = o_{p^*}(a_n)$ if for every $\epsilon > 0$ and $\delta > 0$, we have $\mathbb{P}(\mathbb{P}^*(a_n^{-1}\|\mathbf{A}_n^*\| > \epsilon) > \delta) \to 0$ as $n \to \infty$, where $\mathbb{P}^*(.)$ is a bootstrap-induced measure. Similarly, $\|\mathbf{A}_n^*\| = O_{p^*}(a_n)$ if for every $\delta > 0$ and $\eta > 0$, there exists a constant $C > 0$, such that $\mathbb{P}(\mathbb{P}^*(a_n^{-1}\|\mathbf{A}_n^*\| > C) > \delta) < \eta$ for all $n \geq 1$. Additionally, $\mathbb{E}^*(\cdot)$, $\mathbb{V}ar^*(\cdot)$ and $\mathbb{C}ov^*(\cdot,\cdot)$ represent, respectively, the expectation, variance and covariance taken with respect to the induced measure $\mathbb{P}^*$, and $\mathbf{A}_n^* = \mathbf{A}^* + o_{p^*}(1)$ means $\|\mathbf{A}_n^* - \mathbf{A}^*\| = o_{p^*}(1)$ for the limiting bootstrap matrix $\mathbf{A}^*$. Lastly, $\to^{p*}$ ($\to^{p}$) and $\xrightarrow{d^*}$ ($\xrightarrow{d}$) represent convergence in probability and distribution with respect to the induced (generic) probability measure.

2 Original Data

Note that

$$\overline{\mathbf{Z}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i = \mathbf{A}_T\mathbf{Z} = \mathbf{A}_T\underline{\mathbf{F}}\ \underline{\mathbf{C}} + \mathbf{A}_T\mathbf{U} = \mathbf{F}\mathbf{A}_m\underline{\mathbf{C}} + \mathbf{A}_T\mathbf{U} = \mathbf{F}\overline{\mathbf{C}} + \overline{\mathbf{U}}$$

such that also

$$\mathbf{F} = (\overline{\mathbf{Z}} - \overline{\mathbf{U}})\overline{\mathbf{C}}^{\dagger} \tag{2.1}$$

As Karabiyik et al. (2017) point out, the pseudo inverse $(T^{-1}\overline{\mathbf{Z}}'\overline{\mathbf{Z}})^{\dagger}$ when used in the projection matrix $\mathbf{P}_{\widehat{\mathbf{F}}} = \overline{\mathbf{Z}}(\overline{\mathbf{Z}}'\overline{\mathbf{Z}})^{\dagger}\overline{\mathbf{Z}}'$ is unbounded asymptotically since the $T \times 1 + k$ matrix $\overline{\mathbf{Z}}$ converges to a reduced rank matrix as $(N, T) \to \infty$ when $m < 1 + k$. This requires the use of the $\mathbf{R} = \mathbf{T}\overline{\mathbf{H}}\mathbf{D}_N$ rotation matrix in the analysis. Here, let $\mathbf{T}$ be the $(1 + k) \times (1 + k)$ matrix that partitions/reshuffles $\overline{\mathbf{C}}$ in an $m \times m$ full rank matrix $\overline{\mathbf{C}}_m$ and an $m \times (k + 1 - m)$ matrix $\overline{\mathbf{C}}_{-m}$ as $\overline{\mathbf{C}}\mathbf{T} = [\overline{\mathbf{C}}_m, \overline{\mathbf{C}}_{-m}]$ and yields the corresponding partitioning of the error terms $\overline{\mathbf{U}}\mathbf{T} = [\overline{\mathbf{U}}_m, \overline{\mathbf{U}}_{-m}]$. The remaining terms are

$$\overline{\mathbf{H}} = [\overline{\mathbf{H}}_m, \overline{\mathbf{H}}_{-m}] = \begin{bmatrix} \overline{\mathbf{C}}_m^{-1} & -\overline{\mathbf{C}}_m^{-1}\overline{\mathbf{C}}_{-m} \\ \mathbf{0}_{(k+1-m)\times m} & \mathbf{I}_{k+1-m} \end{bmatrix}, \qquad \mathbf{D}_N = \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times(k+1-m)} \\ \mathbf{0}_{(k+1-m)\times m} & \sqrt{N}\mathbf{I}_{k+1-m} \end{bmatrix} \tag{2.2}$$

where since it is easily seen under Ass.3 that $\overline{\mathbf{C}} = \mathbf{C} + O_p(N^{-1/2})$ and $rk(\overline{\mathbf{C}}_m) = m$ under Ass.4, we have $\|\overline{\mathbf{H}} - \mathbf{H}\| = O_p(N^{-1/2})$ with $\mathbf{H} = [\mathbf{H}_m, \mathbf{H}_{-m}] = \begin{bmatrix} \mathbf{C}_m^{-1} & -\mathbf{C}_m^{-1}\mathbf{C}_{-m} \\ \mathbf{0}_{(k+1-m)\times m} & \mathbf{I}_{k+1-m} \end{bmatrix}$ and $\mathbf{C}_m$ and $\mathbf{C}_{-m}$ denoting the partitioning following from $\mathbf{C}\mathbf{T} = [\mathbf{C}_m, \mathbf{C}_{-m}]$.
This gives in turn

$$\widehat{\mathbf{F}}^0 = \overline{\mathbf{Z}}\mathbf{R} = \overline{\mathbf{Z}}^0 = \widehat{\mathbf{F}}\mathbf{R} = [\mathbf{F}\overline{\mathbf{C}} + \overline{\mathbf{U}}]\mathbf{R} = \mathbf{F}^0 + \overline{\mathbf{U}}^0 \tag{2.3}$$

with $\mathbf{F}^0 = \mathbf{F}\mathbf{R} = [\mathbf{F}, \mathbf{0}_{T\times(k+1-m)}]$ and $\overline{\mathbf{U}}^0 = \overline{\mathbf{U}}\mathbf{R} = [\overline{\mathbf{U}}_m^0, \overline{\mathbf{U}}_{-m}^0]$, where $\overline{\mathbf{U}}_m^0 = \overline{\mathbf{U}}_m\overline{\mathbf{C}}_m^{-1}$ and $\overline{\mathbf{U}}_{-m}^0 = \sqrt{N}\overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_{-m} = \sqrt{N}(\overline{\mathbf{U}}_m - \overline{\mathbf{U}}_m\overline{\mathbf{C}}_m^{-1}\overline{\mathbf{C}}_{-m})$. Here we note that since $\mathbf{R}$ is full rank, $\mathbf{P}_{\widehat{\mathbf{F}}^0} = \overline{\mathbf{Z}}\mathbf{R}(\mathbf{R}'\overline{\mathbf{Z}}'\overline{\mathbf{Z}}\mathbf{R})^{\dagger}\mathbf{R}'\overline{\mathbf{Z}}' = \overline{\mathbf{Z}}(\overline{\mathbf{Z}}'\overline{\mathbf{Z}})^{\dagger}\overline{\mathbf{Z}}' = \mathbf{P}_{\widehat{\mathbf{F}}}$ and analyzing $\mathbf{P}_{\widehat{\mathbf{F}}^0}$ is equivalent to analyzing $\mathbf{P}_{\widehat{\mathbf{F}}}$.
Substituting in $\mathbf{F} = (\overline{\mathbf{Z}} - \overline{\mathbf{U}})\overline{\mathbf{C}}^{\dagger}$ from (2.1) into the DGP of $\mathbf{y}_i$ and $\mathbf{X}_i$ yields

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{F}\gamma_i + \varepsilon_i = \mathbf{X}_i\boldsymbol{\beta}_i + \overline{\mathbf{Z}}\overline{\mathbf{C}}^{\dagger}\gamma_i + \varepsilon_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma_i \tag{2.4}$$

$$\mathbf{X}_i = \mathbf{F}\boldsymbol{\Gamma}_i + \mathbf{V}_i = \overline{\mathbf{Z}}\overline{\mathbf{C}}^{\dagger} + \mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i \tag{2.5}$$

2.1 Preliminary results

Let $\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0} = T^{-1}(\widehat{\mathbf{F}}^0)'\widehat{\mathbf{F}}^0$ and

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u} = \begin{bmatrix} \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}} & \mathbf{0}_{m\times(1+k-m)} \\ \mathbf{0}_{(1+k-m)\times m} & \widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0} \end{bmatrix} \tag{2.6}$$

where $\widehat{\mathbf{\Sigma}}_{\mathbf{F}} = T^{-1}\mathbf{F}'\mathbf{F}$ and $\widehat{\mathbf{\Sigma}}_{\mathbf{u}^0_{-m}} = T^{-1}(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}}^0_{-m}$ and

$$\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0} = T^{-1}\overline{\mathbf{U}}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0}(\overline{\mathbf{U}}^0)' + T^{-1}\overline{\mathbf{U}}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0}(\mathbf{F}^0)' + T^{-1}\mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0}(\overline{\mathbf{U}}^0)' + T^{-1}\mathbf{F}^0\left[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - [T^{-1}(\mathbf{F}^0)'\mathbf{F}^0]^{\dagger}\right](\mathbf{F}^0)'. \tag{2.7}$$

By using the definition of $\overline{\mathbf{U}}^0$, the two first terms on the right-hand side of the expansion above can be written as

$$\begin{aligned} T^{-1}\overline{\mathbf{U}}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0}(\overline{\mathbf{U}}^0)' &= T^{-1}[\overline{\mathbf{U}}^0_m, \overline{\mathbf{U}}^0_{-m}]\begin{bmatrix} T^{-1}\mathbf{F}'\mathbf{F} & \mathbf{0}_{m\times(k+1-m)} \\ \mathbf{0}_{(k+1-m)\times m} & T^{-1}(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}}^0_{-m}\end{bmatrix}^{\dagger}\begin{bmatrix}\overline{\mathbf{U}}^{0\prime}_m \\ \overline{\mathbf{U}}^{0\prime}_{-m}\end{bmatrix} + \overline{\mathbf{U}}^0\left[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right](\overline{\mathbf{U}}^0)' \\ &= T^{-1}\overline{\mathbf{U}}^0_m\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}\overline{\mathbf{U}}^{0\prime}_m + T^{-1}\overline{\mathbf{U}}^0_{-m}(T^{-1}(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}}^0_{-m})^{\dagger}\overline{\mathbf{U}}^{0\prime}_{-m} + T^{-1}\overline{\mathbf{U}}^0\left[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right](\overline{\mathbf{U}}^0)', \end{aligned} \tag{2.8}$$

and

$$\begin{aligned} &T^{-1}\overline{\mathbf{U}}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0}(\mathbf{F}^0)' \\ &= T^{-1}[\overline{\mathbf{U}}^0_m, \overline{\mathbf{U}}^0_{-m}]\begin{bmatrix} T^{-1}\mathbf{F}'\mathbf{F} & \mathbf{0}_{m\times(k+1-m)} \\ \mathbf{0}_{(k+1-m)\times m} & T^{-1}(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}}^0_{-m}\end{bmatrix}^{\dagger}\begin{bmatrix}\mathbf{F}' \\ \mathbf{0}_{(k+1-m)\times T}\end{bmatrix} + T^{-1}\overline{\mathbf{U}}^0\left[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right](\mathbf{F}^0)' \\ &= T^{-1}\overline{\mathbf{U}}^0_m\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}\mathbf{F}' + T^{-1}\overline{\mathbf{U}}^0\left[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right](\mathbf{F}^0)' \end{aligned} \tag{2.9}$$

The third term is just a transpose of the second. The fourth can be rewritten using

$$\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u} = (T^{-1}(\mathbf{F}^0)'\mathbf{F}^0)^{\dagger} + \begin{bmatrix} \mathbf{0}_{m\times m} & \mathbf{0}_{m\times(k+1-m)} \\ \mathbf{0}_{(k+1-m)\times m} & (T^{-1}(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}}^0_{-m})^{\dagger}\end{bmatrix}.$$

Hence, because the last $k+1-m$ rows of $\mathbf{F}^0$ are zero,

$$\begin{aligned} &T^{-1}\mathbf{F}^0[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - (T^{-1}(\mathbf{F}^0)'\mathbf{F}^0)^{\dagger}](\mathbf{F}^0)' \\ &= T^{-1}\mathbf{F}^0\left[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right](\mathbf{F}^0)' + T^{-1}\mathbf{F}^0\begin{bmatrix} \mathbf{0}_{m\times m} & \mathbf{0}_{m\times(k+1-m)} \\ \mathbf{0}_{(k+1-m)\times m} & (T^{-1}(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}}^0_{-m})^{\dagger}\end{bmatrix}(\mathbf{F}^0)' \\ &= T^{-1}\mathbf{F}^0\left[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right](\mathbf{F}^0)' \end{aligned} \tag{2.10}$$

By substituting (2.8)–(2.10) into (2.7), and using the definition of $\widehat{\mathbf{F}}^0$, we obtain

$$\begin{aligned} \mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0} &= T^{-1}\overline{\mathbf{U}}^0_{-m}(T^{-1}(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}}^0_{-m})^{\dagger}(\overline{\mathbf{U}}^0_{-m})' + T^{-1}\overline{\mathbf{U}}^0_m\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}(\overline{\mathbf{U}}^0_m)' + T^{-1}\overline{\mathbf{U}}^0_m\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}\mathbf{F}' + T^{-1}\mathbf{F}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}(\overline{\mathbf{U}}^0_m)' \\ &\quad + T^{-1}\widehat{\mathbf{F}}^0\left[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right]\widehat{\mathbf{F}}^{0\prime} \\ &= T^{-1}\overline{\mathbf{U}}^0_{-m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{-m}}(\overline{\mathbf{U}}^0_{-m})' + T^{-1}\overline{\mathbf{U}}^0_m\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}(\overline{\mathbf{U}}^0_m)' + T^{-1}\mathbf{F}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}(\overline{\mathbf{U}}^0_m)' + T^{-1}\overline{\mathbf{U}}^0_m\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}\mathbf{F}' \\ &\quad + T^{-1}\widehat{\mathbf{F}}^0\left[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right](\widehat{\mathbf{F}}^0)', \end{aligned} \tag{2.11}$$

This expression will play an important role in the analysis that follows.

Next we establish the following auxiliary lemmas.

Lemma B-1 *Under Ass.1-3, 5 and 6, it follows as* $(N, T) \to \infty$ *that*

$$T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}} = O_p(N^{-1}) \qquad T^{-1}\mathbf{F}'\overline{\mathbf{U}} = O_p((NT)^{-1/2})$$
$$T^{-1}\mathbf{F}'\mathbf{U}_i = O_p(T^{-1/2}) \qquad T^{-1}\overline{\mathbf{U}}'\mathbf{U}_i = O_p(N^{-1}) + O_p((NT)^{-1/2})$$

Proof of Lemma B-1. Under Ass.1-3, 5 and 6 the proof of this lemma is identical to Lemmas 1 and 2 in Pesaran (2006). Details are therefore omitted.

Lemma B-2 *Under Ass.1-6 it follows as* $(N, T) \to \infty$ *that*

$$T^{-1}(\overline{\mathbf{U}}^0_m)'\overline{\mathbf{U}}^0_m = O_p(N^{-1}) \qquad T^{-1}(\overline{\mathbf{U}}^0_m)'\overline{\mathbf{U}}^0_{-m} = O_p(N^{-1/2})$$
$$T^{-1}\mathbf{F}'\overline{\mathbf{U}}^0_m = O_p((NT)^{-1/2}) \qquad T^{-1}\mathbf{F}'\overline{\mathbf{U}}^0_{-m} = O_p(T^{-1/2})$$
$$T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_m = O_p(N^{-1}) \qquad T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_{-m} = O_p(N^{-1/2})$$
$$T^{-1}(\overline{\mathbf{U}}^0_m)'\mathbf{U}_i = O_p(N^{-1}) + O_p((NT)^{-1/2})$$
$$T^{-1}(\overline{\mathbf{U}}^0_{-m})'\mathbf{U}_i = O_p(N^{-1/2}) + O_p(T^{-1/2})$$
$$T^{-1}\widehat{\mathbf{F}}^{0\prime}\overline{\mathbf{U}} = O_p(N^{-1/2}) \qquad T^{-1}\widehat{\mathbf{F}}^{0\prime}\mathbf{U}_i = O_p(N^{-1/2}) + O_p(T^{-1/2})$$

moreover, with $\widehat{\mathbf{\Sigma}}_{\mathbf{u}^0_{-m}} = T^{-1}(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}}^0_{-m}$ *and* $\widehat{\mathbf{\Sigma}}_{\mathbf{F}_u}$ *defined in* (2.6)

$$NT^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}} = \mathbf{\Sigma}_{\mathbf{u},h} + O_p(T^{-1/2}) \tag{2.12}$$
$$\widehat{\mathbf{\Sigma}}_{\mathbf{u}^0_{-m}} = \mathbf{\Sigma}_{\mathbf{u}^0_{-m}} + O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{2.13}$$
$$\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{-m}} = \mathbf{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}} + O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{2.14}$$
$$\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right\| = O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{2.15}$$

where $\mathbf{\Sigma}_{\mathbf{u}^0_{-m}} = \mathbf{H}'_{-m}\mathbf{T}'\mathbf{\Sigma}_{\mathbf{u},h}\mathbf{T}\mathbf{H}_{-m}$, $\mathbf{\Sigma}_{\mathbf{u},h} = \mathbf{\Sigma}_{\mathbf{u}} + \begin{pmatrix} \mathbf{\Omega}_{\boldsymbol{v},\otimes} vec(\mathbf{\Sigma}) & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} \end{pmatrix}$ *and where* $\mathbf{\Omega}_{\boldsymbol{v},\otimes} = \mathbb{E}(\boldsymbol{v}'_i \otimes \boldsymbol{v}'_i)$. *If* $\boldsymbol{v}_i = \mathbf{0}_{k\times 1} \forall i$ *(homogeneous slopes), then* $\mathbf{\Sigma}_{\mathbf{u},h} = \mathbf{\Sigma}_{\mathbf{u}}$.

Proof of Lemma B-2.

From Lemma B-1 we have $\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\right\| = O_p(N^{-1})$ and $\left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}\right\| = O_p((NT)^{-1/2})$, such that substituting in the respective definitions and noting that $\|\mathbf{T}\| = O_p(1)$, $\left\|\overline{\mathbf{H}}\right\| = O_p(1)$ (and therefore also its partitioning) gives

$$\left\|T^{-1}(\overline{\mathbf{U}}^0_m)'\overline{\mathbf{U}}^0_m\right\| = \left\|\overline{\mathbf{H}}'_m\mathbf{T}'T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_m\right\| \leq \left\|\overline{\mathbf{H}}_m\right\|^2\|\mathbf{T}\|^2\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\right\| = O_p(N^{-1})$$
$$\left\|T^{-1}(\overline{\mathbf{U}}^0_m)'\overline{\mathbf{U}}^0_{-m}\right\| = \sqrt{N}\left\|\overline{\mathbf{H}}'_m\mathbf{T}'T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_{-m}\right\| \leq \sqrt{N}\left\|\overline{\mathbf{H}}_m\right\|\left\|\overline{\mathbf{H}}_{-m}\right\|\|\mathbf{T}\|^2\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\right\| = O_p(N^{-1/2})$$
$$\left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}^0_m\right\| = \left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_m\right\| \leq \left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}\right\|\|\mathbf{T}\|\left\|\overline{\mathbf{H}}_m\right\| = O_p((NT)^{-1/2})$$
$$\left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}^0_{-m}\right\| = \sqrt{N}\left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_{-m}\right\| \leq \sqrt{N}\left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}\right\|\|\mathbf{T}\|\left\|\overline{\mathbf{H}}_{-m}\right\| = O_p(T^{-1/2})$$
$$\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_m\right\| = \left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_m\right\| \leq \left\|\overline{\mathbf{H}}_m\right\|\|\mathbf{T}\|\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\right\| = O_p(N^{-1})$$
$$\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_{-m}\right\| = \sqrt{N}\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_{-m}\right\| \leq \sqrt{N}\left\|\overline{\mathbf{H}}_{-m}\right\|\|\mathbf{T}\|\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\right\| = O_p(N^{-1/2})$$

Similarly making use of $\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{U}_i\right\| = O_p(N^{-1}) + O_p((NT)^{-1/2})$ from lemma B-1

$$\left\|T^{-1}(\overline{\mathbf{U}}^0_m)'\mathbf{U}_i\right\| \leq \left\|\mathbf{T}\overline{\mathbf{H}}_m\right\| \left\|T^{-1}\overline{\mathbf{U}}'\mathbf{U}_i\right\| = O_p(N^{-1}) + O_p((NT)^{-1/2})$$
$$\left\|T^{-1}(\overline{\mathbf{U}}^0_{-m})'\mathbf{U}_i\right\| \leq \sqrt{N}\left\|\mathbf{T}\overline{\mathbf{H}}_{-m}\right\| \left\|T^{-1}\overline{\mathbf{U}}'\mathbf{U}_i\right\| = O_p(N^{-1/2}) + O_p(T^{-1/2})$$

Next, noting that $\mathbf{F}^0 = [\mathbf{F}, \mathbf{0}_{T\times 1+k-m}]$ and now inserting $\left\|T^{-1}\mathbf{F}'\mathbf{U}_i\right\|$ making use of the orders in Lemma B-1 gives

$$\left\|T^{-1}\widehat{\mathbf{F}}^{0\prime}\overline{\mathbf{U}}\right\| \leq \left\|T^{-1}(\mathbf{F}^0)'\overline{\mathbf{U}}\right\| + \sqrt{N}\left\|\mathbf{T}\overline{\mathbf{H}}\right\| \left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\right\| = O_p(N^{-1/2})$$
$$\left\|T^{-1}\widehat{\mathbf{F}}^{0\prime}\mathbf{U}_i\right\| \leq \left\|T^{-1}(\mathbf{F}^0)'\mathbf{U}_i\right\| + \sqrt{N}\left\|\mathbf{T}\overline{\mathbf{H}}\right\| \left\|T^{-1}\overline{\mathbf{U}}'\mathbf{U}_i\right\| = O_p(N^{-1/2}) + O_p(T^{-1/2})$$

This establishes the first set of results in the lemma.

Next, for eq.(2.12) we get

$$\begin{aligned}\frac{N\overline{\mathbf{U}}'\overline{\mathbf{U}}}{T} &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\mathbf{U}_i'\mathbf{U}_j}{T} = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{U}_i'\mathbf{U}_i}{T} + \frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\mathbf{U}_i'\mathbf{U}_j}{T} \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{U}_i'\mathbf{U}_i}{T} + O_p(T^{-1/2}) \\ &\longrightarrow^p \boldsymbol{\Sigma}_{\mathbf{u},h}\end{aligned} \tag{2.16}$$

with $\boldsymbol{\Sigma}_{\mathbf{u},h} = \boldsymbol{\Sigma}_{\mathbf{u}} + \begin{pmatrix}\boldsymbol{\Omega}_{\boldsymbol{v},\otimes}vec(\boldsymbol{\Sigma}) & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k}\end{pmatrix}$ and where $\boldsymbol{\Omega}_{\boldsymbol{v},\otimes} = \mathbb{E}(\boldsymbol{v}_i' \otimes \boldsymbol{v}_i')$. The remainder in T in the second line follows from

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\mathbf{U}_i'\mathbf{U}_j}{T}\right\| = O_p(T^{-1/2}), \tag{2.17}$$

because

$$\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\mathbf{U}_i'\mathbf{U}_j}{T}\right) = \mathbf{0}_{(1+k)\times(1+k)} \tag{2.18}$$

and

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\mathbf{U}_i'\mathbf{U}_j}{T}\right)\left(\frac{1}{N}\sum_{k=1}^{N}\sum_{l\neq k}^{N}\frac{\mathbf{U}_k'\mathbf{U}_l}{T}\right)'\right] &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k=1}^{N}\sum_{l\neq k}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left(\mathbf{u}_{i,t}\mathbf{u}_{j,t}'\mathbf{u}_{l,s}\mathbf{u}_{k,s}'\right) \\ = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left(\mathbf{u}_{i,t}\mathbf{u}_{j,t}'\mathbf{u}_{j,s}\mathbf{u}_{i,s}'\right) &+ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left(\mathbf{u}_{i,t}\mathbf{u}_{j,t}'\mathbf{u}_{i,s}\mathbf{u}_{j,s}'\right) \\ = \frac{1}{(NT)^2}O_p(N^2T) = O_p(T^{-1}),\end{aligned} \tag{2.19}$$

since due to cross-sectional independence of $\boldsymbol{\varepsilon}_i, \mathbf{V}_i, \boldsymbol{v}_i$ stipulated in Ass.1 and 6, the expectation is non-zero only if $i = k, j = l$ or $i = l, j = k$ and the final step comes from the following argument. Recalling that

$i \neq j$:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left(\mathbf{u}_{i,t}\mathbf{u}_{j,t}'\mathbf{u}_{i,s}\mathbf{u}_{j,s}'\right) = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{u}_{i,t}\mathbf{u}_{j,t}'\mathbf{u}_{i,t}\mathbf{u}_{j,t}'\right) + \frac{1}{T}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\mathbb{E}\left(\mathbf{u}_{i,t}\mathbf{u}_{j,t}'\mathbf{u}_{i,s}\mathbf{u}_{j,s}'\right) = O(1) + O(1) \tag{2.20}$$

because finite second moments of $\mathbf{u}_{i,t}$ and absolute summable autocovariances of $\mathbf{u}_{i,t}$ ensure that

$$\sum_{s\neq t}^{T}\mathbb{E}\left(\mathbf{u}_{i,t}\mathbf{u}_{j,t}'\mathbf{u}_{i,s}\mathbf{u}_{j,s}'\right) = O(1). \tag{2.21}$$

We will use this argument to deduce the orders of the similar terms. For the leading term in (2.16), making use of $\mathbf{B}_i = \mathbf{B} + \widetilde{\mathbf{B}}_i$ gives

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{U}_i'\mathbf{U}_i}{T} &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{B}'\left(\frac{\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i}{T}\right)\mathbf{B} + \frac{1}{N}\sum_{i=1}^{N}\mathbf{B}'\left(\frac{\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i}{T}\right)\widetilde{\mathbf{B}}_i + \frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{B}}_i'\left(\frac{\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i}{T}\right)\mathbf{B} + \frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{B}}_i'\left(\frac{\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i}{T}\right)\widetilde{\mathbf{B}}_i \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbf{B}'\left(\frac{\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i}{T}\right)\mathbf{B} + \frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{B}}_i'\left(\frac{\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i}{T}\right)\widetilde{\mathbf{B}}_i + O_p\left(\frac{1}{\sqrt{N}}\right) \\
&\longrightarrow^p \mathbf{\Sigma}_{\mathbf{u}} + \begin{pmatrix} \mathbf{\Omega}_{\boldsymbol{v},\otimes} vec(\mathbf{\Sigma}) & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} \end{pmatrix} = \mathbf{\Sigma}_{\mathbf{u},h}
\end{aligned}$$

as $(N,T)\to\infty$. On the second line, since by Ass.6 we have $\mathbb{E}(\boldsymbol{v}_i) = \mathbf{0}_{k\times 1}$ so that $\mathbb{E}(\widetilde{\mathbf{B}}_i) = \mathbf{0}_{(1+k)\times(1+k)}$ and since also $\boldsymbol{v}_i$ is independent of the other variables and over i,

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{B}'\left(\frac{\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i}{T}\right)\widetilde{\mathbf{B}}_i\right] = \mathbf{0}_{(1+k)\times(1+k)}$$

and also

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{B}'\left(\frac{\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i}{T}\right)\widetilde{\mathbf{B}}_i\right]\left[\frac{1}{N}\sum_{j=1}^{N}\mathbf{B}'\left(\frac{\widetilde{\mathbf{U}}_j'\widetilde{\mathbf{U}}_j}{T}\right)\widetilde{\mathbf{B}}_j\right]' = \mathbb{E}\left[\frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{B}'\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i\widetilde{\mathbf{B}}_i\widetilde{\mathbf{B}}_j'\widetilde{\mathbf{U}}_j'\widetilde{\mathbf{U}}_j\mathbf{B}\right] \\
&= \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{B}'\mathbb{E}\left[\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i\mathbb{E}(\widetilde{\mathbf{B}}_i\widetilde{\mathbf{B}}_j'|\widetilde{\mathbf{U}}_i,\widetilde{\mathbf{U}}_j)\widetilde{\mathbf{U}}_j'\widetilde{\mathbf{U}}_j\right]\mathbf{B} \\
&= \frac{1}{(NT)^2}\sum_{i=1}^{N}\mathbf{B}'\mathbb{E}\left[\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i\mathbb{E}(\widetilde{\mathbf{B}}_i\widetilde{\mathbf{B}}_i')\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i\right]\mathbf{B} = O\left(\frac{1}{N}\right)
\end{aligned}$$

results in

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{B}}_i'\left(\frac{\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i}{T}\right)\mathbf{B}\right\| = O_p\left(\frac{1}{\sqrt{N}}\right),$$

The final line follows from

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\mathbf{B}'\left(\frac{\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i}{T}\right)\mathbf{B} &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{\Sigma}_{\mathbf{u},i} + O_p(T^{-1/2}) \\
&\longrightarrow^p \mathbf{\Sigma}_{\mathbf{u}}
\end{aligned}$$

as $(N,T)\to\infty$ by Ass.1 and because $\widetilde{\mathbf{U}}_i\widetilde{\mathbf{B}}_i=[\mathbf{V}_i\boldsymbol{v}_i,\mathbf{0}_{T\times k}]$ leads to

$$\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{B}}_i'\left(\frac{\widetilde{\mathbf{U}}_i'\widetilde{\mathbf{U}}_i}{T}\right)\widetilde{\mathbf{B}}_i=\begin{pmatrix}\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{v}_i'\mathbf{V}_i'\mathbf{V}_i\boldsymbol{v}_i & \mathbf{0}_{1\times k}\\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k}\end{pmatrix}=\begin{pmatrix}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{v}_i'\boldsymbol{\Sigma}_i\boldsymbol{v}_i & \mathbf{0}_{1\times k}\\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k}\end{pmatrix}+O_p(T^{-1/2})$$
$$\longrightarrow^p\begin{pmatrix}\boldsymbol{\Omega}_{v,\otimes}vec(\boldsymbol{\Sigma}) & \mathbf{0}_{1\times k}\\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k}\end{pmatrix}$$

as $(N,T)\to\infty$, with $\boldsymbol{\Omega}_{v,\otimes}=\mathbb{E}(\boldsymbol{v}_i'\otimes\boldsymbol{v}_i')$ and $T^{-1}\mathbf{V}_i'\mathbf{V}_i=\boldsymbol{\Sigma}_i+O_p(T^{-1/2})$ from Ass.1 and 6. Note that the term is zero when $\boldsymbol{v}_i=\mathbf{0}_{k\times 1}$. This establishes (2.12).

Next, again making use of $\overline{\mathbf{U}}^0_{-m}=\sqrt{N}\overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_{-m}$ and substituting in (2.12) gives, as $(N,T)\to\infty$

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}^0_{-m}}&=T^{-1}(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}}^0_{-m}=\overline{\mathbf{H}}'_{-m}\mathbf{T}'\frac{N\overline{\mathbf{U}}'\overline{\mathbf{U}}}{T}\mathbf{T}\overline{\mathbf{H}}_{-m}=\overline{\mathbf{H}}'_{-m}\mathbf{T}'\boldsymbol{\Sigma}_{\mathbf{u},h}\mathbf{T}\overline{\mathbf{H}}_{-m}+O_p(T^{-1/2})\\&=\mathbf{H}'_{-m}\mathbf{T}'\boldsymbol{\Sigma}_{\mathbf{u},h}\mathbf{T}\mathbf{H}_{-m}+O_p(N^{-1/2})+O_p(T^{-1/2})\\&=\boldsymbol{\Sigma}_{\mathbf{u}^0_{-m}}+O_p(N^{-1/2})+O_p(T^{-1/2})\end{aligned}$$

because $\|\overline{\mathbf{H}}_{-m}-\mathbf{H}_{-m}\|=O_p(N^{-1/2})$ and note that $\boldsymbol{\Sigma}_{\mathbf{u}^0_{-m}}=\mathbf{H}'_{-m}\mathbf{T}'\boldsymbol{\Sigma}_{\mathbf{u},h}\mathbf{T}\mathbf{H}_{-m}$ is a $(1+k-m)\times(1+k-m)$ positive definite matrix. It is positive definite, because $\boldsymbol{\Sigma}_{\mathbf{u},h}$ is positive definite and hence by Exercise 8.26 in Abadir and Magnus (2005), $rk(\mathbf{T}'\boldsymbol{\Sigma}_{\mathbf{u},h}\mathbf{T})=rk(\mathbf{T})$. Therefore, by part (b) of the same exercise, $\text{rk}(\mathbf{B}'\mathbf{A}\mathbf{B})=\text{rk}(\mathbf{B})=q$ for $\mathbf{B}\in\mathbb{R}^{p\times q}$ and positive definite $\mathbf{A}\in\mathbb{R}^{p\times p}$ implies that the whole matrix quadratic form is positive definite. The result follows by applying the same argument again, now taking $\mathbf{H}_{-m}$ as $\mathbf{B}$ and $\mathbf{T}'\boldsymbol{\Sigma}_{\mathbf{u},h}\mathbf{T}$ as $\mathbf{A}$. This establishes (2.13), and since $rk(\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}^0_{-m}})-rk(\boldsymbol{\Sigma}_{\mathbf{u}^0_{-m}})\overset{a.s.}{\to}0$, it follows from Theorem 1 in Karabiyik et al. (2017)

$$\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{u}^0_{-m}}=\boldsymbol{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}+O_p(N^{-1/2})+O_p(T^{-1/2})$$

This is (2.14) of the lemma.

Next, consider

$$\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u}=\frac{1}{T}\begin{bmatrix}\mathbf{F}'\overline{\mathbf{U}}^0_m+(\overline{\mathbf{U}}^0_m)'\mathbf{F} & \mathbf{F}'\overline{\mathbf{U}}^0_{-m}\\(\overline{\mathbf{U}}^0_{-m})'\mathbf{F} & \mathbf{0}_{(1+k-m)\times(1+k-m)}\end{bmatrix}+\frac{1}{T}\begin{bmatrix}(\overline{\mathbf{U}}^0_m)'\overline{\mathbf{U}}^0_m & (\overline{\mathbf{U}}^0_m)'\overline{\mathbf{U}}^0_{-m}\\(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}}^0_m & \mathbf{0}_{(1+k-m)\times(1+k-m)}\end{bmatrix}$$

where substituting in the results established in the first part of the lemma results in

$$\left\|\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u}\right\|=O_p(N^{-1/2})+O_p(T^{-1/2})$$

Noting then that $rk(\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0})=1+k$, and since $\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u}$ is a block diagonal matrix also $rk(\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u})=rk(\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}})+rk(\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}^0_{-m}})=1+k$ even as $(N,T)\to\infty$ from Ass.1 and 2. Since then $rk(\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0})-rk(\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u})\overset{a.s.}{\to}0$, we have by Theorem 1 in Karabiyik et al. (2017)

$$\left\|\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0}-\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right\|=O_p(N^{-1/2})+O_p(T^{-1/2})$$

which establishes the last statement of the lemma in (2.15).

2.2 Homogeneous Slopes

In the homogeneous slope setting, we impose a common slope by setting $\boldsymbol{v}_i = \mathbf{0}_{k\times 1}$ so that $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ and $\mathbf{B}_i = \mathbf{B}$ for all $i = 1, \ldots, N$. In this setting, given that $\mathbf{M}_{\widehat{\mathbf{F}}}\overline{\mathbf{Z}} = \mathbf{0}_{T\times(1+k)}$, the scaled deviation of the CCEP estimator is

$$
\begin{aligned}
\sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= -\sqrt{NT}\boldsymbol{\beta} + ((NT)^{-1}\mathbf{X}'\underline{\mathbf{M}}_{\widehat{\mathbf{F}}}\mathbf{X})^{-1}(NT)^{-1/2}\mathbf{X}'\underline{\mathbf{M}}_{\widehat{\mathbf{F}}}\mathbf{y} \\
&= -\sqrt{NT}\boldsymbol{\beta} + \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{y}_i \\
&= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}[\varepsilon_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma_i] \\
&= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}[\varepsilon_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}(\gamma + \widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y)] \\
&= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}[\varepsilon_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y] \\
&= \widehat{\mathbf{Q}}^{-1}\widehat{\mathbf{q}}, \qquad (2.22)
\end{aligned}
$$

where use was made of $\gamma_i = \mathbf{C}_i\mathbf{B}^{-1}\mathbf{q}_y = (\mathbf{C} + \boldsymbol{\eta}_i)\mathbf{B}^{-1}\mathbf{q}_y = \gamma + \widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y$ and $\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma = N\overline{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{F}}}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma = \mathbf{0}_{k\times 1}$, because $\overline{\mathbf{X}} \subset \overline{\mathbf{Z}}$. Making use of $\mathbf{M}_{\widehat{\mathbf{F}}} = \mathbf{M}_{\widehat{\mathbf{F}}^0}$ and $\mathbf{M}_{\widehat{\mathbf{F}}^0} = \mathbf{M}_{\mathbf{F}^0} - [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]$, the denominator in the final expression is

$$
\begin{aligned}
\widehat{\mathbf{Q}} &= \frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i = \frac{1}{NT}\sum_{i=1}^{N}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i]'\mathbf{M}_{\widehat{\mathbf{F}}}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i] \\
&= \frac{1}{NT}\sum_{i=1}^{N}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i]'\mathbf{M}_{\mathbf{F}^0}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i] - \frac{1}{NT}\sum_{i=1}^{N}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i]'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}][\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i] \\
&= \widehat{\mathbf{Q}}_{\mathbf{M}_{\mathbf{F}^0}} - \widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]}. \qquad (2.23)
\end{aligned}
$$

For a stated subscript $\mathbf{A}$, we define the further decomposition

$$
\begin{aligned}
\widehat{\mathbf{Q}}_{\mathbf{A}} &= \widehat{\mathbf{Q}}_{\mathbf{A},VV} - \widehat{\mathbf{Q}}_{\mathbf{A},V\Gamma} - (\widehat{\mathbf{Q}}_{\mathbf{A},V\Gamma})' + \widehat{\mathbf{Q}}_{\mathbf{A},\Gamma\Gamma} \\
\widehat{\mathbf{Q}}_{\mathbf{A},VV} &= \frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{A}\mathbf{V}_i \\
\widehat{\mathbf{Q}}_{\mathbf{A},V\Gamma} &= \frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{A}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i \\
\widehat{\mathbf{Q}}_{\mathbf{A},\Gamma\Gamma} &= \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'\mathbf{A}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i
\end{aligned}
$$

Next, similar arguments yield for the numerator

$$
\widehat{\mathbf{q}} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i]'\mathbf{M}_{\widehat{\mathbf{F}}}[\varepsilon_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y] = \widehat{\mathbf{q}}_{\mathbf{I}} - \widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}} - \widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]} \qquad (2.24)
$$

where for a given subscript $\mathbf{A}$ the respective terms are decomposed as

$$\begin{aligned}
\widehat{\mathbf{q}}_{\mathbf{A}} &= \widehat{\mathbf{q}}_{\mathbf{A},V\varepsilon} - \widehat{\mathbf{q}}_{\mathbf{A},V\eta} - \widehat{\mathbf{q}}_{\mathbf{A},\Gamma\varepsilon} + \widehat{\mathbf{q}}_{\mathbf{A},\Gamma\eta} \\
\widehat{\mathbf{q}}_{\mathbf{A},V\varepsilon} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{V}_i' \mathbf{A} \boldsymbol{\varepsilon}_i \\
\widehat{\mathbf{q}}_{\mathbf{A},V\eta} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{V}_i' \mathbf{A} \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger} \widetilde{\boldsymbol{\eta}}_i \mathbf{q}_y \\
\widehat{\mathbf{q}}_{\mathbf{A},\Gamma\varepsilon} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \boldsymbol{\Gamma}_i' (\overline{\mathbf{C}}^{\dagger})' \overline{\mathbf{U}}' \mathbf{A} \boldsymbol{\varepsilon}_i \\
\widehat{\mathbf{q}}_{\mathbf{A},\Gamma\eta} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \boldsymbol{\Gamma}_i' (\overline{\mathbf{C}}^{\dagger})' \overline{\mathbf{U}}' \mathbf{A} \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger} \widetilde{\boldsymbol{\eta}}_i \mathbf{q}_y
\end{aligned}$$

2.2.1 Lemmas

Lemma B-3 *Under Ass.1-5 we have as* $(N,T) \to \infty$ *such that* $T/N = \tau_{N,T} \to \tau < \infty$ *that*

$$\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]} = \sqrt{\tau}(\mathbf{d}_1 + \mathbf{d}_2) + O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{2.25}$$

where $\mathbf{d}_1 = \mathbf{d}_2 = \mathbf{0}_{k\times 1}$ *provided* $m = 1 + k$*, whereas in case* $m < 1 + k$

$$\mathbf{d}_1 = \mathbf{q}_{xy}' \boldsymbol{\Sigma}_{\boldsymbol{\eta}}' vec\left((\mathbf{C}^{\dagger})' \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{D}_{-m} \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{C}^{\dagger} \right) \tag{2.26}$$

$$\mathbf{d}_2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Sigma}_i [\boldsymbol{\beta}, \mathbf{I}_k] \mathbf{D}_{-m} [\sigma_i^2, \mathbf{0}_{1\times k}]' - \boldsymbol{\Gamma}' (\mathbf{C}^{\dagger})' \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{D}_{-m} [\sigma^2, \mathbf{0}_{1\times k}]' \tag{2.27}$$

and with $\mathbf{q}_{xy} = (\mathbf{q}_y \otimes \mathbf{q}_x)$*,* $\mathbf{q}_y = [1, \mathbf{0}_{k\times 1}']'$*,* $\mathbf{q}_x = [\mathbf{0}_{k\times 1}, \mathbf{I}_k]'$ *and* $\mathbf{D}_{-m} = \mathbf{T}\mathbf{H}_{-m} \boldsymbol{\Sigma}_{\mathbf{u}_{-m}^0}^{\dagger} \mathbf{H}_{-m}' \mathbf{T}'$*.*

Proof of Lemma B-3

Recall that $\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]} = \widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon} - \widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}],V\eta} - \widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\varepsilon} + \widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\eta}$. We first consider

$$\begin{aligned}
\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\eta} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \boldsymbol{\Gamma}_i' (\overline{\mathbf{C}}^{\dagger})' \overline{\mathbf{U}}' [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}] \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger} \widetilde{\boldsymbol{\eta}}_i \mathbf{q}_y \\
&= \sqrt{\frac{T}{N}} \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Gamma}_i' (\overline{\mathbf{C}}^{\dagger})' NT^{-1} \overline{\mathbf{U}}' [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}] \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger} \widetilde{\boldsymbol{\eta}}_i \mathbf{q}_y
\end{aligned}$$

Inserting (2.11) in $\overline{\mathbf{U}}' [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}] \overline{\mathbf{U}}$ gives

$$\begin{aligned}
NT^{-1} \overline{\mathbf{U}}' [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}] \overline{\mathbf{U}} &= NT^{-1} \overline{\mathbf{U}}' \overline{\mathbf{U}}_{-m}^0 \widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0}^{\dagger} T^{-1} (\overline{\mathbf{U}}_{-m}^0)' \overline{\mathbf{U}} + NT^{-1} \overline{\mathbf{U}}' \overline{\mathbf{U}}_m^0 \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger} T^{-1} (\overline{\mathbf{U}}_m^0)' \overline{\mathbf{U}} \\
&\quad + NT^{-1} \overline{\mathbf{U}}' \mathbf{F} \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger} T^{-1} (\overline{\mathbf{U}}_m^0)' \overline{\mathbf{U}} + T^{-1} \overline{\mathbf{U}}' \overline{\mathbf{U}}_m^0 \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger} T^{-1} \mathbf{F}' \overline{\mathbf{U}} \\
&\quad + NT^{-1} \overline{\mathbf{U}}' \widehat{\mathbf{F}}^0 \left[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0}^{\dagger} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u}^{\dagger} \right] T^{-1} (\widehat{\mathbf{F}}^0)' \overline{\mathbf{U}} \\
&= NT^{-1} \overline{\mathbf{U}}' \overline{\mathbf{U}}_{-m}^0 \widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0}^{\dagger} T^{-1} (\overline{\mathbf{U}}_{-m}^0)' \overline{\mathbf{U}} + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{T} \mathbf{H}_{-m} \boldsymbol{\Sigma}_{\mathbf{u}_{-m}^0}^{\dagger} \mathbf{H}_{-m}' \mathbf{T}' \boldsymbol{\Sigma}_{\mathbf{u}} + O_p(N^{-1/2}) + O_p(T^{-1/2})
\end{aligned} \tag{2.28}$$

because by Ass.1 and $\overline{\mathbf{H}}_{-m} = \mathbf{H}_{-m} + O_p(N^{-1/2})$ we have

$$\sqrt{N}T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_{-m} = \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\mathbf{U}_i'\mathbf{U}_j}{T}\right)\mathbf{T}\overline{\mathbf{H}}_{-m} = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Sigma}_{\mathbf{u},i}\right)\mathbf{T}\overline{\mathbf{H}}_{-m} + O_p(T^{-1/2})$$
$$= \mathbf{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m} + O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{2.29}$$

which implies together with (2.14) that the first term is

$$NT^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_{-m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{-m}}T^{-1}(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}} = \sqrt{N}T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_{-m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{-m}}\sqrt{N}T^{-1}(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}}$$
$$= \mathbf{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\mathbf{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'\mathbf{\Sigma}_{\mathbf{u}} + O_p(N^{-1/2}) + O_p(T^{-1/2})$$

whereas for the other terms, making use of $\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}} = O_p(1)$ and the orders established in Lemmas B-1 and B-2 gives

$$\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_m\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_m)'\overline{\mathbf{U}}\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_m\right\|^2\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}\right\| = O_p(N^{-2})$$
$$\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_m)'\overline{\mathbf{U}}\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}\right\|\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}\right\|\left\|T^{-1}(\overline{\mathbf{U}}^0_m)'\overline{\mathbf{U}}\right\| = O_p(T^{-1/2}N^{-3/2})$$
$$\left\|T^{-1}\overline{\mathbf{U}}'\widehat{\mathbf{F}}^0\left[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right]T^{-1}(\widehat{\mathbf{F}}^0)'\overline{\mathbf{U}}\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}'\widehat{\mathbf{F}}^0\right\|^2\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right\| = O_p(N^{-3/2}) + O_p(N^{-1}T^{-1/2})$$

Hence, substituting in (2.28), using $\tau_{N,T} = T/N = O(1)$ and $\text{vec}\,(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\text{vec}\,(\mathbf{B})$ (Abadir and Magnus, 2005, Exercise 10.18) yield

$$\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\eta}$$
$$= \sqrt{\frac{T}{N}}\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'NT^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y$$
$$= \sqrt{\tau_{N,T}}\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\mathbf{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\mathbf{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'\mathbf{\Sigma}_{\mathbf{u}}\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y + O_p(N^{-1/2}) + O_p(T^{-1/2})$$
$$= \sqrt{\tau_{N,T}}\left(\frac{1}{N}\sum_{i=1}^{N}(\mathbf{q}_y'\widetilde{\boldsymbol{\eta}}_i' \otimes \mathbf{\Gamma}_i')\right)vec\left((\overline{\mathbf{C}}^{\dagger})'\mathbf{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\mathbf{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'\mathbf{\Sigma}_{\mathbf{u}}\overline{\mathbf{C}}^{\dagger}\right) + O_p(N^{-1/2}) + O_p(T^{-1/2})$$
$$= \sqrt{\tau_{N,T}}\left(\frac{1}{N}\sum_{i=1}^{N}(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y \otimes \mathbf{\Gamma}_i)'\right)vec\left((\overline{\mathbf{C}}^{\dagger})'\mathbf{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\mathbf{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'\mathbf{\Sigma}_{\mathbf{u}}\overline{\mathbf{C}}^{\dagger}\right) + O_p(N^{-1/2}) + O_p(T^{-1/2})$$
$$= \sqrt{\tau}\mathbf{q}_{xy}'\mathbf{\Sigma}_{\eta}'vec\left((\mathbf{C}^{\dagger})'\mathbf{\Sigma}_{\mathbf{u}}\mathbf{D}_{-m}\mathbf{\Sigma}_{\mathbf{u}}\mathbf{C}^{\dagger}\right) + O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{2.30}$$

with $\mathbf{D}_{-m} = \mathbf{T}\mathbf{H}_{-m}\mathbf{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'$, $\mathbf{\Sigma}_{\boldsymbol{\eta}} = \mathbb{E}(\widetilde{\boldsymbol{\eta}}_i \otimes \widetilde{\boldsymbol{\eta}}_i)$ and $\mathbf{q}_{xy} = (\mathbf{q}_y \otimes \mathbf{q}_x)$, and where we also made use of the following facts: $\tau_{N,T} \to \tau < \infty$, and from Ass.3 that $\overline{\mathbf{C}}^{\dagger} = \mathbf{C}^{\dagger} + O_p(N^{-1/2})$, $\frac{1}{N}\sum_{i=1}^{N}\widetilde{\boldsymbol{\eta}}_i'\mathbf{\Gamma} = \overline{\widetilde{\boldsymbol{\eta}}}'\mathbf{\Gamma} = O_p(N^{-1/2})$ and $\frac{1}{N}\sum_{i=1}^{N}(\widetilde{\boldsymbol{\eta}}_i \otimes \widetilde{\boldsymbol{\eta}}_i) = \mathbf{\Sigma}_{\boldsymbol{\eta}} + O_p(N^{-1/2})$, which lead to

$$\frac{1}{N}\sum_{i=1}^{N}(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y \otimes \mathbf{\Gamma}_i) = \frac{1}{N}\sum_{i=1}^{N}(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y \otimes \mathbf{\Gamma} + \widetilde{\boldsymbol{\eta}}_i\mathbf{q}_x) = \frac{1}{N}\sum_{i=1}^{N}(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y \otimes \widetilde{\boldsymbol{\eta}}_i\mathbf{q}_x) + O_p(N^{-1/2})$$
$$= \frac{1}{N}\sum_{i=1}^{N}(\widetilde{\boldsymbol{\eta}}_i \otimes \widetilde{\boldsymbol{\eta}}_i)(\mathbf{q}_y \otimes \mathbf{q}_x) + O_p(N^{-1/2})$$
$$= \mathbf{\Sigma}_{\boldsymbol{\eta}}\mathbf{q}_{xy} + O_p(N^{-1/2}) \tag{2.31}$$

Next up is $\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\varepsilon} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{\Gamma}_i'(\overline{\mathbf{C}}^\dagger)'\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\boldsymbol{\varepsilon}_i$. Making use of $\boldsymbol{\varepsilon}_i = \mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$ and with (2.11) follows the decomposition

$$\begin{aligned}
&T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\boldsymbol{\varepsilon}_i \\
&= T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_{-m}\widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{u}^0_{-m}}T^{-1}(\overline{\mathbf{U}}^0_{-m})'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y + T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_m\widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_m)'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y \\
&\quad + T^{-1}\overline{\mathbf{U}}'\mathbf{F}\widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_m)'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y + T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_m\widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{F}}T^{-1}\mathbf{F}'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y \\
&\quad + T^{-1}\overline{\mathbf{U}}'\widehat{\mathbf{F}}^0\left[\widehat{\mathbf{\Sigma}}^\dagger_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{F}_u}\right]T^{-1}(\widehat{\mathbf{F}}^0)'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y \\
&= T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_{-m}\widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{u}^0_{-m}}T^{-1}(\overline{\mathbf{U}}^0_{-m})'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y + O_p(N^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2})
\end{aligned} \tag{2.32}$$

because substituting in results from Lemma B-1 and B-2 gives

$$\begin{aligned}
\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_m\widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_m)'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right\| &\le \left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_m\right\|\left\|\widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{F}}\right\|\left\|T^{-1}(\overline{\mathbf{U}}^0_m)'\mathbf{U}_i\right\|\left\|\mathbf{B}^{-1}\mathbf{q}_y\right\| \\
&= O_p(N^{-2}) + O_p(N^{-3/2}T^{-1/2}) \\
\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}\widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_m)'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right\| &\le \left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}\right\|\left\|\widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{F}}\right\|\left\|T^{-1}(\overline{\mathbf{U}}^0_m)'\mathbf{U}_i\right\|\left\|\mathbf{B}^{-1}\mathbf{q}_y\right\| \\
&= O_p(N^{-3/2}T^{-1/2}) + O_p((NT)^{-1}) \\
\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_m\widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{F}}T^{-1}\mathbf{F}'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right\| &\le \left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_m\right\|\left\|\widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{F}}\right\|\left\|T^{-1}\mathbf{F}'\mathbf{U}_i\right\|\left\|\mathbf{B}^{-1}\mathbf{q}_y\right\| = O_p(N^{-1}T^{-1/2}) \\
\left\|T^{-1}\overline{\mathbf{U}}'\widehat{\mathbf{F}}^0\left[\widehat{\mathbf{\Sigma}}^\dagger_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{F}_u}\right]T^{-1}(\widehat{\mathbf{F}}^0)'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right\| &\le \left\|T^{-1}\overline{\mathbf{U}}'\widehat{\mathbf{F}}^0\right\|\left\|\widehat{\mathbf{\Sigma}}^\dagger_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{F}_u}\right\|\left\|T^{-1}(\widehat{\mathbf{F}}^0)'\mathbf{U}_i\right\|\left\|\mathbf{B}^{-1}\mathbf{q}_y\right\| \\
&= O_p(N^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2})
\end{aligned}$$

and we note that

$$\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_{-m}\widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{u}^0_{-m}}T^{-1}(\overline{\mathbf{U}}^0_{-m})'\mathbf{U}_i\right\| = O_p(N^{-1}) + O_p((NT)^{-1/2}) \tag{2.33}$$

Making use of (2.32) and $T/N = O(1)$, scaling it by $\sqrt{NT}$ gives

$$\begin{aligned}
\sqrt{N}T^{-1/2}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\boldsymbol{\varepsilon}_i &= T^{-1}\sqrt{N}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_{-m}\widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{u}^0_{-m}}T^{-1/2}(\overline{\mathbf{U}}^0_{-m})'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= \mathbf{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\mathbf{\Sigma}^\dagger_{\mathbf{u}^0_{-m}}T^{-1/2}(\overline{\mathbf{U}}^0_{-m})'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= \mathbf{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\mathbf{\Sigma}^\dagger_{\mathbf{u}^0_{-m}}\overline{\mathbf{H}}'_{-m}\mathbf{T}'\sqrt{N}T^{-1/2}\overline{\mathbf{U}}'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y + O_p(N^{-1/2}) + O_p(T^{-1/2})
\end{aligned}$$

where in the second line we have substituted in (2.14) and (2.29) and in the final line $\overline{\mathbf{U}}^0_{-m} = \sqrt{N}\overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_{-m}$. Then

$$\begin{aligned}
\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\varepsilon} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{\Gamma}_i'(\overline{\mathbf{C}}^\dagger)'\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\boldsymbol{\varepsilon}_i \\
&= \frac{1}{N}\sum_{i=1}^N \mathbf{\Gamma}_i'(\overline{\mathbf{C}}^\dagger)'\sqrt{N}T^{-1/2}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\boldsymbol{\varepsilon}_i \\
&= \frac{1}{N}\sum_{i=1}^N \mathbf{\Gamma}_i'(\overline{\mathbf{C}}^\dagger)'\mathbf{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\mathbf{\Sigma}^\dagger_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'\sqrt{N}T^{-1/2}\overline{\mathbf{U}}'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= \frac{1}{N}\sum_{i=1}^N \mathbf{\Gamma}_i'\mathbf{D}\sqrt{N}T^{-1/2}\overline{\mathbf{U}}'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y + O_p(N^{-1/2}) + O_p(T^{-1/2})
\end{aligned}$$

where we have defined $\mathbf{D} = (\overline{\mathbf{C}}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'$. Noting that $\overline{\mathbf{U}} = N^{-1}\mathbf{U}_i + \frac{1}{N}\sum_{j\neq i}^{N}\mathbf{U}_j$, the remaining term can be written as

$$
\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\Gamma}'_i\mathbf{D}\overline{\mathbf{U}}'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y &= \sqrt{\frac{T}{N}}\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Gamma}'_i\mathbf{D}\mathbf{U}'_i\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y + \sqrt{\frac{T}{N}}\frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\boldsymbol{\Gamma}'_i\mathbf{D}\mathbf{U}'_j\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y \\
&= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}'_i\mathbf{D}(T^{-1}\mathbf{U}'_i\mathbf{U}_i)\mathbf{B}^{-1}\mathbf{q}_y + \frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\boldsymbol{\Gamma}'_i\mathbf{D}\mathbf{U}'_j\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right] \\
&= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}'_i\mathbf{D}[\sigma_i^2,\mathbf{0}_{1\times k}]'\right] + O_p(T^{-1/2}) \\
&= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}\left([\sigma_i^2,\mathbf{0}_{1\times k}]\otimes\boldsymbol{\Gamma}'_i\right)\right]vec(\mathbf{D}) + O_p(T^{-1/2}) \\
&= \sqrt{\tau}\left([\sigma^2,\mathbf{0}_{1\times k}]\otimes\boldsymbol{\Gamma}'\right)vec((\mathbf{C}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}') + O_p(N^{-1/2}) + O_p(T^{-1/2})
\end{aligned}
$$

because $T^{-1}\mathbf{U}'_i\mathbf{U}_i = \boldsymbol{\Sigma}_{\mathbf{u},i} + O_p(T^{-1/2})$ from Ass.1, $\boldsymbol{\Sigma}_{\mathbf{u},i}\mathbf{B}^{-1}\mathbf{q}_y = [\sigma_i^2,\mathbf{0}_{1\times k}]'$ and $\frac{1}{N}\sum_{i=1}^{N}\left([\sigma_i^2,\mathbf{0}_{1\times k}]\otimes\boldsymbol{\Gamma}'_i\right) \to^p \left([\sigma^2,\mathbf{0}_{1\times k}]\otimes\boldsymbol{\Gamma}'\right)$ by Ass.1 and 3. We also have that

$$
\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\boldsymbol{\Gamma}'_i\mathbf{D}\mathbf{U}'_j\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right\| = O_p\left(\frac{1}{\sqrt{T}}\right) \tag{2.34}
$$

due to

$$
\mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\boldsymbol{\Gamma}'_i\mathbf{D}\mathbf{U}'_j\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right] = \mathbf{0}_{k\times 1}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\boldsymbol{\Gamma}'_i\mathbf{D}\mathbf{U}'_j\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right]\left[\frac{1}{NT}\sum_{k=1}^{N}\sum_{l\neq k}^{N}\boldsymbol{\Gamma}'_k\mathbf{D}\mathbf{U}'_l\mathbf{U}_k\mathbf{B}^{-1}\mathbf{q}_y\right]' \\
&= \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k=1}^{N}\sum_{l\neq k}^{N}\mathbb{E}\left[\boldsymbol{\Gamma}'_i\mathbf{D}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}[\mathbf{u}_{j,t}\mathbf{u}'_{i,t}\mathbf{B}^{-1}\mathbf{q}_y\mathbf{q}'_y(\mathbf{B}^{-1})'\mathbf{u}_{k,s}\mathbf{u}'_{l,s}]\right)\mathbf{D}'\boldsymbol{\Gamma}_k\right] \\
&= \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbb{E}\left[\boldsymbol{\Gamma}'_i\mathbf{D}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}[\mathbf{u}_{j,t}\mathbf{u}'_{i,t}\mathbf{B}^{-1}\mathbf{q}_y\mathbf{q}'_y(\mathbf{B}^{-1})'\mathbf{u}_{i,s}\mathbf{u}'_{j,s}]\right)\mathbf{D}'\boldsymbol{\Gamma}_i\right] \\
&\quad + \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbb{E}\left[\boldsymbol{\Gamma}'_i\mathbf{D}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}[\mathbf{u}_{j,t}\mathbf{u}'_{i,t}\mathbf{B}^{-1}\mathbf{q}_y\mathbf{q}'_y(\mathbf{B}^{-1})'\mathbf{u}_{j,s}\mathbf{u}'_{i,s}]\right)\mathbf{D}'\boldsymbol{\Gamma}_j\right] \\
&= \frac{1}{(NT)^2}O\left(N^2T\right) = O\left(\frac{1}{T}\right)
\end{aligned}
$$

where the second equality follows from noting that under Ass.1 (independence over cross-sections) the inner expectation of errors is only non-zero when $(k=i, l=j)$ or $(k=j, l=i)$ (because $i\neq j$ and $l\neq k$). The final order result follows from (2.20). Therefore, combining the results lead to the conclusion

$$
\begin{aligned}
\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\varepsilon} &= \sqrt{\tau_{N,T}}\left([\sigma^2,\mathbf{0}_{1\times k}]\otimes\boldsymbol{\Gamma}'\right)vec((\mathbf{C}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}') + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= \sqrt{\tau_{N,T}}\boldsymbol{\Gamma}'(\mathbf{C}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'[\sigma^2,\mathbf{0}_{1\times k}]' + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= \sqrt{\tau}\boldsymbol{\Gamma}'(\mathbf{C}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{D}_{-m}[\sigma^2,\mathbf{0}_{1\times k}]' + O_p(N^{-1/2}) + O_p(T^{-1/2})
\end{aligned} \tag{2.35}
$$

where $\mathbf{D}_{-m} = \mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'$.

The analysis of $\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\eta}$ is near identical to that of $\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\varepsilon}$. Noting that $\mathbf{V}_i = \mathbf{U}_i\mathbf{q}_x$, we obtain

$$\begin{aligned}\sqrt{N}T^{-1/2}\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}} &= T^{-1/2}\mathbf{q}_x'\mathbf{U}_i'\overline{\mathbf{U}}^0_{-m}\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{u}^0_{-m}}T^{-1}\sqrt{N}(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}} + O_p(N^{-1/2}) + O_p(T^{-1/2})\\ &= T^{-1/2}\sqrt{N}\mathbf{q}_x'\mathbf{U}_i'\overline{\mathbf{U}}\mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'\boldsymbol{\Sigma}_{\mathbf{u}} + O_p(N^{-1/2}) + O_p(T^{-1/2})\end{aligned}$$

because from $\mathbf{V}_i = \mathbf{U}_i\mathbf{q}_x$ the terms in this decomposition contain the exact same variables that drive the orders as for $T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\boldsymbol{\varepsilon}_i = T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$ in (2.32) and we again substituted in (2.14) and (2.29) and $\overline{\mathbf{U}}^0_{-m} = \sqrt{N}\overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_{-m}$ on the final line. Then

$$\begin{aligned}\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\eta} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\\ &= \frac{1}{N}\sum_{i=1}^{N}\sqrt{N}T^{-1/2}\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\\ &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{q}_x'\mathbf{U}_i'\overline{\mathbf{U}}\mathbf{D}'\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y + O_p(N^{-1/2}) + O_p(T^{-1/2})\end{aligned}$$

where we again made use of $\mathbf{D} = (\overline{\mathbf{C}}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'$ and obtain using analogous arguments as above

$$\begin{aligned}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{q}_x'\mathbf{U}_i'\overline{\mathbf{U}}\mathbf{D}'\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y &= \sqrt{\tau_{N,T}}\left[\frac{1}{NT}\sum_{i=1}^{N}\mathbf{q}_x'(T^{-1}\mathbf{U}_i'\mathbf{U}_i)\mathbf{D}'\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y + \frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbf{q}_x'\mathbf{U}_i'\mathbf{U}_j\mathbf{D}'\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\right]\\ &= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{q}_x'\boldsymbol{\Sigma}_{\mathbf{u},i}\mathbf{D}'\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\right] + O_p(T^{-1/2})\\ &= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{q}_y'\widetilde{\boldsymbol{\eta}}_i'\otimes\mathbf{q}_x'\boldsymbol{\Sigma}_{\mathbf{u},i}\right]vec(\mathbf{D}') + O_p(T^{-1/2})\\ &= O_p(N^{-1/2}) + O_p(T^{-1/2})\end{aligned}$$

because by Ass.3

$$\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{q}_y'\widetilde{\boldsymbol{\eta}}_i'\otimes\mathbf{q}_x'\boldsymbol{\Sigma}_{\mathbf{u},i}\right] = (\mathbf{q}_y'\otimes\mathbf{q}_x')\frac{1}{N}\sum_{i=1}^{N}\left[\widetilde{\boldsymbol{\eta}}_i'\otimes\boldsymbol{\Sigma}_{\mathbf{u},i}\right] = O_p(N^{-1/2})$$

and where

$$\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbf{q}_x'\mathbf{U}_i'\mathbf{U}_j\mathbf{D}'\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\right\| = O_p\left(\frac{1}{\sqrt{T}}\right) \tag{2.36}$$

from its zero expectation and

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbf{q}_x'\mathbf{U}_i'\mathbf{U}_j\mathbf{D}'\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\right]\left[\frac{1}{NT}\sum_{k=1}^{N}\sum_{l\neq k}^{N}\mathbf{q}_x'\mathbf{U}_k'\mathbf{U}_l\mathbf{D}'\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\right]'\\
&=\frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k=1}^{N}\sum_{l\neq k}^{N}\mathbf{q}_x'\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[\mathbf{u}_{i,t}\mathbf{u}_{j,t}'\mathbf{D}'\mathbb{E}(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\mathbf{q}_y'\widetilde{\boldsymbol{\eta}}_k')\mathbf{D}\mathbf{u}_{l,s}\mathbf{u}_{k,s}'\right]\mathbf{q}_x\\
&=\frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{l\neq i}^{N}\mathbf{q}_x'\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[\mathbf{u}_{i,t}\mathbf{u}_{j,t}'\mathbf{D}'\mathbb{E}(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\mathbf{q}_y'\widetilde{\boldsymbol{\eta}}_i')\mathbf{D}\mathbf{u}_{l,s}\mathbf{u}_{i,s}'\right]\mathbf{q}_x\\
&=\frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{l\neq i}^{N}\mathbf{q}_x'\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[\mathbf{u}_{i,t}\mathbb{E}\left[\mathbf{u}_{j,t}'\mathbf{D}'\mathbb{E}(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\mathbf{q}_y'\widetilde{\boldsymbol{\eta}}_i')\mathbf{D}\mathbf{u}_{l,s}|\mathbf{U}_i\right]\mathbf{u}_{i,s}'\right]\mathbf{q}_x\\
&=\frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbf{q}_x'\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[\mathbf{u}_{i,t}\mathbf{u}_{j,t}'\mathbf{D}'\mathbb{E}(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\mathbf{q}_y'\widetilde{\boldsymbol{\eta}}_i')\mathbf{D}\mathbf{u}_{j,s}\mathbf{u}_{i,s}'\right]\mathbf{q}_x\\
&=\frac{1}{(NT)^2}O\left(N^2T\right)=O\left(\frac{1}{T}\right)
\end{aligned}$$

since $\mathbb{E}(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\mathbf{q}_y'\widetilde{\boldsymbol{\eta}}_k')=\mathbf{0}$ when $k\neq i$ and one sum in the fourth line is eliminated by noticing zero expectation when $l\neq j$. Otherwise, we use arguments identical to (2.20). We thus conclude that

$$\left\|\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\eta}\right\|=O_p(N^{-1/2})+O_p(T^{-1/2}) \tag{2.37}$$

Last up is

$$\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon}=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\boldsymbol{\varepsilon}_i=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{q}_x'\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$$

Making use of results in Lemma B-2 yields

$$\begin{aligned}
&\left\|T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_{-m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{-m}^0)'\mathbf{U}_i\right\|\leq\left\|T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_{-m}^0\right\|^2\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0}^{\dagger}\right\|=O_p(N^{-1})+O_p(T^{-1})+O_p((NT)^{-1/2})\\
&\left\|T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_m^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}(\overline{\mathbf{U}}_m^0)'\mathbf{U}_i\right\|\leq\left\|T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_m^0\right\|^2\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}\right\|=O_p(N^{-2})+O_p(N^{-3/2}T^{-1/2})+O_p((NT)^{-1})\\
&\left\|T^{-1}\mathbf{U}_i'\mathbf{F}\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}(\overline{\mathbf{U}}_m^0)'\mathbf{U}_i\right\|\leq\left\|T^{-1}\mathbf{U}_i'\mathbf{F}\right\|\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}\right\|\left\|T^{-1}(\overline{\mathbf{U}}_m^0)'\mathbf{U}_i\right\|=O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})\\
&\left\|T^{-1}\mathbf{U}_i'\widehat{\mathbf{F}}^0\left[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0}^{\dagger}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u}^{\dagger}\right]T^{-1}(\widehat{\mathbf{F}}^0)'\mathbf{U}_i\right\|\leq\left\|T^{-1}\mathbf{U}_i'\widehat{\mathbf{F}}^0\right\|^2\left\|\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0}^{\dagger}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u}^{\dagger}\right\|\\
&\qquad=O_p(N^{-3/2})+O_p(N^{-1/2}T^{-1})+O_p(N^{-1}T^{-1/2})+O_p(T^{-3/2})
\end{aligned}$$

so that we obtain for the familiar decomposition

$$\begin{aligned}
&T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\\
&=T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_{-m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{-m}^0)'\mathbf{U}_i+T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_m^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}(\overline{\mathbf{U}}_m^0)'\mathbf{U}_i\\
&+T^{-1}\mathbf{U}_i'\mathbf{F}\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}(\overline{\mathbf{U}}_m^0)'\mathbf{U}_i+T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_m^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}\mathbf{F}'\mathbf{U}_i+T^{-1}\mathbf{U}_i'\widehat{\mathbf{F}}^0\left[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0}^{\dagger}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u}^{\dagger}\right]T^{-1}(\widehat{\mathbf{F}}^0)'\mathbf{U}_i\\
&=T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_{-m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{-m}^0)'\mathbf{U}_i+O_p(N^{-3/2})+O_p(N^{-1/2}T^{-1})+O_p(N^{-1}T^{-1/2})+O_p(T^{-3/2})\\
&=NT^{-2}\mathbf{U}_i'\overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_{-m}\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0}^{\dagger}\overline{\mathbf{H}}_{-m}'\mathbf{T}'\overline{\mathbf{U}}'\mathbf{U}_i+O_p(N^{-3/2})+O_p(N^{-1/2}T^{-1})+O_p(N^{-1}T^{-1/2})+O_p(T^{-3/2})
\end{aligned} \tag{2.38}$$

Substituting in (2.38) and making use of $T/N = O(1)$ gives

$$\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{q}_x'\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y = \sqrt{\frac{T}{N}}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{q}_x' NT^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right]$$

$$= \sqrt{\tau_{N,T}}\left[\frac{N}{T^2}\sum_{i=1}^{N}\mathbf{q}_x'\mathbf{U}_i'\overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_{-m}\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{u}^0_{-m}}\overline{\mathbf{H}}'_{-m}\mathbf{T}'\overline{\mathbf{U}}'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right] + O_p(N^{-1/2}) + O_p(T^{-1/2})$$

$$= \sqrt{\tau_{N,T}}\left[\frac{N}{T^2}\sum_{i=1}^{N}\mathbf{q}_x'\mathbf{U}_i'\overline{\mathbf{U}}\mathbf{D}_{-m}\overline{\mathbf{U}}'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right] + O_p(N^{-1/2}) + O_p(T^{-1/2})$$

$$= \sqrt{\tau_{N,T}}\left[\frac{N}{T^2}\sum_{i=1}^{N}\mathbf{q}_x'\mathbf{U}_i'\left(\frac{1}{N}\sum_{j=1}^{N}\mathbf{U}_j\right)\mathbf{D}_{-m}\left(\frac{1}{N}\sum_{l=1}^{N}\mathbf{U}_l\right)'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right] + O_p(N^{-1/2}) + O_p(T^{-1/2})$$

$$= \sqrt{\tau_{N,T}}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\mathbf{V}_i'\mathbf{U}_j\mathbf{D}_{-m}\mathbf{U}_l'\boldsymbol{\varepsilon}_i\right] + O_p(N^{-1/2}) + O_p(T^{-1/2})$$

where to go to the third line we have made use of (2.14), $\overline{\mathbf{H}}_{-m} = \mathbf{H}_{-m} + O_p(N^{-1/2})$ and have defined $\mathbf{D}_{-m} = \mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'$.

