A Regularization Approach to Common Correlated Effects Estimation
Supplementary Online Appendix

Artūras Juodis*

Amsterdam School of Economics, University of Amsterdam and Tinbergen Institute

---

Contents

*Amsterdam School of Economics, University of Amsterdam, The Netherlands, Roetersstraat 11, 1018 WB, Amsterdam, The Netherlands. Email: a.juodis@uva.nl.

Preprint submitted to Journal of Applied Econometrics *October 25, 2021*

S.1. Further Discussions

S.1.1. Primitive Conditions for the Existence $\boldsymbol{F}_\perp$ Factors

Our DGP for $\boldsymbol{Q}_i = (\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i)$ directly separates all factors into two sets of factors ($\boldsymbol{F}$ and $\boldsymbol{F}_\perp$) with zero contemporaneous correlation. Alternatively, we can consider an observationally equivalent formulation that separates factors in these two groups indirectly. In particular, let the DGP be given as follows:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i, \tag{S.1}$$

$$\boldsymbol{X}_i = \boldsymbol{F}_e\boldsymbol{\Lambda}_{i,e} + \boldsymbol{V}_i, \tag{S.2}$$

$$\boldsymbol{Z}_i = \boldsymbol{F}_e\boldsymbol{C}_{i,e} + \boldsymbol{U}_i. \tag{S.3}$$

Here $\boldsymbol{F}_e$ is a $[T \times R_e]$ matrix of factors. For this formulation, an equivalent rank condition can be stated as follows

$$\text{rk}(\text{E}[\boldsymbol{C}_{i,e}]) = R_f \leq R_e, \text{ s.t. } R \leq R_f, \tag{S.4}$$

and $\text{E}[\boldsymbol{C}_{i,e}] = \boldsymbol{C}_U\boldsymbol{C}_V$, where $\boldsymbol{C}_U$ and $\boldsymbol{C}_V$ are $[R_e \times R_f]$ and $[R_f \times K_z]$ matrices of rank R_f.

Finally, we assume that $\boldsymbol{F}$ lies in the column space of $\boldsymbol{F}_f \equiv \boldsymbol{F}_e\boldsymbol{C}_U$, i.e. $\boldsymbol{F} = \boldsymbol{F}_f\boldsymbol{R}$ for some non-stochastic selection matrix $\boldsymbol{R}$ with $\text{rk}(\boldsymbol{R}) = R$. As it is argued by Juodis and Reese (2021),[1] the modified rank condition in Eq. (7) implies that all composite terms can be rewritten in terms of the $[T \times R_f]$ matrix $\boldsymbol{F}_f$ and $[T \times R_\perp]$ matrix $\boldsymbol{F}_\perp$:

$$\boldsymbol{F}\boldsymbol{\lambda}_i = \boldsymbol{F}_f\widetilde{\boldsymbol{\lambda}}_i, \quad \boldsymbol{F}_e\boldsymbol{\Lambda}_{i,e} = \boldsymbol{F}_f\widetilde{\boldsymbol{\Lambda}}_i + \boldsymbol{F}_\perp\widetilde{\boldsymbol{\Lambda}}_{i,\perp}, \quad \boldsymbol{F}_e\boldsymbol{C}_{i,e} = \boldsymbol{F}_f\widetilde{\boldsymbol{C}}_i + \boldsymbol{F}_\perp\widetilde{\boldsymbol{C}}_{i,\perp}, \tag{S.5}$$

where $R_\perp = R_f - R_e$. This is exactly our formulation for the DGP of $\boldsymbol{Q}_i$ upon redefining the corresponding factor loadings, and appropriately augmenting the loadings vector $\boldsymbol{\lambda}_i$ if $R_f > R$.

S.1.2. Technical Notes

Note 1 (**Normalization**)**.** In Step 4 one can either use $\widehat{\boldsymbol{F}}$ or $\overline{\boldsymbol{Z}}$ as the input to construct $\widehat{\boldsymbol{F}}_r$. The advantage of the latter option is that the columns of $\boldsymbol{Z}_i$ that are not informative about $\boldsymbol{F}$ will not contribute to the variance (or bias) of the estimator. Thus, automatically one also does model selection in terms of the relevant variables for $\boldsymbol{F}$. However, in this case the resulting estimator is not invariant (unlike the original CCE) to non-singular transformations of the columns of $\boldsymbol{Z}_i$. On the other hand, if one uses the normalized version in Step 4, then the resulting estimator is invariant to such transformation, but its variance (or bias) might be affected by the non-informative columns of $\boldsymbol{Z}_i$.

[1] See Lemma S.18 in the Supplementary Appendix of that paper.

Note 2 (**Non-stationary Factors**). Given that normalization is used to construct $\widehat{\boldsymbol{F}}$, under certain regularity conditions our results can be extended towards the non-stationary setups of Kapetanios et al. (2011) and Westerlund (2018). However, under all setups it is crucial to assume $\boldsymbol{F}_{\perp}$ contains factors that are at most weakly integrated. This ensures that the asymptotic distribution of the rCCEP estimator is (mixed-)normal. If some of the unproxied factors in $\boldsymbol{F}_{\perp}$ are $I(1)$, then the asymptotic distribution is no longer normal, as it is completely dominated by the non-stationary component in $\boldsymbol{F}_{\perp}$. For T fixed $T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}$ has a mixed-normal limit, irrespective of the time-series properties of $\boldsymbol{F}$ and $\boldsymbol{F}_{\perp}$. Hence, the setup with non-stationary factors in large T setup is qualitatively similar to fixed T setup with factors of any form.

Note 3 (**Non-linear DGP for $\boldsymbol{Z}_i$**). While in this paper we maintain the assumption that $\boldsymbol{Z}_i$ has an exact factor structure, the fact that we allow $\boldsymbol{F}_{\perp}$ to be in the model is also informative about some model where this is not the case. For example, assume that some of the elements in $\boldsymbol{z}_{i,t}$ are non-linear in factors, e.g. $z_{i,t}^{(\cdot)} = g(\boldsymbol{c}_i, \boldsymbol{d}_t, \nu_{i,t})$. In this case, one can still decompose $z_{i,t}^{(\cdot)} = q_t + u_{i,t}$, where for every pair (i, j) $u_{i,t}$ and $u_{j,t}$ are independent conditional on $\boldsymbol{F}_e$ and have mean zero. Provided that at least some of elements in $\boldsymbol{Z}_i$ have an exact factor structure in $\boldsymbol{F}$, the presence of $u_{i,t}$ (unconditionally correlated over i) should not qualitatively change our conclusions. For example, such DGP was considered in the context of non-parametric regression in Su and Jin (2012), where they assume that $y_{i,t} = g(\boldsymbol{x}_{i,t}) + \boldsymbol{\lambda}_i'\boldsymbol{f}_t + \varepsilon_{i,t}$. The regressor vector $\boldsymbol{x}_{i,t}$ is assumed to be of Eq. (3) form, with $\boldsymbol{y}_i$ included in $\boldsymbol{Z}_i$. Notice that in the aforementioned study the relevance of cross-sectional average of the non-parametric component is not emphasized. Furthermore, their proof strategy ignores potential issues later documented by Karabıyık et al. (2017) when $K_z > R$. Another example is the recent study of De Vos and Westerlund (2019), where they allow some of the regressors to be unrestricted in terms of the error-component structure. Unfortunately, De Vos and Westerlund (2019) do not prove any distributional properties of the CCEP estimator in this case.

Note 4 (**Two-step Estimators**). It is of special interest to compare the CCE estimator with a standard parametric two-step estimator with estimated nuisance parameters in the first step. In the standard textbook setup the number of nuisance parameters is fixed, and the leading term is well behaved, i.e. mixing/martingale difference, or even iid over units/time. However, for the CCEP estimator the situation is substantially different as the number of pre-estimated nuisance parameters (factors) is of order $\mathcal{O}(T)$, and the leading term characterized by $\text{vec}(\boldsymbol{F}_{\perp}\boldsymbol{C}_{i,\perp})$ is correlated with the score of the second step estimator, in the way that any possible mixing and/or martingale difference conditions that are essential for $\sqrt{NT}-$consistency are no longer satisfied.

Note 5 (**Assumption (Errors)**). Assumption 1(c) is somewhat non-standard to the CCE literature, as we assume that the autocovariance matrices for different units i are the same up to a proportionality factor $\boldsymbol{\Omega}_i$. We use this particular formulation as we further assume that $\boldsymbol{\Omega}_i$ (and thus $\boldsymbol{\Gamma}_{i,\boldsymbol{e}}(h)$) can be treated as iid random variable over i. We believe that this formulation is more natural in the random-coefficients setup, as both the factor loadings, as well as the slope coefficients are already assumed to be random. Thus, it is more appropriate also to treat scale parameters as random. In contrast, Pesaran (2006) in the heterogeneous coefficients setup assumes that slope coefficients are iid random variables, while $\boldsymbol{\Gamma}_{i,\boldsymbol{e}}(h)$ is a sequence of constants, see Assumptions 2-3 of the aforementioned paper.[2]

Note 6 (Unit-by-unit Estimators). Regularization ensures that $\boldsymbol{\Sigma}_{X,r,i}$ is non-stochastic (conditional on individual level heterogeneity) in the limit, so that the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{rCCE,i}$ is asymptotically normal for each i. However, despite normality, the distributional properties of $\boldsymbol{\beta}_{rCCE,i}$ differ from any standard textbook example for the following two reasons:

1. For each finite index set $(i_1, \ldots, i_J)$, the joint asymptotic distribution of $(\widehat{\boldsymbol{\beta}}'_{rCCE,i_1}, \ldots, \widehat{\boldsymbol{\beta}}'_{rCCE,i_J})'$ is multivariate normal. However, the covariance matrix is not block-diagonal, as all estimators are correlated through $\sum_{i=1}^{N} \text{vec}(\boldsymbol{C}_{i,\perp})$ in Eq. (33) This result is somewhat similar to the results for the heterogeneous QML estimator, where for certain DGPs of $\boldsymbol{X}_i$ the unit specific estimates are also correlated over the cross-section dimension, see Song (2013) and Remark 4 in Ando and Bai (2015).

2. The asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{rCCE,i}$ is characterized by the sum of the two components: the time-series component $\boldsymbol{b}_{0,i}$ and the cross-section component $\boldsymbol{\xi}_{r,i}$. Standard methods (e.g residual based, block bootstrap, cross-section bootstrap) to estimate the variance-covariance matrix will fail to appropriately account for the two distinct sources of variation.

S.1.3. Non-normal Limiting Distribution: Intuition

In this section, we briefly explain under which conditions the asymptotic distributions of the rCCEP is non-normal. In particular, as with the standard CCEP estimator, the term $\boldsymbol{b}_{2,r}$ contains a term linear in $\overline{\boldsymbol{Z}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}_r}\overline{\boldsymbol{Z}}$, thus $\boldsymbol{b}_{2,r}$ can be stochastic in the limit. To see this, observe that $\overline{\boldsymbol{Z}}$ is a linear function of zero-mean factor loadings $\boldsymbol{\Lambda}_{i,\perp}$, so that

$$\frac{N}{T}\overline{\boldsymbol{Z}}'\boldsymbol{M}_{\widehat{\boldsymbol{F}}_r}\overline{\boldsymbol{Z}} \tag{S.6}$$

converges in distribution to a quadratic form of correlated random normal matrices (plus a constant matrix). Thus, no available bias-correction procedures will be effective in mitigating distributional effects of this term (this is not an issue if $\boldsymbol{\xi}_r = \mathcal{O}_P(1)$, as this term will dominate

[2] Juodis and Reese (2021) provide an example where such inconsistency in terms of the assumptions can lead to an erroneous conclusion in terms of the asymptotic properties of the cross-section dependence testing procedure.

the asymptotic distribution). Furthermore, if all elements in $\boldsymbol{\Lambda}_{\perp,i}$ and $\boldsymbol{\lambda}_i$ are uncorrelated, so that instead

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\lambda}_i \otimes \boldsymbol{\Lambda}_{i,\perp} = \mathcal{O}_P(N^{-1/2}), \tag{S.7}$$

and

$$\sqrt{\frac{T}{N}}\boldsymbol{\xi}_r = \mathcal{O}_P(1). \tag{S.8}$$

Hence, similarly to $\boldsymbol{b}_{2,r}$, $\boldsymbol{\xi}_r$ essentially is another "stochastic bias" component, that vanishes only if $T/N \to 0$ (that we exclude by assumption). As a result, none of the standard bias-correction procedures (analytical or jackknife) will be able to remove it completely from the asymptotic distribution. However, because of regularization, $\boldsymbol{b}_{2,r}$ and $\boldsymbol{\xi}_r$ are at most a second degree polynomial in cross-sectional averages of factor loadings (and not non-linear functions as in CCEP estimator). This in turn is sufficient that after analytical (or half-panel jackknife) bias-correction, the resulting bias term will remain stochastic, but have zero expected value. We elaborate on this more formally in S.2.1.

Non-normal properties of $\boldsymbol{b}_{2,r}$ and $\boldsymbol{\xi}_r$ in the intermediate case Eq. (S.7) further complicate any attempts to derive uniformly valid inference procedures. To be more specific, notice that for the validity of any bootstrap procedure it is necessary that the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{rCCEP} - \boldsymbol{\beta}_0$ can be well approximated by that of:

$$\widetilde{\boldsymbol{\beta}}^*_{rCCEP} - \widehat{\boldsymbol{\beta}}_{rCCEP}. \tag{S.9}$$

For this to be true, it is essential that all the bias-terms in $\boldsymbol{b}_1$ and $\boldsymbol{b}_{2,r}$ can be accounted for, i.e. removed using bias-correction. However, this is not the case due to the stochastic part of $\boldsymbol{b}_{2,r}$, i.e.

$$N\boldsymbol{G}'_{\perp}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{\perp}}\boldsymbol{G}_{\perp}. \tag{S.10}$$

For bootstrap inference to be valid, the distribution of this random variable should be well approximated by that of

$$N(\boldsymbol{G}^*_{\perp})'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{\perp}}\boldsymbol{G}^*_{\perp} - N\boldsymbol{G}'_{\perp}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{\perp}}\boldsymbol{G}_{\perp}, \tag{S.11}$$

where $\boldsymbol{G}^*_{\perp}$ is the bootstrap analogue of $\boldsymbol{G}_{\perp}$. This is clearly infeasible, as usual cross-sectional bootstrap procedures are designed for asymptotically linear estimators (first degree cross-sectional U-statistics), and not for the quadratic forms we encounter in this example (a second degree U-statistic). Unfortunately, in this case it is also unclear under which circumstances the bootstrap procedure can be conservative, which might a desirable property for the procedure to be uniformly valid.

S.1.4. Non-normal Limiting Distribution: Monte Carlo Evidence

In this section, we numerically illustrate the performance of the suggested bootstrap procedure in the setup discussed above. In particular, we consider the same setup as in the main text, except that the factor loadings are generated as follows:

$$\begin{pmatrix} \lambda_i \\ \gamma_{i,\perp} \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ \gamma_{\perp} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right), \tag{S.12}$$

where in all designs we set $\gamma_{\perp} = 0$. This means that the rCCEP estimator is $\sqrt{NT}-$consistent but the asymptotic distribution has several additive non-normal components originated from $\boldsymbol{b}_{2,r}$ and $\boldsymbol{\xi}_r$. The Monte Carlo results for all pooled and MG estimators are summarized in Table S.1 and Table S.2, respectively.

Results. From the results in Table S.1 we can conclude that for larger ratios of T/N bootstrap inference for CCEP is largely oversized, while for the rCCEP and rCCEP-BC estimators size is much better controlled. We can also observe that bias-correction plays almost no role in this setup (despite the $\sqrt{NT}-$convergence rate). This conclusion coincides with that drawn for the equivalent setup in the main text.

In the heterogeneous coefficients setup both the CCE-MG and the rCCE-MG estimators are normally distributed asymptotically, the rCCE-MG estimator has a much smaller finite sample bias. This is especially true for $N = 20$, where finite sample bias makes the CCE-MG based inference substantially oversized. Thus, despite the fact that for N, T large $\boldsymbol{\xi}$ and $\boldsymbol{\xi}_r$ are asymptotically negligible under our assumptions, they still matter unless N is large.

Table S.1: Estimation and inference. Homogeneous setup.

Design			CCE					$rCCE$					$rCCE-BC$				
$\gamma_\perp$	N	T	Bias	RMSE	Power−	Size	Power+	Bias	RMSE	Power−	Size	Power+	Bias	RMSE	Power−	Size	Power+
0	20	20	.10	.23	.87	.14	.10	.03	.20	.64	.06	.32	.01	.19	.66	.05	.32
0	20	50	.10	.17	1	.23	.14	.02	.14	.86	.05	.45	.01	.12	.89	.03	.48
0	20	100	.09	.15	1	.33	.18	.01	.12	.92	.05	.51	.01	.09	.95	.02	.54
0	50	20	.06	.20	.99	.09	.69	.00	.17	.97	.06	.90	.00	.17	.98	.05	.93
0	50	50	.06	.13	1	.17	.89	.00	.11	1	.05	.97	.00	.11	1	.04	.98
0	50	100	.05	.10	1	.23	.94	.00	.08	1	.03	.98	.00	.08	1	.03	.99
0	100	20	.04	.19	1	.06	.99	.00	.17	1	.06	1	.00	.17	1	.03	1
0	100	50	.04	.12	1	.10	1	.00	.10	1	.04	1	.00	.10	1	.04	1
0	100	100	.04	.09	1	.16	1	.00	.07	1	.04	1	.00	.07	1	.04	1

Notes. Here "Bias" is the average of any scaled estimator over $M = 4000$ replications. "RMSE" is the corresponding Root Mean Squared Error of any scaled estimator. CCE is the pooled CCE estimator; $rCCE$ is the pooled regularized CCE estimator; $rCCE-BC$ is the $rCCE$ estimator with analytical bias correction as in the Appendix. "Power−" and "Power+" are respectively the rejection frequencies for the null hypothesis of $\beta_0 = -0.1$ and $\beta_0 = 0.1$. "Size" corresponds to the null hypothesis $\beta_0 = 0$.

Table S.2: Estimation and inference. Heterogeneous setup.

Design			$CCE-MG$					$rCCE-MG$				
$\gamma_\perp$	N	T	Bias	RMSE	Power−	Size	Power+	Bias	RMSE	Power−	Size	Power+
0	20	20	.10	.32	.56	.10	.07	.03	.30	.35	.06	.20
0	20	50	.10	.27	.76	.16	.09	.03	.25	.43	.06	.27
0	20	100	.10	.25	.82	.17	.11	.02	.24	.46	.06	.31
0	50	20	.06	.30	.81	.08	.39	.01	.28	.71	.06	.62
0	50	50	.06	.24	.95	.10	.54	.01	.23	.85	.06	.76
0	50	100	.06	.23	.97	.12	.62	.01	.22	.89	.06	.82
0	100	20	.04	.29	.96	.06	.83	.00	.28	.94	.06	.92
0	100	50	.04	.24	1	.08	.95	.00	.23	.99	.06	.98
0	100	100	.04	.22	1	.08	.97	.00	.22	1	.06	.99

Notes. $CCE-MG$ is the mean-group CCE estimator; $rCCE-MG$ is the mean-group regularized CCE estimator. See Table S.1 for more details.

S.1.5. The Need for Cross-sectional Bootstrap

In what follows we provide an intuitive explanation on why the standard variance-covariance estimators are inconsistent when $\boldsymbol{F}_{\perp}$ factors are present. Consider the heterogeneous coefficients setup with individual specific estimates $\widehat{\boldsymbol{\beta}}_{rCCE,i}$. Further, assume that we can actually directly observe the leading term of this estimator:

$$\begin{aligned}\boldsymbol{\phi}_i &\equiv \boldsymbol{\beta}_0 + \widetilde{\boldsymbol{\beta}}_i + N^{-1/2}\boldsymbol{\Sigma}_{X,r,i}^{-1}\boldsymbol{\xi}_{r,i} \\ &= \boldsymbol{\beta}_0 + \widetilde{\boldsymbol{\beta}}_i + \boldsymbol{\Sigma}_{X,r,i}^{-1}\left(\boldsymbol{\lambda}_i' \otimes \boldsymbol{\Lambda}_{i,\perp}'\right)\left(\boldsymbol{S}' \otimes \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{\perp}}\right)\frac{1}{N}\sum_{i=1}^{N}\operatorname{vec}\left(\boldsymbol{C}_{i,\perp}\right). \end{aligned} \tag{S.13}$$

The (vectorized) sample variance-covariance estimator can be expanded as follows:

$$\begin{aligned}\operatorname{vec}(\widehat{\boldsymbol{\Sigma}}_{\phi}) &= \frac{N}{N-1}\operatorname{vec}((\overline{\boldsymbol{\phi}}-\boldsymbol{\beta}_0)(\overline{\boldsymbol{\phi}}-\boldsymbol{\beta}_0)') \\ &+ \frac{1}{N-1}\sum_{i=1}^{N}\operatorname{vec}(\widetilde{\boldsymbol{\beta}}_i\widetilde{\boldsymbol{\beta}}_i') \\ &+ \frac{1}{N-1}\sum_{i=1}^{N}\left(\widetilde{\boldsymbol{\beta}}_i \otimes \boldsymbol{\Sigma}_{X,r,i}^{-1}\left(\boldsymbol{\lambda}_i' \otimes \boldsymbol{\Lambda}_{i,\perp}'\right)\right)\operatorname{vec}\left(\left(\boldsymbol{S}' \otimes \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{\perp}}\right)\frac{1}{N}\sum_{i=1}^{N}\operatorname{vec}\left(\boldsymbol{C}_{i,\perp}\right)\right) \\ &+ \frac{1}{N-1}\sum_{i=1}^{N}((\boldsymbol{\Sigma}_{X,r,i}^{-1}\left(\boldsymbol{\lambda}_i' \otimes \boldsymbol{\Lambda}_{i,\perp}'\right)) \otimes \widetilde{\boldsymbol{\beta}}_i)\operatorname{vec}\left(\frac{1}{N}\sum_{i=1}^{N}\operatorname{vec}\left(\boldsymbol{C}_{i,\perp}\right)'\left(\boldsymbol{S} \otimes \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{\perp}}'\right)\right) \\ &+ \frac{1}{N-1}\sum_{i=1}^{N}((\boldsymbol{\Sigma}_{X,r,i}^{-1}\left(\boldsymbol{\lambda}_i' \otimes \boldsymbol{\Lambda}_{i,\perp}'\right)) \otimes (\boldsymbol{\Sigma}_{X,r,i}^{-1}\left(\boldsymbol{\lambda}_i' \otimes \boldsymbol{\Lambda}_{i,\perp}'\right))) && \text{(S.14)} \\ &\times \operatorname{vec}\left(\left(\boldsymbol{S}' \otimes \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{\perp}}\right)\frac{1}{N}\sum_{i=1}^{N}\operatorname{vec}\left(\boldsymbol{C}_{i,\perp}\right)\frac{1}{N}\sum_{i=1}^{N}\operatorname{vec}\left(\boldsymbol{C}_{i,\perp}\right)'\left(\boldsymbol{S} \otimes \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{\perp}}'\right)\right). && \text{(S.15)}\end{aligned}$$

Given that all i specific variables are bounded by assumption, it is easy to see that:

$$\operatorname{vec}(\widehat{\boldsymbol{\Sigma}}_{\phi}) = \frac{1}{N-1}\sum_{i=1}^{N}\operatorname{vec}(\widetilde{\boldsymbol{\beta}}_i\widetilde{\boldsymbol{\beta}}_i') + \mathcal{O}_P(N^{-1/2}). \tag{S.16}$$

S.2. Additional Theoretical Results

S.2.1. Analytical Bias-correction Results

Here we briefly discuss how to construct an estimate of $\boldsymbol{b}_{2,r}$, using analytical bias-correction (plug-in principle). The idea of analytical bias-correction for the CCEP estimator was first proposed in Westerlund and Urbain (2015). This section largely follows their methodology of that paper, and extends it appropriately for the purpose of the rCCEP estimator.

