A method for evaluating the rank condition for CCE estimators

De Vos, Ignace and Everaert, Gerdie and Sarafidis, Vasilis

VU Amsterdam, Ghent University, BI Norwegian Business School

1 April 2021

A method for evaluating the rank condition for CCE estimators

Ignace De Vos*
Department of Econometrics and Data Science, VU Amsterdam
and
Gerdie Everaert
Department of Economics, Ghent University
and
Vasilis Sarafidis
Department of Economics, BI Norwegian Business School

Abstract

This paper proposes a binary classifier to evaluate the rank condition (RC) that is required for consistency of the Common Correlated Effects (CCE) estimator. The RC postulates that the number of unobserved factors, m, is not larger than the rank of the unobserved matrix of average factor loadings, ϱ. The key insight in this paper is that ϱ can be consistently estimated with existing techniques through the matrix of cross-sectional averages of the data. Similarly, m can be estimated consistently from the data using existing methods. A binary classifier, constructed by comparing estimates of m and ϱ, correctly determines whether the RC is satisfied or not as $(N, T) \to \infty$. We illustrate the practical relevance of testing the RC by studying the effect of the Dodd-Frank Act on bank profitability.

Keywords: common factors, common correlated effects approach, rank condition

*The authors thank Joakim Westerlund, Alexander Chudik, George Kapetanios and Arturas Juodis for helpful comments and discussions. This paper has also benefited from presentations at the 2017, 2018 International Panel Data Conference, the 2018 Asian and European Meetings of the Econometric Society and the 2019 Panel Data Workshop in Amsterdam. Ignace De Vos acknowledges financial support from the Ghent University BOF research fund. Ignace De Vos and Gerdie Everaert further acknowledge financial support from the National Bank of Belgium.

1 Introduction

In a seminal paper, Pesaran (2006) put forward the Common Correlated Effects (CCE) approach for $\sqrt{NT}$-consistent estimation of panel data models with a multifactor error structure. The method aims to control for the unobserved common factors by augmenting the regression model with cross-sectional averages (CSA) of the observables. The CCE estimator has been applied to a large range of fields[1], and it has also been extended to several theoretical settings.[2] Such popularity of CCE can be attributed to the computational simplicity as well as the excellent finite-sample performance of the estimator in stylised setups.

Nevertheless, CCE comes at a cost. In particular, the CSA of the observables are valid proxies for the unobserved factors only if the number of factors, m, does not exceed the rank of the matrix of averaged factor loadings, ϱ. This so-called "rank condition" (RC) implies that there exist at least as many observables holding linearly independent information about the unobserved factors as there are factors. Westerlund and Urbain (2013) demonstrate that the CCE estimator is inconsistent when the RC fails and the factor loadings are correlated with the regressors. Furthermore, Karabiyik et al. (2019) and Juodis et al. (2021) show that even when the factor loadings are uncorrelated with the regressors, failure of the RC leads to a lower rate of consistency for the CCE estimator. Despite the importance of the RC for the asymptotic properties of the CCE estimator, practitioners typically take this assumption for granted. The main reason is that the matrix of average factor loadings is unobserved and therefore its rank cannot be directly evaluated or estimated.

This paper puts forward a binary classifier that evaluates the rank condition. The key insight is that the rank of the unobserved matrix of average factor loadings, ϱ, can be established from the matrix of CSA of the data. We shall show that ϱ can be estimated consistently using existing procedures developed for determining the true rank of an unknown matrix; see e.g. Camba-Mendez and Kapetanios (2009) and Al-Sadoon (2017) for

[1] A recent search on Google Scholar indicated that the number of empirical applications based on CCE estimation currently exceeds one thousand.

[2] See e.g. Kapetanios et al. (2011), Su and Jin (2012), Chudik and Pesaran (2015), Norkute et al. (2020), Harding et al. (2020) and De Vos and Everaert (2021), to mention a few.

an overview of this literature. Similarly, the number of factors, m, can be estimated from the data in a straightforward manner using existing methods, such as those developed by Onatski (2010), Ahn and Horenstein (2013) and Kapetanios (2010), among many others. Comparing consistent estimates of m and ϱ, $\widehat{m}$ and $\widehat{\varrho}$ respectively, the rank condition is deemed to be satisfied when the classifier $\widehat{RC} \equiv 1 - \mathbb{1}\{\widehat{\varrho} < \widehat{m}\} = 1$, where $\mathbb{1}\{\cdot\}$ is an indicator function that returns 1 when the argument inside the curly brackets holds true and 0 otherwise. $\widehat{RC}$ is shown to be consistent, i.e. it correctly determines whether the rank condition is satisfied or not, with probability 1 as $(N, T) \to \infty$.

When the RC is violated for the standard CCE approach, one can augment the model with additional CSA that contain new information about the factors. Several potential augmentations have been suggested (see e.g. Chudik and Pesaran, 2015; Karabiyik et al., 2019). However, it is not always clear which set of additional CSA to choose, and whether the selected augmentation is sufficient to restore the RC.[3] To address these issues, we put forward a strategy that combines the classifier proposed in the present paper and the IC criterion of Karabiyik et al. (2019). The resulting procedure enables consistent CCE estimation of panel data models with a multifactor error structure, even in cases where the rank condition fails for the original CCE estimator.

We illustrate the practical relevance of our RC classifier and augmentation strategy by studying the effect of the Dodd-Frank Act of 2010 on bank profitability. In particular, based on a random sample of 450 banks we analyse bank profitability conditional on several potential drivers, controlling for macro-risk factors and common shocks. To examine the impact of the Dodd-Frank Act, we estimate the model separately over two subperiods, namely 2006:Q1-2010:Q4 and 2011:Q1-2019:Q4. The RC classifier reveals that the rank condition fails for the first subperiod. By augmenting the standard set of CSA using external variables, our procedure is able to restore the rank condition. This proves to be important because the estimated effect of bank size on profitability is significantly lower

[3]The strategy of choosing all possible additional CSA at hand, in the hope of satisfying the RC, can lead to a different problem. In particular, as shown by Karabiyik et al. (2017) and Juodis et al. (2021), when too many CSA are employed, the CCE estimator can suffer from bias or it may have a reduced convergence rate.

when the RC is restored.

In what follows we will use $\mathbf{A}^{\dagger}$ to denote the Moore-Penrose pseudo-inverse of the matrix $\mathbf{A}$, $rk(\mathbf{A})$ for its rank, $|\mathbf{A}|$ for the determinant and $\|\mathbf{A}\| = [tr(\mathbf{A}\mathbf{A}')]^{1/2}$ for its Euclidean (Frobenius) matrix norm. A vec(.) denotes the vectorization operation. Finally, $\lfloor a \rfloor$ ($\lceil a \rceil$) is the floor (ceiling) function, which yields the largest (smallest) integer less than (greater than) or equal to a.

2 A multifactor panel data model and CCE

2.1 Model and assumptions

We study the following linear regression model with unobserved common factors

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

where $\mathbf{y}_i = [y_{i1}, \ldots, y_{iT}]'$ denotes a $T \times 1$ vector of observations on the dependent variable for individual i, $\mathbf{X}_i = [\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}]'$ denotes a $T \times K$ matrix of covariates, where $\mathbf{x}_{it}$ is $K \times 1$, and $\boldsymbol{\beta}$ is a $K \times 1$ vector of unknown parameters of interest with $\|\boldsymbol{\beta}\| < \infty$. The error term is composite, such that $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_T]'$ denotes a $T \times m$ matrix of unobserved common factors, where $\mathbf{f}_t$ is $m \times 1$, and $\boldsymbol{\lambda}_i$ denotes an $m \times 1$ vector of factor loadings. Finally, $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1}, \ldots, \varepsilon_{iT}]'$ is a $T \times 1$ vector of purely idiosyncratic disturbances.

Following Pesaran (2006), we assume that the covariates are also subject to a common factor structure, such that the data generating process (DGP) for $\mathbf{X}_i$ is given by

$$\mathbf{X}_i = \mathbf{F}\boldsymbol{\Gamma}_i + \mathbf{V}_i, \tag{2}$$

where $\boldsymbol{\Gamma}_i$ denotes an $m \times K$ matrix of factor loadings, and $\mathbf{V}_i = [\mathbf{v}_{i1}, \ldots, \mathbf{v}_{iT}]'$ is a $T \times K$ matrix of idiosyncratic errors.

Replacing $\mathbf{X}_i$ in Eq. (1) by the expression in Eq. (2), and stacking the observables into a

$T \times (K+1)$ matrix $\mathbf{Z}_i = [\mathbf{y}_i, \mathbf{X}_i] \equiv [\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}]'$, yields

$$\mathbf{Z}_i = \mathbf{F}\mathbf{C}_i + \mathbf{U}_i, \tag{3}$$

where $\mathbf{C}_i = [\boldsymbol{\delta}_i, \boldsymbol{\Gamma}_i]$ is of order $m \times (K+1)$ with $\boldsymbol{\delta}_i = \boldsymbol{\lambda}_i + \boldsymbol{\Gamma}_i\boldsymbol{\beta}$, and $\mathbf{U}_i = [\boldsymbol{\varepsilon}_i + \mathbf{V}_i\boldsymbol{\beta}\ , \mathbf{V}_i]$. In what follows, it is important to note that $\mathbf{C}_i$ can be written as $\mathbf{C}_i = \widetilde{\mathbf{C}}_i\mathbf{B}$, with

$$\widetilde{\mathbf{C}}_i = [\boldsymbol{\lambda}_i, \boldsymbol{\Gamma}_i]\,; \qquad \mathbf{B} = \begin{bmatrix} 1 & \mathbf{0}_{1\times K} \\ \boldsymbol{\beta} & \mathbf{I}_K \end{bmatrix}. \tag{4}$$

Therefore, since $\mathbf{B}$ has full rank, the rank of $\mathbf{C}_i$ is solely determined by the matrix of factor loadings $\widetilde{\mathbf{C}}_i$.

The following assumptions are made throughout the paper:

Assumption 1. *(Idiosyncratic errors) ε_{it} and $\mathbf{v}_{it}$ are mean zero, covariance-stationary and independent across i, with $E(\varepsilon_{it}^4) < \infty$ and $E(\|\mathbf{v}_{it}\|^4) < \infty$ for all i and t.*

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

Assumption 3. *(Factor loadings) $\widetilde{\mathbf{C}}_i$ is generated according to*

$$\widetilde{\mathbf{C}}_i = \widetilde{\mathbf{C}} + \boldsymbol{\Xi}_i; \qquad \boldsymbol{\xi}_i \sim i.i.d.(\mathbf{0}_{m(K+1)}, \boldsymbol{\Omega}_\xi), \tag{5}$$

where $\widetilde{\mathbf{C}} = E(\widetilde{\mathbf{C}}_i) \equiv [\boldsymbol{\lambda}, \boldsymbol{\Gamma}]$ such that $\left\|\widetilde{\mathbf{C}}\right\| < \infty$, $\boldsymbol{\xi}_i = vec(\boldsymbol{\Xi}_i)$ and $\boldsymbol{\Omega}_\xi = E(\boldsymbol{\xi}_i\boldsymbol{\xi}_i')$ with $\|\boldsymbol{\Omega}_\xi\| < \infty$. In addition, $\frac{1}{N}\sum_{i=1}^N \mathbf{C}_i\mathbf{C}_i' \to \boldsymbol{\Sigma}_C$ as $N \to \infty$, with $\boldsymbol{\Sigma}_C$ positive definite.

Assumption 4. *(Independence) $\mathbf{f}_t, \varepsilon_{is}, \mathbf{v}_{jl}, \boldsymbol{\xi}_h$ are mutually independent for all t, i, s, j, l, h.*

The setup described by the DGP in Eq. (3) together with Assumptions 1-4, is similar to that in Pesaran (2006) but deviates in the following respects. First, we focus on a model with homogeneous slope coefficients and without fixed effects. This is for ease of exposition only, as the results below also follow through under the assumption of independent random

coefficients with a common mean, as in Pesaran (2006). Second, following Westerlund and Urbain (2013) and Karabiyik et al. (2019), Assumption 3 generalizes Pesaran (2006) by allowing $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Gamma}_i$ to be correlated within (but not across) cross-section units i. Third, we introduce more explicit regularity conditions on the factors and their loadings compared to what is typically the case in the CCE literature. In particular, the non-central second moments are assumed to converge to a positive definite matrix. Such regularity conditions are common in the factor literature (see e.g. Bai and Ng, 2002) and allow us to consistently estimate m. Lastly, note that $rk(\mathbf{F}) = m$ of Assumption 2 implies that $T \geq m$.

2.2 CCE and the rank condition

Since $\mathbf{F}$ enters into the data generating process of both $\mathbf{y}_i$ and $\mathbf{X}_i$, and $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Gamma}_i$ are allowed to be mutually correlated, $\mathbf{X}_i$ is endogenous. Therefore, standard panel data estimators, such as the two-way fixed estimator, fail to be consistent for the parameters of interest, $\boldsymbol{\beta}$. The key idea of CCE is to replace $\mathbf{F}$ with CSA of the observables in Eq. (3).

In particular, taking sample averages over i in Eq. (3), we obtain

$$\underset{T\times(K+1)}{\overline{\mathbf{Z}}} = \underset{T\times m}{\mathbf{F}} \; \underset{m\times(K+1)}{\overline{\mathbf{C}}} + \underset{T\times(K+1)}{\overline{\mathbf{U}}}, \tag{6}$$

where $\overline{\mathbf{Z}} = [\overline{\mathbf{z}}_1, \ldots, \overline{\mathbf{z}}_T]'$, $\overline{\mathbf{U}} = [\overline{\mathbf{u}}_1, \ldots, \overline{\mathbf{u}}_T]'$ and bars denote CSA as in $\overline{\mathbf{Z}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{Z}_i$.

Under Assumptions 1-4 it is easy to show that $\overline{\mathbf{C}} = \mathbf{C} + O_p(N^{-1/2})$, where $\mathbf{C} = E(\mathbf{C}_i)$, and $\|\overline{\mathbf{u}}_t\| = O_p(N^{-1/2})$ for all $t = 1, \ldots, T$. As a result, the observed CSA converge to a linear combination of the m common factors at every $t = 1, \ldots, T$:

$$\overline{\mathbf{z}}_t = \mathbf{C}'\mathbf{f}_t + (\overline{\mathbf{C}} - \mathbf{C})'\mathbf{f}_t + \overline{\mathbf{u}}_t = \mathbf{C}'\mathbf{f}_t + O_p(N^{-1/2}). \tag{7}$$

Suppose that $\mathbf{C}$ has full rank such that $\mathbf{C}\mathbf{C}'$ is invertible and bounded by Assumption 3. Pre-multiplying Eq. (7) by $\mathbf{C}$ and solving for $\mathbf{f}_t$ yields

$$\mathbf{f}_t = (\mathbf{C}\mathbf{C}')^{-1}\mathbf{C}\overline{\mathbf{z}}_t + O_p(N^{-1/2}). \tag{8}$$

Hence, as $N \to \infty$, the common factor component at time t can be controlled for (or estimated) with the cross-section averages $\overline{\mathbf{z}}_t$.

