Online Appendix to "Testing Factors in CCE"

Nicholas Brown
Queen's University

Joakim Westerlund*
Lund University
and
Deakin University

October 17, 2022

Abstract

This online appendix provides (i) the formal assumptions that underlie all of our theoretical results, (ii) some theoretical results that are not in the main paper, (iii) the proofs of the asymptotic results provided in Section 3 of the same paper, and (iv) a Monte Carlo study.

A Assumptions

Here and throughout this appendix, $\operatorname{tr}\mathbf{A}$, $\operatorname{rank}\mathbf{A}$ and $\|\mathbf{A}\| = \sqrt{\operatorname{tr}(\mathbf{A}'\mathbf{A})}$ denote the trace, the rank, and the Frobenius (Euclidean) norm of the matrix $\mathbf{A}$, respectively. The symbols $\to_d$ and $\to_p$ signify convergence in distribution and convergence in probability, respectively.

Assumption A.1. $\boldsymbol{\varepsilon}_i$ and $\mathbf{V}_i$ are independently distributed across i and with zero mean, and finite fourth-order cumulants.

Assumption A.2. $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\nu}_i$, where $\boldsymbol{\nu}_i$ is independently distributed across i and with zero mean, and finite fourth-order cumulants.

Assumption A.3. $\mathbf{F}$, $\boldsymbol{\varepsilon}_i$, $\mathbf{V}_j$ and $\boldsymbol{\nu}_n$ are mutually independent for all i, j and n. Also, $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Gamma}_i$ are non-random.

*Corresponding author: Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. Telephone: +46 46 222 8997. Fax: +46 46 222 4613. E-mail address: joakim.westerlund@nek.lu.se.

Assumption A.4. $m_0 \leq k+1 < T$.

Before we take the next assumption, Assumption A.5, it is useful first combine equations (1) and (2) of the main paper. This leads to the following static factor model for $\mathbf{Z}_i = [\mathbf{y}_i, \mathbf{X}_i]$;

$$\mathbf{Z}_i = \mathbf{F}\mathbf{C}_i + \mathbf{U}_i, \tag{A.1}$$

where $\mathbf{C}_i = [\gamma_i + \mathbf{\Gamma}_i\boldsymbol{\beta}_i, \mathbf{\Gamma}_i]$ is $m_0 \times (k+1)$, and $\mathbf{U}_i = [\boldsymbol{\varepsilon}_i + \mathbf{V}_i\boldsymbol{\beta}_i, \mathbf{V}_i]$ is $T \times (k+1)$. Assumption A.5 places restrictions on $\overline{\mathbf{C}}$, the average $\mathbf{C}_i$.

Assumption A.5. $\operatorname{rank} \overline{\mathbf{C}} = m_0$ for all N. Also, there is an unique index set M_0 with $|M_0| = m_0$ such that $\operatorname{rank}(\overline{\mathbf{C}}\mathbf{S}_{M_0}) = m_0$.

Assumption A.6. $\operatorname{rank} \mathbf{F} = m_0$.

Assumption A.7. $N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{F}_M}\mathbf{X}_i \to_p \mathbf{\Phi}$ as $N \to \infty$ for all M, where $\mathbf{\Phi}$ is positive definite.

The required independence of $\varepsilon_{i,t}$ and $\mathbf{v}_{i,t}$ over i in Assumption A.1 is not necessary, and can be relaxed at the expense of additional "high-level" moment conditions.

Assumption A.2 is largely the same as Assumption 4 in Pesaran (2006). It relaxes the otherwise so common equal slope condition (see, for example, Bai, 2009). The slopes are not required to be different, though, as the covariance matrix of $\boldsymbol{\nu}_i$ need not be positive definite. This means that researchers are spared the problem of having to test the homogeneity restriction.

We only consider non-random loadings (Assumption A.3), which represent a more general consideration than random loadings.

The condition that $\overline{\mathbf{C}}$ has full row rank m_0 (Assumption A.5) is the same as condition (21) in Pesaran (2006) and is standard in the CCE literature. Together with $m_0 \leq k+1$ (Assumption A.4), it ensures that the space spanned by $\mathbf{F}$ can be consistently estimated using (a subset of) $\overline{\mathbf{Z}}$. This condition can be relaxed by assuming that some of the factors are observed and appending those factors to $\widehat{\mathbf{F}}_M$, as explained in the empirical illustration.

Assumption A.5 defines the correct index set M_0. This set has the property that $\mathbf{S}_{M_0}$ uniquely selects the averages in $\overline{\mathbf{Z}}$ that are rotationally consistent for $\mathbf{F}$. The uniqueness condition here is analogous to the literature on moment selection (see, for example, Andrews and Lu, 2001). In general, however, depending on $\overline{\mathbf{C}}$, uniqueness may or may not hold. Whether M_0 is unique or not does not affect the properties of the proposed test; however, it may affect the interpretation of the test outcome. Note in particular that if M_0 is not unique then we are no longer looking for a unique set of averages that span the space of $\mathbf{F}$ but rather we may end up with multiple such sets.

B Additional results

The T_M test is useful when the researcher has a preferred set of averages M in mind that he or she wants to test. However, sometimes there is no natural choice of M and there might be several candidates that seem equally reasonable. Fortunately, there is an easy way out of this dilemma. If one of the sets of averages under consideration includes M_0, then its T_M statistic has an asymptotic distribution and the statistics based on the other sets diverge. If, instead, M_0 is not included in any of the sets considered, then all statistics will diverge. Thus, only the set with the smallest $|T_M|$ statistic can possibly include M_0 and we can reject H_0 when the smallest $|T_M|$ is large. This discussion motivates the following test statistic:

$$MT = \min_{j=1,\dots,n} |T_{M_j}|, \tag{B.2}$$

where $T_{M_j} = T(\widehat{\mathbf{F}}_{M_j})$ with $M_1, \dots, M_n$ being the sets of averages considered. The MT test is an instance of an intersection-union test (see Berger, 1982). An important feature of this type of test is that there is no need to size correct for the multiplicity of tests used, as is made clear in the following proposition.

Proposition B.1. *Suppose that Assumptions A.1–A.7 are met. Then,*

$$\lim_{N\to\infty} \mathbb{P}(MT > z_{\alpha/2}) = \alpha. \tag{B.3}$$

Proof: The proof of this proposition is a direct consequence of Theorem 1 of Berger (1982). It is therefore omitted. ■

According to Proposition B.1, the appropriate critical value to use with MT is the same as for each of the individual $|T_M|$ tests making up MT, and yet the overall significance level is still α.

C Proofs

Proof of Theorem 1.

We start with (a). Since H_0 holds in this part of the theorem, we have $M \supseteq M_0$. This means that rank $(\overline{\mathbf{C}}\mathbf{S}_M) = m_0$. We may without loss of generality partition $\mathbf{S}_M$ as $\mathbf{S}_M = [\mathbf{S}_1, \mathbf{S}_{2,M}]$, where $\mathbf{S}_1$ and $\mathbf{S}_{2,M}$ are $(k+1) \times m_0$ and $(k+1) \times (m - m_0)$ matrices, respectively, that selects the full and reduced rank submatrices of $\overline{\mathbf{C}}$. Here $M_0^c = M \setminus M_0$ is the complement of M_0, $|M_0| = m_0$ and $|M_0^c| = m - m_0$. Note also that because M_0 is unique, unlike $\mathbf{S}_{2,M}$, $\mathbf{S}_1$ is independent of M. Thus, $\overline{\mathbf{C}}\mathbf{S}_M = [\overline{\mathbf{C}}\mathbf{S}_1, \overline{\mathbf{C}}\mathbf{S}_{2,M}] = [\overline{\mathbf{C}}_1, \overline{\mathbf{C}}_{2,M}] = \overline{\mathbf{C}}_M$, where $\overline{\mathbf{C}}_1$ is an $m_0 \times m_0$ full rank matrix and $\overline{\mathbf{C}}_{2,M}$ is $m_0 \times (m - m_0)$. The $T \times m$ matrix $\overline{\mathbf{U}}\mathbf{S}_M$ is partitioned similarly as $\overline{\mathbf{U}}\mathbf{S}_M = [\overline{\mathbf{U}}_1, \overline{\mathbf{U}}_{2,M}] = \overline{\mathbf{U}}_M$, where $\overline{\mathbf{U}}_1$ is $T \times m_0$ and $\overline{\mathbf{U}}_{2,M}$ is $T \times (m - m_0)$. In this notation,

$$\widehat{\mathbf{F}}_M = \mathbf{F}\overline{\mathbf{C}}_M + \overline{\mathbf{U}}_M = [\mathbf{F}\overline{\mathbf{C}}_1, \mathbf{F}\overline{\mathbf{C}}_{2,M}] + [\overline{\mathbf{U}}_1, \overline{\mathbf{U}}_{2,M}]. \tag{C.4}$$