Next, to study the remainder, let $d_{v,g}$ denote the element on row v and column g of $\mathbf{D}_{-m}$ and let $\mathbf{q}_a$ be a $k\times 1$ vector of zeros with a 1 on its $a-th$ row. Then $\mathbf{U}_i^{(a)} = \mathbf{U}_i\mathbf{q}_a$ ($\mathbf{V}_i^{(a)} = \mathbf{V}_i\mathbf{q}_a$) is the $a-th$ column of $\mathbf{U}_i$ ($\mathbf{V}_i$) and $\mathrm{v}_{i,t}^{(a)} = \mathbf{v}_{i,t}'\mathbf{q}_a$ denotes the $a-th$ row of $\mathbf{v}_{it}$. Note that if $g=1$ then by definition $\mathbf{U}_j^{(1)} = \boldsymbol{\varepsilon}_j + \mathbf{V}_j\boldsymbol{\beta}$ whereas for $g>1$ we have $\mathbf{U}_j^{(g)} = \mathbf{V}_j^{(g-1)}$. Then we can unpack the remaining expression between brackets as follows

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\mathbf{V}_i'\mathbf{U}_j\mathbf{D}_{-m}\mathbf{U}_l'\boldsymbol{\varepsilon}_i$$

$$= \frac{1}{NT^2}\sum_{v=1}^{1+k}\sum_{g=1}^{1+k}d_{v,g}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\mathbf{V}_i'\mathbf{U}_j^{(v)}\mathbf{U}_l^{(g)\prime}\boldsymbol{\varepsilon}_i$$

$$= \frac{1}{NT^2}\sum_{v=1}^{1+k}d_{v,1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\mathbf{V}_i'\mathbf{U}_j^{(v)}[\boldsymbol{\varepsilon}_l + \mathbf{V}_l\boldsymbol{\beta}]'\boldsymbol{\varepsilon}_i + \frac{1}{NT^2}\sum_{v=1}^{1+k}\sum_{g=2}^{1+k}d_{v,g}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\mathbf{V}_i'\mathbf{U}_j^{(v)}\mathbf{V}_l^{(g-1)\prime}\boldsymbol{\varepsilon}_i$$

and now also unpacking the first $\mathbf{U}^{(v)}$ with

$$
\begin{aligned}
&= \frac{1}{NT^2} d_{1,1} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} \mathbf{V}_i'[\boldsymbol{\varepsilon}_j + \mathbf{V}_j\boldsymbol{\beta}][\boldsymbol{\varepsilon}_l + \mathbf{V}_l\boldsymbol{\beta}]'\boldsymbol{\varepsilon}_i + \frac{1}{NT^2}\sum_{v=2}^{1+k} d_{v,1} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} \mathbf{V}_i'\mathbf{V}_j^{(v-1)}[\boldsymbol{\varepsilon}_l + \mathbf{V}_l\boldsymbol{\beta}]'\boldsymbol{\varepsilon}_i \\
&+ \frac{1}{NT^2}\sum_{g=2}^{1+k} d_{1,g} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} \mathbf{V}_i'[\boldsymbol{\varepsilon}_j + \mathbf{V}_j\boldsymbol{\beta}]\mathbf{V}_l^{(g-1)\prime}\boldsymbol{\varepsilon}_i + \frac{1}{NT^2}\sum_{v=2}^{1+k}\sum_{g=2}^{1+k} d_{v,g} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} \mathbf{V}_i'\mathbf{V}_j^{(v-1)}\mathbf{V}_l^{(g-1)\prime}\boldsymbol{\varepsilon}_i \\
&= \frac{1}{NT^2} d_{1,1} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} \left[\mathbf{V}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_l'\boldsymbol{\varepsilon}_i + \mathbf{V}_i'\boldsymbol{\varepsilon}_j\boldsymbol{\beta}'\mathbf{V}_l'\boldsymbol{\varepsilon}_i + \mathbf{V}_i'\mathbf{V}_j\boldsymbol{\beta}\boldsymbol{\varepsilon}_l'\boldsymbol{\varepsilon}_i + \mathbf{V}_i'\mathbf{V}_j\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}_l'\boldsymbol{\varepsilon}_i\right] \\
&+ \frac{1}{NT^2}\sum_{v=2}^{1+k} d_{v,1} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} \left[\mathbf{V}_i'\mathbf{V}_j^{(v-1)}\boldsymbol{\varepsilon}_l'\boldsymbol{\varepsilon}_i + \mathbf{V}_i'\mathbf{V}_j^{(v-1)}\boldsymbol{\beta}'\mathbf{V}_l'\boldsymbol{\varepsilon}_i\right] \\
&+ \frac{1}{NT^2}\sum_{g=2}^{1+k} d_{1,g} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} \left[\mathbf{V}_i'\boldsymbol{\varepsilon}_j\mathbf{V}_l^{(g-1)\prime}\boldsymbol{\varepsilon}_i + \mathbf{V}_i'\mathbf{V}_j\boldsymbol{\beta}\mathbf{V}_l^{(g-1)\prime}\boldsymbol{\varepsilon}_i\right] \\
&+ \frac{1}{NT^2}\sum_{v=2}^{1+k}\sum_{g=2}^{1+k} d_{v,g} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} \mathbf{V}_i'\mathbf{V}_j^{(v-1)}\mathbf{V}_l^{(g-1)\prime}\boldsymbol{\varepsilon}_i
\end{aligned}
$$

Next, also unpacking over time and bundling related terms yields

$$
\begin{aligned}
=& d_{1,1}\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\mathbf{v}_{i,t}\varepsilon_{j,t}\varepsilon_{l,s}\varepsilon_{i,s} + \mathbf{v}_{i,t}\mathbf{v}_{l,s}'\boldsymbol{\beta}\varepsilon_{j,t}\varepsilon_{i,s} + \mathbf{v}_{i,t}\mathbf{v}_{j,t}'\boldsymbol{\beta}\varepsilon_{l,s}\varepsilon_{i,s} + \mathbf{v}_{i,t}\mathbf{v}_{j,t}'\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{v}_{l,s}\varepsilon_{i,s}\right]\right\} \\
&+ \sum_{v=2}^{1+k} d_{v,1}\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\mathbf{v}_{i,t}\mathrm{v}_{j,t}^{(v-1)}\varepsilon_{l,s}\varepsilon_{i,s} + \mathbf{v}_{i,t}\mathrm{v}_{j,t}^{(v-1)}\mathbf{v}_{l,s}'\boldsymbol{\beta}\varepsilon_{i,s}\right]\right\} \\
&+ \sum_{g=2}^{1+k} d_{1,g}\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\mathbf{v}_{i,t}\mathrm{v}_{l,s}^{(g-1)}\varepsilon_{i,s}\varepsilon_{j,t} + \mathbf{v}_{i,t}\mathbf{v}_{j,t}'\boldsymbol{\beta}\mathrm{v}_{l,s}^{(g-1)}\varepsilon_{i,s}\right]\right\} \\
&+ \sum_{v=2}^{1+k}\sum_{g=2}^{1+k} d_{v,g}\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{v}_{i,t}\mathrm{v}_{j,t}^{(v-1)}\mathrm{v}_{l,s}^{(g-1)}\varepsilon_{i,s}\right\} \qquad (2.39)
\end{aligned}
$$

Consider then that we can write the second term in the first set of brackets as

$$
\begin{aligned}
&\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{v}_{i,t}\mathbf{v}_{l,s}'\boldsymbol{\beta}\varepsilon_{j,t}\varepsilon_{i,s} \\
&= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{\Sigma}_{i,t,s}\boldsymbol{\beta}\sigma_{i,t,s} + \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}(\mathbf{v}_{i,t}\mathbf{v}_{l,s}' - \mathbf{1}_{(l=i)}\boldsymbol{\Sigma}_{i,t,s})\boldsymbol{\beta}(\varepsilon_{j,t}\varepsilon_{i,s} - \mathbf{1}_{(j=i)}\sigma_{i,t,s}) \\
&= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{\Sigma}_{i,t,s}\boldsymbol{\beta}\sigma_{i,t,s} + \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{V}_{i,l}^{t,s}\boldsymbol{\beta}e_{i,j}^{t,s}
\end{aligned}
$$

where $\boldsymbol{\Sigma}_{i,t,s} = \mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}_{i,s}')$, $\sigma_{i,t,s} = \mathbb{E}(\boldsymbol{\varepsilon}_{i,t}\boldsymbol{\varepsilon}_{i,s})$, $\mathbf{V}_{i,l}^{t,s} = \mathbf{v}_{i,t}\mathbf{v}_{l,s}' - \mathbf{1}_{(l=i)}\boldsymbol{\Sigma}_{i,t,s}$ and $e_{i,j}^{t,s} = \varepsilon_{j,t}\varepsilon_{i,s} - \mathbf{1}_{(j=i)}\sigma_{i,t,s}$. The second term has expectation zero and we can write given independence of $\mathbf{V}_{i,l}^{t,s}$ and $e_{i,j}^{t,s}$

$$
\begin{aligned}
&\mathbb{E}\left(\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{V}_{i,l}^{t,s}\boldsymbol{\beta}e_{i,j}^{t,s}\right)\left(\frac{1}{NT^2}\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{o=1}^{N}\sum_{r=1}^{T}\sum_{q=1}^{T}\mathbf{V}_{m,o}^{r,q}\boldsymbol{\beta}e_{m,n}^{r,q}\right)' \\
&= \frac{1}{N^2T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{o=1}^{N}\mathbb{E}\left(\mathbf{V}_{i,l}^{t,s}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}_{m,o}^{r,q\prime}\right)\mathbb{E}\left(e_{i,j}^{t,s}e_{m,n}^{r,q}\right)
\end{aligned}
$$

As the analysis is highly tedious but features a repeating pattern, we shall sketch out the rationale. First, consider that either of these expectations is zero when one of the indices differs from all others. Then, the rightmost expectation is only non-zero if $(n = m = j = i)$, $(n = i, m = j, j \neq i)$ or $(n = j, m = i, j \neq i)$. (Note that $(n = m, j = i, i \neq m)$ equals zero since by definition $\mathbb{E}\left(e_{i,i}^{t,s} e_{m,m}^{r,q}\right) = 0$ when $i \neq m$). The $(n = m = j = i)$ case yields

$$\frac{1}{N^2 T^4} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{r=1}^{T} \sum_{q=1}^{T} \sum_{i=1}^{N} \sum_{l=1}^{N} \sum_{o=1}^{N} \mathbb{E}\left(\mathbf{V}_{i,l}^{t,s} \boldsymbol{\beta}\boldsymbol{\beta}' \mathbf{V}_{m,o}^{r,q\prime}\right) \mathbb{E}\left(e_{i,j}^{t,s} e_{m,n}^{r,q}\right)$$

for which the left expectation is zero unless also $(o = l = i)$ or $(o = l, i \neq l)$. These two cases give

$$(o = l, i \neq l) \qquad \frac{1}{N^2 T^4} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{r=1}^{T} \sum_{q=1}^{T} \sum_{i=1}^{N} \sum_{l=1}^{N} \mathbb{E}\left(\mathbf{V}_{i,l}^{t,s} \boldsymbol{\beta}\boldsymbol{\beta}' \mathbf{V}_{i,l}^{r,q\prime}\right) \mathbb{E}\left(e_{i,i}^{t,s} e_{i,i}^{r,q}\right) = \frac{1}{N^2 T^4} O\left(N^2 T^2\right) = O\left(\frac{1}{T^2}\right)$$

$$(o = l = i) \qquad \frac{1}{N^2 T^4} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{r=1}^{T} \sum_{q=1}^{T} \sum_{i=1}^{N} \mathbb{E}\left(\mathbf{V}_{i,i}^{t,s} \boldsymbol{\beta}\boldsymbol{\beta}' \mathbf{V}_{i,i}^{r,q\prime}\right) \mathbb{E}\left(e_{i,i}^{t,s} e_{i,i}^{r,q}\right) = \frac{1}{N^2 T^4} O\left(N T^4\right) = O\left(\frac{1}{N}\right)$$

since $\varepsilon_{i,t}$ has finite fourth moments and because $i \neq l$ and stationarity (absolute summable autocovariances) of $\mathbf{v}_{i,t}$ results in

$$vec\left(\sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{r=1}^{T} \sum_{q=1}^{T} \mathbb{E}\left(\mathbf{V}_{i,l}^{t,s} \boldsymbol{\beta}\boldsymbol{\beta}' \mathbf{V}_{i,l}^{r,q\prime}\right)\right) = \left(\sum_{t=1}^{T} \sum_{r=1}^{T} \mathbb{E}\left[\mathbf{v}_{i,r} \otimes \mathbf{v}_{i,t}\right]\right) \left(\sum_{s=1}^{T} \sum_{q=1}^{T} \mathbb{E}\left[\mathbf{v}_{l,q} \otimes \mathbf{v}_{l,s}\right]\right) vec(\boldsymbol{\beta}\boldsymbol{\beta}') = O(T^2)$$

Consider then $(n = i, m = j, i \neq m)$, which gives

$$\frac{1}{N^2 T^4} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{r=1}^{T} \sum_{q=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{o=1}^{N} \mathbb{E}\left(\mathbf{V}_{i,l}^{t,s} \boldsymbol{\beta}\boldsymbol{\beta}' \mathbf{V}_{j,o}^{r,q\prime}\right) \mathbb{E}\left(e_{i,j}^{t,s} e_{j,i}^{r,q}\right)$$

where the left expectation is again zero unless one of 2 additional restrictions are imposed, either $(o = i, j = l, l \neq i)$ or $(l = j = o = i)$. The latter was already covered above and the first yields

$$\frac{1}{N^2 T^4} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{r=1}^{T} \sum_{q=1}^{T} \sum_{i=1}^{N} \sum_{l=1}^{N} \mathbb{E}\left(\mathbf{V}_{i,l}^{t,s} \boldsymbol{\beta}\boldsymbol{\beta}' \mathbf{V}_{l,i}^{r,q\prime}\right) \mathbb{E}\left(e_{i,l}^{t,s} e_{l,i}^{r,q}\right) = \frac{1}{N^2 T^4} O\left(N^2 T^2\right) = O\left(\frac{1}{T^2}\right)$$

since $i \neq l$ and $e_{i,l}^{t,s} e_{l,i}^{r,q} = \varepsilon_{i,t}\varepsilon_{i,q}\varepsilon_{l,s}\varepsilon_{l,r}$ gives by $\varepsilon_{i,t}$ having absolute summable autocovariances that

$$\sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{r=1}^{T} \sum_{q=1}^{T} \mathbb{E}(e_{i,l}^{t,s} e_{l,i}^{r,q}) = \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{r=1}^{T} \sum_{q=1}^{T} \mathbb{E}(\varepsilon_{i,t}\varepsilon_{i,q})\mathbb{E}(\varepsilon_{l,s}\varepsilon_{l,r}) = \left(\sum_{t=1}^{T} \sum_{q=1}^{T} \mathbb{E}(\varepsilon_{i,t}\varepsilon_{i,q})\right) \left(\sum_{s=1}^{T} \sum_{r=1}^{T} \mathbb{E}(\varepsilon_{l,s}\varepsilon_{l,r})\right)$$

$$= \left(\sum_{t=1}^{T} \sum_{q=1}^{T} \sigma_{i,t,q}\right) \left(\sum_{s=1}^{T} \sum_{r=1}^{T} \sigma_{l,s,r}\right) = O(T^2)$$

Finally, if $(n = j, m = i, j \neq i)$, then we require either $(o = i, l = i)$ or $(o = l)$, which gives respectively

$$(o = i, l = i) \qquad \frac{1}{N^2 T^4} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{r=1}^{T} \sum_{q=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{E}\left(\mathbf{V}_{i,i}^{t,s} \boldsymbol{\beta}\boldsymbol{\beta}' \mathbf{V}_{i,i}^{r,q\prime}\right) \mathbb{E}\left(e_{i,j}^{t,s} e_{i,j}^{r,q}\right) = \frac{1}{N^2 T^4} O\left(N^2 T^2\right) = O\left(\frac{1}{T^2}\right)$$

$$(o = l) \qquad \frac{1}{N^2 T^4} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{r=1}^{T} \sum_{q=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \mathbb{E}\left(\mathbf{V}_{i,l}^{t,s} \boldsymbol{\beta}\boldsymbol{\beta}' \mathbf{V}_{i,l}^{r,q\prime}\right) \mathbb{E}\left(e_{i,j}^{t,s} e_{i,j}^{r,q}\right) = \frac{1}{N^2 T^4} O\left(N^3 T^2\right) = O\left(\frac{N}{T^2}\right)$$

$$= O\left(\frac{1}{T}\right)$$

by similar arguments as above, and $T/N = O(1)$. Hence, combining results yields

$$\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{V}_{i,l}^{t,s} \boldsymbol{\beta} e_{i,j}^{t,s} \right\| = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \tag{2.40}$$

and therefore

$$\begin{aligned}\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{v}_{i,t} \mathbf{v}_{l,s}' \boldsymbol{\beta} \varepsilon_{j,t} \varepsilon_{i,s} &= \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \boldsymbol{\Sigma}_{i,t,s} \boldsymbol{\beta} \sigma_{i,t,s} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \\ &= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)\end{aligned}$$

since by Ass.1 (summable autocovariances) also $\left\| T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \boldsymbol{\Sigma}_{i,t,s} \boldsymbol{\beta} \sigma_{i,t,s} \right\| = O(1)$. The exact same arguments also yield

$$\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{v}_{i,t} \mathbf{v}_{l,s}^{(g-1)} \varepsilon_{i,s} \varepsilon_{j,t} \right\| = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

Next, for combinations such as $\mathbf{v}_{i,t} \mathbf{v}_{j,t}' \boldsymbol{\beta} \varepsilon_{l,s} \varepsilon_{i,s}$ we can write with $\mathbf{V}_{i,j}^{t} = \mathbf{v}_{i,t} \mathbf{v}_{j,t}' - \mathbf{1}_{(j=i)} \boldsymbol{\Sigma}_i$ and $e_{i,l}^{s} = \varepsilon_{i,s} \varepsilon_{l,s} - \mathbf{1}_{(l=i)} \sigma_i^2$

$$\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{v}_{i,t} \mathbf{v}_{j,t}' \boldsymbol{\beta} \varepsilon_{l,s} \varepsilon_{i,s} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Sigma}_i \boldsymbol{\beta} \sigma_i^2 + \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{V}_{i,j}^{t} \boldsymbol{\beta} e_{i,l}^{s}$$

Note then that

$$\mathbb{E}\left[\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{V}_{i,j}^{t} \boldsymbol{\beta} e_{i,l}^{s} \right] = \mathbf{0}_{k \times 1}$$

Here, the variance is

$$\begin{aligned}&\mathbb{E}\left[\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{V}_{i,j}^{t} \boldsymbol{\beta} e_{i,l}^{s} \right] \left[\frac{1}{NT^2} \sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{o=1}^{N} \sum_{r=1}^{T} \sum_{q=1}^{T} \mathbf{V}_{m,n}^{r} \boldsymbol{\beta} e_{m,o}^{q} \right]' \\ &= \frac{1}{N^2 T^4} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{r=1}^{T} \sum_{q=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{o=1}^{N} \mathbb{E}\left(\mathbf{V}_{i,j}^{t} \boldsymbol{\beta} \boldsymbol{\beta}' \mathbf{V}_{m,n}^{r\prime} \right) \mathbb{E}\left(e_{i,l}^{s} e_{m,o}^{q} \right)\end{aligned}$$

We see that the left expectation is zero unless $(m = j = n = i)$, or either $(n = m, j = i, i \neq m)$, $(n = j, m = i, i \neq j)$ or $(n = i, m = j, i \neq j)$. Depending on this choice, the free indices o and l in the second expectation must be chosen such that the whole product is non zero. In what follows we proceed with the systematic elimination using absolute summability and the arguments similar to the ones used above. We indicate the appropriate index choice for the first expectation in the first parentheses and the relevant

o, l adjustment for the second expectation in the second ones.

$$(i=j=m=n),(o=l) \quad \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{l=1}^{N}\mathbb{E}\left(\mathbf{V}_{i,i}^{t}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}_{i,i}^{r\prime}\right)\mathbb{E}\left(e_{i,l}^{s}e_{i,l}^{q}\right)=O(N^2T^3)$$

$$(i=j=m=n),(o=l=i) \quad \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\mathbb{E}\left(\mathbf{V}_{i,i}^{t}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}_{i,i}^{r\prime}\right)\mathbb{E}\left(e_{i,i}^{s}e_{i,i}^{q}\right)=O(NT^4)$$

$$(n=j,m=i,i\neq j),(o=l) \quad \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\mathbb{E}\left(\mathbf{V}_{i,j}^{t}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}_{i,j}^{r\prime}\right)\mathbb{E}\left(e_{i,l}^{s}e_{i,l}^{q}\right)=O(N^3T^2)$$

$$(n=j,m=i,i\neq j),(o=l=i) \quad \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left(\mathbf{V}_{i,j}^{t}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}_{i,j}^{r\prime}\right)\mathbb{E}\left(e_{i,i}^{s}e_{i,i}^{q}\right)=O(N^2T^3)$$

$$(m=j,n=i,i\neq j),(l=j,o=i) \quad \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left(\mathbf{V}_{i,j}^{t}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}_{j,i}^{r\prime}\right)\mathbb{E}\left(e_{i,j}^{s}e_{j,i}^{q}\right)=O(N^2T^2)$$

$$(m=j,n=i,i\neq j),(l=i,o=j) \quad \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left(\mathbf{V}_{i,j}^{t}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}_{j,i}^{r\prime}\right)\mathbb{E}\left(e_{i,i}^{s}e_{j,j}^{q}\right)=O(N^2T^2)$$

$$(m=n,j=i,i\neq m),(o=i,l=m) \quad \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left(\mathbf{V}_{i,i}^{t}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}_{m,m}^{r\prime}\right)\mathbb{E}\left(e_{i,m}^{s}e_{m,i}^{q}\right)=O(N^2T^3)$$

$$(m=n,j=i,i\neq m),(o=m,l=i) \quad \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{m=1}^{N}\mathbb{E}\left(\mathbf{V}_{i,i}^{t}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}_{m,m}^{r\prime}\right)\mathbb{E}\left(e_{i,i}^{s}e_{m,m}^{q}\right)=O(N^2T^3)$$

which by making use of $T/N=O(1)$, leads to the conclusion that

$$\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{V}_{i,j}^{t}\boldsymbol{\beta}e_{i,l}^{s}\right\|=O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right) \tag{2.41}$$

such that

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{v}_{i,t}\mathbf{v}_{j,t}'\boldsymbol{\beta}\varepsilon_{l,s}\varepsilon_{i,s}=\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i\boldsymbol{\beta}\sigma_i^2+O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)$$

and also by near identical arguments

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{v}_{i,t}\mathrm{v}_{j,t}^{(v-1)}\varepsilon_{l,s}\varepsilon_{i,s}=\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i\mathbf{q}_{(v-1)}\sigma_i^2+O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)$$

It remains to analyze the terms with triples of the same variable, and a single occurrence of an independent variable, such as the very first term $\mathbf{v}_{i,t}\varepsilon_{j,t}\varepsilon_{l,s}\varepsilon_{i,s}$. For such terms we obtain that

$$\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{v}_{i,t}\varepsilon_{j,t}\varepsilon_{l,s}\varepsilon_{i,s}\right\|=O_p\left(\frac{1}{\sqrt{T}}\right) \tag{2.42}$$

since, by independence of all $\mathbf{v}_{i,t}$ and $\varepsilon_{j,t}$,

$$\mathbb{E}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{v}_{i,t}\varepsilon_{j,t}\varepsilon_{l,s}\varepsilon_{i,s}\right]=\mathbf{0}_{k\times 1}$$

and also

$$\mathbb{E}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{v}_{i,t}\varepsilon_{j,t}\varepsilon_{l,s}\varepsilon_{i,s}\right]\left[\frac{1}{NT^2}\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{o=1}^{N}\sum_{r=1}^{T}\sum_{q=1}^{T}\mathbf{v}_{m,r}\varepsilon_{n,r}\varepsilon_{o,q}\varepsilon_{m,q}\right]'$$

$$= \frac{1}{N^2T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{o=1}^{N}\mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}'_{m,r})\mathbb{E}(\varepsilon_{j,t}\varepsilon_{l,s}\varepsilon_{i,s}\varepsilon_{n,r}\varepsilon_{o,q}\varepsilon_{m,q})$$

$$= \frac{1}{N^2T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{n=1}^{N}\sum_{o=1}^{N}\mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}'_{i,r})\mathbb{E}(\varepsilon_{j,t}\varepsilon_{l,s}\varepsilon_{i,s}\varepsilon_{n,r}\varepsilon_{o,q}\varepsilon_{i,q})$$

because $\mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}'_{m,r}) = \mathbf{0}_{k\times k}$ for all $m \neq i$. The further analysis of the remaining term is too extensive and cumbersome to fully write down here, but careful elimination making use of the independence over cross-sections in Ass.1 shows that four typical terms remain in the expression above, of which the orders are given next

$$\frac{1}{N^2T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}'_{i,r})\mathbb{E}(\varepsilon_{i,t}\varepsilon_{i,r}\varepsilon^2_{i,s}\varepsilon^2_{i,q}) = \frac{1}{N^2T^4}O(NT^3) = O\left(\frac{1}{NT}\right)$$

$$\frac{1}{N^2T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{l\neq j,i}^{N}\mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}'_{i,r})\mathbb{E}(\varepsilon_{j,t}\varepsilon_{j,r})\mathbb{E}(\varepsilon_{i,s}\varepsilon_{i,q})\mathbb{E}(\varepsilon_{l,q}\varepsilon_{l,s}) = \frac{1}{N^2T^4}O(N^3T^2) = O\left(\frac{1}{T}\right)$$

$$\frac{1}{N^2T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{l\neq i}^{N}\mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}'_{i,r})\mathbb{E}(\varepsilon_{i,s}\varepsilon_{i,r}\varepsilon_{i,q})\mathbb{E}(\varepsilon_{l,t}\varepsilon_{l,s}\varepsilon_{l,q}) = \frac{1}{N^2T^4}O(N^2T^3) = O\left(\frac{1}{T}\right)$$

$$\frac{1}{N^2T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{l\neq i}^{N}\mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}'_{i,r})\mathbb{E}(\varepsilon_{i,t}\varepsilon_{i,s}\varepsilon_{i,r}\varepsilon_{i,q})\mathbb{E}(\varepsilon_{l,q}\varepsilon_{l,s}) = \frac{1}{N^2T^4}O(N^2T^3) = O\left(\frac{1}{T^1}\right),$$

because of stationarity of both involved variables, $T/N = O(1)$, and $E(\varepsilon^6_{i,t}) < \infty$ by Ass.1. Here, for example, the order of the last term can be deduced from

$$\left\|\frac{1}{N^2T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{l\neq i}^{N}\mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}'_{i,r})\mathbb{E}(\varepsilon_{i,t}\varepsilon_{i,s}\varepsilon_{i,r}\varepsilon_{i,q})\mathbb{E}(\varepsilon_{l,q}\varepsilon_{l,s})\right\|$$

$$= N\sup_{i,r,t}\|\mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}'_{i,r})\| \times \sup_{l,q,s}|\mathbb{E}(\varepsilon_{l,s}\varepsilon_{l,q})| \times \frac{1}{N^2T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}|\mathbb{E}(\varepsilon_{i,t}\varepsilon_{i,s}\varepsilon_{i,r}\varepsilon_{i,q})|$$

$$= \frac{1}{N^2T^4}O(N^2T^3) = O(T^{-1}).$$

This results in the original statement above, and we similarly obtain

$$\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{v}_{i,t}\mathbf{v}'_{j,t}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{v}_{l,s}\varepsilon_{i,s}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right)$$

$$\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{v}_{i,t}\mathrm{v}^{(v-1)}_{j,t}\mathbf{v}'_{l,s}\boldsymbol{\beta}\varepsilon_{i,s}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right)$$

$$\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{v}_{i,t}\mathbf{v}'_{j,t}\boldsymbol{\beta}\mathrm{v}^{(g-1)}_{l,s}\varepsilon_{i,s}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right)$$

$$\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{v}_{i,t}\mathrm{v}^{(v-1)}_{j,t}\mathrm{v}^{(g-1)}_{l,s}\varepsilon_{i,s}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right)$$

Combining then all these results in (2.39) gives

$$
\begin{aligned}
\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\mathbf{V}_i'\mathbf{U}_j\mathbf{D}_{-m}\mathbf{U}_l'\boldsymbol{\varepsilon}_i &= d_{1,1}\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\mathbf{v}_{i,t}\mathbf{v}_{j,t}'\boldsymbol{\beta}\varepsilon_{l,s}\varepsilon_{i,s}\right]\right\}\\
&\quad+\sum_{v=2}^{1+k}d_{v,1}\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\mathbf{v}_{i,t}\mathrm{v}_{j,t}^{(v-1)}\varepsilon_{l,s}\varepsilon_{i,s}\right]\right\}\\
&\quad+O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{N}}\right)\\
&= d_{1,1}\left\{\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i\boldsymbol{\beta}\sigma_i^2\right\}+\sum_{v=2}^{1+k}d_{v,1}\left\{\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i\mathbf{q}_{v-1}\sigma_i^2\right\}\\
&\quad+O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{N}}\right)\\
&= \frac{1}{N}\sum_{i=1}^{N}d_{1,1}\boldsymbol{\Sigma}_i\boldsymbol{\beta}\sigma_i^2+\frac{1}{N}\sum_{i=1}^{N}\sum_{v=2}^{k+1}d_{v,1}\boldsymbol{\Sigma}_i\mathbf{q}_{v-1}\sigma_i^2\\
&\quad+O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{N}}\right)\\
&= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i[\boldsymbol{\beta},\mathbf{I}_k]\mathbf{D}_{-m}[\sigma_i^2,\mathbf{0}_{1\times k}]'+O_p(T^{-1/2})+O_p(N^{-1/2}), \qquad (2.43)
\end{aligned}
$$

such that finally

$$
\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon}=\sqrt{\tau}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i[\boldsymbol{\beta},\mathbf{I}_k]\mathbf{D}_{-m}[\sigma_i^2,\mathbf{0}_{1\times k}]'+O_p(T^{-1/2})+O_p(N^{-1/2}) \qquad (2.44)
$$

Combining then (2.30), (2.35), (2.37) and (2.44) into $\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]}=\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon}-\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\eta}-\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\varepsilon}+\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\eta}$ gives the final result in (2.25).

It remains to show that $\mathbf{d}_1=\mathbf{d}_2=\mathbf{0}_{k\times 1}$ when $m=1+k$. In this case, given $T^{-1}\overline{\mathbf{Z}}'\overline{\mathbf{Z}}=\overline{\mathbf{C}}'T^{-1}\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}+O_p(N^{-1})+O_p((NT)^{-1/2})$ and $rk(T^{-1}\overline{\mathbf{Z}}'\overline{\mathbf{Z}})-rk((\overline{\mathbf{C}}'T^{-1}\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}))\xrightarrow{a.s.}0$ it follows by Theorem 1 in Karabiyik et al. (2017) that $\left\|(T^{-1}\overline{\mathbf{Z}}'\overline{\mathbf{Z}})^{\dagger}-(\overline{\mathbf{C}}'T^{-1}\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^{\dagger}\right\|=O_p(N^{-1})+O_p((NT)^{-1/2})$ and $\left\|(T^{-1}\overline{\mathbf{Z}}'\overline{\mathbf{Z}})^{\dagger}\right\|=O_p(1)$. Hence, whereas $m=1+k$ yields by definition $\mathbf{R}=\overline{\mathbf{C}}^{-1}$ so that $\mathbf{M}_{\mathbf{F}^0}=\mathbf{M}_{\mathbf{F}}$, the fact that by the properties of the generalized inverse we have $\mathbf{M}_{\mathbf{F}^0}=\mathbf{M}_{\mathbf{F}}=\mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}}$ and also $\mathbf{M}_{\widehat{\mathbf{F}}^0}=\mathbf{M}_{\widehat{\mathbf{F}}}$, of which crucially all components are well behaved, lets us simplify and analyze the decomposition in (2.11) (given $m=1+k$) as

$$
\begin{aligned}
\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}=\mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}}-\mathbf{M}_{\widehat{\mathbf{F}}} &= T^{-1}\overline{\mathbf{U}}(T^{-1}\overline{\mathbf{Z}}'\overline{\mathbf{Z}})^{\dagger}\overline{\mathbf{U}}'+T^{-1}\overline{\mathbf{U}}(T^{-1}\overline{\mathbf{Z}}'\overline{\mathbf{Z}})^{\dagger}\overline{\mathbf{C}}'\mathbf{F}'+T^{-1}\mathbf{F}\overline{\mathbf{C}}(T^{-1}\overline{\mathbf{Z}}'\overline{\mathbf{Z}})^{\dagger}\overline{\mathbf{U}}'\\
&\quad+T^{-1}\mathbf{F}\overline{\mathbf{C}}[(T^{-1}\overline{\mathbf{Z}}'\overline{\mathbf{Z}})^{\dagger}-(\overline{\mathbf{C}}'T^{-1}\mathbf{F}'\mathbf{F}\overline{\mathbf{C}})^{\dagger}]\overline{\mathbf{C}}'\mathbf{F}' \qquad (2.45)
\end{aligned}
$$

Then, substituting in this decomposition yields

$$
\left\|\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\eta}\right\|\leq\sqrt{\tau_{N,T}}\left\|\frac{1}{N}\sum_{i=1}^{N}(\mathbf{q}_y'\widetilde{\boldsymbol{\eta}}_i\otimes\boldsymbol{\Gamma}_i')\right\|\left\|\overline{\mathbf{C}}^{\dagger}\right\|^2 N\left\|T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}\right\|=O_p(N^{-1})+O_p((NT)^{-1/2})
$$

because by application of the results in lemma B-1 we now obtain

$$\begin{aligned}\left\|T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}\right\| &\leq \left\|\frac{\overline{\mathbf{U}}'\overline{\mathbf{U}}}{T}\right\|^2\left\|\left(\frac{\overline{\mathbf{Z}}'\overline{\mathbf{Z}}}{T}\right)^{\dagger}\right\|+2\left\|\frac{\overline{\mathbf{U}}'\overline{\mathbf{U}}}{T}\right\|\left\|\left(\frac{\overline{\mathbf{Z}}'\overline{\mathbf{Z}}}{T}\right)^{\dagger}\right\|\|\overline{\mathbf{C}}\|\left\|\frac{\mathbf{F}'\overline{\mathbf{U}}}{T}\right\| \\ &\quad+\|\overline{\mathbf{C}}\|^2\left\|\frac{\mathbf{F}'\overline{\mathbf{U}}}{T}\right\|^2\left\|\left(\frac{\overline{\mathbf{Z}}'\overline{\mathbf{Z}}}{T}\right)^{\dagger}-\left(\overline{\mathbf{C}}'\frac{\mathbf{F}'\mathbf{F}}{T}\overline{\mathbf{C}}\right)^{\dagger}\right\| \\ &= O_p(N^{-2})+O_p(N^{-3/2}T^{-1/2})\end{aligned}$$

Next up are $\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\varepsilon}$ and $\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\eta}$. Making use of $\boldsymbol{\varepsilon}_i=\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$, $\mathbf{V}_i=\mathbf{U}_i\mathbf{q}_x$ and $T/N=O(1)$

$$\left\|\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\varepsilon}\right\| \leq \sqrt{\tau_{N,T}}\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{\Gamma}_i\|\left\|\overline{\mathbf{C}}^{\dagger}\right\|N\left\|T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\right\|\left\|\mathbf{B}^{-1}\mathbf{q}_y\right\|=O_p(N^{-1})+O_p((NT)^{-1/2})$$

$$\left\|\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\eta}\right\| \leq \sqrt{\tau_{N,T}}\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{q}_x\|N\left\|T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\right\|\left\|\overline{\mathbf{C}}^{\dagger}\right\|\|\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\|=O_p(N^{-1})+O_p((NT)^{-1/2})$$

since from (2.45) and lemma B-1 follows

$$\begin{aligned}&\left\|T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\right\| \\ &\leq \left\|\frac{\overline{\mathbf{U}}'\overline{\mathbf{U}}}{T}\right\|\left\|\left(\frac{\overline{\mathbf{Z}}'\overline{\mathbf{Z}}}{T}\right)^{\dagger}\right\|\left\|\frac{\overline{\mathbf{U}}'\mathbf{U}_i}{T}\right\|+\left\|\frac{\overline{\mathbf{U}}'\overline{\mathbf{U}}}{T}\right\|\left\|\left(\frac{\overline{\mathbf{Z}}'\overline{\mathbf{Z}}}{T}\right)^{\dagger}\right\|\|\overline{\mathbf{C}}\|\left\|\frac{\mathbf{F}'\mathbf{U}_i}{T}\right\|+\left\|\frac{\mathbf{F}'\overline{\mathbf{U}}}{T}\right\|\|\overline{\mathbf{C}}\|\left\|\left(\frac{\overline{\mathbf{Z}}'\overline{\mathbf{Z}}}{T}\right)^{\dagger}\right\|\left\|\frac{\overline{\mathbf{U}}'\mathbf{U}_i}{T}\right\| \\ &\quad+\|\overline{\mathbf{C}}\|^2\left\|\frac{\mathbf{F}'\overline{\mathbf{U}}}{T}\right\|\left\|\left(\frac{\overline{\mathbf{Z}}'\overline{\mathbf{Z}}}{T}\right)^{\dagger}-\left(\overline{\mathbf{C}}'\frac{\mathbf{F}'\mathbf{F}}{T}\overline{\mathbf{C}}\right)^{\dagger}\right\|\left\|\frac{\mathbf{F}'\mathbf{U}_i}{T}\right\| \\ &= O_p(N^{-2})+O_p(N^{-3/2}T^{-1/2})\end{aligned}$$

Finally, for $\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon}$ we find

$$\left\|\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon}\right\| \leq \sqrt{\tau_{N,T}}\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{q}_x\|N\left\|T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\right\|\left\|\mathbf{B}^{-1}\mathbf{q}_y\right\|=O_p(N^{-1})+O_p(T^{-1/2})$$

from $T/N=O(1)$ and

$$\begin{aligned}&\left\|T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\right\| \\ &\leq \left\|\frac{\overline{\mathbf{U}}'\mathbf{U}_i}{T}\right\|^2\left\|\left(\frac{\overline{\mathbf{Z}}'\overline{\mathbf{Z}}}{T}\right)^{\dagger}\right\|+2\left\|\frac{\overline{\mathbf{U}}'\mathbf{U}_i}{T}\right\|\left\|\left(\frac{\overline{\mathbf{Z}}'\overline{\mathbf{Z}}}{T}\right)^{\dagger}\right\|\|\overline{\mathbf{C}}\|\left\|\frac{\mathbf{F}'\mathbf{U}_i}{T}\right\|+\|\overline{\mathbf{C}}\|^2\left\|\frac{\mathbf{F}'\mathbf{U}_i}{T}\right\|^2\left\|\left(\frac{\overline{\mathbf{Z}}'\overline{\mathbf{Z}}}{T}\right)^{\dagger}-\left(\overline{\mathbf{C}}'\frac{\mathbf{F}'\mathbf{F}}{T}\overline{\mathbf{C}}\right)^{\dagger}\right\| \\ &= O_p(N^{-2})+O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})\end{aligned}$$

Hence, by combining results we have when $m=1+k$ as $(N,T)\to\infty$

$$\left\|\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]}\right\|=O_p(N^{-1})+O_p(T^{-1/2})$$

which implies that $\mathbf{d}_1=\mathbf{d}_2=\mathbf{0}_{k\times 1}$ in eq.(2.25) of the lemma, as needed to be shown.

Lemma B-4 *Under Ass.1-5 we have*

$$\left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}}\right\| = O_p(T^{-1/2}) \tag{2.46}$$

as $(N,T) \to \infty$.

Proof of Lemma B-4

Recall that $\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}} = \widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon} - \widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\eta} - \widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},\Gamma\varepsilon} + \widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},\Gamma\eta}$. We begin the proof by noting that $\mathbf{P}_{\mathbf{F}^0} = \mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}$ and $\mathbf{F}^0 = [\mathbf{F}, \mathbf{0}_{T,1+k-m}]$. Rewriting the last term gives

$$\begin{aligned}\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},\Gamma\eta} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y \\ &= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}(\mathbf{q}_y'\widetilde{\boldsymbol{\eta}}_i'\otimes\mathbf{\Gamma}_i')\right]\left[\overline{\mathbf{C}}^{\dagger}\otimes\overline{\mathbf{C}}^{\dagger}\right]' vec\left(NT^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}\right) = O_p(T^{-1})\end{aligned} \tag{2.47}$$

because $\left\|\frac{1}{N}\sum_{i=1}^{N}(\mathbf{q}_y'\widetilde{\boldsymbol{\eta}}_i'\otimes\mathbf{\Gamma}_i')\right\| = O_p(1)$ from Ass.3 and $\left\|T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}\right\| = O_p((NT)^{-1/2})$ from Lemma B-1 yields

$$\left\|NT^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}\right\| \leq N\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\right\|\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}\right\|\left\|T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}\right\| = O_p(T^{-1})$$

Next, with $\left\|T^{-1}\mathbf{F}^{0\prime}\mathbf{U}_i\right\| = O_p(T^{-1/2})$ from Lemma B-1 and $\mathbf{V}_i = \mathbf{U}_i\mathbf{q}_x$, $\boldsymbol{\varepsilon}_i = \mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$ follows

$$\begin{aligned}\left\|\sqrt{N}T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\boldsymbol{\varepsilon}_i\right\| &\leq \sqrt{N}\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\right\|\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}\right\|\left\|T^{-1}\mathbf{F}^{0\prime}\mathbf{U}_i\right\|\left\|\mathbf{B}^{-1}\mathbf{q}_y\right\| = O_p(T^{-1}) \\ \left\|\sqrt{N}T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\mathbf{V}_i\right\| &\leq \sqrt{N}\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\right\|\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}\right\|\left\|T^{-1}\mathbf{F}^{0\prime}\mathbf{U}_i\right\|\left\|\mathbf{q}_x\right\| = O_p(T^{-1})\end{aligned}$$

such that

$$\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},\Gamma\varepsilon} = \sqrt{T}\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\sqrt{N}T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\boldsymbol{\varepsilon}_i = O_p(T^{-1/2}) \tag{2.48}$$

$$\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\eta} = \sqrt{T}\frac{1}{N}\sum_{i=1}^{N}T^{-1}\mathbf{V}_i'\mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}\sqrt{N}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y = O_p(T^{-1/2}) \tag{2.49}$$

Next, by independence in Ass.5 and expectation zero error terms in Ass.1

$$\mathbb{E}\left(\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon}\right) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{E}\left(\mathbf{V}_i\right)'\mathbb{E}\left(\mathbf{P}_{\mathbf{F}^0}\right)\mathbb{E}\left(\boldsymbol{\varepsilon}_i\right) = \mathbf{0}_{k\times 1}$$

and also

$$\mathbb{E}\left(\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon}\right)\left(\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon}\right)' = \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left(\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j')\mathbf{P}_{\mathbf{F}^0}\mathbf{V}_j\right) = \frac{1}{NT}\sum_{i=1}^{N}\mathbb{E}\left(\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')\mathbf{P}_{\mathbf{F}^0}\mathbf{V}_i\right)$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\mathbb{E}\left(\mathbf{V}_i'\mathbf{F}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')\mathbf{F}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\mathbf{V}_i\right)$$

$$= \frac{1}{NT^3}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\mathbb{E}\left(\mathbf{v}_{i,t}\mathbf{f}_t^{0\prime}\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}\mathbf{f}_s^0\mathbb{E}(\varepsilon_{i,s}\varepsilon_{i,r})\mathbf{f}_r^{0\prime}\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}\mathbf{f}_q^0\mathbf{v}_{i,q}'\right)$$

$$= \frac{1}{NT^3}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{q=1}^{T}\mathbb{E}\left(\mathbf{v}_{i,t}\mathbf{f}_t^{0\prime}\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}\left[\sum_{s=1}^{T}\sum_{r=1}^{T}\sigma_{i,s,r}\mathbf{f}_s^0\mathbf{f}_r^{0\prime}\right]\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}\mathbf{f}_q^0\mathbf{v}_{i,q}'\right)$$

$$= \frac{1}{NT^3}O(NT^2) = O\left(\frac{1}{T}\right)$$

by the stationarity of $\mathbf{f}_t, \varepsilon_{i,t}, \mathbf{v}_{i,t}$ and their mutual independence, implies that

$$\left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon}\right\| = O_p(T^{-1/2}) \tag{2.50}$$

Combining all the results above then leads to

$$\left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}}\right\| \leq \left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon}\right\| + \left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\eta}\right\| + \left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},\Gamma\varepsilon}\right\| + \left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},\Gamma\eta}\right\| = O_p(T^{-1/2})$$

which is what needed to be shown.

Lemma B-5 *Under Ass.1-5 as $(N,T)\to\infty$ such that $\tau_{N,T}\to\tau<\infty$,*

$$\widehat{\mathbf{q}}_{\mathbf{I}} = \widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon} + \sqrt{\tau}(\mathbf{b}_1 - \mathbf{b}_2) + o_p(1) \tag{2.51}$$

$$\widehat{\mathbf{q}}_{\mathbf{I}} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_{k\times 1}, \boldsymbol{\Psi}\right) + \sqrt{\tau}(\mathbf{b}_1 - \mathbf{b}_2) \tag{2.52}$$

with $\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_i$ *and*

$$\mathbf{b}_1 = \mathbf{q}_{xy}'\boldsymbol{\Sigma}_{\eta}' vec((\mathbf{C}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{C}^{\dagger})$$

$$\mathbf{b}_2 = \boldsymbol{\Gamma}'(\mathbf{C}^{\dagger})'[\sigma^2, \mathbf{0}_{1\times k}]'$$

and where $\boldsymbol{\Psi} = \text{plim}_{(N,T)\to\infty}\frac{1}{N}\sum_{i=1}^{N}(T^{-1}\mathbf{V}_i'\boldsymbol{\Omega}_i\mathbf{V}_i)$.

Proof of Lemma B-5

Recall that

$$\widehat{\mathbf{q}}_{\mathbf{I}} = \widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon} - \widehat{\mathbf{q}}_{\mathbf{I},V\eta} - \widehat{\mathbf{q}}_{\mathbf{I},\Gamma\varepsilon} + \widehat{\mathbf{q}}_{\mathbf{I},\Gamma\eta}$$

For the last term in this decomposition we find using familiar operations

$$\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\eta} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y = \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\otimes\boldsymbol{\Gamma}_i)\right]' vec\left((\overline{\mathbf{C}}^{\dagger})'T^{-1}N\overline{\mathbf{U}}'\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\right)$$

$$= \sqrt{\tau}\mathbf{q}_{xy}'\boldsymbol{\Sigma}_{\eta}' vec\left((\mathbf{C}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{C}^{\dagger}\right) + O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{2.53}$$

where we have substituted in (2.31) and $T^{-1}N\overline{\mathbf{U}}'\overline{\mathbf{U}} = \mathbf{\Sigma_u} + O_p(T^{-1/2})$. Similarly making use of $\boldsymbol{\varepsilon}_i = \mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$ yields

$$\begin{aligned}
\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\varepsilon} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'\boldsymbol{\varepsilon}_i = \sqrt{\tau_{N,T}}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'T^{-1}\overline{\mathbf{U}}'\boldsymbol{\varepsilon}_i \\
&= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'(T^{-1}\mathbf{U}_i'\mathbf{U}_i)\mathbf{B}^{-1}\mathbf{q}_y\right] + \sqrt{\tau_{N,T}}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\mathbf{U}_j'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right] \\
&= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'[\sigma_i^2,\mathbf{0}_{1\times k}]'\right] + O_p(T^{-1/2}) \\
&= \sqrt{\tau_{N,T}}\mathbf{\Gamma}'(\overline{\mathbf{C}}^{\dagger})'\frac{1}{N}\sum_{i=1}^{N}[\sigma_i^2,\mathbf{0}_{1\times k}]' + \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{q}_x'\widetilde{\boldsymbol{\eta}}_i'(\overline{\mathbf{C}}^{\dagger})'[\sigma_i^2,\mathbf{0}_{1\times k}]'\right] + O_p(T^{-1/2}) &\quad (2.54)\\
&= \sqrt{\tau}\mathbf{\Gamma}'(\mathbf{C}^{\dagger})'[\sigma^2,\mathbf{0}_{1\times k}]' + O_p(N^{-1/2}) + O_p(T^{-1/2}) &\quad (2.55)
\end{aligned}$$

where we substituted in (2.34) and made use of $T^{-1}\mathbf{U}_i'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y = \mathbf{\Sigma}_{\mathbf{u},i}\mathbf{B}^{-1}\mathbf{q}_y + O_p(T^{-1/2}) = [\sigma_i^2,\mathbf{0}_{1\times k}]' + O_p(T^{-1/2})$. For the next term, making use of $\mathbf{V}_i = \mathbf{U}_i\mathbf{q}_x$ and substituting in the same results as above leads to

$$\begin{aligned}
\widehat{\mathbf{q}}_{\mathbf{I},V\eta} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y = \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{q}_x'(T^{-1}\mathbf{U}_i'\mathbf{U}_i)\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\right] + \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbf{q}_x'(T^{-1}\mathbf{U}_i'\mathbf{U}_j)\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\right] \\
&= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{q}_x'\mathbf{\Sigma}_{\mathbf{u},i}\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\right] + O_p(T^{-1/2}) = \sqrt{\tau_{N,T}}\mathbf{q}_{xy}'\left[\frac{1}{N}\sum_{i=1}^{N}(\widetilde{\boldsymbol{\eta}}_i\otimes\mathbf{\Sigma}_{\mathbf{u},i})\right]'vec(\overline{\mathbf{C}}^{\dagger}) + O_p(T^{-1/2}) \\
&= O_p(N^{-1/2}) + O_p(T^{-1/2}) &\quad (2.56)
\end{aligned}$$

because $\frac{1}{N}\sum_{i=1}^{N}\widetilde{\boldsymbol{\eta}}_i = O_p(N^{-1/2})$ by Ass.3. Finally, as $\|\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}\| = O_p(1)$ and fourth moments are finite it follows under Ass.1 that

$$\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_i \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k\times 1},\mathbf{\Psi}) \quad (2.57)$$

as $(N,T)\to\infty$, with $\mathbf{\Psi} = \text{plim}_{(N,T)\to\infty}\frac{1}{N}\sum_{i=1}^{N}(T^{-1}\mathbf{V}_i'\mathbf{\Omega}_i\mathbf{V}_i)$ and $\mathbf{\Omega}_i = \mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')$. Hence, combining (2.53)-(2.57) in the decomposition of $\widehat{\mathbf{q}}_{\mathbf{I}}$ then leads to

$$\widehat{\mathbf{q}}_{\mathbf{I}} \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k\times 1},\mathbf{\Psi}) + \sqrt{\tau}(\mathbf{b}_1 - \mathbf{b}_2)$$

with $\mathbf{b}_1 = \mathbf{q}_{xy}'\mathbf{\Sigma}_{\eta}'vec((\mathbf{C}^{\dagger})'\mathbf{\Sigma_u}\mathbf{C}^{\dagger})$ and $\mathbf{b}_2 = \mathbf{\Gamma}'(\mathbf{C}^{\dagger})'[\sigma^2,\mathbf{0}_{1\times k}]'$, which is what needed to be shown.

Lemma B-6 *Under Ass.1-5 we have that*

$$\widehat{\mathbf{Q}} \longrightarrow^p \mathbf{\Sigma} \quad (2.58)$$

$$\widehat{\mathbf{Q}}^{-1} \longrightarrow^p \mathbf{\Sigma}^{-1} \quad (2.59)$$

as $(N,T)\to\infty$.

Proof of Lemma B-6

Recall that

$$\widehat{\mathbf{Q}} = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i = \frac{1}{NT}\sum_{i=1}^{N}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\mathbf{\Gamma}_i]'\mathbf{M}_{\widehat{\mathbf{F}}}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\mathbf{\Gamma}_i] = \widehat{\mathbf{Q}}_{\mathbf{I}} - \widehat{\mathbf{Q}}_{\mathbf{M}_{\mathbf{F}^0}} - \widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]}$$

which for a given subscript $\mathbf{A}$ is in turn decomposed as

$$\widehat{\mathbf{Q}}_{\mathbf{A}} = \widehat{\mathbf{Q}}_{\mathbf{A},VV} - \widehat{\mathbf{Q}}_{\mathbf{A},V\Gamma} - (\widehat{\mathbf{Q}}_{\mathbf{A},V\Gamma})' + \widehat{\mathbf{Q}}_{\mathbf{A},\Gamma\Gamma}$$

$$\widehat{\mathbf{Q}}_{\mathbf{A},VV} = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{A}\mathbf{V}_i$$

$$\widehat{\mathbf{Q}}_{\mathbf{A},V\Gamma} = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{A}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\mathbf{\Gamma}_i$$

$$\widehat{\mathbf{Q}}_{\mathbf{A},\Gamma\Gamma} = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'\mathbf{A}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\mathbf{\Gamma}_i$$

Recall that $\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\right\| = O_p(N^{-1})$ and $\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\right\| = O_p((NT)^{-1/2})$ by Lemma B-1, and note that (2.28) implies $\left\|T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}\right\| = O_p(N^{-1})$. Then we have

$$\left\|\widehat{\mathbf{Q}}_{\mathbf{I},\Gamma\Gamma}\right\| \leq \left\|\frac{1}{N}\sum_{i=1}^{N}(\mathbf{\Gamma}_i'\otimes\mathbf{\Gamma}_i')\right\| \left\|\overline{\mathbf{C}}^{\dagger}\right\|^2 \left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\right\| = O_p(N^{-1})$$

$$\left\|\widehat{\mathbf{Q}}_{\mathbf{P}_{\mathbf{F}^0},\Gamma\Gamma}\right\| \leq \left\|\frac{1}{N}\sum_{i=1}^{N}(\mathbf{\Gamma}_i'\otimes\mathbf{\Gamma}_i')\right\| \left\|\overline{\mathbf{C}}^{\dagger}\right\|^2 \left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\right\|^2 \left\|\widehat{\mathbf{\Sigma}}_{\mathbf{F}^0}\right\| = O_p((NT)^{-1})$$

$$\left\|\widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\Gamma}\right\| \leq \left\|\frac{1}{N}\sum_{i=1}^{N}(\mathbf{\Gamma}_i'\otimes\mathbf{\Gamma}_i')\right\| \left\|\overline{\mathbf{C}}^{\dagger}\right\|^2 \left\|T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}\right\| = O_p(N^{-1})$$

Next, the fact that $\mathbf{V}_i = \mathbf{U}_i\mathbf{q}_x$ and using also $\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{U}_i\right\| = O_p(N^{-1}) + O_p((NT)^{-1/2})$ and $\|T^{-1}\mathbf{F}^{0\prime}\mathbf{U}_i\| = O_p(T^{-1/2})$ of Lemma B-1 reveal that

$$\left\|\widehat{\mathbf{Q}}_{\mathbf{I},V\Gamma}\right\| \leq \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{q}_x\| \left\|T^{-1}\overline{\mathbf{U}}'\mathbf{U}_i\right\| \left\|\overline{\mathbf{C}}^{\dagger}\right\| \|\mathbf{\Gamma}_i\| = O_p(N^{-1}) + O_p((NT)^{-1/2})$$

$$\left\|\widehat{\mathbf{Q}}_{\mathbf{P}_{\mathbf{F}^0},V\Gamma}\right\| \leq \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{q}_x\| \left\|T^{-1}\mathbf{F}^{0\prime}\mathbf{U}_i\right\| \left\|\widehat{\mathbf{\Sigma}}_{\mathbf{F}^0}\right\| \left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\right\| \left\|\overline{\mathbf{C}}^{\dagger}\right\| \|\mathbf{\Gamma}_i\| = O_p(N^{-1/2}T^{-1})$$

$$\left\|\widehat{\mathbf{Q}}_{\mathbf{P}_{\mathbf{F}^0},VV}\right\| \leq \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{q}_x\|^2 \left\|T^{-1}\mathbf{F}^{0\prime}\mathbf{U}_i\right\|^2 \left\|\widehat{\mathbf{\Sigma}}_{\mathbf{F}^0}\right\| = O_p(T^{-1})$$

For the next result, given that we made use of $\varepsilon_i = \mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$ to derive the results in equations (2.32) and (2.33), they imply directly that

$$\left\|T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\right\| = O_p(N^{-1}) + O_p((NT)^{-1/2}) \tag{2.60}$$

so that in turn

$$\left\|\widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\Gamma}\right\| \leq \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{q}_x\| \left\|T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\right\| \left\|\overline{\mathbf{C}}^{\dagger}\right\| \|\mathbf{\Gamma}_i\| = O_p(N^{-1}) + O_p((NT)^{-1/2})$$

Next, we have obtained in (2.38) that

$$T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i = T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_{-m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{-m}^0)'\mathbf{U}_i$$
$$+ O_p(N^{-3/2}) + O_p(N^{-1/2}T^{-1}) + O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2})$$

which using $\left\|T^{-1}(\overline{\mathbf{U}}_{-m}^0)'\mathbf{U}_i\right\| = O_p(N^{-1/2}) + O_p(T^{-1/2})$ from Lemma B-2 leads to

$$\begin{aligned}\left\|T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\right\| &\leq \left\|T^{-1}(\overline{\mathbf{U}}_{-m}^0)'\mathbf{U}_i\right\|^2 \left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0}^{\dagger}\right\| \\ &\quad + O_p(N^{-3/2}) + O_p(N^{-1/2}T^{-1}) + O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2}) \\ &= O_p(N^{-1}) + O_p(T^{-1}) + O_p((NT)^{-1/2}) \end{aligned} \tag{2.61}$$

and substituting in this result yields

$$\left\|\widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}],VV}\right\| \leq \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{q}_x\|^2\left\|T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\right\| = O_p(N^{-1}) + O_p(T^{-1}) + O_p((NT)^{-1/2})$$

For the last remaining term it follows from Ass.1 that

$$\widehat{\mathbf{Q}}_{\mathbf{I},VV} = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{V}_i'\mathbf{V}_i}{T} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i + O_p(T^{-1/2})$$

Finally, by combining then all the previous results

$$\widehat{\mathbf{Q}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i + O_p(T^{-1/2}) + O_p(N^{-1})$$

Equation (2.58) of the lemma follows from this and $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i \longrightarrow^p \boldsymbol{\Sigma}$ as $N \to \infty$ by Ass.1, with $\boldsymbol{\Sigma}$ positive definite, which in turn, given that $rk(\widehat{\mathbf{Q}}) - rk(\boldsymbol{\Sigma}) \xrightarrow{a.s.} 0$, then leads to (2.59) by Theorem 1 of Karabiyik et al. (2017).

2.2.2 Theorems and Corollaries

Theorem 1 *Under Ass.1-5 we have as* $(N,T) \to \infty$ *such that* $\tau_{N,T} = T/N \to \tau < \infty$ *that*

$$\sqrt{NT}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}) + \sqrt{\tau}\boldsymbol{\Sigma}^{-1}(\mathbf{b} - \mathbf{d})$$

where $\mathbf{b} = \mathbf{b}_1 - \mathbf{b}_2$, $\mathbf{d} = \mathbf{d}_1 + \mathbf{d}_2$ *are given in Lemmas B-3 and B-5, with* $\mathbf{d} = \mathbf{0}_{k\times 1}$ *when* $m = 1 + k$*.*

Proof of Theorem 1

Recall that the scaled CCEP estimator is

$$\sqrt{NT}(\widehat{\beta} - \beta) = \widehat{\mathbf{Q}}^{-1}\widehat{\mathbf{q}}$$

Substituting in lemmas B-3, B-4 and B-5 into $\widehat{\mathbf{q}} = \widehat{\mathbf{q}}_{\mathbf{I}} - \widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}} - \widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]}$ results in

$$\widehat{\mathbf{q}} \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Psi}) + \sqrt{\tau}(\mathbf{b}_1 - \mathbf{b}_2) - \sqrt{\tau}(\mathbf{d}_1 + \mathbf{d}_2)$$

as $(N,T) \to \infty$ such that $\tau_{N,T} \to \tau < \infty$. This together with $\widehat{\mathbf{Q}}^{-1} \longrightarrow^p \boldsymbol{\Sigma}^{-1}$ as $(N,T) \to \infty$ from Lemma B-6 then leads to

$$\sqrt{NT}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}) + \sqrt{\tau}\boldsymbol{\Sigma}^{-1}(\mathbf{b} - \mathbf{d})$$

where $\mathbf{b} = \mathbf{b}_1 - \mathbf{b}_2$ and $\mathbf{d} = \mathbf{d}_1 + \mathbf{d}_2$, as was to be shown.

Corollary 1 *Under Ass.1-3, 5 and 7, we have as* $(N,T) \to \infty$ *such that* $\tau_{N,T} \to \tau < \infty$ *that*

$$\sqrt{NT}(\widehat{\beta}_{\mathbf{x}} - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}) + \sqrt{\tau}\boldsymbol{\Sigma}^{-1}\mathbf{g}$$

where

$$\mathbf{g} = \mathbf{q}_{xy}'\boldsymbol{\Sigma}_\eta' vec((\boldsymbol{\Gamma}^\dagger)'\boldsymbol{\Sigma}(\mathbf{I}_k - \mathbf{D}_{\mathbf{x},-m}\boldsymbol{\Sigma})\boldsymbol{\Gamma}^\dagger)$$

and with $\mathbf{D}_{\mathbf{x},-m} = \mathbf{T}_{\mathbf{x}}\mathbf{H}_{\mathbf{x},-m}(\mathbf{H}_{\mathbf{x},-m}'\mathbf{T}_{\mathbf{x}}'\boldsymbol{\Sigma}\mathbf{T}_{\mathbf{x}}\mathbf{H}_{\mathbf{x},-m})^\dagger\mathbf{H}_{\mathbf{x},-m}'\mathbf{T}_{\mathbf{x}}'$.

Proof of Corollary 1

The asymptotic distribution of the CCEP estimator when $\overline{\mathbf{y}}$ is excluded in the estimation of the factors can be studied by replacing $\overline{\mathbf{Z}}, \overline{\mathbf{U}}, \overline{\mathbf{C}}$ with $\overline{\mathbf{X}}, \overline{\mathbf{V}}$ and $\overline{\boldsymbol{\Gamma}}$, respectively, such that $\mathbf{F} = (\overline{\mathbf{X}} - \overline{\mathbf{V}})\overline{\boldsymbol{\Gamma}}^\dagger$ and $\mathbf{P}_{\widehat{\mathbf{F}}} = \overline{\mathbf{X}}(\overline{\mathbf{X}}'\overline{\mathbf{X}})^\dagger\overline{\mathbf{X}}'$. The appropriate rotation matrix in case $m < k$ is then $\mathbf{R}_{\mathbf{x}} = \mathbf{T}_{\mathbf{x}}\overline{\mathbf{H}}_{\mathbf{x}}\mathbf{D}_{N,\mathbf{x}}$ where $\mathbf{T}_{\mathbf{x}}$ is the $k \times k$ partitioning such that $\overline{\boldsymbol{\Gamma}}\mathbf{T}_{\mathbf{x}} = [\overline{\boldsymbol{\Gamma}}_m, \overline{\boldsymbol{\Gamma}}_{-m}]$ with $\overline{\boldsymbol{\Gamma}}_m$ an $m \times m$ full rank matrix and also $\overline{\mathbf{V}}\mathbf{T}_{\mathbf{x}} = [\overline{\mathbf{V}}_m, \overline{\mathbf{V}}_{-m}]$. The remaining matrices are now

$$\overline{\mathbf{H}}_{\mathbf{x}} = [\overline{\mathbf{H}}_{\mathbf{x},m}, \overline{\mathbf{H}}_{\mathbf{x},-m}] = \begin{bmatrix} \overline{\boldsymbol{\Gamma}}_m^{-1} & -\overline{\boldsymbol{\Gamma}}_m^{-1}\overline{\boldsymbol{\Gamma}}_{-m} \\ \mathbf{0}_{(k-m)\times m} & \mathbf{I}_{k-m} \end{bmatrix}, \qquad \mathbf{D}_{N,\mathbf{x}} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times(k-m)} \\ \mathbf{0}_{(k-m)\times m} & \sqrt{N}\mathbf{I}_{k-m} \end{bmatrix} \tag{2.62}$$

with also $\mathbf{H_x} = [\mathbf{H}_{\mathbf{x},m}, \mathbf{H}_{\mathbf{x},-m}] = \begin{bmatrix} \mathbf{\Gamma}_m^{-1} & -\mathbf{\Gamma}_m^{-1}\mathbf{\Gamma}_{-m} \\ \mathbf{0}_{(k-m)\times m} & \mathbf{I}_{k-m} \end{bmatrix}$ and $\mathbf{\Gamma}_m$ and $\mathbf{\Gamma}_{-m}$ denoting the partitioning following from $\mathbf{\Gamma T_x} = [\mathbf{\Gamma}_m, \mathbf{\Gamma}_{-m}]$.