The pooled CCE estimator for $\boldsymbol{\beta}$ is the least-squares estimator given by

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M} \mathbf{y}_i, \tag{9}$$

where $\mathbf{M} = \mathbf{I}_T - \overline{\mathbf{Z}}(\overline{\mathbf{Z}}'\overline{\mathbf{Z}})^{\dagger}\overline{\mathbf{Z}}'$.[4]

The above idea of estimating factors with CSA crucially relies upon the assumption that $\mathbf{C}$ has full rank. This restriction, known as the "rank condition" (RC), corresponds to

$$\varrho = m \tag{10}$$

where $\varrho = rk(\mathbf{C})$. When $\varrho < m$, the RC fails and the CCE estimator is generally inconsistent. This is because the CSA do not contain enough information on $\mathbf{f}_t$, which implies that the factor estimator in Eq. (8) does not exist.

There are several cases where the RC may fail. To begin with, such failure occurs when $m > K + 1$, i.e. the number of factors is larger than the number of CSA, in which case $\varrho \leq min\{m, K+1\} = K + 1 < m$. In addition, although $K + 1 \geq m$ is a necessary condition for the RC to hold, it is by no means sufficient. For example, certain columns of $\overline{\mathbf{Z}}$ can be asymptotically uninformative because the corresponding observables: (i) do not load on the common factors (e.g. some of the columns in $\boldsymbol{\Gamma}_i$ equal zero); (ii) have factor loadings that average out (e.g. $\overline{\boldsymbol{\Gamma}} = O_p(N^{-1/2})$); or (iii) do not contain information on the common factors that is distinct from that already provided by other observables. In all these cases, the number of columns that are informative to estimate $\mathbf{f}_t$, as measured by ϱ, can be lower than m.[5]

[4] When the model contains fixed effects, then one may set $\mathbf{M} = \mathbf{I}_T - \overline{\mathbf{H}}(\overline{\mathbf{H}}'\overline{\mathbf{H}})^{\dagger}\overline{\mathbf{H}}'$, where $\overline{\mathbf{H}} = [\boldsymbol{\iota}_T, \overline{\mathbf{Z}}]$ and $\boldsymbol{\iota}_T$ is a $T \times 1$ vector of ones.

[5] Strictly speaking, the dimension of the vector space spanned by $\overline{\mathbf{Z}}$ will be lower than m in these cases.

3 Evaluation of the rank condition

Despite the importance of the RC for the properties of the CCE estimator, this assumption is typically taken for granted. The main reason is that the population mean of the matrix of factor loadings, $\mathbf{C}$, is unobserved and therefore its rank cannot be directly evaluated or estimated. The key insight of this paper is that ϱ can be determined by estimating the rank of $\overline{\mathbf{Z}}$ using existing techniques. Given a consistent estimate of ϱ, the RC is evaluated by direct comparison of that value with a consistent estimate for m. The latter can be determined from the observed data in a straightforward manner, based on e.g. Bai and Ng (2002), Alessi et al. (2010), Onatski (2010) and Ahn and Horenstein (2013).

The following two sections provide details for consistent estimation of ϱ and m as $(N,T) \to \infty$. Section 3.3 puts forward a binary classifier that evaluates the rank condition correctly with probability 1 as $(N,T) \to \infty$. Section 3.4 discusses a strategy for obtaining a consistent CCE estimator when the RC fails.

3.1 Consistent estimation of ϱ

We make use of the fact that the rank of an unobserved matrix $\mathbf{A}$ can be determined through a $\sqrt{N}$-consistent estimator of that matrix $\mathbf{A}_N = \mathbf{A} + O_p(N^{-1/2})$; see e.g. Robin and Smith (2000) and Kleibergen and Paap (2006). Noting that $\overline{\mathbf{Z}} = \mathbf{FC} + O_p(N^{-1/2})$ is $\sqrt{N}$-consistent for $\mathbf{FC}$ (for T fixed), then from the rank equivalence

$$rk\left(\mathbf{FC}\right) = rk\left(\mathbf{C}\right) = \varrho, \tag{11}$$

it follows that ϱ can be consistently estimated by applying a rank estimator to $\overline{\mathbf{Z}}$.[6]

Many popular rank estimators are based on sequential testing procedures or on information criteria (IC). Camba-Mendez and Kapetanios (2009) provide an overview and conclude that sequential testing procedures have an advantage over IC methods under several modeling scenarios. Therefore, in the remainder of this section, we closely follow the sequential

[6]Note that the first equality in (11) follows from the fact that $\mathbf{F}$ has full column rank by Assumption 2. See exercise 4.25 on pg. 85 in Abadir and Magnus (2005).

testing procedure developed by Robin and Smith (2000). This is easy to implement and relies on relatively mild assumptions, in that it does not require the variance-covariance of the estimator of the unknown matrix $\mathbf{FC}$ to be full rank, or its rank to be known.

A major complication that arises in our setting is that unlike Robin and Smith (2000), where the dimensions of the target matrix are fixed as the sample size grows, here $\mathbf{FC}$ and its estimator $\overline{\mathbf{Z}}$ are of order $T \times n$.[7] Therefore, the number of rows increases with the time dimension such that $\overline{\mathbf{Z}} = \mathbf{FC} + O_p(\sqrt{T}N^{-1/2})$ is not $\sqrt{N}$-consistent when $(N,T) \to \infty$. To circumvent this issue, we introduce a narrow matrix $\mathbf{\Psi}$ of order $n \times T$, such that $\mathbf{\Psi}\overline{\mathbf{Z}}$ is $n \times n$ and $rk(\mathbf{\Psi FC}) = rk(\mathbf{FC})$. That is, $\mathbf{\Psi}$ has the role of reducing the dimension of $\overline{\mathbf{Z}}$ without altering the rank of the matrix it estimates. The following assumption is imposed:

Assumption 5. *(Dimension reduction matrix)* $\mathbf{\Psi}$ *satisfies as* $(N,T) \to \infty$

$$
\begin{aligned}
&(i) \quad \|\mathbf{\Psi F}\| = O_p(1); \qquad (ii) \quad \left\|\mathbf{\Psi}\overline{\mathbf{U}}\right\| = O_p(N^{-1/2}); \\
&(iii) \ \sqrt{N}\mathrm{vec}(\mathbf{\Psi}\overline{\mathbf{Z}} - \mathbf{\Psi FC}) \to^{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{\Omega}).
\end{aligned}
$$

Assumption 5 places additional restrictions on the potential choices for $\mathbf{\Psi}$, besides it being rank preserving. Assumption 5(i) requires that the entries of $\mathbf{\Psi}$ are sufficiently bounded. Assumption 5(ii) states that $\mathbf{\Psi}$ is asymptotically uncorrelated with $\overline{\mathbf{U}}$, the error term in Eq. (6). Assumption 5(iii) ensures that, by application of a suitable central limit theorem, $\mathbf{\Psi}\overline{\mathbf{Z}}$ remains $\sqrt{N}$-consistent for $\mathbf{\Psi FC}$ and asymptotically normally distributed as $(N,T) \to \infty$. This assumption is identical to Assumption 2.2 in Robin and Smith (2000), except that it is imposed on $\mathbf{\Psi}\overline{\mathbf{Z}}$ rather than $\overline{\mathbf{Z}}$ itself.

In practice, there exist several options for $\mathbf{\Psi}$ that satisfy the rank preservation condition and Assumption 5. One option is to set $\mathbf{\Psi} = T^{-1/2}\mathbf{\Phi}$, with the entries of $\mathbf{\Phi}$ drawn from the standard normal distribution. The following theorem confirms that this choice is rank-preserving and satisfies the required consistency and boundedness conditions of Assumption

[7] Hereafter, we change the notation for the number of columns of $\overline{\mathbf{Z}}$ from $K+1$ to n. This is in order to accommodate for the need to augment $\overline{\mathbf{Z}}$ with additional CSA when the rank condition fails, as described in Section 3.4.

5. In turn, Assumption 5(*iii*) is easily seen to hold by application of a CLT.

Theorem 1. *Let $T > n$ and $\mathbf{\Phi}$ be a $n \times T$ random matrix with i.i.d. standard normal entries.*

(i) For a $T \times n$ matrix $\mathbf{A}$, it holds that

$$Pr\left[rk(\mathbf{\Phi A}) = rk(\mathbf{A})\right] = 1.$$

(ii) Let $\mathbf{\Psi} = T^{-1/2}\mathbf{\Phi}$. Under Assumptions 1-4, as $(N,T) \to \infty$, it follows that

$$\mathbf{\Psi}\overline{\mathbf{Z}} = \mathbf{\Psi F C} + O_p(N^{-1/2}),$$

where $\|\mathbf{\Psi F C}\| = O_p(1)$.

The proof of Theorem 1 is in Appendix B.

Remark 3.1. An alternative stochastic option would be $\mathbf{\Psi} = \overline{\mathbf{Z}}'/T$. However, this is ruled out because even though $\mathbf{\Psi}\overline{\mathbf{Z}} = \overline{\mathbf{Z}}'\overline{\mathbf{Z}}/T$ is stochastically bounded and has the same rank as $\overline{\mathbf{Z}}$, it does not have an asymptotic normal distribution, i.e. it violates Assumption 5(*iii*).

The $\mathbf{\Psi}$ matrix can also be deterministic. Since n time periods contain the same information on the rank of $\mathbf{C}$ as do T observations, an obvious candidate is $\mathbf{\Psi} = [\mathbf{0}_{n\times(T-n)},\ \mathbf{I}_n]$, which considers only the last n rows of $\overline{\mathbf{Z}}$. One can also take averages over every n-th row in $\overline{\mathbf{Z}}$ by setting $\mathbf{\Psi} = \frac{1}{\lceil T/n \rceil}[\boldsymbol{\iota}'_{\lceil T/n \rceil} \otimes \mathbf{I}_n]\mathbf{I}_{\lceil T/n \rceil n, T}$, where $\boldsymbol{\iota}_a$ is an $a \times 1$ vector of ones.

Given the above, we propose estimating the rank of the $n \times n$ matrix $\mathbf{\Psi}\overline{\mathbf{Z}}$ by sequentially testing the null hypothesis $H_0 : \varrho = \varrho^*$ against the alternative $H_a : \varrho > \varrho^*$, using the following statistic:

$$\tau = N \sum_{\ell=\varrho^*+1}^{n} \lambda_\ell(\mathbf{A}), \tag{12}$$

where $\lambda_1(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$ are the ordered eigenvalues of $\mathbf{A} \equiv \mathbf{\Psi}\overline{\mathbf{Z}}\overline{\mathbf{Z}}'\mathbf{\Psi}'$. The procedure is implemented sequentially for $\varrho^* = 0, \ldots, n-1$ and the estimated rank $\widehat{\varrho}$ corresponds to the smallest value of ϱ^* for which the null hypothesis is not rejected. Under the null, τ has

a limiting distribution which is a weighted sum of independent χ^2_1 variables, with weights given by the $(n-\varrho^*)^2$ largest eigenvalues of $(\mathbf{D}'_{\varrho^*}\otimes\mathbf{R}'_{\varrho^*})\mathbf{\Omega}(\mathbf{D}_{\varrho^*}\otimes\mathbf{R}_{\varrho^*})$, where $\mathbf{D}_{\varrho^*}$ and $\mathbf{R}_{\varrho^*}$ denote the eigenvectors corresponding to the $n-\varrho^*$ smallest eigenvalues of $\overline{\mathbf{Z}}'\mathbf{\Psi}'\mathbf{\Psi}\overline{\mathbf{Z}}$ and $\mathbf{A}$, respectively. This is summarized in the following proposition:

Proposition 1. *Suppose that Assumption 5 holds true. Then, as $N\to\infty$,*

$$\tau \to^{\mathcal{L}} \sum_{\ell=1}^{(n-\varrho^*)^2} \varpi_\ell \mathcal{Z}^2_\ell, \tag{13}$$

where ϖ_ℓ is the ℓth largest eigenvalue of $(\mathbf{D}'_{\varrho^*}\otimes\mathbf{R}'_{\varrho^*})\mathbf{\Omega}(\mathbf{D}_{\varrho^*}\otimes\mathbf{R}_{\varrho^*})$ and $\{\mathcal{Z}_\ell\}_{\ell=1}^{(n-\varrho^*)(n-\varrho^*)}$ are independent standard normal variates such that $\mathcal{Z}^2_\ell\sim\chi^2_1$ is independent across ℓ.

The proof of Proposition 1 follows from similar arguments as in Robin and Smith (2000), mutatis mutandis. We omit the proof to save space.

Note that although $\mathbf{\Omega}$ is unknown, it can be estimated consistently by[8]:

$$\widehat{\mathbf{\Omega}} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{vec}(\mathbf{\Psi}\mathbf{Z}_i - \mathbf{\Psi}\overline{\mathbf{Z}})\mathrm{vec}(\mathbf{\Psi}\mathbf{Z}_i - \mathbf{\Psi}\overline{\mathbf{Z}})'. \tag{14}$$

As discussed in Robin and Smith (2000), the estimator for ϱ obtained from the test sequence is consistent when the employed significance level α_N vanishes at an appropriate rate with N. This is because α_N is the probability of over-estimating the true rank, $P(\widehat{\varrho}>\varrho)$, which must tend to zero for consistency. The authors show that $\alpha_N = o(1)$ and $-N^{-1}\ln\alpha_N = o(1)$ are sufficient for consistency.

Remark 3.2. Clearly, α_N has to vanish sufficiently fast with N to limit the over-estimation frequency, but not too fast, as this results in under-estimation when N is small. We suggest to specify the nominal level as $\alpha_N = \alpha c N^{-1/\gamma}$. This way, for a given choice of α and γ, the small N significance level is controlled through $c>1$, whereas the speed at which α_N decreases with N is governed by $\gamma>0$. For instance, choosing $\alpha = 5\%$ and setting $c=20$,

[8] $\mathbf{\Omega}$ can also be estimated using bootstrap techniques. When the model contains fixed constants, $\mathbf{Z}_i$ should be time-demeaned, i.e., pre-multiplied with $\mathbf{Q} = \mathbf{I}_T - \boldsymbol{\iota}_T\boldsymbol{\iota}'_T/T$.