Define

$$\overline{\mathbf{B}}_M = \begin{bmatrix} \overline{\mathbf{C}}_1^{-1} & -\overline{\mathbf{C}}_1^{-1}\overline{\mathbf{C}}_{2,M} \\ \mathbf{0}_{(m-m_0)\times m_0} & \mathbf{I}_{m-m_0} \end{bmatrix} = [\overline{\mathbf{B}}_{1,M}, \overline{\mathbf{B}}_{2,M}], \tag{C.5}$$

with obvious definitions of $\overline{\mathbf{B}}_{1,M}$ and $\overline{\mathbf{B}}_{2,M}$. Note that while $\overline{\mathbf{B}}_{1,M}$ is $m \times m_0$, $\overline{\mathbf{B}}_{2,M}$ is $m \times (m - m_0)$, which means that $\overline{\mathbf{B}}_M$ is $m \times m$. The matrix $\overline{\mathbf{B}}_M$ is also full rank, because rank $\overline{\mathbf{B}}_M =$ rank $\overline{\mathbf{C}}_1^{-1} +$ rank $\mathbf{I}_{m-m_0} = m$ (see Abadir and Magnus, 2005, Exercise 5.43), and can therefore be inverted. This is very important, as will soon become clear. Another very useful property of $\overline{\mathbf{B}}_M$ is that $\overline{\mathbf{C}}_M\overline{\mathbf{B}}_M = [\mathbf{I}_{m_0}, \mathbf{0}_{m_0\times(m-m_0)}]$, which we can use to establish the following:

$$\widehat{\mathbf{F}}_M\overline{\mathbf{B}}_M = \mathbf{F}\overline{\mathbf{C}}_M\overline{\mathbf{B}}_M + \overline{\mathbf{U}}_M\overline{\mathbf{B}}_M = [\mathbf{F}, \mathbf{0}_{T\times(m-m_0)}] + [\overline{\mathbf{U}}_M\overline{\mathbf{B}}_{1,M}, \overline{\mathbf{U}}_M\overline{\mathbf{B}}_{2,M}]. \tag{C.6}$$

Because $\|\overline{\mathbf{U}}\| = O_p(N^{-1/2})$ for a fixed T under general conditions, the last $m - m_0$ columns of $\widehat{\mathbf{F}}_M\overline{\mathbf{B}}_M$ converge to zero and are in this sense degenerate. Also, since $\overline{\mathbf{B}}_M$ is invertible, we have $\mathbf{P}_{\widehat{F}_M} = \mathbf{P}_{\widehat{F}_M\overline{B}_M}$. The degeneracy in $\widehat{\mathbf{F}}_M\overline{\mathbf{B}}_M$ therefore causes an asymptotic singularity in $\mathbf{P}_{\widehat{F}_M\overline{B}_M}$. In order to address this issue, we introduce $m \times m$ normalization matrix $\mathbf{D}_M = \text{diag}(\mathbf{I}_{m_0}, \sqrt{N}\mathbf{I}_{m-m_0})$, which is such that if we let $\widehat{\mathbf{F}}_M^0 = \widehat{\mathbf{F}}_M\overline{\mathbf{B}}_M\mathbf{D}_M$, $\mathbf{F}_M^0 = \mathbf{F}\overline{\mathbf{C}}_M\overline{\mathbf{B}}_M\mathbf{D}_M = [\mathbf{F}, \mathbf{0}_{T\times(m-m_0)}]$ and $\overline{\mathbf{U}}_M^0 = \overline{\mathbf{U}}_M\overline{\mathbf{B}}_M\mathbf{D}_M = [\overline{\mathbf{U}}\mathbf{S}_M\overline{\mathbf{B}}_{1,M}, \sqrt{N}\overline{\mathbf{U}}\mathbf{S}_M\overline{\mathbf{B}}_{2,M}] = [\overline{\mathbf{U}}_{1,M}^0, \overline{\mathbf{U}}_{2,M}^0]$, then

$$\widehat{\mathbf{F}}_M^0 = \mathbf{F}_M^0 + \overline{\mathbf{U}}_M^0 = [\mathbf{F}, \mathbf{0}_{T\times(m-m_0)}] + [\overline{\mathbf{U}}_{1,M}^0, \overline{\mathbf{U}}_{2,M}^0]. \tag{C.7}$$

It is important to realize that since now $\|\overline{\mathbf{U}}^0_{1,M}\| = O_p(N^{-1/2})$ and $\|\overline{\mathbf{U}}^0_{2,M}\| = O_p(1)$, letting $\mathbf{F}^+_M = [\mathbf{F}, \overline{\mathbf{U}}^0_{2,M}]$, we have

$$\widehat{\mathbf{F}}^0_M = [\mathbf{F}, \overline{\mathbf{U}}^0_{2,M}] + O_p(N^{-1/2}) = \mathbf{F}^+_M + O_p(N^{-1/2}), \tag{C.8}$$

which means that in contrast to $\widehat{\mathbf{F}}_M \overline{\mathbf{B}}_M$, all the columns of $\widehat{\mathbf{F}}^0_M$ are non-degenerate. This is therefore the appropriate estimator to consider in the asymptotic analysis. Moreover, since $\mathbf{M}_{\widehat{F}_M} = \mathbf{M}_{\widehat{F}^0_M}$, having $\widehat{\mathbf{F}}_M$ is just as good as having $\widehat{\mathbf{F}}^0_M$, although in practice the latter estimator is of course unobservable.

The above notation can be extended to cover also the case when $M = M_0$. Note in particular that if we define $\overline{\mathbf{C}}_M = \overline{\mathbf{C}}\mathbf{S}_M = \overline{\mathbf{C}}\mathbf{S}_1 = \overline{\mathbf{C}}_1$, $\overline{\mathbf{B}}_M = \overline{\mathbf{B}}_{1,M} = \overline{\mathbf{C}}_1^{-1}$ and $\mathbf{D}_M = \mathbf{D}_{1,M} = \mathbf{I}_{m_0}$, we have $\widehat{\mathbf{F}}^0_M = \widehat{\mathbf{F}}^0_{1,M} = \widehat{\mathbf{F}}_{1,M}\overline{\mathbf{B}}_{1,M}\mathbf{D}_{1,M} = \widehat{\mathbf{F}}_{1,M}\overline{\mathbf{C}}_1^{-1}$, $\mathbf{F}^0_M = \mathbf{F}^0_{1,M} = \mathbf{F}\overline{\mathbf{C}}\mathbf{S}_1\overline{\mathbf{B}}_{1,M}\mathbf{D}_{1,M} = \mathbf{F}\overline{\mathbf{C}}_1\overline{\mathbf{C}}_1^{-1} = \mathbf{F}$ and $\overline{\mathbf{U}}^0_M = \overline{\mathbf{U}}^0_{1,M} = \overline{\mathbf{U}}\mathbf{S}_1\overline{\mathbf{B}}_{1,M}\mathbf{D}_{1,M} = \overline{\mathbf{U}}\mathbf{S}_1\overline{\mathbf{C}}_1^{-1} = \overline{\mathbf{U}}^0_{M_0,1}$, and hence

$$\widehat{\mathbf{F}}^0_M = \widehat{\mathbf{F}}^0_{1,M} = \widehat{\mathbf{F}}_{1,M}\overline{\mathbf{C}}_1^{-1} = \mathbf{F}^0_{1,M} + \overline{\mathbf{U}}^0_{1,M} = \mathbf{F} + \overline{\mathbf{U}}^0_{M_0,1} = \mathbf{F} + O_p(N^{-1/2}). \tag{C.9}$$

We now make use of the above consistency results to evaluate $\widehat{\mathbf{F}}_M\widehat{\mathbf{g}}_{i,M}$. Letting $\widehat{\mathbf{g}}^0_{i,M} = (\mathbf{D}_N\overline{\mathbf{B}}'_M)^{-1\prime}\widehat{\mathbf{g}}_{i,M} = (\overline{\mathbf{B}}_M\mathbf{D}_N)^{-1}\widehat{\mathbf{g}}_{i,M}$ and $\gamma^0_i = (\overline{\mathbf{C}}\mathbf{S}_M\overline{\mathbf{B}}_M\mathbf{D}_N)^+\gamma_i = \mathbf{D}_N\overline{\mathbf{B}}'_M\overline{\mathbf{C}}'\gamma_i = [\gamma_i', \mathbf{0}_{1\times(m-m_0)}]'$,

$$\begin{aligned}
\widehat{\mathbf{F}}_M\widehat{\mathbf{g}}_{i,M} - \mathbf{F}\gamma_i &= \widehat{\mathbf{F}}_M\overline{\mathbf{B}}_M\mathbf{D}_N(\overline{\mathbf{B}}_M\mathbf{D}_N)^{-1}\widehat{\mathbf{g}}_{i,M} - \mathbf{F}\overline{\mathbf{C}}_M\overline{\mathbf{B}}_M\mathbf{D}_N(\overline{\mathbf{C}}_M\overline{\mathbf{B}}_M\mathbf{D}_N)^+\gamma_i \\
&= \widehat{\mathbf{F}}^0_M\widehat{\mathbf{g}}^0_{i,M} - \mathbf{F}^0_M\gamma^0_i \\
&= (\widehat{\mathbf{F}}^0_M - \mathbf{F}^0_M)\gamma^0_i + \widehat{\mathbf{F}}^0_M(\widehat{\mathbf{g}}^0_{i,M} - \gamma^0_i).
\end{aligned} \tag{C.10}$$