Replacing in the analysis of Theorem 1 the $\mathbf{R}, \mathbf{T}, \overline{\mathbf{H}}, \mathbf{H}$ respectively with $\mathbf{R_x}, \mathbf{T_x}, \overline{\mathbf{H}}_\mathbf{x}$ and $\mathbf{H_x}$ and allows us to study the CCEP estimator, with for completeness now

$$\widehat{\mathbf{F}}^0 = [\mathbf{F}\overline{\mathbf{\Gamma}} + \overline{\mathbf{V}}]\mathbf{R_x} = \mathbf{F}^0 + \overline{\mathbf{V}}^0 \tag{2.63}$$

where $\mathbf{F}^0 = \mathbf{F}\mathbf{R_x} = [\mathbf{F}, \mathbf{0}_{T\times(k-m)}]$, $\overline{\mathbf{V}}^0 = [\overline{\mathbf{V}}_m^0, \overline{\mathbf{V}}_{-m}^0]$, and with $\overline{\mathbf{V}}_m^0 = \overline{\mathbf{V}}_m\overline{\mathbf{\Gamma}}_m^{-1}$ and $\overline{\mathbf{V}}_{-m}^0 = \sqrt{N}\overline{\mathbf{V}}\mathbf{T_x}\overline{\mathbf{H}}_{\mathbf{x},-m} = \sqrt{N}(\overline{\mathbf{V}}_m - \overline{\mathbf{V}}_m\overline{\mathbf{\Gamma}}_m^{-1}\overline{\mathbf{\Gamma}}_{-m})$.

Denote now the scaled deviation of CCEP estimator

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}_\mathbf{x} - \boldsymbol{\beta}) = \widehat{\mathbf{Q}}_\mathbf{x}^{-1}\widehat{\mathbf{q}}_\mathbf{x} \tag{2.64}$$

where we will employ the same decompositions as introduced in (2.23) and (2.24) but we denote with the additional **x** subscript the fact that in the decomposition $\overline{\mathbf{Z}}, \overline{\mathbf{U}}, \overline{\mathbf{C}}$ are replaced with $\overline{\mathbf{X}}, \overline{\mathbf{V}}$ and $\overline{\mathbf{\Gamma}}$ and the rotation has also been redefined as above.

Consider then that since $\overline{\mathbf{V}} \subset \overline{\mathbf{U}}, \overline{\mathbf{X}} \subset \overline{\mathbf{Z}}$ and $\overline{\mathbf{\Gamma}} \subset \overline{\mathbf{C}}$, all the derived orders in Lemmas B-1 to B-6 are upper bounds for the analysis here. Hence, it follows directly from Lemma B-6

$$\widehat{\mathbf{Q}}_\mathbf{x}^{-1} \longrightarrow^p \mathbf{\Sigma}^{-1} \tag{2.65}$$

whereas from Lemma B-4

$$\left\| \widehat{\mathbf{q}}_{\mathbf{x},\mathbf{P}_{\mathbf{F}^0}} \right\| = O_p(T^{-1/2}) \tag{2.66}$$

and from (2.37) in Lemma B-3

$$\left\| \widehat{\mathbf{q}}_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}],V\eta} \right\| = O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{2.67}$$

Also, employing the same arguments as for (2.30) but setting $\overline{\mathbf{U}} = \overline{\mathbf{V}}, \overline{\mathbf{C}} = \overline{\mathbf{\Gamma}}$ and $\mathbf{T} = \mathbf{T_x}, \overline{\mathbf{H}} = \overline{\mathbf{H}}_\mathbf{x}$ reveals

$$\widehat{\mathbf{q}}_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\eta} = \sqrt{\tau}\mathbf{q}_{xy}'\mathbf{\Sigma}_\eta' vec((\mathbf{\Gamma}^\dagger)'\mathbf{\Sigma}\mathbf{D}_{\mathbf{x},-m}\mathbf{\Sigma}\mathbf{\Gamma}^\dagger) + O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{2.68}$$

with $\mathbf{D}_{\mathbf{x},-m} = \mathbf{T_x}\mathbf{H}_{\mathbf{x},-m}(\mathbf{H}_{\mathbf{x},-m}'\mathbf{T}_\mathbf{x}'\mathbf{\Sigma}\mathbf{T_x}\mathbf{H}_{\mathbf{x},-m})^\dagger\mathbf{H}_{\mathbf{x},-m}'\mathbf{T}_\mathbf{x}'$. The latter follows since setting $\overline{\mathbf{U}} = \overline{\mathbf{V}}$ yields in eq.(2.29) that $\sqrt{N}T^{-1}\overline{\mathbf{U}}\overline{\mathbf{U}}_{-m}^0 = \mathbf{\Sigma}\mathbf{T_x}\mathbf{H}_{\mathbf{x},-m} + O_p(N^{-1/2}) + O_p(T^{-1/2})$ as eq.(2.14) of Lemma B-2 becomes

$$\widehat{\mathbf{\Sigma}}_{\mathbf{u}_{-m}^0}^\dagger = (\mathbf{H}_{\mathbf{x},-m}'\mathbf{T}_\mathbf{x}'\mathbf{\Sigma}\mathbf{T_x}\mathbf{H}_{\mathbf{x},-m})^\dagger + O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{2.69}$$

Next, we have for $\widehat{\mathbf{q}}_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\varepsilon}$ by substituting in the same results as in the proof for $\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\varepsilon}$ in (2.70) of Lemma B-3, defining also $\mathbf{D}_{\mathbf{x}} = (\overline{\mathbf{\Gamma}}^{\dagger})'\mathbf{\Sigma}\mathbf{D}_{\mathbf{x},-m}$ and using $\overline{\mathbf{V}} = N^{-1}(\mathbf{V}_i + \sum_{j\neq i}^{N}\mathbf{V}_j)$

$$\begin{aligned}
\widehat{\mathbf{q}}_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\varepsilon} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{\Gamma}}^{\dagger})'\overline{\mathbf{V}}'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\boldsymbol{\varepsilon}_i \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{\Gamma}}^{\dagger})'\mathbf{\Sigma}\mathbf{D}_{\mathbf{x},-m}\sqrt{N}T^{-1/2}\overline{\mathbf{V}}'\boldsymbol{\varepsilon}_i + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i'\mathbf{D}_{\mathbf{x}}(T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i) + \frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbf{\Gamma}_i'\mathbf{D}_{\mathbf{x}}\mathbf{V}_j'\boldsymbol{\varepsilon}_i\right] + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i'\mathbf{D}_{\mathbf{x}}(T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i)\right] + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= O_p(N^{-1/2}) + O_p(T^{-1/2}) \qquad (2.70)
\end{aligned}$$

where on the fourth line the upper bound for the rightmost term on line three, as derived in (2.34), was substituted in, and the fifth line makes use of $\|\mathbf{\Gamma}_i\| = O_p(1)$ and $\|T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i\| = O_p(T^{-1/2})$ from Ass. 1. Next up is $\widehat{\mathbf{q}}_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon}$. Following the same steps, notation, and making use of the same results as below (2.38) gives, with $d_{v,g}^{\mathbf{x}}$ denoting row v and column g of $\mathbf{D}_{\mathbf{x},-m}$,

$$\begin{aligned}
\widehat{\mathbf{q}}_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon} &= \sqrt{\tau_{N,T}}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\mathbf{V}_i'\mathbf{V}_j\mathbf{D}_{\mathbf{x},-m}\mathbf{V}_l'\boldsymbol{\varepsilon}_i\right] + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= \sqrt{\tau_{N,T}}\left[\sum_{v=1}^{k}\sum_{g=1}^{k}d_{v,g}^{\mathbf{x}}\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{v}_{i,t}\mathrm{v}_{j,t}^{(v)}\mathrm{v}_{l,s}^{(g)}\varepsilon_{i,s}\right\}\right] + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= O_p(N^{-1/2}) + O_p(T^{-1/2}) \qquad (2.71)
\end{aligned}$$

Where we used that fact that since the result in (2.42) holds for sums of products of $\varepsilon_{i,t}$ and $\mathbf{v}_{i,t}$ which feature three occurrences of one and a single of the other, the order obtained in (2.42) is the same as for the leading term here, specifically $\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{v}_{i,t}\mathrm{v}_{j,t}^{(v)}\mathrm{v}_{l,s}^{(g)}\varepsilon_{i,s}\right\| = O_p(T^{-1/2})$. Combining then (2.67), (2.68), (2.70) and (2.71) in $\widehat{\mathbf{q}}_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]} = \widehat{\mathbf{q}}_{\mathbf{x},\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon} - \widehat{\mathbf{q}}_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\eta} - \widehat{\mathbf{q}}_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\varepsilon} + \widehat{\mathbf{q}}_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\eta}$ gives

$$\widehat{\mathbf{q}}_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]} \longrightarrow^{p} \sqrt{\tau}\mathbf{d}_{\mathbf{x}} \qquad (2.72)$$

with $\mathbf{d}_{\mathbf{x}} = \mathbf{q}_{xy}'\mathbf{\Sigma}_{\eta}'vec((\mathbf{\Gamma}^{\dagger})'\mathbf{\Sigma}\mathbf{D}_{\mathbf{x},-m}\mathbf{\Sigma}\mathbf{\Gamma}^{\dagger})$.
Consider next

$$\widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I}} = \widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},V\varepsilon} - \widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},V\eta} - \widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},\Gamma\varepsilon} + \widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},\Gamma\eta}$$

Recalling that $\overline{\mathbf{V}} \subset \overline{\mathbf{U}}$ implies that the earlier derived orders in Lemma B-5 are upper bounds for the analysis here, it follows directly from (2.56) of the proof of Lemma B-5 that

$$\left\|\widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},V\eta}\right\| = O_p(N^{-1/2}) + O_p(T^{-1/2}) \qquad (2.73)$$

Then, for the last term in this decomposition,

$$\widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},\Gamma\eta} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{\Gamma}}^{\dagger})'\overline{\mathbf{V}}'\overline{\mathbf{V}}\overline{\mathbf{\Gamma}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y = \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\otimes\mathbf{\Gamma}_i)\right]' vec\left((\overline{\mathbf{\Gamma}}^{\dagger})'NT^{-1}\overline{\mathbf{V}}'\overline{\mathbf{V}}\overline{\mathbf{\Gamma}}^{\dagger}\right)$$

$$= \sqrt{\tau}\mathbf{q}_{xy}'\mathbf{\Sigma}_{\boldsymbol{\eta}}'vec\left((\mathbf{\Gamma}^{\dagger})'\mathbf{\Sigma}\mathbf{\Gamma}^{\dagger}\right) + O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{2.74}$$

where we have substituted in (2.31) and $(NT^{-1}\overline{\mathbf{V}}'\overline{\mathbf{V}}) = \mathbf{\Sigma} + O_p(T^{-1/2})$. Next,

$$\widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},\Gamma\varepsilon} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{\Gamma}}^{\dagger})'\overline{\mathbf{V}}'\boldsymbol{\varepsilon}_i = \sqrt{\tau_{N,T}}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{\Gamma}}^{\dagger})'T^{-1}\overline{\mathbf{V}}'\boldsymbol{\varepsilon}_i$$

$$= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{\Gamma}}^{\dagger})'(T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i)\right] + \sqrt{\tau_{N,T}}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{\Gamma}}^{\dagger})'\mathbf{V}_j'\boldsymbol{\varepsilon}_i\right]$$

$$= O_p(T^{-1/2}) \tag{2.75}$$

since by the same arguments as for (2.34) the rightmost term on the second line is $O_p(T^{-1/2})$, and for the left term we have used $T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i = O_p(T^{-1/2})$. Then, since by definition $\widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},V\varepsilon} = \widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}$ equation (2.57) directly applies and

$$\widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},V\varepsilon} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_i \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_{k\times 1},\mathbf{\Psi}\right) \tag{2.76}$$

combining (2.73), (2.74), (2.75) and (2.76) in the decomposition of $\widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I}}$ gives

$$\widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I}} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_{k\times 1},\mathbf{\Psi}\right) + \sqrt{\tau}\mathbf{b}_{\mathbf{x}} \tag{2.77}$$

with $\mathbf{b}_{\mathbf{x}} = \mathbf{q}_{xy}'\mathbf{\Sigma}_{\boldsymbol{\eta}}'vec((\mathbf{\Gamma}^{\dagger})'\mathbf{\Sigma}\mathbf{\Gamma}^{\dagger})$. In turn combining (2.66), (2.72) and (2.77) into $\widehat{\mathbf{q}}_{\mathbf{x}} = \widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I}} - \widehat{\mathbf{q}}_{\mathbf{x},\mathbf{P}_{\mathbf{F}0}} - \widehat{\mathbf{q}}_{\mathbf{x},[\mathbf{M}_{\mathbf{F}0}-\mathbf{M}_{\widehat{\mathbf{F}}0}]}$ results in

$$\widehat{\mathbf{q}}_{\mathbf{x}} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_{k\times 1},\mathbf{\Psi}\right) + \sqrt{\tau}(\mathbf{b}_{\mathbf{x}} - \mathbf{d}_{\mathbf{x}}) \tag{2.78}$$

such that with also (2.65) substituted into (2.64) we get

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_{k\times 1},\mathbf{\Sigma}^{-1}\mathbf{\Psi}\mathbf{\Sigma}^{-1}\right) + \sqrt{\tau}\mathbf{\Sigma}^{-1}\mathbf{g}$$

where $\mathbf{g} = \mathbf{b}_{\mathbf{x}} - \mathbf{d}_{\mathbf{x}} = \mathbf{q}_{xy}'\mathbf{\Sigma}_{\boldsymbol{\eta}}'vec((\mathbf{\Gamma}^{\dagger})'\mathbf{\Sigma}(\mathbf{I}_k - \mathbf{D}_{\mathbf{x},-m}\mathbf{\Sigma})\mathbf{\Gamma}^{\dagger})$. This is the stated result.

2.3 Heterogeneous Slopes

We consider here the heterogeneous slope DGP where $\boldsymbol{\beta}_i$ is characterized by Ass.6 such that $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{v}_i$, and it is understood that also the cross-section averages $\overline{\mathbf{U}}, \overline{\mathbf{C}}$ represent the heterogeneous slope variants. Note that all the results in Section 2.1 are derived under Ass.6 and hence apply here as well. In this DGP, we obtain from substituting in (2.4), $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{v}_i$ and $\mathbf{M}_{\widehat{\mathbf{F}}}\overline{\mathbf{Z}} = \mathbf{0}_{T\times(1+k)}$ for the scaled deviation of the Mean Group CCE estimator

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{mg} - \boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}[\widehat{\mathbf{q}}_{v,i} + \widehat{\mathbf{q}}_i] = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{v}_i + \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\widehat{\mathbf{q}}_i \tag{2.79}$$

and in turn, for the scaled deviation of the CCEP estimator, making use of $\gamma_i = \mathbf{C}_i\mathbf{B}_i^{-1}\mathbf{q}_y = \gamma + \widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y$ and $\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma = N\overline{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{F}}}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma = \mathbf{0}_{k\times 1}$, because $\overline{\mathbf{X}} \subset \overline{\mathbf{Z}}$,

$$\begin{aligned}
\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= -\sqrt{N}\boldsymbol{\beta} + \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}[\mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma_i] \\
&= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}[\mathbf{X}_i\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y] \\
&= \overline{\mathbf{Q}}^{-1}[\overline{\mathbf{q}} + \overline{\mathbf{q}}_v],
\end{aligned} \tag{2.80}$$

where in (2.79) and (2.80) we have defined

$$\begin{aligned}
\overline{\mathbf{Q}} &= \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i, & \widehat{\mathbf{Q}}_i &= \frac{\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i}{T} \\
\overline{\mathbf{q}} &= \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{q}}_i, & \widehat{\mathbf{q}}_i &= \frac{\sqrt{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}[\boldsymbol{\varepsilon}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma_i]}{T} \\
\overline{\mathbf{q}}_v &= \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{q}}_{v,i} & \widehat{\mathbf{q}}_{v,i} &= \frac{\sqrt{N}\mathbf{X}_i\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i}{T}\boldsymbol{v}_i
\end{aligned}$$

Making use of (2.5), $\mathbf{M}_{\widehat{\mathbf{F}}}\overline{\mathbf{Z}} = \mathbf{0}_{T\times 1+k}$, $\mathbf{M}_{\widehat{\mathbf{F}}} = \mathbf{M}_{\widehat{\mathbf{F}}^0}$ and $\mathbf{M}_{\widehat{\mathbf{F}}^0} = \mathbf{M}_{\mathbf{F}^0} - [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]$, let the following be the familiar decomposition at the individual level

$$\begin{aligned}
\widehat{\mathbf{Q}}_i &= T^{-1}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i \\
&= T^{-1}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i]'\mathbf{M}_{\mathbf{F}^0}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i] - T^{-1}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i]'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}][\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i] \\
&= \widehat{\mathbf{Q}}_{\mathbf{M}_{\mathbf{F}^0},i} - \widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}],i}
\end{aligned} \tag{2.81}$$

where for a stated subscript $\mathbf{A}$, we define the further decomposition

$$\begin{aligned}
\widehat{\mathbf{Q}}_{\mathbf{A},i} &= \widehat{\mathbf{Q}}_{\mathbf{A},VV,i} - \widehat{\mathbf{Q}}_{\mathbf{A},V\Gamma,i} - (\widehat{\mathbf{Q}}_{\mathbf{A},V\Gamma,i})' + \widehat{\mathbf{Q}}_{\mathbf{A},\Gamma\Gamma,i} \\
\widehat{\mathbf{Q}}_{\mathbf{A},VV,i} &= T^{-1}\mathbf{V}_i'\mathbf{A}\mathbf{V}_i \\
\widehat{\mathbf{Q}}_{\mathbf{A},V\Gamma,i} &= T^{-1}\mathbf{V}_i'\mathbf{A}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i \\
\widehat{\mathbf{Q}}_{\mathbf{A},\Gamma\Gamma,i} &= T^{-1}\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'\mathbf{A}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i
\end{aligned}$$

and where barred variants with an omitted i subscript denote averages over i as $\overline{\mathbf{Q}}_{\mathbf{A},VV} = \frac{1}{N}\sum_{i=1}^{N} \widehat{\mathbf{Q}}_{\mathbf{A},VV,i}$. Next, for the individual-specific numerators

$$\widehat{\mathbf{q}}_i = \sqrt{N}T^{-1}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\mathbf{\Gamma}_i]'\mathbf{M}_{\widehat{\mathbf{F}}}[\varepsilon_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma_i] = \widehat{\mathbf{q}}_{\mathbf{I},i} - \widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},i} - \widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],i} \tag{2.82}$$

where for a given subscript $\mathbf{A}$ the respective terms are decomposed as

$$\begin{aligned}
\widehat{\mathbf{q}}_{\mathbf{A},i} &= \widehat{\mathbf{q}}_{\mathbf{A},V\varepsilon,i} - \widehat{\mathbf{q}}_{\mathbf{A},V\gamma,i} - \widehat{\mathbf{q}}_{\mathbf{A},\Gamma\varepsilon,i} + \widehat{\mathbf{q}}_{\mathbf{A},\Gamma\gamma,i} \\
\widehat{\mathbf{q}}_{\mathbf{A},V\varepsilon,i} &= \sqrt{N}T^{-1}\mathbf{V}_i'\mathbf{A}\boldsymbol{\varepsilon}_i \\
\widehat{\mathbf{q}}_{\mathbf{A},V\gamma,i} &= \sqrt{N}T^{-1}\mathbf{V}_i'\mathbf{A}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma_i \\
\widehat{\mathbf{q}}_{\mathbf{A},\Gamma\varepsilon,i} &= \sqrt{N}T^{-1}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'\mathbf{A}\boldsymbol{\varepsilon}_i \\
\widehat{\mathbf{q}}_{\mathbf{A},\Gamma\gamma,i} &= \sqrt{N}T^{-1}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'\mathbf{A}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma_i
\end{aligned}$$

where barred terms will similarly be defined as $\overline{\mathbf{q}}_{\mathbf{A},V\varepsilon} = \frac{1}{N}\sum_{i=1}^{N} \widehat{\mathbf{q}}_{\mathbf{A},V\varepsilon,i}$. Finally, $\widehat{\mathbf{q}}_{\boldsymbol{v}}$ features only in (2.80) so we can directly define the averaged term

$$\begin{aligned}
\overline{\mathbf{q}}_{\boldsymbol{v}} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbf{X}_i\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i}{T}\boldsymbol{v}_i = \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\mathbf{\Gamma}_i]'\mathbf{M}_{\widehat{\mathbf{F}}}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\mathbf{\Gamma}_i]\boldsymbol{v}_i \\
&= \overline{\mathbf{q}}_{\mathbf{I},\boldsymbol{v}} - \overline{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},\boldsymbol{v}} - \overline{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\boldsymbol{v}}
\end{aligned} \tag{2.83}$$

with, given a matrix $\mathbf{A}$,

$$\begin{aligned}
\overline{\mathbf{q}}_{\mathbf{A},\boldsymbol{v}} &= \overline{\mathbf{q}}_{\mathbf{A},VV,\boldsymbol{v}} - \overline{\mathbf{q}}_{\mathbf{A},V\Gamma,\boldsymbol{v}} - (\overline{\mathbf{q}}_{\mathbf{A},V\Gamma,\boldsymbol{v}})' + \overline{\mathbf{q}}_{\mathbf{A},\Gamma\Gamma,\boldsymbol{v}} \\
\overline{\mathbf{q}}_{\mathbf{A},VV,\boldsymbol{v}} &= \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{A}\mathbf{V}_i\boldsymbol{v}_i \\
\overline{\mathbf{q}}_{\mathbf{A},V\Gamma,\boldsymbol{v}} &= \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{A}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\mathbf{\Gamma}_i\boldsymbol{v}_i \\
\overline{\mathbf{q}}_{\mathbf{A},\Gamma\Gamma,\boldsymbol{v}} &= \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'\mathbf{A}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\mathbf{\Gamma}_i\boldsymbol{v}_i
\end{aligned}$$

We next establish the distributions under heterogeneous slopes

2.3.1 Analysis of CCEP

Theorem 4 *Under Ass.1-6, with in addition* $\mathbb{E}(\|\mathbf{v}_{it}\|^8) < \infty$ *and* $\mathbb{E}(\|\boldsymbol{v}_i\|^6) < \infty$, *we have as* $(N,T) \to \infty$ *that*

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k\times 1}, \mathbf{\Sigma}^{-1}\mathbf{\Psi}_h\mathbf{\Sigma}^{-1})$$

with $\mathbf{\Psi}_h = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Sigma}_i\mathbf{\Omega}_{\boldsymbol{v}}\mathbf{\Sigma}_i$.

Proof of Theorem 4

Recall the scaled CCEP deviation in the heterogeneous slope model defined in (2.80). Note that $\overline{\mathbf{Q}} = \widehat{\mathbf{Q}}$ so that the decomposition of $\overline{\mathbf{Q}}$ is the same as that for $\widehat{\mathbf{Q}}$ employed in Lemma B-6. Given then that Lemmas B-1 and B-2 apply equally under the slope heterogeneity characterized by Ass.6, the asymptotic orders derived in Lemma B-6 apply directly to the heterogeneous slope setting and we have from the exact same arguments as in that proof

$$\overline{\mathbf{Q}}^{-1} \longrightarrow^p \mathbf{\Sigma}^{-1} \tag{2.84}$$

Similarly, since heterogeneity does not impact the orders derived in Lemmas B-1 and B-2 (only limit statements are affected) and we have by definition $\overline{\mathbf{q}}_{\mathbf{I}} = \frac{1}{\sqrt{T}}\widehat{\mathbf{q}}_{\mathbf{I}}, \overline{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}} = \frac{1}{\sqrt{T}}\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}}, \overline{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]} = \frac{1}{\sqrt{T}}\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]}$ (so that we have scaled up by $\sqrt{N}$ rather than $\sqrt{NT}$), the results from Lemmas B-3, B-5, B-4 that $\|\widehat{\mathbf{q}}_{\mathbf{I}}\| = O_p(1), \left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}}\right\| = O_p(T^{-1/2}), \left\|\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]}\right\| = O_p(1)$, imply that $\|\overline{\mathbf{q}}_{\mathbf{I}}\| = O_p(T^{-1/2}), \left\|\overline{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}}\right\| = O_p(T^{-1}), \left\|\overline{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]}\right\| = O_p(T^{-1/2})$. Hence, $\|\overline{\mathbf{q}}\| = O_p(T^{-1/2})$ and $\overline{\mathbf{q}}_v$ is the leading term in the asymptotic expansion.

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \overline{\mathbf{Q}}^{-1}\overline{\mathbf{q}}_v + O_p(T^{-1/2})$$

For $\overline{\mathbf{q}}_v$ we start the analysis with the terms containing the deviations $\mathbf{A} = [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]$. For the last term in the decomposition we have

$$\begin{aligned}
\overline{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\Gamma,v} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\mathbf{\Gamma}_i\boldsymbol{v}_i \\
&= \left[\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{v}_i'\mathbf{\Gamma}_i' \otimes \mathbf{\Gamma}_i')\right] vec\left((\overline{\mathbf{C}}^{\dagger})'\sqrt{N}T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\right) \\
&\longrightarrow^p \mathbf{0}_{k\times 1}
\end{aligned}$$

because inserting (2.11) in $\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}$ gives

$$\begin{aligned}
T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}} &= T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_{-m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{-m}}T^{-1}(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}} + T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_m\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_m)'\overline{\mathbf{U}} \\
&\quad + T^{-1}\overline{\mathbf{U}}'\mathbf{F}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_m)'\overline{\mathbf{U}} + T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_m\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}\mathbf{F}'\overline{\mathbf{U}} \\
&\quad + T^{-1}\overline{\mathbf{U}}'\widehat{\mathbf{F}}^0\left[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right]T^{-1}(\widehat{\mathbf{F}}^0)'\overline{\mathbf{U}} \\
&= O_p(N^{-1})
\end{aligned} \tag{2.85}$$

which follows analogously to earlier results by application of Lemmas B-1 and B-2 to get

$$\begin{aligned}
&\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_{-m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{-m}}T^{-1}(\overline{\mathbf{U}}^0_{-m})'\overline{\mathbf{U}}\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_{-m}\right\|^2\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{-m}}\right\| = O_p(N^{-1}) \\
&\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_m\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_m)'\overline{\mathbf{U}}\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}^0_m\right\|^2\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}\right\| = O_p(N^{-2}) \\
&\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_m)'\overline{\mathbf{U}}\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}\right\|\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}\right\|\left\|T^{-1}(\overline{\mathbf{U}}^0_m)'\overline{\mathbf{U}}\right\| = O_p(T^{-1/2}N^{-3/2}) \\
&\left\|T^{-1}\overline{\mathbf{U}}'\widehat{\mathbf{F}}^0\left[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right]T^{-1}(\widehat{\mathbf{F}}^0)'\overline{\mathbf{U}}\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}'\widehat{\mathbf{F}}^0\right\|^2\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_u}\right\| = O_p(N^{-3/2}) + O_p(N^{-1}T^{-1/2})
\end{aligned}$$

and also, since by Ass.3 and 6 the mean zero and independence of $\boldsymbol{v}_i$ implies

$$\left\|\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{v}_i'\mathbf{\Gamma}_i' \otimes \mathbf{\Gamma}_i')\right\| = O_p(N^{-1/2})$$

Next, noting $\mathbf{V}_i = \mathbf{U}_i \mathbf{q}_x$ and substituting in (2.60) from the proof of Lemma B-6 gives

$$\left\| \sqrt{N} T^{-1} \mathbf{V}_i' [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}] \overline{\mathbf{U}} \right\| = O_p(N^{-1/2}) + O_p(T^{-1/2})$$

so that

$$\begin{aligned}
\left\| \overline{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}], V\Gamma, \boldsymbol{v}} \right\| &= \left\| \frac{1}{\sqrt{N}T} \sum_{i=1}^{N} \mathbf{V}_i' [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}] \overline{\mathbf{U}} \overline{\mathbf{C}}^{\dagger} \boldsymbol{\Gamma}_i \boldsymbol{v}_i \right\| \\
&\leq \frac{1}{N} \sum_{i=1}^{N} \left\| \sqrt{N} T^{-1} \mathbf{V}_i' [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}] \overline{\mathbf{U}} \right\| \left\| \overline{\mathbf{C}}^{\dagger} \right\| \|\boldsymbol{\Gamma}_i\| \|\boldsymbol{v}_i\| \\
&= O_p(N^{-1/2}) + O_p(T^{-1/2})
\end{aligned}$$

which is not the sharpest possible bound, yet sufficient for our purposes. Then for the final term of this kind

$$\begin{aligned}
\overline{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}], VV, \boldsymbol{v}} &= \frac{1}{\sqrt{N}T} \sum_{i=1}^{N} \mathbf{V}_i' [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}] \mathbf{V}_i \boldsymbol{v}_i = \frac{1}{N} \sum_{i=1}^{N} \sqrt{N} T^{-1} \mathbf{V}_i' [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}] \mathbf{V}_i \boldsymbol{v}_i \\
&\longrightarrow^p \mathbf{0}_{k \times 1}
\end{aligned}$$

This can be seen from the following expansion obtained by substituting (2.11) into $\overline{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}], VV, \boldsymbol{v}}$, and making use of the same arguments as for (2.38), but not approximating terms that are $O_p(\sqrt{N}T^{-1})$,

$$\begin{aligned}
&\frac{1}{N} \sum_{i=1}^{N} \sqrt{N} T^{-1} \mathbf{V}_i' [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}] \mathbf{V}_i \boldsymbol{v}_i \\
&= \frac{1}{N} \sum_{i=1}^{N} \sqrt{N} \left(\frac{\mathbf{V}_i' \overline{\mathbf{U}}_{-m}^0}{T} \right) \widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0}^{\dagger} \left(\frac{\overline{\mathbf{U}}_{-m}^{0\prime} \mathbf{V}_i}{T} \right) \boldsymbol{v}_i + \frac{1}{N} \sum_{i=1}^{N} \sqrt{N} \left(\frac{\mathbf{V}_i' \widehat{\mathbf{F}}^0}{T} \right) [\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0}^{\dagger} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u}^{\dagger}] \left(\frac{\widehat{\mathbf{F}}^{0\prime} \mathbf{V}_i}{T} \right) \boldsymbol{v}_i \\
&\quad + O_p\left(\frac{1}{N^{3/2}} \right) + O_p\left(\frac{1}{T} \right) + O_p\left(\frac{1}{\sqrt{NT}} \right)
\end{aligned}$$

Recall that $\overline{\mathbf{U}}_{-m}^0 = \sqrt{N} \overline{\mathbf{U}} \mathbf{T} \overline{\mathbf{H}}_{-m}$, such that with $\widehat{\mathbf{D}} = \mathbf{T} \overline{\mathbf{H}}_{-m} \widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0}^{\dagger} \overline{\mathbf{H}}_{-m}' \mathbf{T}' = O_p(1)$ the first term in this expansion can be rewritten as

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^{N} \sqrt{N} \left(\frac{\mathbf{V}_i' \overline{\mathbf{U}}_{-m}^0}{T} \right) \widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0}^{\dagger} \left(\frac{\overline{\mathbf{U}}_{-m}^{0\prime} \mathbf{V}_i}{T} \right) \boldsymbol{v}_i &= \frac{1}{N} \sum_{i=1}^{N} N^{3/2} \left(\frac{\mathbf{V}_i' \overline{\mathbf{U}}}{T} \right) \widehat{\mathbf{D}} \left(\frac{\overline{\mathbf{U}}' \mathbf{V}_i}{T} \right) \boldsymbol{v}_i \\
&= \frac{1}{N^{3/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \left(\frac{\mathbf{V}_i' \mathbf{U}_j}{T} \right) \widehat{\mathbf{D}} \left(\frac{\mathbf{U}_k' \mathbf{V}_i}{T} \right) \boldsymbol{v}_i \\
&= \sum_{v=1}^{1+k} \sum_{g=1}^{1+k} \widehat{d}_{v,g} \frac{1}{N^{3/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \left(\frac{\mathbf{V}_i' \mathbf{U}_j^{(v)}}{T} \right) \left(\frac{\mathbf{U}_k^{(g)\prime} \mathbf{V}_i}{T} \right) \boldsymbol{v}_i
\end{aligned}$$

where $\widehat{d}_{v,g}$ denotes the element on row v and column g of $\widehat{\mathbf{D}}$, and $\mathbf{U}_i^{(l)}$ denotes column l of $\mathbf{U}_i$. Hence,

$$\left\| \frac{1}{N} \sum_{i=1}^{N} \sqrt{N} \left(\frac{\mathbf{V}_i' \overline{\mathbf{U}}_{-m}^0}{T} \right) \widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0}^{\dagger} \left(\frac{\overline{\mathbf{U}}_{-m}^{0\prime} \mathbf{V}_i}{T} \right) \boldsymbol{v}_i \right\| \leq \sum_{v=1}^{1+k} \sum_{g=1}^{1+k} |\widehat{d}_{v,g}| \left\| \frac{1}{N^{3/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \left(\frac{\mathbf{V}_i' \mathbf{U}_j^{(v)}}{T} \right) \left(\frac{\mathbf{U}_k^{(g)\prime} \mathbf{V}_i}{T} \right) \boldsymbol{v}_i \right\|$$

where we recall that $|\widehat{d}_{v,g}| = O_p(1)$ and that k is fixed and finite. Noting that $\mathbf{U}_i = [\boldsymbol{\varepsilon}_i + \mathbf{V}_i(\boldsymbol{\beta} + \boldsymbol{v}_i), \mathbf{V}_i]$, the term with the highest degree of dependence (and hence the driver of the asymptotic order) occurs when $v = g = 1$. In that case, the leading term is (since $\boldsymbol{\varepsilon}_i$ is independent of all other terms)

$$\frac{1}{N^{3/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \left(\frac{\mathbf{V}_i' \mathbf{V}_j}{T} \right) \boldsymbol{v}_j \boldsymbol{v}_k' \left(\frac{\mathbf{V}_k' \mathbf{V}_i}{T} \right) \boldsymbol{v}_i$$

Its expectation is zero unless $i = j = k$, and in the latter case, given finite moments

$$\mathbf{A}_i = \mathbb{E} \left[\frac{1}{N^{3/2}} \sum_{i=1}^{N} \left(\frac{\mathbf{V}_i' \mathbf{V}_i}{T} \right) \boldsymbol{v}_i \boldsymbol{v}_i' \left(\frac{\mathbf{V}_i' \mathbf{V}_i}{T} \right) \boldsymbol{v}_i \right] = O\left(\frac{1}{\sqrt{N}} \right)$$

Also, by the cross-section independence, and independence of $\mathbf{V}_i$ and $\boldsymbol{v}_j$ for all i, j

$$\frac{1}{N^3 T^4} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \mathbb{E} \left(\left[\mathbf{V}_i' \mathbf{V}_j \boldsymbol{v}_j \boldsymbol{v}_k' \mathbf{V}_k' \mathbf{V}_i \boldsymbol{v}_i - \mathbf{1}_{(i=j=k)} \mathbf{A}_i \right] \left[\mathbf{V}_l' \mathbf{V}_m \boldsymbol{v}_m \boldsymbol{v}_n' \mathbf{V}_n' \mathbf{V}_l \boldsymbol{v}_l - \mathbf{1}_{(l=m=n)} \mathbf{A}_l \right]' \right)$$
$$= O\left(\frac{1}{T^2} \right) + O\left(\frac{1}{N} \right)$$

which is notably not the sharpest possible order, but sufficient for our purposes. The argument for this result is as follows: the cross-section independence of $\mathbf{V}_i$ and $\boldsymbol{v}_i$ implies that the expectation is zero when one of the indices (k, l, m, n, i, j) differs from the others. This means that the expectation is zero when the product features more than 3 different CS-indices (6 options means at most 3 different index pairs can be constructed without having at least one index differ from all the others). The non-zero part of this expectation can thus be split up into cases with sums over either 3, 2 or 1 distinct CS-indices.
For the case with sums over 3 distinct index pairs, the following situations arise:

- if $i = l$, the structure takes the following form, with for example $j = k$ and $m = n$

$$\frac{1}{N^3 T^4} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \sum_{m \neq i,j}^{N} \mathbb{E} \left(\left[\mathbf{V}_i' \mathbf{V}_j \boldsymbol{v}_j \boldsymbol{v}_j' \mathbf{V}_j' \mathbf{V}_i \boldsymbol{v}_i \right] \left[\mathbf{V}_i' \mathbf{V}_m \boldsymbol{v}_m \boldsymbol{v}_m' \mathbf{V}_m' \mathbf{V}_i \boldsymbol{v}_i \right]' \right)$$

  The summation is always over one set of 4 $\mathbf{V}$'s with a common index, and two pairs of 2 $\mathbf{V}$'s with a common CS-index. Let without loss of generality $k = 1$ for expositional convenience, such that we can explicitly unpack the sums over time. Using also CS-independence, we have

$$\frac{1}{N^3 T^4} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \sum_{m \neq i,j}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{r=1}^{T} \mathbb{E} \left(\left[\mathrm{v}_{it} \mathrm{v}_{jt} v_j^2 \mathrm{v}_{js} \mathrm{v}_{is} v_i \right] \left[\mathrm{v}_{iq} \mathrm{v}_{mq} v_m^2 \mathrm{v}_{mr} \mathrm{v}_{ir} v_i \right] \right)$$
$$= \frac{1}{N^3} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \sum_{m \neq i,j}^{N} \mathbb{E}(v_i^2) \mathbb{E}(v_j^2) \mathbb{E}(v_m^2) \frac{1}{T^4} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{r=1}^{T} \mathbb{E}(\mathrm{v}_{it} \mathrm{v}_{is} \mathrm{v}_{iq} \mathrm{v}_{ir}) \mathbb{E}(\mathrm{v}_{jt} \mathrm{v}_{js}) \mathbb{E}(\mathrm{v}_{mq} \mathrm{v}_{mr}) = O(T^{-2})$$

  because $E(\|\mathbf{v}_{it}\|^4) < \infty$ for all (i, t) and $\sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{r=1}^{T} \mathbb{E}(\mathrm{v}_{jt} \mathrm{v}_{js}) \mathbb{E}(\mathrm{v}_{mq} \mathrm{v}_{mr}) = O(T^2)$ by the finite summability of autocovariances by Ass.1.

- If $i \neq l$, then the structure takes the form below, with for example $i = j$, $l = k$ and $m = n$,

$$\frac{1}{N^3T^4}\sum_{i=1}^{N}\sum_{l\neq i}^{N}\sum_{m\neq i,l}^{N}\mathbb{E}\left(\left[\mathbf{V}_i'\mathbf{V}_i\boldsymbol{v}_i\boldsymbol{v}_l'\mathbf{V}_l'\mathbf{V}_i\boldsymbol{v}_i\right]\left[\mathbf{V}_l'\mathbf{V}_m\boldsymbol{v}_m\boldsymbol{v}_m'\mathbf{V}_m'\mathbf{V}_l\boldsymbol{v}_l\right]'\right)$$

  Summations always contain two sets of 3 $\mathbf{V}$'s with a common index, and one set of 2 $\mathbf{V}$'s with common index (two 3rd moments and one 2nd moment). Setting again $k = 1$ for expositional convenience gives that

$$\frac{1}{N^3T^4}\sum_{i=1}^{N}\sum_{l\neq i}^{N}\sum_{m\neq i,l}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{q=1}^{T}\sum_{r=1}^{T}\mathbb{E}\left(\left[\mathrm{v}_{it}\mathrm{v}_{it}v_i^2v_l\mathrm{v}_{ls}\mathrm{v}_{is}\right]\left[\mathrm{v}_{lq}\mathrm{v}_{mq}v_m^2\mathrm{v}_{mr}\mathrm{v}_{lr}v_l\right]\right)$$
$$=\frac{1}{N^3}\sum_{i=1}^{N}\sum_{l\neq i}^{N}\sum_{m\neq i,l}^{N}\mathbb{E}(v_i^2)\mathbb{E}(v_l^2)\mathbb{E}(v_m^2)\frac{1}{T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{q=1}^{T}\sum_{r=1}^{T}\mathbb{E}(\mathrm{v}_{it}^2\mathrm{v}_{is})\mathbb{E}(\mathrm{v}_{ls}\mathrm{v}_{lq}\mathrm{v}_{lr})\mathbb{E}(\mathrm{v}_{mq}\mathrm{v}_{mr}) = O(T^{-2})$$

  because $\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{q=1}^{T}\sum_{r=1}^{T}\mathbb{E}(\mathrm{v}_{it}^2\mathrm{v}_{is})\mathbb{E}(\mathrm{v}_{ls}\mathrm{v}_{lq}\mathrm{v}_{lr})\mathbb{E}(\mathrm{v}_{mq}\mathrm{v}_{mr}) = O(T^2)$ follows from the stationarity of $\mathbf{v}_{it}$ under Ass.1.

This covers the cases with three distinct indices. For cases with 2 or less distinct CS indices, it is easily seen given the $N^{-3}T^{-4}$ scaling, $\mathbb{E}(\|\mathbf{v}_{it}\|^8) < \infty$ and $\mathbb{E}(\|\boldsymbol{v}_i\|^6) < \infty$, that the sum of expectations can be *at most* of order $O(N^{-1})$. Consequently, $\left\|\frac{1}{N^{3/2}T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\mathbf{V}_i'\mathbf{V}_j\boldsymbol{v}_j\boldsymbol{v}_k'\mathbf{V}_k'\mathbf{V}_i\boldsymbol{v}_i\right\| = o_p(1)$ as $(N,T)\to\infty$, and given that this is also the leading term in the inequality above (where the other terms can be analyzed with near similar arguments), we have

$$\frac{1}{N}\sum_{i=1}^{N}\sqrt{N}\left(\frac{\mathbf{V}_i'\overline{\mathbf{U}}_{-m}^0}{T}\right)\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{-m}^0}^{\dagger}\left(\frac{\overline{\mathbf{U}}_{-m}^{0\prime}\mathbf{V}_i}{T}\right)\boldsymbol{v}_i \longrightarrow^p \mathbf{0}_{k\times 1} \tag{2.86}$$

Next, making use of $\widehat{\mathbf{F}}^0 = \mathbf{F}^0 + [\overline{\mathbf{U}}_m^0, \overline{\mathbf{U}}_{-m}^0]$, and substituting it into the second term of the expansion, it is easily seen that from $\overline{\mathbf{U}}_m^0 = \overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_m$ and $\overline{\mathbf{U}}_{-m}^0 = \sqrt{N}\overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_{-m}$, the two drivers with the slowest decay are respectively

$$\frac{1}{N}\sum_{i=1}^{N}\sqrt{N}\left(\frac{\mathbf{V}_i'\overline{\mathbf{U}}_{-m}^0}{T}\right)\left[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0}^{\dagger}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u}^{\dagger}\right]\left(\frac{(\overline{\mathbf{U}}_{-m}^0)'\mathbf{V}_i}{T}\right)\boldsymbol{v}_i \longrightarrow^p \mathbf{0}_{k\times 1}$$

from the same arguments as (2.86), but noting that the rate is faster since also $\left\|\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0}^{\dagger}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u}^{\dagger}\right\| = O_p(N^{-1/2}) + O_p(T^{-1/2})$ from lemma B-2. For the second driver we similarly find

$$\frac{1}{N}\sum_{i=1}^{N}\sqrt{N}\left(\frac{\mathbf{V}_i'\mathbf{F}^0}{T}\right)\left[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0}^{\dagger}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u}^{\dagger}\right]\left(\frac{\mathbf{F}^{0\prime}\mathbf{V}_i}{T}\right)\boldsymbol{v}_i \longrightarrow^p \mathbf{0}_{k\times 1}$$

because

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\sqrt{N}\left(\frac{\mathbf{V}_i'\mathbf{F}^0}{T}\right)\left[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0}^{\dagger}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u}^{\dagger}\right]\left(\frac{\mathbf{F}^{0\prime}\mathbf{V}_i}{T}\right)\boldsymbol{v}_i\right\| \leq \left\|\frac{1}{\sqrt{N}T^2}\sum_{i=1}^{N}\left(\boldsymbol{v}_i'\mathbf{V}_i'\mathbf{F}^0\otimes\mathbf{V}_i'\mathbf{F}^0\right)\right\|\left\|\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^0}^{\dagger}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_u}^{\dagger}\right\|$$

where $\left\| \frac{1}{\sqrt{N}T^2} \sum_{i=1}^N \left(\boldsymbol{v}_i' \mathbf{V}_i' \mathbf{F}^0 \otimes \mathbf{V}_i' \mathbf{F}^0 \right) \right\| = O_p(T^{-1})$ due to $\left\| T^{-1} \mathbf{F}' \mathbf{V}_i \right\| = O_p(T^{-1/2})$ and because $\boldsymbol{v}_i$ is independent of $(\mathbf{V}_i, \mathbf{F}^0)$, with $\frac{1}{N} \sum_{i=1}^N \sqrt{N} \boldsymbol{v}_i = O_p(1)$. Rigorously, by using $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$ (Abadir and Magnus, 2005, Exercise 10.3) in connection to $a \otimes \mathbf{A} = a\mathbf{A}$ for a scalar a, we obtain

$$\begin{aligned}
&\mathbb{E}\left(\left\| \frac{1}{\sqrt{N}T^2} \sum_{i=1}^N \boldsymbol{v}_i' \mathbf{V}_i' \mathbf{F}^0 \otimes \mathbf{V}_i' \mathbf{F}^0 \right\|^2 \right) \\
&= \mathbb{E}\left(\operatorname{tr}\left[\left(\frac{1}{\sqrt{N}T^2} \sum_{j=1}^N \boldsymbol{v}_j' \mathbf{V}_j' \mathbf{F}^0 \otimes \mathbf{V}_j' \mathbf{F}^0 \right)' \left(\frac{1}{\sqrt{N}T^2} \sum_{i=1}^N \boldsymbol{v}_i' \mathbf{V}_i' \mathbf{F}^0 \otimes \mathbf{V}_i' \mathbf{F}^0 \right) \right] \right) \\
&= \mathbb{E}\left(\operatorname{tr}\left[\left(\frac{1}{\sqrt{N}T^2} \sum_{i=1}^N \boldsymbol{v}_i' \mathbf{V}_i' \mathbf{F}^0 \otimes \mathbf{V}_i' \mathbf{F}^0 \right) \left(\frac{1}{\sqrt{N}T^2} \sum_{j=1}^N \boldsymbol{v}_j' \mathbf{V}_j' \mathbf{F}^0 \otimes \mathbf{V}_j' \mathbf{F}^0 \right)' \right] \right) \\
&= \frac{1}{NT^4} \sum_{i=1}^N \sum_{j=1}^N \operatorname{tr}\left(\mathbb{E}\left[\left(\boldsymbol{v}_i' \mathbf{V}_i' \mathbf{F}^0 \otimes \mathbf{V}_i' \mathbf{F}^0 \right) \left(\mathbf{F}^{0\prime} \mathbf{V}_j \boldsymbol{v}_j \otimes \mathbf{F}^{0\prime} \mathbf{V}_j \right) \right] \right) \\
&= \frac{1}{NT^4} \sum_{i=1}^N \operatorname{tr}\left(\mathbb{E}\left[\left(\boldsymbol{v}_i' \mathbf{V}_i' \mathbf{F}^0 \otimes \mathbf{V}_i' \mathbf{F}^0 \right) \left(\mathbf{F}^{0\prime} \mathbf{V}_i \boldsymbol{v}_i \otimes \mathbf{F}^{0\prime} \mathbf{V}_i \right) \right] \right) \\
&= \frac{1}{NT^4} \sum_{i=1}^N \operatorname{tr}\left(\mathbb{E}\left[\boldsymbol{v}_i' \mathbf{V}_i' \mathbf{F}^0 \mathbf{F}^{0\prime} \mathbf{V}_i \boldsymbol{v}_i \otimes \mathbf{V}_i' \mathbf{F}^0 \mathbf{F}^{0\prime} \mathbf{V}_i \right] \right) \\
&= \frac{1}{NT^4} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \sum_{p=1}^T \sum_{r=1}^T \operatorname{tr}\left(\mathbb{E}\left[(\mathbf{f}_t^0)' \mathbf{f}_s^0 \boldsymbol{v}_i' \mathbf{v}_{i,t} \mathbf{v}_{i,s}' \boldsymbol{v}_i \otimes \mathbf{v}_{i,p} \mathbf{v}_{i,r}' (\mathbf{f}_p^0)' \mathbf{f}_r^0 \right] \right) \\
&= \operatorname{tr}\left(\frac{1}{NT^4} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \sum_{p=1}^T \sum_{r=1}^T \mathbb{E}\left[(\mathbf{f}_t^0)' \mathbf{f}_s^0 (\mathbf{f}_p^0)' \mathbf{f}_r^0 \right] \mathbb{E}\left[\mathbf{v}_{i,p} \mathbf{v}_{i,r}' \boldsymbol{v}_i' \mathbf{v}_{i,t} \mathbf{v}_{i,s}' \boldsymbol{v}_i \right] \right) \\
&= \operatorname{tr}\left(\frac{1}{NT^4} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \sum_{p=1}^T \sum_{r=1}^T \mathbb{E}\left[(\mathbf{f}_t^0)' \mathbf{f}_s^0 (\mathbf{f}_p^0)' \mathbf{f}_r^0 \right] \mathbb{E}\left[\mathbf{v}_{i,p} \mathbf{v}_{i,r}' \operatorname{tr}(\mathbf{v}_{i,t} \mathbf{v}_{i,s}' \boldsymbol{v}_i \boldsymbol{v}_i') \right] \right) \\
&= O(T^{-1}),
\end{aligned} \tag{2.87}$$

because

$$\begin{aligned}
&\left\| \frac{1}{NT^4} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \sum_{p=1}^T \sum_{r=1}^T \mathbb{E}\left[(\mathbf{f}_t^0)' \mathbf{f}_s^0 (\mathbf{f}_p^0)' \mathbf{f}_r^0 \right] \mathbb{E}\left[\mathbf{v}_{i,p} \mathbf{v}_{i,r}' \operatorname{tr}(\mathbf{v}_{i,t} \mathbf{v}_{i,s}' \boldsymbol{v}_i \boldsymbol{v}_i') \right] \right\| \\
&\leq \sup_{t,s,p,r} \left| \mathbb{E}\left[(\mathbf{f}_t^0)' \mathbf{f}_s^0 (\mathbf{f}_p^0)' \mathbf{f}_r^0 \right] \right| \\
&\times \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{T^4} \sum_{p=1}^T \sum_{r=1}^T \sum_{t=1}^T \sum_{s=1}^T \left\| \mathbb{E}\left[\mathbf{v}_{i,p} \mathbf{v}_{i,r}' \operatorname{tr}(\mathbf{v}_{i,t} \mathbf{v}_{i,s}' \boldsymbol{v}_i \boldsymbol{v}_i') \right] \right\| \right) = O(T^{-1}),
\end{aligned}$$

since Æ$_i$ is independent from the rest of the terms. In conclusion,

$$\overline{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}],VV,\boldsymbol{v}} \longrightarrow^p \mathbf{0}_{k \times 1}$$

and combining results leads to

$$\overline{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}],\boldsymbol{v}} \longrightarrow^p \mathbf{0}_{k \times 1}.$$

In similar fashion as above, we have given $\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}\right\| = O_p(1)$ that

$$\overline{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},VV,\boldsymbol{v}} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(T^{-1}\mathbf{V}_i'\mathbf{F}^0)\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}(T^{-1}\mathbf{F}^{0\prime}\mathbf{V}_i)\boldsymbol{v}_i = O_p(T^{-1})$$

also, given $\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}\| = O_p((NT)^{-1/2})$

$$\begin{aligned}\left\|\overline{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\Gamma,\boldsymbol{v}}\right\| &= \left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}T^{-1}\mathbf{V}_i'\mathbf{F}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i\boldsymbol{v}_i\right\| \\ &\leq \frac{1}{N}\sum_{i=1}^{N}\left\|T^{-1}\mathbf{V}_i'\mathbf{F}^0\right\|\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}\right\|\left\|\sqrt{N}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}\right\|\left\|\overline{\mathbf{C}}^{\dagger}\right\|\|\boldsymbol{\Gamma}_i\|\|\boldsymbol{v}_i\| = O_p(T^{-1})\end{aligned}$$

which could again be sharpened noting that $\boldsymbol{\Gamma}_i$ and $\boldsymbol{v}_i$ are independent of the other variables and $\left\|\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i\boldsymbol{v}_i\right\| = O_p(N^{-1/2})$. Finally,

$$\begin{aligned}\left\|\overline{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},\Gamma\Gamma,\boldsymbol{v}}\right\| &= \left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\boldsymbol{\Gamma}_i\boldsymbol{v}_i\right\| \\ &\leq \sqrt{N}\left\|\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{v}_i'\boldsymbol{\Gamma}_i'\otimes\boldsymbol{\Gamma}_i')\right\|\left\|\overline{\mathbf{C}}^{\dagger}\right\|^2\left\|T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}\right\|^2\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}\right\| = O_p((NT)^{-1})\end{aligned}$$

because $\left\|\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{v}_i'\boldsymbol{\Gamma}_i'\otimes\boldsymbol{\Gamma}_i')\right\| = O_p(N^{-1/2})$ by Ass. 3 and 6. Therefore,

$$\overline{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},\boldsymbol{v}} \longrightarrow^{p} \mathbf{0}_{k\times 1}$$

This establishes that both $\overline{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\boldsymbol{v}}$ and $\overline{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},\boldsymbol{v}}$ are asymptotically negligible. What remains is

$$\overline{\mathbf{q}}_{\mathbf{I},\boldsymbol{v}} = \overline{\mathbf{q}}_{\mathbf{A},VV,\boldsymbol{v}} - \overline{\mathbf{q}}_{\mathbf{I},V\Gamma,\boldsymbol{v}} - (\overline{\mathbf{q}}_{\mathbf{A},V\Gamma,\boldsymbol{v}})' + \overline{\mathbf{q}}_{\mathbf{A},\Gamma\Gamma,\boldsymbol{v}}$$

For the final two terms, given that $\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{U}_i\right\| = O_p(N^{-1}) + O_p((NT)^{-1/2})$ and $\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\right\| = O_p(N^{-1})$

$$\begin{aligned}\|\overline{\mathbf{q}}_{\mathbf{I},V\Gamma,\boldsymbol{v}}\| &\leq \frac{1}{N}\sum_{i=1}^{N}\left\|\sqrt{N}T^{-1}\mathbf{V}_i'\overline{\mathbf{U}}\right\|\left\|\overline{\mathbf{C}}^{\dagger}\right\|\|\boldsymbol{\Gamma}_i\|\|\boldsymbol{v}_i\| = O_p(N^{-1/2}) + O_p(T^{-1/2}) \\ \|\overline{\mathbf{q}}_{\mathbf{I},\Gamma\Gamma,\boldsymbol{v}}\| &\leq \sqrt{N}\left\|\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{v}_i'\boldsymbol{\Gamma}_i'\otimes\boldsymbol{\Gamma}_i')\right\|\left\|\overline{\mathbf{C}}^{\dagger}\right\|^2\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\right\| = O_p(N^{-1})\end{aligned}$$

Hence, combining all the results so far yields as $(N,T)\to\infty$

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \overline{\mathbf{Q}}^{-1}\overline{\mathbf{q}}_{\mathbf{I},VV,\boldsymbol{v}} + o_p(1) \tag{2.88}$$

where for the leading term we find as $(N,T)\to\infty$

$$\overline{\mathbf{q}}_{\mathbf{I},VV,\boldsymbol{v}} \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k\times 1},\boldsymbol{\Psi}_h) \tag{2.89}$$

with $\mathbf{\Psi}_h = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbf{\Sigma}_i \mathbf{\Omega}_v \mathbf{\Sigma}_i$ because $\mathbb{E}(\overline{\mathbf{q}}_{\mathbf{I},VV,\boldsymbol{v}}) = \mathbf{0}_{k\times 1}$ and by cross-section and mutual independence of $\boldsymbol{v}_i$ and $\mathbf{V}_i$

$$\begin{aligned}
\mathbb{V}ar(\overline{\mathbf{q}}_{\mathbf{I},VV,\boldsymbol{v}}) &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} \mathbb{E}\left[\left(\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\right)\mathbb{E}(\boldsymbol{v}_i\boldsymbol{v}_j'|\mathbf{V}_i,\mathbf{V}_j)\left(\frac{\mathbf{V}_j'\mathbf{V}_j}{T}\right)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left[\left(\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\right)\mathbb{E}(\boldsymbol{v}_i\boldsymbol{v}_i')\left(\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\right)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left[\left(\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\right)\mathbf{\Omega}_v\left(\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\right)\right] = O(1)
\end{aligned}$$

so that the result in (2.89) follows from applying a CLT to the leading term in

$$\overline{\mathbf{q}}_{\mathbf{I},VV,\boldsymbol{v}} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\right)\boldsymbol{v}_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{\Sigma}_i\boldsymbol{v}_i + O_p(T^{-1/2})$$

Combining then (2.84) and (2.89) into (2.88) gives

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k\times 1}, \mathbf{\Sigma}^{-1}\mathbf{\Psi}_h\mathbf{\Sigma}^{-1}) \tag{2.90}$$

as $(N,T)\to\infty$, which is the result stated in the theorem.

2.3.2 Analysis of CCEMG

Theorem 6 *Under Ass.1-6 as* $(N,T)\to\infty$

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{mg} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{k\times 1}, \mathbf{\Omega}_v)$$

Proof of Theorem 6

Recall the scaled deviation of the CCEMG estimator defined in (2.79) and the decomposition of its components given in (2.81) and (2.82). For the analysis of the denominator, $\widehat{\mathbf{Q}}_i = T^{-1}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i = T^{-1}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^0}\mathbf{X}_i$ can be decomposed into

$$T^{-1}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\mathbf{\Gamma}_i]'\mathbf{M}_{\widehat{\mathbf{F}}^0}[\mathbf{V}_i - \overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\mathbf{\Gamma}_i] = \widehat{\mathbf{Q}}_{\mathbf{I},i} - \widehat{\mathbf{Q}}_{\mathbf{M}_{\mathbf{F}^0},i} - \widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}^0,i} - \mathbf{M}_{\widehat{\mathbf{F}}^0}],i}, \tag{2.91}$$

which is an identical decomposition as in the proof of Lemma B-6, but focused on the summands for individual i only. Because averaging over $i = 1,\ldots,N$ does not alter the order of the remainder, it is immediate from the same lemma that $\widehat{\mathbf{Q}}_i = T^{-1}\mathbf{V}_i'\mathbf{V}_i + O_p(N^{-1}) + O_p(T^{-1}) + O_p((NT)^{-1/2})$. Because $T^{-1}\mathbf{V}_i'\mathbf{V}_i = \mathbf{\Sigma}_i + O_p(T^{-1/2})$, we thus obtain

$$\widehat{\mathbf{Q}}_i = \mathbf{\Sigma}_i + O_p(T^{-1/2}) + O_p(N^{-1}) \tag{2.92}$$

and hence, because $rk(\widehat{\mathbf{Q}}_i) - rk(\mathbf{\Sigma}_i) \overset{a.s.}{\to} 0$, we come to

$$\widehat{\mathbf{Q}}_i^{-1} = \mathbf{\Sigma}_i^{-1} + O_p(T^{-1/2}) + O_p(N^{-1}). \tag{2.93}$$

where we note that $\mathbf{\Sigma}_i$ is positive definite by Ass.1.

Next, we analyze the numerator and use its decomposition $\widehat{\mathbf{q}}_{\mathbf{A},i} = \widehat{\mathbf{q}}_{\mathbf{A},V\varepsilon,i} - \widehat{\mathbf{q}}_{\mathbf{A},V\gamma,i} - \widehat{\mathbf{q}}_{\mathbf{A},\Gamma\varepsilon,i} + \widehat{\mathbf{q}}_{\mathbf{A},\Gamma\gamma,i}$. Then, letting $\mathbf{A} = [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]$, we obtain

$$\begin{aligned}\left\|\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\gamma,i}\right\| &= \left\|\sqrt{N}T^{-1}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma_i\right\| \\ &\leq \sqrt{N}\left\|\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\right\|\left\|\overline{\mathbf{C}}^{\dagger}\gamma_i\right\|\left\|T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}\right\| = O_p(N^{-1/2}),\end{aligned} \tag{2.94}$$

using the fact that $\left\|T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}\right\| = O_p(N^{-1})$ from (2.85). Further, with $\boldsymbol{\varepsilon}_i = \mathbf{U}_i\mathbf{B}_i^{-1}\mathbf{q}_y$ and the result $\left\|T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\right\| = O_p(N^{-1}) + O_p((NT)^{-1/2})$ from (2.60)

$$\begin{aligned}\left\|\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\varepsilon,i}\right\| &= \left\|\sqrt{N}T^{-1}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\boldsymbol{\varepsilon}_i\right\| \\ &\leq \sqrt{N}\left\|\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\right\|\left\|T^{-1}\overline{\mathbf{U}}'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\right\|\left\|\mathbf{B}_i^{-1}\mathbf{q}_y\right\| = O_p(N^{-1/2}) + O_p(T^{-1/2}),\end{aligned} \tag{2.95}$$

Moving on, with $\mathbf{V}_i = \mathbf{U}_i\mathbf{q}_x$, we also immediately obtain

$$\begin{aligned}\left\|\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\gamma,i}\right\| &= \left\|\sqrt{N}T^{-1}\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma_i\right\| \\ &\leq \sqrt{N}\left\|\overline{\mathbf{C}}^{\dagger}\gamma_i\right\|\|\mathbf{q}_x\|\left\|T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}\right\| = O_p(N^{-1/2}) + O_p(T^{-1/2})\end{aligned} \tag{2.96}$$

using the same argument. To proceed, we let $\mathbf{A} = \mathbf{P}_{\mathbf{F}^0}$. This leads to

$$\begin{aligned}\left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},\Gamma\gamma,i}\right\| &= \left\|\sqrt{N}T^{-1}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma_i\right\| \leq \sqrt{N}\left\|\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\right\|\left\|\overline{\mathbf{C}}^{\dagger}\gamma_i\right\|\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\widehat{\mathbf{\Sigma}}_{\mathbf{F}^0}^{\dagger}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}\right\| \\ &\leq \sqrt{N}\left\|\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\right\|\left\|\overline{\mathbf{C}}^{\dagger}\gamma_i\right\|\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\right\|^2\left\|\widehat{\mathbf{\Sigma}}_{\mathbf{F}^0}^{\dagger}\right\| = O_p(N^{-1/2}T^{-1}),\end{aligned} \tag{2.97}$$

which comes from the fact that $\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\right\| = O_p((NT)^{-1/2})$ from Lemma B-1. Further on,

$$\begin{aligned}\left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},\Gamma\varepsilon,i}\right\| &= \left\|\sqrt{N}T^{-1}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'\mathbf{P}_{\mathbf{F}^0}\boldsymbol{\varepsilon}_i\right\| \leq \sqrt{N}\left\|\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\right\|\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\widehat{\mathbf{\Sigma}}_{\mathbf{F}^0}^{\dagger}T^{-1}\mathbf{F}^{0\prime}\boldsymbol{\varepsilon}_i\right\| \\ &\leq \sqrt{N}\left\|\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\right\|\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\right\|\left\|\widehat{\mathbf{\Sigma}}_{\mathbf{F}^0}^{\dagger}\right\|\left\|T^{-1}\mathbf{F}^{0\prime}\boldsymbol{\varepsilon}_i\right\| = O_p(T^{-1}),\end{aligned} \tag{2.98}$$

using the facts that $\left\|T^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\right\| = O_p((NT)^{-1/2})$ and $\|T^{-1}\mathbf{F}^{0\prime}\boldsymbol{\varepsilon}_i\| = O_p(T^{-1/2})$ from $\boldsymbol{\varepsilon}_i = \mathbf{U}_i\mathbf{B}_i^{-1}\mathbf{q}_y$ and $\|T^{-1}\mathbf{F}^{0\prime}\mathbf{U}_i\| = O_p(T^{-1/2})$ in Lemma B-1. Using the latter result again with $\mathbf{V}_i = \mathbf{U}_i\mathbf{q}_x$ gives $\|T^{-1}\mathbf{V}_i'\mathbf{F}^0\| = O_p(T^{-1/2})$, so that in the same fashion,

$$\begin{aligned}\left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\gamma,i}\right\| &= \left\|\sqrt{N}T^{-1}\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma_i\right\| \leq \sqrt{N}\left\|\overline{\mathbf{C}}^{\dagger}\gamma_i\right\|\left\|T^{-1}\mathbf{V}_i'\mathbf{F}^0\widehat{\mathbf{\Sigma}}_{\mathbf{F}^0}^{\dagger}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}\right\| \\ &\leq \sqrt{N}\left\|\overline{\mathbf{C}}^{\dagger}\gamma_i\right\|\left\|T^{-1}\mathbf{V}_i'\mathbf{F}^0\right\|\left\|\widehat{\mathbf{\Sigma}}_{\mathbf{F}^0}^{\dagger}\right\|\left\|T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}\right\| = O_p(T^{-1})\end{aligned} \tag{2.99}$$

Further, we let $\mathbf{A} = \mathbf{I}_T$. Firstly, this leads to

$$\|\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\gamma,i}\| = \left\|\sqrt{N}T^{-1}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma_i\right\| \leq \sqrt{N}\left\|\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\right\|\left\|\overline{\mathbf{C}}^{\dagger}\gamma_i\right\|\left\|T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}\right\| = O_p(N^{-1/2}),$$

because $\left\| T^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}} \right\| = O_p(N^{-1})$. Also,

$$\begin{aligned}\|\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\varepsilon,i}\| &= \left\| \sqrt{N}T^{-1}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})'\overline{\mathbf{U}}'\boldsymbol{\varepsilon}_i \right\| \leq \sqrt{N}\left\| \mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})' \right\| \left\| T^{-1}\overline{\mathbf{U}}'\boldsymbol{\varepsilon}_i \right\| \\ &\leq \sqrt{N}\left\| \mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger})' \right\| \left\| \mathbf{B}_i^{-1} \right\| \|\mathbf{q}_y\| \left\| T^{-1}\overline{\mathbf{U}}'\mathbf{U}_i \right\| = O_p(N^{-1/2}) + O_p(T^{-1/2}),\end{aligned}$$

because $\left\| T^{-1}\overline{\mathbf{U}}'\mathbf{U}_i \right\| = O_p(N^{-1}) + O_p((NT)^{-1/2})$. Eventually, we obtain

$$\begin{aligned}\|\widehat{\mathbf{q}}_{\mathbf{I},V\gamma,i}\| &= \left\| \sqrt{N}T^{-1}\mathbf{V}_i'\overline{\mathbf{U}}\overline{\mathbf{C}}^{\dagger}\gamma_i \right\| \leq \sqrt{N}\left\| \overline{\mathbf{C}}^{\dagger}\gamma_i \right\| \left\| T^{-1}\mathbf{V}_i'\overline{\mathbf{U}} \right\| \leq \sqrt{N}\left\| \overline{\mathbf{C}}^{\dagger}\gamma_i \right\| \|\mathbf{q}_x\| \left\| T^{-1}\mathbf{U}_i'\overline{\mathbf{U}} \right\| \\ &= O_p(N^{-1/2}) + O_p(T^{-1/2}) \end{aligned} \tag{2.100}$$

using the same argument as for the term above. Summarizing the order results for the 3 different versions of $\mathbf{A}$, we come to

$$\widehat{\mathbf{q}}_i = \widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon,i} - \widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon,i} + \widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon,i} + O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{2.101}$$

which in combination with $\left\| \widehat{\mathbf{Q}}_i^{-1} \right\| = O_p(1)$ by (2.93) yields

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\widehat{\mathbf{q}}_i = \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\left[\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon,i} - \widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon,i} + \widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon,i}\right] + O_p(N^{-1/2}) + O_p(T^{-1/2}).$$

Next, consider $\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon,i} = \frac{1}{NT}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\sqrt{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_i$. Clearly, given that by (2.93) $\widehat{\mathbf{Q}}_i^{-1}$ is bounded with a well behaved fixed limit as $(N,T) \to \infty$, the order of this term is driven by $\frac{1}{NT}\sum_{i=1}^{N}\sqrt{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_i$. For the latter,

$$\mathbb{E}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}T^{-1/2}\mathbf{V}_i'\boldsymbol{\varepsilon}_i\right) = \mathbf{0}_{k\times 1} \tag{2.102}$$

and

$$\begin{aligned}&\mathbb{E}\left[\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}T^{-1/2}\mathbf{V}_i'\boldsymbol{\varepsilon}_i\right)\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}T^{-1/2}\mathbf{V}_j'\boldsymbol{\varepsilon}_j\right)'\right] \\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}T^{-1}\mathbb{E}(\mathbf{V}_i'\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j')\mathbf{V}_j) \\ &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(T^{-1}\mathbf{V}_i'\mathbf{\Omega}_i\mathbf{V}_i) = O(1),\end{aligned} \tag{2.103}$$

by the independence of $\mathbf{V}_i$ and $\boldsymbol{\varepsilon}_i$ implies that $\left\| \frac{1}{\sqrt{N}}\sum_{i=1}^{N}T^{-1/2}\mathbf{V}_i'\boldsymbol{\varepsilon}_i \right\| = O_p(1)$, and therefore, by insertion into the term above (and noting that the normalisation is $N^{-1/2}T^{-1}$)

$$\left\| \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon,i} \right\| = O_p(T^{-1/2})$$

Next up is $\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon,i}$. Substituting in (2.11) and making use of the same arguments as for (2.38), but sharpening the approximation (by not expanding $O_p(N^aT^{-b})$ terms with $a,b>0$) gives

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon,i}=\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\sqrt{N}T^{-1}\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\boldsymbol{\varepsilon}_i$$
$$=\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\sqrt{N}\left(\frac{\mathbf{V}_i'\overline{\mathbf{U}}_{-m}^0}{T}\right)\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{u}^0_{-m}}\left(\frac{(\overline{\mathbf{U}}_{-m}^0)'\boldsymbol{\varepsilon}_i}{T}\right)+\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\sqrt{N}\left(\frac{\mathbf{V}_i'\widehat{\mathbf{F}}^0}{T}\right)[\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0}-\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}_u}]\left(\frac{(\widehat{\mathbf{F}}^0)'\boldsymbol{\varepsilon}_i}{T}\right)$$
$$+O_p(N^{-3/2})+O_p(T^{-1})+O_p((NT)^{-1/2})$$

Consider the first term of this expansion. Making use of $\overline{\mathbf{U}}_{-m}^0=\sqrt{N}\overline{\mathbf{U}}\mathbf{T}\overline{\mathbf{H}}_{-m}$ and $\widehat{\mathbf{D}}=\mathbf{T}\overline{\mathbf{H}}_{-m}\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{u}^0_{-m}}\overline{\mathbf{H}}_{-m}'\mathbf{T}'$ gives

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\sqrt{N}\left(\frac{\mathbf{V}_i'\overline{\mathbf{U}}_{-m}^0}{T}\right)\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{u}^0_{-m}}\left(\frac{(\overline{\mathbf{U}}_{-m}^0)'\boldsymbol{\varepsilon}_i}{T}\right)=\frac{1}{N^{3/2}}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\sum_{j=1}^{N}\sum_{k=1}^{N}\left(\frac{\mathbf{V}_i'\mathbf{U}_j}{T}\right)\widehat{\mathbf{D}}\left(\frac{\mathbf{U}_k'\boldsymbol{\varepsilon}_i}{T}\right)$$

Since $\left\|\widehat{\mathbf{D}}\right\|=O_p(1)$, $\left\|\widehat{\mathbf{Q}}_i^{-1}\right\|=O_p(1)$ and both matrices have well behaved limits as $(N,T)\to\infty$ (see e.g. (2.93) and Lemma B-3), the asymptotic order is driven by $N^{-3/2}T^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\mathbf{V}_i'\mathbf{U}_j\widehat{\mathbf{D}}\mathbf{U}_k'\boldsymbol{\varepsilon}_i$. As such, making use of $\left\|N^{-1}T^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\mathbf{V}_i'\mathbf{U}_j\widehat{\mathbf{D}}\mathbf{U}_k'\boldsymbol{\varepsilon}_i\right\|=O_p(1)$, which is obtained by the exact same arguments as for (2.43), we have as $(N,T)\to\infty$

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\sqrt{N}\left(\frac{\mathbf{V}_i'\overline{\mathbf{U}}_{-m}^0}{T}\right)\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{u}^0_{-m}}\left(\frac{(\overline{\mathbf{U}}_{-m}^0)'\boldsymbol{\varepsilon}_i}{T}\right)\right\|=O_p\left(\frac{1}{\sqrt{N}}\right)\tag{2.104}$$

Recall that (2.43) was proven under two restrictions which we do not employ here: $\boldsymbol{\beta}_i=\boldsymbol{\beta}$ and $T/N=O(1)$. For completeness, we will briefly argue that the arguments used to obtain the orders for (2.43) also follow through in the current setting. Firstly, concerning $\boldsymbol{\beta}_i=\boldsymbol{\beta}$, note that Lemma B-2 was derived under heterogeneous slopes, which thus enables its use in preliminary steps. In addition, the $\boldsymbol{v}_i$ are cross-sectionally independent mean zero variables that are also independent from the rest of the model primitives. Hence, given that the $\boldsymbol{v}_i$ also appear in a frequency that is *always lower* than the number of $\mathbf{v}_i$ in any of the expressions leading up to (2.43) (with indices also shared with $\mathbf{v}_i$), the determinant of the dependence structure remains $(\mathbf{v},\varepsilon)$ as in the previous analysis, and all the employed arguments to obtain the asymptotic orders therefore follow through in the heterogeneous setting. Second, concerning $T/N=O(1)$, we note that in the current case the terms are multiplied by an additional $N^{-1/2}$ scaling term. Therefore, given that the highest order remainders from the variance calculations leading to (2.43) took the form $\frac{1}{N^2T^4}O(N^3T^2)$, which vanishes when $T/N=O(1)$, the additional scaling by $N^{-1/2}$ here already brings down the order such that the relative rate restriction is no longer required for the terms to vanish. In conclusion, the heterogeneity does not alter the order results.