$\gamma = 1$ fixes the nominal level to 5% for $N = 20$ and lets it decrease at rate N. Given that over-estimating ϱ may lead to false conclusions that the rank condition holds, we prefer a conservative estimator through a fast decrease with N (i.e., requiring strong evidence against the null before rejecting it in favor of a higher rank estimate).

3.2 Consistent estimation of m

Existing methods to estimate the number of factors from observed data rely on one of the following three approaches: looking at differences or ratios of adjacent eigenvalues (Onatski, 2010; Ahn and Horenstein, 2013), specifying threshold functions to separate bounded from unbounded eigenvalues (Bai and Ng, 2002; Alessi et al., 2010) or sequential tests to determine which eigenvalues are unbounded (Trapani, 2018; Kapetanios, 2010). Preliminary simulation evidence conducted for this paper shows that the Growth Ratio (GR) by Ahn and Horenstein (2013) performs well in finite samples and outperforms other estimators.[9] Therefore, in what follows we propose estimating m using the GR statistic.

In particular, let $\mathbf{Z} = [\mathbf{Z}_1, \ldots, \mathbf{Z}_N]$ denote a $T \times (K+1)N$ matrix, where $\mathbf{Z}_i$ (defined in Eq. (3)) collects all observables for individual i in a $T \times (K+1)$ matrix. Also, let m_{max} denote the maximum value of m considered in estimation, such that $m_{max} \geq m$. We define

$$\widehat{m} = \underset{j \in \{1,\ldots,m_{max}\}}{\arg\max}\, GR(j); \quad GR(j) = \frac{ln(V(j-1)/V(j))}{ln(V(j)/V(j+1))}, \tag{15}$$

where $V(j) = \sum_{k=j+1}^{h} \lambda_j \left(\mathbf{ZZ}'/NT\right)$ with $h = min\{T, N(K+1)\}$, and $\lambda_j \left(\mathbf{ZZ}'/NT\right)$ denotes the jth largest eigenvalue of $(\mathbf{ZZ}'/NT)$.

The GR statistic is easy to compute because it involves maximizing the "growth ratio" of two adjacent eigenvalues arranged in descending order. The main intuition is that the growth ratios of two adjacent eigenvalues of $\mathbf{ZZ}'/NT$ are asymptotically bounded, except for the growth ratio involving the mth and $(m+1)$th eigenvalues, which diverges to infinity.

[9] Juodis and Sarafidis (2022) provide additional evidence that confirms the good performance of the GR statistic in finite samples.

Under regularity conditions implied by Assumptions 1-4, Ahn and Horenstein (2013) show

$$lim_{min\{N,T\}\to\infty} Pr\left(\widehat{m} = m\right) = 1, \tag{16}$$

for any $m_{max} \in \{m, (d^c min\{N,T\}) - m - 1\}$, where $d^c \in (0,1]$.

Remark 3.3. In exactly the same way as described above, the number of factors can also be estimated based on the $T \times T$ matrix $\mathbf{YY}'/NT$, where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$ is of dimension $T \times N$. However, since both $\mathbf{y}_i$ and $\mathbf{X}_i$ share the same factors by assumption, it is natural to combine them together in order to increase the information set used to construct proxies for $\mathbf{F}$. This strategy is in line with the rationale behind the CCE approach, which involves solving a system of equations, such that Eq. (3) includes LHS variables (observables) that are solely driven by a common factor component and purely idiosyncratic noise. Moreover, this strategy is consistent with Westerlund and Urbain (2015), who also estimate factors based on $\mathbf{ZZ}'/NT$.

3.3 A consistent classifier for the rank condition

Given consistent estimates $\widehat{\varrho}$ and $\widehat{m}$ of the rank of $\mathbf{C}$ and the number of factors, the rank condition is deemed to be violated when $\widehat{\varrho} < \widehat{m}$. We define the following classifier:

$$\widehat{RC} \equiv 1 - \mathbb{1}\{\widehat{\varrho} < \widehat{m}\}, \tag{17}$$

where $\mathbb{1}\{\cdot\}$ is an indicator function that returns 1 when the argument inside the curly brackets holds true, and 0 otherwise. Hence, if $\widehat{RC} = 1$ the rank condition is considered to be satisfied, whereas $\widehat{RC} = 0$ indicates that (10) may be violated. The definition in (17) shows that we also take $\widehat{\varrho} > \widehat{m}$ as a sign that (10) is satisfied.[10]

The following proposition summarizes the asymptotic properties of the proposed classifier:

Proposition 2. *Let Assumptions 1-5 hold true. Suppose also that ϱ is determined based on*

[10] $\widehat{\varrho} > \widehat{m}$ can only occur in finite samples due to estimation error but not at the population level.

the sequential testing procedure outlined in Section 3.1, with $\alpha_N = o(1)$*, and* $-N^{-1}\ln\alpha_N = o(1)$*, and* m *is determined by Eq. (15) with* $m_{max} \geq m$*. Then, as* $(N,T) \to \infty$*,*

$$Pr\left[\left(\widehat{RC} = 1|\varrho = m\right) \cup \left(\widehat{RC} = 0|\varrho < m\right)\right] \to 1. \tag{18}$$

That is, the probability that the classifier correctly identifies whether the rank condition is satisfied or not, converges to unity. The result follows directly from the consistency of $\widehat{\varrho}$ as $N \to \infty$ under Assumptions 1-5, given an appropriate rate of decay for α_N, and the consistency of $\widehat{m}$ as $(N,T) \to \infty$ given appropriate specification of m_{max}.

3.4 What if the rank condition is violated?

When $\widehat{RC} = 0$, the standard CCE estimator is generally inconsistent unless the regressors are uncorrelated with the unobserved factor loadings. One may seek to restore the RC by augmenting the model with additional CSA (see Appendix A for several options). This brings about two important issues.

The first one is how to choose relevant additional CSA from a set of candidate expansions, as not all candidates are necessarily informative about $\mathbf{F}$. The second one is whether the selected additional CSA are also able to restore the RC.

To tackle the first question, Karabiyik et al. (2019) have proposed an IC selection procedure. To illustrate, let $\overline{\mathbf{Z}}_+$ be the matrix of available expansions

$$\overline{\mathbf{Z}}_+ = \{\overline{\mathbf{Z}}_+^{(1)}, \overline{\mathbf{Z}}_+^{(2)}, \overline{\mathbf{Z}}_+^{(3)}\}, \tag{19}$$

where (say) $\overline{\mathbf{Z}}_+^{(1)} = \overline{\mathbf{Z}}^{(e)}$ contains CSA of new exogenous variables, $\overline{\mathbf{Z}}_+^{(2)} = \overline{\mathbf{Z}}_{w_1}$ contains a matrix of CSA arising from a new weighting variable w_1, and similarly $\overline{\mathbf{Z}}_+^{(3)} = \overline{\mathbf{Z}}_{w_2}$ for a weight w_2 (see Appendix A for details). The appropriate set of expansion CSA can be selected from $\overline{\mathbf{Z}}_+$ by minimizing

$$\boldsymbol{\ell}^* = \arg\min_{\boldsymbol{\ell}} IC(\boldsymbol{\ell}), \tag{20}$$

where $C_{NT} = \min\{N, \sqrt{T}\}$ in

$$IC(\boldsymbol{\ell}) = \ln|\Sigma_{i=1}^{N} \mathbf{Z}_i' \mathbf{M}_A^{(\boldsymbol{\ell})} \mathbf{Z}_i / NT| + g(n); \qquad g(n) = n(K+1)\frac{\ln(C_{NT})}{C_{NT}}, \tag{21}$$

and $\boldsymbol{\ell} = \{\ell_1, \ell_2, ...\}$ gathers the indices of the considered expansions from $\overline{\mathbf{Z}}_+$ such that for (say) $\boldsymbol{\ell} = \{\ell_1, \ell_3\}$, $\overline{\mathbf{Z}}_A^{(\boldsymbol{\ell})} = [\overline{\mathbf{Z}}, \overline{\mathbf{Z}}_+^{(1)}, \overline{\mathbf{Z}}_+^{(3)}] = [\overline{\mathbf{Z}}, \overline{\mathbf{Z}}^{(e)}, \overline{\mathbf{Z}}_{w_2}]$. The number n denotes the number of columns in $\overline{\mathbf{Z}}_A^{(\boldsymbol{\ell})}$, and $\mathbf{M}_A^{(\boldsymbol{\ell})} = \mathbf{I}_T - \overline{\mathbf{Z}}_A^{(\boldsymbol{\ell})} \left(\overline{\mathbf{Z}}_A^{(\boldsymbol{\ell})\prime} \overline{\mathbf{Z}}_A^{(\boldsymbol{\ell})} \right)^{\dagger} \overline{\mathbf{Z}}_A^{(\boldsymbol{\ell})\prime}$.

A desirable property of the IC selection procedure is that it identifies the CSA that bring in new information about the factors in $\mathbf{Z}_i$ given what is already present in $\overline{\mathbf{Z}}$. Candidates that are uninformative, or informative on factors that do not feature in $\mathbf{Z}_i$, will be excluded (asymptotically). However, as is there is no guarantee that the RC will be fully restored for the chosen CSA. For example, if the IC does not select additional CSA besides $\overline{\mathbf{Z}}$, this could be either because the rank condition is satisfied with $\overline{\mathbf{Z}}$, or because no further informative CSA are available in the proposal set $\overline{\mathbf{Z}}_+$. To overcome this problem, we propose combining the IC with our RC classifier, as outlined in Algorithm 1.

Remark 3.4. An alternative strategy would be to combine our classifier with the regularization approach proposed by Juodis (2021). We leave this possibility for future research.

Algorithm 1: CCE_A algorithm

(1) Estimate the model parameters using the standard CCE approach and calculate $IC_0 = \ln|\Sigma_{i=1}^N \mathbf{Z}_i' \mathbf{M} \mathbf{Z}_i / NT| + g(n)$. Proceed to step 2;

(2) Evaluate the rank condition for $\overline{\mathbf{Z}}$. If $\widehat{RC} = 1$, no further steps are required. If $\widehat{RC} = 0$, proceed to step 3;

(3) Employ the IC in Eq. (21) to select from $\overline{\mathbf{Z}}_+ = \{\overline{\mathbf{Z}}_+^{(1)}, \overline{\mathbf{Z}}_+^{(2)}, \overline{\mathbf{Z}}_+^{(3)}, \dots\}$ the set of CSA that are relevant for the factors in $\mathbf{Z}_i$. That is, define $\boldsymbol{\ell}^* = \arg\min_{\boldsymbol{\ell}} IC(\boldsymbol{\ell})$;

(4) If $IC(\boldsymbol{\ell}^*) \leq IC_0$, evaluate the rank condition for $\overline{\mathbf{Z}}_A = [\overline{\mathbf{Z}}, \overline{\mathbf{Z}}_+^{(\boldsymbol{\ell}^*)}]$ and proceed to step 5, else proceed to step 6;

(5) If $\widehat{RC}(\overline{\mathbf{Z}}_A) = 1$, estimate the model with the CCE_A estimator based on $\overline{\mathbf{Z}}_A$. No further steps are required. If $\widehat{RC}(\overline{\mathbf{Z}}_A) = 0$, proceed to step 6;

(6) $\overline{\mathbf{Z}}_+$ does not contain sufficient informative expansions to restore the rank condition in the model. Add new potential expansions to $\overline{\mathbf{Z}}_+$ and return to step 3;

Remark 3.5. It is also possible to evaluate the RC for all potential augmentations until $\widehat{RC} = 1$. However, this strategy bares the risk of selecting CSA that load on different factors than those in $\mathbf{Z}_i$, and so they are irrelevant for approximating the factor space. This is because such CSA will increase the rank of the augmented loading matrix, despite being irrelevant, and they will therefore be incorrectly favored by the classifier. A preliminary pass-through by the IC selection, as in Algorithm 1, eliminates such irrelevant options.

Remark 3.6. We refer to Appendix D for a discussion and results pertaining to inference with the augmented CCE estimator after the application of Algorithm 1.

4 Monte Carlo Simulation

In this section we investigate the small sample performance of the rank condition classifier proposed in Section 3 using Monte Carlo simulations.

4.1 Design

Data are generated from Eq. (3), broadly following Westerlund and Urbain (2013). We set $m = 2$, $K = 1$, $\beta = 3$ and sample the time series in $\mathbf{F}$, $\boldsymbol{\varepsilon}_i$ and $\mathbf{V}_i$ assuming independent au-

toregressive processes with a common AR coefficient $\rho = 0.8$ and normally distributed mean zero innovations with variance $(1-\rho^2)$ for the factors and $(1-\rho^2)/2$ for the idiosyncratic errors. For the factor loadings $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Gamma}_i$, we specify the following three scenarios:

- Experiment 1: $\boldsymbol{\lambda}_i = [3,2]' + \boldsymbol{\eta}_i$, $\boldsymbol{\eta}_i \sim N(\mathbf{0}_2, \mathbf{I}_2)$, and $\boldsymbol{\Gamma}_i = \boldsymbol{\lambda}_i + [-2,0]'$.
- Experiment 2: $\boldsymbol{\lambda}_i = \begin{cases} [0,2]' + \boldsymbol{\eta}_i & \text{for } i = 1,\ldots,\lfloor N/2 \rfloor \\ [2,0]' + \boldsymbol{\eta}_i & \text{for } i = \lfloor N/2 \rfloor + 1,\ldots,N \end{cases}$ with $\boldsymbol{\eta}_i \sim N(\mathbf{0}_2, \mathbf{I}_2)$ and $\boldsymbol{\Gamma}_i = \boldsymbol{\lambda}_i$.
- Experiment 3: $\boldsymbol{\lambda}_i \sim N(\mathbf{0}_2, \mathbf{I}_2)$ and $\boldsymbol{\Gamma}_i = \boldsymbol{\lambda}_i$.

Thus, in Experiment 1 the RC is satisfied for the simple CSA $\overline{\mathbf{Z}}$ ($\varrho = m = 2$). In Experiment 2, the basic CSA contain some information for estimating the factors ($\varrho = 1$), yet not sufficient to satisfy the RC. Since the loadings in $\mathbf{y}_i$ and $\mathbf{X}_i$ are (perfectly) correlated, the standard CCE estimator is not consistent. In Experiment 3 the standard CSA contain no information at all about the factors ($\varrho = 0 < m$), in which case consistent CCE estimation is also not possible with $\overline{\mathbf{Z}}$.