Clearly, $(\widehat{\mathbf{F}}^0_M - \mathbf{F}^0_M)\gamma^0_i = \overline{\mathbf{U}}^0_{1,M}\gamma_i$. Consider the second term on the right-hand side. By using $\overline{\mathbf{C}}_M\overline{\mathbf{B}}_{1,M} = \mathbf{I}_{m_0}$ and defining $\mathbf{g}_{i,M} = \overline{\mathbf{B}}_{1,M}\gamma_i$, the model for $\mathbf{y}_i$ can be written as

$$\begin{aligned}
\mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta}_i + \widehat{\mathbf{F}}_M\overline{\mathbf{B}}_{1,M}\gamma_i - (\widehat{\mathbf{F}}_M - \mathbf{F}\overline{\mathbf{C}}_M)\overline{\mathbf{B}}_{1,M}\gamma_i + \varepsilon_i \\
&= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\boldsymbol{\nu}_i + \widehat{\mathbf{F}}_M\mathbf{g}_{i,M} - \overline{\mathbf{U}}^0_{1,M}\gamma_i + \varepsilon_i.
\end{aligned} \tag{C.11}$$

By inserting this into the above expression for $\widehat{\mathbf{g}}_{i,M}$,

$$\begin{aligned}
\widehat{\mathbf{g}}_{i,M} &= (\widehat{\mathbf{F}}'_M\widehat{\mathbf{F}}_M)^+\widehat{\mathbf{F}}'_M(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_M) \\
&= (\widehat{\mathbf{F}}'_M\widehat{\mathbf{F}}_M)^+\widehat{\mathbf{F}}'_M(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\boldsymbol{\nu}_i + \widehat{\mathbf{F}}_M\mathbf{g}_i - \overline{\mathbf{U}}^0_{1,M}\gamma_i + \varepsilon_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_M) \\
&= \mathbf{g}_{i,M} + (\widehat{\mathbf{F}}'_M\widehat{\mathbf{F}}_M)^+\widehat{\mathbf{F}}'_M[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,M}\gamma_i + \varepsilon_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})],
\end{aligned} \tag{C.12}$$

implying

$$\begin{aligned}
\widehat{\mathbf{g}}^0_{i,M} &= (\overline{\mathbf{B}}_M\mathbf{D}_N)^{-1}\widehat{\mathbf{g}}_{i,M} \\
&= (\overline{\mathbf{B}}_M\mathbf{D}_N)^{-1}\mathbf{g}_{i,M} + (\overline{\mathbf{B}}_M\mathbf{D}_N)^{-1}(\widehat{\mathbf{F}}'_M\widehat{\mathbf{F}}_M)^{+}\widehat{\mathbf{F}}'_M[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,M}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] \\
&= (\overline{\mathbf{B}}_M\mathbf{D}_N)^{-1}\mathbf{g}_{i,M} + (\mathbf{D}_N\overline{\mathbf{B}}'_M\widehat{\mathbf{F}}'_M\widehat{\mathbf{F}}_M\overline{\mathbf{B}}_M\mathbf{D}_N)^{-1}\mathbf{D}_N\overline{\mathbf{B}}'_M\widehat{\mathbf{F}}'[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,M}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] \\
&= (\overline{\mathbf{B}}_M\mathbf{D}_N)^{-1}\mathbf{g}_{i,M} + (\widehat{\mathbf{F}}^{0\prime}_M\widehat{\mathbf{F}}^0_M)^{+}\widehat{\mathbf{F}}^{0\prime}_M[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,M}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})]. \qquad (C.13)
\end{aligned}$$

Consider the first term on the right-hand side. A direct calculation using the rules for the inverse of a partitioned matrix (see, for example, Abadir and Magnus, 2005, Exercise 5.16) reveals that

$$(\mathbf{D}_N\overline{\mathbf{B}}_M)^{-1} = \begin{bmatrix} \overline{\mathbf{C}}_1 & \overline{\mathbf{C}}_{2,M} \\ \mathbf{0}_{(m-m_0)\times m_0} & N^{-1/2}\mathbf{I}_{m-m_0} \end{bmatrix}, \qquad (C.14)$$

so that

$$\begin{aligned}
(\overline{\mathbf{B}}_M\mathbf{D}_N)^{-1}\overline{\mathbf{B}}_{1,M} &= \begin{bmatrix} \overline{\mathbf{C}}_1 & \overline{\mathbf{C}}_{2,M} \\ \mathbf{0}_{(m-m_0)\times m_0} & N^{-1/2}\mathbf{I}_{m-m_0} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{C}}_1^{-1} \\ \mathbf{0}_{(m-m_0)\times m_0} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I}_{1,M} \\ \mathbf{0}_{(m-m_0)\times m_0} \end{bmatrix}. \qquad (C.15)
\end{aligned}$$

This implies

$$(\overline{\mathbf{B}}_M\mathbf{D}_N)^{-1}\mathbf{g}_{i,M} = \begin{bmatrix} \gamma_i \\ \mathbf{0}_{(m-m_0)\times 1} \end{bmatrix} = \gamma_i^0, \qquad (C.16)$$

leading to the following expression for $\widehat{\mathbf{g}}^0_{i,M} - \gamma_i^0$:

$$\widehat{\mathbf{g}}^0_{i,M} - \gamma_i^0 = (\widehat{\mathbf{F}}^{0\prime}_M\widehat{\mathbf{F}}^0_M)^{+}\widehat{\mathbf{F}}^{0\prime}_M[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,M}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})]. \qquad (C.17)$$

It follows that

$$\widehat{\mathbf{F}}^0_M(\widehat{\mathbf{g}}^0_{i,M} - \gamma_i^0) = \widehat{\mathbf{F}}^0_M(\widehat{\mathbf{F}}^{0\prime}_M\widehat{\mathbf{F}}^0_M)^{+}\widehat{\mathbf{F}}^{0\prime}_M[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,M}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})]. \qquad (C.18)$$

We have already shown that $\widehat{\mathbf{F}}^0_M = \mathbf{F}^+_M + O_p(N^{-1/2})$. By using this and the results provided in the proof of Lemma A.1 in Westerlund et al. (2019), we have that $\|\widehat{\mathbf{F}}^{0\prime}_M\widehat{\mathbf{F}}^0_M - \mathbf{F}^{+\prime}_M\mathbf{F}^+_M\| = O_p(N^{-1/2})$ and, more importantly,

$$\|(\widehat{\mathbf{F}}^{0\prime}_M\widehat{\mathbf{F}}^0_M)^{+} - (\mathbf{F}^{+\prime}_M\mathbf{F}^+_M)^{+}\| = O_p(N^{-1/2}), \qquad (C.19)$$