Ideally we would like to construct an estimator $\widehat{\boldsymbol{b}}_{2,r}$ such that:

$$\widehat{\boldsymbol{b}}_{2,r} = \boldsymbol{b}_{2,r} + \mathrm{o}_P(1). \tag{S.17}$$

From Theorem 1 this bias term consists of three components

$$\boldsymbol{b}_{2,r} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{G}^{+})'\widehat{\boldsymbol{\Sigma}}^{-1/2}\left(N\boldsymbol{G}_{\perp}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{\perp}}\boldsymbol{G}_{\perp} + \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_i[\boldsymbol{u}_{i,t}\boldsymbol{u}_{i,t}']\right)(\widehat{\boldsymbol{\Sigma}}^{-1/2})'\boldsymbol{G}^{+}\boldsymbol{\lambda}_i$$
$$- \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{G}^{+})'\widehat{\boldsymbol{\Sigma}}^{-1/2}\,\mathrm{E}_i[\boldsymbol{u}_{i,t}\varepsilon_{i,t}] - \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_i[\boldsymbol{v}_{i,t}\boldsymbol{u}_{i,t}'](\widehat{\boldsymbol{\Sigma}}^{-1/2})'\boldsymbol{G}^{+}\boldsymbol{\lambda}_i. \tag{S.18}$$

Here $\boldsymbol{G}^{+} = \widehat{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{G}'(\boldsymbol{G}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{G}')^{-1}$ is the Moore-Penrose generalized inverse of $\boldsymbol{G}(\widehat{\boldsymbol{\Sigma}}^{-1/2})'$. The first component consists of a component $\boldsymbol{G}_{\perp}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{\perp}}\boldsymbol{G}_{\perp}$ which is a random variable even in the limit $N \to \infty$. Thus, the requirement in Eq. (S.17), cannot be satisfied for this simple reason. Instead, we can (at best) only consistently estimate all terms that involve elements of $\mathrm{E}_i[\boldsymbol{e}_{i,t}\boldsymbol{e}_{i,t}']$. In particular, for all $i = 1, \ldots, N$ consider the following plug-in estimators:

$$\widehat{\boldsymbol{V}}_i = \boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\boldsymbol{X}_i, \tag{S.19}$$
$$\widehat{\boldsymbol{U}}_i = \boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\boldsymbol{Z}_i, \tag{S.20}$$
$$\widehat{\boldsymbol{\varepsilon}}_i = \boldsymbol{M}_{\widetilde{\boldsymbol{F}}}(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_{rCCEP}), \tag{S.21}$$
$$\widehat{\boldsymbol{\Lambda}}_i = (\widetilde{\boldsymbol{F}}'\widetilde{\boldsymbol{F}})^{-1}\widetilde{\boldsymbol{F}}'\boldsymbol{X}_i, \tag{S.22}$$
$$\widehat{\boldsymbol{\lambda}}_i = (\widetilde{\boldsymbol{F}}'\widetilde{\boldsymbol{F}})^{-1}\widetilde{\boldsymbol{F}}'(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_{rCCEP}). \tag{S.23}$$

Clearly $\widehat{\boldsymbol{\lambda}}_i$ and $\widehat{\boldsymbol{\Lambda}}_i$ are not consistent for their population counterparts because of the rotational indeterminacy. However, this is not problematic, as we are only interested in their rotated versions, i.e. $\boldsymbol{G}^{+}\boldsymbol{\lambda}_i$ and $\boldsymbol{G}^{+}\boldsymbol{\Lambda}_i'$. For this purpose, we propose the following plug-in estimator $\widehat{\boldsymbol{G}}^{+} = (T^{-1}\widehat{\boldsymbol{F}}'\widetilde{\boldsymbol{F}})\boldsymbol{V}_{N,T}^{-1}$. It is not that difficult to show that:

$$\widehat{\boldsymbol{G}}^{+}\widehat{\boldsymbol{\lambda}}_i = \boldsymbol{G}^{+}\boldsymbol{\lambda}_i + \mathrm{o}_P(1), \tag{S.24}$$

and similarly for $\widehat{\boldsymbol{G}}^{+}\widehat{\boldsymbol{\Lambda}}_i$. Combining all the ingredients we propose the following plug in estimate of $\boldsymbol{b}_{2,r}$:

$$\widehat{\boldsymbol{b}}_{2,r} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\Lambda}}_i'(\widehat{\boldsymbol{G}}^{+})'\left(\frac{1}{NT}\sum_{i=1}^{N}\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{U}}_i'\widehat{\boldsymbol{U}}_i(\widehat{\boldsymbol{\Sigma}}^{-1/2})'\right)\widehat{\boldsymbol{G}}^{+}\widehat{\boldsymbol{\lambda}}_i$$
$$- \frac{1}{NT}\sum_{i=1}^{N}\widehat{\boldsymbol{\Lambda}}_i'(\widehat{\boldsymbol{G}}^{+})'\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{U}}_i'\widehat{\boldsymbol{\varepsilon}}_i - \frac{1}{NT}\sum_{i=1}^{N}\widehat{\boldsymbol{V}}_i'\widehat{\boldsymbol{U}}_i(\widehat{\boldsymbol{\Sigma}}^{-1/2})'\widehat{\boldsymbol{G}}^{+}\widehat{\boldsymbol{\lambda}}_i. \tag{S.25}$$

The next result characterizes the asymptotic properties of the plug-in bias-correction procedure:

Proposition S.1. *Under Assumptions 4.1-4.5:*

$$\boldsymbol{b}_{2,r} - \widehat{\boldsymbol{b}}_{2,r} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{G}^{+})'\widehat{\boldsymbol{\Sigma}}^{-1/2}\left(N\boldsymbol{G}_{\perp}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{\perp}}\boldsymbol{G}_{\perp} - \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{C}_{i,\perp}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{\perp}}\boldsymbol{C}_{i,\perp}\right)(\widehat{\boldsymbol{\Sigma}}^{-1/2})'\boldsymbol{G}^{+}\boldsymbol{\lambda}_i$$
$$+ \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Lambda}_{i,\perp}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{\perp}}\boldsymbol{C}_{i,\perp}(\widehat{\boldsymbol{\Sigma}}^{-1/2})'\boldsymbol{G}^{+}\boldsymbol{\lambda}_i + \mathrm{o}_P(1). \tag{S.26}$$

Here we briefly discuss the implications of this proposition. First of all, as long as factor loadings are sufficiently correlated *and* $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}$ is of full rank asymptotically (Assumption 4.2), then $\boldsymbol{b}_{2,r} - \widehat{\boldsymbol{b}}_{2,r} = \mathcal{O}_P(1)$, i.e. bias-correction is not fully effective.

Secondly, note that from $\boldsymbol{G}_\perp = N^{-1}\sum_{i=1}^N \boldsymbol{C}_{i,\perp}$, it follows that:

$$\mathrm{E}\left[\left(N\boldsymbol{G}'_\perp \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{G}_\perp - \frac{1}{N}\sum_{i=1}^N \boldsymbol{C}'_{i,\perp}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{C}_{i,\perp}\right)|\boldsymbol{F}_\perp\right] = \mathbf{O}. \tag{S.27}$$

Thus, the expression in the first line of Proposition S.1 converges in distribution to a zero-mean random variable. In that respect, while not completely removing the incidental parameter bias, the analytical bias-correction centres the overall bias term at zero. Moreover, the second line in Proposition S.1 also centres the $\boldsymbol{\xi}_r$ component, as it ensures that in DGPs with $\boldsymbol{\xi}_r = \mathcal{O}_P(N^{-1/2})$ the total stochastic "bias":

$$\begin{aligned}
\frac{1}{\sqrt{N}}\boldsymbol{\xi}_r + \boldsymbol{b}_{2,r} - \widehat{\boldsymbol{b}}_{2,r} &= \frac{1}{N}\sum_{i=1}^N \boldsymbol{\Lambda}'_i(\boldsymbol{G}^+)'\widehat{\boldsymbol{\Sigma}}^{-1/2}\left(N\boldsymbol{G}'_\perp\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{G}_\perp - \frac{1}{N}\sum_{i=1}^N \boldsymbol{C}'_{i,\perp}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{C}_{i,\perp}\right)(\widehat{\boldsymbol{\Sigma}}^{-1/2})'\boldsymbol{G}^+\boldsymbol{\lambda}_i \\
&- \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \boldsymbol{\Lambda}'_{i,\perp}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{C}_{j,\perp}(\widehat{\boldsymbol{\Sigma}}^{-1/2})'\boldsymbol{G}^+\boldsymbol{\lambda}_i + \frac{1}{N^2}\sum_{i=1}^N \boldsymbol{\Lambda}'_{i,\perp}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{C}_{i,\perp}(\widehat{\boldsymbol{\Sigma}}^{-1/2})'\boldsymbol{G}^+\boldsymbol{\lambda}_i + \mathrm{o}_P(1) \\
&= \frac{1}{N}\sum_{i=1}^N \boldsymbol{\Lambda}'_i(\boldsymbol{G}^+)'\widehat{\boldsymbol{\Sigma}}^{-1/2}\left(\frac{1}{N}\sum_{i=1}^N\sum_{j\neq i}^N \boldsymbol{C}'_{i,\perp}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{C}_{j,\perp}\right)(\widehat{\boldsymbol{\Sigma}}^{-1/2})'\boldsymbol{G}^+\boldsymbol{\lambda}_i \\
&- \frac{1}{N}\sum_{i=1}^N\sum_{j\neq i}^N \boldsymbol{\Lambda}'_{i,\perp}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{C}_{j,\perp}(\widehat{\boldsymbol{\Sigma}}^{-1/2})'\boldsymbol{G}^+\boldsymbol{\lambda}_i + \mathrm{o}_P(1),
\end{aligned} \tag{S.28}$$

converges in distribution to a zero-mean random vector. Notice that the asymptotic limit of this expression is not normal. Instead, the second-degree U-statistic in

$$\left(\frac{1}{N}\sum_{i=1}^N\sum_{j\neq i}^N \boldsymbol{C}'_{i,\perp}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{C}_{j,\perp}\right), \tag{S.29}$$

has a shifted Wishart-type asymptotic distribution.

S.2.2. Unit-specific rCCE Estimator

The next theorem characterizes the large sample distribution of the unit-by-unit regularized CCE estimator $\widehat{\boldsymbol{\beta}}_{rCCE,i}$.

Theorem S.1. *Under Assumptions 4.1-4.5 for all i with $\ell \geq 4$:*

$$\sqrt{T}(\widehat{\boldsymbol{\beta}}_{rCCE,i} - \boldsymbol{\beta}_i) = \boldsymbol{\Sigma}^{-1}_{X,r,i}\left(\boldsymbol{b}_{0,i} + \sqrt{\frac{T}{N}}\boldsymbol{\xi}_{r,i}\right) + \mathrm{o}_P(1), \tag{S.30}$$

where:

$$\boldsymbol{\Sigma}_{X,r,i} = \mathrm{E}_i[\boldsymbol{v}_{i,t}\boldsymbol{v}_{i,t}'] + \boldsymbol{\Lambda}_{i,\perp}'\boldsymbol{\Gamma}_{\perp,\perp}(0)\boldsymbol{\Lambda}_{i,\perp}, \tag{S.31}$$

$$\boldsymbol{b}_{0,i} = \frac{1}{\sqrt{T}}\boldsymbol{V}_{i,+}'\boldsymbol{\varepsilon}_i, \tag{S.32}$$

$$\boldsymbol{\xi}_{r,i} = \left(\boldsymbol{\lambda}_i' \otimes \boldsymbol{\Lambda}_{i,\perp}'\right)\left(\boldsymbol{S}' \otimes \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathrm{vec}\left(\boldsymbol{C}_{i,\perp}\right). \tag{S.33}$$

Here $\boldsymbol{S}$ is defined as in Theorem 1.

Most of the expressions in Theorem S.1 are simplified versions of the corresponding expressions in Theorem 1. Among other things, the reduced convergence rate $\sqrt{T}$ directly implies that the analogues of the $\boldsymbol{b}_1$ and $\boldsymbol{b}_{2,r}$ bias terms are always asymptotically negligible for unit-by-unit estimates.

For any fixed i, the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{rCCE,i}$ is expected to be multivariate mixed-normal, with the variance-covariance matrix that can be expressed as a sum of the two variance-covariance matrices generated by $\boldsymbol{b}_{0,i}$ and $\boldsymbol{\xi}_{r,i}$, respectively, as by Assumptions 4.1 and 4.3 these two components are asymptotically uncorrelated (thus also independent by joint normality). Unfortunately all standard (time-series or cross-section) CLTs, are not suitable to study the joint asymptotic distribution of $(\boldsymbol{b}_{0,i}', \boldsymbol{\xi}_{r,i}')'$. Instead, we conjecture that following Hahn et al. (2020), one has to construct an appropriate increasing sequence of sigma-fields in both dimensions to make use of the general CLT for MDS arrays from Hall and Heyde (1980).

S.2.3. Fixed T Inference

In this section we summarize the sufficient conditions for the fixed T setup. In what follows, we denote the stacked matrix of idiosyncratic components by $\boldsymbol{E}_i = (\boldsymbol{\varepsilon}_i, \boldsymbol{V}_i, \boldsymbol{U}_i)$.

Assumption S.2.1 (Errors (Fixed T)). *(a) $\boldsymbol{E}_i$ and $\boldsymbol{E}_j$ are independent for all $i \neq j$; (b) $\mathrm{E}_i[\boldsymbol{E}_i] = \boldsymbol{0}$ and $\mathrm{E}_i[\|\boldsymbol{e}_{i,t}\|^{4+\delta}] < \infty$; (c) (i) $\boldsymbol{E}_i$ admits factorization $\boldsymbol{E}_i = \widetilde{\boldsymbol{E}}_i\boldsymbol{\Omega}_i'$; (ii) $\mathrm{E}[\widetilde{\boldsymbol{E}}_i'\widetilde{\boldsymbol{E}}_i]$ are identical for all i; (d) $\boldsymbol{\Gamma}_{i,\varepsilon,\boldsymbol{v}}(h) = \boldsymbol{0}_K'$ for h; (e) (i) $\mathrm{rk}[\mathrm{E}[\boldsymbol{V}_i'\boldsymbol{V}_i]] = K$.; (ii) $\mathrm{rk}[\mathrm{E}[\boldsymbol{U}_i'\boldsymbol{U}_i]] = K_z$;(iii) $\sqrt{N}\overline{\boldsymbol{U}} \xrightarrow{d} \boldsymbol{\Xi}_T$ where $\mathrm{rk}[\boldsymbol{\Xi}_T'\boldsymbol{\Xi}_T] \geq K_z - R$ almost surely.*

Assumption S.2.2 (Factors (Fixed T)). $\mathrm{rk}[\boldsymbol{F}'\boldsymbol{F}] = R$ *almost surely.*

Assumption S.2.3 (Unit Heterogeneity (Fixed T)). *(a) The random vector $\boldsymbol{h}_i$ is independent and identically distributed over i with $\mathrm{E}[\|\boldsymbol{h}_i\|^{4+\delta}] < \infty$; (b) $\mathrm{E}[\boldsymbol{C}_i] = \boldsymbol{C}$ such that $\mathrm{rk}(\boldsymbol{C}) = R$; (c) $\boldsymbol{h}_i$ and $\boldsymbol{d}_t$ are mutually independent for all i and t; (d) $\boldsymbol{h}_i$ and $\widetilde{\boldsymbol{E}}_j$ are mutually independent for all i, j.*

Assumption S.2.4 (Eigenvalues). *The eigenvalues of the $[R \times R]$ matrix $\boldsymbol{C}\,\mathrm{E}[\widehat{\boldsymbol{\Sigma}}]^{-1}\boldsymbol{C}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}}$ are distinct almost surely.*

In comparison to the assumptions discussed in the main text, we notice that for T fixed we leave the time-series properties of the underlying time-varying stochastic variables almost completely unspecified. In particular, both $\boldsymbol{E}_i$ and $\boldsymbol{F}$ ($\boldsymbol{F}_\perp$) can be non-stationary random processes. The most visible non-standard difference is part $(e)(iii)$ of Assumption S.2.1, where we require that the $[T \times K_z]$ limiting matrix $\boldsymbol{\Xi}_T$ satisfies a certain rank restriction. This is high-level condition is sufficient to establish the convergence rate for zero eigenvalues of the $\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}$ matrix. In the large T setup this condition is conveniently implied by part $(e)(ii)$ of the corresponding assumption $\mathrm{rk}[\mathrm{E}[\boldsymbol{U}_i'\boldsymbol{U}_i]] = K_z$.

S.3. Additional Monte Carlo Study: rCCE vs. IFE

In this section, we compare the performance of the bias-corrected rCCEP against the bias-corrected Interactive Fixed Effects estimator of Bai (2009). We focus on three distinct MC designs to highlight some of the problems of every procedure:

1. (Default) The setup of the main paper with $\gamma_\perp = 0$.
2. (Weak identification) The setup of the main paper with $\gamma_\perp = 0.2$.
3. (Heteroscedasticity) The setup of the main paper with $\gamma_\perp = 0$ but with $\varepsilon_{i,t}$ heteroscedastic both in i and t.

In every sub-section we present the corresponding Monte Carlo results for both estimators. We initialize the iterative estimation procedure of the IFE estimator using the rCCEP estimator. We use the simple iterative procedure to find the corresponding local minimum of the. For the IFE estimator we take the number of factors as given $R = 1$, i.e. in this way the estimator is infeasible. This way, we abstract away from the problem of estimating the true number of factors, as there are several alternative methods that we do not want to compare in this paper.

S.3.1. Default Setup

In what follows, we compare the two statistical approaches under exactly the same circumstances as in the main text. We note that while the rCCEP estimator is only $\sqrt{N}-$consistent, the IFE estimator is $\sqrt{NT}-$consistent in this design. In particular, we can expect a smaller variance and higher power from the IFE estimator. The results are presented in Table S.3. Below we summarize the main findings that mostly confirm our *ex ante* expectations.

- Both estimators are almost completely free from any bias. The IFE-BC estimator also dominates the rCCE-BC in terms of the RMSE. This is especially visible for larger values of T, as the latter estimator is only $\sqrt{N}-$consistent.
- Bootstrap based CIs provide good coverage for both procedures. However, size control is marginally better for the rCCEP-BC estimator.
- As dictated by a faster convergence rate, the power of the IFE-BC based testing procedures is strictly larger than that of the rCCEP-BC estimator. This difference is especially pronounced for $N = 20$.

Table S.3: Estimation and inference. Default setup.

Design			$rCCEP - BC$					$IFE - BC$				
$\gamma_\perp$	N	T	Bias	RMSE	Power−	Size	Power+	Bias	RMSE	Power−	Size	Power+
0	20	20	.02	.23	.58	.05	.31	.01	.19	.71	.07	.49
0	20	50	.02	.16	.78	.04	.44	.00	.11	.97	.06	.90
0	20	100	.01	.14	.85	.04	.50	.00	.08	1	.07	.99
0	50	20	.01	.19	.94	.05	.83	.01	.18	.97	.07	.95
0	50	50	.01	.14	.98	.05	.91	.00	.11	1	.08	1
0	50	100	.01	.11	1	.05	.92	.00	.08	1	.06	1
0	100	20	.00	.19	1	.06	.99	.00	.18	1	.07	1
0	100	50	.00	.14	1	.05	.99	.00	.11	1	.05	1
0	100	100	.01	.11	1	.05	1	.00	.08	1	.06	1

Notes. $rCCEP - BC$ is the bias-corrected pooled rCCE estimator; $IFE - BC$ is bias-corrected IFE estimator. See Table S.1 for more details.

S.3.2. Weak Identification in Mean Loadings

In this section, we can consider a setup where R is only "weakly" identified from cross-sectional averages as $\gamma_\perp = 0.3$. In particular, even if the true number of factors even in the limit is $R = 2$, the ER criterion might fail to pick up the correct number of factors as the two non-zero eigenvalues are far away from each other. For this reason, this setup is unfavourable with respect to the rCCEP-BC estimator, while the IFE-BC estimator is completely invariant to $\gamma_\perp$. [3] We note that in this case it is not the estimator rCCEP-BC itself that is sensitive to the small values of $\gamma_\perp$ (for $\gamma_\perp = 0.3$ it is simply CCEP with as many factors as observables), but the proposed ER criterion.

Table S.4: Estimation and inference. Weak mean loadings.

Design			$rCCEP-BC$					$IFE-BC$				
$\gamma_\perp$	N	T	Bias	RMSE	Power−	Size	Power+	Bias	RMSE	Power−	Size	Power+
0.3	20	20	-.07	.24	.40	.05	.34	.01	.19	.71	.07	.49
0. 3	20	50	-.08	.18	.54	.05	.50	.00	.11	.97	.06	.90
0. 3	20	100	-.09	.16	.58	.04	.57	.00	.08	1	.07	1
0. 3	50	20	-.14	.27	.72	.09	.91	.00	.18	.97	.07	.96
0. 3	50	50	-.15	.22	.87	.12	.99	.00	.11	1	.08	1
0. 3	50	100	-.16	.21	.91	.15	.99	.00	.08	1	.06	1
0. 3	100	20	-.19	.31	.95	.16	1	.00	.18	1	.07	1
0. 3	100	50	-.21	.28	.99	.28	1	.00	.11	1	.05	1
0. 3	100	100	-.22	.27	1	.38	1	.00	.06	1	.06	1

Notes. See Table S.3 for more details.