For the second term in the expansion, decomposing it with $\widehat{\mathbf{F}}^0=\mathbf{F}^0+[\overline{\mathbf{U}}_m^0,\overline{\mathbf{U}}_{-m}^0]$ reveals that there are two leading terms. For the first we obtain from the same reasoning as for (2.104), but noting also that

$$\left\| \widehat{\boldsymbol{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}_u} \right\| = O_p(N^{-1/2}) + O_p(T^{-1/2}),$$

$$\left\| \frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbf{Q}}_i^{-1} \sqrt{N} \left(\frac{\mathbf{V}_i' \overline{\mathbf{U}}^0_{-m}}{T} \right) [\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}_u}] \left(\frac{(\overline{\mathbf{U}}^0_{-m})' \boldsymbol{\varepsilon}_i}{T} \right) \right\| = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)$$

and for the second

$$\left\| \frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbf{Q}}_i^{-1} \sqrt{N} \left(\frac{\mathbf{V}_i' \mathbf{F}^0}{T} \right) [\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}_u}] \left(\frac{(\mathbf{F}^0)' \boldsymbol{\varepsilon}_i}{T} \right) \right\| = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T^{3/2}}\right) \tag{2.105}$$

because $\left\| \widehat{\mathbf{Q}}_i^{-1} \right\| = O_p(1)$ and $\left\| \widehat{\boldsymbol{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^0} - \widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}_u} \right\| = O_p(N^{-1/2}) + O_p(T^{-1/2})$ imply that the asymptotic order is driven by

$$\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{\mathbf{V}_i' \mathbf{F}^0}{T} \right) \left(\frac{(\mathbf{F}^0)' \boldsymbol{\varepsilon}_i}{T} \right) \right\| = O_p\left(\frac{1}{T}\right) \tag{2.106}$$

which follows from $\mathbb{E}[N^{-1/2} T^{-2} \sum_{i=1}^{N} \mathbf{V}_i' \mathbf{F}^0 (\mathbf{F}^0)' \boldsymbol{\varepsilon}_i] = \mathbf{0}_{k \times 1}$ by Ass.5, and also by making use of the cross-section independence of $\boldsymbol{\varepsilon}_i$ and $\mathbf{V}_i$

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{\mathbf{V}_i' \mathbf{F}^0}{T} \right) \left(\frac{(\mathbf{F}^0)' \boldsymbol{\varepsilon}_i}{T} \right) \right] \left[\frac{1}{\sqrt{N}} \sum_{j=1}^{N} \left(\frac{\mathbf{V}_j' \mathbf{F}^0}{T} \right) \left(\frac{(\mathbf{F}^0)' \boldsymbol{\varepsilon}_j}{T} \right) \right]' \\
&= \frac{1}{NT^4} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{E}\left[\mathbf{V}_i' \mathbf{F}^0 (\mathbf{F}^0)' \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j' \mathbf{F}^0 (\mathbf{F}^0)' \mathbf{V}_j \right] = \frac{1}{NT^4} \sum_{i=1}^{N} \mathbb{E}\left[\mathbf{V}_i' \mathbf{F}^0 (\mathbf{F}^0)' \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' \mathbf{F}^0 (\mathbf{F}^0)' \mathbf{V}_i \right] \\
&= \frac{1}{NT^4} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{l=1}^{T} \sum_{r=1}^{T} \mathbb{E}\left[\mathbf{v}_{i,t} (\mathbf{f}_t^0)' \mathbf{f}_s^0 \varepsilon_{i,s} \varepsilon_{i,l} (\mathbf{f}_l^0)' \mathbf{f}_r^0 \mathbf{v}_{i,r}' \right] \\
&= \frac{1}{NT^4} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{l=1}^{T} \sum_{r=1}^{T} \mathbb{E}(\varepsilon_{i,s} \varepsilon_{i,l}) \mathbb{E}(\mathbf{v}_{i,t} \mathbf{v}_{i,r}') \mathbb{E}\left[(\mathbf{f}_t^0)' \mathbf{f}_s^0 (\mathbf{f}_l^0)' \mathbf{f}_r^0 \right] = O\left(\frac{1}{T^2}\right)
\end{aligned}$$

where the final line employs the absolute summability of autocovariances in Ass.1 and the bounded fourth moments of factors by assumption 2. Hence,

$$\left\| \frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbf{Q}}_i^{-1} \widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}], V\varepsilon, i} \right\| = O_p(N^{-1/2}) + O_p(T^{-1})$$

Finally, given the well behaved limit of $\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}^0}$, and the fact that the driving terms are identical to (2.105), we have by the same arguments

$$\frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbf{Q}}_i^{-1} \widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}, V\varepsilon, i} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbf{Q}}_i^{-1} \sqrt{N} \left(\frac{\mathbf{V}_i' \mathbf{F}^0}{T} \right) \widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}^0} \left(\frac{(\mathbf{F}^0)' \boldsymbol{\varepsilon}_i}{T} \right) = O_p\left(\frac{1}{T}\right) \tag{2.107}$$

Combining all the results above we come to

$$\left\| \frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbf{Q}}_i^{-1} \widehat{\mathbf{q}}_i \right\| = O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{2.108}$$

Therefore,

$$\begin{aligned}\sqrt{N}(\widehat{\boldsymbol{\beta}}_{mg} - \boldsymbol{\beta}) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{v}_i + o_p(1) \\ &\xrightarrow{d} \mathcal{N}(\mathbf{0}_{k \times 1}, \boldsymbol{\Omega}_v)\end{aligned} \tag{2.109}$$

as $(N, T) \to \infty$.

3 Pairs bootstrap

3.1 The bootstrap resampling matrix W and its properties

Let $\mathbf{w}_i = [w_{i,1}, w_{i,2}, \ldots, w_{i,N}]$ be a $1 \times N$ Boolean selection vector ($[0, 0, 1, \ldots, 0]$) drawn from a multinomial distribution with 1 trial and $k = N$ events, with all event probabilities equal to $p_j = N^{-1}$. It can also be interpreted as Bernoulli random vector, such that it has probability mass function for $\boldsymbol{\theta} \in \mathbb{R}^N$:

$$p_{\mathbf{w}_i} = \begin{cases} \theta_{i,j}, \ w_{i,j} = 1 \text{ and } x_{i,k} = 0 \text{ for } j \neq k, \\ 0, \ \mathbf{w}_i \text{ is not a unit vector.} \end{cases}$$

Note that this implies that $\|\mathbf{w}_i\| = 1$ for all $i = 1, \ldots, N$ and for the scalar elements $w_{i,j}$ in this vector, further using $\theta_{i,j} := \mathbb{P}^*\left(w_{i,j} = 1\right)$, we have

$$\begin{aligned} &\mathbb{E}^*(w_{i,j}) = \mathbb{P}^*\left(w_{i,j} = 1\right) = N^{-1} \\ &\mathbb{P}^*\left(w_{i,j} = 0\right) = 1 - N^{-1} \\ &\mathbb{V}ar^*(w_{i,j}) = N^{-1}(1 - N^{-1}) \\ &\mathbb{C}ov^*(w_{i,i}, w_{i,j}) = -N^{-2} \qquad \text{for} \qquad i \neq j \end{aligned}$$

Next, gather these vectors in the $N \times N$ matrix

$$\underset{(N\times N)}{\mathbf{w}} = [\mathbf{w}_1', \ldots, \mathbf{w}_N']' \tag{3.1}$$

Then, it holds that $\|\mathbf{w}\|^2 = N$ (deterministically) and we can also define the important $1 \times N$ vector

$$\boldsymbol{\iota}_N'\mathbf{w} = \sum_{i=1}^{N} \mathbf{w}_i = [s_1, s_2, \ldots, s_N] = \mathbf{s} \tag{3.2}$$

and additionally the property that

$$\mathbf{w}'\mathbf{w} = diag(\mathbf{s}) \tag{3.3}$$

The scalar elements s_i of the $\mathbf{s}$ vector indicate the frequency with which cross-section i has been resampled in the bootstrap dataset. The s_i give the total resampling counts and hence have the known properties that follow from the multinomial distribution with N trials and N events. That is, following Chatterjee (1998) or Bose and Chatterjee (2002), the sums are $\{s_i \in \mathbb{N}_0 | s_i \leq N\}$, such that $\mathbb{P}^*(s_i = x_i) = N^{-1}$ for some non-random $\{x_i \in \mathbb{N}_0 | x_i \leq N\}$, $\mathbb{E}^*(s_i) = 1$, for all i. Moreover, $\mathbb{V}ar^*(s_i) = 1 - N^{-1}$ and $\mathbb{C}ov^*(s_i, s_j) = -N^{-1}$ for all i and $i \neq j$, respectively. Consequently, $\mathbb{E}^*(s_i^2) = \mathbb{V}ar^*(s_i) + (\mathbb{E}^*(s_i))^2 = 2 - N^{-1}$. Ultimately, we have $\mathbb{P}^*\left(\sum_{i=1}^{N} s_i = N\right) = 1$, because this is how the support of the multinomial is defined. The corresponding matrix which permutes a stack of N different $a-$rowed matrices according to the cross-section re-allocation weights given in $\mathbf{w}$ is then

$$\underset{(aN\times aN)}{\mathbf{W}_a} = (\mathbf{w} \otimes \mathbf{I}_a) \tag{3.4}$$

where $\|\mathbf{W}_a\|^2 = aN$ since

$$\mathbf{W}_a'\mathbf{W}_a = (\mathbf{w}' \otimes \mathbf{I}_a)(\mathbf{w} \otimes \mathbf{I}_a) = (\mathbf{w}'\mathbf{w} \otimes \mathbf{I}_a) = diag(\mathbf{s} \otimes \boldsymbol{\iota}_a') \tag{3.5}$$
$$\boldsymbol{\iota}_{aN}'\mathbf{W}_a'\mathbf{W}_a\boldsymbol{\iota}_{aN} = aN$$

With (1.4) and (3.2) we can then establish the important relation with the averaging matrix

$$\mathbf{A}_a\mathbf{W}_a = N^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{I}_a)(\mathbf{w} \otimes \mathbf{I}_a) = N^{-1}(\boldsymbol{\iota}_N'\mathbf{w} \otimes \mathbf{I}_a) = N^{-1}(\mathbf{s} \otimes \mathbf{I}_a),$$

which illustrates that units making up the CA are weighted according to the CS-counts $\mathbf{s}$, and implies, when pre-multiplied with an $aN \times z$ matrix $\mathbf{G}$ that stacks the N cross-section specific, $a-$rowed matrices $\mathbf{G}_i$

$$\mathbf{A}_a\mathbf{W}_a\mathbf{G} = N^{-1}(\mathbf{s} \otimes \mathbf{I}_a)\mathbf{G} = \frac{1}{N}\sum_{i=1}^{N} s_i\mathbf{G}_i = \overline{\mathbf{G}}_w. \tag{3.6}$$

where we will use a w subscript to denote the dependence on the resampling weights.

Finally, we will repeatedly make use of the key property for multiplication between 'whole cross-section' permutation matrices as in (3.4) when multiplied with matrices repeated over individuals, such as $\underline{\mathbf{G}} = (\mathbf{I}_N \otimes \mathbf{G})$, with $\mathbf{G}$ an $a \times b$, matrix that

$$\mathbf{W}_a\underline{\mathbf{G}} = \underline{\mathbf{G}}\mathbf{W}_b. \tag{3.7}$$

which is easily shown using familiar Kronecker properties

$$\mathbf{W}_a\underline{\mathbf{G}} = (\mathbf{w} \otimes \mathbf{I}_a)(\mathbf{I}_N \otimes \mathbf{G}) = (\mathbf{w}\mathbf{I}_N \otimes \mathbf{I}_a\mathbf{G}) = (\mathbf{I}_N\mathbf{w} \otimes \mathbf{G}\mathbf{I}_b) = (\mathbf{I}_N \otimes \mathbf{G})(\mathbf{w} \otimes \mathbf{I}_b) = \underline{\mathbf{G}}\mathbf{W}_b$$

Note that (3.7) implies commutation for $\mathbf{W}_T\underline{\mathbf{M}}_{\hat{\mathbf{F}}}$ and for instance $\mathbf{W}_T\underline{\mathbf{F}} = \underline{\mathbf{F}}\mathbf{W}_m$.
Next, we establish the important lemma of the resampling counts

Lemma C-1 *(Higher moments of permutation weights)*

a) $\mu_2 = \mathbb{E}^*(s_i^2) = 2 - N^{-1}$

b) $\mu_3 = \mathbb{E}^*(s_i^3) = 5 + r$

c) $\mu_4 = \mathbb{E}^*(s_i^4) = 15 + \rho$

d) μ_5 *and* μ_6 *are both* $O(1)$

e) $0 \leq \mathbb{E}^*(s_i^2, s_j^2) < \infty$ *for* $i \neq j$

f) $\mathbb{V}ar^*(s_i^2) = 11 + q$

g) $\mathbb{E}^*(s_i, s_j) = 1 - N^{-1}$

where r, ρ, q are remainders of order $O(N^{-1})$.

Proof. a) This simply follows from

$$\mu_2 = \mathbb{V}ar^*(s_i) + \mathbb{E}^*(s_i)^2 = 1 - N^{-1} + 1 = 2 - N^{-1},$$

using the known expressions of the variance and the mean.

We derive the higher moments of the multinomial random variable s_i using the Moment Generating Function (MGF). The *joint* MGF of the vector $\mathbf{s} = (s_1, ..., s_N)$ for $\mathbf{t} \in \mathbb{R}^N$ in neighbourhood of $\mathbf{0}_{N\times 1}$ is given by

$$M_{\mathbf{s}}(\mathbf{t}) = \left(\sum_{i=1}^{N} \mathbb{P}^*(s_i = x_i) e^{t_i} \right)^N .$$

The marginal MGF for an arbitrary s_i is obtained by setting $\mathbf{t} = (0, ..., t_i, ..., 0)$, i.e by focusing only on the i-th coordinate. Using this together with the fact that probability masses are $\mathbb{P}^*(s_i = x_i) = N^{-1}\ \forall i$, we obtain

$$M_{s_i}(t_i) = \left(\frac{1}{N} e^{t_i} + \frac{1}{N} \sum_{i=1}^{N-1} 1 \right)^N = \left(\frac{1}{N} e^{t_i} + \frac{N-1}{N} \right)^N .$$

The third and the fourth (μ_3 and μ_4) moments are obtained by taking the respective derivatives and evaluating at $t_i = 0$.

b) We show that for the third moment, the derivative at $t_i = 0$ is:

$$\begin{aligned}
\mu_3 = \left.\frac{d^3 M_{s_i}(t_i)}{dt_i^3}\right|_{t_i=0} &= \Bigg[(N-1)(N-2) \left(\frac{1}{N} e^{t_i} + \frac{N-1}{N} \right)^{N-3} \frac{1}{N^2} e^{3t_i} \\
&+ (N-1) \left(\frac{1}{N} e^{t_i} + \frac{N-1}{N} \right)^{N-2} \frac{2}{N} e^{2t_i} + (N-1) \left(\frac{1}{N} e^{t_i} + \frac{N-1}{N} \right)^{N-2} \frac{1}{N} e^{2t_i} \\
&+ \left(\frac{1}{N} e^{t_i} + \frac{N-1}{N} \right)^{N-1} e^{t_i} \Bigg]_{t_i=0} \\
&= \left(N^2 - 3N + 3\right) N^{-2} + (N-1) 2N^{-1} + 2 - N^{-1} \\
&= 5 + r
\end{aligned}$$

where r is the remainder independent of i and it is of the order $O(N^{-1})$. The result is obtained using the fact that $\left(\frac{1}{N} e^{t_i} + \frac{N-1}{N}\right)^N \Big|_{t_i=0} = 1$ for all N. Also, observe that

$$\mu_2 = \left.\frac{d^2 M_{s_i}(t_i)}{dt_i^2}\right|_{t_i=0} = \left[(N-1) \left(\frac{1}{N} e^{t_i} + \frac{N-1}{N} \right)^{N-2} \frac{1}{N} e^{2t_i} + \left(\frac{1}{N} e^{t_i} + \frac{N-1}{N} \right)^{N-1} e^{t_i} \right]_{t_i=0} = 2 - N^{-1},$$

therefore, the third moment can be alternatively represented, $\mu_3 = a + \mu_2$, where a is a finite constant.

c) We show that for the fourth moment, the derivative at $t_i = 0$ is:

$$\mu_4 = \left.\frac{d^4 M_{s_i}(t_i)}{dt_i^4}\right|_{t_i=0} = \left[(N-1)(N-2)(N-3)\left(\frac{1}{N}e^{t_i} + \frac{N-1}{N}\right)^{N-4}\frac{1}{N^3}e^{4t_i}\right.$$
$$+ (N-1)(N-2)\left(\frac{1}{N}e^{t_i} + \frac{N-1}{N}\right)^{N-3}\frac{3}{N^2}e^{3t_i}$$
$$+ (N-1)(N-2)\left(\frac{1}{N}e^{t_i} + \frac{N-1}{N}\right)^{N-3}\frac{2}{N^2}e^{3t_i} + (N-1)\left(\frac{1}{N}e^{t_i} + \frac{N-1}{N}\right)^{N-2}\frac{4}{N}e^{2t_i}$$
$$+ (N-1)(N-2)\left(\frac{1}{N}e^{t_i} + \frac{N-1}{N}\right)^{N-3}\frac{1}{N^2}e^{3t_i}$$
$$+ (N-1)\left(\frac{1}{N}e^{t_i} + \frac{N-1}{N}\right)^{N-2}\frac{2}{N}e^{2t_i} + (N-1)\left(\frac{1}{N}e^{t_i} + \frac{N-1}{N}\right)^{N-2}\frac{1}{N}e^{2t_i}$$
$$\left. + \left(\frac{1}{N}e^{t_i} + \frac{N-1}{N}\right)^{N-1} e^{t_i}\right]_{t_i=0}$$
$$= \left(N^3 - 6N^2 + 12N - 9\right)N^{-3} + \left(N^2 - 3N + 3\right)3N^{-2} + \left(N^2 - 3N + 3\right)2N^{-2}$$
$$+ (N-1)4N^{-1} + 5 + r$$
$$= 15 + \rho,$$

where the result follows, since the last 4 terms in the sum represent μ_3, which means that the same recursion applies: $\mu_4 = a' + \mu_3$. Similarly, ρ is is the remainder independent of i and it has the order of $O(N^{-1})$.

d) Because it is enough to demonstrate finiteness of μ_5 and μ_6, it is sufficient to use the recursion established in b) and c). In particular, $\mu_5 = a'' + \mu_4$, where a'' is a finite constant and μ_4 is established to be finite. Thus, also, $\mu_6 = a''' + \mu_5$, where a''' is another finite constant and μ_5 is established to be finite.

e) with result c) and the Cauchy-Schwarz inequality, we obtain the following bounds:

$$0 \leq \mathbb{E}^*(s_i^2 s_j^2) \leq \sqrt{\mathbb{E}^*(s_i^4)}\sqrt{\mathbb{E}^*(s_j^4)} < \infty \tag{3.8}$$

for $i \neq j$. This implies that $\mathbb{C}ov^*(s_i^2, s_j^2) < \infty$, as well.

f) Next up, we derive the variance of s_i^2. In particular,

$$\mathbb{V}ar^*(s_i^2) = \mathbb{E}^*(s_i^4) - \mathbb{E}^*(s_i^2)^2 = 15 - 4 + O(N^{-1}) = 11 + q, \tag{3.9}$$

where $q = O(N^{-1})$ and is independent of i. Hence, this variance is bounded for all i and N.

g) Lastly, we deduce $\mathbb{E}^*(s_i s_j)$ for $i \neq q$. From the covariance formula, we obtain

$$\mu_{s_i s_j} = \mathbb{E}^*(s_i s_j) = \mathbb{C}ov^*(s_i, s_j) + \mathbb{E}^*(s_i)\mathbb{E}^*(s_j) = 1 - N^{-1}, \tag{3.10}$$

3.2 Preliminary results

Given the bootstrap resampling matrix $\mathbf{W}_T$ defined in the previous section, the bootstrap observables are

$$\mathbf{y}^* = \mathbf{W}_T\mathbf{y} = [\mathbf{y}_1^{*\prime}, \ldots, \mathbf{y}_N^{*\prime}]' \tag{3.11}$$

$$\mathbf{X}^* = \mathbf{W}_T\mathbf{X} = [\mathbf{X}_1^{*\prime}, \ldots, \mathbf{X}_N^{*\prime}]' \tag{3.12}$$

so that the entire bootstrap data matrix $\mathbf{Z}^* = [\mathbf{y}^*, \mathbf{X}^*]$ can in turn be written as

$$\mathbf{Z}^* = \mathbf{W}_T[\mathbf{y}, \mathbf{X}] = \mathbf{W}_T\mathbf{Z} = [\mathbf{Z}_1^{*\prime}, \ldots, \mathbf{Z}_N^{*\prime}]' \tag{3.13}$$

Making use of (1.5) and (3.6) reveals that the employed CA of the bootstrap observables are

$$\overline{\mathbf{Z}}^* = \mathbf{A}_T\mathbf{Z}^* = \mathbf{A}_T\mathbf{W}_T\mathbf{Z} = \mathbf{A}_T\mathbf{W}_T(\underline{\mathbf{F}}\ \underline{\mathbf{C}} + \mathbf{U}) = \mathbf{F}\ \mathbf{A}_m\mathbf{W}_m\underline{\mathbf{C}} + \mathbf{A}_T\mathbf{W}_T\mathbf{U} = \mathbf{F}\overline{\mathbf{C}}_w + \overline{\mathbf{U}}_w$$

with $\overline{\mathbf{C}}_w = \frac{1}{N}\sum_{i=1}^N s_i\mathbf{C}_i$ and $\overline{\mathbf{U}}_w = \frac{1}{N}\sum_{i=1}^N s_i\mathbf{U}_i$, or generally that $\overline{\mathbf{Z}}^* = \frac{1}{N}\sum_{i=1}^N s_i\mathbf{Z}_i$, i.e. the bootstrap CA are a simple reweighting of the original CA. Factors in the bootstrap world can in turn be expressed as

$$\mathbf{F} = (\overline{\mathbf{Z}}^* - \overline{\mathbf{U}}_w)\overline{\mathbf{C}}_w^{\dagger} \tag{3.14}$$

It will be convenient for the analysis that follows to have expressions for the original data orthogonalized on the bootstrap CA's ($\overline{\mathbf{Z}}^*$). That is, substituting (3.14) into the DGPs of $\mathbf{y}_i, \mathbf{X}_i$,

$$\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{y}_i = \mathbf{M}_{\widehat{\mathbf{F}}^*}[\mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\gamma_i] \tag{3.15}$$

$$\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{X}_i = \mathbf{M}_{\widehat{\mathbf{F}}^*}[\mathbf{V}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\boldsymbol{\Gamma}_i] \tag{3.16}$$

because $\mathbf{M}_{\widehat{\mathbf{F}}^*}\overline{\mathbf{Z}}^* = \mathbf{0}_{T\times(1+k)}$ where $\mathbf{M}_{\widehat{\mathbf{F}}^*} = \mathbf{I}_T - \mathbf{P}_{\widehat{\mathbf{F}}^*}$ and $\mathbf{P}_{\widehat{\mathbf{F}}^*} = \overline{\mathbf{Z}}^*(\overline{\mathbf{Z}}^{*\prime}\overline{\mathbf{Z}}^*)^{\dagger}\overline{\mathbf{Z}}^{*\prime}$. It is then easily seen that since $rk(\mathbf{C}) = m$, $\|\overline{\mathbf{C}}_w - \mathbf{C}\| = O_{p^*}(N^{-1/2})$ and $\|\overline{\mathbf{U}}_w\| = O_{p^*}(N^{-1/2})$ (for a fixed T) also $\overline{\mathbf{Z}}^*$ asymptotically converges to a reduced rank matrix when $m < 1 + k$ and a rotation will need to be employed in the analysis. To that end, let $\mathbf{R}_w = \mathbf{T}\overline{\mathbf{H}}_w\mathbf{D}_N$ be the rotation matrix in the pairs bootstrap world, with $\mathbf{T}$ and $\mathbf{D}_N$ defined as in Section 2 such that $\overline{\mathbf{C}}_w\mathbf{T} = [\overline{\mathbf{C}}_{w,m}, \overline{\mathbf{C}}_{w,-m}]$ are the alternatively weighted $\overline{\mathbf{C}}_m$ and $\overline{\mathbf{C}}_{-m}$ respectively, and similarly for $\overline{\mathbf{U}}_w\mathbf{T} = [\overline{\mathbf{U}}_{w,m}, \overline{\mathbf{U}}_{w,-m}]$. The bootstrap transformation matrix $\overline{\mathbf{H}}_w$ has a similar, but not identical, form as in (2.2)

$$\overline{\mathbf{H}}_w = [\overline{\mathbf{H}}_{w,m}, \overline{\mathbf{H}}_{w,-m}] = \begin{bmatrix} \overline{\mathbf{C}}_{w,m}^{-1} & -\overline{\mathbf{C}}_{w,m}^{-1}\overline{\mathbf{C}}_{w,-m} \\ \mathbf{0}_{(k+1-m)\times m} & \mathbf{I}_{k+1-m} \end{bmatrix} \tag{3.17}$$

such that the rotated quantities in the bootstrap world become

$$\widehat{\mathbf{F}}^{0*} = \overline{\mathbf{Z}}^*\mathbf{R}_w = \widehat{\mathbf{F}}^*\mathbf{R}_w = [\mathbf{F}\overline{\mathbf{C}}_w + \overline{\mathbf{U}}_w]\mathbf{R}_w = \mathbf{F}^0 + \overline{\mathbf{U}}_w^0 \tag{3.18}$$

where $\mathbf{F}^0$ is identical to that in Section 2 and $\overline{\mathbf{U}}_w^0 = \overline{\mathbf{U}}_w\mathbf{R}_w = [\overline{\mathbf{U}}_{w,m}^0, \overline{\mathbf{U}}_{w,-m}^0]$ with $\overline{\mathbf{U}}_{w,m}^0 = \overline{\mathbf{U}}_{w,m}\overline{\mathbf{C}}_{w,m}^{-1}$ and $\overline{\mathbf{U}}_{w,-m}^0 = \sqrt{N}(\overline{\mathbf{U}}_{w,m} - \overline{\mathbf{U}}_{w,m}\overline{\mathbf{C}}_{w,m}^{-1}\overline{\mathbf{C}}_{w,-m})$. Similarly, since $\mathbf{R}_w$ is full rank we have analogously to Section 2 that $\mathbf{P}_{\widehat{\mathbf{F}}^{0*}} = \overline{\mathbf{Z}}^*\mathbf{R}_w(\mathbf{R}_w'\overline{\mathbf{Z}}^{*\prime}\overline{\mathbf{Z}}^*\mathbf{R}_w)^\dagger\mathbf{R}_w'\overline{\mathbf{Z}}^{*\prime} = \overline{\mathbf{Z}}^*(\overline{\mathbf{Z}}^{*\prime}\overline{\mathbf{Z}}^*)^\dagger\overline{\mathbf{Z}}^{*\prime} = \mathbf{P}_{\widehat{\mathbf{F}}^*}$ and analyzing $\mathbf{P}_{\widehat{\mathbf{F}}^{0*}}$ is equivalent to analyzing $\mathbf{P}_{\widehat{\mathbf{F}}^*}$.

Let now $\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^{0*}} = T^{-1}(\widehat{\mathbf{F}}^{0*})'\widehat{\mathbf{F}}^{0*}$ and define also

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}} = \begin{bmatrix} \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}} & \mathbf{0}_{m\times(1+k-m)} \\ \mathbf{0}_{(1+k-m)\times m} & \widehat{\boldsymbol{\Sigma}}_{\mathbf{u}^0_{w,-m}} \end{bmatrix} \tag{3.19}$$

where $\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}} = T^{-1}\mathbf{F}'\mathbf{F}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}^0_{w,-m}} = T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\overline{\mathbf{U}}^0_{w,-m}$. We then have using familiar steps

$$\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}} = T^{-1}\overline{\mathbf{U}}^0_w\widehat{\boldsymbol{\Sigma}}^\dagger_{\widehat{\mathbf{F}}^{0*}}(\overline{\mathbf{U}}^0_w)' + T^{-1}\overline{\mathbf{U}}^0_w\widehat{\boldsymbol{\Sigma}}^\dagger_{\widehat{\mathbf{F}}^{0*}}(\mathbf{F}^0)' + T^{-1}\mathbf{F}^0\widehat{\boldsymbol{\Sigma}}^\dagger_{\widehat{\mathbf{F}}^{0*}}(\overline{\mathbf{U}}^0_w)' + T^{-1}\mathbf{F}^0\left[\widehat{\boldsymbol{\Sigma}}^\dagger_{\widehat{\mathbf{F}}^{0*}} - [T^{-1}(\mathbf{F}^0)'\mathbf{F}^0]^\dagger\right](\mathbf{F}^0)'$$

which given the definitions above corresponds to

$$\begin{aligned}\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}} &= T^{-1}\overline{\mathbf{U}}^0_{w,-m}\widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{u}^0_{w,-m}}(\overline{\mathbf{U}}^0_{w,-m})' + T^{-1}\overline{\mathbf{U}}^0_{w,m}\widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}}(\overline{\mathbf{U}}^0_{w,m})' + T^{-1}\mathbf{F}\widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}}(\overline{\mathbf{U}}^0_{w,m})' + T^{-1}\overline{\mathbf{U}}^0_{w,m}\widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}}\mathbf{F}' \\ &\quad + T^{-1}\widehat{\mathbf{F}}^{0*}\left[\widehat{\boldsymbol{\Sigma}}^\dagger_{\widehat{\mathbf{F}}^{0*}} - \widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}_{w,u}}\right](\widehat{\mathbf{F}}^{0*})'\end{aligned} \tag{3.20}$$

Next we establish the following auxiliary lemmas in the bootstrap world.

Lemma C-2 *Under Ass.1-3, 5 and 6, it follows as $(N,T)\to\infty$ that*

$$T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w = O_{p^*}(N^{-1}) \tag{3.21}$$

$$NT^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w = 2\boldsymbol{\Sigma}_{\mathbf{u},h} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \tag{3.22}$$

$$T^{-1}\mathbf{F}'\overline{\mathbf{U}}_w = O_{p^*}((NT)^{-1/2}) \tag{3.23}$$

$$T^{-1}\overline{\mathbf{U}}_w'\mathbf{U}_i = O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2}) \tag{3.24}$$

where $\boldsymbol{\Sigma}_{\mathbf{u},h} = \boldsymbol{\Sigma}_{\mathbf{u}} + \begin{pmatrix} \boldsymbol{\Omega}_{\boldsymbol{v},\otimes}vec(\boldsymbol{\Sigma}) & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k}\end{pmatrix}$ *and* $\boldsymbol{\Sigma}_{\mathbf{u},h} = \boldsymbol{\Sigma}_{\mathbf{u}}$ *in case* $\boldsymbol{v}_i = \mathbf{0}_{k\times 1}$ *(homogeneous slopes).*

Lemma C-3 *Under Ass.1-6 it follows as $(N,T)\to\infty$ that*

$$T^{-1}(\overline{\mathbf{U}}^0_{w,m})'\overline{\mathbf{U}}^0_{w,m} = O_{p^*}(N^{-1}) \qquad T^{-1}(\overline{\mathbf{U}}^0_{w,m})'\overline{\mathbf{U}}^0_{w,-m} = O_{p^*}(N^{-1/2})$$

$$T^{-1}\mathbf{F}'\overline{\mathbf{U}}^0_{w,m} = O_{p^*}((NT)^{-1/2}) \qquad T^{-1}\mathbf{F}'\overline{\mathbf{U}}^0_{w,-m} = O_{p^*}(T^{-1/2})$$

$$T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}^0_{w,m} = O_{p^*}(N^{-1}) \qquad T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}^0_{w,-m} = O_{p^*}(N^{-1/2})$$

$$T^{-1}(\overline{\mathbf{U}}^0_{w,m})'\mathbf{U}_i = O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2})$$

$$T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\mathbf{U}_i = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$$

$$T^{-1}(\widehat{\mathbf{F}}^{0*})'\overline{\mathbf{U}}_w = O_{p^*}(N^{-1/2}) \qquad T^{-1}(\widehat{\mathbf{F}}^{0*})'\mathbf{U}_i = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$$

moreover, with $\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}^0_{w,-m}} = T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\overline{\mathbf{U}}^0_{w,-m}$

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}^0_{w,-m}} = 2\boldsymbol{\Sigma}_{\mathbf{u}^0_{-m}} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \tag{3.25}$$

$$\widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{u}^0_{w,-m}} = (1/2)\boldsymbol{\Sigma}^\dagger_{\mathbf{u}^0_{-m}} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \tag{3.26}$$

$$\left\|\widehat{\boldsymbol{\Sigma}}^\dagger_{\widehat{\mathbf{F}}^{0*}} - \widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}_{w,u}}\right\| = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \tag{3.27}$$

where $\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}}$ is defined in (3.19) and $\boldsymbol{\Sigma}_{\mathbf{u}^0_{-m}}$ is stated in Lemma B-2.

Proof of Lemma C-2.

Defining $\widetilde{\mathbf{u}}_{i,j,t} = \mathbf{u}_{i,t}\mathbf{u}'_{j,t} - \mathbf{1}_{(j=i)}\boldsymbol{\Sigma}_{\mathbf{u},i}$, with $\mathbf{1}_{(a)}$ the indicator function that returns 1 when condition a inside the brackets is true, and zero otherwise, and with $\boldsymbol{\Sigma}_{\mathbf{u},h,i} = \boldsymbol{\Sigma}_{\mathbf{u},i} + \begin{pmatrix} \boldsymbol{\Omega}_{\boldsymbol{v},\otimes} vec(\boldsymbol{\Sigma}_i) & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} \end{pmatrix}$ and $\boldsymbol{\Omega}_{\boldsymbol{v},\otimes} = \mathbb{E}(\boldsymbol{v}'_i \otimes \boldsymbol{v}'_i)$, we can write

$$
\begin{aligned}
T^{-1}\overline{\mathbf{U}}'_w\overline{\mathbf{U}}_w &= \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \frac{s_i s_j \mathbf{U}'_i\mathbf{U}_j}{T} = \frac{1}{N^2}\sum_{i=1}^N s_i^2\boldsymbol{\Sigma}_{\mathbf{u},h,i} + \frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N s_i s_j\left[\mathbf{U}'_i\mathbf{U}_j - \mathbf{1}_{(i=j)}T\boldsymbol{\Sigma}_{\mathbf{u},h,i}\right] \\
&= \frac{1}{N^2}\sum_{i=1}^N s_i^2\boldsymbol{\Sigma}_{\mathbf{u},h,i} + \frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N s_i s_j\sum_{t=1}^T\widetilde{\mathbf{u}}_{i,j,t} \\
&= \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^N s_i^2\boldsymbol{\Sigma}_{\mathbf{u},h,i}\right) + O_{p^*}\left(\frac{1}{N^{3/2}}\right) + O_{p^*}\left(\frac{1}{N\sqrt{T}}\right) \\
&= N^{-1}2\boldsymbol{\Sigma}_{\mathbf{u},h} + O_{p^*}(N^{-3/2}) + O_{p^*}(N^{-1}T^{-1/2}) \\
&= O_{p^*}(N^{-1})
\end{aligned}
$$

which made use of $\frac{1}{N}\sum_{i=1}^N s_i^2\boldsymbol{\Sigma}_{\mathbf{u},h,i} \to \mu_2\boldsymbol{\Sigma}_{\mathbf{u},h} = 2\boldsymbol{\Sigma}_{\mathbf{u},h} + o(1)$ as $N\to\infty$ by Ass.1,6, where $\boldsymbol{\Sigma}_{\mathbf{u},h} = \boldsymbol{\Sigma}_{\mathbf{u}} + \begin{pmatrix} \boldsymbol{\Omega}_{\boldsymbol{v},\otimes} vec(\boldsymbol{\Sigma}) & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} \end{pmatrix}$ and $\mu_2 = 2 - N^{-1}$ from Lemma C-1, with $\|\boldsymbol{\Sigma}_{\mathbf{u},h}\| = O(1)$ and

$$
\left\|\frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N s_i s_j\sum_{t=1}^T\widetilde{\mathbf{u}}_{i,j,t}\right\| = O_{p^*}\left(\frac{1}{N^{3/2}}\right) + O_{p^*}\left(\frac{1}{N\sqrt{T}}\right)
$$

because the independence of s_i and $\mathbf{u}_{i,t}$ implies independence between s_i and $\widetilde{\mathbf{u}}_{i,j,t}$, and it follows in turn from $\mathbb{E}^*(\widetilde{\mathbf{u}}_{i,j,t}) = \mathbf{0}$ for all i,j,t under Ass.1,6 that $\mathbb{E}^*\left[\sum_{i=1}^N\sum_{j=1}^N s_i s_j\sum_{t=1}^T\widetilde{\mathbf{u}}_{i,j,t}\right] = \mathbf{0}$. Also,

$$
\begin{aligned}
&\mathbb{E}^*\left(\frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N s_i s_j\sum_{t=1}^T\widetilde{\mathbf{u}}_{i,j,t}\right)\left(\frac{1}{N^2T}\sum_{l=1}^N\sum_{r=1}^N s_l s_r\sum_{s=1}^T\widetilde{\mathbf{u}}_{l,r,s}\right)' \\
&= \frac{1}{N^4T^2}\sum_{i=1}^N\sum_{j=1}^N\sum_{l=1}^N\sum_{r=1}^N\mathbb{E}^*(s_i s_j s_l s_r)\sum_{t=1}^T\sum_{s=1}^T\mathbb{E}^*\left(\widetilde{\mathbf{u}}_{i,j,t}\widetilde{\mathbf{u}}'_{l,r,s}\right) \\
&= \frac{1}{N^4T^2}\sum_{i=1}^N\sum_{j=1}^N\mathbb{E}^*(s_i^2s_j^2)\sum_{t=1}^T\sum_{s=1}^T\mathbb{E}^*\left(\widetilde{\mathbf{u}}_{i,j,t}\widetilde{\mathbf{u}}'_{i,j,s}\right) + \frac{1}{N^4T^2}\sum_{i=1}^N\sum_{j=1}^N\mathbb{E}^*(s_i^2s_j^2)\sum_{t=1}^T\sum_{s=1}^T\mathbb{E}^*\left(\widetilde{\mathbf{u}}_{i,j,t}\widetilde{\mathbf{u}}'_{j,i,s}\right) \\
&\quad + \frac{1}{N^4T^2}\sum_{i=1}^N\mathbb{E}^*(s_i^4)\sum_{t=1}^T\sum_{s=1}^T\mathbb{E}^*\left(\widetilde{\mathbf{u}}_{i,i,t}\widetilde{\mathbf{u}}'_{i,i,s}\right) \\
&= O\left(\frac{1}{N^3}\right) + O\left(\frac{1}{N^2T}\right)
\end{aligned}
$$

because s_i has finite moments up to the fourth order from Lemma C-1 and the second equality is obtained by noting that by Ass.1,6 and the cross-section independence of the error terms and slope heterogeneity $\mathbb{E}^*(\widetilde{\mathbf{u}}_{i,j,t}\widetilde{\mathbf{u}}'_{l,r,s}) = \mathbf{0}$ when at least one of the indices (i,j,l,r) differs from the others or when $(j=i, r=l, l\neq i)$. This implies then by definition that either $(l=i, r=j, i\neq j)$, $(l=j, r=i, i\neq j)$ or $(l=r=j=i)$

give non-zero expectation, which are $O(N^2)$, $O(N^2)$ and $O(N)$ sums respectively. To obtain the last line we make use of

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}^*\left(\widetilde{\mathbf{u}}_{i,j,t}\widetilde{\mathbf{u}}_{i,j,s}'\right) = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}^*\left(\widetilde{\mathbf{u}}_{i,j,t}\widetilde{\mathbf{u}}_{i,j,t}'\right) + \frac{1}{T}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\mathbb{E}^*\left(\widetilde{\mathbf{u}}_{i,j,t}\widetilde{\mathbf{u}}_{i,j,s}'\right) = O(1)$$

for all $i \neq j$, because $\mathbb{E}^*\left(\widetilde{\mathbf{u}}_{i,j,t}\widetilde{\mathbf{u}}_{i,j,t}'\right) = \mathbb{E}^*(\mathbf{u}_{i,t}\mathbb{E}^*(\mathbf{u}_{j,t}'\mathbf{u}_{j,t})\mathbf{u}_{i,t}') = O(1)$ since $\mathbf{u}_{i,t}$ and $\mathbf{u}_{j,t}$ are independent with finite second moments, and the second term follows from $\sum_{s\neq t}^{T}\mathbb{E}^*(\widetilde{\mathbf{u}}_{i,j,t}\widetilde{\mathbf{u}}_{i,j,s}') = \mathbb{E}^*(\mathbf{u}_{i,t}\sum_{s\neq t}^{T}\mathbb{E}^*(\mathbf{u}_{j,t}'\mathbf{u}_{j,s})\mathbf{u}_{i,s}') = O(1)$ by $\mathbf{u}_{j,s}$ having absolute summable autocovariances. This with $\left\|T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}^*(\widetilde{\mathbf{u}}_{i,i,t}\widetilde{\mathbf{u}}_{i,i,s}')\right\| = O(1)$ by $\mathbb{E}\left\|\mathbf{u}_{i,t}\right\|^4 < \infty$ from Ass.1 implies

$$\frac{1}{N^4T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}^*(s_i^2s_j^2)\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}^*\left(\widetilde{\mathbf{u}}_{i,j,t}\widetilde{\mathbf{u}}_{i,j,s}'\right) = O\left(\frac{1}{N^2T}\right)$$
$$\frac{1}{N^4T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}^*(s_i^2s_j^2)\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}^*\left(\widetilde{\mathbf{u}}_{i,j,t}\widetilde{\mathbf{u}}_{j,i,s}'\right) = O\left(\frac{1}{N^2T}\right)$$
$$\frac{1}{N^4T^2}\sum_{i=1}^{N}\mathbb{E}^*(s_i^4)\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}^*\left(\widetilde{\mathbf{u}}_{i,i,t}\widetilde{\mathbf{u}}_{i,i,s}'\right) = O\left(\frac{1}{N^3}\right)$$

The second result follows directly from the first

$$NT^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w = N\left[N^{-1}2\mathbf{\Sigma}_{\mathbf{u},h} + O_{p^*}(N^{-3/2}) + O_{p^*}(N^{-1}T^{-1/2})\right] = 2\mathbf{\Sigma}_{\mathbf{u},h} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$$

Next consider

$$T^{-1}\mathbf{F}'\overline{\mathbf{U}}_w = \frac{1}{NT}\sum_{i=1}^{N}s_i\sum_{t=1}^{T}\mathbf{f}_t\mathbf{u}_{t,i}'$$

where given the independence of s_i and Ass.5 we have $\mathbb{E}^*((NT)^{-1}\sum_{i=1}^{N}s_i\sum_{t=1}^{T}\mathbf{f}_t\mathbf{u}_{t,i}') = \mathbf{0}$ and since $\mathbb{E}^*(\mathbf{u}_{t,i}'\mathbf{u}_{t,j}) = 0$ (scalar) for $i \neq j$

$$\begin{aligned}
\mathbb{E}^*\left[T^{-2}\mathbf{F}'\overline{\mathbf{U}}_w\overline{\mathbf{U}}_w'\mathbf{F}\right] &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}^*(s_is_j)\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}^*\left(\mathbf{f}_t\mathbf{u}_{t,i}'\mathbf{u}_{s,j}\mathbf{f}_s'\right)\\
&= \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}^*(s_is_j)\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}^*\left(\mathbf{f}_t\mathbf{u}_{t,i}'\mathbf{u}_{s,i}\mathbf{f}_s'\right)\\
&= \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}^*(s_is_j)\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}^*\left[tr\left(\mathbf{u}_{t,i}\mathbf{u}_{s,i}'\right)\right]\mathbb{E}^*\left(\mathbf{f}_t\mathbf{f}_s'\right)\\
&= O\left(\frac{1}{T}\right)\left\{\frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}^*(s_is_j)\right\}\\
&= O\left(\frac{1}{NT}\right)
\end{aligned}$$

were use was made of $\sum_{s=1}^{T}\mathbb{E}^*\left[tr\left(\mathbf{u}_{t,i}\mathbf{u}_{s,i}'\right)\right]\mathbb{E}^*\left(\mathbf{f}_t\mathbf{f}_s'\right) = O(1)$ since $\mathbf{f}_t$ and $\mathbf{u}_{i,t}$ are stationary with finite summable autocovariances, and second moments are finite for all variables. Hence, $\left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}_w\right\| =$

$O_{p^*}((NT)^{-1/2})$. Similarly, making use of Lemma B-1 and the independence of s_i from all other variables

$$
\begin{aligned}
T^{-1}\overline{\mathbf{U}}_w'\mathbf{U}_i &= \frac{1}{N}\sum_{j=1}^{N} s_j\left(\frac{\mathbf{U}_j'\mathbf{U}_i}{T}\right) = \frac{1}{N}s_i\left(\frac{\mathbf{U}_i'\mathbf{U}_i}{T}\right) + \frac{1}{N}\sum_{j\neq i}^{N} s_j\left(\frac{\mathbf{U}_j'\mathbf{U}_i}{T}\right)\\
&= N^{-1}s_i\left(\mathbf{B}_i'\begin{bmatrix}\sigma_i^2 & \mathbf{0}_{1\times k}\\ \mathbf{0}_{k\times 1} & \mathbf{\Sigma}_i\end{bmatrix}\mathbf{B}_i\right) + N^{-1}s_i O_{p^*}\left(\frac{1}{\sqrt{T}}\right) + O_{p^*}\left(\frac{1}{\sqrt{NT}}\right)\\
&= O_{p^*}\left(\frac{1}{N}\right) + O_{p^*}\left(\frac{1}{N\sqrt{T}}\right) + O_{p^*}\left(\frac{1}{\sqrt{NT}}\right)\\
&= O_{p^*}\left(\frac{1}{N}\right) + O_{p^*}\left(\frac{1}{\sqrt{NT}}\right)
\end{aligned}
$$

because $T^{-1}\mathbf{U}_i'\mathbf{U}_i = \mathbf{B}_i'\begin{bmatrix}\sigma_i^2 & \mathbf{0}_{1\times k}\\ \mathbf{0}_{k\times 1} & \mathbf{\Sigma}_i\end{bmatrix}\mathbf{B}_i + O_p(T^{-1/2}) = O_p(1)$ and $s_i = O(1)$ for all i and N. This establishes (3.24).

Proof of Lemma C-3.

From Lemma C-2 we have $\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\right\| = O_{p^*}(N^{-1})$ and $\left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}_w\right\| = O_{p^*}((NT)^{-1/2})$, such that substituting in the definitions $\overline{\mathbf{U}}^0_{w,-m} = \sqrt{N}\overline{\mathbf{U}}_w\mathbf{T}\overline{\mathbf{H}}_{w,-m}$, $\overline{\mathbf{U}}^0_{w,m} = \overline{\mathbf{U}}_w\mathbf{T}\overline{\mathbf{H}}_{w,m}$ and noting that $\|\mathbf{T}\| = O_{p^*}(1)$, $\|\overline{\mathbf{H}}_w\| = O_{p^*}(1)$ (and therefore also its partitioning) gives

$$
\begin{aligned}
&\left\|T^{-1}(\overline{\mathbf{U}}^0_{w,m})'\overline{\mathbf{U}}^0_{w,m}\right\| = \left\|\overline{\mathbf{H}}_{w,m}'\mathbf{T}'T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\mathbf{T}\overline{\mathbf{H}}_{w,m}\right\| \le \left\|\overline{\mathbf{H}}_{w,m}\right\|^2\|\mathbf{T}\|^2\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\right\| = O_{p^*}(N^{-1})\\
&\left\|T^{-1}(\overline{\mathbf{U}}^0_{m})'\overline{\mathbf{U}}^0_{w,-m}\right\| \le \sqrt{N}\left\|\overline{\mathbf{H}}_{w,m}\right\|\left\|\overline{\mathbf{H}}_{w,-m}\right\|\|\mathbf{T}\|^2\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\right\| = O_{p^*}(N^{-1/2})\\
&\left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}^0_{w,m}\right\| = \left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}_w\mathbf{T}\overline{\mathbf{H}}_{w,m}\right\| \le \left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}_w\right\|\|\mathbf{T}\|\left\|\overline{\mathbf{H}}_{w,m}\right\| = O_{p^*}((NT)^{-1/2})\\
&\left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}^0_{w,-m}\right\| = \sqrt{N}\left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}_w\mathbf{T}\overline{\mathbf{H}}_{w,-m}\right\| \le \sqrt{N}\left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}_w\right\|\|\mathbf{T}\|\left\|\overline{\mathbf{H}}_{w,-m}\right\| = O_{p^*}(T^{-1/2})\\
&\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}^0_{w,m}\right\| = \left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\mathbf{T}\overline{\mathbf{H}}_{w,m}\right\| \le \left\|\overline{\mathbf{H}}_{w,m}\right\|\|\mathbf{T}\|\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\right\| = O_{p^*}(N^{-1})\\
&\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}^0_{w,-m}\right\| = \sqrt{N}\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\mathbf{T}\overline{\mathbf{H}}_{w,-m}\right\| \le \sqrt{N}\left\|\overline{\mathbf{H}}_{w,-m}\right\|\|\mathbf{T}\|\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\right\| = O_{p^*}(N^{-1/2})
\end{aligned}
$$

Similarly making use of $\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{U}_i\right\| = O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2})$ from lemma C-2

$$
\begin{aligned}
&\left\|T^{-1}(\overline{\mathbf{U}}^0_{w,m})'\mathbf{U}_i\right\| \le \left\|\mathbf{T}\overline{\mathbf{H}}_{w,m}\right\|\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{U}_i\right\| = O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2})\\
&\left\|T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\mathbf{U}_i\right\| \le \sqrt{N}\left\|\mathbf{T}\overline{\mathbf{H}}_{w,-m}\right\|\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{U}_i\right\| = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})
\end{aligned}
$$

Next, noting that $\mathbf{F}^0 = [\mathbf{F}, \mathbf{0}_{T\times 1+k-m}]$, $\left\|T^{-1}\mathbf{F}'\mathbf{U}_i\right\| = O_{p^*}(T^{-1/2})$ and making use of the orders in Lemma C-2

$$
\begin{aligned}
&\left\|T^{-1}(\widehat{\mathbf{F}}^{0*})'\overline{\mathbf{U}}_w\right\| \le \left\|T^{-1}(\mathbf{F}^0)'\overline{\mathbf{U}}_w\right\| + \sqrt{N}\left\|\mathbf{T}\overline{\mathbf{H}}\right\|\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\right\| = O_{p^*}(N^{-1/2})\\
&\left\|T^{-1}(\widehat{\mathbf{F}}^{0*})'\mathbf{U}_i\right\| \le \left\|T^{-1}(\mathbf{F}^0)'\mathbf{U}_i\right\| + \sqrt{N}\left\|\mathbf{T}\overline{\mathbf{H}}\right\|\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{U}_i\right\| = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})
\end{aligned}
$$

This establishes the first set of results in the lemma. Next, again making use of $\overline{\mathbf{U}}^0_{w,-m} = \sqrt{N}\overline{\mathbf{U}}_w \mathbf{T}\overline{\mathbf{H}}_{w,-m}$ and substituting in eq.(3.22) of Lemma C-2 establishes that

$$\begin{aligned}
\widehat{\mathbf{\Sigma}}_{\mathbf{u}^0_{w,-m}} &= T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\overline{\mathbf{U}}^0_{w,-m} = \overline{\mathbf{H}}'_{w,-m}\mathbf{T}'NT^{-1}\overline{\mathbf{U}}_w\overline{\mathbf{U}}_w\mathbf{T}\overline{\mathbf{H}}_{w,-m} \\
&= \overline{\mathbf{H}}'_{w,-m}\mathbf{T}'\left(2\mathbf{\Sigma}_{\mathbf{u},h}\right)\mathbf{T}\overline{\mathbf{H}}_{w,-m} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\
&= 2\mathbf{H}'_{-m}\mathbf{T}'\mathbf{\Sigma}_{\mathbf{u},h}\mathbf{T}\mathbf{H}_{-m} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\
&= 2\mathbf{\Sigma}_{\mathbf{u}^0_{-m}} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})
\end{aligned}$$

because $\left\|\overline{\mathbf{H}}_{w,-m} - \mathbf{H}_{-m}\right\| = O_{p^*}(N^{-1/2})$ and where $\mathbf{\Sigma}_{\mathbf{u}^0_{-m}}$ is the $(1+k-m)\times(1+k-m)$ positive definite matrix defined in Lemma B-2. This establishes (3.25). Given that then $rk(\widehat{\mathbf{\Sigma}}_{\mathbf{u}^0_{w,-m}}) - rk(\mathbf{\Sigma}_{\mathbf{u}^0_{-m}}) \stackrel{a.s.}{\to} 0$ it also follows that

$$\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}} = (1/2)\mathbf{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$$

by Theorem 1 in Karabiyik et al. (2017). This is (3.26) of the lemma. Next, consider

$$\widehat{\mathbf{\Sigma}}_{\widehat{\mathbf{F}}^{0*}} - \widehat{\mathbf{\Sigma}}_{\mathbf{F}_{w,u}} = \frac{1}{T}\begin{bmatrix} \mathbf{F}'\overline{\mathbf{U}}^0_{w,m} + (\overline{\mathbf{U}}^0_{w,m})'\mathbf{F} & \mathbf{F}'\overline{\mathbf{U}}^0_{w,-m} \\ (\overline{\mathbf{U}}^0_{w,-m})'\mathbf{F} & \mathbf{0}_{1+k-m\times 1+k-m} \end{bmatrix} + \frac{1}{T}\begin{bmatrix} (\overline{\mathbf{U}}^0_{w,m})'\overline{\mathbf{U}}^0_{w,m} & (\overline{\mathbf{U}}^0_{w,m})'\overline{\mathbf{U}}^0_{w,-m} \\ (\overline{\mathbf{U}}^0_{w,-m})'\overline{\mathbf{U}}^0_{w,m} & \mathbf{0}_{1+k-m\times 1+k-m} \end{bmatrix}$$

where substituting in the results established in the first part of the lemma results in

$$\left\|\widehat{\mathbf{\Sigma}}_{\widehat{\mathbf{F}}^{0*}} - \widehat{\mathbf{\Sigma}}_{\mathbf{F}_{w,u}}\right\| = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$$

Noting then that as in the original setting $rk(\widehat{\mathbf{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}) = 1+k$, and $rk(\widehat{\mathbf{\Sigma}}_{\mathbf{F}_{w,u}}) = rk(\widehat{\mathbf{\Sigma}}_{\mathbf{F}}) + rk(\widehat{\mathbf{\Sigma}}_{\mathbf{u}^0_{w,-m}}) = 1+k$ also as $(N,T)\to\infty$ under Ass.1-2, it follows that $rk(\widehat{\mathbf{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}) - rk(\widehat{\mathbf{\Sigma}}_{\mathbf{F}_{w,u}}) \stackrel{a.s.}{\to} 0$. This allows us to invoke Theorem 1 in Karabiyik et al. (2017) and obtain

$$\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^{0*}} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_{w,u}}\right\| = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$$

which establishes the last statement of the lemma in (3.27).

3.3 Homogeneous Slopes

In the homogeneous slope setting, we impose $\boldsymbol{v}_i = \mathbf{0}_{k\times 1}$ so that $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ and $\mathbf{B}_i = \mathbf{B}$ for all $i = 1, \ldots, N$. In addition, recall that this implies that $\boldsymbol{\Sigma}_{\mathbf{u},h} = \boldsymbol{\Sigma}_{\mathbf{u}}$ in Lemmas C-2 and C-3. The bootstrap CCEP estimator is

$$\widehat{\boldsymbol{\beta}}^* = \left(\mathbf{X}^{*\prime}\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\mathbf{y}^* \tag{3.28}$$

with $\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*} = (\mathbf{I}_N \otimes \mathbf{M}_{\widehat{\mathbf{F}}^*})$. Then, making use of $\mathbf{y}^* = \mathbf{W}_T\mathbf{y}$ and $\mathbf{X}^* = \mathbf{W}_T\mathbf{X}$ and substituting (3.11) and (1.1) into (3.28) gives

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^* &= \left(\mathbf{X}^{*\prime}\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\mathbf{W}_T\mathbf{y} \\
&= \left(\mathbf{X}^{*\prime}\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\mathbf{W}_T\left[\mathbf{X}\boldsymbol{\beta} + \underline{\mathbf{F}}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\right] \\
&= \boldsymbol{\beta} + \left(\mathbf{X}^{*\prime}\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\mathbf{W}_T\left[\underline{\mathbf{F}}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\right]
\end{aligned}$$

such that

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}) = \widehat{\mathbf{Q}}^{*-1}\widehat{\mathbf{q}}^* \tag{3.29}$$

where

$$\widehat{\mathbf{Q}}^* = (NT)^{-1}\mathbf{X}^{*\prime}\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\mathbf{X}^* \tag{3.30}$$

$$\widehat{\mathbf{q}}^* = (NT)^{-1/2}\mathbf{X}^{*\prime}\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\left[\mathbf{W}_T\underline{\mathbf{F}}\boldsymbol{\gamma} + \mathbf{W}_T\boldsymbol{\varepsilon}\right] \tag{3.31}$$

The denominator $\widehat{\mathbf{Q}}^*$ in the bootstrap world can, by making use of (3.5) and (3.7), be expressed as

$$\begin{aligned}
\widehat{\mathbf{Q}}^* &= (NT)^{-1}\mathbf{X}^{*\prime}\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\mathbf{X}^* = (NT)^{-1}\mathbf{X}'\mathbf{W}_T'\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\mathbf{W}_T\mathbf{X} = (NT)^{-1}\mathbf{X}'\mathbf{W}_T'\mathbf{W}_T\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\mathbf{X} \\
&= (NT)^{-1}\mathbf{X}'diag(\mathbf{s}\otimes\boldsymbol{\iota}_T')\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\mathbf{X} \\
&= \frac{1}{NT}\sum_{i=1}^{N} s_i\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{X}_i
\end{aligned}$$

and in turn substituting in (3.16) gives

$$\begin{aligned}
\widehat{\mathbf{Q}}^* &= \frac{1}{NT}\sum_{i=1}^{N} s_i\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{X}_i = \frac{1}{NT}\sum_{i=1}^{N} s_i[\mathbf{V}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\boldsymbol{\Gamma}_i]'\mathbf{M}_{\widehat{\mathbf{F}}^*}[\mathbf{V}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\boldsymbol{\Gamma}_i] \\
&= \widehat{\mathbf{Q}}^*_{\mathbf{M}_{\mathbf{F}^0}} - \widehat{\mathbf{Q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]}
\end{aligned} \tag{3.32}$$

where for a given subscript $\mathbf{A}$ we define the following decomposition

$$\begin{aligned}
\widehat{\mathbf{Q}}^*_{\mathbf{A}} &= \widehat{\mathbf{Q}}^*_{\mathbf{A},VV} - \widehat{\mathbf{Q}}^*_{\mathbf{A},V\Gamma} - (\widehat{\mathbf{Q}}^*_{\mathbf{A},V\Gamma})' + \widehat{\mathbf{Q}}^*_{\mathbf{A},\Gamma\Gamma} \\
\widehat{\mathbf{Q}}^*_{\mathbf{A},VV} &= \frac{1}{NT}\sum_{i=1}^{N} s_i\mathbf{V}_i'\mathbf{A}\mathbf{V}_i \\
\widehat{\mathbf{Q}}^*_{\mathbf{A},V\Gamma} &= \frac{1}{NT}\sum_{i=1}^{N} s_i\mathbf{V}_i'\mathbf{A}\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\boldsymbol{\Gamma}_i \\
\widehat{\mathbf{Q}}^*_{\mathbf{A},\Gamma\Gamma} &= \frac{1}{NT}\sum_{i=1}^{N} s_i\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'\mathbf{A}\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\boldsymbol{\Gamma}_i
\end{aligned}$$

Making use of (3.5), (3.7) and (3.14) yields with similar operations for the bootstrap numerator

$$
\begin{aligned}
\widehat{\mathbf{q}}^* &= (NT)^{-1/2}\mathbf{X}'\mathbf{W}_T'\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\mathbf{W}_T\left[\underline{\mathbf{F}}\underline{\gamma}+\varepsilon\right] = (NT)^{-1/2}\mathbf{X}'\mathbf{W}_T'\mathbf{W}_T\underline{\mathbf{M}}_{\widehat{\mathbf{F}}^*}\left[\underline{\mathbf{F}}\underline{\gamma}+\varepsilon\right] \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}s_i\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\left[\mathbf{F}\gamma_i+\varepsilon_i\right] = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}s_i\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\left[\mathbf{F}(\gamma+\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y)+\varepsilon_i\right] \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}s_i[\mathbf{V}_i-\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\boldsymbol{\Gamma}_i]'\mathbf{M}_{\widehat{\mathbf{F}}^*}[\varepsilon_i-\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y] \\
&= \widehat{\mathbf{q}}_{\mathbf{I}}^*-\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}0}}^*-\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}0}-\mathbf{M}_{\widehat{\mathbf{F}}0*}]}^*
\end{aligned} \tag{3.33}
$$

where we also made use of $\gamma_i = \mathbf{C}_i\mathbf{B}^{-1}\mathbf{q}_y = (\mathbf{C}+\boldsymbol{\eta}_i)\mathbf{B}^{-1}\mathbf{q}_y = \gamma+\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y$ under Ass.3, and $\sum_{i=1}^{N}s_i\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{F}\gamma = N\overline{\mathbf{X}}^*\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{F}\gamma = \mathbf{0}_{k\times 1}$ because $\overline{\mathbf{X}}^* \subset \overline{\mathbf{Z}}^*$. For a given subscript $\mathbf{A}$ we employ the following decomposition

$$
\begin{aligned}
\widehat{\mathbf{q}}_{\mathbf{A}}^* &= \widehat{\mathbf{q}}_{\mathbf{A},V\varepsilon}^*-\widehat{\mathbf{q}}_{\mathbf{A},V\eta}^*-\widehat{\mathbf{q}}_{\mathbf{A},\Gamma\varepsilon}^*+\widehat{\mathbf{q}}_{\mathbf{A},\Gamma\eta}^* \\
\widehat{\mathbf{q}}_{\mathbf{A},V\varepsilon}^* &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}s_i\mathbf{V}_i'\mathbf{A}\varepsilon_i \\
\widehat{\mathbf{q}}_{\mathbf{A},V\eta}^* &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}s_i\mathbf{V}_i'\mathbf{A}\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y \\
\widehat{\mathbf{q}}_{\mathbf{A},\Gamma\varepsilon}^* &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}s_i\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'\mathbf{A}\varepsilon_i \\
\widehat{\mathbf{q}}_{\mathbf{A},\Gamma\eta}^* &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}s_i\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'\mathbf{A}\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y
\end{aligned}
$$

3.3.1 Lemmas

Lemma C-4 *Under Ass.1-5 we have as $(N,T)\to\infty$ such that $\tau_{N,T}\to\tau<\infty$ that*

$$
\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}0}-\mathbf{M}_{\widehat{\mathbf{F}}0*}]}^* \longrightarrow^{p^*} 2\sqrt{\tau}(\mathbf{d}_1+\mathbf{d}_2)+\sqrt{\tau}\mathbf{d}^+ \tag{3.34}
$$

where $\mathbf{d}_1=\mathbf{d}_2=\mathbf{d}^+=\mathbf{0}_{k\times 1}$ when $m=1+k$, whereas if $m<1+k$

$$
\mathbf{d}^+ = (1/2)\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i[\boldsymbol{\beta},\mathbf{I}_k]\mathbf{D}_{-m}[\sigma_i^2,\mathbf{0}_{1\times k}]' \tag{3.35}
$$

and $\mathbf{d}_1$, $\mathbf{d}_2$ are defined in eqs.(2.26)-(2.27) of Lemma B-3, respectively.