By using this last result and $\widehat{\mathbf{F}}_M^0 = \mathbf{F}_M^+ + O_p(N^{-1/2})$, we can show that

$$\begin{aligned}
\widehat{\mathbf{F}}_M^0(\widehat{\mathbf{F}}_M^{0\prime}\widehat{\mathbf{F}}_M^0)^+\widehat{\mathbf{F}}_M^{0\prime} &= \widehat{\mathbf{F}}_M^0[(\widehat{\mathbf{F}}_M^{0\prime}\widehat{\mathbf{F}}_M^0)^+ - (\mathbf{F}_M^{+\prime}\mathbf{F}_M^+)^+]\widehat{\mathbf{M}}_M^{0\prime} + \widehat{\mathbf{F}}_M^0(\mathbf{F}_M^{+\prime}\mathbf{F}_M^+)^+\widehat{\mathbf{F}}_M^{0\prime} \\
&= \widehat{\mathbf{F}}_M^0(\mathbf{F}_M^{+\prime}\mathbf{F}_M^+)^+\widehat{\mathbf{F}}_M^{0\prime} + O_p(N^{-1/2}) \\
&= \mathbf{F}_M^+(\mathbf{F}_M^{+\prime}\mathbf{F}_M^+)^+\mathbf{F}^{+\prime} + O_p(N^{-1/2}), \qquad \text{(C.20)}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{F}_M^+(\mathbf{F}_M^{+\prime}\mathbf{F}_M^+)^+\mathbf{F}_M^{+\prime} &= [\mathbf{F}, \overline{\mathbf{U}}_{2,M}^0] \begin{bmatrix} (\mathbf{F}'\mathbf{F})^{-1} + (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\overline{\mathbf{U}}_{2,M}^0(\overline{\mathbf{U}}_{2,M}^{0\prime}\mathbf{M}_F\overline{\mathbf{U}}_{2,M}^0)^{-1}\overline{\mathbf{U}}_{2,M}^{0\prime}\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1} & -(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\overline{\mathbf{U}}_{2,M}^0(\overline{\mathbf{U}}_{2,M}^{0\prime}\mathbf{M}_F\overline{\mathbf{U}}_{2,M}^0)^{-1} \\ -(\overline{\mathbf{U}}_{2,M}^{0\prime}\mathbf{M}_F\overline{\mathbf{U}}_{2,M}^0)^{-1}\overline{\mathbf{U}}_{2,M}^{0\prime}\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1} & (\overline{\mathbf{U}}_{2,M}^{0\prime}\mathbf{M}_F\overline{\mathbf{U}}_{2,M}^0)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{F}' \\ \overline{\mathbf{U}}_{2,M}^{0\prime} \end{bmatrix} \\
&= \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}' - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\overline{\mathbf{U}}_{2,M}^0(\overline{\mathbf{U}}_{2,M}^{0\prime}\mathbf{M}_F\overline{\mathbf{U}}_{2,M}^0)^{-1}\overline{\mathbf{U}}_{2,M}^{0\prime}\mathbf{M}_F \\
&\quad + \overline{\mathbf{U}}_{2,M}^0(\overline{\mathbf{U}}_{2,M}^{0\prime}\mathbf{M}_F\overline{\mathbf{U}}_{2,M}^0)^{-1}\overline{\mathbf{U}}_{2,M}^{0\prime}\mathbf{M}_F \\
&= \mathbf{P}_F + \mathbf{M}_F\overline{\mathbf{U}}_{2,M}^0(\overline{\mathbf{U}}_{2,M}^{0\prime}\mathbf{M}_F\overline{\mathbf{U}}_{2,M}^0)^{-1}\overline{\mathbf{U}}_{2,M}^{0\prime}\mathbf{M}_F \\
&= \mathbf{P}_F + \mathbf{P}_{M_F\overline{U}_{2,M}^0}. \qquad \text{(C.21)}
\end{aligned}$$

Insertion into $\widehat{\mathbf{F}}^0(\widehat{\mathbf{g}}_{i,M}^0 - \gamma_i^0)$ gives

$$\begin{aligned}
\widehat{\mathbf{F}}_M^0(\widehat{\mathbf{g}}_{i,M}^0 - \gamma_i^0) &= \widehat{\mathbf{F}}_M^0(\widehat{\mathbf{F}}_M^{0\prime}\widehat{\mathbf{F}}_M^0)^+\widehat{\mathbf{F}}_M^{0\prime}[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}_{1,M}^0\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] \\
&= \mathbf{F}_M^+(\mathbf{F}_M^{+\prime}\mathbf{F}_M^+)^+\mathbf{F}_M^{+\prime}[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}_{1,M}^0\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] + O_p(N^{-1/2}) \\
&= (\mathbf{P}_F + \mathbf{P}_{M_F\overline{U}_{2,M}^0})[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}_{1,M}^0\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] + O_p(N^{-1/2}), \quad \text{(C.22)}
\end{aligned}$$

which in turn implies

$$\begin{aligned}
&\widehat{\mathbf{F}}_M\widehat{\mathbf{g}}_{i,M} - \mathbf{F}\gamma_i \\
&= (\widehat{\mathbf{F}}_M^0 - \mathbf{F}_M^0)\gamma_i^0 + \widehat{\mathbf{F}}_M^0(\widehat{\mathbf{g}}_{i,M}^0 - \gamma_i^0) \\
&= \overline{\mathbf{U}}_{1,M}^0\gamma_i + (\mathbf{P}_F + \mathbf{P}_{M_F\overline{U}_{2,M}^0})[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}_{1,M}^0\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] + O_p(N^{-1/2}). \qquad \text{(C.23)}
\end{aligned}$$

The above result holds for any $M \supseteq M_0$, including $M = \overline{M}$. Hence,

$$\begin{aligned}
\widehat{\mathbf{F}}_{\overline{M}}\widehat{\mathbf{g}}_{i,\overline{M}} - \mathbf{F}\gamma_i &= \overline{\mathbf{U}}_{1,\overline{M}}^0\gamma_i + (\mathbf{P}_F + \mathbf{P}_{M_F\overline{U}_{2,\overline{M}}^0})[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}_{1,\overline{M}}^0\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \boldsymbol{\beta})] \\
&\quad + O_p(N^{-1/2}), \qquad \text{(C.24)}
\end{aligned}$$

leading to the following expression for $\mathbf{\Delta}_M$:

$$\begin{aligned}
\mathbf{\Delta}_{i,M} &= \widehat{\mathbf{F}}_{\overline{M}}\widehat{\mathbf{g}}_{i,\overline{M}} - \widehat{\mathbf{F}}_M\widehat{\mathbf{g}}_{i,M} \\
&= \overline{\mathbf{U}}^0_{1,\overline{M}}\gamma_i + (\mathbf{P}_F + \mathbf{P}_{M_F\overline{U}^0_{2,\overline{M}}})[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,\overline{M}}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \boldsymbol{\beta})] \\
&- \overline{\mathbf{U}}^0_{1,M}\gamma_i - (\mathbf{P}_F + \mathbf{P}_{M_F\overline{U}^0_{2,M}})[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,M}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] + O_p(N^{-1/2}) \\
&= \mathbf{M}_F(\overline{\mathbf{U}}^0_{1,\overline{M}} - \overline{\mathbf{U}}^0_{1,M})\gamma_i - \mathbf{P}_F\mathbf{X}_i(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \widehat{\boldsymbol{\beta}}_M) \\
&+ \mathbf{P}_{M_F\overline{U}^0_{2,\overline{M}}}[\mathbf{V}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,\overline{M}}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{V}_i(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \boldsymbol{\beta})] \\
&- \mathbf{P}_{M_F\overline{U}^0_{2,M}}[\mathbf{V}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,M}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{V}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] + O_p(N^{-1/2}) \\
&= \mathbf{E}_{i,M} + O_p(N^{-1/2}), \qquad \text{(C.25)}
\end{aligned}$$

where $\mathbf{E}_{i,M}$ is implicitly defined and the second equality holds because $\mathbf{P}_{M_F\overline{U}^0_{2,\overline{M}}}\mathbf{F} = \mathbf{0}_{T\times m_0}$.

It is important to note that the order of the reminder in the above expression for $\widehat{\mathbf{F}}_{\overline{M}}\widehat{\mathbf{g}}_{i,\overline{M}} - \widehat{\mathbf{F}}_M\widehat{\mathbf{g}}_{i,M}$, which incurred when replacing $\widehat{\mathbf{F}}^0_M(\widehat{\mathbf{F}}^{0\prime}_M\widehat{\mathbf{F}}^0_M)^+\widehat{\mathbf{F}}^{0\prime}_M$ with $\mathbf{F}^+_M(\mathbf{F}^{+\prime}_M\mathbf{F}^+_M)^+\mathbf{F}^{+\prime}_M$, is the same even after averaging over i and multiplying by $\sqrt{N}$. In order to appreciate this, we make use of the fact that $\sqrt{N}(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})$ is asymptotically mixed normal by Theorem 1 of Westerlund and Kaddoura (2022), and hence $\|\sqrt{N}(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})\| = O_p(1)$. Moreover, $\|\sqrt{N}\overline{\mathbf{U}}^0_{1,M}\| = O_p(1)$, and since $\mathbf{X}_i$ and $\boldsymbol{\nu}_i$ are independent with $\boldsymbol{\nu}_i$ mean zero and independent also across i, we also have $\|N_g^{-1/2}\sum_{i=1}^N \mathbf{X}_i\boldsymbol{\nu}_i\| = O_p(1)$. It follows that

$$\begin{aligned}
&\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,M}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})]\right\| \\
&\leq \left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{X}_i\boldsymbol{\nu}_i\right\| + \|\sqrt{N}\overline{\mathbf{U}}^0_{1,M}\|\left\|\frac{1}{N}\sum_{i=1}^N\gamma_i\right\| + \left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N\boldsymbol{\varepsilon}_i\right\| \\
&+ \left\|\frac{1}{N}\sum_{i=1}\mathbf{X}_i\right\|\|\sqrt{N}(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})\| = O_p(1). \qquad \text{(C.26)}
\end{aligned}$$