The results are presented in Table S.4. Below we summarize the main findings.

- The results for the IFE-BC estimator are almost identical those in the default setup.
- The rCCEP-BC estimator is severely biased, where the bias is increasing in T, and especially in N.
- For larger values of N, bootstrap inference for rCCEP-BC is highly unreliable.

Overall, we can conclude that for a small value of $\gamma_\perp$ we consider, standard asymptotic approximation is unreliable. As we suggested above, the main reason behind the substantial bias of the rCCEP-BC estimator is the inability of the ER procedure to pick the right number of factors, see Table

[3]Thus, *a priori* results similar to those as in the default scenario are expected for this estimator, but we note that deviations can occur as we use the rCCEP estimator as a starting values for numerical iterations.

S.5. From this table we can conclude that a substantially larger N is needed (i.e. $N \approx 1000$) for the reported numbers to close to those reported in the main text. The corresponding estimation and inference results are provided in Table S.6. As expected the results improve as N increases.

Table S.5: Eigenvalue Ratio (ER) based estimates of R. Weak mean loadings.

		$\#\widehat{R} = 1$
N	T	$\gamma_{\perp} = 0.3$
20	20	.7890
20	50	.8520
20	100	.8610
50	20	.7910
50	50	.8448
50	100	.8610
100	20	.7328
100	50	.7925
100	100	.8130
500	20	.2530
500	50	.2295
500	100	.2245
1000	20	.0675
1000	50	.0335
1000	100	.0195

Notes. The ER statistic is implemented as in Eq. (15) using the normalized cross-section averages, and the dummy-column as in Remark 6. For $\gamma_{\perp} = 0.3$, the true number of factors is $R = 2$. $M = 4000$.

Table S.6: Estimation and inference. Weak mean loadings.

Design			$rCCEP - BC$				
$\gamma_{\perp}$	N	T	Bias	RMSE	Power−	Size	Power+
0.3	500	20	-.15	.36	1	.22	1
0.3	500	50	-.13	.30	1	.26	1
0.3	500	100	-.12	.28	1	.26	1
0.3	1000	20	-.05	.30	1	.11	1
0.3	1000	50	-.02	.19	1	.08	1
0.3	1000	100	-.01	.13	1	.07	1

Notes. See Table S.3 for more details.

S.3.3. Heteroscedastic Errors

In this section, we consider a modification to the default setup where the DGP for $y_{i,t}$ is now of the following form:

$$y_{i,t} = x_{i,t}\beta + f_t\lambda_i + (\lambda_i\pi_t)\varepsilon_{i,t}, \tag{S.34}$$

where $\pi_t \sim N(0,1)$ iid over t. This design is advantageous for the rCCEP estimator, as consistency properties of this estimator do not depend on the second moment of the combined error term $(\lambda_i\pi_t)\varepsilon_{i,t}$. The IFE-BC estimator, while being asymptotically unbiased even under heteroscedasticity, can be severely biased for small values of N, T.

The results are presented in Table S.7. Below we summarize the main findings.

- Similarly to the default setup, the rCCEP-BC estimator is almost completely free from bias. On the other hand, the IFE-BC estimator is severely biased unless $T = 100$. Hence, while the bias-correction removes leading bias terms, the higher order terms (as well as estimation error from the bias-correction) have sizeable effects on both the bias and the RMSE.

- rCCEP-BC based bootstrap CIs have good coverage. On the other hand, the same is not true for the IFE-BC estimator that has empirical rejection frequencies that are substantially higher than the nominal size of 5%. As with the bias, the situation improves substantially only for $T = 100$.

- For the rCCEP-BC estimator the power is somewhat lower than in the default scenario. Due to the substantial bias, the power analysis of the IFE-BC is inappropriate.

Table S.7: Estimation and inference. Heteroscedastic Errors.

Design			$rCCEP - BC$					$IFE - BC$				
$\gamma_\perp$	N	T	Bias	RMSE	Power−	Size	Power+	Bias	RMSE	Power−	Size	Power+
0	20	20	-.02	.30	.24	.04	.34	.44	.30	.35	.18	.06
0	20	50	-.02	.21	.27	.04	.48	.23	.25	.43	.20	.08
0	20	100	-.02	.19	.28	.05	.52	.13	.24	.46	.22	.08
0	50	20	.00	.28	.62	.05	.72	.53	.28	.71	.21	.18
0	50	50	-.01	.19	.78	.04	.85	.11	.23	.85	.13	.31
0	50	100	.00	.17	.87	.05	.87	.03	.22	.89	.11	.47
0	100	20	-.01	.27	.89	.05	.94	.58	.28	.94	.20	.34
0	100	50	.00	.20	.98	.06	.98	.07	.23	.99	.10	.68
0	100	100	.00	.17	.99	.06	.99	.01	.22	1	.07	.91

Notes. See Table S.3 for more details.

S.4. Empirical Application

S.4.1. Regularized Estimation with $R = 2$

In this section, we investigate the sensitivity of empirical results to the number of estimated factors, R. In Tables S.8-S.9 we present point estimates and the corresponding CIs for the regularized CCE estimators with $R = 2$. This is the number of factors suggested by the Growth Ratio statistic for the "Full" model. For the sake of completeness, we also report results for the "Narrow" model.

Mean-group estimation. From Table S.8 we observe that results for $rCCE - MG$ estimator are comparable to those obtained in the main text with $R = 1$. The only notable difference is that the CI for $\ln(\sigma_{i,t})$ in the "Narrow" model no longer includes zero, while in the "Full" model the estimated coefficient decreases even further. This moves the estimated effect even further away from the estimate using the CCC-MG procedure. Regarding the control variables, we notice that the CI for $OS_{i,t}^{narr}$ contains zero (in contrast to all other estimating procedures).

Pooled estimation. We observe that inclusion of an additional regularized factor has a sizeable effect on the magnitude of the estimated coefficient for $\ln\left(\frac{H_{i,t}}{L_{i,t}}\right)$. In particular, in the "Full" model the coefficient increases from 0.37 to 0.41, a sizeable change given the width of the corresponding CIs. Thus, in contrast to rCCE-MG estimation, inclusion of an additional regularized factor brings the rCCEP results much closer to those of CCEP (at least for the main parameters interests). As for the coefficients for included control variables, we see that some of the estimated coefficients change signs. However, in all cases the corresponding CIs are still sufficiently wide and include zero.

Table S.8: Estimation results using the Mean-Group CCE for $R = 2$.

	Narrow	Full
	$rCCE-MG$	$rCCE-MG$
$\ln(\sigma_{i,t})$	-0.55	-1.10
	(-1.01;-0.13)	(-1.54;-0.71)
$\ln\left(\frac{H_{i,t}}{L_{i,t}}\right)$	0.33	0.42
	(0.30;0.36)	(0.40;0.45)
$k_{i,t}^{equip}$		-0.57
		(-1.61;0.64)
$(OCAM/K)_{i,t}$		-2.35
		(-4.28;-0.15)
$(HT/K - OCAM/K)_{i,t}$		2.40
		(0.82;3.77)
$RD_{lag.i,t}$		1.09
		(-0.46;2.76)
$OS_{i,t}^{narr}$		-1.15
		(-3.17;0.07)
$(OS^{broad} - OS^{narr})_{i,t}$		0.15
		(-0.46;0.74)

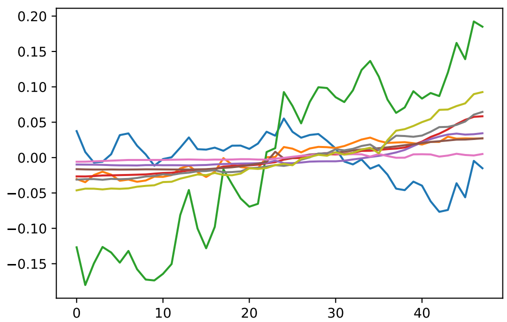
Notes. 95% percentile bootstrap confidence intervals in the parentheses. $\sigma_{i,t}$, the input skill intensity measure; $\frac{H_{i,t}}{L_{i,t}}$ the ratio of high and low skilled workers in the sector; $k_{i,t}^{equip}$ capital equipment per worker; $RD_{lag.i,t}$ research and development intensity; $(HT/K)_{i,t}$ the sectoral share of high-technology capital $(OCAM/K)_{i,t}$ the sectoral share of office, computing, and accounting equipment; $OS_{i,t}^{broad}$ and $OS_{i,t}^{narr}$ are the broad and narrow measures of outsourcing.

Table S.9: Estimation results based on the Pooled CCE for $R = 2$.

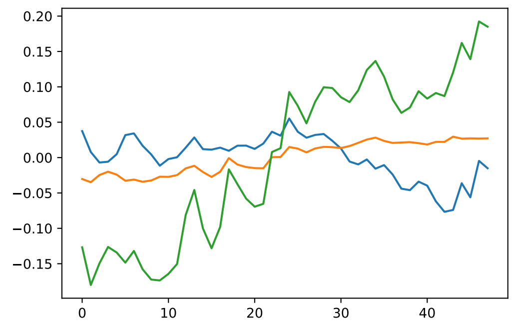
	Narrow		Full	
	$rCCE$	$rCCE_{BC}$	$rCCE$	$rCCE_{BC}$
$\ln(\sigma_{i,t})$	-0.66	-0.67	-0.99	-0.99
	(-1.03;-0.25)	(-1.04;-0.26)	(-1.33;-0.61)	(-1.33;0.61)
$\ln\left(\frac{H_{i,t}}{L_{i,t}}\right)$	0.37	0.37	0.41	0.41
	(0.30;0.43)	(0.30;0.43)	(0.34;0.48)	(0.34;0.48)
$k_{i,t}^{equip}$			0.05	0.05
			(-0.12;0.37)	(-0.12;0.37)
$(OCAM/K)_{i,t}$			0.05	0.05
			(-0.89;0.93)	(-0.89;0.93)
$(HT/K - OCAM/K)_{i,t}$			0.92	0.91
			(0.19;1.63)	(0.18;1.62)
$RD_{lag.i,t}$			-0.16	-0.15
			(-0.57;0.32)	(-0.57;0.32)
$OS_{i,t}^{narr}$			-0.13	0.13
			(-0.27;-0.01)	(-0.27;-0.01)
$(OS^{broad} - OS^{narr})_{i,t}$			0.00	0.00
			(-0.17;0.18)	(-0.17;0.18)

Notes. See Table S.8 for a detailed description of the regressors.

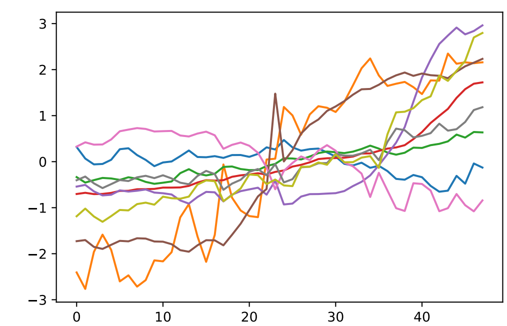
S.4.2. Figures



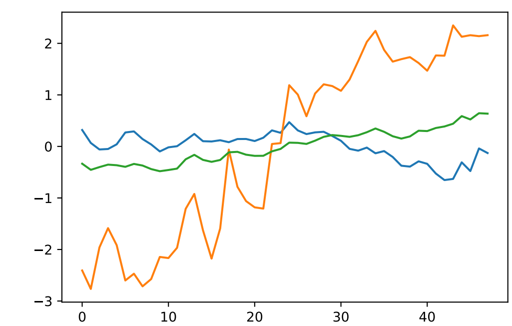
(a) Cross-sectional averages for "Full" model.



(b) Cross-sectional averages for "Narrow" model.



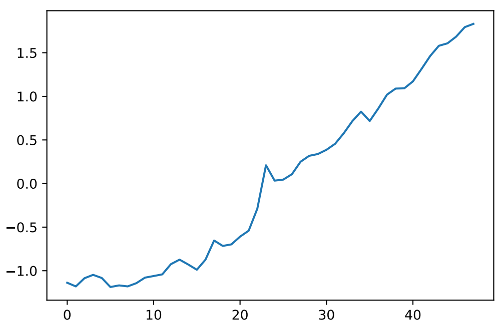
(c) Normalized averages $\widehat{\boldsymbol{F}}$ for "Full" model.



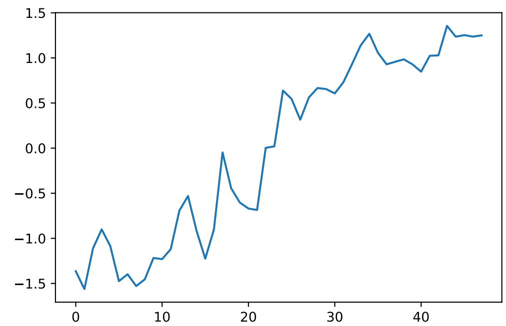
(d) Normalized averages $\widehat{\boldsymbol{F}}$ for "Narrow" model.

Figure S.1: Cross-sectional averages $\overline{\boldsymbol{Z}}$ and their Normalized versions $\widehat{\boldsymbol{F}}$ evolving over time.

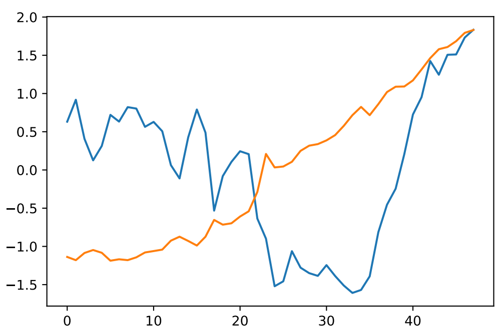
Figure S.1 illustrates the effects of normalization. As we see from sub-figures (c) and (d) of Figure S.1, the upward trending pattern dominates most of the columns in $\widehat{\boldsymbol{F}}$. Furthermore, most variables are well synchronized in terms of cycles, thus it is not surprising that ER criterion mostly indicate the presence of only one factor. Estimates of the regularized factor proxies $\widehat{\boldsymbol{F}}_r$ are provided in Figure S.2. The estimated factor for the "Full" model captures common upward trending behaviour visible for most of the series. For the "Narrow" model, on the other hand, the regularized factor proxy $\widehat{\boldsymbol{F}}_r$ almost perfectly correlates with two of the normalized cross-sectional averages.



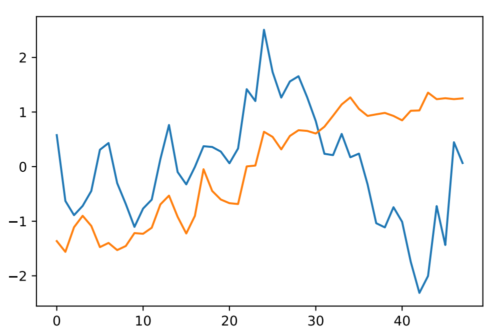
(a) $\widehat{\boldsymbol{F}}_r$ for "Full" model with $R = 1$.



(b) $\widehat{\boldsymbol{F}}_r$ for "Narrow" model with $R = 1$.



(c) $\widehat{\boldsymbol{F}}_r$ for "Full" model with $R = 2$.



(d) $\widehat{\boldsymbol{F}}_r$ for "Narrow" model with $R = 2$.

Figure S.2: Regularized cross-sectional averages $\widehat{\boldsymbol{F}}_r$.

S.5. Proofs

S.5.1. Notes

Throughout this appendix we denote by $\widetilde{\boldsymbol{F}} = \widehat{\boldsymbol{F}}_r$ the regularized estimator of factors with the number of factors $R = R_0$ taken as given (except, for an obvious reason, in the proof of Proposition 1). Furthermore, we will ignore the fact that $\widehat{\boldsymbol{F}}$ is constructed using normalized factor proxies, i.e. $\widehat{\boldsymbol{F}} = \overline{\boldsymbol{Z}}(\widehat{\boldsymbol{\Sigma}}^{-1/2})'$, and instead consider $\widehat{\boldsymbol{F}} = \overline{\boldsymbol{Z}}$. This greatly simplifies the notation, without any immediate effects on the results as by Assumption $\widehat{\boldsymbol{\Sigma}}$ is a positive definite matrix in the limit. Scaling by $\widehat{\boldsymbol{\Sigma}}^{-1/2}$ can be brought back into all expressions by appropriately post-multiplying $\boldsymbol{G}$ (and $\boldsymbol{G}_\perp$) with $(\widehat{\boldsymbol{\Sigma}}^{-1/2})'$, and pre-multiplying $\boldsymbol{u}_{i,t}$ by $\widehat{\boldsymbol{\Sigma}}^{-1/2}$.

In what follows we use the following decomposition:

$$\overline{\boldsymbol{Z}} = \boldsymbol{F}\overline{\boldsymbol{C}} + \boldsymbol{F}_\perp\overline{\boldsymbol{C}_\perp} + \overline{\boldsymbol{U}} = \boldsymbol{F}\boldsymbol{G} + \overline{\boldsymbol{U}}_e, \tag{S.35}$$

with $\overline{\boldsymbol{U}}_e = \boldsymbol{F}_\perp\boldsymbol{G}_\perp + \overline{\boldsymbol{U}}$ where we define $\boldsymbol{G} = \overline{\boldsymbol{C}}$, and similarly for $\boldsymbol{G}_\perp$. Furthermore, be define the combined error term of $\boldsymbol{X}_i$ by $\boldsymbol{V}_{i,+}$ such that:

$$\boldsymbol{V}_{i,+} = \boldsymbol{V}_i + \boldsymbol{F}_\perp\boldsymbol{\Lambda}_{i,\perp}. \tag{S.36}$$

Notice that by definition of $\widetilde{\boldsymbol{F}}$:

$$\widetilde{\boldsymbol{F}} = \overline{\boldsymbol{Z}}\left(\frac{1}{T}\overline{\boldsymbol{Z}}'\widetilde{\boldsymbol{F}}\right)\boldsymbol{V}_{N,T}^{-1}, \tag{S.37}$$

it follows that one can decompose (see also Bai 2003):

$$\widetilde{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{H} = \overline{\boldsymbol{\Psi}}. \tag{S.38}$$

Here

$$\overline{\boldsymbol{\Psi}} = \frac{1}{T}\left(\boldsymbol{F}\boldsymbol{G}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}} + \overline{\boldsymbol{U}}_e\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}} + \overline{\boldsymbol{U}}_e\boldsymbol{G}'\boldsymbol{F}'\widetilde{\boldsymbol{F}}\right)\boldsymbol{V}_{N,T}^{-1}, \tag{S.39}$$

and

$$\boldsymbol{H} = (\boldsymbol{G}\boldsymbol{G}')\left(\frac{1}{T}\boldsymbol{F}'\widetilde{\boldsymbol{F}}\right)\boldsymbol{V}_{N,T}^{-1}. \tag{S.40}$$

As the proofs are similar to those in Westerlund and Urbain (2015), Karabıyık et al. (2019), and Juodis et al. (2021) some of the derivations are streamlined. This allows us to focus on the details essential for the approach of this paper.

S.5.2. Auxiliary Lemmas

Lemma 1. *Under Assumptions 4.1-4.5 with $\ell = 4$:*

$$\|\boldsymbol{V}_{N,T}^{-1}\| = \mathcal{O}_P(1). \tag{S.41}$$

Proof of Lemma 1.

Note that $\boldsymbol{V}_{N,T}$ contains the first R eigenvalues of $T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}$ in the descending order. Under the maintained assumptions:

$$\frac{1}{T}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}} = \mathrm{E}[\boldsymbol{C}_i]'\,\mathrm{E}[\boldsymbol{f}_t\boldsymbol{f}_t']\,\mathrm{E}[\boldsymbol{C}_i] + \mathrm{o}_P(1). \tag{S.42}$$

The matrix on the right hand side of Eq. (S.42) by Assumption 4.4 has exactly R distinct eigenvalues. Thus by the continuity of the eigenvalues

$$\boldsymbol{V}_{N,T} = \boldsymbol{V}_R + \mathrm{o}_P(1), \tag{S.43}$$

where $\boldsymbol{V}_R$ contains R eigenvalues of the right hand side matrix in Eq. (S.42) on its diagonal in the descending order. Finally,

$$\|\boldsymbol{V}_{N,T}^{-1}\| = \mathcal{O}_P(1), \tag{S.44}$$

by the continuity of the Frobenius norm. □

Lemma 2. *Under Assumptions 4.1-4.5 with $\ell = 4$:*

$$\|T^{-1}\overline{\boldsymbol{Z}}'\widetilde{\boldsymbol{F}}\| = \mathcal{O}_P(1). \tag{S.45}$$

Proof of Lemma 2.

Using the definition of $\overline{\boldsymbol{Z}} = \boldsymbol{F}_e\overline{\boldsymbol{C}} + \overline{\boldsymbol{U}}$, notice that:

$$\|T^{-1}\overline{\boldsymbol{Z}}'\widetilde{\boldsymbol{F}}\| < \|T^{-1}\boldsymbol{F}_e'\widetilde{\boldsymbol{F}}\|\|\overline{\boldsymbol{C}}\| + \|T^{-1}\overline{\boldsymbol{U}}'\widetilde{\boldsymbol{F}}\|. \tag{S.46}$$

Here:

$$\|T^{-1}\boldsymbol{D}'\widetilde{\boldsymbol{F}}\| < \sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{d}_t\|^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\widetilde{\boldsymbol{f}}_t\|^2} = \mathcal{O}_P(1), \tag{S.47}$$

as by assumption $\boldsymbol{d}_t$ has a finite fourth moment, while $\sum_{t=1}^{T}\|\widetilde{\boldsymbol{f}}_t\|^2 = T$. Similarly for the second component:

$$\|T^{-1}\overline{\boldsymbol{U}}'\widetilde{\boldsymbol{F}}\| < \sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\overline{\boldsymbol{u}}_t\|^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\widetilde{\boldsymbol{f}}_t\|^2} = \mathcal{O}_P(N^{-1/2}), \tag{S.48}$$

where the result follows from the fact that:

$$\frac{1}{T}\sum_{t=1}^{T}\|\overline{\boldsymbol{u}}_t\|^2 = \mathcal{O}_P(N^{-1}), \tag{S.49}$$

as can be shown using similar steps to those in Pesaran (2006). The final result follows by noting that also $\|\overline{\boldsymbol{C}}\| = \mathcal{O}_P(1)$ by assumption. □

Lemma 3. *Under Assumptions 4.1-4.5 with* $\ell = 4$*:*

(i) $\|T^{-1}\boldsymbol{V}_{i,+}'\boldsymbol{F}\| = \mathcal{O}_P(T^{-1/2})$,

(ii) $\|T^{-1}\boldsymbol{V}_{i,+}'\boldsymbol{F}_{\perp}\| = \mathcal{O}_P(1)$,

(iii) $\|T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{U}}\| = \mathcal{O}_P(N^{-1}) + \mathcal{O}_P((NT)^{-1/2})$,

(iv) $\|T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{U}}_e\| = \mathcal{O}_P(N^{-1/2})$,

(v) $\|T^{-1}\boldsymbol{F}'\overline{\boldsymbol{U}}_e\| = \mathcal{O}_P((NT)^{-1/2})$,

(vi) $\|T^{-1}\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}_e\| = \mathcal{O}_P(N^{-1}) + \mathcal{O}_P((NT)^{-1/2})$,

(vii) $\|T^{-1}\overline{\boldsymbol{U}}'\overline{\boldsymbol{U}}_e\| = \mathcal{O}_P(N^{-1})$.