Proof of Lemma C-4

For convenience, we restate here the decomposition of $\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]}$

$$
\begin{aligned}
\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]} &= \widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon} - \widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\eta} - \widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\varepsilon} + \widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\eta} \\
\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N s_i \mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\boldsymbol{\varepsilon}_i \\
\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\eta} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N s_i \mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y \\
\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\varepsilon} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N s_i \boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\boldsymbol{\varepsilon}_i \\
\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\eta} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N s_i \boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y
\end{aligned}
$$

First up is

$$
\begin{aligned}
\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0}}],\Gamma\eta} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N s_i \boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y \\
&= \sqrt{\frac{T}{N}}\frac{1}{N}\sum_{i=1}^N s_i \boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})' NT^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y
\end{aligned}
$$

Substituting (3.20) into $\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w$ gives

$$
\begin{aligned}
NT^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w &= NT^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}^0_{w,-m}\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}}T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\overline{\mathbf{U}}_w + NT^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}^0_{w,m}\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_{w,m})'\overline{\mathbf{U}}_w \\
&\quad + NT^{-1}\overline{\mathbf{U}}_w'\mathbf{F}\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_{w,m})'\overline{\mathbf{U}}_w + T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}^0_{w,m}\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}\mathbf{F}'\overline{\mathbf{U}}_w \\
&\quad + NT^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^{0*}\left[\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^{0*}} - \widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}_{w,u}}\right]T^{-1}(\widehat{\mathbf{F}}^{0*})'\overline{\mathbf{U}}_w \\
&= NT^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}^0_{w,-m}\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}}T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\overline{\mathbf{U}}_w + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\
&= 2\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'\boldsymbol{\Sigma}_{\mathbf{u}} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \qquad (3.36)
\end{aligned}
$$

because for the first term in the decomposition we obtain

$$
\begin{aligned}
NT^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}^0_{w,-m}\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}}T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\overline{\mathbf{U}}_w &= \sqrt{N}T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}^0_{w,-m}\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}}\sqrt{N}T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\overline{\mathbf{U}}_w \\
&= (4/2)\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'\boldsymbol{\Sigma}_{\mathbf{u}} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\
&= 2\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'\boldsymbol{\Sigma}_{\mathbf{u}} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})
\end{aligned}
$$

where we have substituted in (3.26) of Lemma C-3 together with the result obtained from $\overline{\mathbf{U}}^0_{w,-m} = \sqrt{N}\overline{\mathbf{U}}_w\mathbf{T}\overline{\mathbf{H}}_{w,-m}$

$$
\sqrt{N}T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}^0_{w,-m} = NT^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\mathbf{T}\overline{\mathbf{H}}_{w,-m} = 2\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \qquad (3.37)
$$

where use was made $NT^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w = 2\boldsymbol{\Sigma}_{\mathbf{u}} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$ in (3.22) and $\overline{\mathbf{H}}_{w,-m} = \mathbf{H}_{-m} +$

$O_{p^*}(N^{-1/2})$. The results of Lemma C-3 reveal for the other terms

$$\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{w,m}^0)'\overline{\mathbf{U}}_w\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,m}^0\right\|^2\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}\right\| = O_{p^*}(N^{-2})$$
$$\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{w,m}^0)'\overline{\mathbf{U}}_w\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}\right\|\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}\right\|\left\|T^{-1}(\overline{\mathbf{U}}_{w,m}^0)'\overline{\mathbf{U}}_w\right\| = O_{p^*}(T^{-1/2}N^{-3/2})$$
$$\left\|T^{-1}\overline{\mathbf{U}}_w'\widehat{\mathbf{F}}^{0*}\left[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}^{\dagger}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}}^{\dagger}\right]T^{-1}(\widehat{\mathbf{F}}^{0*})'\overline{\mathbf{U}}_w\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}_w'\widehat{\mathbf{F}}^{0*}\right\|^2\left\|\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}^{\dagger}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}}^{\dagger}\right\| = O_{p^*}(N^{-3/2})+O_{p^*}(N^{-1}T^{-1/2})$$

Substituting in the result above, noting $\overline{\mathbf{C}}_w^{\dagger} = \mathbf{C}^{\dagger} + O_{p^*}(N^{-1/2})$ and $\tau_{N,T} = T/N = O(1)$ yields

$$\begin{aligned}
\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\eta} &= \sqrt{\frac{T}{N}}\frac{1}{N}\sum_{i=1}^{N}s_i\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'NT^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y \\
&= \sqrt{\tau_{N,T}}\left(\frac{1}{N}\sum_{i=1}^{N}s_i(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\otimes\boldsymbol{\Gamma}_i)'\right)vec\left((\overline{\mathbf{C}}_w^{\dagger})'NT^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\right) \\
&= 2\sqrt{\tau}\mathbf{q}_{xy}'\boldsymbol{\Sigma}_{\boldsymbol{\eta}}'vec\left((\mathbf{C}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}_{\mathbf{u}^0_{-m}}^{\dagger}\mathbf{H}_{-m}'\mathbf{T}'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{C}^{\dagger}\right)+O_{p^*}(N^{-1/2}) \\
&= 2\sqrt{\tau}\mathbf{q}_{xy}'\boldsymbol{\Sigma}_{\boldsymbol{\eta}}'vec\left((\mathbf{C}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{D}_{-m}\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{C}^{\dagger}\right)+O_{p^*}(N^{-1/2}) \qquad (3.38)
\end{aligned}$$

where we recall that $\mathbf{D}_{-m} = \mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}_{\mathbf{u}^0_{-m}}^{\dagger}\mathbf{H}_{-m}'\mathbf{T}'$, $\boldsymbol{\Sigma}_{\boldsymbol{\eta}} = \mathbb{E}(\widetilde{\boldsymbol{\eta}}_i\otimes\widetilde{\boldsymbol{\eta}}_i)$ and $\mathbf{q}_{xy} = (\mathbf{q}_y\otimes\mathbf{q}_x)$ and use was made of

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}s_i(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\otimes\boldsymbol{\Gamma}_i) &= \frac{1}{N}\sum_{i=1}^{N}s_i(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\otimes\boldsymbol{\Gamma}+\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_x) = \frac{1}{N}\sum_{i=1}^{N}(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\otimes\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_x)+O_{p^*}(N^{-1/2}) \\
&= \frac{1}{N}\sum_{i=1}^{N}s_i(\widetilde{\boldsymbol{\eta}}_i\otimes\widetilde{\boldsymbol{\eta}}_i)(\mathbf{q}_y\otimes\mathbf{q}_x)+O_{p^*}(N^{-1/2}) \\
&= \mu_1\boldsymbol{\Sigma}_{\boldsymbol{\eta}}\mathbf{q}_{xy}+O_{p^*}(N^{-1/2})
\end{aligned}$$

because the independence between s_i and $\widetilde{\boldsymbol{\eta}}_i$, as well as the independence of $\widetilde{\boldsymbol{\eta}}_i$ over i imply that under Ass.3 $\frac{1}{N}\sum_{i=1}^{N}s_i\widetilde{\boldsymbol{\eta}}_i'\boldsymbol{\Gamma} = O_{p^*}(N^{-1/2})$ and $\frac{1}{N}\sum_{i=1}^{N}s_i(\widetilde{\boldsymbol{\eta}}_i\otimes\widetilde{\boldsymbol{\eta}}_i) = \mu_1\boldsymbol{\Sigma}_{\boldsymbol{\eta}}+O_{p^*}(N^{-1/2})$ with $\mu_1 = \mathbb{E}^*(s_i) = 1$ from Section 3.1.

Next, making use of $\boldsymbol{\varepsilon}_i = \mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$ gives $\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\varepsilon} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}s_i\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$, and with (3.20) follows the decomposition

$$\begin{aligned}
&T^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i \\
&= T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,-m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}^0_{w,-m}}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{w,-m}^0)'\mathbf{U}_i + T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{w,m}^0)'\mathbf{U}_i \\
&+ T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{w,m}^0)'\mathbf{U}_i + T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}\mathbf{F}'\mathbf{U}_i \\
&+ T^{-1}\overline{\mathbf{U}}_w'\widehat{\mathbf{F}}^{0*}\left[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}^{\dagger}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}}^{\dagger}\right]T^{-1}(\widehat{\mathbf{F}}^{0*})'\mathbf{U}_i \\
&= T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,-m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}^0_{w,-m}}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{w,-m}^0)'\mathbf{U}_i + O_{p^*}(N^{-3/2}) + O_{p^*}(N^{-1}T^{-1/2}) + O_{p^*}(T^{-3/2}) \qquad (3.39)
\end{aligned}$$

because substituting in results from lemmas C-2 and C-3 gives

$$\begin{aligned}
&\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{w,m}^0)'\mathbf{U}_i\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,m}^0\right\|\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}\right\|\left\|T^{-1}(\overline{\mathbf{U}}_{w,m}^0)'\mathbf{U}_i\right\| = O_{p^*}(N^{-2}) + O_{p^*}(N^{-3/2}T^{-1/2})\\
&\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{w,m}^0)'\mathbf{U}_i\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}\right\|\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}\right\|\left\|T^{-1}(\overline{\mathbf{U}}_{w,m}^0)'\mathbf{U}_i\right\| = O_{p^*}(N^{-3/2}T^{-1/2}) + O_{p^*}((NT)^{-1})\\
&\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}\mathbf{F}'\mathbf{U}_i\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,m}^0\right\|\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}\right\|\left\|T^{-1}\mathbf{F}'\mathbf{U}_i\right\| = O_{p^*}(N^{-1}T^{-1/2})\\
&\left\|T^{-1}\overline{\mathbf{U}}_w'\widehat{\mathbf{F}}^{0*}\left[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}^{\dagger} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}}^{\dagger}\right]T^{-1}(\widehat{\mathbf{F}}^{0*})'\mathbf{U}_i\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}_w'\widehat{\mathbf{F}}^{0*}\right\|\left\|\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}^{\dagger} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}}^{\dagger}\right\|\left\|T^{-1}(\widehat{\mathbf{F}}^{0*})'\mathbf{U}_i\right\|\\
&\qquad = O_{p^*}(N^{-3/2}) + O_{p^*}(N^{-1}T^{-1/2}) + O_{p^*}(T^{-3/2})
\end{aligned}$$

and we note that

$$\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,-m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{w,-m}^0}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{w,-m}^0)'\mathbf{U}_i\right\| = O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2}) \tag{3.40}$$

Hence, making use of (3.39), $T/N = O(1)$ and substituting in (3.26) and (3.37)

$$\begin{aligned}
\sqrt{N}T^{-1/2}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i &= T^{-1}\sqrt{N}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,-m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{w,-m}^0}^{\dagger}T^{-1/2}(\overline{\mathbf{U}}_{w.-m}^0)'\mathbf{U}_i + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})\\
&= (2/2)\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}_{\mathbf{u}_{-m}^0}^{\dagger}T^{-1/2}(\overline{\mathbf{U}}_{w,-m}^0)'\mathbf{U}_i + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})\\
&= \boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}_{\mathbf{u}_{-m}^0}^{\dagger}\mathbf{H}_{-m}'\mathbf{T}'\sqrt{N}T^{-1/2}\overline{\mathbf{U}}_w'\mathbf{U}_i + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})\\
&= \boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{D}_{-m}\sqrt{N}T^{-1/2}\overline{\mathbf{U}}_w'\mathbf{U}_i + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})
\end{aligned} \tag{3.41}$$

where as before $\mathbf{D}_{-m} = \mathbf{T}\mathbf{H}_{-m}\boldsymbol{\Sigma}_{\mathbf{u}_{-m}^0}^{\dagger}\mathbf{H}_{-m}'\mathbf{T}'$. Then

$$\begin{aligned}
\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\varepsilon}^* &= \frac{1}{N}\sum_{i=1}^N s_i\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\sqrt{N}T^{-1/2}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\\
&= \frac{1}{N}\sum_{i=1}^N s_i\boldsymbol{\Gamma}_i'(\mathbf{C}_w^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{D}_{-m}\sqrt{N}T^{-1/2}\overline{\mathbf{U}}_w'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})
\end{aligned}$$

Next, with the shorthand $\mathbf{D} = (\mathbf{C}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{D}_{-m}$ and $\overline{\mathbf{U}}_w = N^{-1}s_i\mathbf{U}_i + \frac{1}{N}\sum_{j\neq i}^N s_j\mathbf{U}_j$, the remaining term is

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^N s_i\boldsymbol{\Gamma}_i'\overline{\mathbf{U}}'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y &= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^N s_i^2\boldsymbol{\Gamma}_i'\mathbf{D}(T^{-1}\mathbf{U}_i'\mathbf{U}_i)\mathbf{B}^{-1}\mathbf{q}_y + \frac{1}{NT}\sum_{i=1}^N\sum_{j\neq i}^N s_is_j\boldsymbol{\Gamma}_i'\mathbf{D}\mathbf{U}_j'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right]\\
&= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^N s_i^2\boldsymbol{\Gamma}_i'\mathbf{D}[\sigma_i^2, \mathbf{0}_{1\times k}]'\right] + O_{p^*}(T^{-1/2})
\end{aligned}$$

where the order of the rightmost term on the first line follows from $\tau_{N,T} = O(1)$ and the fact that s_i are independent of $\mathbf{U}_i$ and have finite fourth moments (Lemma C-1) so that their presence does not alter the order established before in (2.34). Also, $T^{-1}\mathbf{U}_i'\mathbf{U}_i = \boldsymbol{\Sigma}_{\mathbf{u},i} + O_{p^*}(T^{-1/2})$ from Ass. 1 and by definition $\boldsymbol{\Sigma}_{\mathbf{u},i}\mathbf{B}^{-1}\mathbf{q}_y = [\sigma_i^2, \mathbf{0}_{1\times k}]'$. Then, substituting in the result leads to

$$\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\varepsilon}^* = \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^N s_i^2\boldsymbol{\Gamma}_i'\mathbf{D}[\sigma_i^2, \mathbf{0}_{1\times k}]'\right] + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$$

and noting that $\frac{1}{N}\sum_{i=1}^{N} s_i^2 \mathbf{\Gamma}_i' \mathbf{D}[\sigma_i^2, \mathbf{0}_{1\times k}]'$ using $\mathbb{E}^*\left(s_i^2\right) = \mu_2 = 2 - N^{-1}$ follows

$$\frac{1}{N}\sum_{i=1}^{N} s_i^2 \mathbf{\Gamma}_i' \mathbf{D}[\sigma_i^2, \mathbf{0}_{1\times k}]' \longrightarrow^{p^*} 2\mathbf{\Gamma}' \mathbf{D}[\sigma^2, \mathbf{0}_{1\times k}]'$$

by a) of lemma C-1 and Ass.1 and 3, we come to conclusion that

$$\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\varepsilon} \longrightarrow^{p^*} 2\sqrt{\tau}\mathbf{\Gamma}'(\mathbf{C}^{\dagger})'\mathbf{\Sigma}_{\mathbf{u}}\mathbf{D}_{-m}[\sigma^2, \mathbf{0}_{1\times k}]' \tag{3.42}$$

For the next term, $\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\eta}$, noting that $\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}_w = \mathbf{q}_x'\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}}_w$ we obtain from substituting in (3.41)

$$\sqrt{N}T^{-1/2}\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\overline{\mathbf{U}} = T^{-1/2}\sqrt{N}\mathbf{q}_x'\mathbf{U}_i'\overline{\mathbf{U}}_w\mathbf{D}_{-m}'\mathbf{\Sigma}_{\mathbf{u}} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$$

which yields when inserted into $\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\eta}$ that

$$\begin{aligned}
\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\eta} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} s_i \mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y = \frac{1}{N}\sum_{i=1}^{N} s_i\sqrt{N}T^{-1/2}\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} s_i\mathbf{q}_x'\mathbf{U}_i'\overline{\mathbf{U}}\mathbf{D}'\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})
\end{aligned}$$

where we again made use of $\mathbf{D} = (\overline{\mathbf{C}}^{\dagger})'\mathbf{\Sigma}_{\mathbf{u}}\mathbf{D}_{-m}$ and obtain using analogous arguments as above

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{q}_x' s_i\mathbf{U}_i'\overline{\mathbf{U}}\mathbf{D}'\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y &= \sqrt{\tau_{N,T}}\mathbf{q}_x'\left[\frac{1}{NT}\sum_{i=1}^{N} s_i^2(T^{-1}\mathbf{U}_i'\mathbf{U}_i)\mathbf{D}'\widetilde{\boldsymbol{\eta}}_i + \frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N} s_i s_j\mathbf{U}_i'\mathbf{U}_j\mathbf{D}'\widetilde{\boldsymbol{\eta}}_i\right]\mathbf{q}_y \\
&= \sqrt{\tau_{N,T}}\mathbf{q}_x'\left[\frac{1}{N}\sum_{i=1}^{N} s_i^2\mathbf{\Sigma}_{\mathbf{u},i}\mathbf{D}'\widetilde{\boldsymbol{\eta}}_i\right]\mathbf{q}_y + O_{p^*}(T^{-1/2}) \\
&= O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})
\end{aligned}$$

because the presence of the s_i again does not change the arguments for the order of the second term as obtained before in (2.36) and by independence and Ass.3 follows

$$\frac{1}{N}\sum_{i=1}^{N} s_i^2\mathbf{\Sigma}_{\mathbf{u},i}\mathbf{D}'\widetilde{\boldsymbol{\eta}}_i = \left[\frac{1}{N}\sum_{i=1}^{N} s_i^2(\widetilde{\boldsymbol{\eta}}_i\otimes\mathbf{\Sigma}_{\mathbf{u},i})'\right]vec(\mathbf{D}') = O_{p^*}(N^{-1/2})$$

Therefore

$$\left\|\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\eta}\right\| = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \tag{3.43}$$

To analyze the final term, note

$$\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} s_i\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\boldsymbol{\varepsilon}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} s_i\mathbf{q}_x'\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$$

For the middle term, making use of results in Lemma C-3 yields

$$\left\|T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}^0_{w,-m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}}T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\mathbf{U}_i\right\| \leq \left\|T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}^0_{w,-m}\right\|^2\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}}\right\| = O_{p^*}(N^{-1}) + O_{p^*}(T^{-1}) + O_{p^*}((NT)^{-1/2})$$

$$\left\|T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}^0_{w,m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_{w,m})'\mathbf{U}_i\right\| \leq \left\|T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}^0_{w,m}\right\|^2\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}\right\| = O_{p^*}(N^{-2}) + O_{p^*}((NT)^{-1}) + O_{p^*}(N^{-3/2}T^{-1/2})$$

$$\left\|T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}^0_{w,m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}\mathbf{F}'\mathbf{U}_i\right\| \leq \left\|T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}^0_{w,m}\right\|\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}\right\|\left\|T^{-1}\mathbf{F}'\mathbf{U}_i\right\| = O_{p^*}(N^{-1}T^{-1/2}) + O_{p^*}(N^{-1/2}T^{-1})$$

$$\left\|T^{-1}\mathbf{U}_i'\mathbf{F}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_{w,m})'\mathbf{U}_i\right\| \leq \left\|T^{-1}\mathbf{U}_i'\mathbf{F}\right\|\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}\right\|\left\|T^{-1}(\overline{\mathbf{U}}^0_{w,m})'\mathbf{U}_i\right\| = O_{p^*}(N^{-1}T^{-1/2}) + O_{p^*}(N^{-1/2}T^{-1})$$

$$\begin{aligned}\left\|T^{-1}\mathbf{U}_i'\widehat{\mathbf{F}}^{0*}\left[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^{0*}} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_{w,u}}\right]T^{-1}(\widehat{\mathbf{F}}^{0*})'\mathbf{U}_i\right\| &\leq \left\|T^{-1}\mathbf{U}'\widehat{\mathbf{F}}^{0*}\right\|\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^{0*}} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_{w,u}}\right\|\left\|T^{-1}(\widehat{\mathbf{F}}^{0*})'\mathbf{U}_i\right\| \\ &= O_{p^*}(N^{-3/2}) + O_{p^*}(N^{-1/2}T^{-1}) + O_{p^*}(N^{-1}T^{-1/2}) + O_{p^*}(T^{-3/2})\end{aligned}$$

so that we obtain for the decomposition

$$\begin{aligned}T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i &= T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}^0_{w,-m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}}T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\mathbf{U}_i + T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}^0_{w,m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_{w,m})'\mathbf{U}_i \\ &\quad + T^{-1}\mathbf{U}_i'\mathbf{F}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}(\overline{\mathbf{U}}^0_{w,m})'\mathbf{U}_i + T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}^0_{w,m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}}T^{-1}\mathbf{F}'\mathbf{U}_i \\ &\quad + T^{-1}\mathbf{U}_i'\widehat{\mathbf{F}}^0\left[\widehat{\mathbf{\Sigma}}^{\dagger}_{\widehat{\mathbf{F}}^{0*}} - \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}_{w,u}}\right]T^{-1}(\widehat{\mathbf{F}}^{0*})'\mathbf{U}_i \\ &= T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}^0_{w,-m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}}T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\mathbf{U}_i \\ &\qquad + O_{p^*}(N^{-3/2}) + O_{p^*}(N^{-1/2}T^{-1}) + O_{p^*}(N^{-1}T^{-1/2}) + O_{p^*}(T^{-3/2}) \\ &= (1/2)T^{-1}\sqrt{N}\mathbf{U}_i'\overline{\mathbf{U}}_w\mathbf{T}\mathbf{H}_{-m}\mathbf{\Sigma}^{\dagger}_{\mathbf{u}^0_{-m}}\mathbf{H}'_{-m}\mathbf{T}'T^{-1}\sqrt{N}\overline{\mathbf{U}}'_w\mathbf{U}_i + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\ &= (1/2)NT^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_w\mathbf{D}_{-m}T^{-1}\overline{\mathbf{U}}'_w\mathbf{U}_i + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \qquad (3.44)\end{aligned}$$

where the before last line substitutes in (3.26). This also gives, with $\overline{\mathbf{U}}^0_{w,-m} = \sqrt{N}\overline{\mathbf{U}}_w\mathbf{T}\overline{\mathbf{H}}_{-m}$, $(NT^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_w) = O_{p^*}(1)$ by Lemma C-2, substituting in (3.26) and using $T/N = O(1)$

$$\begin{aligned}&NT^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i \\ &= NT^{-1}\mathbf{U}_i'\overline{\mathbf{U}}^0_{w,-m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}}T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\mathbf{U}_i + O_{p^*}(N^{-1/2}) + O_{p^*}(N^{1/2}T^{-1}) + O_{p^*}(T^{-1/2}) + O_{p^*}(NT^{-3/2}) \\ &= \sqrt{N}T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}^0_{w,-m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}}\sqrt{N}T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\mathbf{U}_i + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\ &= (NT^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_w)\mathbf{T}\overline{\mathbf{H}}_{-m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}}\overline{\mathbf{H}}'_{-m}\mathbf{T}'(NT^{-1}\overline{\mathbf{U}}'_w\mathbf{U}_i) + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\ &= (1/2)(NT^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_w)\mathbf{D}_{-m}(NT^{-1}\overline{\mathbf{U}}'_w\mathbf{U}_i) + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\ &= (1/2)N^2T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_w\mathbf{D}_{-m}T^{-1}\overline{\mathbf{U}}'_w\mathbf{U}_i + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \qquad (3.45)\end{aligned}$$

which is $O_{p^*}(1)$, and substituting in the result gives

$$\begin{aligned}
\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{q}_x' s_i \mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y \\
&= \sqrt{\tau_{N,T}}\mathbf{q}_x'\left[\frac{1}{N}\sum_{i=1}^{N}s_i NT^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i\right]\mathbf{B}^{-1}\mathbf{q}_y \\
&= \frac{1}{2}\sqrt{\tau_{N,T}}\mathbf{q}_x'\left[\frac{1}{N}\sum_{i=1}^{N}s_i N^2T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_w\mathbf{D}_{-m}T^{-1}(\overline{\mathbf{U}}_w)'\mathbf{U}_i\right]\mathbf{B}^{-1}\mathbf{q}_y + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\
&= \frac{1}{2}\sqrt{\tau_{N,T}}\left[\frac{N}{T^2}\sum_{i=1}^{N}s_i\mathbf{V}_i'\overline{\mathbf{U}}_w\mathbf{D}_{-m}\overline{\mathbf{U}}_w'\boldsymbol{\varepsilon}_i\right] + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\
&= \frac{1}{2}\sqrt{\tau_{N,T}}\left[\frac{N}{T^2}\sum_{i=1}^{N}s_i\mathbf{V}_i'\left(\frac{1}{N}\sum_{j=1}^{N}s_j\mathbf{U}_j\right)\mathbf{D}_{-m}\left(\frac{1}{N}\sum_{l=1}^{N}s_l\mathbf{U}_l\right)'\boldsymbol{\varepsilon}_i\right] + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\
&= \frac{1}{2}\sqrt{\tau_{N,T}}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}s_is_js_l\mathbf{V}_i'\mathbf{U}_j\mathbf{D}_{-m}\mathbf{U}_l'\boldsymbol{\varepsilon}_i\right] + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \qquad (3.46)
\end{aligned}$$

The leading term in the brackets can be rewritten in similar form using the notation introduced in the proof of lemma B-3

$$\begin{aligned}
&\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}s_is_js_l\mathbf{V}_i'\mathbf{U}_j\mathbf{D}_{-m}\mathbf{U}_l'\boldsymbol{\varepsilon}_i \qquad (3.47)\\
&= d_{1,1}\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}s_is_js_l\left[\mathbf{v}_{i,t}\varepsilon_{j,t}\varepsilon_{l,s}\varepsilon_{i,s}+\mathbf{v}_{i,t}\mathbf{v}_{l,s}'\boldsymbol{\beta}\varepsilon_{j,t}\varepsilon_{i,s}+\mathbf{v}_{i,t}\mathbf{v}_{j,t}'\boldsymbol{\beta}\varepsilon_{l,s}\varepsilon_{i,s}+\mathbf{v}_{i,t}\mathbf{v}_{j,t}'\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{v}_{l,s}\varepsilon_{i,s}\right]\right\}\\
&+\sum_{v=2}^{1+k}d_{v,1}\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}s_is_js_l\left[\mathbf{v}_{i,t}\mathrm{v}_{j,t}^{(v-1)}\varepsilon_{l,s}\varepsilon_{i,s}+\mathbf{v}_{i,t}\mathrm{v}_{j,t}^{(v-1)}\mathbf{v}_{l,s}'\boldsymbol{\beta}\varepsilon_{i,s}\right]\right\}\\
&+\sum_{g=2}^{1+k}d_{1,g}\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}s_is_js_l\left[\mathbf{v}_{i,t}\mathrm{v}_{l,s}^{(g-1)}\varepsilon_{i,s}\varepsilon_{j,t}+\mathbf{v}_{i,t}\mathbf{v}_{j,t}'\boldsymbol{\beta}\mathrm{v}_{l,s}^{(g-1)}\varepsilon_{i,s}\right]\right\}\\
&+\sum_{v=2}^{1+k}\sum_{g=2}^{1+k}d_{v,g}\left\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}s_is_js_l\mathbf{v}_{i,t}\mathrm{v}_{j,t}^{(v-1)}\mathrm{v}_{l,s}^{(g-1)}\varepsilon_{i,s}\right\} \qquad (3.48)
\end{aligned}$$

Consider then that we can write the second term in the first set of brackets as

$$\begin{aligned}
&\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}s_is_js_l\mathbf{v}_{i,t}\mathbf{v}_{l,s}'\boldsymbol{\beta}\varepsilon_{j,t}\varepsilon_{i,s}\\
&=\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}s_i^3\boldsymbol{\Sigma}_{i,t,s}\boldsymbol{\beta}\sigma_{i,t,s}+\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}s_is_js_l(\mathbf{v}_{i,t}\mathbf{v}_{l,s}'-\mathbf{1}_{(l=i)}\boldsymbol{\Sigma}_{i,t,s})\boldsymbol{\beta}(\varepsilon_{j,t}\varepsilon_{i,s}-\mathbf{1}_{(j=i)}\sigma_{i,t,s})\\
&=\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}s_i^3\boldsymbol{\Sigma}_{i,t,s}\boldsymbol{\beta}\sigma_{i,t,s}+\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}s_is_js_l\mathbf{V}_{i,l}^{t,s}\boldsymbol{\beta}e_{i,j}^{t,s}
\end{aligned}$$

where $\boldsymbol{\Sigma}_{i,t,s}=\mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}_{i,s}')$, $\sigma_{i,t,s}=\mathbb{E}(\boldsymbol{\varepsilon}_{i,t}\boldsymbol{\varepsilon}_{i,s})$, $\mathbf{V}_{i,l}^{t,s}=\mathbf{v}_{i,t}\mathbf{v}_{l,s}'-\mathbf{1}_{(l=i)}\boldsymbol{\Sigma}_{i,t,s}$ and $e_{i,j}^{t,s}=\varepsilon_{j,t}\varepsilon_{i,s}-\mathbf{1}_{(j=i)}\sigma_{i,t,s}$. Given that $s_is_js_l$, $\mathbf{V}_{i,l}^{t,s}$ and $e_{i,j}^{t,s}$ are mutually independent for all i,j,l,t,s

$$\mathbb{E}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}s_is_js_l\mathbf{V}_{i,l}^{t,s}\boldsymbol{\beta}e_{i,j}^{t,s}\right]=\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}(s_is_js_l)\mathbb{E}(\mathbf{V}_{i,l}^{t,s})\boldsymbol{\beta}\mathbb{E}(e_{i,j}^{t,s})=\mathbf{0}_{k\times 1}$$

and similarly, the independence of $s_i s_j s_l s_m s_n s_o$ and these weights having finite sixth moments by d) of Lemma C-1 implies

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} s_i s_j s_l \mathbf{V}_{i,l}^{t,s}\boldsymbol{\beta} e_{i,j}^{t,s}\right)\left(\frac{1}{NT^2}\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{o=1}^{N}\sum_{r=1}^{T}\sum_{q=1}^{T} s_m s_n s_o \mathbf{V}_{m,o}^{r,q}\boldsymbol{\beta} e_{m,n}^{r,q}\right)' \\
&= \frac{1}{N^2T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{o=1}^{N}\mathbb{E}\left(s_i s_j s_l s_m s_n s_o\right)\mathbb{E}\left(\mathbf{V}_{i,l}^{t,s}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}_{m,o}^{r,q\prime}\right)\mathbb{E}\left(e_{i,j}^{t,s}e_{m,n}^{r,q}\right) \\
&= O\left(\frac{1}{N}\right) + O\left(\frac{1}{T}\right)
\end{aligned}$$

by the exact same arguments as those for (2.40) in the proof of Lemma B-3. This leads to

$$\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} s_i s_j s_l \mathbf{V}_{i,l}^{t,s}\boldsymbol{\beta} e_{i,j}^{t,s}\right\| = O_{p^*}\left(\frac{1}{\sqrt{N}}\right) + O_{p^*}\left(\frac{1}{\sqrt{T}}\right)$$

and therefore

$$\begin{aligned}
\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} s_i s_j s_l \mathbf{v}_{i,t}\mathbf{v}_{l,s}'\boldsymbol{\beta}\varepsilon_{j,t}\varepsilon_{i,s} &= \frac{1}{N}\sum_{i=1}^{N} s_i^3 \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{\Sigma}_{i,t,s}\boldsymbol{\beta}\sigma_{i,t,s} + O_{p^*}\left(\frac{1}{\sqrt{N}}\right) + O_{p^*}\left(\frac{1}{\sqrt{T}}\right) \\
&= O_{p^*}\left(\frac{1}{\sqrt{N}}\right) + O_{p^*}\left(\frac{1}{\sqrt{T}}\right)
\end{aligned}$$

since $\left\|T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{\Sigma}_{i,t,s}\boldsymbol{\beta}\sigma_{i,t,s}\right\| = O(1)$ for all i due to $\varepsilon_{i,t}, \mathbf{v}_{i,t}$ having absolute summable autocovariances. The same arguments yield also

$$\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} s_i s_j s_l \mathbf{v}_{i,t}\mathbf{v}_{l,s}^{(g-1)}\varepsilon_{i,s}\varepsilon_{j,t}\right\| = O_{p^*}\left(\frac{1}{\sqrt{N}}\right) + O_{p^*}\left(\frac{1}{\sqrt{T}}\right)$$

Next, for combinations such as $\mathbf{v}_{i,t}\mathbf{v}_{j,t}'\boldsymbol{\beta}\varepsilon_{l,s}\varepsilon_{i,s}$ we can write as before with $\mathbf{V}_{i,j}^{t} = \mathbf{v}_{i,t}\mathbf{v}_{j,t}' - \mathbf{1}_{(j=i)}\boldsymbol{\Sigma}_i$ and $e_{i,l}^{s} = \varepsilon_{i,s}\varepsilon_{l,s} - \mathbf{1}_{(l=i)}\sigma_i^2$

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} s_i s_j s_l \mathbf{v}_{i,t}\mathbf{v}_{j,t}'\boldsymbol{\beta}\varepsilon_{l,s}\varepsilon_{i,s} = \frac{1}{N}\sum_{i=1}^{N} s_i^3\boldsymbol{\Sigma}_i\boldsymbol{\beta}\sigma_i^2 + \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} s_i s_j s_l \mathbf{V}_{i,j}^{t}\boldsymbol{\beta} e_{i,l}^{s}$$

where again by independence of $s_i s_j s_l$, $\mathbf{V}_{i,j}^{t}$ and $e_{i,l}^{s}$

$$\mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} s_i s_j s_l \mathbf{V}_{i,j}^{t}\boldsymbol{\beta} e_{i,l}^{s}\right] = \mathbf{0}_{k\times 1}$$

and the independence of the s_i's together with them having finite sixth moments implies that by the exact same arguments as for (2.41) in the proof of Lemma B-3,

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} s_i s_j s_l \mathbf{V}_{i,j}^{t}\boldsymbol{\beta} e_{i,l}^{s}\right]\left[\frac{1}{NT^2}\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{o=1}^{N}\sum_{r=1}^{T}\sum_{q=1}^{T} s_m s_n s_o \mathbf{V}_{m,n}^{r}\boldsymbol{\beta} e_{m,o}^{q}\right]' \\
&= \frac{1}{N^2T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{o=1}^{N}\mathbb{E}\left(s_i s_j s_l s_m s_n s_o\right)\mathbb{E}\left(\mathbf{V}_{i,j}^{t}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}_{m,n}^{r\prime}\right)\mathbb{E}\left(e_{i,l}^{s}e_{m,o}^{q}\right) \\
&= O\left(\frac{1}{T}\right) + O\left(\frac{1}{N}\right)
\end{aligned}$$

Hence,

$$\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} s_i s_j s_l \mathbf{V}_{i,j}^{t} \boldsymbol{\beta} e_{i,l}^{s} \right\| = O_{p^*}\left(\frac{1}{\sqrt{N}}\right) + O_{p^*}\left(\frac{1}{\sqrt{T}}\right)$$

which substituted in leads to

$$\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} s_i s_j s_l \mathbf{v}_{i,t} \mathbf{v}_{j,t}' \boldsymbol{\beta} \varepsilon_{l,s} \varepsilon_{i,s} = \frac{1}{N} \sum_{i=1}^{N} s_i^3 \boldsymbol{\Sigma}_i \boldsymbol{\beta} \sigma_i^2 + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$$

Then, from $\frac{1}{N} \sum_{i=1}^{N} s_i^3 \boldsymbol{\Sigma}_i \boldsymbol{\beta} \sigma_i^2 \to^{p^*} \mu_3 \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Sigma}_i \boldsymbol{\beta} \sigma_i^2$ as $N \to \infty$, where $\mu_3 = \mathbb{E}^*(s_i^3) = 5 + O(N^{-1})$ by b) in Lemma C-1, follows the conclusion

$$\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} s_i s_j s_l \mathbf{v}_{i,t} \mathbf{v}_{j,t}' \boldsymbol{\beta} \varepsilon_{l,s} \varepsilon_{i,s} \longrightarrow^{p^*} 5 \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Sigma}_i \boldsymbol{\beta} \sigma_i^2$$

and by the same reasoning also

$$\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} s_i s_j s_l \mathbf{v}_{i,t} \mathrm{v}_{j,t}^{(v-1)} \varepsilon_{l,s} \varepsilon_{i,s} \longrightarrow^{p^*} 5 \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Sigma}_i \mathbf{q}_{(v-1)} \sigma_i^2$$

Finally, the independence of the s_i, s_j, s_l from the other variables, with $\mathbb{E}(s_i^6) < \infty$, allows us to use the same arguments as for (2.42) in the proof of Lemma B-3 and get

$$\begin{aligned}
&\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} s_i s_j s_l \mathbf{v}_{i,t} \varepsilon_{j,t} \varepsilon_{l,s} \varepsilon_{i,s} \right\| = O_{p^*}\left(\frac{1}{\sqrt{T}}\right) \\
&\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} s_i s_j s_l \mathbf{v}_{i,t} \mathbf{v}_{j,t}' \boldsymbol{\beta} \boldsymbol{\beta}' \mathbf{v}_{l,s} \varepsilon_{i,s} \right\| = O_{p^*}\left(\frac{1}{\sqrt{T}}\right) \\
&\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} s_i s_j s_l \mathbf{v}_{i,t} \mathrm{v}_{j,t}^{(v-1)} \mathbf{v}_{l,s}' \boldsymbol{\beta} \varepsilon_{i,s} \right\| = O_{p^*}\left(\frac{1}{\sqrt{T}}\right) \\
&\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} s_i s_j s_l \mathbf{v}_{i,t} \mathbf{v}_{j,t}' \boldsymbol{\beta} \mathrm{v}_{l,s}^{(g-1)} \varepsilon_{i,s} \right\| = O_{p^*}\left(\frac{1}{\sqrt{T}}\right) \\
&\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} s_i s_j s_l \mathbf{v}_{i,t} \mathrm{v}_{j,t}^{(v-1)} \mathrm{v}_{l,s}^{(g-1)} \varepsilon_{i,s} \right\| = O_{p^*}\left(\frac{1}{\sqrt{T}}\right)
\end{aligned} \tag{3.49}$$

Combining then all these results in (3.47) gives

$$\begin{aligned}
\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} s_i s_j s_l \mathbf{V}_i' \mathbf{U}_j \mathbf{D}_{-m} \mathbf{U}_l' \boldsymbol{\varepsilon}_i = d_{1,1} &\left\{ \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} s_i s_j s_l \left[\mathbf{v}_{i,t} \mathbf{v}_{j,t}' \boldsymbol{\beta} \varepsilon_{l,s} \varepsilon_{i,s} \right] \right\} \\
&+ \sum_{v=2}^{1+k} d_{v,1} \left\{ \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} s_i s_j s_l \left[\mathbf{v}_{i,t} \mathrm{v}_{j,t}^{(v-1)} \varepsilon_{l,s} \varepsilon_{i,s} \right] \right\} \\
&+ O_{p^*}\left(\frac{1}{\sqrt{T}}\right) + O_{p^*}\left(\frac{1}{\sqrt{N}}\right)
\end{aligned} \tag{3.50}$$

and subsequently

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} s_i s_j s_l \mathbf{V}_i'\mathbf{U}_j\mathbf{D}_{-m}\mathbf{U}_l'\boldsymbol{\varepsilon}_i \longrightarrow^{p^*} 5d_{1,1}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i\boldsymbol{\beta}\sigma_i^2 + 5\sum_{v=2}^{1+k} d_{v,1}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i\mathbf{q}_{v-1}\sigma_i^2$$

$$= 5\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i[\boldsymbol{\beta},\mathbf{I}_k]\mathbf{D}_{-m}[\sigma_i^2,\mathbf{0}_{1\times k}]'$$

so that from insertion into (3.46) follows

$$\begin{aligned}\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon} &\longrightarrow^{p^*} \frac{5}{2}\sqrt{\tau}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i[\boldsymbol{\beta},\mathbf{I}_k]\mathbf{D}_{-m}[\sigma_i^2,\mathbf{0}_{1\times k}]' \\ &= 2\sqrt{\tau}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i[\boldsymbol{\beta},\mathbf{I}_k]\mathbf{D}_{-m}[\sigma_i^2,\mathbf{0}_{1\times k}]' \\ &+ (1/2)\sqrt{\tau}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i[\boldsymbol{\beta},\mathbf{I}_k]\mathbf{D}_{-m}[\sigma_i^2,\mathbf{0}_{1\times k}]', \end{aligned} \tag{3.51}$$

where we will use the second term to formulate the bias as in that of Lemma C-4. Combining then (3.38), (3.42), (3.43) and (3.51) into $\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]} = \widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon} - \widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\eta} - \widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\varepsilon} + \widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\eta}$, and recalling from their definitions in (2.26)-(2.27) of Lemma B-3 that

$$\begin{aligned}\mathbf{d}_1 &= \mathbf{q}_{xy}'\boldsymbol{\Sigma}_{\boldsymbol{\eta}}' vec\left((\mathbf{C}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{D}_{-m}\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{C}^{\dagger}\right) \\ \mathbf{d}_2 &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i[\boldsymbol{\beta},\mathbf{I}_k]\mathbf{D}_{-m}[\sigma_i^2,\mathbf{0}_{1\times k}]' - \boldsymbol{\Gamma}'(\mathbf{C}^{\dagger})'\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{D}_{-m}[\sigma^2,\mathbf{0}_{1\times k}]'\end{aligned}$$

we come to the conclusion that

$$\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]} \longrightarrow^{p^*} 2\sqrt{\tau}(\mathbf{d}_1+\mathbf{d}_2) + \sqrt{\tau}\mathbf{d}^+$$

with $\mathbf{d}^+ = (1/2)\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i[\boldsymbol{\beta},\mathbf{I}_k]\mathbf{D}_{-m}[\sigma_i^2,\mathbf{0}_{1\times k}]'$, as needed to be shown.

It remains to show that $\mathbf{d}_1 = \mathbf{d}_2 = \mathbf{d}^+ = \mathbf{0}$ when $m = 1+k$. To see this, note first that by Lemma C-2

$$\begin{aligned}T^{-1}(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^* &= \overline{\mathbf{C}}_w' T^{-1}\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}_w + \overline{\mathbf{C}}_w' T^{-1}\mathbf{F}'\overline{\mathbf{U}}_w + T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}\overline{\mathbf{C}}_w + T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w \\ &= \overline{\mathbf{C}}_w' T^{-1}\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}_w + O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2})\end{aligned}$$

and also $rk(T^{-1}(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*) - rk(\overline{\mathbf{C}}_w' T^{-1}\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}_w) \xrightarrow{a.s.} 0$ implies that Theorem 1 in Karabiyik et al. (2017) can directly be applied to yield $\left\|(T^{-1}(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*)^{\dagger} - (\overline{\mathbf{C}}_w' T^{-1}\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}_w)^{\dagger}\right\| = O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2})$ and $\left\|(T^{-1}(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*)^{\dagger}\right\| = O_{p^*}(1)$, thereby foregoing the need for a rotation. Hence, since if $m = 1+k$ we get using the earlier definition that $\mathbf{R}_w = \overline{\mathbf{C}}_w^{-1}$ such that $\mathbf{M}_{\mathbf{F}^0} = \mathbf{M}_{\mathbf{F}}$ and $\mathbf{M}_{\widehat{\mathbf{F}}^{0*}} = \mathbf{I}_T - \overline{\mathbf{Z}}^*\mathbf{R}_w(\mathbf{R}_w'\overline{\mathbf{Z}}^{*\prime}\overline{\mathbf{Z}}^*\mathbf{R}_w)^{\dagger}\mathbf{R}_w'\overline{\mathbf{Z}}^{*\prime}$, but that by the properties of the generalized inverse $\mathbf{M}_{\mathbf{F}} = \mathbf{I}_T - \mathbf{F}(\mathbf{F}'\mathbf{F})^{\dagger}\mathbf{F}' = \mathbf{I}_T - \mathbf{F}\overline{\mathbf{C}}_w(\overline{\mathbf{C}}_w' T^{-1}\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}_w)^{\dagger}\overline{\mathbf{C}}_w'\mathbf{F}' = \mathbf{M}_{\mathbf{F}\overline{\mathbf{C}}_w}$ and $\mathbf{M}_{\widehat{\mathbf{F}}^{0*}} = \mathbf{M}_{\widehat{\mathbf{F}}^*}$, the components of which are all well behaved, equation (3.20) can be simplified and analyzed as

$$\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}} = T^{-1}\overline{\mathbf{U}}_w(T^{-1}(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*)^\dagger\overline{\mathbf{U}}_w' + T^{-1}\overline{\mathbf{U}}_w(T^{-1}(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*)^\dagger\overline{\mathbf{C}}_w'\mathbf{F}' + T^{-1}\mathbf{F}\overline{\mathbf{C}}_w(T^{-1}(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*)^\dagger\overline{\mathbf{U}}_w' \\ + T^{-1}\mathbf{F}\overline{\mathbf{C}}_w[(T^{-1}(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*)^\dagger - (\overline{\mathbf{C}}_w'T^{-1}\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}_w)^\dagger]\overline{\mathbf{C}}_w'\mathbf{F}' \tag{3.52}$$

Then, substituting in this decomposition yields

$$\begin{aligned}\left\|\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\eta}\right\| &\leq \sqrt{\tau_{N,T}}\left\|\frac{1}{N}\sum_{i=1}^N s_i(\mathbf{q}_y'\widetilde{\boldsymbol{\eta}}_i'\otimes\boldsymbol{\Gamma}_i')\right\|\left\|\overline{\mathbf{C}}_w^\dagger\right\|^2 N\left\|T^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\right\| \\ &= O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2})\end{aligned}$$

because by application of the results in lemma C-2 we now obtain

$$\begin{aligned}\left\|T^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\right\| \leq{}& \left\|\frac{\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w}{T}\right\|^2\left\|\left(\frac{(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*}{T}\right)^\dagger\right\| + 2\left\|\frac{\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w}{T}\right\|\left\|\left(\frac{(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*}{T}\right)^\dagger\right\|\left\|\overline{\mathbf{C}}_w\right\|\left\|\frac{\mathbf{F}'\overline{\mathbf{U}}_w}{T}\right\| \\ &+ \left\|\overline{\mathbf{C}}_w\right\|^2\left\|\frac{\mathbf{F}'\overline{\mathbf{U}}_w}{T}\right\|^2\left\|\left(\frac{(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*}{T}\right)^\dagger - \left(\overline{\mathbf{C}}_w'\frac{\mathbf{F}'\mathbf{F}}{T}\overline{\mathbf{C}}_w\right)^\dagger\right\| \\ ={}& O_{p^*}(N^{-2}) + O_{p^*}(N^{-3/2}T^{-1/2})\end{aligned}$$

Next up are $\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\varepsilon}$ and $\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\eta}$. Making use of $\boldsymbol{\varepsilon}_i = \mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$, $\mathbf{V}_i = \mathbf{U}_i\mathbf{q}_x$ and $T/N = O(1)$

$$\begin{aligned}\left\|\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\varepsilon}\right\| &\leq \sqrt{\tau_{N,T}}\frac{1}{N}\sum_{i=1}^N s_i\left\|\boldsymbol{\Gamma}_i\right\|\left\|\overline{\mathbf{C}}_w^\dagger\right\| N\left\|T^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i\right\|\left\|\mathbf{B}^{-1}\mathbf{q}_y\right\| \\ &= O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2}) \\ \left\|\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\eta}\right\| &\leq \sqrt{\tau_{N,T}}\frac{1}{N}\sum_{i=1}^N s_i\left\|\mathbf{q}_x\right\| N\left\|T^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i\right\|\left\|\overline{\mathbf{C}}_w^\dagger\right\|\left\|\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\right\| \\ &= O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2})\end{aligned}$$

since from (3.52) and lemma C-2 follows

$$\begin{aligned}&\left\|T^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i\right\| \\ &\leq \left\|\frac{\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w}{T}\right\|\left\|\left(\frac{(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*}{T}\right)^\dagger\right\|\left\|\frac{\overline{\mathbf{U}}_w'\mathbf{U}_i}{T}\right\| + \left\|\frac{\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w}{T}\right\|\left\|\left(\frac{(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*}{T}\right)^\dagger\right\|\left\|\overline{\mathbf{C}}_w\right\|\left\|\frac{\mathbf{F}'\mathbf{U}_i}{T}\right\| \\ &+ \left\|\frac{\mathbf{F}'\overline{\mathbf{U}}_w}{T}\right\|\left\|\overline{\mathbf{C}}_w\right\|\left\|\left(\frac{(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*}{T}\right)^\dagger\right\|\left\|\frac{\overline{\mathbf{U}}_w'\mathbf{U}_i}{T}\right\| + \left\|\overline{\mathbf{C}}_w\right\|^2\left\|\frac{\mathbf{F}'\overline{\mathbf{U}}_w}{T}\right\|\left\|\left(\frac{(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*}{T}\right)^\dagger - \left(\overline{\mathbf{C}}_w'\frac{\mathbf{F}'\mathbf{F}}{T}\overline{\mathbf{C}}_w\right)^\dagger\right\|\left\|\frac{\mathbf{F}'\mathbf{U}_i}{T}\right\| \\ &= O_{p^*}(N^{-2}) + O_{p^*}(N^{-3/2}T^{-1/2})\end{aligned}$$

Finally, for $\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon}$ we find

$$\left\|\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon}\right\| \leq \sqrt{\tau_{N,T}}\frac{1}{N}\sum_{i=1}^N s_i\left\|\mathbf{q}_x\right\| N\left\|T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i\right\|\left\|\mathbf{B}^{-1}\mathbf{q}_y\right\| = O_{p^*}(N^{-1}) + O_{p^*}(T^{-1/2})$$

from $T/N = O(1)$ and

$$\begin{aligned}
&\left\| T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i \right\| \\
&\leq \left\| \frac{\overline{\mathbf{U}}_w'\mathbf{U}_i}{T} \right\|^2 \left\| \left(\frac{(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*}{T} \right)^{\dagger} \right\| + 2 \left\| \frac{\overline{\mathbf{U}}_w'\mathbf{U}_i}{T} \right\| \left\| \left(\frac{(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*}{T} \right)^{\dagger} \right\| \|\overline{\mathbf{C}}_w\| \left\| \frac{\mathbf{F}'\mathbf{U}_i}{T} \right\| \\
&\quad + \|\overline{\mathbf{C}}_w\|^2 \left\| \frac{\mathbf{F}'\mathbf{U}_i}{T} \right\|^2 \left\| \left(\frac{(\overline{\mathbf{Z}}^*)'\overline{\mathbf{Z}}^*}{T} \right)^{\dagger} - \left(\overline{\mathbf{C}}_w' \frac{\mathbf{F}'\mathbf{F}}{T} \overline{\mathbf{C}}_w \right)^{\dagger} \right\| \\
&= O_{p^*}(N^{-2}) + O_{p^*}(N^{-1}T^{-1/2}) + O_{p^*}(N^{-1/2}T^{-1})
\end{aligned}$$

Hence, by combining results we have when $m = 1 + k$ as $(N, T) \to \infty$

$$\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]} \longrightarrow^{p^*} \mathbf{0}_{k\times 1}$$

which translates to $\mathbf{d}_1 = \mathbf{d}_2 = \mathbf{d}^+ = \mathbf{0}_{k\times 1}$ in eq.(3.34) of the lemma, as needed to be shown.

Lemma C-5 *Under Ass.1-5 we have*

$$\left\| \widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0}} \right\| = O_{p^*}(T^{-1/2}) \tag{3.53}$$

as $(N, T) \to \infty$.

Proof of Lemma C-5

The proof of this lemma is nearly identical to that of lemma B-4. That is, since $\widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0}} = \widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon} - \widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},V\eta} - \widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},\Gamma\varepsilon} + \widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},\Gamma\eta}$, and noting that $\mathbf{P}_{\mathbf{F}^0} = \mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}$ and $\mathbf{F}^0 = [\mathbf{F}, \mathbf{0}_{T,1+k-m}]$, rewriting the last term gives

$$\begin{aligned}\widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},\Gamma\eta} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} s_i\mathbf{\Gamma}_i'(\overline{\mathbf{C}}^{\dagger}_w)'\overline{\mathbf{U}}_w'\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}_w\overline{\mathbf{C}}^{\dagger}_w\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y \\ &= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N} s_i(\mathbf{q}_y'\widetilde{\boldsymbol{\eta}}_i'\otimes\mathbf{\Gamma}_i')\right]\left[\overline{\mathbf{C}}^{\dagger}_w\otimes\overline{\mathbf{C}}^{\dagger}_w\right]' vec\left(NT^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}_w\right) = O_{p^*}(T^{-1})\end{aligned} \tag{3.54}$$

because the independence of s_i from the other variables and $\mathbb{E}^*(s_i) = O(1)$ implies $\left\|\frac{1}{N}\sum_{i=1}^{N} s_i(\mathbf{q}_y'\widetilde{\boldsymbol{\eta}}_i'\otimes\mathbf{\Gamma}_i')\right\| = O_{p^*}(1)$ from Ass.3 and $\left\|T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}_w\right\| = O_{p^*}((NT)^{-1/2})$ from Lemma C-2 yields

$$\left\|NT^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}_w\right\| \le N\left\|T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}_w\right\|^2\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}\right\| = O_{p^*}(T^{-1})$$

For the next two terms, using also $\left\|T^{-1}\mathbf{F}^{0\prime}\mathbf{U}_i\right\| = O_{p^*}(T^{-1/2})$ from Lemma B-1 and $\mathbf{V}_i = \mathbf{U}_i\mathbf{q}_x$, $\boldsymbol{\varepsilon}_i = \mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$ follows

$$\begin{aligned}\left\|\sqrt{N}T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\boldsymbol{\varepsilon}_i\right\| &\le \sqrt{N}\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0\right\|\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}\right\|\left\|T^{-1}\mathbf{F}^{0\prime}\mathbf{U}_i\right\|\left\|\mathbf{B}^{-1}\mathbf{q}_y\right\| = O_{p^*}(T^{-1}) \\ \left\|\sqrt{N}T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\mathbf{V}_i\right\| &\le \sqrt{N}\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0\right\|\left\|\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}\right\|\left\|T^{-1}\mathbf{F}^{0\prime}\mathbf{U}_i\right\|\left\|\mathbf{q}_x\right\| = O_{p^*}(T^{-1})\end{aligned}$$

such that

$$\left\|\widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},\Gamma\varepsilon}\right\| \le \sqrt{T}\frac{1}{N}\sum_{i=1}^{N}\|s_i\|\,\|\mathbf{\Gamma}_i\|\left\|\overline{\mathbf{C}}^{\dagger}_w\right\|\left\|\sqrt{N}T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\boldsymbol{\varepsilon}_i\right\| = O_{p^*}(T^{-1/2}) \tag{3.55}$$

$$\left\|\widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},V\eta}\right\| = \sqrt{T}\frac{1}{N}\sum_{i=1}^{N}\|s_i\|\left\|T^{-1}\mathbf{V}_i'\mathbf{F}^0\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{F}^0}\sqrt{N}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}_w\right\|\left\|\overline{\mathbf{C}}^{\dagger}_w\right\|\|\widetilde{\boldsymbol{\eta}}_i\|\left\|\mathbf{q}_y\right\| = O_{p^*}(T^{-1/2}) \tag{3.56}$$

Next, by the independence of s_i and mutual independence in Ass.5 with expectation zero error terms in Ass.1

$$\mathbb{E}^*\left(\widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon}\right) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{E}^*(s_i)\mathbb{E}^*(\mathbf{V}_i)'\mathbb{E}^*(\mathbf{P}_{\mathbf{F}^0})\mathbb{E}^*(\boldsymbol{\varepsilon}_i) = \mathbf{0}_{k\times 1}$$

and also, since $\mathbf{V}_i$ and $\boldsymbol{\varepsilon}_i$ are independent over i with $\mathbb{E}^*(s_i^2) < \infty$

$$\begin{aligned}
\mathbb{E}^*(\widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon})(\widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon})' &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}^*(s_i s_j)\mathbb{E}^*\left(\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\mathbb{E}^*(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j')\mathbf{P}_{\mathbf{F}^0}\mathbf{V}_j\right) \\
&= \frac{1}{NT}\sum_{i=1}^{N}\mathbb{E}^*(s_i^2)\mathbb{E}^*\left(\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\mathbb{E}^*(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')\mathbf{P}_{\mathbf{F}^0}\mathbf{V}_i\right) \\
&= \frac{1}{NT}\sum_{i=1}^{N}\mathbb{E}^*(s_i^2)\mathbb{E}^*\left(\mathbf{V}_i'\mathbf{F}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\mathbb{E}^*(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')\mathbf{F}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\mathbf{V}_i\right) \\
&= \frac{1}{NT^3}\sum_{i=1}^{N}\mathbb{E}^*(s_i^2)\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\mathbb{E}^*\left(\mathbf{v}_{i,t}\mathbf{f}_t^{0\prime}\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}^0}\mathbf{f}_s^0\mathbb{E}^*(\varepsilon_{i,s}\varepsilon_{i,r})\mathbf{f}_r^{0\prime}\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}^0}\mathbf{f}_q^0\mathbf{v}_{i,q}'\right) \\
&= \frac{1}{NT^3}\sum_{i=1}^{N}\mathbb{E}^*(s_i^2)\sum_{t=1}^{T}\sum_{q=1}^{T}\mathbb{E}^*\left(\mathbf{v}_{i,t}\mathbf{f}_t^{0\prime}\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}^0}\left[\sum_{s=1}^{T}\sum_{r=1}^{T}\sigma_{i,s,r}\mathbf{f}_s^0\mathbf{f}_r^{0\prime}\right]\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}^0}\mathbf{f}_q^0\mathbf{v}_{i,q}'\right) \\
&= \frac{1}{NT^3}O(NT^2) = O\left(\frac{1}{T}\right)
\end{aligned}$$

by the stationarity of $\mathbf{f}_t, \varepsilon_{i,t}, \mathbf{v}_{i,t}$ and their mutual independence. This implies that

$$\left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon}\right\| = O_{p^*}(T^{-1/2}) \tag{3.57}$$

Combining (3.54), (3.55), (3.56) and (3.57) then leads to the conclusion

$$\left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}}\right\| \leq \left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon}\right\| + \left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\eta}\right\| + \left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},\Gamma\varepsilon}\right\| + \left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},\Gamma\eta}\right\| = O_{p^*}(T^{-1/2})$$

which is what needed to be shown.

Lemma C-6 *Under Ass.1-5 as* $(N,T)\to\infty$ *such that* $\tau_{N,T}\to\tau<\infty$,

$$\widehat{\mathbf{q}}_{\mathbf{I}}^{*}=\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}^{*}+2\sqrt{\tau}(\mathbf{b}_1-\mathbf{b}_2)+o_p(1) \tag{3.58}$$

$$\widehat{\mathbf{q}}_{\mathbf{I}}^{*}\xrightarrow{d^*}\mathcal{N}\left(\mathbf{0}_{k\times 1},2\mathbf{\Psi}\right)+2\sqrt{\tau}(\mathbf{b}_1-\mathbf{b}_2) \tag{3.59}$$

with $\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}^{*}=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}s_i\mathbf{V}_i'\boldsymbol{\varepsilon}_i$ *and* $\mathbf{b}_1$, $\mathbf{b}_2$ *and* $\mathbf{\Psi}$ *defined in lemma B-5.*

Proof of Lemma C-6

Recall that

$$\widehat{\mathbf{q}}_{\mathbf{I}}^{*}=\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}^{*}-\widehat{\mathbf{q}}_{\mathbf{I},V\eta}^{*}-\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\epsilon}^{*}+\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\eta}^{*}$$

For the last term in this decomposition we find from

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}s_i(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\otimes\mathbf{\Gamma}_i)&=\frac{1}{N}\sum_{i=1}^{N}s_i(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\otimes(\mathbf{C}+\widetilde{\boldsymbol{\eta}}_i)\mathbf{q}_x)=\mathbb{E}^*(s_i)\mathbb{E}^*(\widetilde{\boldsymbol{\eta}}_i\otimes\widetilde{\boldsymbol{\eta}}_i)(\mathbf{q}_y\otimes\mathbf{q}_x)+O_{p^*}(N^{-1/2})\\&=\mathbf{\Sigma}_{\boldsymbol{\eta}}\mathbf{q}_{xy}+O_{p^*}(N^{-1/2})\end{aligned}$$

by the independence of s_i from the other variables, $\mathbb{E}^*(s_i)=1$ and Ass.3, together with substituting in (3.22) from Lemma C-2 and $\overline{\mathbf{C}}_w^{\dagger}=\mathbf{C}^{\dagger}+O_{p^*}(N^{-1/2})$ that

$$\begin{aligned}\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\eta}^{*}&=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}s_i\mathbf{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y=\sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}s_i(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\otimes\mathbf{\Gamma}_i)\right]'vec\left((\overline{\mathbf{C}}_w^{\dagger})'NT^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\right)\\&=2\sqrt{\tau}\mathbf{q}_{xy}'\mathbf{\Sigma}_{\boldsymbol{\eta}}'vec\left((\mathbf{C}^{\dagger})'\mathbf{\Sigma}_{\mathbf{u}}\mathbf{C}^{\dagger}\right)+O_{p^*}(N^{-1/2})+O_{p^*}(T^{-1/2})\end{aligned} \tag{3.60}$$

Next, making use of $\boldsymbol{\varepsilon}_i=\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$ and $\overline{\mathbf{U}}_w=N^{-1}(s_i\mathbf{U}_i+\sum_{j\neq i}^{N}s_j\mathbf{U}_j)$,

$$\begin{aligned}\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\varepsilon}^{*}&=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}s_i\mathbf{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'\boldsymbol{\varepsilon}_i=\sqrt{\tau_{N,T}}\sum_{i=1}^{N}s_i\mathbf{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'T^{-1}\overline{\mathbf{U}}_w'\boldsymbol{\varepsilon}_i\\&=\sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}s_i^2\mathbf{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'(T^{-1}\mathbf{U}_i'\mathbf{U}_i)\mathbf{B}^{-1}\mathbf{q}_y\right]+\sqrt{\tau_{N,T}}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N}s_is_j\mathbf{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\mathbf{U}_j'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y\right]\\&=\sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N}s_i^2\mathbf{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'[\sigma_i^2,\mathbf{0}_{1\times k}]'\right]+O_{p^*}(T^{-1/2})\end{aligned}$$

where the order of the rightmost term on the second line equals that of (2.34) due to the independence of s_i,s_j from the other variables, and we used $T^{-1}\mathbf{U}_i'\mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y=[\sigma_i^2,\mathbf{0}_{1\times k}]'+O_{p^*}(T^{-1/2})$. It thus follows under Ass.1,3 and given that $\mathbb{E}^*(s_i^2)=2+O(N^{-1})$ that

$$\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\varepsilon}^{*}\longrightarrow^{p^*}2\sqrt{\tau}\mathbf{\Gamma}'(\mathbf{C}^{\dagger})'[\sigma^2,\mathbf{0}_{1\times k}]' \tag{3.61}$$

For $\widehat{\mathbf{q}}^*_{\mathbf{I},V\eta}$, making use of $\mathbf{V}_i = \mathbf{U}_i\mathbf{q}_x$ and substituting in the same results as above leads to

$$
\begin{aligned}
\widehat{\mathbf{q}}^*_{\mathbf{I},V\eta} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} s_i \mathbf{V}_i'\overline{\mathbf{U}}_w\overline{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y \\
&= \sqrt{\tau_{N,T}}\mathbf{q}_x'\left[\frac{1}{N}\sum_{i=1}^{N} s_i(T^{-1}\mathbf{U}_i'\mathbf{U}_i)\overline{\mathbf{C}}_w^{\dagger}\widetilde{\boldsymbol{\eta}}_i\right]\mathbf{q}_y + \sqrt{\tau_{N,T}}\mathbf{q}_x'\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}^{N} s_i s_j(T^{-1}\mathbf{U}_i'\mathbf{U}_j)\overline{\mathbf{C}}_w^{\dagger}\widetilde{\boldsymbol{\eta}}_i\right]\mathbf{q}_y \\
&= \sqrt{\tau_{N,T}}\mathbf{q}_x'\left[\frac{1}{N}\sum_{i=1}^{N} s_i\boldsymbol{\Sigma}_{\mathbf{u},i}\overline{\mathbf{C}}_w^{\dagger}\widetilde{\boldsymbol{\eta}}_i\right]\mathbf{q}_y + O_{p^*}(T^{-1/2}) \\
&= \sqrt{\tau_{N,T}}\mathbf{q}_{xy}'\left[\frac{1}{N}\sum_{i=1}^{N} s_i(\widetilde{\boldsymbol{\eta}}_i\otimes\boldsymbol{\Sigma}_{\mathbf{u},i})\right]' vec(\overline{\mathbf{C}}_w^{\dagger}) + O_{p^*}(T^{-1/2}) \\
&= O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})
\end{aligned}
\tag{3.62}
$$

because $\frac{1}{N}\sum_{i=1}^{N}\widetilde{\boldsymbol{\eta}}_i = O_{p^*}(N^{-1/2})$ by Ass.3 and the independence from the s_i (with the latter having finite variance) implies that their presence will not change this order.

Finally, recalling that

$$
\widehat{\mathbf{q}}^*_{\mathbf{I},V\varepsilon} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} s_i\mathbf{V}_i'\boldsymbol{\varepsilon}_i
$$

we have $\left\|\widehat{\mathbf{q}}^*_{\mathbf{I},V\varepsilon}\right\| = O_{p^*}(1)$, and it follows under Ass.1 and the independence of the s_i with respect to $\mathbf{V}_i$ and $\boldsymbol{\varepsilon}_i$ that $\mathbb{E}^*(\widehat{\mathbf{q}}^*_{\mathbf{I},V\varepsilon}) = \mathbf{0}_{k\times 1}$. The cross-section independence of $\mathbf{V}_i$ and $\boldsymbol{\varepsilon}_i$ with $\mathbb{E}^*(s_i^2) = 2 + O(N^{-1})$ in lemma C-1 leads to

$$
\begin{aligned}
\mathbb{V}ar^*\left(\widehat{\mathbf{q}}^*_{\mathbf{I},V\varepsilon}\right) &= \mathbb{E}^*\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N} s_i s_j\mathbf{V}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j'\mathbf{V}_j\right] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}^*(s_i s_j)\mathbb{E}^*(\mathbf{V}_i'\mathbb{E}^*(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j')\mathbf{V}_j) \\
&= \frac{1}{NT}\sum_{i=1}^{N}\mathbb{E}^*(s_i^2)\mathbb{E}^*(\mathbf{V}_i'\mathbb{E}^*(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')\mathbf{V}_i) \\
&= \frac{1}{NT}\sum_{i=1}^{N} 2\mathbb{E}^*(\mathbf{V}_i'\boldsymbol{\Omega}_i\mathbf{V}_i) + O(N^{-1})
\end{aligned}
$$

with $\boldsymbol{\Omega}_i = \mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')$. Hence, given that all 4th order moments are finite and $\{s_i\mathbf{v}_{i,t}\varepsilon_{i,t}\}$ are stationary and cross-section independent we have by a CLT for independent heterogeneous variables as $(N,T)\to\infty$

$$
\widehat{\mathbf{q}}^*_{\mathbf{I},V\varepsilon} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} s_i\mathbf{V}_i'\boldsymbol{\varepsilon}_i \xrightarrow{d^*} \mathcal{N}\left(\mathbf{0}_{k\times 1}, 2\boldsymbol{\Psi}\right)
\tag{3.63}
$$

with $\boldsymbol{\Psi}$ as defined in Lemma B-5. Combining (3.60)-(3.63) in the decomposition of $\widehat{\mathbf{q}}^*_{\mathbf{I}}$ then gives

$$
\widehat{\mathbf{q}}^*_{\mathbf{I}} \xrightarrow{d^*} \mathcal{N}\left(\mathbf{0}_{k\times 1}, 2\boldsymbol{\Psi}\right) + 2\sqrt{\tau}(\mathbf{b}_1 - \mathbf{b}_2)
$$

with $\mathbf{b}_1$ and $\mathbf{b}_2$ as defined in Lemma B-5. This is what needed to be shown.