We can therefore show that

$$\begin{aligned}
&\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N[\mathbf{F}^+_M(\mathbf{F}^{+\prime}_M\mathbf{F}^+_M)^+\mathbf{F}^{+\prime}_M - \widehat{\mathbf{F}}^0_M(\widehat{\mathbf{F}}^{0\prime}_M\widehat{\mathbf{F}}^0_M)^+\widehat{\mathbf{F}}^{0\prime}_M][\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,M}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})]\right\| \\
&\leq \|\mathbf{F}^+_M(\mathbf{F}^{+\prime}_M\mathbf{F}^+_M)^+\mathbf{F}^{+\prime}_M - \widehat{\mathbf{F}}^0_M(\widehat{\mathbf{F}}^{0\prime}_M\widehat{\mathbf{F}}^0_M)^+\widehat{\mathbf{F}}^{0\prime}_M\|\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,M}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})]\right\| \\
&= O_p(N^{-1/2}). \qquad \text{(C.27)}
\end{aligned}$$

Note in addition how

$$\begin{aligned}
\sqrt{N}\overline{\mathbf{E}}_M &= \mathbf{M}_F\sqrt{N}(\overline{\mathbf{U}}^0_{1,\overline{M}} - \overline{\mathbf{U}}^0_{1,M})\overline{\gamma} - \mathbf{P}_F\overline{\mathbf{X}}\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \widehat{\boldsymbol{\beta}}_M) \\
&+ \mathbf{P}_{M_F\overline{U}^0_{2,\overline{M}}}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{V}_i\boldsymbol{\nu}_i - \sqrt{N}\overline{\mathbf{U}}^0_{1,\overline{M}}\overline{\boldsymbol{\gamma}} + \sqrt{N}\overline{\boldsymbol{\varepsilon}} - \sqrt{N}\overline{\mathbf{V}}(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \boldsymbol{\beta})\right) \\
&- \mathbf{P}_{M_F\overline{U}^0_{2,M}}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{V}_i\boldsymbol{\nu}_i - \sqrt{N}\overline{\mathbf{U}}^0_{1,M}\overline{\boldsymbol{\gamma}} + \sqrt{N}\overline{\boldsymbol{\varepsilon}} - \sqrt{N}\overline{\mathbf{V}}(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})\right) \\
&= \mathbf{M}_F\sqrt{N}(\overline{\mathbf{U}}^0_{1,\overline{M}} - \overline{\mathbf{U}}^0_{1,M})\overline{\gamma} - \mathbf{P}_F\mathbf{F}\overline{\boldsymbol{\Gamma}}\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \widehat{\boldsymbol{\beta}}_M) \\
&+ \mathbf{P}_{M_F\overline{U}^0_{2,\overline{M}}}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{V}_i\boldsymbol{\nu}_i - \sqrt{N}\overline{\mathbf{U}}^0_{1,\overline{M}}\overline{\boldsymbol{\gamma}} + \sqrt{N}\overline{\boldsymbol{\varepsilon}}\right) \\
&- \mathbf{P}_{M_F\overline{U}^0_{2,M}}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{V}_i\boldsymbol{\nu}_i - \sqrt{N}\overline{\mathbf{U}}^0_{1,M}\overline{\boldsymbol{\gamma}} + \sqrt{N}\overline{\boldsymbol{\varepsilon}}\right) + O_p(N^{-1/2}) \\
&= \mathbf{M}_F\sqrt{N}(\overline{\mathbf{U}}^0_{1,\overline{M}} - \overline{\mathbf{U}}^0_{1,M})\overline{\gamma} - \mathbf{F}\overline{\boldsymbol{\Gamma}}\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \widehat{\boldsymbol{\beta}}_M) \\
&+ (\mathbf{P}_{M_F\overline{U}^0_{2,\overline{M}}} - \mathbf{P}_{M_F\overline{U}^0_{2,M}})\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{V}_i\boldsymbol{\nu}_i + \sqrt{N}\overline{\boldsymbol{\varepsilon}}\right) \\
&- \mathbf{P}_{M_F\overline{U}^0_{2,\overline{M}}}\sqrt{N}\overline{\mathbf{U}}^0_{1,\overline{M}}\overline{\boldsymbol{\gamma}} + \mathbf{P}_{M_F\overline{U}^0_{2,M}}\sqrt{N}\overline{\mathbf{U}}^0_{1,M}\overline{\boldsymbol{\gamma}} + O_p(N^{-1/2}),
\end{aligned} \tag{C.28}$$

where

$$\overline{\mathbf{U}}^0_{1,M} = \overline{\mathbf{U}}\mathbf{S}_M\overline{\mathbf{B}}_{1,M} = \overline{\mathbf{U}}[\mathbf{S}_1, \mathbf{S}_{2,M}]\begin{bmatrix} \overline{\mathbf{C}}_1^{-1} \\ \mathbf{0}_{(m-m_0)\times m_0} \end{bmatrix} = \overline{\mathbf{U}}\mathbf{S}_1\overline{\mathbf{C}}_1^{-1}. \tag{C.29}$$

This last result holds for any M including $\overline{M}$. Hence, letting

$$\mathbf{E}^0_M = -\mathbf{F}\boldsymbol{\Gamma}_i(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \widehat{\boldsymbol{\beta}}_M) + (\mathbf{P}_{M_F\overline{U}^0_{2,\overline{M}}} - \mathbf{P}_{M_F\overline{U}^0_{2,M}})(\mathbf{V}_i\boldsymbol{\nu}_i + \boldsymbol{\varepsilon}_i - \overline{\mathbf{U}}\mathbf{S}_1\overline{\mathbf{C}}_1^{-1}\gamma_i), \tag{C.30}$$

we have

$$\begin{aligned}
\sqrt{N}\overline{\mathbf{E}}_M &= -\mathbf{F}\overline{\boldsymbol{\Gamma}}\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \widehat{\boldsymbol{\beta}}_M) \\
&+ (\mathbf{P}_{M_F\overline{U}^0_{2,\overline{M}}} - \mathbf{P}_{M_F\overline{U}^0_{2,M}})\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{V}_i\boldsymbol{\nu}_i + \sqrt{N}\overline{\boldsymbol{\varepsilon}} - \sqrt{N}\overline{\mathbf{U}}\mathbf{S}_1\overline{\mathbf{C}}_1^{-1}\overline{\boldsymbol{\gamma}}\right) + O_p(N^{-1/2}) \\
&= \sqrt{N}\overline{\mathbf{E}}^0_M + O_p(N^{-1/2}),
\end{aligned} \tag{C.31}$$

which in turn implies

$$\begin{aligned}
\sqrt{N}\overline{\boldsymbol{\Delta}}_M &= \frac{1}{\sqrt{N}}\sum_{i=1}^N(\widehat{\mathbf{F}}_{\overline{M}}\widehat{\mathbf{g}}_{i,\overline{M}} - \widehat{\mathbf{F}}_M\widehat{\mathbf{g}}_{i,M}) = \frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{E}_{i,M} + O_p(N^{-1/2}) \\
&= \sqrt{N}\overline{\mathbf{E}}_M + O_p(N^{-1/2}) = \sqrt{N}\overline{\mathbf{E}}^0_M + O_p(N^{-1/2}).
\end{aligned} \tag{C.32}$$

Both terms that make up $\mathbf{E}^0_{i,M}$ are mean zero and conditionally independent across i. They are therefore asymptotically mixed normal by a central limit law for conditionally independent variables. However, they are not uncorrelated with each other, which complicates the calculation of the asymptotic variance. Let us therefore define $\mathbf{\Sigma}_M = N\mathbb{E}(\overline{\mathbf{E}}^0_M \overline{\mathbf{E}}^{0\prime}_M | \mathcal{C})$, where $\mathcal{C}$ is the sigma-field generated by $\mathbf{F}$ and (the limit of) $\overline{\mathbf{U}}^0_{2,\overline{M}}$. It follows that

$$\sqrt{N}\overline{\mathbf{E}}^0_M \to_d MN(\mathbf{0}_{T\times 1}, \mathbf{\Sigma}_M) \tag{C.33}$$

as $N \to \infty$, where $MN(\cdot, \cdot)$ signifies a mixed normal distribution that is normal conditionally on $\mathcal{C}$. We can therefore show that

$$\sqrt{N}\overline{\mathbf{\Delta}}_M = \sqrt{N}\overline{\mathbf{E}}^0_M + O_p(N^{-1/2}) \to_d MN(\mathbf{0}_{T\times 1}, \mathbf{\Sigma}_M) \tag{C.34}$$

as $N \to \infty$. Hence,

$$\sqrt{N}1'_{T\times 1}\overline{\mathbf{\Delta}}_M = \sqrt{N}1'_{T\times 1}\overline{\mathbf{E}}^0_M + O_p(N^{-1/2}) \to_d MN(0, 1'_{T\times 1}\mathbf{\Sigma}_M 1_{T\times 1}), \tag{C.35}$$

and so

$$\frac{\sqrt{N}1'_{T\times 1}\overline{\mathbf{\Delta}}_M}{\sqrt{1'_{T\times 1}\mathbf{\Sigma}_M 1_{T\times 1}}} \to_d N(0,1). \tag{C.36}$$

Let us now consider $\widehat{\mathbf{\Sigma}}_M$. Since $\mathbf{\Delta}_{i,M}$ is again conditionally independent across i, by a law of large numbers for conditionally independent variables,

$$\widehat{\mathbf{\Sigma}}_M = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{\Delta}_{i,M} - \overline{\mathbf{\Delta}}_M)(\mathbf{\Delta}_{i,M} - \overline{\mathbf{\Delta}}_M)' \to_p \mathbf{\Sigma}_M \tag{C.37}$$

as $N \to \infty$. The required result under $M \supseteq M_0$ is implies by this.