Proof of Lemma 3.
The results follow directly from Lemmas A.1-A.2 in Juodis et al. (2021) upon noticing that $\boldsymbol{\Gamma}_{f,\perp}(0) = \mathrm{E}[\boldsymbol{f}_t\boldsymbol{f}_{t,\perp}] = \mathbf{O}$ by assumption. □

Lemma 4. *Under Assumptions 4.1-4.5 with* $\ell = 4$*:*

(i) $\|T^{-1}\boldsymbol{V}_{i,+}'\widetilde{\boldsymbol{F}}\| = \mathcal{O}_P(N^{-1/2})$,

(ii) $\|T^{-1}\boldsymbol{\varepsilon}_i'\widetilde{\boldsymbol{F}}\| = \mathcal{O}_P(N^{-1/2})$,

(iii) $\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\| = \mathcal{O}_P(N^{-1})$,

(iv) $\|T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{\Psi}}\| = \mathcal{O}_P(N^{-1/2})$,

(v) $\|T^{-1}\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{\Psi}}\| = \mathcal{O}_P(N^{-1})$,

(vi) $\|T^{-1}\boldsymbol{F}'\overline{\boldsymbol{\Psi}}\| = \mathcal{O}_P(N^{-1})$,

(vii) $\|T^{-1}\widetilde{\boldsymbol{F}}'\overline{\boldsymbol{\Psi}}\| = \mathcal{O}_P(N^{-1})$,

(viii) $\|T^{-1}\overline{\boldsymbol{U}}_e'\overline{\boldsymbol{\Psi}}\| = \mathcal{O}_P(N^{-1})$,

(ix) $\|T^{-1}\overline{\boldsymbol{\Psi}}'\overline{\boldsymbol{\Psi}}\| = \mathcal{O}_P(N^{-1})$.

Proof of Lemma 4.
In what follows we use the following result repeatedly:

$$\|T^{-1}\boldsymbol{R}'\widetilde{\boldsymbol{F}}\| < \|T^{-1}\boldsymbol{R}'\overline{\boldsymbol{Z}}\|\|T^{-1}\overline{\boldsymbol{Z}}'\widetilde{\boldsymbol{F}}\|\|\boldsymbol{V}_{N,T}^{-1}\| < \|T^{-1}\boldsymbol{R}'\overline{\boldsymbol{Z}}\|\mathcal{O}_P(1), \tag{S.50}$$

where $\boldsymbol{R}$ is any "well-behaved" matrix, while $\|\boldsymbol{V}_{N,T}^{-1}\| = \mathcal{O}_P(1)$ follows from Lemma 1.

Part (i).

$$
\begin{aligned}
\|T^{-1}\boldsymbol{V}_{i,+}'\widetilde{\boldsymbol{F}}\| &< \|T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{Z}}\|\mathcal{O}_P(1) \\
&< \left(\|T^{-1}\boldsymbol{V}_{i,+}'\boldsymbol{F}\boldsymbol{G}\| + \|T^{-1}\boldsymbol{V}_{i,+}'\boldsymbol{F}_{\perp}\boldsymbol{G}_{\perp}\| + \|T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{U}}\|\right)\mathcal{O}_P(1) \\
&< \left(\mathcal{O}_P(T^{-1/2}) + \mathcal{O}_P(N^{-1/2}) + \mathcal{O}_P(N^{-1}) + \mathcal{O}_P((NT)^{-1/2})\right)\mathcal{O}_P(1) \\
&= \mathcal{O}_P(T^{-1/2}) + \mathcal{O}_P(N^{-1/2}). \qquad \text{(S.51)}
\end{aligned}
$$

Here the third line follows from Lemma 3. Under $N/T \to c$ the above reduces to $\mathcal{O}_P(N^{-1/2})$.

Part (ii).

$$
\begin{aligned}
\|T^{-1}\boldsymbol{\varepsilon}_i'\widetilde{\boldsymbol{F}}\| &< \|T^{-1}\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{Z}}\|\mathcal{O}_P(1) \\
&< \left(\|T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{F}\boldsymbol{G}\| + \|T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{F}_{\perp}\boldsymbol{G}_{\perp}\| + \|T^{-1}\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}\|\right)\mathcal{O}_P(1) \\
&< \left(\mathcal{O}_P(T^{-1/2}) + \mathcal{O}_P((NT)^{-1/2}) + \mathcal{O}_P(N^{-1}) + \mathcal{O}_P((NT)^{-1/2})\right)\mathcal{O}_P(1) \\
&= \mathcal{O}_P(T^{-1/2}) + \mathcal{O}_P((NT)^{-1/2}) + \mathcal{O}_P(N^{-1}) \\
&= \mathcal{O}_P(N^{-1/2}). \qquad \text{(S.52)}
\end{aligned}
$$

Under $N/T \to c$ the above reduces to $\mathcal{O}_P(N^{-1/2})$.

Part (iii).

$$
\begin{aligned}
\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\| &< \|T^{-1}\overline{\boldsymbol{U}}_e'\overline{\boldsymbol{Z}}\|\mathcal{O}_P(1) \\
&< \left(\|T^{-1}\overline{\boldsymbol{U}}_e\boldsymbol{F}\boldsymbol{G}\| + \|T^{-1}\overline{\boldsymbol{U}}_e'\boldsymbol{F}_{\perp}\boldsymbol{G}_{\perp}\| + \|T^{-1}\overline{\boldsymbol{U}}_e'\overline{\boldsymbol{U}}\|\right)\mathcal{O}_P(1) \\
&< \left(\mathcal{O}_P((NT)^{-1/2}) + \mathcal{O}_P((NT)^{-1/2}) + \mathcal{O}_P(N^{-1}) + \mathcal{O}_P(N^{-1})\right)\mathcal{O}_P(1) \\
&= \mathcal{O}_P(N^{-1}). \qquad \text{(S.53)}
\end{aligned}
$$

Part (iv).

$$
\begin{aligned}
\|T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{\Psi}}\| &< \left(\|T^{-1}\boldsymbol{V}_{i,+}'\boldsymbol{F}\|\|\boldsymbol{G}\|\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\| + \|T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{U}}_e\|\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\| + \|T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{U}}_e\|\|\boldsymbol{G}\|\|T^{-1}\boldsymbol{F}'\widetilde{\boldsymbol{F}}\|\right)\mathcal{O}_P(1) \\
&< \left(\mathcal{O}(T^{-1/2})\,\mathcal{O}_P(N^{-1}) + \mathcal{O}_P(N^{-3/2}) + \mathcal{O}(N^{-1/2})\right)\mathcal{O}_P(1) \\
&= \mathcal{O}(N^{-1/2}). \qquad \text{(S.54)}
\end{aligned}
$$

Here we use the results from part (iii) explicitly in the second line.

Part (v).

$$
\begin{aligned}
\|T^{-1}\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{\Psi}}\| &< \left(\|T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{F}\|\|\boldsymbol{G}\|\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\| + \|T^{-1}\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}_e\|\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\| + \|T^{-1}\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{U}}_e\|\|\boldsymbol{G}\|\|T^{-1}\boldsymbol{F}'\widetilde{\boldsymbol{F}}\|\right)\mathcal{O}_P(1) \\
&< \left(\mathcal{O}(T^{-1/2})\,\mathcal{O}_P(N^{-1}) + \mathcal{O}_P(N^{-2}) + \mathcal{O}((NT)^{-1/2}N^{-1/2}) + \mathcal{O}_P(N^{-1}) + \mathcal{O}((NT)^{-1/2})\right)\mathcal{O}_P(1) \\
&= \mathcal{O}((NT)^{-1/2}) + \mathcal{O}_P(N^{-1}) \\
&= \mathcal{O}_P(N^{-1}). \qquad \text{(S.55)}
\end{aligned}
$$

Here we use the results from part (iii) explicitly in the second line. The final line follows from the $N/T \to c$.

Part (vi).

$$\begin{aligned}\|T^{-1}\boldsymbol{F}'\overline{\boldsymbol{\Psi}}\| &< (\|T^{-1}\boldsymbol{F}'\boldsymbol{F}\|\|\boldsymbol{G}\|\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\|+\|T^{-1}\boldsymbol{F}'\overline{\boldsymbol{U}}_e\|\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\|+\|T^{-1}\boldsymbol{F}'\overline{\boldsymbol{U}}_e\|\|\boldsymbol{G}\|\|T^{-1}\boldsymbol{F}'\widetilde{\boldsymbol{F}}\|)\,\mathcal{O}_P(1)\\ &< (\mathcal{O}_P(N^{-1})+\mathcal{O}_P(N^{-1}(NT)^{-1/2})+\mathcal{O}_P((NT)^{-1/2}))\,\mathcal{O}_P(1)\\ &= \mathcal{O}_P(N^{-1}). \end{aligned} \tag{S.56}$$

Part (vii).

$$\begin{aligned}\|T^{-1}\widetilde{\boldsymbol{F}}'\overline{\boldsymbol{\Psi}}\| &< (\|T^{-1}\widetilde{\boldsymbol{F}}'\boldsymbol{F}\|\|\boldsymbol{G}\|\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\|+\|T^{-1}\widetilde{\boldsymbol{F}}'\overline{\boldsymbol{U}}_e\|\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\|+\|T^{-1}\widetilde{\boldsymbol{F}}'\overline{\boldsymbol{U}}_e\|\|\boldsymbol{G}\|\|T^{-1}\boldsymbol{F}'\widetilde{\boldsymbol{F}}\|)\,\mathcal{O}_P(1)\\ &< (\mathcal{O}_P(N^{-1})+\mathcal{O}_P(N^{-2})+\mathcal{O}_P((NT)^{-1}))\,\mathcal{O}_P(1)\\ &= \mathcal{O}_P(N^{-1}). \end{aligned} \tag{S.57}$$

Part (viii).

$$\begin{aligned}\|T^{-1}\overline{\boldsymbol{U}}_e'\overline{\boldsymbol{\Psi}}\| &< (\|T^{-1}\overline{\boldsymbol{U}}_e'\boldsymbol{F}\|\|\boldsymbol{G}\|\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\|+\|T^{-1}\overline{\boldsymbol{U}}_e'\overline{\boldsymbol{U}}_e\|\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\|+\|T^{-1}\overline{\boldsymbol{U}}_e'\overline{\boldsymbol{U}}_e\|\|\boldsymbol{G}\|\|T^{-1}\boldsymbol{F}'\widetilde{\boldsymbol{F}}\|)\,\mathcal{O}_P(1)\\ &< (\mathcal{O}_P(N^{-1}(NT)^{-1/2})+\mathcal{O}_P(N^{-2})+\mathcal{O}_P(N^{-1}))\,\mathcal{O}_P(1)\\ &= \mathcal{O}_P(N^{-1}). \end{aligned} \tag{S.58}$$

Part (ix).

$$\begin{aligned}\|T^{-1}\overline{\boldsymbol{\Psi}}'\overline{\boldsymbol{\Psi}}\| &< (\|T^{-1}\overline{\boldsymbol{\Psi}}'\boldsymbol{F}\|\|\boldsymbol{G}\|\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\|+\|T^{-1}\overline{\boldsymbol{\Psi}}'\overline{\boldsymbol{U}}_e\|\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\|+\|T^{-1}\overline{\boldsymbol{\Psi}}'\overline{\boldsymbol{U}}_e\|\|\boldsymbol{G}\|\|T^{-1}\boldsymbol{F}'\widetilde{\boldsymbol{F}}\|)\,\mathcal{O}_P(1)\\ &< (\mathcal{O}_P(N^{-3/2})+\mathcal{O}_P(N^{-2})+\mathcal{O}_P(N^{-1}))\,\mathcal{O}_P(1)\\ &= \mathcal{O}_P(N^{-1}). \end{aligned} \tag{S.59}$$

□

S.5.3. Eigenvalue Ratio Statistic

Proof of Proposition 1.

Denote by $\widehat{\boldsymbol{B}} = T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}$ and the corresponding probability limit as $\boldsymbol{B} = \boldsymbol{C}'\boldsymbol{\Gamma}(0)\boldsymbol{C}$. From Lemma 3 we know that $\widehat{\nu}_r \xrightarrow{p} \nu_r$ for all $r = 1, \ldots, K_z$, where ν_r are the eigenvalues of $\boldsymbol{B}$. Thus for all $1 \leq r < R$ the eigenvalue ratio is asymptotically bounded

$$ER(r) \xrightarrow{p} \frac{\nu_r}{\nu_{r+1}} < \infty. \tag{S.60}$$

It remains to be shown that:

$$ER(R) \xrightarrow{p} \infty, \tag{S.61}$$

and

$$ER(r) = \mathcal{O}_P(1), \quad r = R+1, \ldots, K_z - 1. \tag{S.62}$$

While it is easy to show (a simple extension of the corresponding result in Ahn and Horenstein (2013)) that all zero eigenvalues have an upper bound on the convergence rate i.e. $N\widehat{\nu}_r = \mathcal{O}_P(1)$ for all $r = R+1, \ldots, K_z$, their proof cannot be used directly to show that the asymptotic limit is non-degenerate for all $r > R$.[4] Instead, we use the proof strategy similar to that in Theorem 3.1. in Robin and Smith (2000).

By assumption matrix $\boldsymbol{B}$ is of rank R and admits the eigen-decomposition of the form:

$$\boldsymbol{B} = \boldsymbol{Q_B}\boldsymbol{V}\boldsymbol{Q}'_{\boldsymbol{B}}. \tag{S.63}$$

Here $\boldsymbol{V}$ is an $[K_z \times K_z]$ diagonal matrix with the corresponding K_z eigenvalues on the main diagonal. Moreover, $\boldsymbol{Q}'_{\boldsymbol{B}}\boldsymbol{Q_B} = \boldsymbol{I}_{K_z}$. Let us denote all diagonal elements of $\boldsymbol{V}$ as $\nu_1, \ldots, \nu_{K_z}$. By Assumption 4.4: $\nu_1 > \ldots > \nu_R > 0$, while $\nu_r = 0$ for $r = R+1, \ldots, K_z$, so that:

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_R & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}. \tag{S.64}$$

For later reference we note that the Moore-Penrose inverse of $\boldsymbol{B}$ is given by:

$$\boldsymbol{B}^+ = \boldsymbol{Q_B}\boldsymbol{V}^+\boldsymbol{Q}'_{\boldsymbol{B}}, \tag{S.65}$$

where

$$\boldsymbol{V}^+ = \begin{pmatrix} \boldsymbol{V}_R^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}. \tag{S.66}$$

Conformably partition $\boldsymbol{Q_B}$ with respect to the R non-zero and $(K_z - R)$ zero eigenvalues as $\boldsymbol{Q_B} = (\boldsymbol{Q}_{\boldsymbol{B},1}, \boldsymbol{Q}_{\boldsymbol{B},2})$, so that $\boldsymbol{Q}'_{\boldsymbol{B},2}\boldsymbol{Q}_{\boldsymbol{B},1} = \mathbf{O}$. Moreover, as all $\boldsymbol{C}$ lie in the column space of $\boldsymbol{Q}_{\boldsymbol{B},1}$ it also implies that:

$$\boldsymbol{Q}'_{\boldsymbol{B},2}\boldsymbol{C} = \mathbf{O}. \tag{S.67}$$

[4] Ahn and Horenstein (2013) bound $N\widehat{\nu}_r$ from below by the $2r+1$ eigenvalue of $(N/T)\overline{\boldsymbol{U}}_e\overline{\boldsymbol{U}}'_e$, however in our setup this matrix is of finite rank K_z. Thus already for $r = K_z/2$ the lower bound becomes uninformative.

For matrix $\widehat{\boldsymbol{B}}$ denote its corresponding eigenvalue function indexed by λ by $\widehat{\boldsymbol{S}}(\lambda) = \widehat{\boldsymbol{B}} - \lambda \boldsymbol{I}$. Next, we make use of the decomposition of the orthogonal structure in $\boldsymbol{Q}_{\boldsymbol{B}}$ to construct a rotated version of $\widehat{\boldsymbol{S}}(\lambda)$. This step is similar to the corresponding rotation used in Karabıyık et al. (2017) and Juodis et al. (2021). Hence the rotated eigenvalue function can be expanded as:

$$
\begin{aligned}
&(\boldsymbol{Q}_{\boldsymbol{B},1}, \sqrt{N}\boldsymbol{Q}_{\boldsymbol{B},2})'\widehat{\boldsymbol{S}}(\widehat{\nu}_r)(\boldsymbol{Q}_{\boldsymbol{B},1}, \sqrt{N}\boldsymbol{Q}_{\boldsymbol{B},2}) \\
&= \begin{pmatrix} \boldsymbol{V}_R + \mathrm{o}_P(1) & \boldsymbol{Q}'_{\boldsymbol{B},1}\sqrt{N}(\widehat{\boldsymbol{B}} - \boldsymbol{B})\boldsymbol{Q}_{\boldsymbol{B},2} + \mathrm{o}_P(1) \\ \boldsymbol{Q}'_{\boldsymbol{B},2}\sqrt{N}(\widehat{\boldsymbol{B}} - \boldsymbol{B})\boldsymbol{Q}_{\boldsymbol{B},1} + \mathrm{o}_P(1) & \boldsymbol{Q}'_{\boldsymbol{B},2}N(\widehat{\boldsymbol{B}} - \boldsymbol{B})\boldsymbol{Q}'_{\boldsymbol{B},2} - N\widehat{\nu}_r\boldsymbol{I} \end{pmatrix},
\end{aligned} \tag{S.68}
$$

for $r = R+1, \ldots K_z$. Here we explicitly make use of the fact that all eigenvalues are consistent for their population analogues. In particular, $\widehat{\nu}_r = \mathrm{o}_P(1)$ for $r > R$. Using Lemma 2 and similar results in Juodis et al. (2021), it follows that:

$$
\boldsymbol{Q}'_{\boldsymbol{B},1}\sqrt{N}(\widehat{\boldsymbol{B}} - \boldsymbol{B})\boldsymbol{Q}_{\boldsymbol{B},2} = \boldsymbol{Q}'_{\boldsymbol{B},1}\boldsymbol{C}'\boldsymbol{\Gamma}(0)\sqrt{N}(\boldsymbol{G} - \boldsymbol{C})\boldsymbol{Q}_{\boldsymbol{B},2} + \mathrm{o}_P(1), \tag{S.69}
$$

and

$$
\begin{aligned}
\boldsymbol{Q}'_{\boldsymbol{B},2}N(\widehat{\boldsymbol{B}} - \boldsymbol{B})\boldsymbol{Q}_{\boldsymbol{B},2} &= \boldsymbol{Q}'_{\boldsymbol{B},2}\left(\frac{N}{T}\overline{\boldsymbol{U}}'_e\overline{\boldsymbol{U}}_e + \sqrt{N}(\boldsymbol{G} - \boldsymbol{C})'\boldsymbol{\Gamma}(0)\sqrt{N}(\boldsymbol{G} - \boldsymbol{C})\right)\boldsymbol{Q}_{\boldsymbol{B},2} + \mathrm{o}_P(1), \\
&= \boldsymbol{Q}'_{\boldsymbol{B},2}\left(\mathrm{E}[\boldsymbol{u}_{i,t}\boldsymbol{u}'_{i,t}] + \sqrt{N}(\boldsymbol{G} - \boldsymbol{C})'\boldsymbol{\Gamma}(0)\sqrt{N}(\boldsymbol{G} - \boldsymbol{C})\right)\boldsymbol{Q}_{\boldsymbol{B},2} \\
&+ \boldsymbol{Q}'_{\boldsymbol{B},2}\left(\sqrt{N}(\boldsymbol{G}_\perp)'\boldsymbol{\Gamma}_{\perp,\perp}(0)\sqrt{N}(\boldsymbol{G}_\perp)\right)\boldsymbol{Q}_{\boldsymbol{B},2}.
\end{aligned} \tag{S.70}
$$

As by assumption $|\boldsymbol{V}_R| > 0$, then for $r = R+1, \ldots, K_z$:

$$
\begin{aligned}
0 &= |\boldsymbol{V}_R|^{-1}|\widehat{\boldsymbol{S}}(\widehat{\nu}_r)| = |\boldsymbol{V}_R|^{-1}|(\boldsymbol{Q}_{\boldsymbol{B},1}, \sqrt{N}\boldsymbol{Q}_{\boldsymbol{B},2})'\widehat{\boldsymbol{S}}(\widehat{\nu}_r)(\boldsymbol{Q}_{\boldsymbol{B},1}, \sqrt{N}\boldsymbol{Q}_{\boldsymbol{B},2})| \\
&= |\boldsymbol{Q}'_{\boldsymbol{B},2}\left(\mathrm{E}[\boldsymbol{u}_{i,t}\boldsymbol{u}'_{i,t}] + N(\boldsymbol{G}_\perp)'\boldsymbol{\Gamma}_{\perp,\perp}(0)(\boldsymbol{G}_\perp) + N(\boldsymbol{G} - \boldsymbol{C})'\boldsymbol{\Gamma}(0)^{1/2}\boldsymbol{M}(\boldsymbol{\Gamma}(0)^{1/2})'(\boldsymbol{G} - \boldsymbol{C})\right)\boldsymbol{Q}_{\boldsymbol{B},2} - N\widehat{\nu}_r\boldsymbol{I}| \\
&+ \mathrm{o}_P(1).
\end{aligned} \tag{S.71}
$$

Here an $[R \times R]$ matrix $\boldsymbol{M}$ is defined as:

$$
\begin{aligned}
\boldsymbol{M} &= \boldsymbol{I}_R - (\boldsymbol{\Gamma}(0)^{1/2})'\boldsymbol{C}'\boldsymbol{Q}_{\boldsymbol{B},1}\boldsymbol{V}_R^{-1}\boldsymbol{Q}'_{\boldsymbol{B},1}\boldsymbol{C}\boldsymbol{\Gamma}(0)^{1/2} \\
&= \boldsymbol{I}_R - (\boldsymbol{\Gamma}(0)^{1/2})'\boldsymbol{C}'\left(\boldsymbol{C}\boldsymbol{\Gamma}(0)\boldsymbol{C}'\right)^{+}\boldsymbol{C}\boldsymbol{\Gamma}(0)^{1/2}.
\end{aligned} \tag{S.72}
$$

Notice that $\boldsymbol{M}$ is an idempotent (thus also positive semi-definite) projection matrix.