Lemma C-7 *Under Ass.1-5 we have that*

$$\widehat{\mathbf{Q}}^* \longrightarrow^{p^*} \boldsymbol{\Sigma} \tag{3.64}$$

$$(\widehat{\mathbf{Q}}^*)^{-1} \longrightarrow^{p^*} \boldsymbol{\Sigma}^{-1} \tag{3.65}$$

as $(N, T) \to \infty$.

Proof of Lemma C-7

Recall that

$$\widehat{\mathbf{Q}}^* = \frac{1}{NT}\sum_{i=1}^{N} s_i \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^*} \mathbf{X}_i = \widehat{\mathbf{Q}}^*_{\mathbf{I}} - \widehat{\mathbf{Q}}^*_{\mathbf{P}_{\mathbf{F}^0}} - \widehat{\mathbf{Q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]}$$

which for a given matrix $\mathbf{A}$ was decomposed as $\widehat{\mathbf{Q}}^*_{\mathbf{A}} = \widehat{\mathbf{Q}}^*_{\mathbf{A},VV} - \widehat{\mathbf{Q}}^*_{\mathbf{A},V\Gamma} - (\widehat{\mathbf{Q}}^*_{\mathbf{A},V\Gamma})' + \widehat{\mathbf{Q}}^*_{\mathbf{A},\Gamma\Gamma}$ with

$$\widehat{\mathbf{Q}}^*_{\mathbf{A},VV} = \frac{1}{NT}\sum_{i=1}^{N} s_i \mathbf{V}_i' \mathbf{A} \mathbf{V}_i$$

$$\widehat{\mathbf{Q}}^*_{\mathbf{A},V\Gamma} = \frac{1}{NT}\sum_{i=1}^{N} s_i \mathbf{V}_i' \mathbf{A} \overline{\mathbf{U}}_w \overline{\mathbf{C}}_w^{\dagger} \boldsymbol{\Gamma}_i$$

$$\widehat{\mathbf{Q}}^*_{\mathbf{A},\Gamma\Gamma} = \frac{1}{NT}\sum_{i=1}^{N} s_i \boldsymbol{\Gamma}_i' (\overline{\mathbf{C}}_w^{\dagger})' \overline{\mathbf{U}}_w' \mathbf{A} \overline{\mathbf{U}}_w \overline{\mathbf{C}}_w^{\dagger} \boldsymbol{\Gamma}_i$$

For the analysis, recall from Lemma C-2 that $\left\| T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w \right\| = O_{p^*}(N^{-1})$ and $\left\| T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}_w \right\| = O_{p^*}((NT)^{-1/2})$ and note that (3.36) implies $\left\| T^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w \right\| = O_{p^*}(N^{-1})$. Given also that $\left\| \frac{1}{N}\sum_{i=1}^{N} s_i(\boldsymbol{\Gamma}_i' \otimes \boldsymbol{\Gamma}_i') \right\| = O_{p^*}(1)$ by Ass.3 and the independence of s_i and $\boldsymbol{\Gamma}_i$, we have

$$\left\| \widehat{\mathbf{Q}}^*_{\mathbf{I},\Gamma\Gamma} \right\| \leq \left\| \frac{1}{N}\sum_{i=1}^{N} s_i(\boldsymbol{\Gamma}_i' \otimes \boldsymbol{\Gamma}_i') \right\| \left\| \overline{\mathbf{C}}_w^{\dagger} \right\|^2 \left\| T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w \right\| = O_{p^*}(N^{-1})$$

$$\left\| \widehat{\mathbf{Q}}^*_{\mathbf{P}_{\mathbf{F}^0},\Gamma\Gamma} \right\| \leq \left\| \frac{1}{N}\sum_{i=1}^{N} s_i(\boldsymbol{\Gamma}_i' \otimes \boldsymbol{\Gamma}_i') \right\| \left\| \overline{\mathbf{C}}_w^{\dagger} \right\|^2 \left\| T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0 \right\|^2 \left\| \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0} \right\| = O_{p^*}((NT)^{-1})$$

$$\left\| \widehat{\mathbf{Q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\Gamma} \right\| \leq \left\| \frac{1}{N}\sum_{i=1}^{N} s_i(\boldsymbol{\Gamma}_i' \otimes \boldsymbol{\Gamma}_i') \right\| \left\| \overline{\mathbf{C}}_w^{\dagger} \right\|^2 \left\| T^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w \right\| = O_{p^*}(N^{-1})$$

Next, the fact that $\mathbf{V}_i = \mathbf{U}_i\mathbf{q}_x$ and using also $\left\| T^{-1}\overline{\mathbf{U}}_w'\mathbf{U}_i \right\| = O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2})$ from Lemma C-2 and $\left\| T^{-1}\mathbf{F}^{0\prime}\mathbf{U}_i \right\| = O_{p^*}(T^{-1/2})$ of Lemma B-1 reveal that

$$\left\| \widehat{\mathbf{Q}}^*_{\mathbf{I},V\Gamma} \right\| \leq \frac{1}{N}\sum_{i=1}^{N} s_i \left\| \mathbf{q}_x \right\| \left\| T^{-1}\overline{\mathbf{U}}_w'\mathbf{U}_i \right\| \left\| \overline{\mathbf{C}}_w^{\dagger} \right\| \left\| \boldsymbol{\Gamma}_i \right\| = O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2})$$

$$\left\| \widehat{\mathbf{Q}}^*_{\mathbf{P}_{\mathbf{F}^0},V\Gamma} \right\| \leq \frac{1}{N}\sum_{i=1}^{N} s_i \left\| \mathbf{q}_x \right\| \left\| T^{-1}\mathbf{F}^{0\prime}\mathbf{U}_i \right\| \left\| \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0} \right\| \left\| T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0 \right\| \left\| \overline{\mathbf{C}}_w^{\dagger} \right\| \left\| \boldsymbol{\Gamma}_i \right\| = O_{p^*}(N^{-1/2}T^{-1})$$

$$\left\| \widehat{\mathbf{Q}}^*_{\mathbf{P}_{\mathbf{F}^0},VV} \right\| \leq \frac{1}{N}\sum_{i=1}^{N} s_i \left\| \mathbf{q}_x \right\|^2 \left\| T^{-1}\mathbf{F}^{0\prime}\mathbf{U}_i \right\|^2 \left\| \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0} \right\| = O_{p^*}(T^{-1})$$

For the next result, equations (3.39) and (3.40) imply

$$\left\| T^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i \right\| = O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2})$$

so that in turn

$$\left\| \widehat{\mathbf{Q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],VT} \right\| \leq \frac{1}{N}\sum_{i=1}^{N} s_i \left\| \mathbf{q}_x \right\| \left\| T^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i \right\| \left\| \overline{\mathbf{C}}_w^{\dagger} \right\| \left\| \mathbf{\Gamma}_i \right\| = O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2})$$

Next, from two lines above eq.(3.44)

$$T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i = T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}^0_{w,-m}\widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}} T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\mathbf{U}_i + O_{p^*}(N^{-3/2}) + O_{p^*}(N^{-1/2}T^{-1}) + O_{p^*}(N^{-1}T^{-1/2}) + O_{p^*}(T^{-3/2})$$

which using $\left\| T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\mathbf{U}_i \right\| = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$ from Lemma C-3 leads to

$$\begin{aligned} \left\| T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i \right\| &\leq \left\| T^{-1}(\overline{\mathbf{U}}^0_{w,-m})'\mathbf{U}_i \right\|^2 \left\| \widehat{\mathbf{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}} \right\| \\ &\quad + O_{p^*}(N^{-3/2}) + O_{p^*}(N^{-1/2}T^{-1}) + O_{p^*}(N^{-1}T^{-1/2}) + O_{p^*}(T^{-3/2}) \\ &= O_{p^*}(N^{-1}) + O_{p^*}(T^{-1}) + O_{p^*}((NT)^{-1/2}) \end{aligned} \tag{3.66}$$

Substituting this result into $\widehat{\mathbf{Q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],VV}$ gives

$$\left\| \widehat{\mathbf{Q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],VV} \right\| \leq \frac{1}{N}\sum_{i=1}^{N} s_i \left\| \mathbf{q}_x \right\|^2 \left\| T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i \right\| = O_{p^*}(N^{-1}) + O_{p^*}(T^{-1}) + O_{p^*}((NT)^{-1/2})$$

For the last remaining term it follows from Ass.1 and the independence between s_i and $\mathbf{V}_i$ that

$$\widehat{\mathbf{Q}}^*_{\mathbf{I},VV} = \frac{1}{N}\sum_{i=1}^{N} s_i \left(\frac{\mathbf{V}_i'\mathbf{V}_i}{T} \right) = \frac{1}{N}\sum_{i=1}^{N} s_i \mathbf{\Sigma}_i + O_{p^*}(T^{-1/2})$$

Combining then all the previous results into $\widehat{\mathbf{Q}}^* = \widehat{\mathbf{Q}}^*_{\mathbf{I}} - \widehat{\mathbf{Q}}^*_{\mathbf{M}_{\mathbf{F}^0}} - \widehat{\mathbf{Q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]}$ makes

$$\widehat{\mathbf{Q}}^* = \frac{1}{N}\sum_{i=1}^{N} s_i \mathbf{\Sigma}_i + O_{p^*}(T^{-1/2}) + O_{p^*}(N^{-1})$$

Then, given that from $\mathbb{E}^*(s_i) = 1$ and Ass.1

$$\frac{1}{N}\sum_{i=1}^{N} s_i \mathbf{\Sigma}_i \longrightarrow^{p^*} \mathbb{E}^*(s_i)\mathbb{E}^*(\mathbf{\Sigma}_i) = \mathbf{\Sigma}$$

it follows that

$$\widehat{\mathbf{Q}}^* \longrightarrow^{p^*} \mathbf{\Sigma}$$

as in eq.(3.64) of the lemma. As $\mathbf{\Sigma}$ is positive definite by Ass.1, the previous result implies that $rk(\widehat{\mathbf{Q}}^*) - rk(\mathbf{\Sigma}) \overset{a.s.}{\to} 0$, which by application of Theorem 1 in Karabiyik et al. (2017) then leads to eq.(3.65) of the lemma. This finishes the proof.

3.3.2 Theorems and Corollaries

Theorem 2 *Under Ass.1-5 we have as* $(N,T)\to\infty$ *such that* $T/N=\tau_{N,T}\to\tau<\infty$ *that*

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}^*-\widehat{\boldsymbol{\beta}})\xrightarrow{d^*}\mathcal{N}(\mathbf{0}_{k\times1},\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1})+\sqrt{\tau}\boldsymbol{\Sigma}^{-1}(\mathbf{b}-\mathbf{d}-\mathbf{d}^+)$$

where $\mathbf{b}=\mathbf{b}_1-\mathbf{b}_2$ *and* $\mathbf{d}=\mathbf{d}_1+\mathbf{d}_2$ *are given in Lemmas C-4 and C-6 and* $\mathbf{d}^+$ *is defined in Lemma C-4. If in addition either* $\tau=0$ *or* $m=1+k$ *then*

$$\sup_{x\in\mathbb{R}^{k\times1}}\left|\mathbb{P}^*[\sqrt{NT}(\widehat{\boldsymbol{\beta}}^*-\widehat{\boldsymbol{\beta}})\le x]-\mathbb{P}[\sqrt{NT}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\le x]\right|\longrightarrow^p 0,$$

where inequalities are to be interpreted coordinate wise.

Proof of Theorem 2 Consider that from (2.22) and (3.29) we have

$$\begin{aligned}\sqrt{NT}(\widehat{\boldsymbol{\beta}}^*-\boldsymbol{\beta})&=(\widehat{\mathbf{Q}}^*)^{-1}\widehat{\mathbf{q}}^*\\ \sqrt{NT}(\widehat{\boldsymbol{\beta}}^*-\boldsymbol{\beta})-\sqrt{NT}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})&=(\widehat{\mathbf{Q}}^*)^{-1}\widehat{\mathbf{q}}^*-\widehat{\mathbf{Q}}^{-1}\widehat{\mathbf{q}}\\ \sqrt{NT}(\widehat{\boldsymbol{\beta}}^*-\widehat{\boldsymbol{\beta}})&=(\widehat{\mathbf{Q}}^*)^{-1}[\widehat{\mathbf{q}}^*-\widehat{\mathbf{q}}]+[(\widehat{\mathbf{Q}}^*)^{-1}-(\widehat{\mathbf{Q}})^{-1}]\widehat{\mathbf{q}}\end{aligned}\tag{3.67}$$

and using further the definitions in (2.24) and (3.33) we can write, specifically with $\widehat{\mathbf{q}}_{\mathbf{I}}=\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}-\widehat{\mathbf{q}}_{\mathbf{I},V\eta}-\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\varepsilon}+\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\eta}$ and $\widehat{\mathbf{q}}_{\mathbf{I}}^*=\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}^*-\widehat{\mathbf{q}}_{\mathbf{I},V\eta}^*-\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\varepsilon}^*+\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\eta}^*$ that

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}^*-\widehat{\boldsymbol{\beta}})=(\widehat{\mathbf{Q}}^*)^{-1}\widetilde{\mathbf{q}}_{\mathbf{I},V\varepsilon}+[\widehat{\mathbf{k}}^*-\widehat{\mathbf{k}}]+[(\widehat{\mathbf{Q}}^*)^{-1}-(\widehat{\mathbf{Q}})^{-1}]\widehat{\mathbf{q}}\tag{3.68}$$

with $\widetilde{\mathbf{q}}_{\mathbf{I},V\varepsilon}=\mathbf{q}_{\mathbf{I},V\varepsilon}^*-\mathbf{q}_{\mathbf{I},V\varepsilon}$ and we have defined $\widehat{\mathbf{k}}=(\widehat{\mathbf{Q}}^*)^{-1}[\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\eta}-\widehat{\mathbf{q}}_{\mathbf{I},V\eta}-\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\varepsilon}-\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}0}}-\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}0}-\mathbf{M}_{\widehat{\mathbf{F}}0}]}]$, and similarly for the starred bootstrap world equivalent $\widehat{\mathbf{k}}^*=(\widehat{\mathbf{Q}}^*)^{-1}[\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\eta}^*-\widehat{\mathbf{q}}_{\mathbf{I},V\eta}^*-\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\varepsilon}^*-\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}0}}^*-\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}0}-\mathbf{M}_{\widehat{\mathbf{F}}0*}]}^*]$.

Recall then that by definition $\mathbf{q}_{\mathbf{I},V\varepsilon}=\frac{1}{\sqrt{NT}}\sum_{i=1}^N\mathbf{V}_i'\boldsymbol{\varepsilon}_i$ and $\mathbf{q}_{\mathbf{I},V\varepsilon}^*=\frac{1}{\sqrt{NT}}\sum_{i=1}^N s_i\mathbf{V}_i'\boldsymbol{\varepsilon}_i$ and therefore we can write

$$\widetilde{\mathbf{q}}_{\mathbf{I},V\varepsilon}=\mathbf{q}_{\mathbf{I},V\varepsilon}^*-\mathbf{q}_{\mathbf{I},V\varepsilon}=\frac{1}{\sqrt{NT}}\sum_{i=1}^N(s_i-1)\mathbf{V}_i'\boldsymbol{\varepsilon}_i$$

Here, Ass.1 and the mutual independence of s_i, $\mathbf{V}_i$ and $\boldsymbol{\varepsilon}_i$ implies $\mathbb{E}[\widetilde{\mathbf{q}}_{\mathbf{I},V\varepsilon}]=\mathbf{0}_{k\times1}$. We have also by the cross-section independence of $\mathbf{V}_i$ and $\boldsymbol{\varepsilon}_i$ under Ass.1 that

$$\begin{aligned}\mathbb{V}ar^*(\widetilde{\mathbf{q}}_{\mathbf{I},V\varepsilon})=\mathbb{E}^*\left[\widetilde{\mathbf{q}}_{\mathbf{I},V\varepsilon}\widetilde{\mathbf{q}}_{\mathbf{I},V\varepsilon}'\right]&=\frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\mathbb{E}^*[(s_i-1)(s_j-1)]\mathbb{E}^*\left[\mathbf{V}_i'\mathbb{E}^*(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j')\mathbf{V}_j\right]\\&=\frac{1}{NT}\sum_{i=1}^N\mathbb{E}^*[(s_i-1)^2]\mathbb{E}^*\left[\mathbf{V}_i'\mathbb{E}^*(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')\mathbf{V}_i\right]\\&=\frac{1}{NT}\sum_{i=1}^N\mathbb{E}^*\left[\mathbf{V}_i'\boldsymbol{\Omega}_i\mathbf{V}_i\right]+O(N^{-1})\end{aligned}$$

where the final line uses $\mathbb{E}^*(s_i) = 1$ and $\mathbb{E}^*(s_i^2) = 2 - N^{-1}$ from Lemma C-1 such that $\mathbb{E}^*[(s_i - 1)^2] = \mathbb{E}^*(s_i^2) - 2\mathbb{E}^*(s_i) + 1 = 1 - N^{-1}$. Hence, by a CLT for independent but heterogeneous stationary variables as $(N, T) \to \infty$

$$\widetilde{\mathbf{q}}_{\mathbf{I},V\varepsilon} \xrightarrow{d^*} \mathcal{N}(\mathbf{0}_{k\times 1}, \mathbf{\Psi}) \tag{3.69}$$

with $\mathbf{\Psi}$ defined in Theorem 1. Therefore, since by Lemma C-7 we have $(\widehat{\mathbf{Q}}^*)^{-1} \longrightarrow^{p^*} \mathbf{\Sigma}^{-1}$, the leading term in (3.68) is the one that drives the asymptotic distribution

$$(\widehat{\mathbf{Q}}^*)^{-1}\widetilde{\mathbf{q}}_{\mathbf{I},V\varepsilon} \xrightarrow{d^*} \mathcal{N}(\mathbf{0}_{k\times 1}, \mathbf{\Sigma}^{-1}\mathbf{\Psi}\mathbf{\Sigma}^{-1}) \tag{3.70}$$

Next up, since Lemma B-6 shows that also $\widehat{\mathbf{Q}}^{-1} \longrightarrow^{p} \mathbf{\Sigma}^{-1}$ we have $[(\widehat{\mathbf{Q}}^*)^{-1} - \widehat{\mathbf{Q}}^{-1}] \longrightarrow^{p^*} \mathbf{0}_{k\times k}$. As Lemmas B-3, B-4 and B-5 imply $\|\widehat{\mathbf{q}}\| = O_p(1)$, it follows for the last term in (3.68) as $(N, T) \to \infty$ that

$$[(\widehat{\mathbf{Q}}^*)^{-1} - \widehat{\mathbf{Q}}^{-1}]\widehat{\mathbf{q}} \longrightarrow^{p^*} \mathbf{0}_{k\times 1} \tag{3.71}$$

Consider next $[\widehat{\mathbf{k}}^* - \widehat{\mathbf{k}}]$. Lemmas B-3, B-4, B-6 and results (2.53), (2.54), (2.56) in the proof of Lemma B-5 give

$$\widehat{\mathbf{k}} \longrightarrow^{p} \sqrt{\tau}\mathbf{\Sigma}^{-1}(\mathbf{b} - \mathbf{d})$$

whereas from lemmas C-4, C-5, C-7 and (3.60), (3.61), (3.62) in the proof of Lemma B-5 we get

$$\widehat{\mathbf{k}}^* \longrightarrow^{p^*} 2\sqrt{\tau}\mathbf{\Sigma}^{-1}(\mathbf{b} - \mathbf{d}) - \sqrt{\tau}\mathbf{\Sigma}^{-1}\mathbf{d}^+$$

Hence, it follows that

$$[\widehat{\mathbf{k}}^* - \widehat{\mathbf{k}}] \longrightarrow^{p^*} \sqrt{\tau}\mathbf{\Sigma}^{-1}(\mathbf{b} - \mathbf{d} - \mathbf{d}^+) \tag{3.72}$$

such that combining (3.70), (3.71) and (3.72) into (3.68) returns

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \xrightarrow{d^*} \mathcal{N}(\mathbf{0}_{k\times 1}, \mathbf{\Sigma}^{-1}\mathbf{\Psi}\mathbf{\Sigma}^{-1}) + \sqrt{\tau}\mathbf{\Sigma}^{-1}(\mathbf{b} - \mathbf{d} - \mathbf{d}^+) \tag{3.73}$$

which is the reported distribution in the theorem.

For the final statement of the theorem we make use of the fact that provided $m = 1 + k$ we have by lemma B-3 that $\mathbf{d} = \mathbf{d}_1 + \mathbf{d}_2 = \mathbf{0}_{k\times 1}$ in the original sample, and similarly $\mathbf{d} = \mathbf{0}_{k\times 1}, \mathbf{d}^+ = \mathbf{0}_{k\times 1}$ in the bootstrap world by lemma C-4. Therefore, it follows under the additional condition that $m = 1 + k$ from the same arguments as above

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \xrightarrow{d^*} \mathcal{N}(\mathbf{0}_{k\times 1}, \mathbf{\Sigma}^{-1}\mathbf{\Psi}\mathbf{\Sigma}^{-1}) + \sqrt{\tau}\mathbf{\Sigma}^{-1}\mathbf{b} \tag{3.74}$$

From this result and Theorem 1 (when $m = 1 + k$) directly follows that if $m = 1 + k$

$$\sup_{x \in \mathbb{R}^{k \times 1}} \left| \mathbb{P}^*[\sqrt{NT}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \leq x] - \mathbb{P}[\sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq x] \right| \longrightarrow^p 0,$$

where the inequalities are to be interpreted coordinate wise. The statement holds similarly when $T/N \to \tau = 0$, without the requirement that $m = 1 + k$, since the distributions in (3.73) and Theorem 1 are then unbiased.

Corollary 2 *Let Ass. 1-3, 5 and 7 hold. Letting* $\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}}$ *be the CCEP estimator in the bootstrap world with* $\mathbf{P}_{\widehat{\mathbf{F}}^*} = \overline{\mathbf{X}}^*(\overline{\mathbf{X}}^{*\prime}\overline{\mathbf{X}}^*)\overline{\mathbf{X}}^{*\prime}$*, we have as* $(N,T) \to \infty$ *such that* $T/N \to \tau < \infty$ *that*

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}} - \widehat{\boldsymbol{\beta}}) \xrightarrow{d^*} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}) + \sqrt{\tau}\boldsymbol{\Sigma}^{-1}\mathbf{g}$$

with $\mathbf{g}$ *defined in Corollary 1. By consequence,*

$$\sup_{x\in\mathbb{R}^{k\times 1}} \left| \mathbb{P}^*[\sqrt{NT}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}} - \widehat{\boldsymbol{\beta}}_{\mathbf{x}}) \leq x] - \mathbb{P}[\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \boldsymbol{\beta}) \leq x] \right| \longrightarrow^p 0,$$

where inequalities are to be interpreted coordinate wise.

Proof of Corollary 2 As in the proof of Corollary 1, we can study the CCEP estimator with $\overline{\mathbf{y}}^*$ excluded for the estimation of the factors by replacing in all expressions $\overline{\mathbf{Z}}^*, \overline{\mathbf{U}}_w, \overline{\mathbf{C}}_w$ with respectively $\overline{\mathbf{X}}^*, \overline{\mathbf{V}}_w$ and $\overline{\boldsymbol{\Gamma}}_w$. Hence, $\mathbf{F} = (\overline{\mathbf{X}}^* - \overline{\mathbf{V}}_w)\overline{\boldsymbol{\Gamma}}_w^{\dagger}$ and $\mathbf{P}_{\widehat{\mathbf{F}}^*} = \overline{\mathbf{X}}^*(\overline{\mathbf{X}}^{*\prime}\overline{\mathbf{X}}^*)^{+}\overline{\mathbf{X}}^{*\prime}$. The associated rotation matrix is then $\mathbf{R}_{\mathbf{x},w} = \mathbf{T}_{\mathbf{x}}\overline{\mathbf{H}}_{\mathbf{x},w}\mathbf{D}_{N,\mathbf{x}}$ with $\mathbf{T}_{\mathbf{x}}$ and $\mathbf{D}_{N,\mathbf{x}}$ as defined in Corollary 1 with $\overline{\boldsymbol{\Gamma}}_w\mathbf{T}_{\mathbf{x}} = [\overline{\boldsymbol{\Gamma}}_{w,m}, \overline{\boldsymbol{\Gamma}}_{w,-m}]$ and $\overline{\mathbf{V}}_w\mathbf{T}_{\mathbf{x}} = [\overline{\mathbf{V}}_{w,m}, \overline{\mathbf{V}}_{w,-m}]$, and

$$\overline{\mathbf{H}}_{\mathbf{x},w} = [\overline{\mathbf{H}}_{\mathbf{x},w,m}, \overline{\mathbf{H}}_{\mathbf{x},w,-m}] = \begin{bmatrix} \overline{\boldsymbol{\Gamma}}_{w,m}^{-1} & -\overline{\boldsymbol{\Gamma}}_{w,m}^{-1}\overline{\boldsymbol{\Gamma}}_{w,-m} \\ \mathbf{0}_{(k-m)\times m} & \mathbf{I}_{k-m} \end{bmatrix} \tag{3.75}$$

with also $\left\|\overline{\mathbf{H}}_{\mathbf{x},w} - \mathbf{H}_{\mathbf{x}}\right\| = O_{p^*}(N^{-1/2})$ and $\mathbf{H}_{\mathbf{x}}$ was defined in Corollary 1.

Replacing then in addition also everywhere in the analysis of Theorem 2 $\mathbf{R}_w, \mathbf{T}, \overline{\mathbf{H}}_w, \mathbf{H}$ with $\mathbf{R}_{\mathbf{x},w}, \mathbf{T}_{\mathbf{x}}, \overline{\mathbf{H}}_{\mathbf{x},w}, \mathbf{H}_{\mathbf{x}}$ allows us to analyze the CCEP estimator $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ in the pairs bootstrap world, with for completeness now

$$\widehat{\mathbf{F}}^{0*} = [\mathbf{F}\overline{\boldsymbol{\Gamma}}_w + \overline{\mathbf{V}}_w]\mathbf{R}_{\mathbf{x},w} = \mathbf{F}^0 + \overline{\mathbf{V}}_w^0 \tag{3.76}$$

where $\mathbf{F}^0 = [\mathbf{F}, \mathbf{0}_{T\times(k-m)}]$ and $\overline{\mathbf{V}}_w^0 = [\overline{\mathbf{V}}_{w,m}^0, \overline{\mathbf{V}}_{w,-m}^0]$, with $\overline{\mathbf{V}}_{w,m}^0 = \overline{\mathbf{V}}_{w,m}\overline{\boldsymbol{\Gamma}}_{w,m}^{-1}$ and $\overline{\mathbf{V}}_{w,-m}^0 = \sqrt{N}\overline{\mathbf{V}}_w\mathbf{T}_{\mathbf{x}}\overline{\mathbf{H}}_{\mathbf{x},w,-m} = \sqrt{N}(\overline{\mathbf{V}}_{w,m} - \overline{\mathbf{V}}_{w,m}\overline{\boldsymbol{\Gamma}}_{w,m}^{-1}\overline{\boldsymbol{\Gamma}}_{w,-m})$.

Denote now the scaled deviation of the CCEP estimator in the pairs bootstrap

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}} - \boldsymbol{\beta}) = (\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1}\widehat{\mathbf{q}}^*_{\mathbf{x}} \tag{3.77}$$

where the breakdown of $\widehat{\mathbf{Q}}^*_{\mathbf{x}}$ and $\widehat{\mathbf{q}}^*_{\mathbf{x}}$ is identical to that outlined in (3.32) and (3.33), but we will use an additional $\mathbf{x}$ subscript to make explicit that in this breakdown $\overline{\mathbf{Z}}^*, \overline{\mathbf{U}}_w, \overline{\mathbf{C}}_w$ are replaced with $\overline{\mathbf{X}}^*, \overline{\mathbf{V}}_w, \overline{\boldsymbol{\Gamma}}_w$ and the rotated matrices have also been redefined as above. For the analysis, since $\overline{\mathbf{X}}^* \subset \overline{\mathbf{Z}}^*$, $\overline{\mathbf{V}}_w \subset \overline{\mathbf{U}}_w$ and $\overline{\boldsymbol{\Gamma}}_w \subset \overline{\mathbf{C}}_w$, the asymptotic orders derived in Lemmas C-2, C-3, C-4, C-5, C-6, C-7 for $\overline{\mathbf{Z}}^*, \overline{\mathbf{U}}_w, \overline{\mathbf{C}}_w$ are upper bounds for the analysis with $\overline{\mathbf{X}}^*, \overline{\mathbf{V}}_w, \overline{\boldsymbol{\Gamma}}_w$ here (i.e. the terms here converge at a rate at least as fast or faster). Hence, it follows directly from Lemmas C-5 and C-7

$$(\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1} \longrightarrow^{p^*} \boldsymbol{\Sigma}^{-1} \tag{3.78}$$

$$\left\|\widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{P}_{\mathbf{F}^0}}\right\| = O_{p^*}(T^{-1/2}) \tag{3.79}$$

For the analysis of $\widehat{\mathbf{q}}^*_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]} = \widehat{\mathbf{q}}^*_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon} - \widehat{\mathbf{q}}^*_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\eta} - \widehat{\mathbf{q}}^*_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\varepsilon} + \widehat{\mathbf{q}}^*_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\eta}$, we have directly from (3.43) in Lemma C-4

$$\left\|\widehat{\mathbf{q}}^*_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\eta}\right\| = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \tag{3.80}$$

Also, employing the same arguments as for (3.38) but replacing $\overline{\mathbf{U}}_w, \overline{\mathbf{C}}_w, \mathbf{T}, \overline{\mathbf{H}}_w$ with respectively $\overline{\mathbf{V}}_w, \overline{\mathbf{\Gamma}}_w, \mathbf{T}_\mathbf{x}$ and $\overline{\mathbf{H}}_{\mathbf{x},w}$ reveals

$$\widehat{\mathbf{q}}^*_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\eta} = 2\sqrt{\tau}\mathbf{q}'_{xy}\mathbf{\Sigma}'_\eta vec((\mathbf{\Gamma}^\dagger)'\mathbf{\Sigma}\mathbf{D}_{\mathbf{x},-m}\mathbf{\Sigma}\mathbf{\Gamma}^\dagger) + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \tag{3.81}$$

with $\mathbf{D}_{\mathbf{x},-m} = \mathbf{T}_\mathbf{x}\mathbf{H}_{\mathbf{x},-m}(\mathbf{H}'_{\mathbf{x},-m}\mathbf{T}'_\mathbf{x}\mathbf{\Sigma}\mathbf{T}_\mathbf{x}\mathbf{H}_{\mathbf{x},-m})^\dagger\mathbf{H}'_{\mathbf{x},-m}\mathbf{T}'_\mathbf{x}$. The latter follows since replacing $\overline{\mathbf{U}}_w$ with $\overline{\mathbf{V}}_w$ yields in eq.(3.37) that $\sqrt{N}T^{-1}\overline{\mathbf{V}}_w\overline{\mathbf{V}}^0_{w,-m} = 2\mathbf{\Sigma}\mathbf{T}_\mathbf{x}\mathbf{H}_{\mathbf{x},-m} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$ because also replacing $\overline{\mathbf{U}}^0_{w,-m}$ with $\overline{\mathbf{V}}^0_{w,-m} = \sqrt{N}\overline{\mathbf{V}}_w\mathbf{T}_\mathbf{x}\overline{\mathbf{H}}_{\mathbf{x},-m}$ in the proof eq.(3.26) of Lemma C-3 results in

$$\widehat{\mathbf{\Sigma}}^\dagger_{\mathbf{u}^0_{w,-m}} = (1/2)(\mathbf{H}'_{\mathbf{x},-m}\mathbf{T}'_\mathbf{x}\mathbf{\Sigma}\mathbf{T}_\mathbf{x}\mathbf{H}_{\mathbf{x},-m})^\dagger + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \tag{3.82}$$

Next, we have for $\widehat{\mathbf{q}}^*_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\varepsilon}$ by substituting in the same results as in the proof for $\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\varepsilon}$ in (3.42) of Lemma C-4, defining also $\mathbf{D}_\mathbf{x} = (\overline{\mathbf{\Gamma}}^\dagger)'\mathbf{\Sigma}\mathbf{D}_{\mathbf{x},-m}$ and noting $\overline{\mathbf{V}}_w = N^{-1}(s_i\mathbf{V}_i + \sum_{j\neq i}^N s_j\mathbf{V}_j)$ that

$$\begin{aligned}
\widehat{\mathbf{q}}^*_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\varepsilon} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N s_i\mathbf{\Gamma}'_i(\overline{\mathbf{\Gamma}}^\dagger_w)'\overline{\mathbf{V}}'_w[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\boldsymbol{\varepsilon}_i \\
&= \frac{1}{N}\sum_{i=1}^N s_i\mathbf{\Gamma}'_i(\overline{\mathbf{\Gamma}}^\dagger_w)'\mathbf{\Sigma}\mathbf{D}_{\mathbf{x},-m}\sqrt{N}T^{-1/2}\overline{\mathbf{V}}'_w\boldsymbol{\varepsilon}_i + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\
&= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^N s_i^2\mathbf{\Gamma}'_i\mathbf{D}_\mathbf{x}(T^{-1}\mathbf{V}'_i\boldsymbol{\varepsilon}_i) + \frac{1}{NT}\sum_{i=1}^N\sum_{j\neq i}^N s_is_j\mathbf{\Gamma}'_i\mathbf{D}_\mathbf{x}\mathbf{V}'_j\boldsymbol{\varepsilon}_i\right] + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\
&= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^N s_i^2\mathbf{\Gamma}'_i\mathbf{D}_\mathbf{x}(T^{-1}\mathbf{V}'_i\boldsymbol{\varepsilon}_i)\right] + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\
&= O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})
\end{aligned} \tag{3.83}$$

where on the fourth line we made use of $\mathbf{V}_j = \mathbf{U}_j\mathbf{q}_x$, $\boldsymbol{\varepsilon}_i = \mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$ and the independence of s_i, s_j from the other variables (with their 4th moments being finite) to substitute in the order derived in (2.34) for the rightmost term on line three, and the fifth line makes use of $T^{-1}\mathbf{V}'_i\boldsymbol{\varepsilon}_i = O_{p^*}(T^{-1/2})$. Next up is $\widehat{\mathbf{q}}^*_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon}$. Given the relation between $\mathbf{U}_i$ and $\mathbf{V}_i$, we obtain the same result as (3.46) but where $\mathbf{U}_j$ and $\mathbf{U}_l$ are replaced with $\mathbf{V}_j, \mathbf{V}_l$ and $\mathbf{D}_{-m} = \mathbf{D}_{\mathbf{x},-m}$, so that with $d^\mathbf{x}_{v,g}$ denoting row v and column g of $\mathbf{D}_{\mathbf{x},-m}$

$$\begin{aligned}
&\widehat{\mathbf{q}}^*_{\mathbf{x},[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon} \\
&= \sqrt{\tau_{N,T}}\left[\frac{1}{NT^2}\sum_{i=1}^N\sum_{j=1}^N\sum_{l=1}^N s_is_js_l\mathbf{V}'_i\mathbf{V}_j\mathbf{D}_{\mathbf{x},-m}\mathbf{V}'_l\boldsymbol{\varepsilon}_i\right] + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\
&= \sqrt{\tau_{N,T}}\left[\sum_{v=1}^k\sum_{g=1}^k d^\mathbf{x}_{v,g}\left\{\frac{1}{NT^2}\sum_{i=1}^N\sum_{j=1}^N\sum_{l=1}^N s_is_js_l\sum_{t=1}^T\sum_{s=1}^T \mathbf{v}_{i,t}\mathrm{v}^{(v)}_{j,t}\mathrm{v}^{(g)}_{l,s}\varepsilon_{i,s}\right\}\right] + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\
&= O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})
\end{aligned} \tag{3.84}$$

where on the last line (3.49) of the proof of Lemma C-4 was substituted in. Combining then (3.80), (3.81), (3.83) and (3.84) in the definition of $\widehat{\mathbf{q}}^*_{\mathbf{x},[\mathbf{M}_{\mathbf{F}0}-\mathbf{M}_{\widehat{\mathbf{F}}0*}]}$ gives

$$\widehat{\mathbf{q}}^*_{\mathbf{x},[\mathbf{M}_{\mathbf{F}0}-\mathbf{M}_{\widehat{\mathbf{F}}0*}]} \longrightarrow^{p^*} 2\sqrt{\tau}\mathbf{d}_{\mathbf{x}} \tag{3.85}$$

where we recall from Corollary 1 that $\mathbf{d}_{\mathbf{x}} = \mathbf{q}'_{xy}\boldsymbol{\Sigma}'_{\eta} vec((\boldsymbol{\Gamma}^{\dagger})'\boldsymbol{\Sigma}\mathbf{D}_{\mathbf{x},-m}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{\dagger})$.

Consider next

$$\widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I}} = \widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},V\varepsilon} - \widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},V\eta} - \widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},\Gamma\epsilon} + \widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},\Gamma\eta}$$

As before, the fact that $\overline{\mathbf{X}}^* \subset \overline{\mathbf{Z}}^*, \overline{\boldsymbol{\Gamma}}_w \subset \overline{\mathbf{C}}_w, \overline{\mathbf{V}}_w \subset \overline{\mathbf{U}}_w$ implies that the orders derived in Lemma C-6 are upper bounds for the analysis with $\overline{\mathbf{Z}}^*$ replaced by $\overline{\mathbf{X}}^*$, so that it follows directly from (3.62) of the proof of Lemma C-6

$$\left\|\widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},V\eta}\right\| = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \tag{3.86}$$

Then, for the last term in this decomposition

$$\begin{aligned}
\widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},\Gamma\eta} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} s_i \boldsymbol{\Gamma}'_i(\overline{\boldsymbol{\Gamma}}^{\dagger}_w)'\overline{\mathbf{V}}'_w\overline{\mathbf{V}}_w\overline{\boldsymbol{\Gamma}}^{\dagger}_w\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y = \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N} s_i(\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y\otimes\boldsymbol{\Gamma}_i)\right]' vec\left((\overline{\boldsymbol{\Gamma}}^{\dagger}_w)'NT^{-1}\overline{\mathbf{V}}'_w\overline{\mathbf{V}}_w\overline{\boldsymbol{\Gamma}}^{\dagger}_w\right)\\
&= 2\sqrt{\tau}\mathbf{q}'_{xy}\boldsymbol{\Sigma}'_{\eta} vec\left((\boldsymbol{\Gamma}^{\dagger})'\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{\dagger}\right) + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})
\end{aligned}\tag{3.87}$$

where we have substituted (3.22) of Lemma C-2 into $NT^{-1}\overline{\mathbf{V}}'_w\overline{\mathbf{V}}_w = \mathbf{q}'_x(NT^{-1}\overline{\mathbf{U}}'_w\overline{\mathbf{U}}_w)\mathbf{q}_x = 2\mathbf{q}'_x\boldsymbol{\Sigma}_{\mathbf{u}}\mathbf{q}_x + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) = 2\boldsymbol{\Sigma} + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$. Next,

$$\begin{aligned}
\widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},\Gamma\epsilon} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} s_i\boldsymbol{\Gamma}'_i(\overline{\boldsymbol{\Gamma}}^{\dagger}_w)'\overline{\mathbf{V}}'_w\boldsymbol{\varepsilon}_i = \sqrt{\tau_{N,T}}\sum_{i=1}^{N} s_i\boldsymbol{\Gamma}'_i(\overline{\boldsymbol{\Gamma}}^{\dagger}_w)'T^{-1}\overline{\mathbf{V}}'_w\boldsymbol{\varepsilon}_i\\
&= \sqrt{\tau_{N,T}}\left[\frac{1}{N}\sum_{i=1}^{N} s_i^2\boldsymbol{\Gamma}'_i(\overline{\boldsymbol{\Gamma}}^{\dagger}_w)'(T^{-1}\mathbf{V}'_i\boldsymbol{\varepsilon}_i)\right] + \sqrt{\tau_{N,T}}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j\neq i}^{N} s_i s_j\boldsymbol{\Gamma}'_i(\overline{\boldsymbol{\Gamma}}^{\dagger}_w)'\mathbf{V}'_j\boldsymbol{\varepsilon}_i\right]\\
&= O_{p^*}(T^{-1/2})
\end{aligned}\tag{3.88}$$

since given the independence of s_i, s_j from the other variables and $\mathbf{V}_j = \mathbf{U}_j\mathbf{q}_x$, $\boldsymbol{\varepsilon}_i = \mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$ the rightmost term on the second line is $O_{p^*}(T^{-1/2})$ by the same arguments as for (2.34) in the proof of Lemma B-3, and for the left term we have used $T^{-1}\mathbf{V}'_i\boldsymbol{\varepsilon}_i = O_{p^*}(T^{-1/2})$. Then, since $\widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},V\varepsilon} = \widehat{\mathbf{q}}^*_{\mathbf{I},V\varepsilon}$, the result in eq.(3.63) of the proof for Lemma C-6 directly applies and we can combine it together with (3.86), (3.87) and (3.88) in the decomposition of $\widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I}}$ to conclude that as $(N,T)\to\infty$

$$\widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I}} = \widehat{\mathbf{q}}^*_{\mathbf{I},V\varepsilon} + 2\sqrt{\tau}\mathbf{b}_{\mathbf{x}} + o_{p^*}(1) \tag{3.89}$$

which implies in turn

$$\widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I}} \xrightarrow{d^*} \mathcal{N}\left(\mathbf{0}_{k\times 1}, 2\boldsymbol{\Psi}\right) + 2\sqrt{\tau}\mathbf{b}_{\mathbf{x}} \tag{3.90}$$

with $\mathbf{b_x} = \mathbf{q}'_{xy}\boldsymbol{\Sigma}'_{\eta}vec((\boldsymbol{\Gamma}^{\dagger})'\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{\dagger})$.

Consider then the following expansion of $\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}}$ around $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$

$$
\begin{aligned}
\sqrt{NT}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}} - \boldsymbol{\beta}) &= (\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1}\widehat{\mathbf{q}}^*_{\mathbf{x}} \\
\sqrt{NT}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}} - \boldsymbol{\beta}) - \sqrt{NT}(\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \boldsymbol{\beta}) &= (\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1}\widehat{\mathbf{q}}^*_{\mathbf{x}} - \widehat{\mathbf{Q}}^{-1}_{\mathbf{x}}\widehat{\mathbf{q}}_{\mathbf{x}} \\
\sqrt{NT}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}} - \widehat{\boldsymbol{\beta}}_{\mathbf{x}}) &= (\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1}[\widehat{\mathbf{q}}^*_{\mathbf{x}} - \widehat{\mathbf{q}}_{\mathbf{x}}] + [(\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1} - \widehat{\mathbf{Q}}^{-1}_{\mathbf{x}}]\widehat{\mathbf{q}}_{\mathbf{x}} \\
&= (\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1}[\widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},V\varepsilon} - \widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},V\varepsilon}] + [\widehat{\mathbf{k}}^*_{\mathbf{x}} - \widehat{\mathbf{k}}_{\mathbf{x}}] + [(\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1} - \widehat{\mathbf{Q}}^{-1}_{\mathbf{x}}]\widehat{\mathbf{q}}_{\mathbf{x}} \\
&= (\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1}\widetilde{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},V\varepsilon} + [\widehat{\mathbf{k}}^*_{\mathbf{x}} - \widehat{\mathbf{k}}_{\mathbf{x}}] + [(\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1} - \widehat{\mathbf{Q}}^{-1}_{\mathbf{x}}]\widehat{\mathbf{q}}_{\mathbf{x}} \qquad (3.91)
\end{aligned}
$$

where $\widetilde{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},V\varepsilon} = [\widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},V\varepsilon} - \widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},V\varepsilon}]$ and

$$\widehat{\mathbf{k}}_{\mathbf{x}} = (\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1}[\widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},\Gamma\eta} - \widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},V\eta} - \widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},\Gamma\varepsilon} - \widehat{\mathbf{q}}_{\mathbf{x},\mathbf{P}_{\mathbf{F}0}} - \widehat{\mathbf{q}}_{\mathbf{x},[\mathbf{M}_{\mathbf{F}0}-\mathbf{M}_{\widehat{\mathbf{F}}0}]}] \longrightarrow^{p^*} \sqrt{\tau}\boldsymbol{\Sigma}^{-1}\mathbf{g} \qquad (3.92)$$

$$\widehat{\mathbf{k}}^*_{\mathbf{x}} = (\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1}[\widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},\Gamma\eta} - \widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},V\eta} - \widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},\Gamma\varepsilon} - \widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{P}_{\mathbf{F}0}} - \widehat{\mathbf{q}}^*_{\mathbf{x},[\mathbf{M}_{\mathbf{F}0}-\mathbf{M}_{\widehat{\mathbf{F}}0*}]}] \longrightarrow^{p^*} 2\sqrt{\tau}\boldsymbol{\Sigma}^{-1}\mathbf{g} \qquad (3.93)$$

with $\mathbf{g} = \mathbf{b_x} - \mathbf{d_x} = \mathbf{q}'_{xy}\boldsymbol{\Sigma}'_{\eta}vec((\boldsymbol{\Gamma}^{\dagger})'\boldsymbol{\Sigma}(\mathbf{I}_k - \mathbf{D}_{\mathbf{x},-m}\boldsymbol{\Sigma})\boldsymbol{\Gamma}^{\dagger})$. The latter results follow from substituting into both expressions $(\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1} \longrightarrow^{p^*} \boldsymbol{\Sigma}^{-1}$ obtained in (3.78) together with (2.66), (2.72), (2.73), (2.74) and (2.75) obtained in Corollary 1 on the first line, and (3.79), (3.85), (3.86), (3.87) and (3.88) on the second line. Combining (3.92)-(3.93) then gives

$$[\widehat{\mathbf{k}}^*_{\mathbf{x}} - \widehat{\mathbf{k}}_{\mathbf{x}}] \longrightarrow^{p^*} \sqrt{\tau}\boldsymbol{\Sigma}^{-1}\mathbf{g} \qquad (3.94)$$

For the last term in (3.91) we know from (2.65) and (2.78) of Corollary 1 that $\widehat{\mathbf{Q}}^{-1}_{\mathbf{x}} \longrightarrow^{p} \boldsymbol{\Sigma}^{-1}$ and $\|\widehat{\mathbf{q}}_{\mathbf{x}}\| = O_p(1)$, respectively. Hence, with (3.78) this results in

$$[(\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1} - \widehat{\mathbf{Q}}^{-1}_{\mathbf{x}}]\widehat{\mathbf{q}}_{\mathbf{x}} \longrightarrow^{p^*} \mathbf{0}_{k\times 1} \qquad (3.95)$$

For the first term in (3.91) note that $\widetilde{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},V\varepsilon} = [\widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},V\varepsilon} - \widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},V\varepsilon}] = [\widehat{\mathbf{q}}^*_{\mathbf{I},V\varepsilon} - \widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}] = \widetilde{\mathbf{q}}^*_{\mathbf{I},V\varepsilon}$ because we have by definition that $\widehat{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},V\varepsilon} = \widehat{\mathbf{q}}^*_{\mathbf{I},V\varepsilon}$ and $\widehat{\mathbf{q}}_{\mathbf{x},\mathbf{I},V\varepsilon} = \widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}$. Hence, the result in (3.69) of Theorem 2 directly applies and we obtain again making use of (3.78) that

$$(\widehat{\mathbf{Q}}^*)^{-1}\widetilde{\mathbf{q}}_{\mathbf{I},V\varepsilon} \xrightarrow{d^*} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}) \qquad (3.96)$$

Finally, combining (3.94), (3.95) and (3.96) into (3.91) leads to the conclusion that

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}} - \widehat{\boldsymbol{\beta}}_{\mathbf{x}}) \xrightarrow{d^*} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}) + \sqrt{\tau}\boldsymbol{\Sigma}^{-1}\mathbf{g}$$

as was to be shown. This result, together with that of Corollary 1, directly implies

$$\sup_{x\in\mathbb{R}^{k\times 1}} \left| \mathbb{P}^*[\sqrt{NT}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}} - \widehat{\boldsymbol{\beta}}_{\mathbf{x}}) \le x] - \mathbb{P}[\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \boldsymbol{\beta}) \le x] \right| \longrightarrow^{p} 0,$$

where the inequalities are to be interpreted coordinate wise.

Theorem 3 *Under Ass. 1-5 strengthened with* $\mathbb{E}(\|\mathbf{v}_{i,t}\|^8) < \infty$ *we have as* $(N,T) \to \infty$ *such that* $T/N \to \tau < \infty$ *that* $\mathbf{A}^* \longrightarrow^{p^*} \sqrt{\tau}\mathbf{A}$*, where* $\mathbf{A}^* = \mathbb{E}^*(\sqrt{NT}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}))$ *and* $\mathbf{A} = \boldsymbol{\Sigma}^{-1}(\mathbf{b} - \mathbf{d} - \mathbf{d}^+)$*. If in addition Ass. 7 holds, then* $\mathbf{A}^*_{\mathbf{x}} \longrightarrow^{p^*} \sqrt{\tau}\mathbf{A}_{\mathbf{x}}$*, with* $\mathbf{A}^*_{\mathbf{x}} = \mathbb{E}^*(\sqrt{NT}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}} - \widehat{\boldsymbol{\beta}}_{\mathbf{x}}))$ *and* $\mathbf{A}^*_{\mathbf{x}} = \boldsymbol{\Sigma}^{-1}\mathbf{g}$.

Proof of Theorem 3

We begin the proof by verifying that the bootstrap sequences $\left\{\left\|\sqrt{NT}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})\right\|\right\}$ and $\left\{\left\|\sqrt{NT}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}} - \widehat{\boldsymbol{\beta}}_{\mathbf{x}})\right\|\right\}$ are uniformly integrable. To that end, we follow the approach of Gonçalves and Kaffo (2015) (proof of Theorem 3.2) and demonstrate that $\mathbb{E}^*\left(\left\|\sqrt{NT}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})\right\|^{1+\delta}\right) = O_{p^*}(1)$ in probability for some $\delta > 0$, which is a sufficient condition for uniform integrability. Uniform integrability of the sequences then enables Theorem 25.12 in Billingsley (1995), which in combination with Theorem 2 and Corollary 2 then establishes the respective statements of the theorem.

To demonstrate integrability, we set $\delta = 1$ and recall the following expansion from the proof of Theorem 2

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) = (\widehat{\mathbf{Q}}^*)^{-1}\tilde{\mathbf{q}}_{\mathbf{I},V\varepsilon} + [\widehat{\mathbf{k}}^* - \widehat{\mathbf{k}}] + [(\widehat{\mathbf{Q}}^*)^{-1} - (\widehat{\mathbf{Q}})^{-1}]\widehat{\mathbf{q}}.$$

By the Cauchy-Schwarz inequality, $\|\mathbf{x}+\mathbf{y}\|^2 \le (\|\mathbf{x}\| + \|\mathbf{y}\|)^2\ \forall \mathbf{x},\mathbf{y} \in \mathbb{R}^k$ and $(x+y)^2 \le 2(x^2+y^2)\ \forall x, y$, we obtain the following upper bound:

$$\begin{aligned}
\mathbb{E}^*\left(\left\|\sqrt{NT}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})\right\|^2\right) &= \mathbb{E}^*\left(\left\|(\widehat{\mathbf{Q}}^*)^{-1}\tilde{\mathbf{q}}_{\mathbf{I},V\varepsilon} + [\widehat{\mathbf{k}}^* - \widehat{\mathbf{k}}] + [(\widehat{\mathbf{Q}}^*)^{-1} - (\widehat{\mathbf{Q}})^{-1}]\widehat{\mathbf{q}}\right\|^2\right) \\
&\le \mathbb{E}^*\left(\left[\left\|(\widehat{\mathbf{Q}}^*)^{-1}\tilde{\mathbf{q}}_{\mathbf{I},V\varepsilon}\right\| + \left\|[\widehat{\mathbf{k}}^* - \widehat{\mathbf{k}}] + [(\widehat{\mathbf{Q}}^*)^{-1} - (\widehat{\mathbf{Q}})^{-1}]\widehat{\mathbf{q}}\right\|\right]^2\right) \\
&\le 2\mathbb{E}^*\left(\left\|(\widehat{\mathbf{Q}}^*)^{-1}\tilde{\mathbf{q}}_{\mathbf{I},V\varepsilon}\right\|^2\right) + 2\mathbb{E}^*\left(\left\|[\widehat{\mathbf{k}}^* - \widehat{\mathbf{k}}] + [(\widehat{\mathbf{Q}}^*)^{-1} - (\widehat{\mathbf{Q}})^{-1}]\widehat{\mathbf{q}}\right\|^2\right) \\
&\le 2\mathbb{E}^*\left(\left\|(\widehat{\mathbf{Q}}^*)^{-1}\tilde{\mathbf{q}}_{\mathbf{I},V\varepsilon}\right\|^2\right) + 2\mathbb{E}^*\left(\left[\left\|[\widehat{\mathbf{k}}^* - \widehat{\mathbf{k}}]\right\| + \left\|[(\widehat{\mathbf{Q}}^*)^{-1} - (\widehat{\mathbf{Q}})^{-1}]\widehat{\mathbf{q}}\right\|\right]^2\right) \\
&\le 2\left[\mathbb{E}^*\left(\left\|(\widehat{\mathbf{Q}}^*)^{-1}\right\|^4\right)\right]^{1/2} \times \left[\mathbb{E}^*\left(\|\tilde{\mathbf{q}}_{\mathbf{I},V\varepsilon}\|^4\right)\right]^{1/2} \\
&+ 4\mathbb{E}^*\left(\left\|[\widehat{\mathbf{k}}^* - \widehat{\mathbf{k}}]\right\|^2\right) + 4\mathbb{E}^*\left(\left\|[(\widehat{\mathbf{Q}}^*)^{-1} - (\widehat{\mathbf{Q}})^{-1}]\widehat{\mathbf{q}}\right\|^2\right) \\
&\le 2\left[\mathbb{E}^*\left(\left\|(\widehat{\mathbf{Q}}^*)^{-1}\right\|^4\right)\right]^{1/2} \times \left[\mathbb{E}^*\left(\|\tilde{\mathbf{q}}_{\mathbf{I},V\varepsilon}\|^4\right)\right]^{1/2} \\
&+ 4\mathbb{E}^*\left(\left\|[\widehat{\mathbf{k}}^* - \widehat{\mathbf{k}}]\right\|^2\right) + 4\left[\mathbb{E}^*\left(\left\|(\widehat{\mathbf{Q}}^*)^{-1} - (\widehat{\mathbf{Q}})^{-1}\right\|^4\right)\right]^{1/2} \times \left[\mathbb{E}^*\left(\|\widehat{\mathbf{q}}\|^4\right)\right]^{1/2}
\end{aligned} \tag{3.97}$$

where we shall now examine boundedness term by term. Clearly, using the arguments from the proof of Theorem 2 in combination with $\|\mathbf{x}+\mathbf{y}\|^2 \le (\|\mathbf{x}\| + \|\mathbf{y}\|)^2\ \forall \mathbf{x},\mathbf{y} \in \mathbb{R}^k$ and $(x+y)^2 \le 2(x^2+y^2)\ \forall x, y$, we obtain

$$\mathbb{E}^*\left(\left\|[\widehat{\mathbf{k}}^* - \widehat{\mathbf{k}}]\right\|^2\right) = \mathbb{E}^*\left(\left\|\sqrt{\tau}\boldsymbol{\Sigma}^{-1}(\mathbf{b} - \mathbf{d} - \mathbf{d}^+) + \mathbf{r}\right\|^2\right)$$
$$\leq \mathbb{E}^*\left(\left[\left\|\sqrt{\tau}\boldsymbol{\Sigma}^{-1}(\mathbf{b} - \mathbf{d} - \mathbf{d}^+)\right\| + \|\mathbf{r}\|\right]^2\right)$$
$$\leq 2\left\|\sqrt{\tau}\boldsymbol{\Sigma}^{-1}(\mathbf{b} - \mathbf{d} - \mathbf{d}^+)\right\|^2 + 2\mathbb{E}^*\left(\|\mathbf{r}\|^2\right) = O_{p^*}(1), \tag{3.98}$$

where the residual is $\|\mathbf{r}\|^2 = O_{p^*}(N^{-1}) + O_{p^*}(T^{-1})$ in probability by (2.53), (2.54), (2.56), (3.60), (3.61), (3.62). Therefore the second term above vanishes as $(N, T) \to \infty$, while the first term is bounded. Further, using the definition of the Frobenius norm yields

$$\mathbb{E}^*\left(\|\widetilde{\mathbf{q}}_{\mathbf{I},V\varepsilon}\|^4\right) = \mathbb{E}^*\left[\left(\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}(s_i - 1)\mathbf{V}_i'\boldsymbol{\varepsilon}_i\right)' \frac{1}{\sqrt{NT}}\sum_{j=1}^{N}(s_j - 1)\mathbf{V}_j'\boldsymbol{\varepsilon}_j\right)^2\right]$$
$$= \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{k=1}^{N}\mathbb{E}^*\left((s_i - 1)(s_j - 1)(s_l - 1)(s_k - 1)\boldsymbol{\varepsilon}_i'\mathbf{V}_i\mathbf{V}_j'\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_l'\mathbf{V}_l\mathbf{V}_k'\boldsymbol{\varepsilon}_k\right). \tag{3.99}$$

Note that because $\mathbb{E}^*\left(\mathbf{V}_i'\varepsilon_j\right) = \mathbf{0}_k$ for all i, j, the expression above is non-zero only when $i = j = l = k$ or if there are at least two pairs of identical indices (e.g. $i = j$ and $l = k$). Therefore, we further examine these two cases. To begin with, when $i = j = l = k$, we have

$$\frac{1}{N^2T^2}\sum_{i=1}^{N}\mathbb{E}^*\left((s_i - 1)^4\boldsymbol{\varepsilon}_i'\mathbf{V}_i\mathbf{V}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{V}_i\mathbf{V}_i'\boldsymbol{\varepsilon}_i\right)$$
$$= \frac{1}{N^2T^2}\sum_{i=1}^{N}\mathbb{E}^*\left((s_i - 1)^4\right)\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{p=1}^{T}\mathbb{E}^*\left(\mathbf{v}_{i,t}'\mathbf{v}_{i,s}\mathbf{v}_{i,r}'\mathbf{v}_{i,p}\right)\mathbb{E}^*\left(\varepsilon_{i,t}\varepsilon_{i,s}\varepsilon_{i,r}\varepsilon_{i,p}\right),$$

which follows from independence of the bootstrap weights and model primitives. This implies that

$$\left|\frac{1}{N^2T^2}\sum_{i=1}^{N}\mathbb{E}^*\left((s_i - 1)^4\boldsymbol{\varepsilon}_i'\mathbf{V}_i\mathbf{V}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{V}_i\mathbf{V}_i'\boldsymbol{\varepsilon}_i\right)\right|$$
$$\leq \sup_{i,t,s,r,p}\left(\left|\mathbb{E}^*\left(\varepsilon_{i,t}\varepsilon_{i,s}\varepsilon_{i,r}\varepsilon_{i,p}\right)\right|\right) \times \mathbb{E}^*\left((s_i - 1)^4\right) \times \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{p=1}^{T}\left|\mathbb{E}^*\left(\mathbf{v}_{i,t}'\mathbf{v}_{i,s}\mathbf{v}_{i,r}'\mathbf{v}_{i,p}\right)\right|$$
$$= \sup_{i,t,s,r,p}\left(\left|\mathbb{E}^*\left(\varepsilon_{i,t}\varepsilon_{i,s}\varepsilon_{i,r}\varepsilon_{i,p}\right)\right|\right) \times \mathbb{E}^*\left((s_i - 1)^4\right) \times \left(\frac{T}{N}\right)\frac{1}{NT^3}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{p=1}^{T}\left|\mathbb{E}^*\left(\mathbf{v}_{i,t}'\mathbf{v}_{i,s}\mathbf{v}_{i,r}'\mathbf{v}_{i,p}\right)\right|$$
$$= O_{p^*}(1),$$

because $T/N = O(1)$ under our assumptions, $\mathbb{E}^*\left((s_i - 1)^4\right) \in \mathbb{R}_+$ and it is homogeneous across i and, finally,

$$\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{p=1}^{T}\left|\mathbb{E}^*\left(\mathbf{v}_{i,t}'\mathbf{v}_{i,s}\mathbf{v}_{i,r}'\mathbf{v}_{i,p}\right)\right| = O_{p^*}(T^3), \tag{3.100}$$

because 4th moments are finite and we do not have any pairs in (3.100) with a common time index. Letting say $t \to \infty$, this implies that the dependence of $\mathbf{v}_{i,t}$ with the other members of the product dies out (due to stationarity) for any given combination of the other indices s, r, p. Hence, the sum of expectations over one of the indices dies out and is summable as $T \to \infty$, which implies the stated order. The next case occurs when there are two pairs of common indices, say $i = j$ and $l = k$ with $i \neq l$. Here, we obtain

$$\begin{aligned}
&\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}^*\left((s_i-1)^2(s_j-1)^2\boldsymbol{\varepsilon}_i'\mathbf{V}_i\mathbf{V}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j'\mathbf{V}_j\mathbf{V}_j'\boldsymbol{\varepsilon}_j\right)\\
&=\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}^*\left((s_i-1)^2(s_j-1)^2\right)\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{p=1}^{T}\mathbb{E}^*\left(\mathbf{v}_{i,t}'\mathbf{v}_{i,s}\mathbf{v}_{j,r}'\mathbf{v}_{j,p}\right)\mathbb{E}^*\left(\varepsilon_{i,t}\varepsilon_{i,s}\varepsilon_{j,r}\varepsilon_{j,p}\right)\\
&=\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}^*\left((s_i-1)^2(s_j-1)^2\right)\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{p=1}^{T} tr\left[\mathbb{C}ov^*\left(\mathbf{v}_{i,t},\mathbf{v}_{i,s}\right)\right]\times tr\left[\mathbb{C}ov^*\left(\mathbf{v}_{j,r},\mathbf{v}_{j,p}\right)\right]\\
&\quad\times\mathbb{C}ov^*\left(\varepsilon_{i,t},\varepsilon_{i,s}\right)\times\mathbb{C}ov^*\left(\varepsilon_{j,r},\varepsilon_{j,p}\right),
\end{aligned}$$

which implies that

$$\begin{aligned}
&\left|\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}^*\left((s_i-1)^2(s_j-1)^2\boldsymbol{\varepsilon}_i'\mathbf{V}_i\mathbf{V}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j'\mathbf{V}_j\mathbf{V}_j'\boldsymbol{\varepsilon}_j\right)\right|\\
&\leq\sup_{i,j}\left(\mathbb{E}^*\left((s_i-1)^2(s_j-1)^2\right)\right)\times\sup_{i,t,s}\left(\left|\mathbb{C}ov^*\left(\varepsilon_{i,t},\varepsilon_{i,s}\right)\right|\right)\times\sup_{j,r,p}\left(\left|\mathbb{C}ov^*\left(\varepsilon_{i,r},\varepsilon_{j,p}\right)\right|\right)\\
&\quad\times\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{p=1}^{T}\left|tr\left[\mathbb{C}ov^*\left(\mathbf{v}_{i,t},\mathbf{v}_{i,s}\right)\right]\right|\times\left|tr\left[\mathbb{C}ov^*\left(\mathbf{v}_{j,r},\mathbf{v}_{j,p}\right)\right]\right|\\
&=\sup_{i,j}\left(\mathbb{E}^*\left((s_i-1)^2(s_j-1)^2\right)\right)\times\sup_{i,t,s}\left(\left|\mathbb{C}ov^*\left(\varepsilon_{i,t},\varepsilon_{i,s}\right)\right|\right)\times\sup_{j,r,p}\left(\left|\mathbb{C}ov^*\left(\varepsilon_{j,r},\varepsilon_{j,p}\right)\right|\right)\\
&\quad\times\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|tr\left[\mathbb{C}ov^*\left(\mathbf{v}_{i,t},\mathbf{v}_{i,s}\right)\right]\right|\right)\times\left(\frac{1}{T}\sum_{r=1}^{T}\sum_{p=1}^{T}\left|tr\left[\mathbb{C}ov^*\left(\mathbf{v}_{j,r},\mathbf{v}_{j,p}\right)\right]\right|\right)\\
&=O_{p^*}(1)
\end{aligned}$$

because of the absolute summability of the covariances and the fact that the sum over individuals is $O(N^2)$. This implies that overall

$$\mathbb{E}^*\left(\|\widetilde{\mathbf{q}}_{\mathbf{I},V\varepsilon}\|^4\right)=O_{p^*}(1) \tag{3.101}$$

in probability.

Next, we evaluate $\mathbb{E}^*\left(\|\widehat{\mathbf{q}}\|^4\right)$. Here, we use the same convenient decomposition $\widehat{\mathbf{q}}=\widehat{\mathbf{q}}_{\mathbf{I}}-\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}}-\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]}$ and their respective sub-decompositions (recall section 2.2) to write:

$$\widehat{\mathbf{q}}=\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}+\mathbf{a}_1+\mathbf{r}_1$$

where $\mathbf{a}_1=-\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\varepsilon}+\widehat{\mathbf{q}}_{\mathbf{I},\Gamma\eta}-\left(\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon}-\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\varepsilon}+\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],\Gamma\eta}\right)$ collects all the terms that lead to bias, and $\mathbf{r}_1=-\widehat{\mathbf{q}}_{\mathbf{I},V\eta}-\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}}+\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}+\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\eta}$ contains all the vanishing terms. Combining (2.37), (2.56) and Lemma B-4 implies $\mathbf{r}_1 = O_p(N^{-1/2}) + O_p(T^{-1/2})$. In addition, from (2.30), (2.35), (2.44), (2.53) and (2.54) follows $\mathbf{a}_1 = \sqrt{\tau}(\mathbf{b}-\mathbf{d}) + O_p(N^{-1/2}) + O_p(T^{-1/2})$. Hence, we can write

$$\widehat{\mathbf{q}} = \widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon} + \mathbf{a}_1 + \mathbf{r}_1 = \widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon} + \sqrt{\tau}(\mathbf{b}-\mathbf{d}) + \mathbf{r}$$

where $\mathbf{r} = O_p(N^{-1/2}) + O_p(T^{-1/2})$.