We now move on to part (b) of the theorem. Since H_1 holds here, we have $M \subset M_0$. However, regardless of whether $M \supseteq M_0$ or $M \subset M_0$,

$$\widehat{\mathbf{F}}_M = \mathbf{F}\overline{\mathbf{C}}_M + \overline{\mathbf{U}}_M = \mathbf{F}\overline{\mathbf{C}}_M + O_p(N^{-1/2}). \tag{C.38}$$

Hence, consistency in this sense is not impaired by under-specification of the number of averages. However, because $m < m_0$, we have rank $(\overline{\mathbf{C}}_M) = m$. Hence, in contrast to before, now $\overline{\mathbf{C}}_M$ has full column rank, which means that many of the results that held under H_0 cannot be used anymore. Note in particular how $\overline{\mathbf{C}}_M^+ = (\overline{\mathbf{C}}_M'\overline{\mathbf{C}}_M)^{-1}\overline{\mathbf{C}}_M'$ such that $\overline{\mathbf{C}}_M^+\overline{\mathbf{C}}_M = \mathbf{I}_m$ (see, for example, Abadir and Magnus, 2005, Exercise 10.31), which means that $\mathbf{F}$ and γ_i cannot be rotated in the same way as before. However, we still have $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\boldsymbol{\nu}_i + \mathbf{F}\gamma_i + \boldsymbol{\varepsilon}_i$, which means that

$$\begin{aligned}\widehat{\mathbf{g}}_{i,M} &= (\widehat{\mathbf{F}}_M'\widehat{\mathbf{F}}_M)^+\widehat{\mathbf{F}}_M'(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_M)\\ &= (\widehat{\mathbf{F}}_M'\widehat{\mathbf{F}}_M)^+\widehat{\mathbf{F}}_M'(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\boldsymbol{\nu}_i + \mathbf{F}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_M)\\ &= (\widehat{\mathbf{F}}_M'\widehat{\mathbf{F}}_M)^+\widehat{\mathbf{F}}_M'[\mathbf{X}_i\boldsymbol{\nu}_i + \mathbf{F}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})].\end{aligned} \tag{C.39}$$

Here,

$$\widehat{\mathbf{F}}_M'\widehat{\mathbf{F}}_M = \overline{\mathbf{C}}_M'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}_M + O_p(N^{-1/2}), \tag{C.40}$$

where the rank of $\widehat{\mathbf{F}}_M'\widehat{\mathbf{F}}_M$ is equal to m, which is also the rank of $\overline{\mathbf{C}}_M'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}_M$. The fact that the rank does not change as the limit is taken implies that

$$(\widehat{\mathbf{F}}_M'\widehat{\mathbf{F}}_M)^+ = (\overline{\mathbf{C}}_M'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}_M)^+ + O_p(N^{-1/2}) \tag{C.41}$$

(see Karabiyik et al., 2017). Insertion into the above expression for $\widehat{\mathbf{g}}_{i,M}$ yields

$$\widehat{\mathbf{g}}_{i,M} = (\overline{\mathbf{C}}_M'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}_M)^+\overline{\mathbf{C}}_M'\mathbf{F}'[\mathbf{X}_i\boldsymbol{\nu}_i + \mathbf{F}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] + O_p(N^{-1/2}), \tag{C.42}$$

which in turn implies

$$\begin{aligned}\widehat{\mathbf{F}}_M\widehat{\mathbf{g}}_{i,M} &= \mathbf{F}\overline{\mathbf{C}}_M(\overline{\mathbf{C}}_M'\mathbf{F}'\mathbf{F}\overline{\mathbf{C}}_M)^+\overline{\mathbf{C}}_M'\mathbf{F}'[\mathbf{X}_i\boldsymbol{\nu}_i + \mathbf{F}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] + O_p(N^{-1/2})\\ &= \mathbf{P}_{F\overline{C}_M}[\mathbf{X}_i\boldsymbol{\nu}_i + \mathbf{F}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] + O_p(N^{-1/2}).\end{aligned} \tag{C.43}$$

Consider $\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}$. Since $\mathbf{M}_{\widehat{F}_M}\overline{\mathbf{X}} = \mathbf{0}_{T\times k}$, we have $\sum_{i=1}^N \mathbf{M}_{\widehat{F}_M}\mathbf{X}_i = N\mathbf{M}_{\widehat{F}_M}\overline{\mathbf{X}} = \mathbf{0}_{T\times k}$. By using this, $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\boldsymbol{\nu}_i + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i$, $\widehat{\mathbf{F}}_M = \mathbf{F}\overline{\mathbf{C}}_M + O_p(N^{-1/2})$ and the independence of $\boldsymbol{\nu}_i$ and $\boldsymbol{\varepsilon}_i$,

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_M &= \left(\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{\widehat{F}_M}\mathbf{X}_i\right)^{-1}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{\widehat{F}_M}\mathbf{y}_i \\
&= \boldsymbol{\beta} + \left(\frac{1}{N}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{\widehat{F}_M}\mathbf{X}_i\right)^{-1}\frac{1}{N}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{\widehat{F}_M}(\mathbf{X}_i\boldsymbol{\nu}_i + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i) \\
&= \boldsymbol{\beta} + \left(\frac{1}{N}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{\widehat{F}_M}\mathbf{X}_i\right)^{-1}\frac{1}{N}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{\widehat{F}_M}[\mathbf{X}_i\boldsymbol{\nu}_i + \mathbf{F}(\boldsymbol{\gamma}_i - \overline{\boldsymbol{\gamma}}) + \boldsymbol{\varepsilon}_i] \\
&= \boldsymbol{\beta} + \left(\frac{1}{N}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{F\overline{C}_M}\mathbf{X}_i\right)^{-1}\frac{1}{N}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{F\overline{C}_M}[\mathbf{X}_i\boldsymbol{\nu}_i + \mathbf{F}(\boldsymbol{\gamma}_i - \overline{\boldsymbol{\gamma}}) + \boldsymbol{\varepsilon}_i] + O_p(N^{-1/2}) \\
&= \boldsymbol{\beta} + \left(\frac{1}{N}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{F\overline{C}_M}\mathbf{X}_i\right)^{-1}\frac{1}{N}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{F\overline{C}_M}\mathbf{F}(\boldsymbol{\gamma}_i - \overline{\boldsymbol{\gamma}}) + O_p(N^{-1/2}). \qquad \text{(C.44)}
\end{aligned}$$

The second term on the right-hand side here is $O_p(1)$. One exception is if one in addition to the conditions of this paper assumes that $\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_i$, where $\boldsymbol{\eta}_i$ is mean zero, independent across i and also independent of all other random elements of the model. In this case, the second term above is $O_p(N^{-1/2})$. However, we do not require loadings to be random, and therefore

$$\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta} = O_p(1). \qquad \text{(C.45)}$$