We evaluate the RC in each MC iteration, using Algorithm 1 of Section 3.4. The number of factors (m) is estimated by the GR statistic of Ahn and Horenstein (2013), setting $m_{max} = 7$. The rank of the loading matrix (ϱ) is estimated as in Section 3.1 with a random dimension reduction $\boldsymbol{\Psi} = T^{-1/2}\boldsymbol{\Phi}$, $\boldsymbol{\Phi}$ containing i.i.d. standard-normal entries, and the nominal significance level given by $\alpha_N = c\alpha N^{-1/\gamma}$, with $c = 20$, $\gamma = 1$ and $\alpha = 5\%$.

Additional CSA are constructed using the following weighting schemes:

$$\overline{\mathbf{Z}}_{w,1} = \sum_{i=1}^{N} \mathbf{Z}_i w_{i,1}; \qquad w_{i,1} = \begin{cases} 1/N_1 & \text{for } i = 1,\ldots,N/2; \\ 0 & \text{for } i = N/2+1,\ldots,N, \end{cases} \tag{22}$$

$$\overline{\mathbf{Z}}_{w,2} = \sum_{i=1}^{N} \mathbf{Z}_i w_{i,2}; \qquad w_{i,2} = \begin{cases} 0 & \text{for } i = 1,\ldots,N/2; \\ 1/(N-N_1) & \text{for } i = N/2+1,\ldots,N, \end{cases} \tag{23}$$

which results in CSA calculated over the first ($\overline{\mathbf{Z}}_{w,1}$) and second ($\overline{\mathbf{Z}}_{w,2}$) group of $N/2$ cross-

sectional units. This choice of weights presumes the existence of an exogenous grouping of the cross-sectional units, as in Experiment 2. It is not an appropriate RC-restoring expansion for experiments 1 and 3, as no such grouping exists for these experiments.
We also consider candidate CSA originating from the $T \times 2$ matrix of external variables

$$\mathbf{Z}_i^{(e)} = \mathbf{F}\mathbf{C}_i^{(e)} + \boldsymbol{\epsilon}_i^{(e)},$$

where the columns of $\boldsymbol{\epsilon}_i^{(e)}$ are generated as AR(1) processes with autoregressive coefficient $\rho = 0.8$ and mean zero normally distributed innovations with variance $(1-\rho^2)/2$, while

$$\mathbf{C}_i^{(e)} = \begin{bmatrix} 2.5 & 1 \\ 1 & 2.5 \end{bmatrix} + \boldsymbol{\eta}_i^{(e)}; \qquad vec(\boldsymbol{\eta}_i^{(e)}) \sim N(\mathbf{0}_4, \mathbf{I}_4).$$

As the $\mathbf{Z}_i^{(e)}$ load on the same factors as those in $\mathbf{Z}_i$, the matrix $\overline{\mathbf{Z}}^{(e)} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{Z}_i^{(e)}$ is an informative, RC-restoring, expansion in experiments 2 and 3. We also accommodate in our simulations the fact that in practice not all external variables will load on the same factors as those in $\mathbf{Z}_i$. These irrelevant candidates are generated from

$$\mathbf{Z}_i^{(g)} = \mathbf{G}\mathbf{C}_i^{(g)} + \boldsymbol{\epsilon}_i^{(g)},$$

where the factors $\mathbf{G}$, loadings $\mathbf{C}_i^{(g)}$ and innovations $\boldsymbol{\epsilon}_i^{(g)}$ follow the same DGP as $\mathbf{F}$, $\mathbf{C}_i^{(e)}$ and $\boldsymbol{\epsilon}_i^{(e)}$ but are independently generated from the latter. As such, $\mathbf{Z}_i^{(g)}$ is informative about $\mathbf{G}$ but not $\mathbf{F}$, and $\overline{\mathbf{Z}}^{(g)}$ is therefore not an appropriate expansion in any of the considered experiments. The total set of candidate expansions that is fed into Algorithm 1 is thus a mixture of both relevant and uninformative candidates, and is given by

$$\overline{\mathbf{Z}}_+ = [\overline{\mathbf{Z}}_{w,1}, \overline{\mathbf{Z}}_{w,2}, \overline{\mathbf{Z}}^{(e)}, \overline{\mathbf{Z}}^{(g)}]. \tag{24}$$

In accordance with Algorithm 1, the augmented estimator CCE_A selects expansions from $\overline{\mathbf{Z}}_+$ using the Information Criterion by Karabiyik et al. (2019) given in Eq. (21). The RC is re-evaluated when expansions are selected.

We generate $2,000$ datasets for each combination of $N = (20, 50, 100, 200, 500, 1,000)$ and $T = (20, 50, 100, 200)$, and calculate the under/over-estimation frequencies for $\widehat{\varrho}$ and $\widehat{m}$, and the classification accuracy of $\widehat{RC}$, i.e. the % of MC draws where the RC is correctly evaluated. When the RC is not satisfied for the standard CCE estimator (experiments 2 and 3), we also consider the CCE_{A} estimator and compute the 'RC satisfied rate' as the % of MC draws where Algorithm 1 selects expansions that restore the rank condition.

4.2 Estimating ϱ and m

Results for the performance of the estimators for ϱ and m are presented in Table 1 in A/B format, with A and B the percentage of MC iterations where ϱ or m are respectively under- and over-estimated. The left panel contains results for estimating the rank ϱ of the loading matrix and reveals that both the over-and under-estimation frequencies tend to zero as $N \to \infty$. This is consistent with the main result of the paper that ϱ can be estimated consistently from $\overline{\mathbf{Z}}$. It is clear however, that the rank estimator is nevertheless somewhat sensitive to the size of the cross-section dimension, which needs to be sufficiently large (i.e., N of at least 50) to achieve an accuracy of 75%. In contrast, the performance of the rank estimator is largely invariant to the size of T, which supports the projection strategy to guarantee computability of the estimator and large N consistency when also $T \to \infty$. Finally, the rank estimator is conservative in the sense that the true rank is more likely to be under-estimated than over-estimated. This is a consequence of our chosen significance level $\alpha_N = 20\alpha N^{-1}$, of which its fast decay in N implies that strong evidence against the null $\varrho = \varrho^*$ is required before it is rejected in favor of a higher rank $\varrho > \varrho^*$. Yet, the observed under-estimation frequency is reasonable and vanishes sufficiently fast with N.

The right panel of Table 1 reports results for estimating the number of factors $m = 2$ with the GR estimator of Ahn and Horenstein (2013). The estimator performs very well despite the high serial dependence in the generated data, in which case many of its competitors in the literature tend to behave more poorly. The finite sample performance of the GR approach appears to be primarily driven by the time series dimension T. Yet, its small-sample performance is more than adequate as the approach displays low error frequencies

Table 1: Under/over-estimation frequency of the estimators for ϱ and m

		$\widehat{\varrho}$				$\widehat{m}$			
	(N,T)	20	50	100	200	20	50	100	200
Experiment 1	20	35/0	26/0	34/0	33/0	14/15	4/0	0/0	0/0
$\varrho = 2, m = 2$	50	23/0	20/0	20/0	19/0	5/7	1/0	0/0	0/0
	100	14/0	20/0	16/0	14/0	4/8	0/0	0/0	0/0
	200	11/0	12/0	11/0	10/0	7/6	0/0	0/0	0/0
	500	9/0	8/0	8/0	8/0	7/5	0/0	0/0	0/0
	1000	5/0	7/0	7/0	5/0	5/6	0/0	0/0	0/0
Experiment 2	20	32/3	33/3	26/4	27/3	10/6	2/0	0/0	0/0
$\varrho = 1, m = 2$	50	19/2	14/2	16/4	17/1	5/3	0/0	0/0	0/0
	100	13/1	4/1	10/0	8/0	8/2	0/0	0/0	0/0
	200	7/1	8/1	8/0	5/0	6/4	0/0	0/0	0/0
	500	5/0	1/0	3/0	3/0	10/1	0/0	0/0	0/0
	1000	2/0	1/0	2/0	3/0	8/1	0/0	0/0	0/0
Experiment 3	20	0/7	0/7	0/8	0/7	14/15	4/0	0/0	0/0
$\varrho = 0, m = 2$	50	0/3	0/3	0/3	0/2	5/7	1/0	0/0	0/0
	100	0/0	0/1	0/1	0/1	4/8	0/0	0/0	0/0
	200	0/2	0/1	0/1	0/1	7/6	0/0	0/0	0/0
	500	0/0	0/0	0/0	0/0	7/5	0/0	0/0	0/0
	1000	0/0	0/0	0/0	0/0	5/6	0/0	0/0	0/0

Notes: (*i*) Based on 2000 MC iterations. (*ii*) Reported in the left panel is the percentage of under/over- estimation of the true rank ϱ by the rank estimator $\widehat{\varrho}$ applied to $\mathbf{\Psi}\overline{\mathbf{Z}}$, with $\mathbf{\Psi} = T^{-1/2}\mathbf{\Phi}$, $\mathbf{\Phi}$ drawn from the standard-normal distribution $\alpha_N = 20\alpha N^{-1}$ and $\alpha = 5\%$. (*iii*) The right panel is the percentage of under/over estimation of the true number of factors $m = 2$ by the GR estimator with $m_{max} = 7$.

even when $T = 20$, and identifies m without error when $T > 50$.

4.3 Evaluating the rank condition

Experiment 1: rank condition satisfied

In Experiment 1, the RC is satisfied for the CCE estimator that uses the standard set of CSA in $\overline{\mathbf{Z}}$. The classification accuracy reported in Table 2 shows that the $\widehat{RC}$ classifier is reasonably accurate in detecting that the rank condition is indeed satisfied. Even for smaller samples, the RC is correctly confirmed for at least 70% of the MC iterations, the only exception being the smallest $N = 20$ setting where the lowest rate is 59%. As the sample size grows, the accuracy improves and we find that it tends to 1 as both $(N,T) \to \infty$, as required. The results also show that the main determinant for finite sample performance is the cross-section dimension N, rather than T. This is as expected from the results in

Table 1, which show that (i) $\widehat{\varrho}$ is more prone to finite sample error than $\widehat{m}$. The latter is practically error-less when $T \geq 50$; and (ii) $\widehat{\varrho}$ converges at a slower rate and only as N grows. Hence, $\widehat{\varrho}$ is the main driver of the finite sample performance of $\widehat{RC}$. Therefore, in line with the properties of the CCE estimator itself, it will mainly be N that needs to be sufficiently large to be able to correctly assess the rank condition in practice. Finally, note that the samples where we incorrectly obtained $\widehat{RC} = 0$ for the CCE estimator, prompted the application of the augmentation strategy outlined in Algorithm 1 of Section 3.4. As shown in Table 8 in the appendix, an expansion was only selected in the smallest samples and in at most 2% of the MC iterations. Hence, the rank evaluation results for the augmented CCE_{A} estimator reported on the right panel of Table 2 are almost identical to those for the CCE estimator.

Table 2: Evaluating the rank condition: Experiment 1

		CCE				CCE_{A}			
	(N,T)	20	50	100	200	20	50	100	200
Classification	20	0.59	0.73	0.66	0.66	0.64	0.73	0.66	0.66
accuracy	50	0.71	0.80	0.79	0.80	0.73	0.80	0.79	0.80
	100	0.80	0.80	0.84	0.86	0.82	0.80	0.84	0.86
	200	0.84	0.88	0.89	0.90	0.84	0.88	0.89	0.90
	500	0.86	0.92	0.91	0.91	0.86	0.92	0.91	0.91
	1000	0.89	0.92	0.93	0.95	0.89	0.92	0.93	0.95

Notes: (*i*) Based on 2000 MC iterations. (*ii*) Reported is the Classification Accuracy (CA), which is the proportion of MC samples in which the classifier $\widehat{RC}$ defined in Eq. (17) correctly identifies whether the RC is satisfied or not. (*iii*) The RC classifier uses the GR estimator of Ahn and Horenstein (2013) with $m_{max} = 7$ to estimate m, and the Robin and Smith (2000) rank estimator with a standard-normal projection matrix and significance level $\alpha_N = 20\alpha N^{-1}$ to estimate ϱ. (*iv*) The left panel evaluates the rank condition for the standard CCE estimator that uses the matrix of CSA $\overline{\mathbf{Z}}$ to control for the unobserved factors. The right panel evaluates the rank condition for the CCE_{A} estimator, which is the outcome of Algorithm 1 presented in section 3.4. That is, if $\widehat{RC} = 1$ for $\overline{\mathbf{Z}}$, then only $\overline{\mathbf{Z}}$ is employed in the estimation. If on the other hand $\overline{\mathbf{Z}}$ yields $\widehat{RC} = 0$, then expansion CSA are selected from $\overline{\mathbf{Z}}_{+}$ using the IC in (21).

Experiment 2: rank condition violated for basic weights

In Experiment 2, the RC is violated when using the standard set of CSA $\overline{\mathbf{Z}}$. As the factor loadings are (perfectly) correlated, the CCE estimator is inconsistent for β in this setting.[11] The left panel of Table 3 shows that the RC-classifier strongly signals that the RC is violated

[11]This can also be seen from the estimation results in Table 11 in Appendix C.

for the CCE estimator. The proportion of samples where the classifier wrongly concludes that the RC holds quickly diminishes as $(N,T) \to \infty$.

When the RC is found to be violated, Algorithm 1 is applied by letting the IC search among the proposal expansions for additional CSA. In this experiment this leads to the selection of at least one of the valid augmentations $(\overline{\mathbf{Z}}_{w,1}, \overline{\mathbf{Z}}_{w,2}, \overline{\mathbf{Z}}^{(e)})$ in the majority of combinations of N and T (see Table 8 in Appendix C). Accordingly, the rank condition was successfully restored in 91% of the MC iterations even when $T = N = 20$. The proportion of samples where the RC is restored is given in the lower panel of Table 3 for the CCE_A estimator, and can be seen to converge to 1 as $(N,T) \to \infty$. Hence, Algorithm 1 leads to a consistent CCE_A estimator as $(N,T) \to \infty$, when provided with appropriate rank-increasing CSA. [12] In addition, note that the algorithm also performs well in finite samples. The cases where the RC is not satisfied for CCE_A are due to the miss-classification as $\widehat{RC} = 1$ in the 'CCE' panel, which vanishes as the sample size grows.

In practice, selecting expansion CSA with the IC does not guarantee that the rank condition is also satisfied, leaving the researcher unsure about the state of the RC. Hence, Algorithm 1 incorporates a re-evaluation with the classifier after expansions have been chosen. The top right panel of Table 3 reveals that this re-evaluation is able to confirm with good accuracy that the rank condition is satisfied in those cases where the right expansions have been selected. The overall classification accuracy is over 70% in the smallest samples and gradually converges to 1 as $(N,T) \to \infty$. The few cases where the RC remains violated are all due to incorrectly concluding $\widehat{RC} = 1$ conclusion for $\overline{\mathbf{Z}}$ in the first step, which does not prompt action by the algorithm.