Therefore, using $\|\mathbf{x}+\mathbf{y}\|^2 \leq (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 \;\forall \mathbf{x},\mathbf{y} \in \mathbb{R}^k$ and $(x+y)^2 \leq 2(x^2+y^2) \;\forall x,y$ iteratively, we obtain

$$\begin{aligned}
\mathbb{E}^*\left(\|\widehat{\mathbf{q}}\|^4\right) &= \mathbb{E}^*\left(\left\|\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon} + \sqrt{\tau}(\mathbf{b}-\mathbf{d}) + \mathbf{r}\right\|^4\right) \\
&\leq \mathbb{E}^*\left[\left(\left[\|\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}\| + \left\|\sqrt{\tau}(\mathbf{b}-\mathbf{d}) + \mathbf{r}\right\|\right]^2\right)^2\right] \\
&\leq \mathbb{E}^*\left[\left(2\left\|\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}\right\|^2 + 2\left\|\sqrt{\tau}(\mathbf{b}-\mathbf{d}) + \mathbf{r}\right\|^2\right)^2\right] \\
&\leq \mathbb{E}^*\left[4\left\|\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}\right\|^4 + 4\left\|\sqrt{\tau}(\mathbf{b}-\mathbf{d}) + \mathbf{r}\right\|^4\right] \\
&= 4\mathbb{E}^*\left(\left\|\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}\right\|^4\right) + 4\mathbb{E}^*\left(\left\|\sqrt{\tau}(\mathbf{b}-\mathbf{d}) + \mathbf{r}\right\|^4\right) \\
&\leq 4\mathbb{E}^*\left(\left\|\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}\right\|^4\right) + 4\mathbb{E}^*\left[\left(\left(\left\|\sqrt{\tau}(\mathbf{b}-\mathbf{d})\right\| + \|\mathbf{r}\|\right)^2\right)^2\right] \\
&\leq 4\mathbb{E}^*\left(\left\|\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}\right\|^4\right) + 16\mathbb{E}^*\left(\left\|\sqrt{\tau}(\mathbf{b}-\mathbf{d})\right\|^4\right) + 16\mathbb{E}^*\left(\|\mathbf{r}\|^4\right) \\
&= O_{p^*}(1),
\end{aligned} \tag{3.102}$$

since $\mathbb{E}^*(\|\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon}\|^4) = O_{p^*}(1)$ by the same logic as that used to obtain (3.101), $\mathbb{E}^*\left(\left\|\sqrt{\tau}(\mathbf{b}-\mathbf{d})\right\|^4\right) = O(1)$ because $\tau = O(1)$ and $\mathbf{b},\mathbf{d}$ are fixed finite vectors, and $\|\mathbf{r}\|^4 = o_p(1)$ due to $\|\mathbf{r}\| = O_p(N^{-1/2}) + O_p(T^{-1/2})$. Further, we verify that $\mathbb{E}^*\left(\left\|(\widehat{\mathbf{Q}}^*)^{-1}\right\|^4\right) = O_{p^*}(1)$. We will verify this using $\widehat{\mathbf{Q}}^*$, because the result for the inverse is implied by the continuous mapping theorem. We begin by recalling the decomposition

$$\widehat{\mathbf{Q}}^* = \frac{1}{NT}\sum_{i=1}^{N} s_i \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{F}}^*} \mathbf{X}_i = \widehat{\mathbf{Q}}_{\mathbf{I}}^* - \widehat{\mathbf{Q}}_{\mathbf{P}_{\mathbf{F}0}}^* - \widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}0} - \mathbf{M}_{\widehat{\mathbf{F}}0*}]}^*,$$

where $\left\|\widehat{\mathbf{Q}}_{\mathbf{P}_{\mathbf{F}0}}^* + \widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}0} - \mathbf{M}_{\widehat{\mathbf{F}}0*}]}^*\right\| = O_{p^*}(N^{-1}) + O_{p^*}(T^{-1}) + O_{p^*}((NT)^{-1/2})$ in probability was established before in the proof of Lemma C-7. Therefore, we obtain

$$\begin{aligned}
\mathbb{E}^*\left(\|\widehat{\mathbf{Q}}^*\|^4\right) &= \mathbb{E}^*\left(\left\|\widehat{\mathbf{Q}}_{\mathbf{I}}^* + \left(-\widehat{\mathbf{Q}}_{\mathbf{P}_{\mathbf{F}0}}^* - \widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}0} - \mathbf{M}_{\widehat{\mathbf{F}}0}]}^*\right)\right\|^4\right) \\
&\leq \mathbb{E}^*\left(\left(\left[\left\|\widehat{\mathbf{Q}}_{\mathbf{I}}^*\right\| + \left\|-\widehat{\mathbf{Q}}_{\mathbf{P}_{\mathbf{F}0}}^* - \widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}0} - \mathbf{M}_{\widehat{\mathbf{F}}0}]}^*\right\|\right]^2\right)^2\right) \\
&\leq \mathbb{E}^*\left(\left(2\left[\left\|\widehat{\mathbf{Q}}_{\mathbf{I}}^*\right\|^2 + \left\|-\widehat{\mathbf{Q}}_{\mathbf{P}_{\mathbf{F}0}}^* - \widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}0} - \mathbf{M}_{\widehat{\mathbf{F}}0}]}^*\right\|^2\right]\right)^2\right) \\
&\leq \mathbb{E}^*\left(4\left[\left\|\widehat{\mathbf{Q}}_{\mathbf{I}}^*\right\|^4 + \left\|-\widehat{\mathbf{Q}}_{\mathbf{P}_{\mathbf{F}0}}^* - \widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}0} - \mathbf{M}_{\widehat{\mathbf{F}}0}]}^*\right\|^4\right]\right) \\
&= 4\mathbb{E}^*\left(\left\|\widehat{\mathbf{Q}}_{\mathbf{I}}^*\right\|^4\right) + 4\mathbb{E}^*\left(\left\|-\widehat{\mathbf{Q}}_{\mathbf{P}_{\mathbf{F}0}}^* - \widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}0} - \mathbf{M}_{\widehat{\mathbf{F}}0}}^*\right\|^4\right).
\end{aligned} \tag{3.103}$$

Here, $\left\|\widehat{\mathbf{Q}}^*_{\mathbf{P}_{\mathbf{F}0}} + \widehat{\mathbf{Q}}^*_{[\mathbf{M}_{\mathbf{F}0} - \mathbf{M}_{\widehat{\mathbf{F}}0*}]}\right\|^4 = O_{p^*}(N^{-4}) + O_{p^*}(T^{-4}) + O_{p^*}((NT)^{-2})$, while the first term is bounded due to the decomposition $\widehat{\mathbf{Q}}^*_{\mathbf{I}} = \widehat{\mathbf{Q}}^*_{\mathbf{I},VV} - \widehat{\mathbf{Q}}^*_{\mathbf{I},V\Gamma} - (\widehat{\mathbf{Q}}^*_{\mathbf{I},V\Gamma})' + \widehat{\mathbf{Q}}^*_{\mathbf{I},\Gamma\Gamma}$, where $\left\|\widehat{\mathbf{Q}}^*_{\mathbf{I},V\Gamma} + (\widehat{\mathbf{Q}}^*_{\mathbf{I},V\Gamma})' + \widehat{\mathbf{Q}}^*_{\mathbf{I},\Gamma\Gamma}\right\|$ $= O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2})$ in probability (see the proof of Lemma C-7). Next, following the same steps as in (3.103), we obtain

$$\mathbb{E}^*\left(\left\|\widehat{\mathbf{Q}}^*_{\mathbf{I}}\right\|^4\right) \leq 4\mathbb{E}^*\left(\left\|\widehat{\mathbf{Q}}^*_{\mathbf{I},VV}\right\|^4\right) + 4\mathbb{E}^*\left(\left\|\widehat{\mathbf{Q}}^*_{\mathbf{I},V\Gamma} + (\widehat{\mathbf{Q}}^*_{\mathbf{I},V\Gamma})' + \widehat{\mathbf{Q}}^*_{\mathbf{I},\Gamma\Gamma}\right\|^4\right) = O_{p^*}(1),$$

where boundedness stems from the first term. In particular, using definition of the Frobenius norm, we get

$$\begin{aligned}
\mathbb{E}^*\left(\left\|\widehat{\mathbf{Q}}^*_{\mathbf{I},VV}\right\|^4\right) &= \mathbb{E}^*\left(\left[tr\left(\frac{1}{N}\sum_{i=1}^N s_i\left(\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\right)'\frac{1}{N}\sum_{i=1}^N s_j\left(\frac{\mathbf{V}_j'\mathbf{V}_j}{T}\right)\right)\right]^2\right)\\
&= \mathbb{E}^*\left(\left[tr\left(\frac{1}{N^2T^2}\sum_{i=1}^N\sum_{j=1}^N s_i s_j \mathbf{V}_i'\mathbf{V}_i\mathbf{V}_j'\mathbf{V}_j\right)\right]^2\right)\\
&= \mathbb{E}^*\left(\left[tr\left(\frac{1}{N^2T^2}\sum_{i=1}^N\sum_{j=1}^N s_i s_j \sum_{t=1}^T\sum_{s=1}^T \mathbf{v}_{i,t}\mathbf{v}_{i,t}'\mathbf{v}_{j,s}\mathbf{v}_{j,s}'\right)\right]^2\right). \qquad (3.104)
\end{aligned}$$

Now, observe that to show boundedness in probability, it is sufficient to work with the sums inside the trace operator, because explicitly accounting for the trace produces sums over some index $o = 1, \ldots, k$, where k is fixed, and therefore such sums are $O(1)$. Therefore, by squaring the terms inside the trace, we obtain

$$\begin{aligned}
&\left\|\mathbb{E}^*\left(\left[\frac{1}{N^2T^2}\sum_{i=1}^N\sum_{j=1}^N s_i s_j \sum_{t=1}^T\sum_{s=1}^T \mathbf{v}_{i,t}\mathbf{v}_{i,t}'\mathbf{v}_{j,s}\mathbf{v}_{j,s}'\right]^2\right)\right\|\\
&= \left\|\frac{1}{N^4T^4}\sum_{i=1}^N\sum_{j=1}^N\sum_{l=1}^N\sum_{k=1}^N \mathbb{E}^*\left(s_i s_j s_l s_k\right)\sum_{t=1}^T\sum_{s=1}^T\sum_{r=1}^T\sum_{p=1}^T \mathbb{E}^*\left(\mathbf{v}_{i,t}\mathbf{v}_{i,t}'\mathbf{v}_{j,s}\mathbf{v}_{j,s}'\mathbf{v}_{l,r}\mathbf{v}_{l,r}'\mathbf{v}_{k,p}\mathbf{v}_{k,p}'\right)\right\|\\
&\leq \sup_{i,j,l,k}\left(\left|\mathbb{E}^*\left(s_i s_j s_l s_k\right)\right|\right) \times \frac{1}{N^4T^4}\sum_{i=1}^N\sum_{j=1}^N\sum_{l=1}^N\sum_{k=1}^N\sum_{t=1}^T\sum_{s=1}^T\sum_{r=1}^T\sum_{p=1}^T \left\|\mathbb{E}^*\left(\mathbf{v}_{i,t}\mathbf{v}_{i,t}'\mathbf{v}_{j,s}\mathbf{v}_{j,s}'\mathbf{v}_{l,r}\mathbf{v}_{l,r}'\mathbf{v}_{k,p}\mathbf{v}_{k,p}'\right)\right\|\\
&= O_{p^*}(1) \qquad (3.105)
\end{aligned}$$

in probability, by the strengthened assumption that $\mathbb{E}(\|\mathbf{v}_{i,t}\|^8) < \infty$. Therefore, combining the results in (3.103) - (3.105), we conclude that $\mathbb{E}^*(\|\widehat{\mathbf{Q}}^*\|^4) = O_{p^*}(1)$ in probability and thus $\mathbb{E}^*(\|(\widehat{\mathbf{Q}}^*)^{-1}\|^4) = O_{p^*}(1)$ by the continuous mapping theorem.

To finish, we note that $\mathbb{E}^*\left(\left\|(\widehat{\mathbf{Q}}^*)^{-1} - (\widehat{\mathbf{Q}})^{-1}\right\|^4\right) = o_{p^*}(1)$ in probability, because

$$\left\|(\widehat{\mathbf{Q}}^*)^{-1} - (\widehat{\mathbf{Q}})^{-1}\right\| = O_{p^*}(N^{-1}) + O_{p^*}(T^{-1/2})$$

in probability. Ultimately, plugging all results back into (3.97) gives

$$\mathbb{E}^* \left(\left\| \sqrt{NT}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \right\|^2 \right) = O_{p^*}(1) \tag{3.106}$$

in probability, as was required. This establishes the uniform integrability of $\left\{ \left\| \sqrt{NT}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \right\| \right\}$, and in combination with Theorem 25.12 of Billingsley (1995) and Theorem 2, proves the first statement of this theorem. To show uniform integrability of $\left\{ \left\| \sqrt{NT}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}} - \widehat{\boldsymbol{\beta}}_{\mathbf{x}}) \right\| \right\}$, we impose Ass.7 in stead of Ass.4 and similarly obtain, starting from (3.91) of Corollary 2,

$$\mathbb{E}^* \left(\left\| \sqrt{NT}(\widehat{\boldsymbol{\beta}}^*_{\mathbf{x}} - \widehat{\boldsymbol{\beta}}_{\mathbf{x}}) \right\|^2 \right) = \mathbb{E}^* \left(\left\| (\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1} \widetilde{\mathbf{q}}^*_{\mathbf{x},\mathbf{I},V\varepsilon} + [\widehat{\mathbf{k}}^*_{\mathbf{x}} - \widehat{\mathbf{k}}_{\mathbf{x}}] + [(\widehat{\mathbf{Q}}^*_{\mathbf{x}})^{-1} - (\widehat{\mathbf{Q}}_{\mathbf{x}})^{-1}] \widehat{\mathbf{q}}_{\mathbf{x}} \right\|^2 \right) = O_{p^*}(1)$$

which follows from the same steps and arguments as above. This integrability result combined with Corollary 2 proves the second statement of the theorem.

3.4 Heterogeneous Slopes

We consider here the heterogeneous slope DGP where $\boldsymbol{\beta}_i$ is characterized by Ass.6 such that $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{v}_i$ and it is understood that also the cross-section averages $\overline{\mathbf{U}}_w, \overline{\mathbf{C}}_w$ represent the heterogeneous slope variants. Note that all the results in Section 3.2 are derived under Ass.6 and hence apply here as well. The bootstrap CCE estimators are in this setting

$$\widehat{\boldsymbol{\beta}}^* = \left(\frac{1}{NT}\sum_{i=1}^N \mathbf{X}_i^{*\prime}\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{X}_i^*\right)^{-1}\frac{1}{NT}\sum_{i=1}^N \mathbf{X}_i^{*\prime}\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{y}_i^* \qquad \widehat{\boldsymbol{\beta}}_{mg}^* = \frac{1}{N}\sum_{i=1}^N \left(\mathbf{X}_i^{*\prime}\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{X}_i^*\right)^{-1}\mathbf{X}_i^{*\prime}\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{y}_i^*$$

Given that $\mathbf{y}^* = [\mathbf{y}_1^{*\prime}, \ldots, \mathbf{y}_N^{*\prime}]' = \mathbf{W}_T\mathbf{y}$ and $\mathbf{X}^* = [\mathbf{X}_1^{*\prime}, \ldots, \mathbf{X}_N^{*\prime}]' = \mathbf{W}_T\mathbf{X}$ it is equivalent to write the above more explicitly as

$$\widehat{\boldsymbol{\beta}}^* = \left(\frac{1}{NT}\sum_{i=1}^N s_i\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{X}_i\right)^{-1}\frac{1}{NT}\sum_{i=1}^N s_i\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{y}_i, \qquad \widehat{\boldsymbol{\beta}}_{mg}^* = \frac{1}{N}\sum_{i=1}^N s_i\left(\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{y}_i$$

Here, we obtain from substituting in (3.15) and $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{v}_i$ for the scaled deviation of the Mean Group CCE estimator in the bootstrap world

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{mg}^* - \boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^N s_i\widehat{\mathbf{Q}}_i^{*-1}[\widehat{\mathbf{q}}_{v,i}^* + \widehat{\mathbf{q}}_i^*] = \frac{1}{\sqrt{N}}\sum_{i=1}^N s_i\boldsymbol{v}_i + \frac{1}{N}\sum_{i=1}^N s_i\widehat{\mathbf{Q}}_i^{*-1}\widehat{\mathbf{q}}_i^* \tag{3.107}$$

and in turn, for the scaled deviation of the CCEP estimator in the bootstrap, making use of (3.15), $\gamma_i = \mathbf{C}_i\mathbf{B}_i^{-1}\mathbf{q}_y = \gamma + \widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y$ and $\sum_{i=1}^N s_i\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^\dagger\gamma = N\overline{\mathbf{X}}^{*\prime}\mathbf{M}_{\widehat{\mathbf{F}}^*}\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^\dagger\gamma = \mathbf{0}_{k\times 1}$, because $\overline{\mathbf{X}}^* \subset \overline{\mathbf{Z}}^*$,

$$\begin{aligned}\sqrt{N}(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}) &= \left(\frac{1}{NT}\sum_{i=1}^N s_i\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{N}T}\sum_{i=1}^N s_i\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}[\mathbf{X}_i\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^\dagger\widetilde{\boldsymbol{\eta}}_i\mathbf{q}_y] \\ &= \overline{\mathbf{Q}}^{*-1}[\overline{\mathbf{q}}^* + \overline{\mathbf{q}}_v^*],\end{aligned} \tag{3.108}$$

where in (3.107) and (3.108) we have defined

$$\begin{aligned}\overline{\mathbf{Q}}^* &= \frac{1}{N}\sum_{i=1}^N s_i\widehat{\mathbf{Q}}_i^*, \qquad &\widehat{\mathbf{Q}}_i^* &= \frac{\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{X}_i}{T} \\ \overline{\mathbf{q}}^* &= \frac{1}{N}\sum_{i=1}^N s_i\widehat{\mathbf{q}}_i^*, \qquad &\widehat{\mathbf{q}}_i^* &= \frac{\sqrt{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}[\boldsymbol{\varepsilon}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^\dagger\gamma_i]}{T} \\ \overline{\mathbf{q}}_v^* &= \frac{1}{N}\sum_{i=1}^N s_i\widehat{\mathbf{q}}_{v,i}^*, \qquad &\widehat{\mathbf{q}}_{v,i}^* &= \frac{\sqrt{N}\mathbf{X}_i\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{X}_i}{T}\boldsymbol{v}_i\end{aligned}$$

Making use of (3.16), $\mathbf{M}_{\widehat{\mathbf{F}}} = \mathbf{M}_{\widehat{\mathbf{F}}^0}$ and $\mathbf{M}_{\widehat{\mathbf{F}}^0} = \mathbf{M}_{\mathbf{F}^0} - [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^0}]$, let the following be the familiar decomposition at the individual level

$$\begin{aligned}\widehat{\mathbf{Q}}_i^* &= T^{-1}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{X}_i \\ &= T^{-1}[\mathbf{V}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^\dagger\boldsymbol{\Gamma}_i]'\mathbf{M}_{\mathbf{F}^0}[\mathbf{V}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^\dagger\boldsymbol{\Gamma}_i] - T^{-1}[\mathbf{V}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^\dagger\boldsymbol{\Gamma}_i]'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}][\mathbf{V}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^\dagger\boldsymbol{\Gamma}_i] \\ &= \widehat{\mathbf{Q}}_{\mathbf{M}_{\mathbf{F}^0},i}^* - \widehat{\mathbf{Q}}_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],i}^*\end{aligned} \tag{3.109}$$

where for a stated subscript $\mathbf{A}$, we obtain the breakdown

$$\widehat{\mathbf{Q}}^*_{\mathbf{A},i} = \widehat{\mathbf{Q}}^*_{\mathbf{A},VV,i} - \widehat{\mathbf{Q}}^*_{\mathbf{A},V\Gamma,i} - (\widehat{\mathbf{Q}}^*_{\mathbf{A},V\Gamma,i})' + \widehat{\mathbf{Q}}^*_{\mathbf{A},\Gamma\Gamma,i}$$
$$\widehat{\mathbf{Q}}^*_{\mathbf{A},VV,i} = T^{-1}\mathbf{V}_i'\mathbf{A}\mathbf{V}_i$$
$$\widehat{\mathbf{Q}}^*_{\mathbf{A},V\Gamma,i} = T^{-1}\mathbf{V}_i'\mathbf{A}\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\mathbf{\Gamma}_i$$
$$\widehat{\mathbf{Q}}^*_{\mathbf{A},\Gamma\Gamma,i} = T^{-1}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'\mathbf{A}\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\mathbf{\Gamma}_i$$

and where barred variants with an omitted i subscript denote averages over i as $\overline{\mathbf{Q}}^*_{\mathbf{A},VV} = \frac{1}{N}\sum_{i=1}^N \widehat{\mathbf{Q}}^*_{\mathbf{A},VV,i}$. Next, for the individual-specific numerators

$$\widehat{\mathbf{q}}_i^* = \sqrt{N}T^{-1}[\mathbf{V}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\mathbf{\Gamma}_i]'\mathbf{M}_{\widehat{\mathbf{F}}^*}[\boldsymbol{\varepsilon}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\gamma_i] = \widehat{\mathbf{q}}^*_{\mathbf{I},i} - \widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},i} - \widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],i} \tag{3.110}$$

with for a given subscript $\mathbf{A}$ the decomposition

$$\widehat{\mathbf{q}}^*_{\mathbf{A},i} = \widehat{\mathbf{q}}^*_{\mathbf{A},V\varepsilon,i} - \widehat{\mathbf{q}}^*_{\mathbf{A},V\gamma,i} - \widehat{\mathbf{q}}^*_{\mathbf{A},\Gamma\varepsilon,i} + \widehat{\mathbf{q}}^*_{\mathbf{A},\Gamma\gamma,i}$$
$$\widehat{\mathbf{q}}^*_{\mathbf{A},V\varepsilon,i} = \sqrt{N}T^{-1}\mathbf{V}_i'\mathbf{A}\boldsymbol{\varepsilon}_i$$
$$\widehat{\mathbf{q}}^*_{\mathbf{A},V\gamma,i} = \sqrt{N}T^{-1}\mathbf{V}_i'\mathbf{A}\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\gamma_i$$
$$\widehat{\mathbf{q}}^*_{\mathbf{A},\Gamma\varepsilon,i} = \sqrt{N}T^{-1}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'\mathbf{A}\boldsymbol{\varepsilon}_i$$
$$\widehat{\mathbf{q}}^*_{\mathbf{A},\Gamma\gamma,i} = \sqrt{N}T^{-1}\mathbf{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'\mathbf{A}\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\gamma_i$$

where barred terms will similarly be defined as $\overline{\mathbf{q}}^*_{\mathbf{A},V\varepsilon} = \frac{1}{N}\sum_{i=1}^N \widehat{\mathbf{q}}^*_{\mathbf{A},V\varepsilon,i}$. Finally, $\widehat{\mathbf{q}}_v^*$ features only in (3.108) so we can directly define the averaged term

$$\overline{\mathbf{q}}_{\boldsymbol{v}}^* = \frac{1}{\sqrt{N}}\sum_{i=1}^N s_i\frac{\mathbf{X}_i\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{X}_i}{T}\boldsymbol{v}_i = \frac{1}{\sqrt{N}T}\sum_{i=1}^N s_i[\mathbf{V}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\mathbf{\Gamma}_i]'\mathbf{M}_{\widehat{\mathbf{F}}^*}[\mathbf{V}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\mathbf{\Gamma}_i]\boldsymbol{v}_i$$
$$= \overline{\mathbf{q}}^*_{\mathbf{I},\boldsymbol{v}} - \overline{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},\boldsymbol{v}} - \overline{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\boldsymbol{v}} \tag{3.111}$$

with, given a matrix $\mathbf{A}$,

$$\overline{\mathbf{q}}^*_{\mathbf{A},\boldsymbol{v}} = \overline{\mathbf{q}}^*_{\mathbf{A},VV,\boldsymbol{v}} - \overline{\mathbf{q}}^*_{\mathbf{A},V\Gamma,\boldsymbol{v}} - (\overline{\mathbf{q}}^*_{\mathbf{A},V\Gamma,\boldsymbol{v}})' + \overline{\mathbf{q}}^*_{\mathbf{A},\Gamma\Gamma,\boldsymbol{v}}$$
$$\overline{\mathbf{q}}^*_{\mathbf{A},VV,\boldsymbol{v}} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N s_i\mathbf{V}_i'\mathbf{A}\mathbf{V}_i\boldsymbol{v}_i$$
$$\overline{\mathbf{q}}^*_{\mathbf{A},V\Gamma,\boldsymbol{v}} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N s_i\mathbf{V}_i'\mathbf{A}\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\mathbf{\Gamma}_i\boldsymbol{v}_i$$
$$\overline{\mathbf{q}}_{\mathbf{A},\Gamma\Gamma,\boldsymbol{v}} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N s_i\mathbf{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'\mathbf{A}\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\mathbf{\Gamma}_i\boldsymbol{v}_i$$

3.4.1 CCEP with the Pairs bootstrap

Theorem 5 *Under Ass. 1-6, with in addition* $\mathbb{E}(\|\mathbf{v}_{it}\|^8) < \infty$ *and* $\mathbb{E}(\|\boldsymbol{v}_i\|^6) < \infty$, *we have as* $(N,T) \to \infty$

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \xrightarrow{d^*} \mathcal{N}(\mathbf{0}_{k\times 1}, \mathbf{\Sigma}^{-1}\mathbf{\Psi}_h\mathbf{\Sigma}^{-1})$$

with $\boldsymbol{\Psi}_h$ defined in Theorem 4, and under the same conditions

$$\sup_{x\in\mathbb{R}^{k\times 1}}\left|\mathbb{P}^*[\sqrt{N}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \le x] - \mathbb{P}[\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \le x]\right| \longrightarrow^p 0,$$

where inequalities are to be interpreted coordinate-wise.

Proof of Theorem 5

Recall the decomposition of the scaled deviation of the CCEP estimator in (3.108)

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}) = \overline{\mathbf{Q}}^{*-1}(\overline{\mathbf{q}}_{\boldsymbol{v}}^* + \overline{\mathbf{q}}^*)$$

Note that here by definition $\overline{\mathbf{Q}}^* = \widehat{\mathbf{Q}}^*$ so that the decomposition of $\overline{\mathbf{Q}}^*$ is the same as that for $\widehat{\mathbf{Q}}^*$ analyzed in Lemma C-7, save that $\overline{\mathbf{U}}_w, \overline{\mathbf{C}}_w$ contain the heterogeneous slopes. Given that Lemmas C-2 and C-3 have been derived allowing for the slope heterogeneity characterized by Ass.6, the asymptotic orders derived in Lemma C-7 apply directly to the heterogeneous slope setting and we have from the exact same arguments as in that proof

$$\overline{\mathbf{Q}}^{*-1} \longrightarrow^p \boldsymbol{\Sigma}^{-1} \tag{3.112}$$

Similarly, since heterogeneity does not impact the orders derived in Lemmas C-2 and C-3 (only the limit statements are affected, as noted in the lemma) and we have by definition $\overline{\mathbf{q}}_{\mathbf{I}}^* = \frac{1}{\sqrt{T}}\widehat{\mathbf{q}}_{\mathbf{I}}^*, \overline{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}}^* = \frac{1}{\sqrt{T}}\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}}^*, \overline{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]}^* = \frac{1}{\sqrt{T}}\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]}^*$ (so that we have scaled up by $\sqrt{N}$ rather than $\sqrt{NT}$), the results from Lemmas C-4, C-6, C-5 that $\|\widehat{\mathbf{q}}_{\mathbf{I}}^*\| = O_{p^*}(1), \left\|\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}}^*\right\| = O_{p^*}(T^{-1/2}), \left\|\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]}^*\right\| = O_{p^*}(1)$, imply that $\|\overline{\mathbf{q}}_{\mathbf{I}}^*\| = o_{p^*}(1), \left\|\overline{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0}}^*\right\| = o_{p^*}(1), \left\|\overline{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]}^*\right\| = o_{p^*}(1)$. Hence, $\|\overline{\mathbf{q}}^*\| = o_{p^*}(1)$ and we have

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}) = \overline{\mathbf{Q}}^{*-1}\overline{\mathbf{q}}_{\boldsymbol{v}}^* + o_{p^*}(1) \tag{3.113}$$

Consider then the decomposition of $\overline{\mathbf{q}}_{\boldsymbol{v}}^*$ defined in (3.111). We start with terms containing the deviations $\mathbf{A} = [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]$. First up is,

$$\begin{aligned}
\overline{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\Gamma,\boldsymbol{v}}^* &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N s_i \boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^\dagger)'\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^\dagger\boldsymbol{\Gamma}_i\boldsymbol{v}_i \\
&= \left[\frac{1}{N}\sum_{i=1}^N s_i(\boldsymbol{v}_i'\boldsymbol{\Gamma}_i'\otimes\boldsymbol{\Gamma}_i')\right] vec\left((\overline{\mathbf{C}}_w^\dagger)'\sqrt{N}T^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^\dagger\right) \\
&\longrightarrow^{p^*} \mathbf{0}_{k\times 1}
\end{aligned}$$

because $\left\|\overline{\mathbf{C}}_w^\dagger\right\| = O_{p^*}(1)$ and inserting (3.20) in $\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w$ gives

$$\begin{aligned}
T^{-1}\overline{\mathbf{U}}_w'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w &= T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,-m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{w,-m}^0}^\dagger T^{-1}(\overline{\mathbf{U}}_{w,-m}^0)'\overline{\mathbf{U}}_w \\
&+ T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^\dagger T^{-1}(\overline{\mathbf{U}}_{w,m}^0)'\overline{\mathbf{U}}_w \\
&+ T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^\dagger T^{-1}(\overline{\mathbf{U}}_{w,m}^0)'\overline{\mathbf{U}}_w + T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^\dagger T^{-1}\mathbf{F}'\overline{\mathbf{U}}_w \\
&+ T^{-1}\overline{\mathbf{U}}_w'\widehat{\mathbf{F}}^0\left[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}^\dagger - \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}}^\dagger\right]T^{-1}(\widehat{\mathbf{F}}^0)'\overline{\mathbf{U}}_w \\
&= O_{p^*}(N^{-1})
\end{aligned} \tag{3.114}$$

which follows because by Lemmas C-2 and C-3

$$\left\| T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,-m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{w,-m}^0}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{w,-m}^0)'\overline{\mathbf{U}}_w\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,-m}^0\right\|^2\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{w,-m}^0}^{\dagger}\right\| = O_{p^*}(N^{-1})$$
$$\left\| T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,m}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{w,m}^0)'\overline{\mathbf{U}}_w\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_{w,m}^0\right\|^2\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}\right\| = O_{p^*}(N^{-2})$$
$$\left\| T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}T^{-1}(\overline{\mathbf{U}}_{w,m}^0)'\overline{\mathbf{U}}_w\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}\right\|\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}}^{\dagger}\right\|\left\|T^{-1}(\overline{\mathbf{U}}_{w,m}^0)'\overline{\mathbf{U}}_w\right\| = O_{p^*}(T^{-1/2}N^{-3/2})$$
$$\left\| T^{-1}\overline{\mathbf{U}}_w'\widehat{\mathbf{F}}^0\left[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}^{\dagger}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}}^{\dagger}\right]T^{-1}(\widehat{\mathbf{F}}^0)'\overline{\mathbf{U}}_w\right\| \leq \left\|T^{-1}\overline{\mathbf{U}}_w'\widehat{\mathbf{F}}^{0*}\right\|^2\left\|\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}^{\dagger}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}}^{\dagger}\right\| = O_{p^*}(N^{-3/2})+O_{p^*}(N^{-1}T^{-1/2})$$

and also, the independence of s_i of the other variables, Ass.3 and 6 (the mean zero and independence of $\boldsymbol{v}_i$) implies

$$\left\|\frac{1}{N}\sum_{i=1}^N s_i(\boldsymbol{v}_i'\boldsymbol{\Gamma}_i'\otimes\boldsymbol{\Gamma}_i')\right\| = O_{p^*}(N^{-1/2})$$

Next, the exact same arguments as for (3.39) and (3.40) in the proof of Lemma C-4 can be applied to obtain from $\mathbf{V}_i=\mathbf{U}_i\mathbf{q}_x$

$$\left\|\sqrt{N}T^{-1}\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\right\| = O_{p^*}(N^{-1/2})+O_{p^*}(T^{-1/2})$$

so that

$$\begin{aligned}\left\|\overline{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\Gamma,\boldsymbol{v}}\right\| &= \left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^N s_i\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\boldsymbol{\Gamma}_i\boldsymbol{v}_i\right\| \\ &\leq \frac{1}{N}\sum_{i=1}^N\|s_i\|\left\|\sqrt{N}T^{-1}\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\right\|\left\|\overline{\mathbf{C}}_w^{\dagger}\right\|\|\boldsymbol{\Gamma}_i\|\|\boldsymbol{v}_i\| \\ &= O_{p^*}(N^{-1/2})+O_{p^*}(T^{-1/2})\end{aligned}$$

For the final term of this kind

$$\begin{aligned}\overline{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],VV,\boldsymbol{v}} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N s_i\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{V}_i\boldsymbol{v}_i = \frac{1}{N}\sum_{i=1}^N s_i\sqrt{N}T^{-1}\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{V}_i\boldsymbol{v}_i \\ &\longrightarrow^{p^*}\mathbf{0}_{k\times 1}\end{aligned}$$

To obtain this result, note that by substituting (3.20) into $\overline{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],VV,\boldsymbol{v}}$ and following the same arguments as for (3.44) gives

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^N\sqrt{N}T^{-1}s_i\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{V}_i\boldsymbol{v}_i \\ &= \frac{1}{N}\sum_{i=1}^N\sqrt{N}s_i\left(\frac{\mathbf{V}_i'\overline{\mathbf{U}}_{w,-m}^0}{T}\right)\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{w,-m}^0}^{\dagger}\left(\frac{\overline{\mathbf{U}}_{w,-m}^{0\prime}\mathbf{V}_i}{T}\right)\boldsymbol{v}_i+\frac{1}{N}\sum_{i=1}^N\sqrt{N}s_i\left(\frac{\mathbf{V}_i'\widehat{\mathbf{F}}^{0*}}{T}\right)[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}^{\dagger}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}}^{\dagger}]\left(\frac{(\widehat{\mathbf{F}}^{0*})'\mathbf{V}_i}{T}\right)\boldsymbol{v}_i \\ &\quad+O_{p^*}\left(\frac{1}{N^{3/2}}\right)+O_{p^*}\left(\frac{1}{T}\right)+O_{p^*}\left(\frac{1}{\sqrt{NT}}\right)\end{aligned}$$

where it should be noted that the expansion is sharpened by not approximating terms that are $O_p(\sqrt{N}T^{-\lambda})$ for $\lambda > 0$. Making use of $\overline{\mathbf{U}}^0_{w,-m} = \sqrt{N}\overline{\mathbf{U}}_w \mathbf{T}\overline{\mathbf{H}}_{w,-m}$, and defining $\widehat{\mathbf{D}}_w = \mathbf{T}\overline{\mathbf{H}}_{w,-m}\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}}\overline{\mathbf{H}}'_{w,-m}\mathbf{T}' = O_{p^*}(1)$ the first term in the expansion can be rewritten as

$$\frac{1}{N}\sum_{i=1}^{N}\sqrt{N}s_i\left(\frac{\mathbf{V}_i'\overline{\mathbf{U}}^0_{w,-m}}{T}\right)\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}}\left(\frac{\overline{\mathbf{U}}^{0\prime}_{w,-m}\mathbf{V}_i}{T}\right)\boldsymbol{v}_i = \frac{1}{N^{3/2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}s_is_js_k\left(\frac{\mathbf{V}_i'\mathbf{U}_j}{T}\right)\widehat{\mathbf{D}}_w\left(\frac{\mathbf{U}_k'\mathbf{V}_i}{T}\right)\boldsymbol{v}_i$$

Letting next $\widehat{d}^w_{v,g}$ denote the element on row v and column g of $\widehat{\mathbf{D}}_w$, and with $\mathbf{U}_i^{(l)}$ denoting column l of $\mathbf{U}_i$, we obtain

$$\begin{aligned}&\left\|\frac{1}{N}\sum_{i=1}^{N}\sqrt{N}\left(\frac{\mathbf{V}_i'\overline{\mathbf{U}}^0_{w,-m}}{T}\right)\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{u}^0_{w,-m}}\left(\frac{\overline{\mathbf{U}}^{0\prime}_{w,-m}\mathbf{V}_i}{T}\right)\boldsymbol{v}_i\right\| \\ &\qquad \le \sum_{v=1}^{1+k}\sum_{g=1}^{1+k}|\widehat{d}^w_{v,g}|\left\|\frac{1}{N^{3/2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}s_is_js_k\left(\frac{\mathbf{V}_i'\mathbf{U}_j^{(v)}}{T}\right)\left(\frac{\mathbf{U}_k^{(g)\prime}\mathbf{V}_i}{T}\right)\boldsymbol{v}_i\right\|\end{aligned}$$

where, given fixed and finite k, $|\widehat{d}^w_{v,g}| = O_{p^*}(1)$ and $\mathbf{U}_i = [\boldsymbol{\varepsilon}_i + \mathbf{V}_i(\boldsymbol{\beta}+\boldsymbol{v}_i), \mathbf{V}_i]$, further unpacking reveals that the term with the highest degree of dependence, and hence the driver of the asymptotic order, is

$$\left\|\frac{1}{N^{3/2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}s_is_js_k\left(\frac{\mathbf{V}_i'\mathbf{V}_j}{T}\right)\boldsymbol{v}_j\boldsymbol{v}_k'\left(\frac{\mathbf{V}_k'\mathbf{V}_i}{T}\right)\boldsymbol{v}_i\right\|$$

when $v = g = 1$. Note that the expectation of this term is zero unless $i = j = k$ by cross-section independence, and in the case with equal indices we obtain given finite moments that

$$\mathbf{A}_i = \mathbb{E}^*\left[\frac{1}{N^{3/2}}\sum_{i=1}^{N}s_i^3\left(\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\right)\boldsymbol{v}_i\boldsymbol{v}_i'\left(\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\right)\boldsymbol{v}_i\right] = O\left(\frac{1}{\sqrt{N}}\right)$$

Also, by the cross-section independence, and independence of $\mathbf{V}_i$ and $\boldsymbol{v}_j$ for all i, j

$$\begin{aligned}&\frac{1}{N^3T^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}\mathbb{E}^*(s_is_js_ks_ls_ms_n) \\ &\qquad\times\mathbb{E}^*\left\{\left(\mathbf{V}_i'\mathbf{V}_j\boldsymbol{v}_j\boldsymbol{v}_k'\mathbf{V}_k'\mathbf{V}_i\boldsymbol{v}_i - \mathbf{1}_{(i=j=k)}\mathbf{A}_i\right)\left(\mathbf{V}_l'\mathbf{V}_m\boldsymbol{v}_m\boldsymbol{v}_n'\mathbf{V}_n'\mathbf{V}_l\boldsymbol{v}_l - \mathbf{1}_{(l=m=n)}\mathbf{A}_l\right)'\right\} \\ &= O\left(\frac{1}{T^2}\right) + O\left(\frac{1}{N}\right)\end{aligned}$$

since, as argued for the analysis in the original dataset, the cross-section independence of $\boldsymbol{v}_i$ implies that the expectation is zero for each part of the sums for which a single cross-section index differs from the others. This means that the expectation is zero when more than 3 distinct indices appear. That is, in (i, j, k, l, m, n) at least 3 *pairs* of indices need to be equal, and therefore the nonzero part of this sum of expectations runs over at most 3 distinct summation operands. In the case of three operands, the expectations exists either of sums over 4 $\mathbf{V}_i$ with the same index and 2 pairs of 2 $\mathbf{V}$'s with a common index, or summations over 2 sets of 3 $\mathbf{V}$'s with a common index and one pair of 2 with a common index. In each case, unpacking terms over time reveals that the corresponding sum of expectations is of order $O(T^{-2})$ due to the stationarity of the $\mathbf{v}_{it}$ and its finite moments up to the fourth order. For the segments of the summation with two or less operands, we have given that $\mathbb{E}(\|\mathbf{v}_{it}\|^8) < \infty$ and $\mathbb{E}(\|\boldsymbol{v}_i\|^6) < \infty$ that they are *at most* of order $O(N^{-1})$. Consequently, $\left\| \frac{1}{N^{3/2}T^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N s_i s_j s_k \mathbf{V}_i' \mathbf{V}_j \boldsymbol{v}_j \boldsymbol{v}_k' \mathbf{V}_k' \mathbf{V}_i \boldsymbol{v}_i \right\| = o_{p^*}(1)$ as $(N,T) \to \infty$. Since this is also the leading term in the inequality above, we conclude that

$$\frac{1}{N} \sum_{i=1}^N \sqrt{N} s_i \left(\frac{\mathbf{V}_i' \overline{\mathbf{U}}_{w,-m}^0}{T} \right) \widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{w,-m}^0}^{\dagger} \left(\frac{\overline{\mathbf{U}}_{w,-m}^{0\prime} \mathbf{V}_i}{T} \right) \boldsymbol{v}_i \longrightarrow^{p^*} \mathbf{0}_{k\times 1} \tag{3.115}$$

Next, substituting $\widehat{\mathbf{F}}^{0*} = \mathbf{F}^0 + [\overline{\mathbf{U}}_{w,m}^0, \overline{\mathbf{U}}_{w,-m}^0]$ into the second term of the expansion, we obtain for the two leading terms with the slowest decay

$$\frac{1}{N} \sum_{i=1}^N \sqrt{N} s_i \left(\frac{\mathbf{V}_i' \overline{\mathbf{U}}_{w,-m}^0}{T} \right) \left[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}^{\dagger} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}}^{\dagger} \right] \left(\frac{(\overline{\mathbf{U}}_{w,-m}^0)' \mathbf{V}_i}{T} \right) \boldsymbol{v}_i \longrightarrow^{p^*} \mathbf{0}_{k\times 1}$$

$$\frac{1}{N} \sum_{i=1}^N \sqrt{N} s_i \left(\frac{\mathbf{V}_i' \mathbf{F}^0}{T} \right) \left[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}^{\dagger} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}}^{\dagger} \right] \left(\frac{(\mathbf{F}^0)' \mathbf{V}_i}{T} \right) \boldsymbol{v}_i \longrightarrow^{p^*} \mathbf{0}_{k\times 1}$$

where the first result follows from the same arguments as (3.115) (but noting that the rate is faster since $\left\| \widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}^{\dagger} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}}^{\dagger} \right\| = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$ from lemma C-3, while for the second we used $\|T^{-1}\mathbf{V}_i'\mathbf{F}^0\| = O_p(T^{-1/2})$ and the fact that $\boldsymbol{v}_i$ is independent of the other terms with $\frac{1}{\sqrt{N}} \sum_{i=1}^N \boldsymbol{v}_i = O_p(1)$. That is, in the bootstrap world,

$$
\begin{aligned}
&\mathbb{E}^*\left(\left\|\frac{1}{\sqrt{N}T^2}\sum_{i=1}^{N}s_i\boldsymbol{v}_i'\mathbf{V}_i'\mathbf{F}^0\otimes\mathbf{V}_i'\mathbf{F}^0\right\|^2\right)\\
&=\mathbb{E}^*\left(\operatorname{tr}\left[\left(\frac{1}{\sqrt{N}T^2}\sum_{j=1}^{N}s_j\boldsymbol{v}_j'\mathbf{V}_j'\mathbf{F}^0\otimes\mathbf{V}_j'\mathbf{F}^0\right)'\left(\frac{1}{\sqrt{N}T^2}\sum_{i=1}^{N}s_i\boldsymbol{v}_i'\mathbf{V}_i'\mathbf{F}^0\otimes\mathbf{V}_i'\mathbf{F}^0\right)\right]\right)\\
&=\mathbb{E}^*\left(\operatorname{tr}\left[\left(\frac{1}{\sqrt{N}T^2}\sum_{i=1}^{N}s_i\boldsymbol{v}_i'\mathbf{V}_i'\mathbf{F}^0\otimes\mathbf{V}_i'\mathbf{F}^0\right)\left(\frac{1}{\sqrt{N}T^2}\sum_{j=1}^{N}s_j\boldsymbol{v}_j'\mathbf{V}_j'\mathbf{F}^0\otimes\mathbf{V}_j'\mathbf{F}^0\right)'\right]\right)\\
&=\frac{1}{NT^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\operatorname{tr}\left(\mathbb{E}^*\left[\left(s_is_j\boldsymbol{v}_i'\mathbf{V}_i'\mathbf{F}^0\otimes\mathbf{V}_i'\mathbf{F}^0\right)\left(\mathbf{F}^{0\prime}\mathbf{V}_j\boldsymbol{v}_j\otimes\mathbf{F}^{0\prime}\mathbf{V}_j\right)\right]\right)\\
&=\frac{1}{NT^4}\sum_{i=1}^{N}\operatorname{tr}\left(\mathbb{E}^*\left[\left(s_i^2\boldsymbol{v}_i'\mathbf{V}_i'\mathbf{F}^0\otimes\mathbf{V}_i'\mathbf{F}^0\right)\left(\mathbf{F}^{0\prime}\mathbf{V}_i\boldsymbol{v}_i\otimes\mathbf{F}^{0\prime}\mathbf{V}_i\right)\right]\right)\\
&=\frac{1}{NT^4}\sum_{i=1}^{N}\operatorname{tr}\left(\mathbb{E}^*\left[s_i^2\boldsymbol{v}_i'\mathbf{V}_i'\mathbf{F}^0\mathbf{F}^{0\prime}\mathbf{V}_i\boldsymbol{v}_i\otimes\mathbf{V}_i'\mathbf{F}^0\mathbf{F}^{0\prime}\mathbf{V}_i\right]\right)\\
&=\frac{1}{NT^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{p=1}^{T}\sum_{r=1}^{T}\operatorname{tr}\left(\mathbb{E}^*\left[(\mathbf{f}_t^0)'\mathbf{f}_s^0s_i^2\boldsymbol{v}_i'\mathbf{v}_{i,t}\mathbf{v}_{i,s}'\boldsymbol{v}_i\otimes\mathbf{v}_{i,p}\mathbf{v}_{i,r}'(\mathbf{f}_p^0)'\mathbf{f}_r^0\right]\right)\\
&=\operatorname{tr}\left(\frac{1}{NT^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{p=1}^{T}\sum_{r=1}^{T}\mathbb{E}^*\left[(\mathbf{f}_t^0)'\mathbf{f}_s^0(\mathbf{f}_p^0)'\mathbf{f}_r^0\right]\mathbb{E}^*[s_i^2]\mathbb{E}^*\left[\mathbf{v}_{i,p}\mathbf{v}_{i,r}'\boldsymbol{v}_i'\mathbf{v}_{i,t}\mathbf{v}_{i,s}'\boldsymbol{v}_i\right]\right)\\
&=2\operatorname{tr}\left(\frac{1}{NT^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{p=1}^{T}\sum_{r=1}^{T}\mathbb{E}^*\left[(\mathbf{f}_t^0)'\mathbf{f}_s^0(\mathbf{f}_p^0)'\mathbf{f}_r^0\right]\mathbb{E}^*\left[\mathbf{v}_{i,p}\mathbf{v}_{i,r}'\operatorname{tr}(\mathbf{v}_{i,t}\mathbf{v}_{i,s}'\boldsymbol{v}_i\boldsymbol{v}_i')\right]\right)\\
&-\operatorname{tr}\left(\frac{1}{N^2T^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{p=1}^{T}\sum_{r=1}^{T}\mathbb{E}^*\left[(\mathbf{f}_t^0)'\mathbf{f}_s^0(\mathbf{f}_p^0)'\mathbf{f}_r^0\right]\mathbb{E}^*\left[\mathbf{v}_{i,p}\mathbf{v}_{i,r}'\operatorname{tr}(\mathbf{v}_{i,t}\mathbf{v}_{i,s}'\boldsymbol{v}_i\boldsymbol{v}_i')\right]\right)\\
&=O(T^{-1})+O((NT)^{-1})=O(T^{-1}).
\end{aligned}
$$

In conclusion,

$$\overline{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],VV,\boldsymbol{v}}\longrightarrow^{p^*}\mathbf{0}_{k\times 1}$$

so that by combining results we come to

$$\overline{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\boldsymbol{v}}\longrightarrow^{p^*}\mathbf{0}_{k\times 1}$$

Next consider terms where $\mathbf{A}=\mathbf{P}_{\mathbf{F}^0}$. First, using $\left\|T^{-1}\mathbf{F}'\mathbf{V}_i\right\|=O_p(T^{-1/2})$, the independence of s_i from the other variables and its finite moments, and similarly the independence of $\boldsymbol{v}_i$ over i and from the other variables, with $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}s_i\boldsymbol{v}_i=O_{p^*}(1)$ gives

$$\left\|\overline{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},VV,\boldsymbol{v}}\right\|=\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}s_i(T^{-1}\mathbf{V}_i'\mathbf{F}^0)\widehat{\boldsymbol{\Sigma}}^{\dagger}_{\mathbf{F}^0}(T^{-1}\mathbf{F}^{0\prime}\mathbf{V}_i)\boldsymbol{v}_i\right\|=O_{p^*}(T^{-1})$$

also, with $\left\|T^{-1}\mathbf{F}'\overline{\mathbf{U}}_w\right\| = O_{p^*}((NT)^{-1/2})$

$$\left\|\overline{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},V\Gamma,\boldsymbol{v}}\right\| = \left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N s_i T^{-1}\mathbf{V}_i'\mathbf{F}^0\widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}_w\overline{\mathbf{C}}^\dagger_w\boldsymbol{\Gamma}_i\boldsymbol{v}_i\right\|$$
$$\leq \frac{1}{N}\sum_{i=1}^N \|s_i\| \left\|T^{-1}\mathbf{V}_i'\mathbf{F}^0\right\| \left\|\widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}^0}\right\| \left\|\sqrt{N}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}_w\right\| \left\|\overline{\mathbf{C}}^\dagger_w\right\| \|\boldsymbol{\Gamma}_i\| \|\boldsymbol{v}_i\| = O_{p^*}(T^{-1})$$

which could again be sharpened noting that s_i, $\boldsymbol{\Gamma}_i$ and $\boldsymbol{v}_i$ are independent of the other variables and $\frac{1}{N}\sum_{i=1}^N s_i\boldsymbol{\Gamma}_i\boldsymbol{v}_i = O_{p^*}(N^{-1/2})$. Finally, also

$$\left\|\overline{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},\Gamma\Gamma,\boldsymbol{v}}\right\| = \left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N s_i\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}^\dagger_w)'T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0\widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}^0}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}_w\overline{\mathbf{C}}^\dagger_w\boldsymbol{\Gamma}_i\boldsymbol{v}_i\right\|$$
$$\leq \sqrt{N}\left\|\frac{1}{N}\sum_{i=1}^N s_i(\boldsymbol{v}_i'\boldsymbol{\Gamma}_i'\otimes\boldsymbol{\Gamma}_i')\right\| \left\|\overline{\mathbf{C}}^\dagger_w\right\|^2 \left\|T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}_w\right\|^2 \left\|\widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}^0}\right\| = O_{p^*}((NT)^{-1})$$

which again makes use of $\left\|\frac{1}{N}\sum_{i=1}^N(\boldsymbol{v}_i'\boldsymbol{\Gamma}_i'\otimes\boldsymbol{\Gamma}_i')\right\| = O_{p^*}(N^{-1/2})$. Therefore,

$$\overline{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},\boldsymbol{v}} \longrightarrow^{p^*} \mathbf{0}_{k\times 1}$$

This establishes that both $\overline{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\boldsymbol{v}}$ and $\overline{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},\boldsymbol{v}}$ are asymptotically negligible. What remains is the terms with $\mathbf{A} = \mathbf{I}$, specifically $\overline{\mathbf{q}}^*_{\mathbf{I},\boldsymbol{v}}$. In its decomposition, given that $\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{U}_i\right\| = O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2})$, $\mathbf{V}_i \subset \mathbf{U}_i$ and $\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\right\| = O_{p^*}(N^{-1})$

$$\|\overline{\mathbf{q}}^*_{\mathbf{I},V\Gamma,\boldsymbol{v}}\| \leq \frac{1}{N}\sum_{i=1}^N \|s_i\| \left\|\sqrt{N}T^{-1}\mathbf{V}_i'\overline{\mathbf{U}}_w\right\| \left\|\overline{\mathbf{C}}^\dagger_w\right\| \|\boldsymbol{\Gamma}_i\| \|\boldsymbol{v}_i\| = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$$
$$\|\overline{\mathbf{q}}^*_{\mathbf{I},\Gamma\Gamma,\boldsymbol{v}}\| \leq \sqrt{N}\left\|\frac{1}{N}\sum_{i=1}^N s_i(\boldsymbol{v}_i'\boldsymbol{\Gamma}_i'\otimes\boldsymbol{\Gamma}_i')\right\| \left\|\overline{\mathbf{C}}^\dagger_w\right\|^2 \left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\right\| = O_{p^*}(N^{-1})$$

we have $\overline{\mathbf{q}}^*_{\mathbf{I},\boldsymbol{v}} = \overline{\mathbf{q}}^*_{\mathbf{I},VV,\boldsymbol{v}} - \overline{\mathbf{q}}^*_{\mathbf{I},V\Gamma,\boldsymbol{v}} - (\overline{\mathbf{q}}^*_{\mathbf{I},V\Gamma,\boldsymbol{v}})' + \overline{\mathbf{q}}^*_{\mathbf{I},\Gamma\Gamma,\boldsymbol{v}} = \overline{\mathbf{q}}^*_{\mathbf{I},VV,\boldsymbol{v}} + o_{p^*}(1)$ as $(N,T)\to\infty$, and by combining results into (3.113) we come to

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}) = \overline{\mathbf{Q}}^{*-1}\overline{\mathbf{q}}^*_{\mathbf{I},VV,\boldsymbol{v}} + o_{p^*}(1)$$

Subtracting then (2.88) from the proof of Lemma 4 from both sides gives the expansion around $\widehat{\boldsymbol{\beta}}$

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) = \overline{\mathbf{Q}}^{*-1}\overline{\mathbf{q}}^*_{\mathbf{I},VV,\boldsymbol{v}} - \overline{\mathbf{Q}}^{-1}\overline{\mathbf{q}}_{\mathbf{I},VV,\boldsymbol{v}} + o_{p^*}(1) = \overline{\mathbf{Q}}^{*-1}\widetilde{\mathbf{q}}^*_{\mathbf{I},VV,\boldsymbol{v}} - [\overline{\mathbf{Q}}^{-1} - \overline{\mathbf{Q}}^{*-1}]\overline{\mathbf{q}}_{\mathbf{I},VV,\boldsymbol{v}} + o_{p^*}(1)$$

with $\widetilde{\mathbf{q}}^*_{\mathbf{I},VV,\boldsymbol{v}} = \overline{\mathbf{q}}^*_{\mathbf{I},VV,\boldsymbol{v}} - \overline{\mathbf{q}}_{\mathbf{I},VV,\boldsymbol{v}}$.
Given (3.112) and $\overline{\mathbf{Q}}^{-1} \longrightarrow^p \boldsymbol{\Sigma}^{-1}$, $\|\overline{\mathbf{q}}_{\mathbf{I},VV,\boldsymbol{v}}\| = O_p(1)$ from respectively (2.84) and (2.89) in the proof of Lemma 4, we have for the final term

$$[\overline{\mathbf{Q}}^{-1} - \overline{\mathbf{Q}}^{*-1}]\overline{\mathbf{q}}_{\mathbf{I},VV,\boldsymbol{v}} \longrightarrow^{p^*} \mathbf{0}_{k\times 1}$$

whereas we can write by combining definitions for the first term in the numerator of this expansion

$$\widetilde{\mathbf{q}}^*_{\mathbf{I},VV,\boldsymbol{v}} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(s_i-1)\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\boldsymbol{v}_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(s_i-1)\boldsymbol{\Sigma}_i\boldsymbol{v}_i + O_{p^*}(T^{-1/2})$$

by the mutual independence of $\mathbf{V}_i, s_i, \boldsymbol{v}_i$, the fact that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(s_i-1)\boldsymbol{v}_i = O_{p^*}(1)$ by Ass.6 and Lemma C-1, and $T^{-1}\mathbf{V}_i'\mathbf{V}_i = \boldsymbol{\Sigma}_i + O_p(T^{-1/2})$ from Ass.1. Continuing, we have

$$\mathbb{E}^*\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(s_i-1)\boldsymbol{\Sigma}_i\boldsymbol{v}_i\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{E}^*(s_i-1)\mathbb{E}^*(\boldsymbol{\Sigma}_i)\mathbb{E}^*(\boldsymbol{v}_i) = \mathbf{0}_{k\times 1}$$

and by independence over i

$$\mathbb{E}^*\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(s_i-1)\boldsymbol{\Sigma}_i\boldsymbol{v}_i\right)\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}(s_j-1)\boldsymbol{\Sigma}_j\boldsymbol{v}_j\right)'$$
$$= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}^*[(s_i-1)(s_j-1)]\mathbb{E}^*\left[\boldsymbol{\Sigma}_i\mathbb{E}^*(\boldsymbol{v}_i\boldsymbol{v}_j')\boldsymbol{\Sigma}_j\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}^*[(s_i-1)^2]\mathbb{E}^*\left[\boldsymbol{\Sigma}_i\boldsymbol{\Omega}_{\boldsymbol{v}}\boldsymbol{\Sigma}_i\right]$$

such that given $\mathbb{E}^*[(s_i-1)^2] = \mathbb{E}^*(s_i^2) - 2\mathbb{E}^*(s_i) + 1 = 1 + N^{-1}$ from a) of Lemma C-1, we get by application of a CLT that $\widetilde{\mathbf{q}}^*_{\mathbf{I},VV,\boldsymbol{v}} \xrightarrow{d^*} \mathcal{N}(\mathbf{0},\boldsymbol{\Psi}_h)$ as $(N,T)\to\infty$, with $\boldsymbol{\Psi}_h = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i\boldsymbol{\Omega}_{\boldsymbol{v}}\boldsymbol{\Sigma}_i$, which in turn, again making use of (3.112) and the results above, leads to

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \xrightarrow{d^*} \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}_h\boldsymbol{\Sigma}^{-1})$$

which is the result stated in the theorem. It then follows directly from this result and Theorem 4 that

$$\sup_{x\in\mathbb{R}^{k\times 1}}\left|\mathbb{P}^*[\sqrt{N}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \leq x] - \mathbb{P}[\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq x]\right| \longrightarrow^p 0,$$

where inequalities are to be interpreted coordinate-wise.

3.4.2 CCEMG with the Pairs bootstrap

Theorem 7 *Under Ass.1-6 we have as* $(N,T) \to \infty$

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^*_{mg} - \widehat{\boldsymbol{\beta}}_{mg}) \xrightarrow{d^*} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Omega}_v).$$

In addition, under the same conditions

$$\sup_{x\in\mathbb{R}^{k\times 1}} \left| \mathbb{P}^*[\sqrt{N}(\widehat{\boldsymbol{\beta}}^*_{mg} - \widehat{\boldsymbol{\beta}}_{mg}) \leq x] - \mathbb{P}[\sqrt{N}(\widehat{\boldsymbol{\beta}}_{mg} - \boldsymbol{\beta}) \leq x] \right| \longrightarrow^p 0,$$

where inequalities are to be interpreted coordinate-wise.