Finally, as $\mathrm{E}[\boldsymbol{u}_{i,t}\boldsymbol{u}'_{i,t}]$ is positive definite (by assumption), the quadratic form $\boldsymbol{Q}'_{\boldsymbol{B},2}\,\mathrm{E}[\boldsymbol{u}_{i,t}\boldsymbol{u}'_{i,t}]\boldsymbol{Q}_{\boldsymbol{B},2}$ is also positive definite, while all stochastic quadratic forms inside the determinant are positive semi-definite, the whole matrix inside the determinant is almost surely positive definite. As a result, the asymptotic distribution of $N\widehat{\nu}_r$ for all $r = R+1, \ldots, K_z$ is non-degenerate, as otherwise the determinant cannot be zero for some of those r.

In particular, this confirms that:

$$
ER(R) = \frac{\widehat{\nu}_R}{\widehat{\nu}_{R+1}} = N\frac{\widehat{\nu}_R}{N\widehat{\nu}_{R+1}} = \mathcal{O}_P(N), \tag{S.73}
$$

and

$$ER(R+r) = \frac{\widehat{\nu}_{R+r}}{\widehat{\nu}_{R+1+r}} = \frac{N\widehat{\nu}_{R+r}}{N\widehat{\nu}_{R+1+r}} = \mathcal{O}_P(1), \tag{S.74}$$

for $r = 1, \ldots, K_z - R - 1$. □

S.5.4. Regularized Pooled CCE estimator

Before we proceed with the actual proof of Theorem 1, we first expand the regularized CCE estimator as:

$$\widehat{\boldsymbol{\beta}}_{rCCEP} - \boldsymbol{\beta}_0 = \widehat{\boldsymbol{W}}^{-1} \left(\boldsymbol{a}_1 - \boldsymbol{a}_2 - \boldsymbol{a}_3 - \boldsymbol{a}_4 + \boldsymbol{a}_5 \right), \tag{S.75}$$

where

$$\boldsymbol{a}_1 = (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{V}_{i,+}' \boldsymbol{\varepsilon}_i, \tag{S.76}$$

$$\boldsymbol{a}_2 = (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{V}_{i,+}' \boldsymbol{P}_{\widetilde{\boldsymbol{F}}} \boldsymbol{\varepsilon}_i, \tag{S.77}$$

$$\boldsymbol{a}_3 = (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{V}_{i,+}' \boldsymbol{M}_{\widetilde{\boldsymbol{F}}} \overline{\boldsymbol{\Psi}} \boldsymbol{H}^{-1} \boldsymbol{\lambda}_i, \tag{S.78}$$

$$\boldsymbol{a}_4 = (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{\Lambda}_i' (\boldsymbol{H}^{-1})' \overline{\boldsymbol{\Psi}}' \boldsymbol{M}_{\widetilde{\boldsymbol{F}}} \boldsymbol{\varepsilon}_i, \tag{S.79}$$

$$\boldsymbol{a}_5 = (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{\Lambda}_i' (\boldsymbol{H}^{-1})' \overline{\boldsymbol{\Psi}}' \boldsymbol{M}_{\widetilde{\boldsymbol{F}}} \overline{\boldsymbol{\Psi}} \boldsymbol{H}^{-1} \boldsymbol{\lambda}_i, \tag{S.80}$$

$$\widehat{\boldsymbol{W}} = (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_{\widetilde{\boldsymbol{F}}} \boldsymbol{X}_i. \tag{S.81}$$

Note that from Eq. (S.40)

$$T^{-1} \boldsymbol{G}' \boldsymbol{F}' \widetilde{\boldsymbol{F}} \boldsymbol{V}_{N,T}^{-1} \boldsymbol{H}^{-1} = \boldsymbol{G}' (\boldsymbol{G}\boldsymbol{G}')^{-1} = \boldsymbol{G}^{+}, \tag{S.82}$$

where $\boldsymbol{G}^{+} = \boldsymbol{G}'(\boldsymbol{G}\boldsymbol{G}')^{-1}$ is the Moore-Penrose generalized inverse of $\boldsymbol{G}$.

Lemma 5. *Under Assumptions 4.1-4.5 with* $\ell = 4$*:*

$$\boldsymbol{a}_2 = \frac{1}{T} \sum_{h=1}^{\infty} \mathrm{E}[\boldsymbol{\Gamma}_{i,\varepsilon,\boldsymbol{v}}(-h)'] \operatorname{tr}[\boldsymbol{\Gamma}(h)(\boldsymbol{\Gamma}(0))^{-1}] + \mathrm{o}_P(N^{-1}). \tag{S.83}$$

Proof of Lemma 5.
Denote by $\overline{\boldsymbol{\Phi}} = \overline{\boldsymbol{\Psi}} \boldsymbol{H}^{-1}$. Notice that we can expand the projection matrix $\boldsymbol{P}_{\widetilde{\boldsymbol{F}}}$ as follows:

$$\begin{aligned}
\boldsymbol{P}_{\widetilde{\boldsymbol{F}}} &= \left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right) \left(\left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right)' \left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right) \right)^{-1} \left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right)' \\
&= \boldsymbol{F} \left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1} \boldsymbol{F}' + \widetilde{\boldsymbol{F}} \boldsymbol{H}^{-1} \boldsymbol{\Delta} (\widetilde{\boldsymbol{F}} \boldsymbol{H}^{-1})' \\
&\quad + \widetilde{\boldsymbol{F}} \boldsymbol{H}^{-1} (\boldsymbol{F}'\boldsymbol{F})^{-1} \overline{\boldsymbol{\Phi}}' + \overline{\boldsymbol{\Phi}} (\boldsymbol{F}'\boldsymbol{F})^{-1} \boldsymbol{F}',
\end{aligned} \tag{S.84}$$

where $\boldsymbol{\Delta} = \left(\left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right)' \left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right)\right)^{-1} - (\boldsymbol{F}'\boldsymbol{F})^{-1}$. Note that under our assumptions and results in Lemma 4, it follows that:

$$\frac{1}{T}\left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right)' \left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right) - \frac{1}{T}\boldsymbol{F}'\boldsymbol{F} = \mathcal{O}_P(N^{-1}), \tag{S.85}$$

thus by the Continuous Mapping Theorem (CMT) we conclude that $T\boldsymbol{\Delta} = \mathcal{O}_P(N^{-1})$. Next we expand $\boldsymbol{a}_2$ using the expansion of $\boldsymbol{P}_{\widetilde{\boldsymbol{F}}}$.

$$\boldsymbol{a}_2 = \frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{V}'_{i,+}\boldsymbol{P}_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i + \boldsymbol{a}_{2,2} + \boldsymbol{a}_{2,3} + \boldsymbol{a}_{2,4}. \tag{S.86}$$

Using Lemma 4 it is straightforward to show that $\boldsymbol{a}_{2,2} = \mathcal{O}_P((NT)^{-1})$, $\boldsymbol{a}_{2,3} = \mathcal{O}_P(T^{-1/2}N^{-1})$. Unfortunately, as the bounds in Lemma 4 are rather conservative they are only sufficient to show that $\boldsymbol{a}_{2,4} = \mathcal{O}_P(N^{-1})$. A careful inspection of the corresponding term in Lemma 4 reveals that $\boldsymbol{a}_{2,4}$ can be expanded as follows:

$$\boldsymbol{a}_{2,4} = \frac{\sqrt{T}}{NT}\sum_{i=1}^{N} \boldsymbol{\Lambda}'_{i,\perp}(T^{-1}\boldsymbol{F}'_{\perp}\boldsymbol{F}_{\perp})\sqrt{N}\boldsymbol{G}_{\perp}\boldsymbol{G}'(T^{-1}\boldsymbol{F}'\widetilde{\boldsymbol{F}})\boldsymbol{V}^{-1}_{N,T}(T^{-1}\boldsymbol{F}'\boldsymbol{F})^{-1}(NT)^{-1/2}\boldsymbol{F}'\boldsymbol{\varepsilon}_i + \mathcal{O}_P(N^{-3/2}). \tag{S.87}$$

Using the properties of the vec(·) operator, it is straightforward to see that the stochastic order of this component is determined by the order of the following matrix

$$\boldsymbol{Q}_{N,T} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\varepsilon_{i,t}\boldsymbol{f}'_t) \otimes \boldsymbol{\Lambda}'_{i,\perp}. \tag{S.88}$$

Under the maintained assumption one can show that that each element of $\boldsymbol{Q}_{N,T}$ has a finite fourth moment, thus using the Chebyshev's inequality we can conclude that each $\boldsymbol{Q}_{N,T} = \mathcal{O}_P(1)$. From here it follows that $\boldsymbol{a}_{2,4} = \mathcal{O}(N^{-3/2})$ under assumptions that $N/T \to c$.

Finally,

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{V}'_{i,+}\boldsymbol{P}_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i &= \frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{V}'_{i}\boldsymbol{P}_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i + \frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{\Lambda}'_{i,\perp}\boldsymbol{F}'_{\perp}\boldsymbol{P}_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i \\ &= \boldsymbol{a}_{2,1,1} + \boldsymbol{a}_{2,1,2}.\end{aligned} \tag{S.89}$$

Notice that

$$\boldsymbol{a}_{2,1,2} = \frac{1}{T\sqrt{N}}\sum_{i=1}^{N} \boldsymbol{\Lambda}'_{i,\perp}(T^{-1/2}\boldsymbol{F}'_{\perp}\boldsymbol{F})(T^{-1}\boldsymbol{F}'\boldsymbol{F})^{-1}((NT)^{-1/2}\boldsymbol{F}'\boldsymbol{\varepsilon}_i). \tag{S.90}$$

Given that $\|T^{-1/2}\boldsymbol{F}'_{\perp}\boldsymbol{F}\| = \mathcal{O}_P(1)$, the stochastic order of $\boldsymbol{a}_{2,1,2}$ is determined by $\boldsymbol{Q}_{N,T}$ component, the same as in $\boldsymbol{a}_{2,4}$. Hence, $\boldsymbol{a}_{2,1,2} = \mathcal{O}_P(T^{-1}N^{-1/2}) = \mathcal{O}_P(N^{-3/2})$, as desired. The leading term $\boldsymbol{a}_{2,1}$ satisfies:

$$\|\boldsymbol{a}_{2,1}\| < \frac{1}{T}\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}\boldsymbol{V}'_i\boldsymbol{F}\|(T^{-1}\boldsymbol{F}'\boldsymbol{F})^{-1}\|T^{-1}\boldsymbol{F}'\boldsymbol{\varepsilon}_i\| \tag{S.91}$$

$$= \mathcal{O}_P(T^{-1}). \tag{S.92}$$

Finally, using the sequential limit results analogous to those in Juodis et al. (2021) we conclude that:

$$\boldsymbol{a}_{2,1} = \frac{1}{T}\sum_{h=1}^{\infty} \mathrm{E}[\boldsymbol{\Gamma}_{i,\varepsilon,\boldsymbol{v}}(-h)'] \operatorname{tr}[\boldsymbol{\Gamma}(h)(\boldsymbol{\Gamma}(0))^{-1}] + \mathrm{o}_P(T^{-1}). \tag{S.93}$$

Notice that $\mathrm{E}[\boldsymbol{\Gamma}_{i,\varepsilon,\boldsymbol{v}}(-h)']$ is a well defined quantity, as by Assumption 4.1-4.3 each $\boldsymbol{\Gamma}_{i,\varepsilon,\boldsymbol{v}}(-h)'$ can be factorized into am individual specific component independent of h, and a common (non-stochastic) component which is absolutely summable. □

Lemma 6. *Under Assumptions 4.1-4.5 with* $\ell = 4$*:*

$$\boldsymbol{a}_4 = \frac{1}{N^2}\sum_{i=1}^{N} \boldsymbol{\Lambda}_i'(\boldsymbol{G}^{+})' \,\mathrm{E}_i[\boldsymbol{u}_{i,t}\varepsilon_{i,t}] + \mathrm{o}_P(N^{-1}) \tag{S.94}$$

Proof of Lemma 6.
Using the definition of $\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}$:

$$\begin{aligned}\boldsymbol{a}_4 &= \frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}'\boldsymbol{\varepsilon}_i - \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'(T^{-1}\overline{\boldsymbol{\Psi}}'\widetilde{\boldsymbol{F}})(T^{-1}\widetilde{\boldsymbol{F}}'\boldsymbol{\varepsilon}_i)\\ &= \boldsymbol{a}_{4,1} - \boldsymbol{a}_{4,2}.\end{aligned} \tag{S.95}$$

Using Lemma 4 it follows directly:

$$\begin{aligned}\|\boldsymbol{a}_{4,2}\| &< \frac{1}{N}\sum_{i=1}^{N} \mathcal{O}_P(1)\,\mathcal{O}_P(1)\,\mathcal{O}_P(N^{-1})\,\mathcal{O}_P(N^{-1})\\ &= \mathcal{O}_P(N^{-2}).\end{aligned} \tag{S.96}$$

As for the leading term

$$\begin{aligned}\|\boldsymbol{a}_{4,1}\| &< \frac{1}{N}\sum_{i=1}^{N} \mathcal{O}_P(1)\,\mathcal{O}_P(1)\,\mathcal{O}_P(N^{-1})\\ &= \mathcal{O}_P(N^{-1}).\end{aligned} \tag{S.97}$$

Hence, without additional restrictions this component is not asymptotically negligible. Using the decomposition for $\overline{\boldsymbol{\Psi}}$ and relevant stochastic bounds from Lemma 4, it follows that:

$$\begin{aligned}\boldsymbol{a}_{4,1} &= \frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{\Lambda}_i'(\boldsymbol{G}^{+})'\overline{\boldsymbol{U}}'\boldsymbol{\varepsilon}_i + \frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{\Lambda}_i'(\boldsymbol{G}^{+})'\boldsymbol{G}_\perp'\boldsymbol{F}_\perp'\boldsymbol{\varepsilon}_i + \mathcal{O}_P(N^{-3/2})\\ &= \boldsymbol{a}_{4,1,1} + \boldsymbol{a}_{4,1,2} + \mathcal{O}_P(N^{-3/2}).\end{aligned} \tag{S.98}$$

Using similar steps as in Lemma 5, it can be shown that:

$$\boldsymbol{a}_{4,1,2} = \frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{\Lambda}_i'(\boldsymbol{G}^{+})'\boldsymbol{G}_\perp'\boldsymbol{F}_\perp'\boldsymbol{\varepsilon}_i = \mathcal{O}_P(N^{-3/2}). \tag{S.99}$$

Finaly, the leading term in $\boldsymbol{a}_{4,1,1}$ is determined by the "own" cross-products of $\overline{\boldsymbol{U}}$ and $\boldsymbol{\varepsilon}_i$ (see e.g. Westerlund and Urbain 2015) so that:

$$
\begin{aligned}
\boldsymbol{a}_{4,1,1} &= \frac{1}{N^2}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{G}^+)'T^{-1}\sum_{t=1}^{T}\boldsymbol{u}_{i,t}\varepsilon_{i,t} + \mathcal{O}_P(N^{-3/2}) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{G}^+)'\,\mathrm{E}_i[\boldsymbol{u}_{i,t}\varepsilon_{i,t}] + \mathcal{O}_P(N^{-3/2}). \qquad \text{(S.100)}
\end{aligned}
$$

From here the conclusion of this lemma follows immediately. □

Lemma 7. *Under Assumptions 4.1-4.5 with* $\ell = 4$*:*

$$
\begin{aligned}
\boldsymbol{a}_3 &= \frac{1}{N\sqrt{N}}\sum_{i=1}^{N}(\boldsymbol{\lambda}_i'(\boldsymbol{G}^+)' \otimes \boldsymbol{\Lambda}_{i,\perp}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp})\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{vec}(\boldsymbol{C}_{i,\perp}) \\
&+ \frac{1}{N\sqrt{NT}}\sum_{i=1}^{N}(\boldsymbol{\lambda}_i'(\boldsymbol{G}^+)' \otimes \boldsymbol{\Lambda}_{i,\perp}')\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\operatorname{vec}\left(\boldsymbol{f}_{t,\perp}\boldsymbol{u}_{i,t}'\right) \\
&+ \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{E}_i[\boldsymbol{v}_{i,t}\boldsymbol{u}_{i,t}']\boldsymbol{G}^+\boldsymbol{\lambda}_i + \mathrm{o}_P(N^{-1}). \qquad \text{(S.101)}
\end{aligned}
$$

Proof of Lemma 7.
As in Lemma 6

$$
\begin{aligned}
\boldsymbol{a}_3 &= \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i - \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{\Psi}}\boldsymbol{P}_{\widetilde{\boldsymbol{F}}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i \\
&= \boldsymbol{a}_{3,1} - \boldsymbol{a}_{3,2}. \qquad \text{(S.102)}
\end{aligned}
$$

Using Lemma 4:

$$
\|\boldsymbol{a}_{3,2}\| < \frac{1}{N}\sum_{i=1}^{N}\mathcal{O}_P(N^{-1/2})\,\mathcal{O}_P(N^{-1}) = \mathcal{O}_P(N^{-3/2}). \qquad \text{(S.103)}
$$

The leading component on the other hand is of order:

$$
\|\boldsymbol{a}_{3,1}\| < \frac{1}{N}\sum_{i=1}^{N}\mathcal{O}_P(N^{-1/2}) = \mathcal{O}_P(N^{-1/2}). \qquad \text{(S.104)}
$$

Using the definition of $\boldsymbol{V}_{i,+}$ we expand $\boldsymbol{a}_{3,1}$ as:

$$
\begin{aligned}
\boldsymbol{a}_{3,1} &= \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Lambda}_{i,\perp}'\boldsymbol{F}_\perp'\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{V}_i'\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i \\
&= \boldsymbol{a}_{3,1,1} + \boldsymbol{a}_{3,1,2}. \qquad \text{(S.105)}
\end{aligned}
$$

The second term $\boldsymbol{a}_{3,1,2}$ can be handled analogously to $\boldsymbol{a}_{4,1,1}$ in Lemma 6, so that:

$$
\boldsymbol{a}_{3,1,2} = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{E}_i[\boldsymbol{v}_{i,t}\boldsymbol{u}_{i,t}']\boldsymbol{G}^+\boldsymbol{\lambda}_i + \mathcal{O}_P(N^{-3/2}). \qquad \text{(S.106)}
$$

As for $\boldsymbol{a}_{3,1,1}$, we can use derivations in Lemma 4 to isolate the leading term:

$$
\begin{aligned}
\boldsymbol{a}_{3,1,1} &= \frac{1}{N\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\Lambda}_{i,\perp}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\sqrt{N}\boldsymbol{G}_\perp\boldsymbol{G}^+\boldsymbol{\lambda}_i + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Lambda}_{i,\perp}'\boldsymbol{F}_\perp'\overline{\boldsymbol{U}}\boldsymbol{G}^+\boldsymbol{\lambda}_i + \mathcal{O}_P(N^{-3/2}) \\
&= \frac{1}{N\sqrt{N}}\sum_{i=1}^{N}(\boldsymbol{\lambda}_i'(\boldsymbol{G}^+)'\otimes\boldsymbol{\Lambda}_{i,\perp}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp})\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{vec}(\boldsymbol{C}_{i,\perp}) \\
&+ \frac{1}{N\sqrt{NT}}\sum_{i=1}^{N}(\boldsymbol{\lambda}_i'(\boldsymbol{G}^+)'\otimes\boldsymbol{\Lambda}_{i,\perp})\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\operatorname{vec}\left(\boldsymbol{f}_{t,\perp}\boldsymbol{u}_{i,t}'\right) + \mathcal{O}_P(N^{-3/2}). \qquad \text{(S.107)}
\end{aligned}
$$

Here the second equality follows from the fact that $\boldsymbol{a}_{3,1,1}$ is already a vector thus, $\boldsymbol{a}_{3,1,1} = \operatorname{vec}(\boldsymbol{a}_{3,1,1})$. □

Lemma 8. *Under Assumptions 4.1-4.5 with $\ell = 4$:*

$$
\boldsymbol{a}_5 = \frac{1}{N^2}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{G}^+)'\left(N\boldsymbol{G}_\perp'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{G}_\perp + \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_i[\boldsymbol{u}_{i,t}\boldsymbol{u}_{i,t}']\right)\boldsymbol{G}^+\boldsymbol{\lambda}_i + \mathrm{o}_P(N^{-1}) \qquad \text{(S.108)}
$$

Proof of Lemma 8.
As in Lemmas 6-7:

$$
\begin{aligned}
\boldsymbol{a}_5 &= \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}'\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i - \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}'\boldsymbol{P}_{\widetilde{\boldsymbol{F}}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i \\
&= \boldsymbol{a}_{5,1} - \boldsymbol{a}_{5,2}. \qquad \text{(S.109)}
\end{aligned}
$$

From Lemma 4 we can conclude that:

$$
\|\boldsymbol{a}_{5,2}\| < N^{-1}\sum_{i=1}^{N}\mathcal{O}_P(1)\,\mathcal{O}_P(1)\,\mathcal{O}_P(N^{-2})\,\mathcal{O}_P(1)\,\mathcal{O}_P(1) = \mathcal{O}_P(N^{-2}). \qquad \text{(S.110)}
$$

As for the first component from parts $(viii)$-(ix) of Lemma 4, it can be seen that:

$$
\frac{N}{T}(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}'\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1} = (\boldsymbol{G}^+)'\left(\frac{N}{T}\overline{\boldsymbol{U}}_e'\overline{\boldsymbol{U}}_e\right)\boldsymbol{G}^+ + \mathcal{O}_P(N^{-1/2}). \qquad \text{(S.111)}
$$