Experiment 3: rank condition violated

In Experiment 3, the loading matrix $\mathbf{C}$ for the standard CSA is rank zero. Intuitively, the effect of the factors is averaged out in $\overline{\mathbf{Z}}$ such that the CSA are uninformative for estimating the factor space. The top panel of Table 4 reveals that our RC evaluation method is highly accurate in this setting even for very small N. This is due to the large

[12] This is also confirmed by the estimation results for β in Table 11 of Appendix C.

Table 3: Evaluating the rank condition: Experiment 2

		CCE				CCE_A			
	(N,T)	20	50	100	200	20	50	100	200
Classification accuracy	20	0.92	0.94	0.95	0.97	0.73	0.89	0.93	0.95
	50	0.93	0.97	0.96	0.98	0.86	0.93	0.95	0.98
	100	0.92	0.99	1.00	1.00	0.89	0.99	1.00	1.00
	200	0.94	0.98	1.00	0.99	0.90	0.98	1.00	0.99
	500	0.90	0.99	1.00	1.00	0.89	0.99	1.00	1.00
	1000	0.91	1.00	1.00	1.00	0.90	0.99	1.00	1.00
RC satisfied rate	20					0.91	0.94	0.95	0.97
	50					0.93	0.97	0.96	0.98
	100		Always 0			0.92	0.99	1.00	1.00
	200		(by construction)			0.94	0.98	1.00	0.99
	500					0.90	0.99	1.00	1.00
	1000					0.91	1.00	1.00	1.00

See notes to Table 2. The 'RC satisfied rate' is the % of MC samples in which the algorithm behind CCE_A selects CSA augmentations that restore the rank condition.

discrepancy between $m = 2$ and $\varrho = 0$. As we have specified a conservative estimator for ϱ, such an over-estimation almost never occurred (recall the bottom panel of Table 1).

Given the strong signal by the classifier that the RC is violated, Algorithm 1 in the 'CCE_A' panel has led to a search for expansion CSA in nearly all MC samples. We find that the sole rank-restoring expansion $\overline{\mathbf{Z}}^{(e)}$ was selected with high probability, as indicated by the high proportion of samples for which the RC has been restored (see the bottom panel of Table 4). Note, however, that compared to Experiment 2, the classifier appears less capable to confirm that the rank condition is restored when N is very small. Accuracy is only 40% when $N = 20$. Closer analysis reveals that this is caused by a relatively large under-estimation rate (60%) of the true rank in $N = 20$ samples when the correct expansion was selected. A possible cause is that the expanded matrix $\overline{\mathbf{Z}}_A = [\overline{\mathbf{Z}}, \overline{\mathbf{Z}}^{(e)}]$ has a potential rank (number of columns=4) which is twice the true rank (2). This suggests a relatively high level of estimation noise, and a cross-section dimension of $N = 20$ appears too small to estimate the rank accurately in such cases. Yet, the performance of the estimator improves quickly with N, and classification accuracy recovers to 85% or higher for $N = 100$.

As a final experiment, we consider also the empirically relevant scenario where the proposal set $\overline{\mathbf{Z}}_+$ does not contain sufficient informative CSA to restore the rank condition. To that end, we report in the $\text{CCE}_{A,sub}$ panel of Table 4 the outcomes of Algorithm 1 when the set

Table 4: Evaluating the rank condition: Experiment 3

		CCE				CCE_{A}				$\text{CCE}_{\text{A},sub}$			
	(N,T)	20	50	100	200	20	50	100	200	20	50	100	200
Classification accuracy	20	0.98	0.99	0.99	1.00	0.38	0.41	0.38	0.40	0.92	0.96	0.92	0.94
	50	0.99	0.99	0.99	1.00	0.65	0.68	0.67	0.73	0.97	0.97	0.98	0.99
	100	1.00	1.00	1.00	1.00	0.86	0.94	0.96	0.91	0.98	0.99	0.99	1.00
	200	0.99	1.00	1.00	1.00	0.91	0.98	0.98	0.96	0.97	0.99	0.99	0.98
	500	1.00	1.00	1.00	1.00	0.93	0.99	0.99	0.97	0.96	0.99	1.00	1.00
	1000	1.00	1.00	1.00	1.00	0.93	0.99	0.99	0.99	0.98	1.00	0.99	0.99
RC satisfied rate	20					0.92	0.97	0.99	0.99				
	50					0.96	0.99	0.99	1.00				
	100	Always 0				0.96	0.99	1.00	1.00	Always 0			
	200	(by construction)				0.96	1.00	1.00	1.00	(by construction)			
	500					0.98	0.99	1.00	1.00				
	1000					0.98	1.00	1.00	1.00				

See notes to Tables 2 and 3. $\text{CCE}_{\text{A},sub}$ refers to using Algorithm 1, with the set of potential augmentations given by $\overline{\mathbf{Z}}_{+,sub} = \{\overline{\mathbf{Z}}_{w,1}, \overline{\mathbf{Z}}_{w,2}, \overline{\mathbf{Z}}^{(g)}\}$ instead of $\overline{\mathbf{Z}}_{+}$.

of proposal expansions is $\overline{\mathbf{Z}}_{+,sub} = \{\overline{\mathbf{Z}}_{w,1}, \overline{\mathbf{Z}}_{w,2}, \overline{\mathbf{Z}}^{(g)}\}$ instead of $\overline{\mathbf{Z}}_{+}$. Hence, $\overline{\mathbf{Z}}_{+,sub}$ contains insufficient valid expansions to restore the RC, and Algorithm 1 should signal that the RC remains violated even when expansions have been selected from it. It is furthermore important that the IC does not select $\overline{\mathbf{Z}}^{(g)}$, the CSA that load on factors other than those in $\mathbf{Z}_i$, as it would lead to false conclusions that the RC is satisfied by the classifier (see Remark 3.5). This makes the setting particularly challenging. However, the results summarized in the $\text{CCE}_{\text{A},sub}$ panel of Table 4 show that the classifier confirms with high accuracy that the RC fails even after expansions were chosen.

5 Application: the impact of the Dodd-Frank Act on the profitability of U.S. banks

Studies on the profitability of banking institutions are vital for obtaining better understanding of the causes of financial crises, economic recessions and growth. Profits constitute the first line of defense against losses from credit impairment, since retained earnings are an important source of capital. When it comes to large banks, high profitability may also signal excessive market power through stronger brand image or implicit regulatory protection; this is the so-called "too-big-to-fail" (TBTF) hypothesis, which postulates that large financial institutions may be so widely interconnected to the rest of the economy that their

failure would generate a disastrous domino effect for the whole economy. To the extent that governments effectively subsidize downsize risk for financial institutions with TBTF status, large banks face artificially lower costs of capital, and thus reap more profits.

A large number of studies analyse drivers of bank profits (see e.g., Staikouras and Wood, 2004; Iannotta et al., 2007; Goddard et al., 2011; Lee and Hsieh, 2013; Baker and Wurgler, 2015). There is also a fairly substantial literature focusing on the TBTF hypothesis (see e.g., Sironi, 2003; Gropp and Vesala, 2004; Morgan and Stiroh, 2005; Stern and Feldman, 2009; Völz and Wedow, 2011; Hakenes and Schnabel, 2011). The bulk of this literature provides evidence that government bailout guarantees may distort market discipline, inducing excessive risk-taking and morally hazardous behavior (Mattana et al., 2015).

The present illustration contributes to this literature by examining the impact of the well-known "Dodd-Frank Act" (DFA) on profitability in the U.S. banking sector. The DFA is a U.S. federal law enacted in 2010 that has instituted a new failure-resolution regime, which seeks to ensure that losses resulting from bad decisions by managers are absorbed by equity and debt holders, thus potentially reducing moral hazard. Existing empirical evidence on the extent to which the DFA has alleviated the TBTF is relatively sparse and not in agreement. For example, while Baily et al. (2020) conclude on a positive influence of the DFA towards resolving moral hazard, other studies point in the opposite direction (see e.g. Bordo and Duca, 2018). In what follows, we apply the CCE estimator and rank test methodology developed in the present paper to shed some light on this important topic.

5.1 Data and Model Specification

We make use of a panel data set consisting of 450 U.S. banking institutions over the period 2006:Q1–2019:Q4.[13] We analyse the impact of major drivers of bank profitability, with

[13]All data are publicly available and they have been downloaded from the Federal Deposit Insurance Corporation (FDIC) website. See https://www.fdic.gov/.

emphasis on bank size. Thus, we specify the following model:

$$ROA_{it} = \beta_1^{(\ell)} SIZE_{it} + \beta_2^{(\ell)} CAR_{it} + \beta_3^{(\ell)} LIQUIDITY_{it} + \beta_4^{(\ell)} QUALITY_{it} + \beta_5^{(\ell)} RISK_{it} + u_{it};$$
$$u_{it} = \eta_i + \boldsymbol{\lambda}_i' \mathbf{f}_t + \varepsilon_{it},$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$, $\ell = \tau_1 \mathbb{1}\{t < 2011 : Q1\} + \tau_2 \mathbb{1}\{t \geq 2011 : Q1\}$. Essentially, the model is estimated for two sub-periods, namely 2006:Q1–2010:Q4 and 2011:Q1–2019:Q4. The first sub-period belongs to the Basel I-II period, whereas the second corresponds to the DFA and coincides with the introduction of the Basel III internationally.[14]

The variables of the model are defined as follows: ROA_{it} is the return on assets (annualized net income expressed as a percentage of average total assets on a consolidated basis); $SIZE_{it}$ denotes the natural logarithm of bank total assets; CAR_{it} stands for "capital adequacy ratio" (ratio of Tier 1 capital over average total assets minus ineligible intangibles). Higher values of this ratio imply higher levels of capitalisation; $LIQUIDITY_{it}$ is proxied by the loan-to-deposit ratio. Higher values imply a lower level of liquidity; $QUALITY_{it}$ is computed as the total amount of loan loss provisions expressed as a percentage of assets; and $RISK_{it}$ denotes the ratio of non-performing loans to total loans. Higher values of $RISK$ indicate that banks ex-ante took higher lending risk and therefore they have accumulated ex-post more bad loans

The error term u_{it} is composite. In particular, η_i captures bank-specific effects, such as ownership and location. The $m \times 1$ vector $\mathbf{f}_t$ denotes unobserved economy-wide factors that influence bank profits, albeit with heterogeneous intensities $\boldsymbol{\lambda}_i$. Last, ε_{it} is an idiosyncratic error.

The above set of explanatory variables originate from bank accounts (balance sheets and/or profit and loss accounts) and are tied to management decisions. As such, they are viewed as "internal". Bank profitability is also driven by "external" factors that lie beyond the control of management, such as business cycle effects, monetary shocks and financial innovation.

[14]Basel III is an international regulatory framework for capital standards, which incorporates a set of reforms within the banking sector, designed to improve the regulation, supervision and risk management. It requires banks to maintain proper leverage ratios and meet certain minimum capital requirements.

These are absorbed in our model by the common factor component specified in the error term, $\boldsymbol{\lambda}_i'\mathbf{f}_t$. Although in some cases external drivers can be measured and included directly in the model, often the details of measurement may be difficult and/or contentious.[15]

We note that internal and external drivers of bank profitability are likely to be mutually correlated. For example, asset quality may depend on the position of the business cycle, since contractionary phases are typically associated with a higher level of default risk. Therefore, standard panel data approaches that fail to control for external drivers are likely to face an endogeneity problem and, hence, to yield inconsistent parameter estimates. The CCE approach allows for consistent estimation, provided that the rank condition is satisfied such that the external drivers are adequately controlled for.

For notational convenience, let $\mathbf{Z}_i$ denote the $T \times 6$ matrix with the observables

$$\mathbf{Z}_i = \left[\mathbf{y}_i, \mathbf{x}_i^{(1)}, \ldots, \mathbf{x}_i^{(5)}\right], \tag{25}$$

where $\mathbf{y}_i = [y_{i1}, \ldots, y_{iT}]'$ is a $T \times 1$ vector such that $y_{it} \equiv ROA_{it}$, and similarly for the remaining variables, where $\mathbf{x}_i^{(k)}$ denotes the covariate with coefficient $\beta_k^{(\ell)}$.

5.2 Evaluating the RC

Before looking at the CCE estimation results, it is important to test whether the RC is satisfied. The number of factors m is estimated from the $T \times (K+1)N$ matrix $\mathbf{Z} = [\mathbf{Z}_1, \ldots, \mathbf{Z}_N]$, using the Growth Ratio statistic of Ahn and Horenstein (2013). The rank of the matrices of CSA that we consider, to be defined shortly, is determined based on the sequential testing procedure of Robin and Smith (2000). Since T is small in both sub-samples, there is no need to reduce the row-dimensionality using a projection matrix.

We start with the standard CCE estimator based on the unweighted CSA $\overline{\mathbf{Z}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i$. Table 5 reports results for evaluating the RC. The first and second columns correspond

[15] For example, how does one measure monetary shocks? Does one look at interest rates or monetary aggregates? Which monetary aggregates? Similarly, how does one proxy financial innovation? For instance, how does one measure embedded leverage in new financial instruments?

to the standard CCE estimator applied to the periods 2006:Q1–2010:Q4 (Basel I-II) and 2011:Q1–2019:Q4 (Dodd-Frank Act). For the period under Basel I-II, $\widehat{m} = 3$. The standard CSA $\overline{\mathbf{Z}}$ appear unable to proxy these factors as the RC is found to be violated, $\widehat{RC} = 0$. For the period under the Dodd-Frank Act, we obtain $\widehat{m} = 2$ and the rank condition now appears to hold for the standard CCE estimator.

Table 5: US bank profitability: Evaluating the rank condition

	CCE		**CCE$_A$**	
	Basel I-II	**Dodd-Frank Act**	**Basel I-II**	**Dodd-Frank Act**
$\widehat{m}$	3	2	3	2
$\widehat{\varrho}$	1	3	3	2
$\widehat{RC}$	0	1	1	1

Notes: 'Basel I-II' refers to the 2006:Q1–2010:Q4 period, while 'Dodd-Frank Act' to the 2011:Q1–2019:Q4 period. $\widehat{m}$ is the number of factors estimated from the $T \times (K+1)N$ matrix $\mathbf{Z}$, using the GR statistic of Ahn and Horenstein (2013). $\widehat{\varrho}$ is the rank estimator of Robin and Smith (2000), with $\alpha_N = 20\alpha N^{-1}$ and $\mathbf{\Psi} = \mathbf{I}_T$. $\widehat{RC}$ is the rank condition classifier defined in Eq. (17).