It follows that

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^N \widehat{\mathbf{F}}_M\widehat{\mathbf{g}}_{i,M} &= \mathbf{P}_{F\overline{C}_M}\frac{1}{N}\sum_{i=1}^N [\mathbf{X}_i\boldsymbol{\nu}_i + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] + O_p(N^{-1/2}) \\
&= \mathbf{P}_{F\overline{C}_M}\frac{1}{N}\sum_{i=1}^N [\mathbf{F}\boldsymbol{\gamma}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] + O_p(N^{-1/2}) \\
&= \mathbf{P}_{F\overline{C}_M}[\mathbf{F}\overline{\boldsymbol{\gamma}} - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] + O_p(N^{-1/2}), \qquad \text{(C.46)}
\end{aligned}$$

where the first term on the right is $O_p(1)$. Together with the asymptotic expansion of $\widehat{\mathbf{F}}\overline{M}\widehat{\mathbf{g}}_{i,\overline{M}}$

provided in the proof of (a), this last result implies

$$
\begin{aligned}
\overline{\mathbf{\Delta}}_M &= \frac{1}{N}\sum_{i=1}^{N}(\widehat{\mathbf{F}}_{\overline{M}}\widehat{\mathbf{g}}_{i,\overline{M}} - \widehat{\mathbf{F}}_M\widehat{\mathbf{g}}_{i,M}) \\
&= \mathbf{F}\overline{\gamma} + \overline{\mathbf{U}}^0_{1,\overline{M}}\overline{\gamma} + (\mathbf{P}_F + \mathbf{P}_{M_F\overline{U}^0_{2,\overline{M}}})\frac{1}{N}\sum_{i=1}^{N}[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,\overline{M}}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \boldsymbol{\beta})] \\
&\quad - \mathbf{P}_{F\overline{C}_M}[\mathbf{F}\overline{\gamma} - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] + O_p(N^{-1/2}) \\
&= \mathbf{F}\overline{\gamma} - \mathbf{P}_{F\overline{C}_M}[\mathbf{F}\overline{\gamma} - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta})] + O_p(N^{-1/2}) \\
&= \mathbf{M}_{F\overline{C}_M}\mathbf{F}\overline{\gamma} + \mathbf{P}_{F\overline{C}_M}\mathbf{F}\overline{\mathbf{\Gamma}}(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}) + O_p(N^{-1/2}) = O_p(1). \qquad \text{(C.47)}
\end{aligned}
$$

The sought result is implied by this. This establishes (b) and hence the proof of the theorem is complete. ■

Proof of Theorem 2.

This proof is similar to that of Theorem 1 in Fujikoshi and Sakurai (2019), or the same theorem in Fujikoshi (2022). We begin by observing that

$$
\begin{aligned}
\mathbb{P}(\widehat{M} = M_0) &= \mathbb{P}\left(\left\{\bigcap_{j\in M_0}|T_{M_j}| > c_N\right\}\bigcap\left\{\bigcap_{j\notin M_0}|T_{M_j}| < c_N\right\}\right) \\
&= 1 - \mathbb{P}\left(\left\{\bigcup_{j\in M_0}|T_{M_j}| \le c_N\right\}\bigcup\left\{\bigcup_{j\notin M_0}|T_{M_j}| \ge c_N\right\}\right) \\
&\ge 1 - \sum_{j\in M_0}\mathbb{P}(|T_{M_j}| \le c_N) - \sum_{j\notin M_0}\mathbb{P}(|T_{M_j}| \ge c_N). \qquad \text{(C.48)}
\end{aligned}
$$

The second term on the right is the probability that the test does not reject even when truly important averages are kicked out. This probability goes to zero as $|T_{M_j}| = O_p(\sqrt{N})$ by Theorem 1 (b) and $c_N/\sqrt{N} \to 0$ by assumption. The third and last term in (C.48) is the probability that the test rejects when redundant averages are kicked out. Here we use in sequence Chebyshev's inequality and Theorem 1 (b), giving

$$
\mathbb{P}(|T_{M_j}| \ge c_N) \le c_N^{-2}\text{var}(T_{M_j}) = c_N^{-2} + o(c_N^{-2}) = O(c_N^{-2}) \to 0, \qquad \text{(C.49)}
$$

as $c_N \to \infty$. Hence, since the two last terms in (C.48) tend to zero,

$$
\mathbb{P}(\widehat{M} = M_0) \to 1, \qquad \text{(C.50)}
$$

as required. ■

Proof of Theorem 3.

The proof of (a) follows from simple manipulations of that of Theorem 1 (a). We begin by observing that H_0 can be formulated as $\mathbf{G} = \mathbf{FH}$, where $\operatorname{rank}\mathbf{H} = m_0 \leq m$. Since $\mathbf{H}$ has full row rank, we have $\mathbf{H}^+ = \mathbf{H}'(\mathbf{HH}')^{-1}$, a $m \times m_0$ matrix, which is such that $\mathbf{HH}^+ = \mathbf{I}_{m_0}$ (see Abadir and Magnus, 2005, Exercise 10.31). It follows that $\mathbf{F}\gamma_i = \mathbf{FHH}^+\gamma_i = \mathbf{G}\mathbf{g}_{i,G}$, where we define $\mathbf{g}_{i,G} = \mathbf{H}^+\gamma_i$ analogously to $\mathbf{g}_{i,M}$ in the Proof of Theorem 1. Therefore,

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\boldsymbol{\nu}_i + \mathbf{G}\mathbf{g}_{i,G} + \boldsymbol{\varepsilon}_i, \tag{C.51}$$

from which we deduce that

$$\begin{aligned}
\widehat{\mathbf{g}}_{i,G} &= (\mathbf{G}'\mathbf{G})^+\mathbf{G}'(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_G) \\
&= (\mathbf{G}'\mathbf{G})^+\mathbf{G}'(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\boldsymbol{\nu}_i + \mathbf{G}\mathbf{g}_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_G) \\
&= \mathbf{g}_{i,G} + (\mathbf{G}'\mathbf{G})^+\mathbf{G}'[\mathbf{X}_i\boldsymbol{\nu}_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})].
\end{aligned} \tag{C.52}$$

where $\widehat{\boldsymbol{\beta}}_G$ is $\widehat{\boldsymbol{\beta}}_M$ with $\mathbf{G}$ in place of $\widehat{\mathbf{F}}_M$. If $\mathbf{A}$ has full column rank and $\mathbf{B}$ has full row rank, then $(\mathbf{AB})^+ = \mathbf{B}^+\mathbf{A}^+$ (see Abadir and Magnus, 2005, Exercise 10.36). Applying this twice to $(\mathbf{G}'\mathbf{G})^+ = (\mathbf{H}'\mathbf{F}'\mathbf{FH})^+$ yields $(\mathbf{F}'\mathbf{FH})^+\mathbf{H}'^+ = \mathbf{H}^+(\mathbf{F}'\mathbf{F})^{-1}\mathbf{H}'^+$. By putting these results together,

$$\begin{aligned}
\mathbf{G}\widehat{\mathbf{g}}_{i,G} - \mathbf{F}\gamma_i &= \mathbf{G}(\widehat{\mathbf{g}}_{i,G} - \mathbf{g}_{i,G}) \\
&= \mathbf{G}(\mathbf{G}'\mathbf{G})^+\mathbf{G}'[\mathbf{X}_i\boldsymbol{\nu}_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})] \\
&= \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'[\mathbf{X}_i\boldsymbol{\nu}_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})] \\
&= \mathbf{P}_F[\mathbf{X}_i\boldsymbol{\nu}_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})],
\end{aligned} \tag{C.53}$$

which in turn implies

$$
\begin{aligned}
\boldsymbol{\Delta}_{i,G} &= \widehat{\mathbf{F}}_{\overline{M}}\widehat{\mathbf{g}}_{i,\overline{M}} - \mathbf{G}\widehat{\mathbf{g}}_{i,G} \\
&= \overline{\mathbf{U}}^0_{1,\overline{M}}\gamma_i + (\mathbf{P}_F + \mathbf{P}_{M_F\overline{U}^0_{2,\overline{M}}})[\mathbf{X}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,\overline{M}}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \boldsymbol{\beta})] \\
&- \mathbf{P}_F[\mathbf{X}_i\boldsymbol{\nu}_i + \boldsymbol{\varepsilon}_i - \mathbf{X}_i(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \boldsymbol{\beta})] + O_p(N^{-1/2}) \\
&= \mathbf{M}_F\overline{\mathbf{U}}^0_{1,\overline{M}}\gamma_i - \mathbf{P}_F\mathbf{X}_i(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \widehat{\boldsymbol{\beta}}_{\overline{M}}) \\
&+ \mathbf{P}_{M_F\overline{U}^0_{2,\overline{M}}}[\mathbf{V}_i\boldsymbol{\nu}_i - \overline{\mathbf{U}}^0_{1,\overline{M}}\gamma_i + \boldsymbol{\varepsilon}_i - \mathbf{V}_i(\widehat{\boldsymbol{\beta}}_{\overline{M}} - \boldsymbol{\beta})] + O_p(N^{-1/2}) \\
&= \mathbf{E}_{i,G} + O_p(N^{-1/2}),
\end{aligned}
\tag{C.54}
$$

where we again use notation that is analogous to that in Proof of Theorem 1. While not the same, $\mathbf{E}_{i,G}$ has the same properties as $\mathbf{E}_{i,M}$ in that proof. Asymptotic normality therefore follows by the same arguments. The proof of part (b) is almost identical to that of Theorem 1 (b) and is therefore omitted. ■