Using Lemmas A.1-A.2 in Juodis et al. (2021), one can further show that:

$$
\begin{aligned}
\frac{N}{T}\overline{\boldsymbol{U}}_e'\overline{\boldsymbol{U}}_e &= \frac{N}{T}\overline{\boldsymbol{U}}'\overline{\boldsymbol{U}} + N\boldsymbol{G}_\perp'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{G}_\perp + \mathcal{O}_P(N^{-1/2}) \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_i[\boldsymbol{u}_{i,t}\boldsymbol{u}_{i,t}'] + N\boldsymbol{G}_\perp'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{G}_\perp + \mathcal{O}_P(N^{-1/2}). \qquad \text{(S.112)}
\end{aligned}
$$

Combining these intermediate steps we conclude that:

$$
\boldsymbol{a}_5 = \frac{1}{N^2}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{G}^+)'\left(N\boldsymbol{G}_\perp'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{G}_\perp + \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_i[\boldsymbol{u}_{i,t}\boldsymbol{u}_{i,t}']\right)\boldsymbol{G}^+\boldsymbol{\lambda}_i + \mathcal{O}_P(N^{-3/2}). \qquad \text{(S.113)}
$$

□

Lemma 9. *Under Assumptions 4.1-4.5 with $\ell = 4$:*

$$\widehat{\boldsymbol{W}} = \mathrm{E}[\boldsymbol{v}_{i,t}\boldsymbol{v}_{i,t}'] + \mathrm{E}[\boldsymbol{\Lambda}_{i,\perp}'\boldsymbol{\Gamma}_{\perp,\perp}(0)\boldsymbol{\Lambda}_{i,\perp}] + \mathcal{O}_P(N^{-1/2}). \tag{S.114}$$

Proof of Lemma 9.

$$\widehat{\boldsymbol{W}} = \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{V}_{i,+}' - \boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}')\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}(\boldsymbol{V}_{i,+}' - \boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}')'. \tag{S.115}$$

Using the same arguments as in Lemmas 5-8 it is straightforward to show that:

$$\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{W}}_1 + \widehat{\boldsymbol{W}}_2 + \mathcal{O}_P(N^{-1}). \tag{S.116}$$

Here:

$$\widehat{\boldsymbol{W}}_1 = \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{V}_{i,+}'\boldsymbol{V}_{i,+} = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{E}_i[\boldsymbol{v}_{i,t}\boldsymbol{v}_{i,t}'] + \boldsymbol{\Lambda}_{i,\perp}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{\Lambda}_{i,\perp}) + \mathcal{O}_P(N^{-1}). \tag{S.117}$$

Notice that the presence of the heterogeneous expectations, factor loadings, and factors $\boldsymbol{F}_\perp$ imply that in general:

$$\widehat{\boldsymbol{W}}_1 = \mathrm{E}[\boldsymbol{v}_{i,t}\boldsymbol{v}_{i,t}'] + \mathrm{E}[\boldsymbol{\Lambda}_{i,\perp}'\,\mathrm{E}[\boldsymbol{f}_{t,\perp}\boldsymbol{f}_{t,\perp}']\boldsymbol{\Lambda}_{i,\perp}] + \mathcal{O}_P(N^{-1/2}), \tag{S.118}$$

under proportional asymptotics $N/T \to c$. Finally, the second component $\widehat{\boldsymbol{W}}_2$ is driven by a stochastic component similar in its structure to the dominant term in $\boldsymbol{a}_3$, i.e:

$$\widehat{\boldsymbol{W}}_2 = -\widehat{\boldsymbol{W}}_{2,1} - \widehat{\boldsymbol{W}}_{2,1}' + \mathcal{O}_P(N^{-1}), \tag{S.119}$$

where

$$\begin{aligned}\widehat{\boldsymbol{W}}_{2,1} &= \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}'\boldsymbol{F}_\perp\boldsymbol{\Lambda}_{i,\perp}\\ &= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{G}^{+})'\boldsymbol{G}_\perp'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{\Lambda}_{i,\perp} + \mathcal{O}_P(N^{-1})\\ &= \mathcal{O}_P(N^{-1/2}),\end{aligned} \tag{S.120}$$

using the steps analogous to those for $\boldsymbol{a}_3$ and the definition of $\boldsymbol{G}_\perp$. Combining all terms we conclude that:

$$\widehat{\boldsymbol{W}} = \mathrm{E}[\boldsymbol{v}_{i,t}\boldsymbol{v}_{i,t}'] + \mathrm{E}[\boldsymbol{\Lambda}_{i,\perp}'\boldsymbol{\Gamma}_{\perp,\perp}(0)\boldsymbol{\Lambda}_{i,\perp}] + \mathcal{O}_P(N^{-1/2}), \tag{S.121}$$

using the definition $\boldsymbol{\Gamma}_{\perp,\perp}(0) = \mathrm{E}[\boldsymbol{f}_{t,\perp}\boldsymbol{f}_{t,\perp}']$.

□

Proof of Theorem 1.
As before we note that the pooled estimator can be expanded as:

$$\widehat{\boldsymbol{\beta}}_{rCCEP} - \boldsymbol{\beta}_0 = \widehat{\boldsymbol{W}}^{-1}\left(\boldsymbol{a}_1 - (\boldsymbol{a}_2 + \boldsymbol{a}_3 + \boldsymbol{a}_4 - \boldsymbol{a}_5)\right), \tag{S.122}$$

where

$$\boldsymbol{a}_1 = (NT)^{-1}\sum_{i=1}^{N} \boldsymbol{V}_{i,+}'\boldsymbol{\varepsilon}_i \tag{S.123}$$

$$\boldsymbol{a}_2 = (NT)^{-1}\sum_{i=1}^{N} \boldsymbol{V}_{i,+}'\boldsymbol{P}_{\widetilde{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i \tag{S.124}$$

$$\boldsymbol{a}_3 = (NT)^{-1}\sum_{i=1}^{N} \boldsymbol{V}_{i,+}'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i \tag{S.125}$$

$$\boldsymbol{a}_4 = (NT)^{-1}\sum_{i=1}^{N} \boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i \tag{S.126}$$

$$\boldsymbol{a}_5 = (NT)^{-1}\sum_{i=1}^{N} \boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i \tag{S.127}$$

$$\widehat{\boldsymbol{W}} = (NT)^{-1}\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\boldsymbol{X}_i. \tag{S.128}$$

We leave $\boldsymbol{a}_1$ as it is, and use Lemmas 5- 8 to isolate the leading terms in $\boldsymbol{a}_2, \ldots, \boldsymbol{a}_5$. In particular:

$$\begin{aligned}
\sqrt{NT}(\boldsymbol{a}_2 + \boldsymbol{a}_3 + \boldsymbol{a}_4 - \boldsymbol{a}_5) = {} & \sqrt{\frac{N}{T}}\sum_{h=1}^{\infty} E[\boldsymbol{\Gamma}_{i,\varepsilon,\boldsymbol{v}}(-h)']\,\mathrm{tr}[\boldsymbol{\Gamma}(h)(\boldsymbol{\Gamma}(0))^{-1}] \\
& + \sqrt{T}\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\lambda}_i'(\boldsymbol{G}^{+})' \otimes \boldsymbol{\Lambda}_{i,\perp}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp})\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathrm{vec}\,(\boldsymbol{C}_{i,\perp}) \\
& + \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\lambda}_i'(\boldsymbol{G}^{+})' \otimes \boldsymbol{\Lambda}_{i,\perp}')\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{vec}\left(\boldsymbol{f}_{t,\perp}\boldsymbol{u}_{i,t}'\right) \\
& + \sqrt{\frac{T}{N}}\left(\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_i[\boldsymbol{v}_{i,t}\boldsymbol{u}_{i,t}']\boldsymbol{G}^{+}\boldsymbol{\lambda}_i + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{G}^{+})'\,\mathrm{E}_i[\boldsymbol{u}_{i,t}\varepsilon_{i,t}]\right) \\
& - \sqrt{\frac{T}{N}}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{G}^{+})'\left(N\boldsymbol{G}_\perp'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{G}_\perp + \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_i[\boldsymbol{u}_{i,t}\boldsymbol{u}_{i,t}']\right)\boldsymbol{G}^{+}\boldsymbol{\lambda}_i \\
& + o_P(1).
\end{aligned} \tag{S.129}$$

The result of this theorem follows upon setting $\boldsymbol{S} \equiv -\boldsymbol{G}^{+}$, and by recognizing that: the first line of the above expression contributes to $\boldsymbol{b}_1$, the second line is that of $\boldsymbol{\xi}_r$, the third line is the variance component (that together with $\boldsymbol{a}_1$) defines $\boldsymbol{b}_0$. The last two lines define the $\boldsymbol{b}_{2,r}$ bias term. Finally, Lemma 9 can be used to derive $\boldsymbol{\Sigma}_{X,r}$. □

S.5.5. Analytical Bias-Correction

Proof of Proposition S.1.

As this proof almost directly follows from Lemmas 5-9, here we only focus on the main ideas.

Notice that by construction, the estimate of the bias is of the form:

$$\widehat{\boldsymbol{b}}_{2,r} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\Lambda}}_i'(\widehat{\boldsymbol{G}}^+)'\left(\frac{1}{NT}\sum_{i=1}^{N}\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{U}}_i'\widehat{\boldsymbol{U}}_i(\widehat{\boldsymbol{\Sigma}}^{-1/2})'\right)\widehat{\boldsymbol{G}}^+\widehat{\boldsymbol{\lambda}}_i$$
$$- \frac{1}{NT}\sum_{i=1}^{N}\widehat{\boldsymbol{\Lambda}}_i'(\widehat{\boldsymbol{G}}^+)'\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{U}}_i'\widehat{\varepsilon}_i - \frac{1}{NT}\sum_{i=1}^{N}\widehat{\boldsymbol{V}}_i'\widehat{\boldsymbol{U}}_i(\widehat{\boldsymbol{\Sigma}}^{-1/2})'\widehat{\boldsymbol{G}}^+\widehat{\boldsymbol{\lambda}}_i. \tag{S.130}$$

For simplicity, in what follows, we set $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{I}$, as in all previous proofs. At first we argue why $\widehat{\boldsymbol{G}}^+\widehat{\boldsymbol{\lambda}}_i$ is consistent for $\boldsymbol{G}^+\boldsymbol{\lambda}_i$. Observe that

$$\|\widehat{\boldsymbol{G}}^+\| < \|T^{-1}\overline{\boldsymbol{Z}}'\widetilde{\boldsymbol{F}}\|\|\boldsymbol{V}_{N,T}^{-1}\| = \mathcal{O}_P(1), \tag{S.131}$$

by Lemmas 1-2. Furthermore, notice that by construction $\widehat{\boldsymbol{\lambda}}_i$ for each i:

$$\begin{aligned}\|\widehat{\boldsymbol{\lambda}}_i\| &< \|T^{-1}\widetilde{\boldsymbol{F}}'\boldsymbol{\varepsilon}_i\| + \|T^{-1}\widetilde{\boldsymbol{F}}'\boldsymbol{X}_i\|\|\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_{rCCEP}\| + \|T^{-1}\widetilde{\boldsymbol{F}}'\boldsymbol{F}\|\|\boldsymbol{\lambda}_i\| \\ &= \mathcal{O}_P(N^{-1/2}) + \mathcal{O}_P(1)\,\mathcal{O}_P(N^{-1/2}) + \mathcal{O}_P(1) \\ &= \mathcal{O}_P(1),\end{aligned} \tag{S.132}$$

as by Theorem 1 $\sqrt{N}(\widehat{\boldsymbol{\beta}}_{rCCEP} - \boldsymbol{\beta}_0) = \mathcal{O}_P(1)$. From this bound it is clear that:

$$\widehat{\boldsymbol{G}}^+\widehat{\boldsymbol{\lambda}}_i = \boldsymbol{G}'(T^{-1}\boldsymbol{F}'\widetilde{\boldsymbol{F}})\boldsymbol{V}_{N,T}^{-1}(T^{-1}\widetilde{\boldsymbol{F}}'\boldsymbol{F})\boldsymbol{\lambda}_i + \mathcal{O}_P(N^{-1/2}). \tag{S.133}$$

By Proposition 1 of Bai (2003):

$$(T^{-1}\boldsymbol{F}'\widetilde{\boldsymbol{F}})\boldsymbol{V}_{N,T}^{-1}(T^{-1}\widetilde{\boldsymbol{F}}'\boldsymbol{F}) = (\boldsymbol{G}\boldsymbol{G}')^{-1} + \mathrm{o}_P(1). \tag{S.134}$$

Using this result we conclude that

$$\widehat{\boldsymbol{G}}^+\widehat{\boldsymbol{\lambda}}_i = \boldsymbol{G}^+\boldsymbol{\lambda}_i + \mathrm{o}_P(1), \tag{S.135}$$

where the dominant term in $\mathrm{o}_P(1)$ is common for all units i. Analogous steps can be used to show that:

$$\widehat{\boldsymbol{G}}^+\widehat{\boldsymbol{\Lambda}}_i = \boldsymbol{G}^+\boldsymbol{\Lambda}_i' + \mathrm{o}_P(1). \tag{S.136}$$

Taking Eq. (S.135)-(S.136) as given, the term in the middle of the first component of $\widehat{\boldsymbol{b}}_{2,r}$ can be expanded as

$$\frac{1}{NT}\sum_{i=1}^{N}\widehat{\boldsymbol{U}}_i'\widehat{\boldsymbol{U}}_i = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{E}_i[\boldsymbol{u}_{i,t}\boldsymbol{u}_{i,t}'] + \boldsymbol{C}_{i,\perp}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{C}_{i,\perp}) + \mathcal{O}_P(N^{-1/2}), \tag{S.137}$$

directly using Lemma 9 upon changing $\boldsymbol{X}_i$ to $\boldsymbol{Z}_i$. Combining this expansion with results in Eq. (S.135)-(S.135) the desired result follows.

It remains to investigate the limits of the remaining two components of $\widehat{\boldsymbol{b}}_{2,r}$. However, as the derivations are almost identical, we only focus on the third component to outline the main steps. From Lemma 3:

$$\|T^{-1}\widehat{\boldsymbol{V}}_i'\widehat{\boldsymbol{U}}_i\| = \|T^{-1}\boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\boldsymbol{Z}_i\| = \mathcal{O}_P(1). \tag{S.138}$$

Thus, combining these facts we conclude that:

$$\frac{1}{NT}\sum_{i=1}^{N}\widehat{\boldsymbol{V}}_i'\widehat{\boldsymbol{U}}_i\widehat{\boldsymbol{G}}^{+}\widehat{\boldsymbol{\lambda}}_i = \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\boldsymbol{Z}_i\boldsymbol{G}^{+}\boldsymbol{\lambda}_i + \mathrm{o}_P(1). \tag{S.139}$$

The leading term can be then expanded as

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\boldsymbol{Z}_i\boldsymbol{G}^{+}\boldsymbol{\lambda}_i &= \frac{1}{NT}\sum_{i=1}^{N}(\boldsymbol{V}_{i,+} - \overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\Lambda}_i)'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}(\boldsymbol{E}_i - \overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{C}_i)\boldsymbol{G}^{+}\boldsymbol{\lambda}_i \\
&= \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{V}_{i,+}'\boldsymbol{E}_i\boldsymbol{G}^{+}\boldsymbol{\lambda}_i + \mathcal{O}_P(N^{-1/2}) \\
&= \frac{1}{N}\sum_{i=1}^{N}(\mathrm{E}_i[\boldsymbol{v}_{i,t}\boldsymbol{u}_{i,t}'] + \boldsymbol{\Lambda}_{i,\perp}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\boldsymbol{C}_{i,\perp})\boldsymbol{G}^{+}\boldsymbol{\lambda}_i + \mathcal{O}_P(N^{-1/2}).
\end{aligned} \tag{S.140}$$

Here the final two lines from directly from Lemmas 5-9. □

S.5.6. Auxiliary Lemmas for MG Estimator

Lemma 10. *For any matrix valued random variables* $\boldsymbol{X}_i$ *such that* $\mathrm{E}[\|\boldsymbol{X}_i\|^r]$ *for some* $r \geq 1$*, we have*

$$\sup_i \|\boldsymbol{X}_i\| = \mathcal{O}_P(N^{1/r}). \tag{S.141}$$

Proof of Lemma 10.

Observe that:

$$\begin{aligned} \sup_i \|\boldsymbol{X}_i\| &= \left(\sup_i \|\boldsymbol{X}_i\|^r\right)^{1/r} \\ &\leq \left(\sum_i \|\boldsymbol{X}_i\|^r\right)^{1/r} \\ &= \mathcal{O}_P(N^{1/r}). \end{aligned} \tag{S.142}$$

Here the first line follows by definition, the second line follows from the fact that $\sup_i \|\boldsymbol{X}_i\| \leq \sum_i \|\boldsymbol{X}_i\|$ and the Jensen's inequality. Finally, the conclusion follows from the Markov's inequality.

□

Lemma 11. *Under Assumptions 4.1-4.7 with* $\ell = 8$*:*

(i) $\sup_i \|T^{-1}\boldsymbol{V}_i'\overline{\boldsymbol{\Psi}}\| = \mathcal{O}_P(N^{-3/4})$.

(ii) $\sup_i \|T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{\Psi}}\| = \mathcal{O}_P(N^{-1/2})$.

Proof of Lemma 11.

Part (i).

From the definition of $\overline{\boldsymbol{\Psi}}$ and triangle inequality:

$$\begin{aligned} \sup_i \|T^{-1}\boldsymbol{V}_i'\overline{\boldsymbol{\Psi}}\| \leq& \sup_i \|T^{-1}\boldsymbol{V}_i'\boldsymbol{F}\| \|\boldsymbol{G}\| \|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\| \|\boldsymbol{V}_{N,T}^{-1}\| \\ &+ \sup_i \|T^{-1}\boldsymbol{V}_i'\overline{\boldsymbol{U}}_e\| \|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\| \|\boldsymbol{V}_{N,T}^{-1}\| \\ &+ \sup_i \|T^{-1}\boldsymbol{V}_i'\overline{\boldsymbol{U}}_e\| \|\boldsymbol{G}'\| \|T^{-1}\boldsymbol{F}'\widetilde{\boldsymbol{F}}\| \|\boldsymbol{V}_{N,T}^{-1}\|. \end{aligned} \tag{S.143}$$

In order to evaluate the stochastic order of the all three components it is sufficient to evaluate the two $\sup_i$ terms on the RHS of the inequality. In particular, from Assumption 4.7 and Lemma 10 it follows that:

$$\sup_i \|T^{-1}\boldsymbol{V}_i'\boldsymbol{F}\| = \mathcal{O}_P(T^{-1/2}N^{1/4}). \tag{S.144}$$

One the other hand:

$$\begin{aligned}\sup_i\|T^{-1}\boldsymbol{V}_i'\overline{\boldsymbol{U}}_e\| &\leq \sup_i\|T^{-1}\boldsymbol{V}_i'\overline{\boldsymbol{U}}\|+\sup_i\|T^{-1}\boldsymbol{V}_i'\boldsymbol{F}_\perp\boldsymbol{G}_\perp\| \\ &\leq \sup_i\|T^{-1}\boldsymbol{V}_i'\overline{\boldsymbol{U}}-\mathrm{E}[T^{-1}\boldsymbol{V}_i'\overline{\boldsymbol{U}}]\|+\sup_i\|\mathrm{E}[T^{-1}\boldsymbol{V}_i'\overline{\boldsymbol{U}}]\|+\sup_i\|T^{-1}\boldsymbol{V}_i'\boldsymbol{F}_\perp\|\|\boldsymbol{G}_\perp\| \\ &= (NT)^{-1/2}\,\mathcal{O}_P(N^{1/4})+\mathcal{O}_P(N^{-1})+\mathcal{O}_P(T^{-1/2}N^{1/4})\,\mathcal{O}_P(N^{-1/2}) \\ &= \mathcal{O}_P(N^{-1})+\mathcal{O}_P(T^{-1/2}N^{-1/4}). \end{aligned} \tag{S.145}$$

Here the fact that Assumption 4.7 is formulated in terms of $\widetilde{\boldsymbol{E}}_i$ and not in terms of $\boldsymbol{E}_i$ is inconsequential for the above conclusions as we assume that all elements of $\boldsymbol{\Omega}_i$ are bounded by some constant Δ.

Combining all expressions and together with Lemma 3 we conclude that:

$$\begin{aligned}\sup_i\|T^{-1}\boldsymbol{V}_i'\overline{\boldsymbol{\Psi}}\| &= \mathcal{O}_P(T^{-1/2}N^{-3/4})+\mathcal{O}_P(N^{-1})+\mathcal{O}_P(T^{-1/2}N^{-1/4}) \\ &= \mathcal{O}_P(N^{-1})+\mathcal{O}_P(T^{-1/2}N^{-1/4}). \end{aligned} \tag{S.146}$$

Part (ii).

The second part of this lemma follows from the fact that:

$$\begin{aligned}\sup_i\|T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{\Psi}}\| &\leq \sup_i\|T^{-1}\boldsymbol{V}_i'\overline{\boldsymbol{\Psi}}\|+\sup_i\|\boldsymbol{\Lambda}_{i,\perp}\|\|T^{-1}\boldsymbol{F}_\perp'\overline{\boldsymbol{\Psi}}\| \\ &= \mathcal{O}_P(N^{-1})+\mathcal{O}_P(T^{-1/2}N^{-1/4})+\mathcal{O}_P(N^{-1/2}). \end{aligned} \tag{S.147}$$

Here we use the fact that $\sup_i\|\boldsymbol{\Lambda}_{i,\perp}\|<\Delta$ by assumption while the order of $\|T^{-1}\boldsymbol{F}_\perp'\overline{\boldsymbol{\Psi}}\|$ can be derived similarly as in Lemma 3.

The final conclusion of this lemma follows using Assumption **??**. □

Lemma 12. *Under Assumptions 4.1-4.7 with* $\ell=8$*:*

$$\sup_i\|T^{-1}\boldsymbol{V}_{i,+}'\widetilde{\boldsymbol{F}}\|=\mathcal{O}_P(N^{-1/4}). \tag{S.148}$$

Proof of Lemma 12.