Given that the RC is violated for the standard CCE approach in the first sub-period of the sample, we consider a set of potential expansion CSA, given by

$$\overline{\mathbf{Z}}_{+} = \{\overline{\mathbf{Z}}_{+}^{(1)}, \overline{\mathbf{Z}}_{+}^{(2)}, \overline{\mathbf{Z}}_{+}^{(3)}, \overline{\mathbf{Z}}_{+}^{(4)}\}. \tag{26}$$

$\overline{\mathbf{Z}}_{+}^{(1)} \equiv [\overline{\mathbf{x}}^{(6)}, \overline{\mathbf{x}}^{(7)}]$ is a $T \times 2$ matrix, where $\overline{\mathbf{x}}^{(6)}$ and $\overline{\mathbf{x}}^{(7)}$ denote the simple CSA of two external variables, namely the return to equity (ROE), and the tier 1 risk-based capital ratio. ROE is defined as annualized net income expressed as a percent of average total equity on a consolidated basis. The risk-based capital ratio is defined as the tier 1 (core) capital expressed as a percent of risk-weighted assets. As these variables present alternative measures of profitability ($\mathbf{y}_i$) and capitalization ($\mathbf{x}_i^{(2)}$), respectively, they are expected to be driven by the same common factors as those entering into the regression model.

$\overline{\mathbf{Z}}_{+}^{(2)}$ and $\overline{\mathbf{Z}}_{+}^{(3)}$ denote $T \times (K+1)$ matrices of *weighted* CSA, computed from $\mathbf{Z}_i$ in Eq. (25). $\overline{\mathbf{Z}}_{+}^{(2)}$ is calculated using as aggregation weight the initial value of the bank-specific debt ratio (defined as total liabilities over total assets). This variable has been employed in the literature as a measure of interconnectedness of banks (Fernandez, 2011). Thus,

banks with similar levels of debt ratio may be hit by common shocks in an alike manner and therefore they take a similar weight in the computation of the CSA of $\mathbf{Z}_i$. $\overline{\mathbf{Z}}_+^{(3)}$ uses the size of each bank in the beginning of the sample as averaging weight. This implies that banks of similar size get a similar weight in the computation of $\overline{\mathbf{Z}}_+^{(3)}$.

Finally, $\overline{\mathbf{Z}}_+^{(4)}$ is a $T\times 2(K+1)$ matrix of CSA, obtained using two weights that are constructed by grouping banks according to their size. In particular, we take CSA over the bottom and the top quintile of banks.

Table 6 reports IC results for each of the suggested additional CSA. Under Basel I-II, where the RC was found to be violated, the IC selects $\overline{\mathbf{Z}}_+^{(1)}$ as a relevant expansion. The other expansions ($\overline{\mathbf{Z}}_+^{(2)}, \overline{\mathbf{Z}}_+^{(3)}, \overline{\mathbf{Z}}_+^{(4)}$) do not provide new information about the factor space. In the DFA period none of the expansions are selected, since the RC was already satisfied.

Table 6: US bank profitability: IC for additional CSA

	Basel I-II	**Dodd-Frank Act**
	IC	*IC*
$\overline{\mathbf{Z}}$	-4.149	-7.664
$[\overline{\mathbf{Z}}, \overline{\mathbf{Z}}_+^{(1)}]$	-4.260	-5.510
$[\overline{\mathbf{Z}}, \overline{\mathbf{Z}}_+^{(2)}]$	0.588	-0.516
$[\overline{\mathbf{Z}}, \overline{\mathbf{Z}}_+^{(3)}]$	2.045	0.012
$[\overline{\mathbf{Z}}, \overline{\mathbf{Z}}_+^{(4)}]$	1.529	7.044

Note: the IC criterion is specified in Eq. (21).

Given the IC results, we consider the augmented CCE_A estimator with CSA $\overline{\mathbf{Z}}_A = [\overline{\mathbf{Z}}, \overline{\mathbf{Z}}_+^{(1)}]$. Whether this augmented set of CSA is also sufficient to restore the rank condition needs to be verified with the RC classifier. Results are reported in the right panel of Table 5. As we can see, the augmentation has restored the rank condition ($\widehat{RC} = 1$) for the first sub-period. As expected, the RC remains satisfied in the second sub-period should we also augment the CCE estimator with $\overline{\mathbf{Z}}_+^{(1)}$.

5.3 CCE and CCE_A estimation results

Table 7 reports CCE and CCE_A estimates for the two sub-periods 2006:Q1–2010:Q4 and 2011:Q1–2019:Q4. The RC evaluation results imply that in the first sub-period the CCE_A

estimator is consistent, whereas CCE is not. Such discrepancy is mainly noticeable in the estimated coefficients of $SIZE$ and $LIQUIDITY$. In both cases, the inconsistent CCE appears to overestimate the impact of these variables on bank profitability. For the period 2011:Q1–2019:Q4, RC holds for both CCE and CCE_A. Hence, there is no need to augment the model with additional CSA. Notably, the estimated coefficients obtained by the two estimators are not statistically different.

Table 7: US bank profitability: CCE and CCE_A estimation results

	Basel I-II		**Dodd-Frank Act**	
	CCE	CCE_A	CCE	CCE_A
$\widehat{\beta}_1$ (size)	0.959***	0.647***	0.267*	0.331**
	(0.325)	(0.196)	(0.149)	(0.156)
$\widehat{\beta}_2$ (CAR)	-0.035**	-0.038***	-0.027	-0.026
	(0.017)	(0.015)	(0.021)	(0.021)
$\widehat{\beta}_3$ (liquidity)	1.045***	0.646***	0.964***	0.871***
	(0.364)	(0.251)	(0.170)	(0.192)
$\widehat{\beta}_4$ (quality)	-0.943***	-0.914***	-0.890***	-0.905***
	(0.061)	(0.040)	(0.050)	(0.048)
$\widehat{\beta}_5$ (RISK)	0.016	0.017	-0.027***	-0.025**
	(0.012)	(0.011)	(0.009)	(0.010)

Notes: 'Basel I-II' refers to the 2006:Q1–2010:Q4 period, while 'Dodd-Frank Act' to the 2011:Q1–2019:Q4 period. Standard errors, computed based on the parametric sandwich-type formula in Eq. (74) of Pesaran (2006), are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Turning to a comparison of the results across the two sub-periods, $SIZE$ appears to be substantially less important in terms of driving profitability of banks under the DFA period. More specifically, the difference between $\widehat{\beta}_1^{(\tau_1)} = 0.647$ and $\widehat{\beta}_1^{(\tau_2)} = 0.267$ equals 0.38 and is statistically significant at the 10% level of significance, with a p-value that is roughly equal to 0.061 (one-tailed test).[16] That is, if large banks exercised market power and implicitly relied on regulatory protection based on a "too-big-to-fail" presumption, such type of behavior seems to be less prevalent after the introduction of the Dodd-Frank Act. This outcome is consistent with the findings of Gao et al. (2018), Cui et al. (2020) and Zhu et al. (2020) and provides evidence that the regulatory reforms introduced by the DFA have

[16]The t-statistic is $t = (0.647 - 0.267)/\sqrt{0.196^2 + 0.149^2} = 1.54$. Note that since the CCE_A and CCE estimates are based on different samples, it is natural to assume that their covariance equals zero.

succeeded in influencing banks' behavior in a substantial manner. Further, note that if we use the standard CCE estimator in both sub-periods, the difference between $\widehat{\beta}_1^{(\tau_1)}$ and $\widehat{\beta}_1^{(\tau_2)}$ amounts to $0.959 - 0.267 = 0.692$. Hence, the impact of the DFA is estimated to be twice as large as that obtained based on our approach. This further highlights the importance of evaluating the rank condition for CCE-type estimators.

6 Conclusion

It is well known that the so-called Rank Condition is crucial for the statistical properties of the CCE approach developed by Pesaran (2006). However, to date this rank condition could not be verified as it relates to the rank of the unobserved matrix of factor loadings. Therefore, in practice the rank condition is typically assumed to hold true.

In this paper we have outlined a procedure to evaluate whether the rank condition holds in the model of interest given a chosen set of cross-sectional averages. If the rank condition is found to be violated, the procedure can be applied in an augmentation strategy, which combines our proposed classifier with an Information Criterion, to determine the set of CSA that restores the rank condition. Therefore, our approach is generally applicable for checking whether the chosen cross-section averages are sufficient to satisfy the rank condition, or whether additional variables should be explored.

References

Abadir, K. M. and Magnus, J. R. (2005). Matrix Algebra. Cambridge University Press, Cambridge.

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. Econometrica, 81(3):1203–1227.

Al-Sadoon, M. M. (2017). A unifying theory of tests of rank. Journal of Econometrics, 199(1):49–62.

Alessi, L., Barigozzi, M., and Capasso, M. (2010). Improved penalization for determining

the number of factors in approximate factor models. Statistics & Probability Letters, 80(23):1806 – 1813.

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

Baily, M. N., Klein, A., and Schardin, J. (2020). The Impact of the Dodd- Frank Act on Financial Stability and Economic Growth. The russell sage foundation journal of the social sciences, 3:20–47.

Baker, M. and Wurgler, J. (2015). Do Strict Capital Requirements Raise the Cost of Capital? Bank Regulation, Capital Structure, and the Low-Risk Anomaly. American Economic Review, 105(5):315–320.

Bordo, M. D. and Duca, J. V. (2018). The impact of the dodd-frank act on small business. Staff Reports 1806, Federal Reserve Bank of New York.

Camba-Mendez, G. and Kapetanios, G. (2009). Statistical tests and estimators of the rank of a matrix and their applications in econometric modelling. Econometric Reviews, 28(6):581–611.

Chudik, A. and Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. Journal of Econometrics, 188(2):393 – 420.

Cui, G., Sarafidis, V., and Yamagata, T. (2020). Iv estimation of spatial dynamic panels with interactive effects: Large sample theory and an application on bank attitude toward risk. Working Paper Series 11/20, Department of Econometrics and Business Statistics at Monash University.

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

Fan, J. and Liao, Y. (2020). Learning latent factors from diversified projections and its applications to over-estimated and weak factors. Journal of the American Statistical Association.

Fernandez, V. (2011). Spatial linkages in international financial markets. Quantitative Finance, 11:237–245.

Gao, Y., Liao, S., and Wang, X. (2018). Capital markets' assessment of the economic impact of the dodd-frank act on systemically important financial firms. Journal of Banking & Finance, 86:204–223.

Goddard, J., Liu, H., Molyneux, P., and Wilson, J. O. S. (2011). The persistence of bank profit. Journal of Banking & Finance, 35:2881–2890.

Gropp, R. and Vesala, J. (2004). Deposit insurance, moral hazard and market monitoring. Review of Finance, 8:571–602.

Hakenes, H. and Schnabel, I. (2011). Bank size and risk taking under Basel II. Journal of Banking and Finance, 35:1436–1449.

Harding, M., Lamarche, C., and Pesaran, H. (2020). Common correlated effects estimation of heterogeneous dynamic panel quantile regression models. Journal of Applied Econometrics, 35(3):294–314.

Iannotta, G., Nocera, G., and Sironi, A. (2007). Ownership structure, risk and performance in the European banking industry. Journal of Banking & Finance, 31:2127–2149.

Juodis, A. (2021). A regularization approach to common correlated effects estimation. University of Amsterdam.

Juodis, A., Karabiyik, H., and Westerlund, J. (2021). On the Robustness of the Pooled CCE Estimator. Journal of Econometrics, 220(2):325–348.

Juodis, A. and Sarafidis, V. (2021). An incidental parameters free inference approach for panels with common shocks. Journal of Econometrics, forthcoming.

Juodis, A. and Sarafidis, V. (2022). A linear estimator for factor augmented fixed-t panels with endogenous regressors. Journal of Business & Economic Statistics, 40(1):1–15.

Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets. Journal of Business and Economic Statistics, 28(3):397–409.

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

Karabiyik, H., Reese, S., and Westerlund, J. (2017). On the role of the rank condition

in CCE estimation of factor-augmented panel regressions. Journal of Econometrics, 197(1):60 – 64.

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

Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. Journal of Econometrics, 133(1):97–126.

Lee, C. and Hsieh, M. (2013). The impact of bank capital on profitability and risk in Asian banking. Journal of International Money and Finance, 32:251–281.

Mattana, P., Petroni, F., and Rossi, S. P. S. (2015). A test for the too-big-to-fail hypothesis for European banks during the financial crisis. Applied Economics, 47:319–332.

Morgan, D. P. and Stiroh, K. J. (2005). Too big to fail after all these years. Staff Reports 220, Federal Reserve Bank of New York.

Norkute, M., Sarafidis, V., Yamagata, T., and Cui, G. (2020). Instrumental variable estimation of dynamic linear panel data models with defactored regressors and a multifactor error structure. Journal of Econometrics, forthcoming.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. The Review of Economics and Statistics, 92(4):1004–1016.

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

Robin, J.-M. and Smith, R. J. (2000). Tests of rank. Econometric Theory, 16(2):151–175.

Sironi, A. (2003). Testing for market discipline in the European banking industry: evidence from subordinated debt issues. Journal of Money, Credit and Banking, 35:443–472.

Staikouras, C. K. and Wood, G. E. (2004). The Determinants Of European Bank Profitability. International Business & Economics Research Journal, 3(6):57–68.

Stern, G. H. and Feldman, R. J. (2009). Too big to fail: The hazards of bank bailouts. Washington, DC. ISBN: 0-8157-8152-0 220, Brookings Institution Press.

Su, L. and Jin, S. (2012). Sieve estimation of panel data models with cross section dependence. Journal of Econometrics, 169(1):34–47.

Trapani, L. (2018). A randomized sequential procedure to determine the number of factors. Journal of the American Statistical Association, 113(523):1341–1349.

Völz, M. and Wedow, M. (2011). Market discipline and too-big-to-fail in the CDS market: does banks' size reduce market discipline? Journal of Empirical Finance, 18:195–210.

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

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

Zhu, H., Sarafidis, V., and Silvapulle, M. (2020). A new structural break test for panels with common factors. The Econometrics Journal, 23:137–155.