D Monte Carlo study

In this section we report the results of a small-scale Monte Carlo study into the small-sample accuracy of our theoretical results. The data generating process used for this purpose is given by a highly simplified version of the one given in equations (1) and (2) of the main paper, and sets $m_0 = 2$, $\boldsymbol{\beta} = \mathbf{0}_{k\times 1}$ and $\boldsymbol{\nu}_i \sim N(\mathbf{0}_{k\times 1}, \omega^2\mathbf{I}_k)$. Two values of ω^2 are considered, 0 (slope homogeneity) and 0.04 (slope heterogeneity), as in Pesaran (2006). We further set $k = 3$, so that the maximum number of cross-sectional averages is given by $k + 1 = 4$. Similarly to Bai and Ng (2002), the elements of $\boldsymbol{\varepsilon}_i$ are allowed to be weakly serially correlated through the following autoregressive specification:

$$
\varepsilon_{i,t} = \rho\varepsilon_{i,t-1} + u_{i,t}, \tag{D.55}
$$

where $\varepsilon_{1,0} = ... = \varepsilon_{N,0} = 0$, $\rho \in \{0, 0.5\}$ and $u_{i,t} \sim N(0,1)$. The elements of $\mathbf{V}_i$ and $\mathbf{F}$ are independently drawn from $N(0,1)$. For the loadings, we generate

$$
\gamma_i = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} + \xi_i, \tag{D.56}
$$

$$
\boldsymbol{\Gamma}_i = \begin{bmatrix} 0.5 & \mathbf{0}_{1\times(k+1-m_0)} \\ 1 & \mathbf{0}_{1\times(k+1-m_0)} \end{bmatrix} + \xi_i\mathbf{1}_{1\times k}, \tag{D.57}
$$

where $\xi_i \sim N(0,1)$. This way we ensure not only that γ_i and $\mathbf{\Gamma}_i$ are correlated but also that

$$\mathbb{E}(\mathbf{C}_i) = \mathbb{E}([\gamma_i + \mathbf{\Gamma}_i \boldsymbol{\beta}_i, \mathbf{\Gamma}_i]) = \mathbb{E}([\gamma_i, \mathbf{\Gamma}_i]) = \begin{bmatrix} 1 & 0.5 & \mathbf{0}_{1\times(k+1-m_0)} \\ 0.5 & 1 & \mathbf{0}_{1\times(k+1-m_0)} \end{bmatrix}, \tag{D.58}$$

which means that asymptotically only the first $m_0 = 2$ columns of $\overline{\mathbf{Z}}$ load on $\mathbf{F}$, that is, $M_0 = \{1,2\}$. These are therefore the averages we want to single out from the rest.

We begin by considering the results reported in Table 1, which contains empirical rejection frequencies for the T_M test at the 5% significance level. When $M = M_0$, these represent size. An important aspect when it comes to power is the direction at which power is evaluated, which is apparent from the proof of Theorem 1 (b). Intuitively, since the test is based on the estimated model for $\mathbf{y}_i$, the T_M test will have relatively high (low) power in the direction of erroneously excluding $\overline{\mathbf{y}}$ ($\overline{\mathbf{X}}$). In Table 1, we therefore consider both the case when $M = \{1\}$, in which the first column of $\overline{\mathbf{X}}$ is erroneously excluded, and $M = \{2\}$, in which $\overline{\mathbf{y}}$ is erroneously excluded. Hence, in both these cases the reported rejection frequencies represent power. All results are based on 1,000 replications.

We begin by noting that size accuracy is generally quite good, which is just as expected given Theorem 1 (a). Of course, accuracy is not perfect, and there are some distortions. Most of these are, however, not larger than that they can be attributed to simulation uncertainty. As expected given the above discussion, power depends on the direction in which it is evaluated, and it is highest when evaluated in the direction of erroneously excluding $\overline{\mathbf{y}}$. Unreported results confirm that power is even higher if both $\overline{\mathbf{y}}$ and the first column of $\overline{\mathbf{X}}$ are excluded.

We now move on to the sequential procedure to determine M_0. We have argued that in order to eliminate the risk of over-specification asymptotically the critical value, c_N, should be allowed to increase with N. Of course, in practice N and T are always fixed, and hence so is c_N. In the structural break literature it is therefore common to employ fixed critical values even if theory requires that they grow with T (see, for example, Bai, 1999). The fact that in practice N and T are always fixed is one reason for not allowing c_N to grow with N. Another reason is that we have seen that the T_M test has poor power in the direction of erroneously excluding $\overline{\mathbf{X}}$. This will cause the sequential procedure to underestimate M_0, as averages associated with small values of T_M are dropped. For these reasons, in this section the sequential procedure is

Table 1: Size and power of the T_M test at the 5% significance level.

			$\omega^2 = \rho = 0$			$\omega^2 = 0.04, \rho = 0.5$		
N	T	M	M_0	$\{1\}$	$\{2\}$	M_0	$\{1\}$	$\{2\}$
50	10		5.0	8.2	77.1	4.4	5.7	73.8
100	10		5.1	15.8	80.9	6.4	12.7	76.8
200	10		4.9	32.3	87.1	5.3	28.0	84.9
400	10		7.4	49.4	90.0	6.2	42.2	89.7
50	20		3.7	8.0	80.9	3.7	6.3	77.6
100	20		3.0	16.3	82.9	2.0	13.4	82.3
200	20		2.4	32.2	87.4	2.9	29.0	86.3
400	20		2.6	54.4	91.0	2.6	51.2	90.3

Notes: ρ and ω^2 are parameters that measure the degree of error serial correlation and coefficient heterogeneity, respectively. The correct index set is given by $M_0 = \{1, 2\}$. The results for the case when $M = M_0$ represent size, while those for $M = \{1\}$ and $M = \{k+1\}$ represent power.

implemented using the same normal critical value as in the T_M test.

Table 2 report some results on the correct selection frequency for $\widehat{M}$ and $\widehat{m}$, and the average $\widehat{m}$ across the 1,000 replications. As expected given the discussion of the last paragraph, we see that the average $\widehat{m}$ approaches m_0 from below. The procedure therefore has a tendency to under-specify the model in small samples. This is reflected in the correct selection frequencies, which can be quite low. However, we also see that accuracy increases quite quickly as N grows, which is presumably a reflection of Theorem 2 and the consistency of $\hat{M}$.

Table 2: Correct selection frequency for $\widehat{M}$ and $\widehat{m}$, and average $\widehat{m}$.

		$\omega^2 = \rho = 0$			$\omega^2 = 0.04, \rho = 0.5$		
N	T	Corr $\widehat{M}$	Corr $\widehat{m}$	Mean $\widehat{m}$	Corr $\widehat{M}$	Corr $\widehat{m}$	Mean $\widehat{m}$
50	10	0.206	0.110	1.340	0.177	0.085	1.296
100	10	0.285	0.194	1.473	0.253	0.165	1.438
200	10	0.366	0.299	1.637	0.350	0.263	1.641
400	10	0.438	0.384	1.776	0.421	0.359	1.766
50	20	0.205	0.146	1.281	0.191	0.118	1.258
100	20	0.289	0.235	1.391	0.268	0.211	1.359
200	20	0.378	0.319	1.567	0.372	0.305	1.538
400	20	0.521	0.482	1.772	0.507	0.454	1.723

Notes: ρ and ω^2 are parameters that measure the degree of error serial correlation and coefficient heterogeneity, respectively. The correct index set is given by $M_0 = \{1, 2\}$. "Corr $\widehat{M}$" and "Corr $\widehat{m}$" refer to the frequency with which M_0 and m_0 are correctly selected, while "Mean $\widehat{m}$" refers to the average $\widehat{m}$ across replications.

References

Abadir, K. M., and J. R. Magnus (2005). *Matrix algebra, Econometric exercises 1*. Cambridge University Press, New York.

Andrews, D. W. K., and B. Lu (2001). Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics* **101**, 123–164.

Bai, J. (1999). Likelihood ratio tests for multiple structural changes. *Journal of Econometrics* **91**, 299–323.

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

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

Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics* **24**, 295–300.

Fujikoshi, Y. (2022). High-dimensional consistencies of KOO methods in multivariate regression model and discriminant analysis. *Journal of Multivariate Analysis* **188**, 104860.

Fujikoshi, Y., and T. Sakurai (2019). Consistency of test-based method for selection of variables in high-dimensional two-group discriminant analysis. *Japanese Journal of Statistics and Data Science* **2**, 155–171.

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.

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

Westerlund, J., and Y. Kaddoura (2022). CCE in heterogenous fixed-T panels. *Econometrics Journal* **25**, 719–738.

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