Note that

$$\begin{aligned}\sup_i\|T^{-1}\boldsymbol{V}_{i,+}'\widetilde{\boldsymbol{F}}\| &\leq \sup_i\|T^{-1}\boldsymbol{V}_{i,+}'\boldsymbol{F}\boldsymbol{H}\|+\sup_i\|T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{\Psi}}\| \\ &\leq \sup_i\|T^{-1}\boldsymbol{V}_i'\boldsymbol{F}\boldsymbol{H}\|+\sup_i\|T^{-1}\boldsymbol{\Lambda}_{i,\perp}'\boldsymbol{F}_\perp'\boldsymbol{F}\boldsymbol{H}\|+\sup_i\|T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{\Psi}}\| \\ &\leq \sup_i\|T^{-1}\boldsymbol{V}_i'\boldsymbol{F}\|\|\boldsymbol{H}\|+\mathcal{O}_P(N^{-1/2}), \end{aligned} \tag{S.149}$$

where we make use of the corresponding bounds derived in Lemmas 10-11. From here

$$\sup_i\|T^{-1}\boldsymbol{V}_{i,+}'\widetilde{\boldsymbol{F}}\|=\mathcal{O}_P(N^{-1/4}). \tag{S.150}$$

□

Lemma 13. *Under Assumptions 4.1-4.7 with* $\ell = 8$ *for* $\widehat{\boldsymbol{\Sigma}}_{X,r,i} = \frac{1}{T}\boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\boldsymbol{X}_i$*:*

(i) $\sup_i\|\widehat{\boldsymbol{\Sigma}}_{X,r,i} - \boldsymbol{\Sigma}_{X,r,i}\| = \mathcal{O}_P(N^{-1/4})$.

(ii) $\sup_i\|\widehat{\boldsymbol{\Sigma}}^{-1}_{X,r,i} - \boldsymbol{\Sigma}^{-1}_{X,r,i}\| = \mathcal{O}_P(N^{-1/4})$.

Proof of Lemma 13.
Given the definition of $\widetilde{\boldsymbol{F}}$ observe that the corresponding orthogonal projection matrix is given by $\boldsymbol{M}_{\widetilde{\boldsymbol{F}}} = \boldsymbol{I}_T - T^{-1}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}$.

Part (i).
Similarly to the proof of Lemma 9 we expand $\widehat{\boldsymbol{\Sigma}}_{X,r,i}$ as follows:

$$\widehat{\boldsymbol{\Sigma}}_{X,r,i} = \boldsymbol{\Sigma}_{X,r,i} + \boldsymbol{\Sigma}_{1,i} - (\boldsymbol{\Sigma}_{2,i} + \boldsymbol{\Sigma}_{3,i} + \boldsymbol{\Sigma}_{4,i} - \boldsymbol{\Sigma}_{5,i})) \tag{S.151}$$

$$\boldsymbol{\Sigma}_{1,i} = T^{-1}\boldsymbol{V}_{i,+}'\boldsymbol{V}_{i,+} - \boldsymbol{\Sigma}_{X,r,i} \tag{S.152}$$
$$\boldsymbol{\Sigma}_{2,i} = T^{-1}\boldsymbol{V}_{i,+}'\boldsymbol{P}_{\widetilde{\boldsymbol{F}}}\boldsymbol{V}_{i,+} \tag{S.153}$$
$$\boldsymbol{\Sigma}_{3,i} = T^{-1}\boldsymbol{V}_{i,+}'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\Lambda}_i \tag{S.154}$$
$$\boldsymbol{\Sigma}_{4,i} = T^{-1}\boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\boldsymbol{V}_{i,+} \tag{S.155}$$
$$\boldsymbol{\Sigma}_{5,i} = T^{-1}\boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\Lambda}_i. \tag{S.156}$$

In order to derive the main result of this lemma, it is sufficient to show that $\sup_i\|\cdot\|$ of all five components is of order $\mathcal{O}_P(N^{-1/4})$. We evaluate all components individually.

By definition

$$\begin{aligned}\boldsymbol{\Sigma}_{1,i} = &\left(\frac{1}{T}\sum_t \boldsymbol{v}_{i,t}\boldsymbol{v}_{i,t}' - \mathrm{E}_i[\boldsymbol{v}_{i,t}\boldsymbol{v}_{i,t}']\right) + \boldsymbol{\Lambda}_{i,\perp}'(T^{-1}\boldsymbol{F}_\perp'\boldsymbol{F}_\perp - \boldsymbol{\Gamma}_{\perp,\perp}(0))\boldsymbol{\Lambda}_{i,\perp} \\ &+ T^{-1}\boldsymbol{V}_i'\boldsymbol{F}_\perp\boldsymbol{\Lambda}_{i,\perp} + T^{-1}\boldsymbol{\Lambda}_{i,\perp}'\boldsymbol{F}_\perp'\boldsymbol{V}_i.\end{aligned} \tag{S.157}$$

Notice that by Assumption 4.7 and Lemma 10:

$$\sup_i\left\|\left(\frac{1}{T}\sum_t \boldsymbol{v}_{i,t}\boldsymbol{v}_{i,t}' - \mathrm{E}_i[\boldsymbol{v}_{i,t}\boldsymbol{v}_{i,t}']\right)\right\| = \mathcal{O}_P(T^{-1/2}N^{1/4}) = \mathcal{O}_P(N^{-1/4}). \tag{S.158}$$

Similarly using Assumption 4.7 and Lemma 10:

$$\sup_i\|T^{-1}\boldsymbol{V}_i'\boldsymbol{F}_\perp\boldsymbol{\Lambda}_{i,\perp}\| \leq \sup_i\|\boldsymbol{\Lambda}_{i,\perp}\|\sup_i\|T^{-1}\boldsymbol{V}_i'\boldsymbol{F}_\perp\| = \mathcal{O}_P(1)\,\mathcal{O}_P(T^{-1/2}N^{1/4}). \tag{S.159}$$

Finally,

$$\sup_i\|\boldsymbol{\Lambda}_{i,\perp}'(T^{-1}\boldsymbol{F}_\perp'\boldsymbol{F}_\perp - \boldsymbol{\Gamma}_{\perp,\perp}(0))\boldsymbol{\Lambda}_{i,\perp}\| \leq \sup_i\|\boldsymbol{\Lambda}_{i,\perp}\|^2\|T^{-1}\boldsymbol{F}_\perp'\boldsymbol{F}_\perp - \boldsymbol{\Gamma}_{\perp,\perp}(0)\| = \mathcal{O}_P(T^{-1/2}). \tag{S.160}$$

Combining all terms we conclude that:

$$\sup_i \|\boldsymbol{\Sigma}_{1,i}\| = \mathcal{O}_P(N^{-1/4}). \tag{S.161}$$

As for the second term, we can directly use Lemma 12 to conclude that:

$$\sup_i \|\boldsymbol{\Sigma}_{2,i}\| = \mathcal{O}_P(N^{-1/2}). \tag{S.162}$$

As for the third component, notice

$$\boldsymbol{\Sigma}_{3,i} = T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\Lambda}_i - T^{-1}\boldsymbol{V}_{i,+}'\boldsymbol{P}_{\widetilde{\boldsymbol{F}}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\Lambda}_i. \tag{S.163}$$

From here:

$$\begin{aligned} \sup_i \|\boldsymbol{\Sigma}_{3,i}\| &\leq \sup_i \|T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{\Psi}}\|\|\boldsymbol{H}\|\sup_i \|\boldsymbol{\Lambda}_i\| \\ &+ \sup_i \|T^{-1}\boldsymbol{V}_{i,+}'\widetilde{\boldsymbol{F}}\|\|T^{-1}\widetilde{\boldsymbol{F}}'\overline{\boldsymbol{\Psi}}\|\|\boldsymbol{H}\|\sup_i \|\boldsymbol{\Lambda}_i\| \\ &= \mathcal{O}_P(N^{-1/2}) + \mathcal{O}_P(N^{-1/4})\,\mathcal{O}_P(N^{-1}). \end{aligned} \tag{S.164}$$

Here we made use of Lemmas 11-12 to respectively bound the first and the second lines. Analogously we conclude that:

$$\sup_i \|\boldsymbol{\Sigma}_{4,i}\| = \mathcal{O}_P(N^{-1/2}). \tag{S.165}$$

Finally,

$$\begin{aligned} \sup_i \|\boldsymbol{\Sigma}_{5,i}\| &\leq \sup_i \|\boldsymbol{\Lambda}_i\|^2 \|T^{-1}(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\| \\ &= \mathcal{O}_P(N^{-1}) + \mathcal{O}_P(N^{-2}). \end{aligned} \tag{S.166}$$

Combining all the above results we conclude that:

$$\sup_i \|\widehat{\boldsymbol{\Sigma}}_{X,r,i} - \boldsymbol{\Sigma}_{X,r,i}\| = \mathcal{O}_P(N^{-1/4}). \tag{S.167}$$

Part (ii).

We derive the second result using the same steps as in Lemma 18 of Norkutė et al. (2021). In particular, as $\text{rk}[\boldsymbol{\Sigma}_{X,r,i}] = K$ for each i a.s. we can expand:

$$\widehat{\boldsymbol{\Sigma}}_{X,r,i}^{-1} - \boldsymbol{\Sigma}_{X,r,i}^{-1} = -\boldsymbol{\Sigma}_{X,r,i}^{-1}(\widehat{\boldsymbol{\Sigma}}_{X,r,i} - \boldsymbol{\Sigma}_{X,r,i})\boldsymbol{\Sigma}_{X,r,i}^{-1} - \boldsymbol{\Sigma}_{X,r,i}^{-1}(\widehat{\boldsymbol{\Sigma}}_{X,r,i} - \boldsymbol{\Sigma}_{X,r,i})(\widehat{\boldsymbol{\Sigma}}_{X,r,i}^{-1} - \boldsymbol{\Sigma}_{X,r,i}^{-1}). \tag{S.168}$$

Such that:

$$\begin{aligned} \sup_i \|\widehat{\boldsymbol{\Sigma}}_{X,r,i}^{-1} - \boldsymbol{\Sigma}_{X,r,i}^{-1}\| &\leq \sup_i \|\boldsymbol{\Sigma}_{X,r,i}^{-1}\|^2 \sup_i \|\widehat{\boldsymbol{\Sigma}}_{X,r,i} - \boldsymbol{\Sigma}_{X,r,i}\| \\ &+ \sup_i \|\boldsymbol{\Sigma}_{X,r,i}^{-1}\| \sup_i \|\widehat{\boldsymbol{\Sigma}}_{X,r,i} - \boldsymbol{\Sigma}_{X,r,i}\| \sup_i \|\widehat{\boldsymbol{\Sigma}}_{X,r,i}^{-1} - \boldsymbol{\Sigma}_{X,r,i}^{-1}\| \end{aligned} \tag{S.169}$$

Note that as $\sup_i \|\boldsymbol{\Sigma}_{X,r,i}^{-1}\| = \mathcal{O}_P(1)$ and given result (i), we know that the second term cannot be the leading term. From here:

$$\sup_i \|\widehat{\boldsymbol{\Sigma}}_{X,r,i}^{-1} - \boldsymbol{\Sigma}_{X,r,i}^{-1}\| = \mathcal{O}_P(N^{-1/4}). \tag{S.170}$$

□

Lemma 14. *Under Assumptions 4.1-4.7 with $\ell = 8$:*

$$\frac{1}{N}\sum_i \|T^{-1}\boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}(\boldsymbol{\varepsilon}_i + \boldsymbol{F}\boldsymbol{\lambda}_i)\| = \mathcal{O}_P(N^{-1/2}). \tag{S.171}$$

Proof of Lemma 14.
Observe that

$$\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}(\boldsymbol{\varepsilon}_i + \boldsymbol{F}\boldsymbol{\lambda}_i) = \boldsymbol{a}_{1,i} - (\boldsymbol{a}_{2,i} + \boldsymbol{a}_{3,i} + \boldsymbol{a}_{4,i} - \boldsymbol{a}_{5,i}), \tag{S.172}$$

where all $\boldsymbol{a}_{\cdot,i}$'s are defined in the proof of Theorem S.1. The proof is analogous to that of Lemma 13, except for the fact that no $\sup_i$ operator are involved. Thus, the leading terms should be scaled by a factor of $N^{-1/4}$. □

S.5.7. rCCE-MG Estimator

Proof of Theorem S.1.
As with the pooled estimator, for each unit-specific estimator can be expanded

$$\widehat{\boldsymbol{\beta}}_{rCCE,i} - \boldsymbol{\beta}_i = \widehat{\boldsymbol{W}}_i^{-1}\left(\boldsymbol{a}_{1,i} - (\boldsymbol{a}_{2,i} + \boldsymbol{a}_{3,i} + \boldsymbol{a}_{4,i} - \boldsymbol{a}_{5,i})\right), \tag{S.173}$$

where

$$\boldsymbol{a}_{1,i} = T^{-1}\boldsymbol{V}_{i,+}'\boldsymbol{\varepsilon}_i \tag{S.174}$$
$$\boldsymbol{a}_{2,i} = T^{-1}\boldsymbol{V}_{i,+}'\boldsymbol{P}_{\widetilde{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i \tag{S.175}$$
$$\boldsymbol{a}_{3,i} = T^{-1}\boldsymbol{V}_{i,+}'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i \tag{S.176}$$
$$\boldsymbol{a}_{4,i} = T^{-1}\boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i \tag{S.177}$$
$$\boldsymbol{a}_{5,i} = T^{-1}\boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i \tag{S.178}$$
$$\widehat{\boldsymbol{W}}_i = T^{-1}\boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\boldsymbol{X}_i. \tag{S.179}$$

Upon closer inspection of the corresponding proofs in Lemmas 5-9, one can see that the resulting rates of convergence are also applicable to the individual specific quantities. In particular, notice that e.g. $\boldsymbol{a}_2 = N^{-1}\sum_{i=1}^{N}\boldsymbol{a}_{2,i}$, thus:

$$\boldsymbol{a}_{2,i} = \mathcal{O}_P(T^{-1}), \tag{S.180}$$

$$\boldsymbol{a}_{3,i} = \frac{1}{\sqrt{N}}(\boldsymbol{\lambda}_i'(\boldsymbol{G}^{+})' \otimes \boldsymbol{\Lambda}_{i,\perp}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp})\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{vec}(\boldsymbol{C}_{i,\perp}) + \mathcal{O}_P(N^{-1}), \tag{S.181}$$

$$\boldsymbol{a}_{4,i} = \mathcal{O}_P(N^{-1}), \tag{S.182}$$

$$\boldsymbol{a}_{5,i} = \mathcal{O}_P(N^{-1}), \tag{S.183}$$

$$\widehat{\boldsymbol{W}}_i = \mathrm{E}_i[\boldsymbol{v}_{i,t}\boldsymbol{v}_{i,t}'] + \boldsymbol{\Lambda}_{i,\perp}'\boldsymbol{\Gamma}_{\perp,\perp}(0)\boldsymbol{\Lambda}_{i,\perp} + \mathcal{O}_P(N^{-1/2}). \tag{S.184}$$

Combining these results, the final conclusion of this theorem follows directly, as $\mathcal{O}_P(\sqrt{T}/N) = \mathrm{o}_P(1)$ by Assumption 4.5. □

Proof of Theorem 2.

This proof follows closely the proof of Theorem 5 in Norkutė et al. (2021). From the definition of the rCCE-MG estimator:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{rCCEMG} - \boldsymbol{\beta}_0 &= \frac{1}{N}\sum_i \widetilde{\boldsymbol{\beta}}_i + \frac{1}{N}\sum_i \widehat{\boldsymbol{\Sigma}}_{X,r,i}^{-1}\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}(\boldsymbol{\varepsilon}_i + \boldsymbol{F}\boldsymbol{\lambda}_i)\\ &= \frac{1}{N}\sum_i \widetilde{\boldsymbol{\beta}}_i + \frac{1}{N}\sum_i \boldsymbol{\Sigma}_{X,r,i}^{-1}\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}(\boldsymbol{\varepsilon}_i + \boldsymbol{F}\boldsymbol{\lambda}_i)\\ &+ \frac{1}{N}\sum_i\left(\widehat{\boldsymbol{\Sigma}}_{X,r,i}^{-1} - \boldsymbol{\Sigma}_{X,r,i}^{-1}\right)\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}(\boldsymbol{\varepsilon}_i + \boldsymbol{F}\boldsymbol{\lambda}_i).\end{aligned} \tag{S.185}$$

Note that as for each i $\boldsymbol{\Sigma}_{X,r,i}^{-1}$ is just a positive definite random matrix (by assumption also uniformly bounded by some Δ), steps similar to those used in the proof of Theorem 1 can be used to directly conclude that:

$$\begin{aligned}\sqrt{N}\frac{1}{N}\sum_i \boldsymbol{\Sigma}_{X,r,i}^{-1}\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}(\boldsymbol{\varepsilon}_i + \boldsymbol{F}\boldsymbol{\lambda}_i) &= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{X,r,i}^{-1}\left(\boldsymbol{\lambda}_i' \otimes \boldsymbol{\Lambda}_{i,\perp}'\right)\left(\boldsymbol{S}' \otimes \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\operatorname{vec}(\boldsymbol{C}_{i,\perp}) + \mathrm{o}_P(1)\\ &= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{X,r,i}^{-1}\boldsymbol{\xi}_{r,i} + \mathrm{o}_P(1).\end{aligned} \tag{S.186}$$

As for the remaining component note that:

$$\begin{aligned}\|\frac{1}{N}\sum_i\left(\widehat{\boldsymbol{\Sigma}}_{X,r,i}^{-1} - \boldsymbol{\Sigma}_{X,r,i}^{-1}\right)\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}(\boldsymbol{\varepsilon}_i + \boldsymbol{F}\boldsymbol{\lambda}_i)\| &\leq \frac{1}{N}\sum_i\|\left(\widehat{\boldsymbol{\Sigma}}_{X,r,i}^{-1} - \boldsymbol{\Sigma}_{X,r,i}^{-1}\right)\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}(\boldsymbol{\varepsilon}_i + \boldsymbol{F}\boldsymbol{\lambda}_i)\|\\ &\leq \left(\frac{1}{N}\sum_i\|\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}(\boldsymbol{\varepsilon}_i + \boldsymbol{F}\boldsymbol{\lambda}_i)\|\right)\sup_i\|\widehat{\boldsymbol{\Sigma}}_{X,r,i}^{-1} - \boldsymbol{\Sigma}_{X,r,i}^{-1}\|\\ &= \mathcal{O}_P(N^{-1/2})\,\mathcal{O}_P(N^{-1/4})\\ &= \mathrm{o}_P(N^{-1/2}),\end{aligned} \tag{S.187}$$

where we use the result from Lemma 13 to bound the $\sup_i$ term. We can then combine all intermediate results to conclude that:

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{rCCEMG} - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{N}} \sum_i \widetilde{\boldsymbol{\beta}}_i + \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Sigma}_{X,r,i}^{-1} \boldsymbol{\xi}_{r,i} + \mathrm{o}_P(1). \tag{S.188}$$

□

S.5.8. Fixed T Asymptotic Theory

Lemma 15. *Under Assumptions S.2.1-S.2.3:*

(i) $\|T^{-1}\boldsymbol{V}_{i,+}'\widetilde{\boldsymbol{F}}\|= \mathcal{O}_P(1)$,

(ii) $\|T^{-1}\boldsymbol{\varepsilon}_i'\widetilde{\boldsymbol{F}}\|= \mathcal{O}_P(1)$,

(iii) $\|T^{-1}\overline{\boldsymbol{U}}_e'\widetilde{\boldsymbol{F}}\|= \mathcal{O}_P(N^{-1/2})$,

(iv) $\|T^{-1}\boldsymbol{V}_{i,+}'\overline{\boldsymbol{\Psi}}\|= \mathcal{O}_P(N^{-1/2})$,

(v) $\|T^{-1}\boldsymbol{\varepsilon}_i'\overline{\boldsymbol{\Psi}}\|= \mathcal{O}_P(N^{-1/2})$,

(vi) $\|T^{-1}\boldsymbol{F}'\overline{\boldsymbol{\Psi}}\|= \mathcal{O}_P(N^{-1/2})$,

(vii) $\|T^{-1}\widetilde{\boldsymbol{F}}'\overline{\boldsymbol{\Psi}}\|= \mathcal{O}_P(N^{-1/2})$,

(viii) $\|T^{-1}\overline{\boldsymbol{U}}_e'\overline{\boldsymbol{\Psi}}\|= \mathcal{O}_P(N^{-1})$,

(ix) $\|T^{-1}\overline{\boldsymbol{\Psi}}'\overline{\boldsymbol{\Psi}}\|= \mathcal{O}_P(N^{-1})$.

(x) $\|T^{-1/2}\overline{\boldsymbol{\Psi}}\|= \mathcal{O}_P(N^{-1/2})$.

Proof of Lemma 15.
The proof is analogous to that of Lemma 4. □

Lemma 16. *Under Assumptions S.2.1-S.2.3:*

$$\|\boldsymbol{P}_{\widetilde{\boldsymbol{F}}} - \boldsymbol{P}_{\boldsymbol{F}}\|= \mathcal{O}_P(N^{-1/2}). \tag{S.189}$$

Proof of Lemma 16.
As in the proof of Lemma 5 we denote by $\overline{\boldsymbol{\Phi}} = \overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}$. Notice that we can expand the projection matrix $\boldsymbol{P}_{\widetilde{\boldsymbol{F}}}$ as follows:

$$\begin{aligned}\boldsymbol{P}_{\widetilde{\boldsymbol{F}}} &= \left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right)\left(\left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right)'\left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right)\right)^{-1}\left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right)' \\ &= \boldsymbol{P}_{\boldsymbol{F}} + \widetilde{\boldsymbol{F}}\boldsymbol{H}^{-1}\boldsymbol{\Delta}(\widetilde{\boldsymbol{F}}\boldsymbol{H}^{-1})' + \widetilde{\boldsymbol{F}}\boldsymbol{H}^{-1}(\boldsymbol{F}'\boldsymbol{F})^{-1}\overline{\boldsymbol{\Phi}}' + \overline{\boldsymbol{\Phi}}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}',\end{aligned} \tag{S.190}$$

where $\boldsymbol{\Delta} = \left(\left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right)' \left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right)\right)^{-1} - (\boldsymbol{F}'\boldsymbol{F})^{-1}$. Using Lemma 15, it follows that:

$$\frac{1}{T}\left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right)' \left(\boldsymbol{F} + \overline{\boldsymbol{\Phi}}\right) - \frac{1}{T}\boldsymbol{F}'\boldsymbol{F} = \mathcal{O}_P(N^{-1/2}), \tag{S.191}$$

thus by the Continuous Mapping Theorem (CMT) we conclude that $T\boldsymbol{\Delta} = \mathcal{O}_P(N^{-1/2})$. Notice that the convergence rate is slower than in the large N,T case. From here:

$$\begin{aligned} \|\boldsymbol{P}_{\widetilde{\boldsymbol{F}}} - \boldsymbol{P}_{\boldsymbol{F}}\| &< \|\boldsymbol{H}^{-1}\|^2\|\widetilde{\boldsymbol{F}}\|\|\boldsymbol{\Delta}\|\|\widetilde{\boldsymbol{F}}'\| + \|\boldsymbol{H}^{-1}\|\|\widetilde{\boldsymbol{F}}\|\|(\boldsymbol{F}'\boldsymbol{F})^{-1}\|\|\overline{\boldsymbol{\Phi}}'\| + \|\overline{\boldsymbol{\Phi}}\|\|(\boldsymbol{F}'\boldsymbol{F})^{-1}\|\|\boldsymbol{F}'\| \\ &= \mathcal{O}_P(N^{-1/2}) + \mathcal{O}_P(N^{-1/2}) + \mathcal{O}_P(N^{-1/2}), \end{aligned} \tag{S.192}$$

where we use that fact that $\|T^{-1/2}\overline{\boldsymbol{\Psi}}\| = \mathcal{O}_P(N^{-1/2})$ also implies that $\|\overline{\boldsymbol{\Psi}}\| = \mathcal{O}_P(N^{-1/2})$.v

□

Proof of Proposition 2.
The proof of this proposition follows identical step to those of Proposition 1. For the ease of exposition, we highlight only the main differences between the two proofs.