SUPPLEMENTARY MATERIAL

Appendices: Potential CSA for expansions when the RC is not satisfied, Mathematical proofs for the main results and additional simulation results. (.pdf-file)

- Appendix A: Rank condition not satisfied: potential CSA for expansions.
- Appendix B: Mathematical proofs for the main results
- Appendix C: Additional simulation results
- Appendix D: Post-selection inference

Appendices

Appendix A Rank condition not satisfied: potential CSA for expansions

Chudik and Pesaran (2015) advocate expanding $\overline{\mathbf{Z}}$ by adding cross-sectional averages of external variables. This practice requires that these variables load on the same set of factors $\mathbf{F}$ that operate in $\mathbf{Z}_i$, but otherwise have no relation to the dependent variable. To illustrate, consider a setting where $m > K+1$ so that the rank condition is violated for $\overline{\mathbf{Z}}$. Let $\mathbf{Z}_i^{(e)}$ be the $T \times K_e$ matrix gathering the exogenous covariates, given by

$$\mathbf{Z}_i^{(e)} = \mathbf{F}\mathbf{C}_i^{(e)} + \boldsymbol{\epsilon}_i^{(e)}, \tag{A-1}$$

where $\mathbf{C}_i^{(e)}$ denotes an $m \times K_e$ matrix of factor loadings with finite mean $\mathbf{C}^{(e)}$, and $\boldsymbol{\epsilon}_i^{(e)}$ is the $T \times K_e$ matrix of errors. Assuming that the components of this DGP also satisfy Assumptions 1-3 and 4, the augmented matrix of CSA, $\overline{\mathbf{Z}}_A = \left[\overline{\mathbf{Z}}, \overline{\mathbf{Z}}^{(e)}\right]$, may satisfy the rank condition, because it can be written as

$$\overline{\mathbf{Z}}_A = \mathbf{F}\left[\overline{\mathbf{C}}, \overline{\mathbf{C}}^{(e)}\right] + \left[\overline{\mathbf{U}}, \overline{\boldsymbol{\epsilon}}^{(e)}\right] = \mathbf{F}\mathbf{C}_A + O_p(N^{-1/2}), \tag{A-2}$$

where $\mathbf{C}_A = [\mathbf{C}, \mathbf{C}^{(e)}]$. Given that $\mathbf{Z}_i^{(e)}$ loads on the same set of factors $\mathbf{F}$, the augmented loading matrix $\mathbf{C}_A$ is now of order $m \times (1 + K + K_e)$. Therefore, this can restore the RC provided that $m \leq 1 + K + K^{(e)}$ and $\mathbf{C}^{(e)}$ is also sufficiently distinct from $\mathbf{C}$.

An alternative idea is to make use of external variables as additional weights, in order to construct *weighted* CSA. Such an approach has been recently advocated by Juodis and Sarafidis (2022), Fan and Liao (2020), Juodis and Sarafidis (2021) and, in the present context of CCE estimation, by Karabiyik et al. (2019).

To illustrate, let w_i denote an external, time-invariant variable.[17] Multiplying Eq. (3) by

[17]For example, Karabiyik et al. (2019) estimate a gravity equation of bilateral trade flows and construct

w_i and summing over i yields

$$\underset{T\times(K+1)}{\overline{\mathbf{Z}}_w} = \underset{T\times m}{\mathbf{F}} \underset{m\times(K+1)}{\overline{\mathbf{C}}_w} + \underset{T\times(K+1)}{\overline{\mathbf{U}}_w}, \tag{A-3}$$

where $\overline{\mathbf{Z}}_w = \sum_{i=1}^N \mathbf{Z}_i w_i$, $\overline{\mathbf{C}}_w = \sum_{i=1}^N \mathbf{C}_i w_i$, and $\overline{\mathbf{U}}_w = \sum_{i=1}^N \mathbf{U}_i w_i$. As shown by Karabiyik et al. (2019), when $\mathbf{C}_i$ and w_i are correlated, but $\mathbf{U}_i$ and w_i are not, then $\overline{\mathbf{Z}}_w = \mathbf{F}\overline{\mathbf{C}}_w + O_p(N^{-1/2})$ and $\overline{\mathbf{C}}_w$ converges to a nonzero matrix.[18] If $\overline{\mathbf{C}}_w$ is also sufficiently distinct from $\overline{\mathbf{C}}$, the obtained $\overline{\mathbf{Z}}_w$ provides new (i.e. rank increasing) information on $\mathbf{F}$, and the rank of the augmented matrix $\overline{\mathbf{Z}}_A = [\overline{\mathbf{Z}}, \overline{\mathbf{Z}}_w]$ is increased. As the authors point out, w_i effectively acts as an instrument for $\mathbf{C}_i$, and multiple w_i can be combined in an attempt to restore the RC.[19]

Lastly, one can also employ deterministic averaging weights, such as binary indicators that give rise to group-specific cross-sectional averages. For example, in a panel of countries, individual units may be classified as developed, emerging and developing economies; in a panel of firms, units may be grouped according to their size or sector; and so on. In many cases, such group memberships are known and the group-specific averages can be more informative factor proxies than the simple (overall) average.

Appendix B Proofs of theoretical results

B.1 Proof of Theorem 1

Let $\mathbf{M}$ be a given $T \times n$ matrix $(T > n)$ and let $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n]'$ be an $n \times T$ random matrix, where $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n$ are i.i.d.$MN(\mathbf{0}, \mathbf{I}_T)$ in $\mathbb{R}^T$.

weights based on different measures of trade cost.

[18] This property is also utilised by Juodis and Sarafidis (2021), who propose the use of aggregation weights in the context of GMM estimation in panels with T fixed or large.

[19] See Section 2 in Karabiyik et al. (2019) for the formal set of assumptions required to ensure the validity of such weights.

Part (i). We wish to prove that

$$\Pr\left[\text{rank}\left(\mathbf{\Phi M}\right) = \text{rank}\left(\mathbf{M}\right)\right] = 1.$$

Case 1. $\mathbf{M}$ has full column rank, i.e., $\text{rank}(\mathbf{M}) = n$.

Consider the row-matrix representation of a product of two matrices:

$$\mathbf{\Phi M} = \begin{bmatrix} \boldsymbol{\phi}_1'\mathbf{M} \\ \boldsymbol{\phi}_2'\mathbf{M} \\ \vdots \\ \boldsymbol{\phi}_n'\mathbf{M} \end{bmatrix}.$$

It suffices to show that

$$\begin{aligned} \Pr\left[\{\boldsymbol{\phi}_1'\mathbf{M}, \ldots, \boldsymbol{\phi}_n'\mathbf{M}\} \text{ are linearly independent}\right] &= 1 \Leftrightarrow \\ \Pr\left[\{\mathbf{M}'\boldsymbol{\phi}_1, \ldots, \mathbf{M}'\boldsymbol{\phi}_n\} \text{ are linearly independent}\right] &= 1. \end{aligned} \tag{A-4}$$

Let $\mathbf{z}_i = \mathbf{M}'\boldsymbol{\phi}_i$ denote an $n \times 1$ vector, for $i = 1, \ldots, n$. Since $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n$ are i.i.d.$MN\left(\mathbf{0}, \mathbf{I}_T\right)$, it follows that $\mathbf{z}_1, \ldots, \mathbf{z}_n$ are i.i.d.$MN\left(\mathbf{0}, \mathbf{M}'\mathbf{M}\right)$, with $\mathbf{M}'\mathbf{M}$ non-singular because $\mathbf{M}$ has full rank. Let $\widetilde{\mathbf{z}}_i$ denote a specific realization of $\mathbf{z}_i$, where $\widetilde{\mathbf{z}}_i \in \mathbb{R}^n$. Define $\mathbf{y} = \text{vec}\left(\mathbf{z}_1, \ldots, \mathbf{z}_n\right)$ and $\widetilde{\mathbf{y}} = \text{vec}\left(\widetilde{\mathbf{z}}_1, \ldots, \widetilde{\mathbf{z}}_n\right)$, both $n^2 \times 1$ vectors. Let

$$A = \left\{\widetilde{\mathbf{y}} \in \mathbb{R}^{n^2} : \widetilde{\mathbf{z}}_1, \ldots, \widetilde{\mathbf{z}}_n \text{ are linearly dependent}\right\}.$$

We have

$$\begin{aligned} & \Pr\left[\mathbf{z}_1, \ldots, \mathbf{z}_n \text{ are linearly dependent}\right] \\ = \; & E\left(I\left[\mathbf{y} \in A\right]\right) \\ = \; & E\left(E\left(I\left[\mathbf{y} \in A\right] | \mathbf{z}_1, \ldots, \mathbf{z}_{n-1}\right)\right) = 0 \end{aligned} \tag{A-5}$$

because

$$E\left(I\left[\mathbf{y} \in A\right]|\mathbf{z}_1, \ldots, \mathbf{z}_{n-1}\right) = 0.$$

Therefore, we have proved that

$$\Pr\left[\mathbf{z}_1, \ldots, \mathbf{z}_n \text{ are linearly dependent}\right] = 0,$$

and thereby

$$\Pr\left[\text{rank}\left(\mathbf{\Phi M}\right) = \text{rank}\left(\mathbf{M}\right) = n\right] = 1.$$

Case 2. $\mathbf{M}$ has less than full column rank, i.e., $\text{rank}(\mathbf{M}) = n_1 < n$.
Partition $\mathbf{M} = [\mathbf{M}_1 \vdots \mathbf{M}_2]$, where $\mathbf{M}_1$ is $T \times n_1$ and $\mathbf{M}_2$ is $T \times (n - n_1)$, such that $\text{rank}(\mathbf{M}_1) = n_1$. Similarly, partition $\mathbf{\Phi}$ such that

$$\mathbf{\Phi} = \left[\begin{array}{c} \mathbf{\Phi}_1 \\ \hdashline \mathbf{\Phi}_2 \end{array}\right]$$

where $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are $n_1 \times T$ and $(n - n_1) \times T$ respectively with $\text{rank}(\mathbf{\Phi}_1) = n_1$ with probability 1. Thus, the product between $\mathbf{\Phi}$ and $\mathbf{M}$ can be written as

$$\mathbf{\Phi M} = \left[\begin{array}{c:c} \mathbf{\Phi}_1\mathbf{M}_1 & \mathbf{\Phi}_1\mathbf{M}_2 \\ \hdashline \mathbf{\Phi}_2\mathbf{M}_1 & \mathbf{\Phi}_2\mathbf{M}_2 \end{array}\right]_{n \times n}. \tag{A-6}$$

Based on exactly the same arguments as in **Case 1**, it can be shown that $\text{rank}(\mathbf{\Phi}_1\mathbf{M}_1) = n_1$ with probability 1. However, since $\mathbf{\Phi}_1\mathbf{M}_1$ is a submatrix of $\mathbf{\Phi M}$, $\text{rank}(\mathbf{\Phi M}) \geq n_1$. Therefore, we have

$$n_1 = \text{rank}\left(\mathbf{\Phi}_1\mathbf{M}_1\right) \leq \text{rank}\left(\mathbf{\Phi M}\right) \leq \min\left\{\text{rank}\left(\mathbf{\Phi}\right), \text{rank}\left(\mathbf{M}\right)\right\} = n_1.$$

Hence, $\text{rank}(\mathbf{\Phi M}) = n_1 = \text{rank}(\mathbf{M})$ with probability 1. This completes part (i) of the theorem.

Part (ii). We can write by simple addition and subtraction

$$\mathbf{\Psi}\overline{\mathbf{Z}} = T^{-1/2}\mathbf{\Phi}\overline{\mathbf{Z}} = T^{-1/2}(\mathbf{\Phi}\mathbf{F}\overline{\mathbf{C}} + \mathbf{\Phi}\overline{\mathbf{U}}) = T^{-1/2}\mathbf{\Phi}\mathbf{F}\mathbf{C} + T^{-1/2}\mathbf{\Phi}\mathbf{F}(\overline{\mathbf{C}} - \mathbf{C}) + T^{-1/2}\mathbf{\Phi}\overline{\mathbf{U}}$$

Recall that $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n]'$, with its rows given by $\boldsymbol{\phi}_k \sim iidMN(\mathbf{0}_{T\times 1}, \mathbf{I}_T)$ for $k = 1, \ldots, n$. Consider then the kth row of $T^{-1/2}\mathbf{\Phi}\overline{\mathbf{U}}$, and write it as $T^{-1/2}\sum_{t=1}^{T}\phi_{kt}\overline{\mathbf{u}}_t'$, with $\boldsymbol{\phi}_k = [\phi_{k1}, \ldots, \phi_{kT}]'$ and $\overline{\mathbf{U}} = [\overline{\mathbf{u}}_1, \ldots, \overline{\mathbf{u}}_T]'$. By the independence of $\mathbf{\Phi}$ and $\overline{\mathbf{U}}$ we have for every $k = 1, \ldots, n$ that $E(\sum_{t=1}^{T}\phi_{kt}\overline{\mathbf{u}}_t') = \mathbf{0}_{1\times n}$ and

$$\text{Var}\left(\frac{\sum_{t=1}^{T}\phi_{kt}\overline{\mathbf{u}}_t'}{\sqrt{T}}\right) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E(\phi_{kt}\phi_{ks})\,E(\overline{\mathbf{u}}_t\overline{\mathbf{u}}_s') = \frac{1}{T}\sum_{t=1}^{T}E(\overline{\mathbf{u}}_t\overline{\mathbf{u}}_t') = O(N^{-1}),$$

because $E(\phi_{kt}\phi_{ks}) = 0$ for $s \neq t$, $E(\phi_{kt}\phi_{kt}) = E(\phi_{kt}^2) = 1$ and $E(\|\overline{\mathbf{u}}_t\|^2) = O(N^{-1})$ by A.4 of Lemma 1 in Pesaran (2006) under Assumptions 1 and 4. Hence, $\left\|T^{-1/2}\mathbf{\Phi}\overline{\mathbf{U}}\right\| = O_p(N^{-1/2})$ as $(N, T) \to \infty$.