Proof of Theorem 7

Recall the decomposition of the scaled CCEMG deviation introduced in (3.107) and below. To analyze the denominators at the individual level, we use $\widehat{\mathbf{Q}}^*_i = \widehat{\mathbf{Q}}^*_{\mathbf{I},i} - \widehat{\mathbf{Q}}^*_{\mathbf{P}_{\mathbf{F}^0},i} - \widehat{\mathbf{Q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],i}$. This is the same decomposition used to derive the asymptotic representation in Lemma C-7. The fact that the summation over $i = 1, \ldots, N$ is absent and $\mathbb{E}^*(s_i) = 1$ does not change the order of the remainder, therefore we can directly apply the result from Lemma C-7, which leads to

$$\widehat{\mathbf{Q}}^*_i = T^{-1}\mathbf{V}'_i\mathbf{V}_i + O_{p^*}(N^{-1}) + O_{p^*}(T^{-1}) + O_{p^*}((NT)^{-1/2}). \tag{3.116}$$

Also, because $T^{-1}\mathbf{V}'_i\mathbf{V}_i = \boldsymbol{\Sigma}_i + O_{p^*}(T^{-1/2})$ and $rk(\widehat{\mathbf{Q}}^*_i) - rk(\boldsymbol{\Sigma}_i) \xrightarrow{a.s.} 0$, we know that

$$\widehat{\mathbf{Q}}^{*-1}_i = \boldsymbol{\Sigma}^{-1}_i + O_{p^*}(N^{-1}) + O_{p^*}(T^{-1/2}). \tag{3.117}$$

Then, for the numerator we start from $\widehat{\mathbf{q}}^*_i = \sqrt{N}T^{-1}[\mathbf{V}_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}^{\dagger}_w\boldsymbol{\Gamma}_i]'\mathbf{M}_{\widehat{\mathbf{F}}^*}[\varepsilon_i - \overline{\mathbf{U}}_w\overline{\mathbf{C}}^{\dagger}_w\gamma_i] = \widehat{\mathbf{q}}^*_{\mathbf{I},i} - \widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},i} - \widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],i}$, where for $\mathbf{A}$ representing $\mathbf{I}_T$, $\mathbf{P}_{\mathbf{F}^0}$ or $\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}$, we have the same decomposition $\widehat{\mathbf{q}}^*_{\mathbf{A},i} = \widehat{\mathbf{q}}^*_{\mathbf{A},V\varepsilon,i} - \widehat{\mathbf{q}}^*_{\mathbf{A},V\gamma,i} - \widehat{\mathbf{q}}^*_{\mathbf{A},\Gamma\varepsilon,i} + \widehat{\mathbf{q}}^*_{\mathbf{A},\Gamma\gamma,i}$. The order results in the bootstrap world are not altered due to Lemma C-2 and Lemma C-3. Yet, for completeness, letting $\mathbf{A} = [\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]$, we obtain

$$\begin{aligned}\left\|\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\gamma,i}\right\| &= \left\|\sqrt{N}T^{-1}\boldsymbol{\Gamma}'_i(\overline{\mathbf{C}}^{\dagger}_w)'\overline{\mathbf{U}}'_w[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\overline{\mathbf{C}}^{\dagger}_w\gamma_i\right\| \\ &\leq \sqrt{N}\left\|\boldsymbol{\Gamma}'_i(\overline{\mathbf{C}}^{\dagger}_w)'\right\|\left\|\overline{\mathbf{C}}^{\dagger}_w\gamma_i\right\|\left\|T^{-1}\overline{\mathbf{U}}'_w[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\right\| = O_{p^*}(N^{-1/2}),\end{aligned} \tag{3.118}$$

using the fact that $\left\|T^{-1}\overline{\mathbf{U}}'_w[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\right\| = O_{p^*}(N^{-1})$ from (3.114). Further, with $\boldsymbol{\varepsilon}_i = \mathbf{U}_i\mathbf{B}^{-1}_i\mathbf{q}_y$ and the result $\left\|T^{-1}\overline{\mathbf{U}}'_w[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i\right\| = O_{p^*}(N^{-1}) + O_{p^*}((NT)^{-1/2})$, which comes from the exact same arguments as for (3.39) and (3.40) in the proof of Lemma C-4, we arrive at

$$\begin{aligned}\left\|\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],\Gamma\varepsilon,i}\right\| &= \left\|\sqrt{N}T^{-1}\boldsymbol{\Gamma}'_i(\overline{\mathbf{C}}^{\dagger}_w)'\overline{\mathbf{U}}'_w[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\varepsilon_i\right\| \\ &\leq \sqrt{N}\left\|\boldsymbol{\Gamma}'_i(\overline{\mathbf{C}}^{\dagger}_w)'\right\|\left\|T^{-1}\overline{\mathbf{U}}'_w[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i\right\|\left\|\mathbf{B}^{-1}_i\mathbf{q}_y\right\| = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}),\end{aligned} \tag{3.119}$$

Moving on, with $\mathbf{V}_i = \mathbf{U}_i \mathbf{q}_x$, we immediately obtain

$$\begin{aligned}\left\|\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\gamma,i}\right\| &= \left\|\sqrt{N}T^{-1}\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\gamma_i\right\| \\ &\leq \sqrt{N}\left\|\overline{\mathbf{C}}_w^{\dagger}\gamma_i\right\|\|\mathbf{q}_x\|\left\|T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\overline{\mathbf{U}}_w\right\| = O_{p^*}(N^{-1/2})+O_{p^*}(T^{-1/2})\end{aligned} \tag{3.120}$$

using the same argument. To proceed, we let $\mathbf{A} = \mathbf{P}_{\mathbf{F}^0}$. This leads to

$$\begin{aligned}\left\|\widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},\Gamma\gamma,i}\right\| &= \left\|\sqrt{N}T^{-1}\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\gamma_i\right\| \leq \sqrt{N}\left\|\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\right\|\left\|\overline{\mathbf{C}}_w^{\dagger}\gamma_i\right\|\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}_w\right\| \\ &\leq \sqrt{N}\left\|\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\right\|\left\|\overline{\mathbf{C}}_w^{\dagger}\gamma_i\right\|\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0\right\|^2\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}\right\| = O_{p^*}(N^{-1/2}T^{-1}),\end{aligned} \tag{3.121}$$

which comes from the fact that $\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0\right\| = O_{p^*}((NT)^{-1/2})$ from Lemma C-2. Further on,

$$\begin{aligned}\left\|\widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},\Gamma\varepsilon,i}\right\| &= \left\|\sqrt{N}T^{-1}\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'\mathbf{P}_{\mathbf{F}^0}\boldsymbol{\varepsilon}_i\right\| \leq \sqrt{N}\left\|\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\right\|\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}T^{-1}\mathbf{F}^{0\prime}\boldsymbol{\varepsilon}_i\right\| \\ &\leq \sqrt{N}\left\|\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\right\|\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0\right\|\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}\right\|\left\|T^{-1}\mathbf{F}^{0\prime}\boldsymbol{\varepsilon}_i\right\| = O_{p^*}(T^{-1}),\end{aligned} \tag{3.122}$$

using the facts that $\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{F}^0\right\| = O_{p^*}((NT)^{-1/2})$ and $\left\|T^{-1}\mathbf{F}^{0\prime}\boldsymbol{\varepsilon}_i\right\| = O_p(T^{-1/2})$ from $\boldsymbol{\varepsilon}_i = \mathbf{U}_i\mathbf{B}_i^{-1}\mathbf{q}_y$ and $\left\|T^{-1}\mathbf{F}^{0\prime}\mathbf{U}_i\right\| = O_{p^*}(T^{-1/2})$ in Lemma B-1. Using the latter result again with $\mathbf{V}_i = \mathbf{U}_i\mathbf{q}_x$ gives $\left\|T^{-1}\mathbf{V}_i'\mathbf{F}^0\right\| = O_{p^*}(T^{-1/2})$, so that in the same fashion,

$$\begin{aligned}\left\|\widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},V\gamma,i}\right\| &= \left\|\sqrt{N}T^{-1}\mathbf{V}_i'\mathbf{P}_{\mathbf{F}^0}\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\gamma_i\right\| \leq \sqrt{N}\left\|\overline{\mathbf{C}}_w^{\dagger}\gamma_i\right\|\left\|T^{-1}\mathbf{V}_i'\mathbf{F}^0\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}_w\right\| \\ &\leq \sqrt{N}\left\|\overline{\mathbf{C}}_w^{\dagger}\gamma_i\right\|\left\|T^{-1}\mathbf{V}_i'\mathbf{F}^0\right\|\left\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{F}^0}^{\dagger}\right\|\left\|T^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}_w\right\| = O_{p^*}(T^{-1})\end{aligned} \tag{3.123}$$

Further, we let $\mathbf{A} = \mathbf{I}_T$. Firstly, this leads to

$$\left\|\widehat{\mathbf{q}}^*_{\mathbf{I},\Gamma\gamma,i}\right\| = \left\|\sqrt{N}T^{-1}\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'\overline{\mathbf{U}}\overline{\mathbf{C}}_w^{\dagger}\gamma_i\right\| \leq \sqrt{N}\left\|\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\right\|\left\|\overline{\mathbf{C}}_w^{\dagger}\gamma_i\right\|\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\right\| = O_{p^*}(N^{-1/2}),$$

because $\left\|T^{-1}\overline{\mathbf{U}}_w'\overline{\mathbf{U}}_w\right\| = O_{p^*}(N^{-1})$. Also,

$$\begin{aligned}\|\widehat{\mathbf{q}}^*_{\mathbf{I},\Gamma\varepsilon,i}\| &= \left\|\sqrt{N}T^{-1}\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\overline{\mathbf{U}}_w'\boldsymbol{\varepsilon}_i\right\| \leq \sqrt{N}\left\|\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\right\|\left\|T^{-1}\overline{\mathbf{U}}_w'\boldsymbol{\varepsilon}_i\right\| \\ &\leq \sqrt{N}\left\|\boldsymbol{\Gamma}_i'(\overline{\mathbf{C}}_w^{\dagger})'\right\|\left\|\mathbf{B}_i^{-1}\right\|\|\mathbf{q}_y\|\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{U}_i\right\| = O_{p^*}(N^{-1/2})+O_{p^*}(T^{-1/2}),\end{aligned}$$

because $\left\|T^{-1}\overline{\mathbf{U}}_w'\mathbf{U}_i\right\| = O_{p^*}(N^{-1})+O_{p^*}((NT)^{-1/2})$. Eventually, we obtain

$$\begin{aligned}\left\|\widehat{\mathbf{q}}^*_{\mathbf{I},V\gamma,i}\right\| &= \left\|\sqrt{N}T^{-1}\mathbf{V}_i'\overline{\mathbf{U}}_w\overline{\mathbf{C}}_w^{\dagger}\gamma_i\right\| \leq \sqrt{N}\left\|\overline{\mathbf{C}}_w^{\dagger}\gamma_i\right\|\left\|T^{-1}\mathbf{V}_i'\overline{\mathbf{U}}_w\right\| \leq \sqrt{N}\left\|\overline{\mathbf{C}}_w^{\dagger}\gamma_i\right\|\|\mathbf{q}_x\|\left\|T^{-1}\mathbf{U}_i'\overline{\mathbf{U}}_w\right\| \\ &= O_{p^*}(N^{-1/2})+O_{p^*}(T^{-1/2})\end{aligned} \tag{3.124}$$

using the same argument as for the term above. Summarizing the order results for the 3 different versions of $\mathbf{A}$, we obtain the same asymptotic representation as in the original sample space:

$$\widehat{\mathbf{q}}_i^* = \widehat{\mathbf{q}}^*_{\mathbf{I},V\varepsilon,i} - \widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon,i} + \widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon,i} + O_{p^*}(N^{-1/2})+O_{p^*}(T^{-1/2}) \tag{3.125}$$

which in combination with $\left\|\widehat{\mathbf{Q}}_i^{*-1}\right\| = O_{p^*}(1)$ by (3.117) yields

$$\frac{1}{N}\sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^{*-1}\widehat{\mathbf{q}}_i^* = \frac{1}{N}\sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^{*-1}\left[\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon,i}^* - \widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon,i}^* + \widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon,i}^*\right] + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}).$$

To proceed, consider the first term, $\frac{1}{N}\sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^{*-1}\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon,i}^* = \frac{1}{NT}\sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^{*-1}\sqrt{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_i$. Given that by (3.117) $\widehat{\mathbf{Q}}_i^{*-1}$ is bounded with a well behaved fixed limit as $(N,T)\to\infty$, the order of this term is driven by $\frac{1}{NT}\sum_{i=1}^{N}\sqrt{N}s_i\mathbf{V}_i'\boldsymbol{\varepsilon}_i$. Since $\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} s_i\mathbf{V}_i'\boldsymbol{\varepsilon}_i\right\| = O_{p^*}(1)$ by cross-section independence, and mutual independence of $\mathbf{V}_i, \boldsymbol{\varepsilon}_i$ and s_i, we have by insertion into the term above (and noting that the normalisation is $N^{-1/2}T^{-1}$)

$$\left\|\frac{1}{N}\sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^{*-1}\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon,i}^*\right\| = O_{p^*}(T^{-1/2})$$

Next, for $\frac{1}{N}\sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^{*-1}\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon,i}^*$, substituting in (3.20) and making use of the same arguments as for (3.44), but sharpening the approximation (by not expanding terms which are $O_{p^*}(\sqrt{N}T^{-\lambda})$ with $\lambda > 0$) gives

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^{*-1}\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon,i}^* &= \frac{1}{N}\sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^{*-1}\sqrt{N}T^{-1}\mathbf{V}_i'[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\boldsymbol{\varepsilon}_i \\
= \frac{1}{N}\sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^{*-1}\sqrt{N}&\left(\frac{\mathbf{V}_i'\overline{\mathbf{U}}_{w,-m}^0}{T}\right)\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{w,-m}^0}^{\dagger}\left(\frac{(\overline{\mathbf{U}}_{w,-m}^0)'\boldsymbol{\varepsilon}_i}{T}\right) \\
&+ \frac{1}{N}\sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^{*-1}\sqrt{N}\left(\frac{\mathbf{V}_i'\widehat{\mathbf{F}}^{0*}}{T}\right)[\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{F}}^{0*}}^{\dagger} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{F}_{w,u}}^{\dagger}]\left(\frac{(\widehat{\mathbf{F}}^{0*})'\boldsymbol{\varepsilon}_i}{T}\right) \\
&+ O_{p^*}(N^{-3/2}) + O_{p^*}(T^{-1}) + O_{p^*}((NT)^{-1/2})
\end{aligned}$$

For the first term of this expansion, $\overline{\mathbf{U}}_{w,-m}^0 = \sqrt{N}\overline{\mathbf{U}}_w\mathbf{T}\overline{\mathbf{H}}_{w,-m}$ reveals that

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^{*-1}\sqrt{N}&\left(\frac{\mathbf{V}_i'\overline{\mathbf{U}}_{w,-m}^0}{T}\right)\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{w,-m}^0}^{\dagger}\left(\frac{(\overline{\mathbf{U}}_{w,-m}^0)'\boldsymbol{\varepsilon}_i}{T}\right) \\
&= \frac{1}{N^{3/2}}\sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^{*-1}\sum_{j=1}^{N}\sum_{k=1}^{N} s_j s_k\left(\frac{\mathbf{V}_i'\mathbf{U}_j}{T}\right)\widehat{\mathbf{D}}_w\left(\frac{\mathbf{U}_k'\boldsymbol{\varepsilon}_i}{T}\right)
\end{aligned}$$

where $\widehat{\mathbf{D}}_w = \mathbf{T}\overline{\mathbf{H}}_{w,-m}\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{w,-m}^0}^{\dagger}\overline{\mathbf{H}}_{w,-m}'\mathbf{T}'$. Since $\left\|\widehat{\mathbf{D}}_w\right\| = O_{p^*}(1)$, $\left\|\widehat{\mathbf{Q}}_i^{*-1}\right\| = O_{p^*}(1)$ and both matrices have well behaved fixed limits as $(N,T)\to\infty$ (see e.g. (3.117) and Lemma C-4), the asymptotic order is driven by $\left\|\frac{1}{N^{3/2}T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} s_i s_j s_k \mathbf{V}_i'\mathbf{U}_j\widehat{\mathbf{D}}_w\mathbf{U}_k'\boldsymbol{\varepsilon}_i\right\| = O_{p^*}(N^{-1/2})$. The latter result can be seen from the fact that the term is identical to (3.47) save with normalization $N^{-3/2}T^{-2}$ in stead of $N^{-1}T^{-2}$. Hence, the exact same arguments as for the result in (3.50) can be employed to yield $\left\|N^{-1}T^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} s_i s_j s_k \mathbf{V}_i'\mathbf{U}_j\widehat{\mathbf{D}}_w\mathbf{U}_k'\boldsymbol{\varepsilon}_i\right\| = O_{p^*}(1)$. Therefore, as $(N,T)\to\infty$

$$\left\|\frac{1}{N}\sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^{*-1}\sqrt{N}\left(\frac{\mathbf{V}_i'\overline{\mathbf{U}}_{w,-m}^0}{T}\right)\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}_{w,-m}^0}^{\dagger}\left(\frac{(\overline{\mathbf{U}}_{w,-m}^0)'\boldsymbol{\varepsilon}_i}{T}\right)\right\| = O_{p^*}\left(\frac{1}{\sqrt{N}}\right) \tag{3.126}$$

For the second term in the expansion, the fact that $\widehat{\mathbf{F}}^{0*} = \mathbf{F}^0 + [\overline{\mathbf{U}}^0_{w,m}, \overline{\mathbf{U}}^0_{w,-m}]$ reveals that its asymptotic behavior is determined by two leading terms. For the first we obtain from the same arguments as for (3.126) but with $\left\|\widehat{\boldsymbol{\Sigma}}^\dagger_{\widehat{\mathbf{F}}^0} - \widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}_{w,u}}\right\| = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$ that

$$\left\|\frac{1}{N}\sum_{i=1}^N s_i \widehat{\mathbf{Q}}_i^{-1}\sqrt{N}\left(\frac{\mathbf{V}_i'\overline{\mathbf{U}}^0_{w,-m}}{T}\right)[\widehat{\boldsymbol{\Sigma}}^\dagger_{\widehat{\mathbf{F}}^{0*}} - \widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}_{w,u}}]\left(\frac{(\overline{\mathbf{U}}^0_{w,-m})'\boldsymbol{\varepsilon}_i}{T}\right)\right\| = O_{p^*}\left(\frac{1}{N}\right) + O_{p^*}\left(\frac{1}{\sqrt{NT}}\right)$$

and for the second

$$\left\|\frac{1}{N}\sum_{i=1}^N s_i \widehat{\mathbf{Q}}_i^{-1}\sqrt{N}\left(\frac{\mathbf{V}_i'\mathbf{F}^0}{T}\right)[\widehat{\boldsymbol{\Sigma}}^\dagger_{\widehat{\mathbf{F}}^{0*}} - \widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}_{w,u}}]\left(\frac{(\mathbf{F}^0)'\boldsymbol{\varepsilon}_i}{T}\right)\right\| = O_{p^*}\left(\frac{1}{\sqrt{NT}}\right) + O_{p^*}\left(\frac{1}{T^{3/2}}\right) \tag{3.127}$$

because the fixed limit for $\widehat{\mathbf{Q}}_i^{*-1}$ as $(N,T) \to \infty$ obtained in (3.117) and $\left\|\widehat{\boldsymbol{\Sigma}}^\dagger_{\widehat{\mathbf{F}}^{0*}} - \widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}_{w,u}}\right\| = O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$ from Lemma C-3 imply that the asymptotic order is driven by

$$\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N s_i \left(\frac{\mathbf{V}_i'\mathbf{F}^0}{T}\right)\left(\frac{(\mathbf{F}^0)'\boldsymbol{\varepsilon}_i}{T}\right)\right\| = O_{p^*}\left(\frac{1}{T}\right)$$

which due to the finite moments of s_i and its independence from the other variates follows from the same arguments as for (2.106). Consequently,

$$\left\|\frac{1}{N}\sum_{i=1}^N s_i \widehat{\mathbf{Q}}_i^{*-1}\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0} - \mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon,i}\right\| = O_{p^*}\left(\frac{1}{T}\right) + O_{p^*}\left(\frac{1}{\sqrt{N}}\right)$$

Finally, from the same arguments as for (3.127) given the well behaved fixed and finite limit of $\widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}^0}$ by Ass.2

$$\left\|\frac{1}{N}\sum_{i=1}^N s_i \widehat{\mathbf{Q}}_i^{*-1}\widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon,i}\right\| = \left\|\frac{1}{N}\sum_{i=1}^N s_i \widehat{\mathbf{Q}}_i^{*-1}\sqrt{N}\left(\frac{\mathbf{V}_i'\mathbf{F}^0}{T}\right)\widehat{\boldsymbol{\Sigma}}^\dagger_{\mathbf{F}^0}\left(\frac{(\mathbf{F}^0)'\boldsymbol{\varepsilon}_i}{T}\right)\right\| = O_{p^*}\left(\frac{1}{T}\right)$$

Combining then all the results above we come to

$$\left\|\frac{1}{N}\sum_{i=1}^N s_i \widehat{\mathbf{Q}}_i^{*-1}\widehat{\mathbf{q}}_i^*\right\| = O_{p^*}\left(\frac{1}{\sqrt{N}}\right) + O_{p^*}\left(\frac{1}{\sqrt{T}}\right) \tag{3.128}$$

and so

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^*_{mg} - \boldsymbol{\beta}) = \frac{1}{\sqrt{N}}\sum_{i=1}^N s_i \boldsymbol{v}_i + o_{p^*}(1)$$

In turn, subtracting $\sqrt{N}(\widehat{\boldsymbol{\beta}}_{mg} - \boldsymbol{\beta})$ obtained in (2.109) from both sides gives the expansion around $\widehat{\boldsymbol{\beta}}_{mg}$

$$\begin{aligned}\sqrt{N}(\widehat{\boldsymbol{\beta}}^*_{mg} - \widehat{\boldsymbol{\beta}}_{mg}) &= \frac{1}{\sqrt{N}}\sum_{i=1}^N (s_i - 1)\boldsymbol{v}_i + o_{p^*}(1) \\ &\xrightarrow{d^*} \mathcal{N}(\mathbf{0}_{k\times 1}, \boldsymbol{\Omega}_v),\end{aligned}$$

as $(N,T) \to \infty$, because

$$\begin{aligned}
&\mathbb{E}^*\left[\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(s_i-1)\boldsymbol{v}_i\right)\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}(s_j-1)\boldsymbol{v}_j\right)'\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}^*[(s_i-1)(s_j-1)]\mathbb{E}^*(\boldsymbol{v}_i\boldsymbol{v}_j') = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}^*[(s_i-1)^2]\mathbb{E}^*(\boldsymbol{v}_i\boldsymbol{v}_i') \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}^*(\boldsymbol{v}_i\boldsymbol{v}_i') + O(N^{-1}) \to \boldsymbol{\Omega}_v
\end{aligned}$$

again using $\mathbb{E}^*[(s_i-1)^2] = 1 - N^{-1}$. This is the result stated in the theorem. It then follows directly from this result and Theorem 6 that

$$\sup_{x\in\mathbb{R}^{k\times 1}}\left|\mathbb{P}^*[\sqrt{N}(\widehat{\boldsymbol{\beta}}^*_{mg} - \widehat{\boldsymbol{\beta}}_{mg}) \leq x] - \mathbb{P}[\sqrt{N}(\widehat{\boldsymbol{\beta}}_{mg} - \boldsymbol{\beta}) \leq x]\right| \longrightarrow^p 0,$$

where inequalities are to be interpreted coordinate-wise.

4 Variance Estimation

Theorem 8 *Consistency of variance estimators.*

Under Ass.1-5, or assumptions 1-3, 5 and 7 when $(\bar{\mathbf{y}}, \bar{\mathbf{y}}^*)$ are excluded from the CA, we have as $(N, T) \to \infty$ such that $\tau_{N,T} \to \tau < \infty$ that

$$NT\widehat{\mathbf{\Theta}} \longrightarrow^{p} \mathbf{\Sigma}^{-1}\mathbf{\Psi}\mathbf{\Sigma}^{-1}$$
$$NT\widehat{\mathbf{\Theta}}^* \longrightarrow^{p^*} \mathbf{\Sigma}^{-1}\mathbf{\Psi}\mathbf{\Sigma}^{-1}$$

If in addition Ass.6 and the conditions of Theorem 4 hold, then

$$N\widehat{\mathbf{\Theta}} \longrightarrow^{p} \mathbf{\Sigma}^{-1}\mathbf{\Psi}_h\mathbf{\Sigma}^{-1}$$
$$N\widehat{\mathbf{\Theta}}^* \longrightarrow^{p^*} \mathbf{\Sigma}^{-1}\mathbf{\Psi}_h\mathbf{\Sigma}^{-1}$$

and also

$$N\widehat{\mathbf{\Omega}}_{\boldsymbol{v}} \longrightarrow^{p} \mathbf{\Omega}_{\boldsymbol{v}}$$
$$N\widehat{\mathbf{\Omega}}_{\boldsymbol{v}}^* \longrightarrow^{p^*} \mathbf{\Omega}_{\boldsymbol{v}}$$

as $(N, T) \to \infty$.

Proof.

Making use of the notation introduced in sections 2.3 and 3.4, the variance estimators for the CCEP estimates in respectively the original and bootstrap samples are

$$\widehat{\mathbf{\Theta}} = N^{-1}\overline{\mathbf{Q}}^{-1}\widehat{\mathbf{\Psi}}\overline{\mathbf{Q}}^{-1}, \qquad \widehat{\mathbf{\Psi}} = \left[\frac{1}{N-1}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{mg})(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{mg})'\widehat{\mathbf{Q}}_i\right] \tag{4.1}$$

$$\widehat{\mathbf{\Theta}}^* = N^{-1}\overline{\mathbf{Q}}^{*-1}\widehat{\mathbf{\Psi}}^*\overline{\mathbf{Q}}^{*-1}, \qquad \widehat{\mathbf{\Psi}}^* = \left[\frac{1}{N-1}\sum_{i=1}^{N}s_i\widehat{\mathbf{Q}}_i^*(\widehat{\boldsymbol{\beta}}_i^* - \widehat{\boldsymbol{\beta}}_{mg}^*)(\widehat{\boldsymbol{\beta}}_i^* - \widehat{\boldsymbol{\beta}}_{mg}^*)'\widehat{\mathbf{Q}}_i^*\right] \tag{4.2}$$

where $\widehat{\boldsymbol{\beta}}_i = \widehat{\mathbf{Q}}_i^{-1}T^{-1}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{y}_i$ and $\widehat{\boldsymbol{\beta}}_i^* = \widehat{\mathbf{Q}}_i^{*-1}T^{-1}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}^*}\mathbf{y}_i$. The latter can using (2.4) and (2.5) be decomposed as

$$\widehat{\boldsymbol{\beta}}_i = \boldsymbol{\beta} + \boldsymbol{v}_i + \widehat{\mathbf{Q}}_i^{-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i, \qquad \widehat{\boldsymbol{\beta}}_i^* = \boldsymbol{\beta} + \boldsymbol{v}_i + \widehat{\mathbf{Q}}_i^{*-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i^*$$

such that from $\widehat{\boldsymbol{\beta}}_{mg} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_i$ and $\widehat{\boldsymbol{\beta}}_{mg}^* = \frac{1}{N}\sum_{i=1}^{N}s_i\widehat{\boldsymbol{\beta}}_i^*$ follows

$$\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{mg} = (\boldsymbol{v}_i - \bar{\boldsymbol{v}}) + \widehat{\mathbf{Q}}_i^{-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i - \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i \tag{4.3}$$

$$\widehat{\boldsymbol{\beta}}_i^* - \widehat{\boldsymbol{\beta}}_{mg}^* = (\boldsymbol{v}_i - \bar{\boldsymbol{v}}_w) + \widehat{\mathbf{Q}}_i^{*-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i^* - \frac{1}{N}\sum_{i=1}^{N}s_i\widehat{\mathbf{Q}}_i^{*-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i^* \tag{4.4}$$

where $\overline{\boldsymbol{v}}_w = \frac{1}{N}\sum_{i=1}^{N} s_i \boldsymbol{v}_i$.

Consider first the homogeneous slope setting $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ and note that then $\boldsymbol{v}_i = \mathbf{0}_{k\times 1}$. Applying (4.3) gives in this case

$$\widehat{\mathbf{Q}}_i(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{mg}) = \frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i - \widehat{\mathbf{Q}}_i\left[\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i\right] \\ = T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i + O_p(N^{-1}) + O_p(T^{-1}) + O_p((NT)^{-1/2}) \tag{4.5}$$

because making use of (2.108) yields

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i\right\| = N^{-1/2}\left\|\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\widehat{\mathbf{q}}_i\right\| = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) \tag{4.6}$$

such that $\left\|\widehat{\mathbf{Q}}_i\left[\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i\right]\right\| \leq \left\|\widehat{\mathbf{Q}}_i\right\|\left\|\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i^{-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i\right\| = O_p(N^{-1}) + O_p((NT)^{-1/2})$ for the final term on the first line. For the first term we get by substituting in (2.101)

$$N^{-1/2}\widehat{\mathbf{q}}_i = N^{-1/2}(\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon,i} - \widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon,i} + \widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon,i}) + O_p(N^{-1}) + O_p((NT)^{-1/2}) \\ = T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i + O_p(N^{-1}) + O_p(T^{-1}) + O_p((NT)^{-1/2}) \tag{4.7}$$

since $N^{-1/2}\widehat{\mathbf{q}}_{\mathbf{I},V\varepsilon,i} = T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i$,

$$\left\|N^{-1/2}\widehat{\mathbf{q}}_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon,i}\right\| \leq \left\|\frac{\mathbf{V}_i'\mathbf{F}^0}{T}\right\|\left\|\left(\frac{(\mathbf{F}^0)'\mathbf{F}^0}{T}\right)^{\dagger}\right\|\left\|\frac{(\mathbf{F}^0)'\boldsymbol{\varepsilon}_i}{T}\right\| = O_p\left(\frac{1}{T}\right)$$

and because substituting in $\mathbf{V}_i = \mathbf{U}_i\mathbf{q}_x$ and $\boldsymbol{\varepsilon}_i = \mathbf{U}_i\mathbf{B}^{-1}\mathbf{q}_y$ together with (2.61) results in

$$\left\|N^{-1/2}\widehat{\mathbf{q}}_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}],V\varepsilon,i}\right\| \leq \|\mathbf{q}_x'\|\left\|T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^0}]\mathbf{U}_i\right\|\left\|\mathbf{B}^{-1}\mathbf{q}_y\right\| = O_p(N^{-1}) + O_p(T^{-1}) + O_p((NT)^{-1/2})$$

Then, recalling that $\|T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i\| = O_p(T^{-1/2})$ and $\|T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{V}_i\| = O_p(1)$ by Ass.1, substituting in (4.5) yields

$$T\widehat{\boldsymbol{\Psi}} = T\left[\frac{1}{N-1}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{mg})(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{mg})'\widehat{\mathbf{Q}}_i\right] \\ = \frac{1}{N-1}\sum_{i=1}^{N}\frac{\mathbf{V}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{V}_i}{T} + O_p\left(\frac{\sqrt{T}}{N}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) \\ + O_p\left(\frac{T}{N^2}\right) + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{\sqrt{T}}{N^{3/2}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) \\ = \frac{1}{N-1}\sum_{i=1}^{N}\frac{\mathbf{V}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{V}_i}{T} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

where the third line uses $T/N = O(1)$. Note then that for the leading term we have given the mutual independence of $\boldsymbol{\varepsilon}_i$ and $\mathbf{V}_i$ by Ass.5, and their finite fourth moments and cross-section independence under Ass.1 that

$$\frac{1}{N-1}\sum_{i=1}^{N}\frac{\mathbf{V}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{V}_i}{T} \longrightarrow^p \boldsymbol{\Psi}$$

as $(N,T)\rightarrow\infty$. Therefore, it follows provided $T/N\rightarrow\tau<\infty$

$$T\widehat{\boldsymbol{\Psi}} \longrightarrow^p \boldsymbol{\Psi}$$

In the bootstrap world, we get from near identical steps with (4.4)

$$\begin{aligned}\widehat{\mathbf{Q}}_i^*(\widehat{\boldsymbol{\beta}}_i^*-\widehat{\boldsymbol{\beta}}_{mg}^*) &= \frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i^*-\widehat{\mathbf{Q}}_i^*\left[\frac{1}{N}\sum_{i=1}^{N}s_i\widehat{\mathbf{Q}}_i^{*-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i^*\right]\\ &= T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i+O_{p^*}(N^{-1})+O_{p^*}(T^{-1})+O_{p^*}((NT)^{-1/2})\end{aligned}\tag{4.8}$$

For completeness, this follows because we have from eq.(3.128)

$$\left\|\frac{1}{N}\sum_{i=1}^{N}s_i\widehat{\mathbf{Q}}_i^{*-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i^*\right\| = O_{p^*}\left(\frac{1}{N}\right)+O_{p^*}\left(\frac{1}{\sqrt{NT}}\right)\tag{4.9}$$

which with $\left\|\widehat{\mathbf{Q}}_i^*\right\|=O_{p^*}(1)$ leads to $\left\|\widehat{\mathbf{Q}}_i^*\left[\frac{1}{N}\sum_{i=1}^{N}s_i\widehat{\mathbf{Q}}_i^{*-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i^*\right]\right\|\leq\left\|\widehat{\mathbf{Q}}_i^*\right\|\left\|\frac{1}{N}\sum_{i=1}^{N}s_i\widehat{\mathbf{Q}}_i^{*-1}\frac{1}{\sqrt{N}}\widehat{\mathbf{q}}_i^*\right\|=O_{p^*}(N^{-1})+O_{p^*}((NT)^{-1/2})$ for the second term on the first line. For the first term,

$$N^{-1/2}\widehat{\mathbf{q}}_i^* = T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i+O_{p^*}(N^{-1})+O_{p^*}(T^{-1})+O_{p^*}((NT)^{-1/2})\tag{4.10}$$

is obtained by first substituting in (3.125) and subsequently $\left\|N^{-1/2}\widehat{\mathbf{q}}^*_{\mathbf{P}_{\mathbf{F}^0},V\varepsilon,i}\right\|\leq\left\|\frac{\mathbf{V}_i'\mathbf{F}^0}{T}\right\|\left\|\left(\frac{(\mathbf{F}^0)'\mathbf{F}^0}{T}\right)^{\dagger}\right\|\left\|\frac{(\mathbf{F}^0)'\boldsymbol{\varepsilon}_i}{T}\right\|=O_{p^*}(T^{-1})$ and $\left\|N^{-1/2}\widehat{\mathbf{q}}^*_{[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}],V\varepsilon,i}\right\|\leq\|\mathbf{q}_x'\|\left\|T^{-1}\mathbf{U}_i'[\mathbf{M}_{\mathbf{F}^0}-\mathbf{M}_{\widehat{\mathbf{F}}^{0*}}]\mathbf{U}_i\right\|\left\|\mathbf{B}^{-1}\mathbf{q}_y\right\|=O_{p^*}(N^{-1})+O_{p^*}(T^{-1})+O_{p^*}((NT)^{-1/2})$ by eq.(3.66). Substituting in (4.8) and making use of $\tau_{N,T}=O(1)$ and $s_i=O_{p^*}(1)$ then results in

$$\begin{aligned}T\widehat{\boldsymbol{\Psi}}^* &= T\left[\frac{1}{N-1}\sum_{i=1}^{N}s_i\widehat{\mathbf{Q}}_i^*(\widehat{\boldsymbol{\beta}}_i^*-\widehat{\boldsymbol{\beta}}_{mg}^*)(\widehat{\boldsymbol{\beta}}_i^*-\widehat{\boldsymbol{\beta}}_{mg}^*)'\widehat{\mathbf{Q}}_i^*\right]\\ &= \frac{1}{N-1}\sum_{i=1}^{N}s_i\frac{\mathbf{V}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{V}_i}{T}+O_{p^*}\left(\frac{1}{\sqrt{N}}\right)+O_{p^*}\left(\frac{1}{\sqrt{T}}\right)\\ &\longrightarrow^{p^*}\boldsymbol{\Psi}\end{aligned}$$

by $\mathbb{E}^*(s_i)=1$, $\mathbb{E}^*(s_i^2)=O(1)$ and the independence of s_i, $\mathbf{V}_i$ and $\boldsymbol{\varepsilon}_i$.
Finally, with $\overline{\mathbf{Q}}^{-1}\longrightarrow^p\boldsymbol{\Sigma}^{-1}$ and $\overline{\mathbf{Q}}^{*-1}\longrightarrow^{p^*}\boldsymbol{\Sigma}^{-1}$ from Lemmas B-6 and C-7, we come to

$$\begin{aligned}NT\widehat{\boldsymbol{\Theta}} &= \overline{\mathbf{Q}}^{-1}T\widehat{\boldsymbol{\Psi}}\overline{\mathbf{Q}}^{-1}\longrightarrow^p\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\\ NT\widehat{\boldsymbol{\Theta}}^* &= \overline{\mathbf{Q}}^{*-1}T\widehat{\boldsymbol{\Psi}}^*\overline{\mathbf{Q}}^{*-1}\longrightarrow^{p^*}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\end{aligned}$$

as stated in the theorem. Arguments are identical when $\mathbf{M}_{\widehat{\mathbf{F}}_{\mathbf{x}}}$ is employed in stead of $\mathbf{M}_{\widehat{\mathbf{F}}}$, provided that $rk(\overline{\boldsymbol{\Gamma}})=m$.

Next, consider (4.1)-(4.2) under heterogeneous slopes characterized by Ass.6. In this case (4.3)-(4.4) give

$$\widehat{\beta}_i - \widehat{\beta}_{mg} = (\boldsymbol{v}_i - \overline{\boldsymbol{v}}) + \widehat{\mathbf{Q}}_i^{-1} \frac{1}{\sqrt{N}} \widehat{\mathbf{q}}_i - \frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbf{Q}}_i^{-1} \frac{1}{\sqrt{N}} \widehat{\mathbf{q}}_i = (\boldsymbol{v}_i - \overline{\boldsymbol{v}}) + O_p(N^{-1}) + O_p(T^{-1/2}) \tag{4.11}$$

$$\widehat{\beta}_i^* - \widehat{\beta}_{mg}^* = (\boldsymbol{v}_i - \overline{\boldsymbol{v}}_w) + \widehat{\mathbf{Q}}_i^{*-1} \frac{1}{\sqrt{N}} \widehat{\mathbf{q}}_i^* - \frac{1}{N} \sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^{*-1} \frac{1}{\sqrt{N}} \widehat{\mathbf{q}}_i^* = (\boldsymbol{v}_i - \overline{\boldsymbol{v}}_w) + O_{p^*}(N^{-1}) + O_{p^*}(T^{-1/2}) \tag{4.12}$$

which follows from substituting in (4.6) and (4.9) for the last terms in each equation and because $\left\|T^{-1}\mathbf{V}_i'\boldsymbol{\varepsilon}_i\right\| = O_p(T^{-1/2})$, $\left\|\widehat{\mathbf{Q}}_i^{-1}\right\| = O_p(1)$ and $\left\|\widehat{\mathbf{Q}}_i^{*-1}\right\| = O_{p^*}(1)$ together with (4.7) and (4.10) give

$$\left\|\widehat{\mathbf{Q}}_i^{-1} N^{-1/2} \widehat{\mathbf{q}}_i\right\| \leq \left\|\widehat{\mathbf{Q}}_i^{-1}\right\| \left\|N^{-1/2}\widehat{\mathbf{q}}_i\right\| = O_p(T^{-1/2}) + O_p(N^{-1})$$
$$\left\|\widehat{\mathbf{Q}}_i^{*-1} N^{-1/2} \widehat{\mathbf{q}}_i\right\| \leq \left\|\widehat{\mathbf{Q}}_i^{*-1}\right\| \left\|N^{-1/2}\widehat{\mathbf{q}}_i^*\right\| = O_{p^*}(T^{-1/2}) + O_{p^*}(N^{-1})$$

Therefore,

$$\widehat{\mathbf{Q}}_i(\widehat{\beta}_i - \widehat{\beta}_{mg}) = \widehat{\mathbf{Q}}_i \boldsymbol{v}_i + O_p(N^{-1/2}) + O_p(T^{-1/2})$$
$$\widehat{\mathbf{Q}}_i^*(\widehat{\beta}_i^* - \widehat{\beta}_{mg}^*) = \widehat{\mathbf{Q}}_i^* \boldsymbol{v}_i^* + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2})$$

because $\left\|\widehat{\mathbf{Q}}_i\overline{\boldsymbol{v}}\right\| \leq \left\|\widehat{\mathbf{Q}}_i\right\| \|\overline{\boldsymbol{v}}\| = O_p(N^{-1/2})$ by Ass.6 and also $\left\|\widehat{\mathbf{Q}}_i^*\overline{\boldsymbol{v}}_w\right\| \leq \left\|\widehat{\mathbf{Q}}_i^*\right\| \|\overline{\boldsymbol{v}}_w\| = O_{p^*}(N^{-1/2})$ because s_i has finite second moments and is independent from $\boldsymbol{v}_i$. Since in addition $\left\|\widehat{\mathbf{Q}}_i\boldsymbol{v}_i\right\| = O_p(1)$, $\left\|\widehat{\mathbf{Q}}_i^*\boldsymbol{v}_i\right\| = O_{p^*}(1)$ these final results lead to

$$\begin{aligned}
\widehat{\mathbf{\Psi}} &= \frac{1}{N-1} \sum_{i=1}^{N} \widehat{\mathbf{Q}}_i(\widehat{\beta}_i - \widehat{\beta}_{mg})(\widehat{\beta}_i - \widehat{\beta}_{mg})'\widehat{\mathbf{Q}}_i = \frac{1}{N-1} \sum_{i=1}^{N} \widehat{\mathbf{Q}}_i \boldsymbol{v}_i \boldsymbol{v}_i' \widehat{\mathbf{Q}}_i + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= \frac{1}{N-1} \sum_{i=1}^{N} \mathbf{\Sigma}_i \boldsymbol{v}_i \boldsymbol{v}_i' \mathbf{\Sigma}_i + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&\longrightarrow^p \mathbf{\Psi}_h
\end{aligned}$$

as $(N, T) \to \infty$, due to $\mathbb{E}(\|\boldsymbol{v}_i\|^6) < \infty$ under Theorem 4 and cross-section independence of $\boldsymbol{v}_i$ by Ass.6, where $\mathbf{\Psi}_h$ was defined in Theorem 4 and we made use of $\widehat{\mathbf{Q}}_i = \mathbf{\Sigma}_i + o_p(1)$ from (3.116). Similarly, in the bootstrap world, since again s_i is independent from $\boldsymbol{v}_i$, $\mathbb{E}^*(s_i) = 1$ and $\mathbb{E}^*(s_i^2) = O(1)$,

$$\begin{aligned}
\widehat{\mathbf{\Psi}}^* &= \frac{1}{N-1} \sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^*(\widehat{\beta}_i^* - \widehat{\beta}_{mg}^*)(\widehat{\beta}_i^* - \widehat{\beta}_{mg}^*)'\widehat{\mathbf{Q}}_i^* = \frac{1}{N-1} \sum_{i=1}^{N} s_i \widehat{\mathbf{Q}}_i^* \boldsymbol{v}_i \boldsymbol{v}_i' \widehat{\mathbf{Q}}_i^* + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\
&= \frac{1}{N-1} \sum_{i=1}^{N} s_i \mathbf{\Sigma}_i \boldsymbol{v}_i \boldsymbol{v}_i' \mathbf{\Sigma}_i + O_{p^*}(N^{-1/2}) + O_{p^*}(T^{-1/2}) \\
&\longrightarrow^{p^*} \mathbf{\Psi}_h
\end{aligned}$$

as $(N, T) \to \infty$. In conclusion, again making use of Lemmas B-6 and C-7, we come to

$$N\widehat{\mathbf{\Theta}} = \overline{\mathbf{Q}}^{-1}\widehat{\mathbf{\Psi}}\overline{\mathbf{Q}}^{-1} \longrightarrow^p \mathbf{\Sigma}^{-1}\mathbf{\Psi}_h\mathbf{\Sigma}^{-1}$$
$$N\widehat{\mathbf{\Theta}}^* = \overline{\mathbf{Q}}^{*-1}\widehat{\mathbf{\Psi}}^*\overline{\mathbf{Q}}^{*-1} \longrightarrow^{p^*} \mathbf{\Sigma}^{-1}\mathbf{\Psi}_h\mathbf{\Sigma}^{-1}$$

as required.

Finally, eq.(4.11) yields also for the estimator of the Mean Group variance

$$
\begin{aligned}
N\widehat{\boldsymbol{\Omega}}_v &= \frac{1}{N-1}\sum_{i=1}^{N}(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{mg})(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{mg})' = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{v}_i - \overline{\boldsymbol{v}})(\boldsymbol{v}_i - \overline{\boldsymbol{v}})' + O_p(N^{-1}) + O_p(T^{-1/2}) \\
&= \frac{1}{N-1}\sum_{i=1}^{N}\boldsymbol{v}_i\boldsymbol{v}_i' + O_p(N^{-1}) + O_p(T^{-1/2}) \\
&\longrightarrow^{p} \boldsymbol{\Omega}_v
\end{aligned}
$$

as $(N, T) \to \infty$, and similarly from (4.12) for the bootstrap sample estimator

$$
\begin{aligned}
N\widehat{\boldsymbol{\Omega}}_v^* &= \frac{1}{N-1}\sum_{i=1}^{N}s_i(\widehat{\boldsymbol{\beta}}_i^* - \widehat{\boldsymbol{\beta}}_{mg}^*)(\widehat{\boldsymbol{\beta}}_i^* - \widehat{\boldsymbol{\beta}}_{mg}^*)' = \frac{1}{N-1}\sum_{i=1}^{N}s_i(\boldsymbol{v}_i - \overline{\boldsymbol{v}}_w)(\boldsymbol{v}_i - \overline{\boldsymbol{v}}_w)' + O_{p^*}(N^{-1}) + O_{p^*}(T^{-1/2}) \\
&= \frac{1}{N-1}\sum_{i=1}^{N}s_i\boldsymbol{v}_i\boldsymbol{v}_i' + O_{p^*}(N^{-1}) + O_{p^*}(T^{-1/2}) \\
&\longrightarrow^{p^*} \boldsymbol{\Omega}_v
\end{aligned}
$$

since again $\mathbb{E}^*(s_i) = 1$, and making use of earlier definitions $\left\| \frac{1}{N}\sum_{i=1}^{N} s_i\boldsymbol{v}_i\overline{\boldsymbol{v}}_w' \right\| = \|\overline{\boldsymbol{v}}_w\overline{\boldsymbol{v}}_w'\| \leq \|\overline{\boldsymbol{v}}_w\|^2 = O_{p^*}(N^{-1})$. This completes the proof.

References

Abadir, K. M. and Magnus, J. (2005). *Matrix Algebra*. Cambridge University Press.

Billingsley, P. (1995). *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 3rd edition.

Bose, A. and Chatterjee, S. (2002). Comparison of bootstrap and jackknife variance estimators in linear regression: second order results. *Statistica Sinica*, 12(2).

Chatterjee, S. (1998). Another look at the jackknife: further examples of generalized bootstrap. *Statistics & probability letters*, 40(4):307–319.

Galvao, A. F. and Kato, K. (2014). Estimation and inference for linear panel data models under misspecification when both n and t are large. *Journal of Business & Economic Statistics*, 32(2):285–309.

Gonçalves, S. and Kaffo, M. (2015). Bootstrap inference for linear dynamic panel data models with individual fixed effects. *Journal of Econometrics*, 186(2):407–426. High Dimensional Problems in Econometrics.

Karabiyik, H., Reese, S., and Westerlund, J. (2017). On the role of the rank condition in CCE estimation of factor-augmented panel regressions. *Journal of Econometrics*, 197(1):60 – 64.

Pesaran, M. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica*, 74(4):967–1012.

Supplement B: Monte Carlo tables for "Bootstrap-Improved Inference for Factor Augmented Regressions with CCE"*

Ignace De Vos[1,2] and Ovidijus Stauskas[1]

[1]*Lund University, Department of Economics*
[2]*Ghent University, Department of Economics*

Content of the supplement

This supplement contains additional Monte Carlo results that are not reported in the main article. The tables are organized as follows:

- Tables regarding estimation (bias and root mean square error): Section 1
- Tables regarding hypothesis testing (empirical size): Section 2

*The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the Flemish Supercomputer Center, funded by Ghent University; the Hercules Foundation; and the Economy, Science, and Innovation Department of the Flemish Government.

1 Estimation tables

Table A-1: Estimation results: $\beta = 5$ setting, fixed slopes

		bias × 100						*rmse* × 100					
	(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
CCEP	25	3.05	2.67	2.74	2.81	2.70	2.75	6.35	5.09	4.15	3.60	3.09	2.99
	50	1.56	1.50	1.42	1.39	1.37	1.41	4.49	3.28	2.60	2.09	1.69	1.58
	100	0.87	0.72	0.77	0.75	0.75	0.78	2.93	2.19	1.76	1.30	1.05	0.92
	200	0.52	0.38	0.36	0.37	0.38	0.37	2.06	1.52	1.14	0.84	0.61	0.51
	500	0.14	0.11	0.18	0.15	0.15	0.16	1.25	0.94	0.69	0.50	0.34	0.26
	1000	0.05	0.06	0.05	0.07	0.07	0.07	0.88	0.66	0.48	0.34	0.23	0.17
pairs	25	1.22	0.85	0.87	0.95	0.76	0.84	7.06	5.17	3.67	2.66	1.75	1.42
	50	0.33	0.32	0.21	0.17	0.15	0.19	5.11	3.33	2.41	1.68	1.05	0.75
	100	0.24	0.04	0.08	0.07	0.07	0.10	3.33	2.36	1.72	1.13	0.75	0.51
	200	0.20	0.01	-0.01	0.00	0.02	0.01	2.36	1.64	1.15	0.79	0.49	0.36
	500	-0.01	-0.05	0.04	-0.01	0.00	0.01	1.44	1.04	0.71	0.50	0.31	0.21
	1000	-0.04	-0.02	-0.02	-0.01	0.00	-0.01	1.02	0.72	0.51	0.35	0.22	0.15
CCEP_x	25	3.17	2.73	2.80	2.88	2.80	2.85	6.61	5.27	4.22	3.69	3.21	3.09
	50	1.56	1.53	1.44	1.43	1.41	1.44	4.58	3.36	2.63	2.13	1.73	1.61
	100	0.85	0.72	0.78	0.76	0.76	0.79	2.99	2.17	1.76	1.32	1.05	0.93
	200	0.52	0.39	0.37	0.37	0.38	0.38	2.09	1.50	1.15	0.84	0.61	0.51
	500	0.12	0.12	0.17	0.15	0.16	0.16	1.25	0.94	0.69	0.50	0.34	0.26
	1000	0.07	0.07	0.05	0.07	0.07	0.07	0.88	0.66	0.48	0.34	0.23	0.17
pairs_x	25	1.08	0.64	0.65	0.74	0.59	0.64	6.58	4.89	3.38	2.45	1.63	1.24
	50	0.28	0.27	0.15	0.15	0.10	0.14	4.80	3.19	2.28	1.61	1.01	0.72
	100	0.19	0.03	0.08	0.06	0.06	0.09	3.14	2.18	1.63	1.10	0.73	0.50
	200	0.20	0.02	0.01	0.01	0.01	0.01	2.24	1.53	1.12	0.77	0.49	0.35
	500	-0.03	-0.04	0.03	-0.01	0.01	0.01	1.35	0.98	0.69	0.49	0.30	0.21
	1000	-0.01	0.00	-0.02	-0.01	0.00	0.00	0.95	0.69	0.49	0.34	0.22	0.15

Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma^2_\eta, \sigma^2_v) = (10, 5, 1, 1, 0)$, with $m = 2$ factors and $k = 3$ regressors. CCEP and CCEP_x denote respectively the CCEP estimator with and without $\bar{y}$ included in the matrix of CA. 'Pairs' and 'pairs_x' correspond to their respective bootstrap-corrected estimates obtained from 2000 bootstrap replications with the pairs (cross-section) resampling algorithm.

Table A-2: Estimation results: $\sigma^2 = 5$ setting, fixed slopes

		bias × 100						*rmse* × 100					
	(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
CCEP	25	3.57	2.53	2.94	2.87	2.73	2.88	13.02	9.95	7.63	5.75	4.23	3.63
	50	1.77	1.60	1.46	1.43	1.32	1.36	9.42	6.86	4.96	3.70	2.51	2.04
	100	1.13	0.79	0.81	0.81	0.81	0.83	6.32	4.80	3.52	2.52	1.80	1.37
	200	0.70	0.35	0.38	0.36	0.40	0.38	4.45	3.33	2.42	1.72	1.14	0.85
	500	0.14	0.03	0.20	0.14	0.16	0.18	2.82	2.14	1.48	1.06	0.70	0.49
	1000	0.03	0.10	0.10	0.06	0.09	0.08	1.95	1.53	1.07	0.74	0.48	0.34
pairs	25	1.66	0.54	0.94	0.86	0.62	0.78	15.14	10.93	7.77	5.39	3.41	2.38
	50	0.44	0.35	0.20	0.16	0.04	0.08	11.01	7.44	5.10	3.57	2.19	1.55
	100	0.58	0.12	0.12	0.12	0.11	0.13	7.26	5.28	3.65	2.49	1.65	1.12
	200	0.42	0.01	0.02	0.00	0.03	0.01	5.16	3.67	2.51	1.73	1.09	0.77
	500	-0.03	-0.13	0.06	-0.02	0.01	0.03	3.27	2.39	1.55	1.09	0.69	0.47
	1000	-0.07	0.02	0.02	-0.02	0.01	0.00	2.25	1.65	1.12	0.76	0.47	0.34
CCEP$_x$	25	3.63	2.62	2.94	2.89	2.75	2.90	13.02	9.89	7.56	5.74	4.25	3.65
	50	1.76	1.62	1.46	1.46	1.34	1.36	9.49	6.88	4.92	3.71	2.52	2.05
	100	1.04	0.73	0.79	0.83	0.80	0.84	6.42	4.80	3.47	2.52	1.80	1.38
	200	0.72	0.29	0.37	0.37	0.39	0.38	4.46	3.31	2.44	1.72	1.14	0.85
	500	0.16	0.03	0.18	0.14	0.16	0.18	2.81	2.13	1.48	1.06	0.70	0.49
	1000	0.06	0.11	0.09	0.06	0.09	0.08	1.96	1.55	1.08	0.74	0.47	0.34
pairs$_x$	25	1.55	0.57	0.82	0.75	0.53	0.68	14.07	10.28	7.40	5.22	3.33	2.31
	50	0.42	0.34	0.17	0.18	0.03	0.07	10.37	7.15	4.88	3.51	2.16	1.53
	100	0.42	0.05	0.09	0.14	0.10	0.13	6.95	5.07	3.48	2.44	1.63	1.11
	200	0.45	-0.07	0.02	0.01	0.03	0.02	4.86	3.49	2.49	1.71	1.08	0.76
	500	0.00	-0.13	0.03	-0.02	0.01	0.03	3.03	2.27	1.51	1.07	0.68	0.46
	1000	-0.02	0.05	0.02	-0.02	0.01	0.00	2.12	1.61	1.10	0.75	0.47	0.34

Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (10, 1, 5, 1, 0)$, with $m = 2$ factors and $k = 3$ regressors. CCEP and CCEP$_x$ denote respectively the CCEP estimator with and without $\bar{y}$ included in the matrix of CA. 'Pairs' and 'pairs$_x$' correspond to their respective bootstrap-corrected estimates obtained from 2000 bootstrap replications with the pairs (cross-section) resampling algorithm.

Table A-3: Estimation results: heterogeneous slopes ($\sigma_v^2 = 5$ setting)

		bias × 100						*rmse* × 100					
	(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
CCEP	25	-3.97	-5.51	-4.91	-6.10	-5.83	-6.73	58.61	52.09	47.93	47.50	45.68	46.20
	50	-5.09	-4.93	-0.94	-3.12	-1.31	-5.79	42.46	38.71	37.89	35.89	35.54	32.71
	100	-0.85	-1.07	-2.99	-2.43	-2.16	-2.54	32.89	28.77	27.06	27.06	26.75	25.34
	200	-0.75	-0.80	-1.32	-1.29	-1.15	-1.48	23.80	21.76	20.74	19.27	18.43	18.41
	500	-0.74	-1.13	-0.71	-0.28	-0.34	-0.62	14.97	13.78	12.83	12.77	12.25	11.85
	1000	0.39	0.29	0.15	0.09	-0.43	-0.18	10.92	10.18	8.95	8.83	8.84	8.75
pairs	25	-2.75	-3.84	-3.54	-4.29	-4.11	-4.97	69.34	58.40	52.78	51.68	48.98	49.84
	50	-3.04	-3.29	1.10	-1.17	0.57	-4.11	49.14	42.68	41.32	38.54	37.84	34.46
	100	0.54	0.37	-1.55	-0.88	-0.67	-1.06	37.56	31.37	28.72	28.52	28.02	26.56
	200	0.26	0.17	-0.34	-0.34	-0.15	-0.49	26.23	23.30	21.72	19.95	19.09	18.96
	500	-0.08	-0.62	-0.17	0.30	0.22	-0.06	16.35	14.43	13.22	13.08	12.50	12.07
	1000	0.83	0.59	0.49	0.41	-0.11	0.14	11.86	10.64	9.17	8.98	8.96	8.85
CCEP_x	25	5.35	2.59	2.85	1.49	2.33	1.11	67.00	58.67	53.99	53.89	51.94	51.97
	50	-0.73	-0.53	3.92	1.58	3.33	-1.08	45.92	40.98	41.55	38.88	39.33	35.30
	100	1.90	1.85	-0.63	0.17	0.43	-0.02	35.44	30.74	28.40	28.57	28.16	26.72
	200	0.94	0.79	0.07	0.10	0.15	-0.15	24.46	22.87	21.40	19.97	19.01	18.99
	500	-0.40	-0.51	-0.15	0.26	0.26	-0.05	15.18	13.85	13.04	12.94	12.47	12.03
	1000	0.59	0.68	0.46	0.39	-0.14	0.13	11.12	10.30	9.04	8.91	8.90	8.82
pairs_x	25	3.31	0.31	0.49	-0.78	0.05	-1.19	75.69	63.23	57.24	56.77	54.27	54.07
	50	-1.75	-1.73	2.69	0.26	2.09	-2.36	50.07	43.19	43.56	40.25	40.33	36.34
	100	1.07	1.21	-1.27	-0.47	-0.32	-0.74	38.01	32.04	29.13	29.13	28.57	27.17
	200	0.65	0.43	-0.31	-0.33	-0.23	-0.53	25.52	23.62	21.72	20.20	19.20	19.18
	500	-0.61	-0.67	-0.32	0.11	0.11	-0.17	15.85	14.08	13.16	13.02	12.52	12.08
	1000	0.53	0.62	0.41	0.31	-0.20	0.05	11.57	10.48	9.10	8.93	8.94	8.84
CCEMG	25	3.52	3.82	3.85	1.52	2.54	1.74	45.51	44.76	44.64	43.58	44.17	44.31
	50	1.47	-0.08	3.60	1.82	3.17	0.13	31.29	31.46	31.23	31.09	32.77	30.29
	100	1.52	1.93	0.06	0.43	0.36	0.18	22.86	23.13	22.00	23.19	22.94	21.99
	200	0.84	0.73	0.37	0.90	0.59	-0.11	16.28	15.95	15.69	16.18	15.82	15.37
	500	-0.32	-0.08	-0.20	0.38	0.21	-0.13	9.94	10.39	9.93	9.99	10.19	9.72
	1000	0.42	0.52	0.18	0.41	-0.24	0.25	7.05	7.07	7.07	6.88	7.29	7.05
pairs_{MG}	25	1.89	2.31	2.19	0.02	0.95	0.15	45.55	44.76	44.51	43.65	44.18	44.30
	50	0.19	-1.26	2.37	0.60	1.95	-1.05	31.37	31.51	31.07	31.08	32.70	30.36
	100	0.78	1.17	-0.71	-0.31	-0.41	-0.56	22.88	23.12	22.04	23.22	22.97	21.99
	200	0.40	0.30	-0.05	0.48	0.13	-0.55	16.31	15.96	15.69	16.17	15.79	15.37
	500	-0.51	-0.27	-0.38	0.19	0.02	-0.33	9.98	10.40	9.95	10.00	10.18	9.72
	1000	0.32	0.42	0.08	0.31	-0.35	0.16	7.07	7.07	7.08	6.88	7.30	7.04
CCEMG_x	25	3.66	3.87	3.93	1.57	2.64	1.81	45.55	44.76	44.66	43.56	44.17	44.32
	50	1.47	-0.06	3.60	1.85	3.20	0.16	31.29	31.48	31.22	31.09	32.78	30.29
	100	1.54	1.97	0.09	0.44	0.36	0.18	22.87	23.14	22.00	23.19	22.93	21.99
	200	0.87	0.72	0.37	0.90	0.59	-0.11	16.30	15.94	15.70	16.19	15.82	15.37
	500	-0.34	-0.09	-0.19	0.38	0.21	-0.13	9.94	10.40	9.92	9.99	10.20	9.72
	1000	0.42	0.52	0.17	0.41	-0.24	0.25	7.03	7.07	7.07	6.88	7.29	7.05
$\text{pairs}_{MG,x}$	25	1.85	2.14	2.09	-0.18	0.76	-0.02	45.47	44.71	44.46	43.55	44.14	44.30
	50	0.18	-1.32	2.33	0.59	1.92	-1.08	31.38	31.59	31.11	31.04	32.68	30.35
	100	0.80	1.18	-0.67	-0.31	-0.41	-0.60	22.87	23.16	22.02	23.19	22.93	21.96
	200	0.46	0.29	-0.07	0.47	0.15	-0.55	16.29	15.94	15.71	16.17	15.82	15.39
	500	-0.53	-0.28	-0.38	0.19	0.01	-0.30	9.97	10.42	9.92	10.00	10.20	9.74
	1000	0.33	0.41	0.07	0.30	-0.34	0.15	7.04	7.05	7.08	6.88	7.30	7.04

Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (10, 1, 1, 1, 5)$, with $m = 2$ factors and $k = 3$ regressors. CCEP and CCEMG denote the CCE estimators with $\bar{y}$ included in the matrix of CA, whereas CCEP_x, CCEMG_x are the versions without $\bar{y}$. The 'pairs' and 'pairs_x' correspond to the respective bootstrap-corrections for the CCEP/CCEP_x estimators, and 'pairs' with an additional MG subscript, i.e. pairs_{MG}/$\text{pairs}_{MG,x}$, denote the bootstrap corrections of CCEMG/CCEMG_x. All corrections are obtained from 2000 bootstrap replications with the pairs (cross-section) resampling algorithm.

2 Inference tables

Table B-1: Empirical size: $\beta = 5$ setting, fixed slopes

	CCEP											
	t-test						*basic*					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.10	0.13	0.18	0.30	0.47	0.69	0.07	0.07	0.05	0.05	0.03	0.01
50	0.08	0.10	0.11	0.17	0.33	0.50	0.09	0.07	0.05	0.04	0.03	0.01
100	0.07	0.07	0.10	0.11	0.24	0.37	0.06	0.06	0.07	0.05	0.05	0.03
200	0.07	0.05	0.07	0.09	0.14	0.19	0.07	0.06	0.06	0.06	0.05	0.05
500	0.06	0.06	0.05	0.07	0.08	0.10	0.07	0.07	0.05	0.06	0.06	0.04
1000	0.05	0.05	0.06	0.06	0.07	0.08	0.07	0.06	0.06	0.05	0.05	0.05
	Bootstrap$-t$						Bootstrap$-t_c$					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.11	0.09	0.08	0.08	0.09	0.12	0.11	0.10	0.08	0.08	0.07	0.06
50	0.12	0.09	0.08	0.07	0.08	0.07	0.12	0.09	0.07	0.07	0.07	0.04
100	0.10	0.07	0.09	0.06	0.07	0.06	0.10	0.07	0.09	0.07	0.07	0.05
200	0.09	0.08	0.07	0.07	0.06	0.07	0.09	0.08	0.07	0.07	0.06	0.07
500	0.08	0.08	0.06	0.06	0.06	0.05	0.08	0.08	0.06	0.06	0.06	0.05
1000	0.07	0.06	0.07	0.06	0.06	0.06	0.07	0.06	0.06	0.06	0.06	0.06

	CCEP$_x$											
	t-test						*basic*					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.09	0.12	0.17	0.28	0.47	0.68	0.04	0.05	0.03	0.03	0.01	0.00
50	0.08	0.09	0.10	0.17	0.32	0.51	0.07	0.04	0.03	0.03	0.01	0.01
100	0.07	0.06	0.10	0.12	0.23	0.37	0.05	0.05	0.05	0.04	0.04	0.03
200	0.07	0.05	0.08	0.08	0.13	0.19	0.06	0.05	0.05	0.05	0.05	0.04
500	0.05	0.06	0.05	0.06	0.08	0.10	0.05	0.06	0.05	0.05	0.05	0.04
1000	0.05	0.05	0.05	0.06	0.07	0.08	0.06	0.05	0.05	0.05	0.06	0.06
	Bootstrap$-t$						Bootstrap$-t_c$					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.10	0.08	0.07	0.07	0.07	0.09	0.09	0.08	0.07	0.07	0.06	0.03
50	0.12	0.08	0.06	0.07	0.06	0.06	0.12	0.08	0.06	0.06	0.05	0.03
100	0.09	0.06	0.08	0.06	0.06	0.06	0.09	0.06	0.08	0.06	0.06	0.05
200	0.08	0.07	0.06	0.06	0.06	0.06	0.08	0.07	0.06	0.07	0.06	0.06
500	0.06	0.07	0.06	0.06	0.06	0.04	0.06	0.07	0.06	0.06	0.06	0.04
1000	0.06	0.05	0.06	0.06	0.07	0.06	0.06	0.05	0.06	0.06	0.06	0.06

Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (10, 5, 1, 1, 0)$, with $m = 2$ factors and $k = 3$ regressors. CCEP and CCEP$_x$ denote respectively the CCEP estimator with and without $\bar{y}$ included in the matrix of CA. '*t*-test' resports the empirical size for a *t*-test at the $\alpha = 0.05$ significance level. '*basic*' reports empirical size for tests based on the basic ('empirical percentile') bootstrap interval, and bootstrap$-t$ and bootstrap$-t_c$ are respectively empirical size for the plain and corrected bootstrap$-t$ interval. All bootstrap tests are based on $B = 2000$ replications with the pairs (cross-section) resampling algorithm.

Table B-2: Empirical size: $\sigma^2 = 5$ setting, fixed slopes

CCEP

	t-test						*basic*					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.08	0.08	0.11	0.12	0.17	0.26	0.09	0.09	0.09	0.08	0.05	0.04
50	0.07	0.07	0.07	0.08	0.10	0.16	0.09	0.07	0.06	0.06	0.05	0.04
100	0.06	0.07	0.06	0.06	0.11	0.12	0.07	0.07	0.07	0.05	0.07	0.05
200	0.07	0.05	0.06	0.06	0.07	0.08	0.09	0.06	0.06	0.06	0.06	0.06
500	0.06	0.06	0.06	0.05	0.06	0.05	0.08	0.07	0.06	0.05	0.05	0.04
1000	0.05	0.06	0.05	0.05	0.05	0.05	0.07	0.06	0.06	0.05	0.05	0.06

	Bootstrap$-t$						Bootstrap$-t_c$					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.09	0.08	0.08	0.08	0.06	0.05	0.09	0.07	0.08	0.08	0.06	0.05
50	0.10	0.07	0.06	0.05	0.05	0.05	0.09	0.07	0.05	0.05	0.05	0.05
100	0.07	0.07	0.07	0.05	0.07	0.06	0.07	0.06	0.07	0.05	0.07	0.05
200	0.09	0.07	0.06	0.06	0.05	0.05	0.09	0.07	0.06	0.06	0.05	0.06
500	0.08	0.08	0.06	0.05	0.05	0.04	0.08	0.08	0.05	0.05	0.05	0.04
1000	0.07	0.06	0.06	0.05	0.04	0.06	0.07	0.06	0.06	0.05	0.04	0.06

CCEP_x

	t-test						*basic*					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.07	0.07	0.09	0.12	0.16	0.25	0.08	0.07	0.08	0.06	0.04	0.03
50	0.06	0.06	0.06	0.07	0.09	0.15	0.08	0.06	0.05	0.05	0.05	0.04
100	0.06	0.06	0.05	0.06	0.10	0.11	0.06	0.07	0.05	0.04	0.07	0.05
200	0.05	0.05	0.06	0.06	0.07	0.08	0.07	0.05	0.06	0.06	0.05	0.06
500	0.05	0.06	0.05	0.05	0.06	0.05	0.05	0.06	0.05	0.04	0.05	0.05
1000	0.04	0.06	0.06	0.05	0.05	0.05	0.06	0.06	0.06	0.05	0.04	0.05

	Bootstrap$-t$						Bootstrap$-t_c$					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.08	0.06	0.07	0.07	0.06	0.05	0.08	0.06	0.07	0.07	0.06	0.05
50	0.08	0.07	0.05	0.06	0.06	0.04	0.08	0.07	0.05	0.06	0.06	0.04
100	0.07	0.06	0.05	0.04	0.07	0.06	0.07	0.06	0.06	0.05	0.07	0.05
200	0.07	0.05	0.06	0.06	0.06	0.05	0.07	0.05	0.06	0.06	0.06	0.05
500	0.06	0.06	0.05	0.04	0.05	0.04	0.06	0.06	0.05	0.04	0.05	0.04
1000	0.06	0.07	0.06	0.05	0.04	0.05	0.06	0.07	0.06	0.05	0.04	0.05

Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma^2_\eta, \sigma^2_v) = (10, 5, 1, 1, 0)$, with $m = 2$ factors and $k = 3$ regressors. CCEP and CCEP_x denote respectively the CCEP estimator with and without $\bar{y}$ included in the matrix of CA. '*t*-test' resports the empirical size for a *t*-test at the $\alpha = 0.05$ significance level. '*basic*' reports empirical size for tests based on the basic ('empirical percentile') bootstrap interval, and bootstrap$-t$ and bootstrap$-t_c$ are respectively empirical size for the plain and corrected bootstrap$-t$ interval. All bootstrap tests are based on $B = 2000$ replications with the pairs (cross-section) resampling algorithm.

Table B-3: Empirical size: CCEP with heterogeneous slopes ($\sigma_v^2 = 5$)

	CCEP											
	t-test						*basic*					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.12	0.12	0.12	0.13	0.13	0.14	0.22	0.21	0.20	0.20	0.21	0.20
50	0.09	0.11	0.09	0.10	0.10	0.11	0.14	0.17	0.15	0.16	0.15	0.16
100	0.08	0.07	0.09	0.10	0.10	0.10	0.13	0.11	0.14	0.13	0.15	0.13
200	0.07	0.08	0.09	0.08	0.07	0.07	0.10	0.11	0.13	0.10	0.10	0.10
500	0.06	0.05	0.05	0.07	0.06	0.06	0.07	0.07	0.08	0.08	0.08	0.07
1000	0.05	0.05	0.04	0.05	0.06	0.06	0.07	0.07	0.05	0.06	0.06	0.07
	Bootstrap$-t$						Bootstrap$-t_c$					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.18	0.15	0.12	0.10	0.10	0.10	0.17	0.14	0.11	0.10	0.10	0.10
50	0.12	0.12	0.10	0.10	0.09	0.08	0.12	0.12	0.10	0.10	0.09	0.08
100	0.13	0.09	0.10	0.09	0.07	0.08	0.12	0.09	0.10	0.09	0.08	0.08
200	0.10	0.10	0.10	0.08	0.08	0.06	0.10	0.10	0.11	0.08	0.08	0.06
500	0.07	0.06	0.06	0.07	0.07	0.06	0.08	0.06	0.06	0.07	0.07	0.06
1000	0.07	0.07	0.04	0.05	0.07	0.06	0.07	0.07	0.04	0.05	0.07	0.06

	CCEP$_x$											
	t-test						*basic*					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.10	0.08	0.09	0.11	0.11	0.11	0.19	0.18	0.16	0.18	0.19	0.19
50	0.08	0.09	0.08	0.08	0.09	0.09	0.13	0.14	0.12	0.13	0.13	0.14
100	0.07	0.06	0.08	0.09	0.08	0.08	0.10	0.10	0.12	0.11	0.12	0.12
200	0.06	0.07	0.08	0.07	0.06	0.06	0.08	0.10	0.10	0.09	0.07	0.08
500	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.07
1000	0.05	0.05	0.04	0.04	0.06	0.06	0.06	0.06	0.04	0.05	0.06	0.06
	Bootstrap$-t$						Bootstrap$-t_c$					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.18	0.13	0.11	0.11	0.10	0.09	0.16	0.12	0.10	0.11	0.09	0.09
50	0.12	0.12	0.10	0.09	0.09	0.08	0.12	0.11	0.09	0.09	0.08	0.07
100	0.10	0.08	0.09	0.08	0.07	0.07	0.10	0.08	0.09	0.08	0.07	0.06
200	0.09	0.08	0.09	0.07	0.07	0.06	0.09	0.08	0.09	0.07	0.07	0.06
500	0.07	0.06	0.05	0.06	0.06	0.06	0.07	0.06	0.05	0.06	0.06	0.06
1000	0.07	0.06	0.04	0.05	0.07	0.06	0.07	0.06	0.04	0.05	0.07	0.06

Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (10, 1, 1, 1, 5)$, with $m = 2$ factors and $k = 3$ regressors. CCEP and CCEP$_x$ denote respectively the CCEP estimator with and without $\bar{y}$ included in the matrix of CA. '*t*-test' resports the empirical size for a *t*-test at the $\alpha = 0.05$ significance level. '*basic*' reports empirical size for tests based on the basic ('empirical percentile') bootstrap interval, and bootstrap$-t$ and bootstrap$-t_c$ are respectively empirical size for the plain and corrected bootstrap$-t$ interval. All bootstrap tests are based on $B = 2000$ replications with the pairs (cross-section) resampling algorithm.

Table B-4: Empirical size: CCEMG with heterogeneous slopes ($\sigma_v^2 = 5$)

	CCEMG											
	t-test						*basic*					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.09	0.09	0.10	0.09	0.09	0.11	0.12	0.12	0.13	0.13	0.13	0.13
50	0.07	0.08	0.06	0.07	0.08	0.07	0.09	0.10	0.08	0.08	0.09	0.09
100	0.07	0.07	0.06	0.07	0.07	0.06	0.08	0.08	0.08	0.09	0.08	0.08
200	0.05	0.06	0.05	0.06	0.06	0.06	0.06	0.07	0.06	0.07	0.06	0.07
500	0.05	0.06	0.05	0.06	0.06	0.05	0.06	0.07	0.06	0.06	0.07	0.05
1000	0.04	0.04	0.05	0.05	0.05	0.04	0.04	0.05	0.05	0.05	0.06	0.04
	Bootstrap$-t$						Bootstrap$-t_c$					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.06	0.05	0.06	0.06	0.05	0.05	0.06	0.05	0.06	0.06	0.05	0.05
50	0.06	0.06	0.05	0.05	0.06	0.05	0.06	0.06	0.05	0.05	0.05	0.04
100	0.05	0.06	0.05	0.06	0.05	0.05	0.05	0.06	0.05	0.06	0.05	0.05
200	0.05	0.06	0.04	0.05	0.05	0.05	0.05	0.06	0.04	0.05	0.05	0.05
500	0.05	0.06	0.05	0.05	0.05	0.04	0.05	0.06	0.05	0.05	0.05	0.04
1000	0.04	0.04	0.05	0.05	0.05	0.05	0.04	0.04	0.05	0.05	0.05	0.05

	CCEMG$_x$											
	t-test						*basic*					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.09	0.09	0.10	0.09	0.09	0.11	0.12	0.12	0.13	0.13	0.13	0.13
50	0.07	0.08	0.06	0.07	0.08	0.07	0.09	0.10	0.08	0.09	0.09	0.09
100	0.07	0.07	0.06	0.07	0.07	0.06	0.08	0.08	0.07	0.09	0.07	0.08
200	0.05	0.06	0.05	0.06	0.06	0.06	0.06	0.07	0.06	0.08	0.06	0.06
500	0.05	0.07	0.05	0.06	0.06	0.05	0.06	0.06	0.06	0.06	0.06	0.05
1000	0.04	0.04	0.05	0.05	0.05	0.04	0.04	0.05	0.05	0.05	0.06	0.05
	Bootstrap$-t$						Bootstrap$-t_c$					
(N,T)	25	50	100	200	500	1000	25	50	100	200	500	1000
25	0.06	0.05	0.06	0.07	0.05	0.05	0.06	0.05	0.06	0.06	0.05	0.05
50	0.06	0.06	0.05	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.05	0.05
100	0.05	0.06	0.05	0.06	0.05	0.05	0.05	0.06	0.04	0.06	0.05	0.05
200	0.05	0.06	0.04	0.05	0.05	0.05	0.05	0.06	0.04	0.05	0.05	0.04
500	0.05	0.06	0.05	0.05	0.05	0.04	0.05	0.06	0.05	0.05	0.05	0.04
1000	0.04	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.05

Notes: The DGP is $(d_u, \beta, \sigma^2, \sigma_\eta^2, \sigma_v^2) = (10, 1, 1, 1, 5)$, with $m = 2$ factors and $k = 3$ regressors. CCEMG and CCEMG$_x$ denote respectively the CCEMG estimator with and without $\bar{y}$ included in the matrix of CA. 't-test' resports the empirical size for a t-test at the $\alpha = 0.05$ significance level. '*basic*' reports empirical size for tests based on the basic ('empirical percentile') bootstrap interval, and bootstrap$-t$ and bootstrap$-t_c$ are respectively empirical size for the plain and corrected bootstrap$-t$ interval. All bootstrap tests are based on $B = 2000$ replications with the pairs (cross-section) resampling algorithm.