Denote by $\widehat{\boldsymbol{B}} = T^{-1}\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}$ and the corresponding fixed T probability limit as $\boldsymbol{B} = \boldsymbol{C}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}}\boldsymbol{C}$. Analogously to Lemma 3 we can show that $\widehat{\nu}_r \xrightarrow{p} \nu_r$ for all $r = 1, \ldots, K_z$, where ν_r are the eigenvalues of $\boldsymbol{B}$. Thus for all $1 \leq r < R$ the eigenvalue ratio is asymptotically bounded

$$ER(r) \xrightarrow{p} \frac{\nu_r}{\nu_{r+1}} < \infty. \tag{S.193}$$

It remains to be shown that:

$$ER(R) \xrightarrow{p} \infty, \tag{S.194}$$

and

$$ER(r) = \mathcal{O}_P(1), \quad r = R+1, \ldots, K_z - 1. \tag{S.195}$$

By assumption matrix $\boldsymbol{B}$ is of rank R almost surely and admits the eigen-decomposition of the form:

$$\boldsymbol{B} = \boldsymbol{Q}_{\boldsymbol{B}}\boldsymbol{V}\boldsymbol{Q}_{\boldsymbol{B}}'. \tag{S.196}$$

Upon direct evaluation of $\widehat{\boldsymbol{B}}$ (note here simple cross-sectional CLT are applicable):

$$\boldsymbol{Q}_{\boldsymbol{B},1}'\sqrt{N}(\widehat{\boldsymbol{B}} - \boldsymbol{B})\boldsymbol{Q}_{\boldsymbol{B},2} = \boldsymbol{Q}_{\boldsymbol{B},1}'\boldsymbol{C}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}}\sqrt{N}(\boldsymbol{G} - \boldsymbol{C})\boldsymbol{Q}_{\boldsymbol{B},2} + o_P(1), \tag{S.197}$$

and

$$\begin{aligned} \boldsymbol{Q}_{\boldsymbol{B},2}'N(\widehat{\boldsymbol{B}} - \boldsymbol{B})\boldsymbol{Q}_{\boldsymbol{B},2} &= \boldsymbol{Q}_{\boldsymbol{B},2}'\left(\frac{N}{T}\overline{\boldsymbol{U}}_e'\overline{\boldsymbol{U}}_e + \sqrt{N}(\boldsymbol{G} - \boldsymbol{C})'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}}\sqrt{N}(\boldsymbol{G} - \boldsymbol{C})\right)\boldsymbol{Q}_{\boldsymbol{B},2} + o_P(1), \\ &= \boldsymbol{Q}_{\boldsymbol{B},2}'\left(\frac{N}{T}\overline{\boldsymbol{U}}'\overline{\boldsymbol{U}} + \sqrt{N}(\boldsymbol{G} - \boldsymbol{C})'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}}\sqrt{N}(\boldsymbol{G} - \boldsymbol{C})\right)\boldsymbol{Q}_{\boldsymbol{B},2} \\ &+ \boldsymbol{Q}_{\boldsymbol{B},2}'\left(\sqrt{N}(\boldsymbol{G}_{\perp})'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_{\perp}}(0)\sqrt{N}(\boldsymbol{G}_{\perp})\right)\boldsymbol{Q}_{\boldsymbol{B},2}. \end{aligned} \tag{S.198}$$

As by assumption $|\boldsymbol{V}_R|> 0$, then for $r = R+1,\dots,K_z$:

$$
\begin{aligned}
0 &= |\boldsymbol{V}_R|^{-1}|\widehat{\boldsymbol{S}}(\widehat{\nu}_r)|= |\boldsymbol{V}_R|^{-1}|(\boldsymbol{Q}_{\boldsymbol{B},1},\sqrt{N}\boldsymbol{Q}_{\boldsymbol{B},2})'\widehat{\boldsymbol{S}}(\widehat{\nu}_r)(\boldsymbol{Q}_{\boldsymbol{B},1},\sqrt{N}\boldsymbol{Q}_{\boldsymbol{B},2})| \\
&= |\boldsymbol{Q}'_{\boldsymbol{B},2}\left(\frac{N}{T}\overline{\boldsymbol{U}}'\overline{\boldsymbol{U}} + N(\boldsymbol{G}_\perp)'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}_\perp}(\boldsymbol{G}_\perp) + N(\boldsymbol{G}-\boldsymbol{C})'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}}^{1/2}\boldsymbol{M}(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}}^{1/2})'(\boldsymbol{G}-\boldsymbol{C})\right)\boldsymbol{Q}_{\boldsymbol{B},2} - N\widehat{\nu}_r\boldsymbol{I}| \\
&\quad + o_P(1). \qquad \text{(S.199)}
\end{aligned}
$$

Here an $[R\times R]$ matrix $\boldsymbol{M}$ is defined as:

$$
\begin{aligned}
\boldsymbol{M} &= \boldsymbol{I}_R - (\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}}^{1/2})'\boldsymbol{C}'\boldsymbol{Q}_{\boldsymbol{B},1}\boldsymbol{V}_R^{-1}\boldsymbol{Q}'_{\boldsymbol{B},1}\boldsymbol{C}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}}^{1/2} \\
&= \boldsymbol{I}_R - (\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}}^{1/2})'\boldsymbol{C}'\left(\boldsymbol{C}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}}\boldsymbol{C}'\right)^{+}\boldsymbol{C}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}}^{1/2}. \qquad \text{(S.200)}
\end{aligned}
$$

Notice that $\boldsymbol{M}$ is an idempotent (thus positive semi-definite) projection matrix.

Finally, as the limiting random variable $N\overline{\boldsymbol{U}}'\overline{\boldsymbol{U}} \xrightarrow{d} \boldsymbol{\Xi}'_T\boldsymbol{\Xi}_T$ has rank of at least $K_z - R$ (by assumption), the quadratic form $\boldsymbol{Q}'_{\boldsymbol{B},2}N\overline{\boldsymbol{U}}'\overline{\boldsymbol{U}}\boldsymbol{Q}_{\boldsymbol{B},2}$ is also positive definite in the limit, while all stochastic quadratic forms inside the determinant are positive semi-definite, the whole matrix inside the determinant is positive definite. As a result, the asymptotic distribution of $N\widehat{\nu}_r$ for all $r = R+1,\dots,K_z$ is non-degenerate, as otherwise the determinant cannot be zero for some of those r. From there the conclusion analogous to that of Proposition 1 follows. □

Proof of Theorem 3.
As with Theorem 1 we first expand the regularized CCE estimator as follows:

$$
\widehat{\boldsymbol{\beta}}_{rCCEP} - \boldsymbol{\beta}_0 = \widehat{\boldsymbol{W}}^{-1}\left(\boldsymbol{a}_1 - \boldsymbol{a}_2 - \boldsymbol{a}_3 - \boldsymbol{a}_4 + \boldsymbol{a}_5\right), \qquad \text{(S.201)}
$$

where

$$
\boldsymbol{a}_1 = (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{V}'_{i,+}\boldsymbol{\varepsilon}_i, \qquad \text{(S.202)}
$$

$$
\boldsymbol{a}_2 = (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{V}'_{i,+}\boldsymbol{P}_{\widetilde{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i, \qquad \text{(S.203)}
$$

$$
\boldsymbol{a}_3 = (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{V}'_{i,+}\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i, \qquad \text{(S.204)}
$$

$$
\boldsymbol{a}_4 = (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\Lambda}'_i(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\boldsymbol{\varepsilon}_i, \qquad \text{(S.205)}
$$

$$
\boldsymbol{a}_5 = (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\Lambda}'_i(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}'\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i, \qquad \text{(S.206)}
$$

$$
\widehat{\boldsymbol{W}} = (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{X}'_i\boldsymbol{M}_{\widetilde{\boldsymbol{F}}}\boldsymbol{X}_i. \qquad \text{(S.207)}
$$

We leave $\boldsymbol{a}_1$ as it is and analyse all remaining components individually.

(i) $\boldsymbol{a}_2$*.*

Notice that we can expand this component as follows:

$$\begin{aligned}\boldsymbol{a}_2 &= (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{V}_{i,+}'\boldsymbol{P}_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i + (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{V}_{i,+}'(\boldsymbol{P}_{\widetilde{\boldsymbol{F}}} - \boldsymbol{P}_{\boldsymbol{F}})\boldsymbol{\varepsilon}_i \\ &= \boldsymbol{a}_{21} + \boldsymbol{a}_{22}. \end{aligned} \tag{S.208}$$

Note that using the vec($\cdot$) operator we can express $\boldsymbol{a}_{22}$ as:

$$\begin{aligned}\boldsymbol{a}_{22} &= \left((NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i' \otimes \boldsymbol{V}_{i,+}'\right)\mathrm{vec}(\boldsymbol{P}_{\widetilde{\boldsymbol{F}}} - \boldsymbol{P}_{\boldsymbol{F}}) \\ &= \mathcal{O}_P(N^{-1/2})\,\mathcal{O}_P(N^{-1/2}). \end{aligned} \tag{S.209}$$

Here the first $\mathcal{O}_P(N^{-1/2})$ term follows from the strict exogeneity assumption between $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{V}_i$ as well as independence between $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\Lambda}_{i,\perp}$ such that $\mathrm{E}[\boldsymbol{\varepsilon}_i' \otimes \boldsymbol{V}_{i,+}'] = \mathbf{O}$. The second $\mathcal{O}_P(N^{-1/2})$ follows directly from Lemma 16.

(ii) $\boldsymbol{a}_3$*.*

Using similar steps as above

$$\begin{aligned}\boldsymbol{a}_3 &= (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{V}_{i,+}'\boldsymbol{M}_{\boldsymbol{F}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i + (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{V}_{i,+}'(\boldsymbol{M}_{\widetilde{\boldsymbol{F}}} - \boldsymbol{M}_{\boldsymbol{F}})\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i \\ &= \boldsymbol{a}_{31} + \boldsymbol{a}_{32}. \end{aligned} \tag{S.210}$$

At first we show that $\boldsymbol{a}_{32}$ is asymptotically negligible. In particular:

$$\begin{aligned}\boldsymbol{a}_{32} &= \left((NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i' \otimes \boldsymbol{V}_{i,+}'\right)\mathrm{vec}((\boldsymbol{M}_{\widetilde{\boldsymbol{F}}} - \boldsymbol{M}_{\boldsymbol{F}})\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}) \\ &= \left((NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i' \otimes \boldsymbol{V}_{i,+}'\right)\mathrm{vec}((\boldsymbol{P}_{\boldsymbol{F}} - \boldsymbol{P}_{\widetilde{\boldsymbol{F}}})\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}) \\ &= \mathcal{O}_P(1)\,\mathcal{O}_P(N^{-1/2})\,\mathcal{O}_P(N^{-1/2}). \end{aligned} \tag{S.211}$$

Here the first $\mathcal{O}_P(1)$ term can be directly established by the Chebyshev's inequality, while the $\mathcal{O}_P(N^{-1/2})$ components follow from Lemmas 15 and 16 respectively.

Now expand $\boldsymbol{a}_{31}$ using the definition of $\boldsymbol{V}_{i,+}$ as follows:

$$\boldsymbol{a}_{31} = (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{V}_i'\boldsymbol{M}_{\boldsymbol{F}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i + (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\Lambda}_{i,\perp}'\boldsymbol{F}_\perp'\boldsymbol{M}_{\boldsymbol{F}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i \tag{S.212}$$

$$= \mathcal{O}_P(N^{-1/2})\,\mathcal{O}_P(N^{-1/2}) + \left((NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i' \otimes \boldsymbol{\Lambda}_{i,\perp}'\right)\mathrm{vec}(\boldsymbol{F}_\perp'\boldsymbol{M}_{\boldsymbol{F}}\overline{\boldsymbol{\Psi}}\boldsymbol{H}^{-1}). \tag{S.213}$$

Here the first $\mathcal{O}_P(N^{-1/2})$ term can be established from the zero-mean condition $\mathrm{E}[\boldsymbol{\lambda}_i' \otimes \boldsymbol{V}_i'] = \mathbf{O}$, while the second one follows from $\|\overline{\boldsymbol{\Psi}}\| = \mathcal{O}_P(N^{-1/2})$ as implied by Lemma 16. As we allow all factor loadings to be arbitrarily correlated, in general:

$$\left((NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i' \otimes \boldsymbol{\Lambda}_{i,\perp}'\right) = \mathcal{O}_P(1). \tag{S.214}$$

As for the remaining component $\boldsymbol{F}_\perp' \boldsymbol{M}_{\boldsymbol{F}} \overline{\boldsymbol{\Psi}} \boldsymbol{H}^{-1}$ we notice that from the definition of $\overline{\boldsymbol{\Psi}}$ and $\boldsymbol{H}$ it follows that:

$$\begin{aligned}
\boldsymbol{F}_\perp' \boldsymbol{M}_{\boldsymbol{F}} \overline{\boldsymbol{\Psi}} \boldsymbol{H}^{-1} &= \frac{1}{T}\boldsymbol{F}_\perp' \boldsymbol{M}_{\boldsymbol{F}} \overline{\boldsymbol{U}}_e \boldsymbol{G}^{+} + \mathcal{O}_P(N^{-1}) \\
&= \frac{1}{T}\boldsymbol{F}_\perp' \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{F}_\perp \boldsymbol{G}_\perp \boldsymbol{G}^{+} + \frac{1}{T}\boldsymbol{F}_\perp' \boldsymbol{M}_{\boldsymbol{F}} \overline{\boldsymbol{U}} \boldsymbol{G}^{+} + \mathcal{O}_P(N^{-1}).
\end{aligned}$$

Combining all terms:

$$\begin{aligned}
\boldsymbol{a}_3 = &\left((NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i' \otimes \boldsymbol{\Lambda}_{i,\perp}'\right) \mathrm{vec}\left(\frac{1}{T}\boldsymbol{F}_\perp' \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{F}_\perp \boldsymbol{G}_\perp \boldsymbol{G}^{+}\right) \\
&+ \left((NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i' \otimes \boldsymbol{\Lambda}_{i,\perp}'\right) \mathrm{vec}\left(\frac{1}{T}\boldsymbol{F}_\perp' \boldsymbol{M}_{\boldsymbol{F}} \overline{\boldsymbol{U}} \boldsymbol{G}^{+}\right) \\
&+ \mathcal{O}_P(N^{-1}).
\end{aligned} \tag{S.215}$$

Note that for T fixed both components are of the same asymptotic order, and thus contributed to the asymptotic distribution of the rCCEP estimator. However, as implied by independence assumption between the idiosyncratic components and the factor loadings, the two components are asymptotically independent.

(iii) $\boldsymbol{a}_4$.
As in previous derivations:

$$\begin{aligned}
\boldsymbol{a}_4 &= (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}' \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{\varepsilon}_i + (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\Lambda}_i'(\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}' (\boldsymbol{M}_{\widetilde{\boldsymbol{F}}} - \boldsymbol{M}_{\boldsymbol{F}}) \boldsymbol{\varepsilon}_i \\
&= \boldsymbol{a}_{41} + \boldsymbol{a}_{42}.
\end{aligned} \tag{S.216}$$

Using the result for $\boldsymbol{a}_{32}$ one can immediately get a conservative bound of $\boldsymbol{a}_{42} = \mathcal{O}_P(N^{-1})$, while for $\boldsymbol{a}_{41}$ we note that:

$$\boldsymbol{a}_{41} = \left((NT)^{-1}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i' \otimes \boldsymbol{\Lambda}_i'\right) \mathrm{vec}\left((\boldsymbol{H}^{-1})'\overline{\boldsymbol{\Psi}}' \boldsymbol{M}_{\boldsymbol{F}}\right) \tag{S.217}$$

$$= \mathcal{O}_P(N^{-1/2})\,\mathcal{O}_P(N^{-1/2}). \tag{S.218}$$

Combining all expressions we conclude that:

$$\boldsymbol{a}_4 = \mathcal{O}_P(N^{-1}). \tag{S.219}$$

(iv) $\boldsymbol{a}_5$.

Notice that:

$$\boldsymbol{a}_5 = \left((NT)^{-1} \sum_{i=1}^{N} \boldsymbol{\lambda}_i' \otimes \boldsymbol{\Lambda}_i' \right) \operatorname{vec}\left((\boldsymbol{H}^{-1})' \overline{\boldsymbol{\Psi}}' \boldsymbol{M}_{\widetilde{\boldsymbol{F}}} \overline{\boldsymbol{\Psi}} \boldsymbol{H}^{-1} \right). \tag{S.220}$$

Based on Lemma 15 the term inside the $\operatorname{vec}(\cdot)$ is $\mathcal{O}_P(N^{-1})$. Thus

$$\boldsymbol{a}_5 = \mathcal{O}_P(N^{-1}). \tag{S.221}$$

All $\boldsymbol{a}$ *components together.*

Next we combine all $\boldsymbol{a}$'s together:

$$\begin{aligned}
\boldsymbol{a} &= \boldsymbol{a}_1 - \boldsymbol{a}_2 - \boldsymbol{a}_3 - \boldsymbol{a}_4 + \boldsymbol{a}_5 \\
&= (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{V}_{i,+}' \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{\varepsilon}_i \\
&- \left((NT)^{-1} \sum_{i=1}^{N} \boldsymbol{\lambda}_i' \otimes \boldsymbol{\Lambda}_{i,\perp}' \right) \operatorname{vec}\left(\frac{1}{T} \boldsymbol{F}_\perp' \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{F}_\perp \boldsymbol{G}_\perp \boldsymbol{G}^+ \right) \\
&- \left((NT)^{-1} \sum_{i=1}^{N} \boldsymbol{\lambda}_i' \otimes \boldsymbol{\Lambda}_{i,\perp}' \right) \operatorname{vec}\left(\frac{1}{T} \boldsymbol{F}_\perp' \boldsymbol{M}_{\boldsymbol{F}} \overline{\boldsymbol{U}} \boldsymbol{G}^+ \right) \\
&+ \mathcal{O}_P(N^{-1}).
\end{aligned}$$

The Hessian matrix $\widehat{\boldsymbol{W}}$.

We can use all previous derivations to immediately conclude that:

$$\begin{aligned}
\widehat{\boldsymbol{W}} &= (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{X}_i + \mathcal{O}_P(N^{-1/2}) \\
&= (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{V}_{i,+}' \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{V}_{i,+} + \mathcal{O}_P(N^{-1/2}).
\end{aligned} \tag{S.222}$$

Moreover, as $\boldsymbol{V}_i$ are uncorrelated from $\boldsymbol{\Lambda}_{i,\perp}$, the above expression can be further simplified as follows:

$$\widehat{\boldsymbol{W}} = (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{V}_i' \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{V}_i + (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{\Lambda}_{i,\perp}' \boldsymbol{F}_\perp' \boldsymbol{M}_{\boldsymbol{F}} \boldsymbol{F}_\perp \boldsymbol{\Lambda}_{i,\perp} + \mathcal{O}_P(N^{-1/2}). \tag{S.223}$$

The final conclusion is established by means of the conditional (on $\boldsymbol{F}$ and $\boldsymbol{F}_\perp$) Law of Large Numbers. □

References

Ahn, S. C. and A. R. Horenstein (2013): "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81, 1203–1227.

Ando, T. and J. Bai (2015): "A Simple New Test for Slope Homogeneity in Panel Data Models with Interactive Effects," *Economics Letters*, 136, 112–117.

Bai, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171.

——— (2009): "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279.

De Vos, I. and J. Westerlund (2019): "On CCE Estimation of Factor-augmented Models when Regressors are not Linear in the Factors," *Economics Letters*, 178, 5 – 7.

Hahn, J., G. Kuersteiner, and M. Mazzocco (2020): "Central Limit Theory for Combined Cross-Section and Time Series," *Econometric Theory*, (forthcoming).

Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory and Its Application*, Probability and Mathematical Statistics, Academic Press.

Juodis, A., H. Karabiyik, and J. Westerlund (2021): "On the Robustness of the Pooled CCE Estimator," *Journal of Econometrics*, 220, 325–348.

Juodis, A. and S. Reese (2021): "The Incidental Parameters Problem in Testing for Remaining Cross-section Correlation," *Journal of Business and Economic Statistics*, (forthcoming).

Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011): "Panels with Non-stationary Multifactor Error Structures," *Journal of Econometrics*, 160, 326 – 348.

Karabiyik, H., S. Reese, and J. Westerlund (2017): "On The Role of The Rank Condition in CCE Estimation of Factor-augmented Panel Regressions," *Journal of Econometrics*, 197, 60–64.

Karabiyik, H., J.-P. Urbain, and J. Westerlund (2019): "CCE Estimation of Factor-Augmented Regression Models with more Factors than Observables," *Journal of Applied Econometrics*, 34, 268–284.

Norkutė, M., V. Sarafidis, T. Yamagata, and G. Cui (2021): "Instrumental Variable Estimation of Dynamic Linear Panel Data Models with Defactored Regressors and a Multifactor Error Structure," *Journal of Econometrics*, 220, 416–446.

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

Robin, J. M. and R. J. Smith (2000): "Tests of Rank," *Econometric Theory*, 16, 151–175.

Song, M. (2013): "Asymptotic Theory for Dynamic Heterogeneous Panels with Cross-sectional Dependence and its Applications," Unpublished Manuscript, Department of Economics, Columbia University.

Su, L. and S. Jin (2012): "Sieve Estimation of Panel Data Models with Cross-section Dependence," *Journal of Econometrics*, 169, 34 – 47.

Westerlund, J. (2018): "CCE in Panels with General Unknown factors," *The Econometrics Journal*, 21, 264–276.

Westerlund, J. and J.-P. Urbain (2015): "Cross-sectional Averages Versus Principal Components," *Journal of Econometrics*, 185, 372 – 377.