Consider next $T^{-1/2}\mathbf{\Phi}\mathbf{F}$. By the independence of $\mathbf{\Phi}$ and $\mathbf{F}$ we have $E(\mathbf{\Phi}\mathbf{F}) = \mathbf{0}_{n\times m}$, and since $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_T]'$ we can write the k-th row of $\mathbf{\Phi}\mathbf{F}$ as $\sum_{t=1}^{T}\phi_{kt}\mathbf{f}_t'$. Therefore,

$$\text{Var}\left(\frac{\sum_{t=1}^{T}\phi_{kt}\mathbf{f}_t'}{\sqrt{T}}\right) = \frac{1}{T}\sum_{t=1}^{T}E(\phi_{kt}\phi_{kt})\,E(\mathbf{f}_t\mathbf{f}_t') = \frac{1}{T}\sum_{t=1}^{T}E(\mathbf{f}_t\mathbf{f}_t') = O(1),$$

because $E(\mathbf{f}_t\mathbf{f}_t') = O(1)$ for every t (Assumption 2). Hence, we have $\left\|T^{-1/2}\mathbf{\Phi}\mathbf{F}\right\| = O_p(1)$. Noting then that $\left\|\overline{\mathbf{C}} - \mathbf{C}\right\| = O_p(N^{-1/2})$ under Assumption 3, it follows that

$$\left\|T^{-1/2}\mathbf{\Phi}\mathbf{F}(\overline{\mathbf{C}} - \mathbf{C})\right\| \leq \left\|T^{-1/2}\mathbf{\Phi}\mathbf{F}\right\|\left\|\overline{\mathbf{C}} - \mathbf{C}\right\| = O_p(N^{-1/2}).$$

Thus, combining the results above yields as $(N, T) \to \infty$,

$$\mathbf{\Psi}\overline{\mathbf{Z}} = T^{-1/2}\mathbf{\Phi}\mathbf{F}\mathbf{C} + T^{-1/2}\mathbf{\Phi}\mathbf{F}(\overline{\mathbf{C}} - \mathbf{C}) + T^{-1/2}\mathbf{\Phi}\overline{\mathbf{U}} = T^{-1/2}\mathbf{\Phi}\mathbf{F}\mathbf{C} + O_p(N^{-1/2}),$$

where also $\left\|T^{-1/2}\mathbf{\Phi}\mathbf{F}\mathbf{C}\right\| \leq \left\|T^{-1/2}\mathbf{\Phi}\mathbf{F}\right\|\|\mathbf{C}\| = O_p(1)$ since $\|\mathbf{C}\| < \infty$ under Assumption 3. Hence, the proof of part (ii) of the theorem is complete. □

Appendix C Additional simulation results

Table 8: Algorithm 1: Selection percentages for expansion CSA

		$\overline{\mathbf{Z}}_{w,1}$				$\overline{\mathbf{Z}}_{w,2}$				$\overline{\mathbf{Z}}^{(e)}$				$\overline{\mathbf{Z}}^{(g)}$			
	(N,T)	20	50	100	200	20	50	100	200	20	50	100	200	20	50	100	200
Experiment 1	20	2	0	0	0	1	0	0	0	2	0	0	0	1	0	0	0
	50	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
	100	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
	200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Experiment 2	20	37	42	46	51	33	43	48	47	23	10	1	0	0	0	0	0
	50	34	39	43	47	28	40	50	52	32	19	4	0	0	0	0	0
	100	32	32	41	50	27	30	47	49	35	38	11	1	1	0	0	0
	200	27	32	35	46	26	30	33	47	42	37	32	7	1	0	0	0
	500	22	25	34	39	27	29	32	39	42	46	34	22	0	0	0	0
	1000	21	25	29	27	22	26	25	35	50	49	46	38	1	0	0	0
Experiment 3	20	4	1	0	0	3	0	0	0	93	98	100	100	1	0	0	0
	50	3	0	0	0	2	0	0	0	96	99	100	100	2	0	0	0
	100	1	0	0	0	2	0	0	0	96	100	100	100	1	0	0	0
	200	1	0	0	0	1	0	0	0	96	100	100	100	2	0	0	0
	500	0	0	0	0	0	0	0	0	98	99	100	100	1	0	0	0
	1000	1	0	0	0	1	0	0	0	98	100	100	100	1	0	0	0
Experiment 3	20	26	20	19	27	23	18	24	18	0	0	0	0	2	0	0	0
$\overline{\mathbf{Z}}_{+,sub} = \overline{\mathbf{Z}}_{+} \backslash \overline{\mathbf{Z}}^{(e)}$	50	22	13	20	21	17	17	13	18	0	0	0	0	2	0	0	0
	100	22	11	12	15	16	16	15	19	0	0	0	0	2	0	0	0
	200	24	13	15	16	18	14	12	19	0	0	0	0	4	0	0	0
	500	22	11	12	16	18	11	14	19	0	0	0	0	3	1	0	0
	1000	19	14	14	18	19	14	17	16	0	0	0	0	2	0	0	0

Notes: Reported are percentages out of 2000 Monte Carlo iterations that the CSA stated in the column has been selected as an expansion by the IC given in Eq. (21). Since multiple expansions can be selected on each sample size, the percentages do not necessarily sum to 100. The bottom panel displays selection frequencies in Experiment 3 when $\overline{\mathbf{Z}}^{(e)}$ is not a selectable option. That is, the set of proposal expansions is $\overline{\mathbf{Z}}_{+,sub} = \{\overline{\mathbf{Z}}_{w,1}, \overline{\mathbf{Z}}_{w,2}, \overline{\mathbf{Z}}^{(g)}\}$.

Table 9: Algorithm 1: Sensitivity and Specificity

		RC satisfied rate				Sensitivity				Specificity			
	(N,T)	20	50	100	200	20	50	100	200	20	50	100	200
Experiment 1	20	1.00	1.00	1.00	1.00	0.65	0.74	0.67	0.67	1.00	1.00	1.00	1.00
	50	1.00	1.00	1.00	1.00	0.73	0.81	0.80	0.81	1.00	1.00	1.00	1.00
	100	1.00	1.00	1.00	1.00	0.83	0.80	0.84	0.86	1.00	1.00	1.00	1.00
	200	1.00	1.00	1.00	1.00	0.85	0.88	0.89	0.91	1.00	1.00	1.00	1.00
	500	1.00	1.00	1.00	1.00	0.87	0.92	0.92	0.92	1.00	1.00	1.00	1.00
	1000	1.00	1.00	1.00	1.00	0.89	0.93	0.93	0.95	1.00	1.00	1.00	1.00
Experiment 2	20	0.92	0.95	0.96	0.97	0.80	0.94	0.98	0.98	0.00	0.00	0.00	0.00
	50	0.94	0.98	0.96	0.99	0.92	0.96	0.99	1.00	0.00	0.00	0.00	0.00
	100	0.93	0.99	1.00	1.00	0.96	1.00	1.00	1.00	0.00	0.00	1.00	1.00
	200	0.94	0.99	1.00	1.00	0.96	1.00	1.00	1.00	0.00	0.00	1.00	0.00
	500	0.91	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.00	0.00	1.00	1.00
	1000	0.92	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.00	1.00	1.00	1.00
Experiment 3	20	0.92	0.97	0.99	0.99	0.36	0.39	0.38	0.40	0.63	0.95	0.22	0.00
	50	0.96	0.99	0.99	1.00	0.64	0.68	0.67	0.73	0.94	0.52	0.00	1.00
	100	0.96	0.99	1.00	1.00	0.85	0.93	0.96	0.91	1.00	1.00	1.00	1.00
	200	0.96	1.00	1.00	1.00	0.91	0.98	0.98	0.96	0.89	1.00	1.00	1.00
	500	0.98	0.99	1.00	1.00	0.93	0.99	0.99	0.97	1.00	1.00	1.00	1.00
	1000	0.98	1.00	1.00	1.00	0.93	0.99	0.99	0.99	1.00	1.00	1.00	1.00
Experiment 3	20	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.92	0.96	0.92	0.94
$\overline{\mathbf{Z}}_{+,sub} = \overline{\mathbf{Z}}_{+} \backslash \overline{\mathbf{Z}}^{(e)}$	50	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.97	0.97	0.98	0.99
	100	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.98	0.99	0.99	1.00
	200	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.97	0.99	0.99	0.98
	500	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.96	0.99	1.00	1.00
	1000	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.98	1.00	0.99	0.99

Notes: (*i*) Reported in the left panel is the fraction of MC samples where the rank condition ($\varrho = m$) is satisfied (restored) after application of Algorithm 1. The middle panel displays the 'Sensitivity' of the RC classifier, or the rate of correctly obtaining $\widehat{RC} = 1$ for the cases where the RC is satisfied/restored ($\frac{\#\text{true RC=1 conclusions}}{\#\text{true RC=1 conclusions}+\#\text{false RC=0 conclusions}}$), and the rightmost panel gives the 'Specificity', or the rate of correctly obtaining $\widehat{RC} = 0$ when the RC is indeed violated/not restored ($\frac{\#\text{true RC=0 conclusions}}{\#\text{true RC=0 conclusions}+\#\text{false RC=1 conclusions}}$). Note that when there are no $RC = 0$ cases and also no $\widehat{RC} = 0$ conclusions, then Specificity = 1, and similarly for the Sensitivity. The inverse of Sensitivity and Specificity give respectively the false positive (false RC holds conclusions) and false negative rates (false RC violated conclusions). (*ii*) The RC classifier employs the GR estimator with $m_{max} = 7$ to estimate m, and the rank estimator employs the random projection with $\alpha_N = 20\alpha N^{-1}$. (*iii*) The bottom panel gives outcomes for Algorithm 1 when the rank-restoring expansion CSA $\overline{\mathbf{Z}}^{(e)}$ is not among the set of proposal expansions such that it is impossible to restore the RC. (*iv*) Note that the classifier Sensitivity/Specificity are not separately reported when evaluating the RC based on $\overline{\mathbf{Z}}$, because they are identical to the 'Classification Accuracy' reported in the main text. That is, when the RC is satisfied for $\overline{\mathbf{Z}}$ (experiment 1), then Specificity=1 and Sensitivity equals the classification accuracy reported in table 2. Conversely, when RC is violated for $\overline{\mathbf{Z}}$, then Sensitivity=1 and Specificity is the reported accuracy.

C.1 Estimation results for $\beta = 3$

Table 10: Estimation results for $\beta = 3$ in Experiment 1

		bias				*rmse*			
	(N,T)	20	50	100	200	20	50	100	200
CCE	20	0.053	0.051	0.051	0.047	0.120	0.089	0.075	0.059
	50	0.021	0.020	0.021	0.023	0.065	0.048	0.040	0.032
	100	0.011	0.011	0.011	0.013	0.050	0.032	0.024	0.020
	200	0.005	0.004	0.007	0.006	0.034	0.021	0.017	0.013
	500	0.002	0.001	0.002	0.003	0.020	0.013	0.010	0.007
	1000	0.002	0.002	0.001	0.001	0.015	0.009	0.007	0.005
CCE_A	20	0.052	0.051	0.051	0.047	0.118	0.089	0.075	0.059
	50	0.021	0.020	0.021	0.023	0.065	0.048	0.040	0.032
	100	0.011	0.011	0.011	0.013	0.051	0.032	0.024	0.020
	200	0.005	0.004	0.007	0.006	0.034	0.021	0.017	0.013
	500	0.002	0.001	0.002	0.003	0.020	0.013	0.010	0.007
	1000	0.002	0.002	0.001	0.001	0.015	0.009	0.007	0.005

Note: Reported are estimation bias for β and root mean square error (rmse).

Table 11: Estimation results for $\beta = 3$ in Experiment 2

		bias				*rmse*			
	(N,T)	20	50	100	200	20	50	100	200
CCE	20	0.819	0.839	0.839	0.844	0.826	0.842	0.841	0.846
	50	0.815	0.839	0.849	0.851	0.820	0.841	0.850	0.852
	100	0.825	0.835	0.848	0.853	0.830	0.837	0.849	0.854
	200	0.825	0.845	0.850	0.854	0.829	0.847	0.851	0.854
	500	0.824	0.832	0.849	0.852	0.828	0.834	0.850	0.852
	1000	0.823	0.841	0.848	0.852	0.826	0.843	0.848	0.852
CCE_A	20	0.090	0.068	0.062	0.043	0.246	0.201	0.186	0.138
	50	0.063	0.029	0.040	0.022	0.219	0.137	0.162	0.101
	100	0.066	0.016	0.006	0.006	0.237	0.080	0.022	0.016
	200	0.057	0.016	0.005	0.006	0.211	0.104	0.017	0.058
	500	0.082	0.003	0.002	0.002	0.263	0.049	0.010	0.006
	1000	0.069	0.001	0.001	0.001	0.237	0.009	0.007	0.005

Note: Reported are estimation bias for β and root mean square error (rmse).

Table 12: Estimation results for $\beta = 3$ in Experiment 3

		bias				*rmse*			
	(N,T)	20	50	100	200	20	50	100	200
CCE	20	0.655	0.677	0.680	0.682	0.672	0.685	0.687	0.687
	50	0.670	0.689	0.694	0.698	0.679	0.695	0.698	0.701
	100	0.669	0.689	0.698	0.707	0.680	0.695	0.701	0.708
	200	0.678	0.694	0.704	0.705	0.687	0.700	0.707	0.707
	500	0.681	0.682	0.701	0.711	0.691	0.687	0.705	0.713
	1000	0.671	0.690	0.700	0.705	0.680	0.694	0.702	0.707
CCE_{A}	20	0.051	0.033	0.029	0.020	0.149	0.083	0.060	0.041
	50	0.026	0.010	0.012	0.010	0.096	0.059	0.050	0.024
	100	0.015	0.009	0.006	0.007	0.069	0.040	0.022	0.017
	200	0.016	0.002	0.004	0.003	0.079	0.021	0.016	0.012
	500	0.006	0.002	0.001	0.002	0.056	0.035	0.010	0.006
	1000	0.008	0.001	0.000	0.000	0.055	0.009	0.007	0.005
$\text{CCE}_{\text{A},sub}$	20	0.504	0.544	0.535	0.535	0.539	0.574	0.566	0.566
	50	0.556	0.591	0.590	0.581	0.580	0.614	0.613	0.604
	100	0.550	0.592	0.604	0.597	0.577	0.616	0.627	0.622
	200	0.548	0.602	0.614	0.600	0.575	0.625	0.635	0.622
	500	0.551	0.604	0.614	0.600	0.578	0.625	0.635	0.625
	1000	0.549	0.592	0.594	0.591	0.577	0.616	0.619	0.617

Note: Reported are estimation bias for β and root mean square error (rmse). $\text{CCE}_{\text{A},sub}$ denotes the outcome of Algorithm 1 with the set of potential augmentations given by $\overline{\mathbf{Z}}_{+,sub} = \{\overline{\mathbf{Z}}_{w,1}, \overline{\mathbf{Z}}_{w,2}, \overline{\mathbf{Z}}^{(g)}\}$ in stead of $\overline{\mathbf{Z}}_{+}$.

Appendix D Post-selection Inference

One remaining question is how to perform inference with the CCE_A estimator after application of Algorithm 1. To address this, let $\widehat{\boldsymbol{\beta}}_A$ be the estimated parameter vector that follows from Algorithm 1, calculated with the (potentially) augmented set of cross-section averages $\overline{\mathbf{Z}}_A$. We shall assume that $\overline{\mathbf{Z}}_+$ contains sufficient valid augmentations to restore the rank condition through $\overline{\mathbf{Z}}_A$ in case the RC is violated with $\overline{\mathbf{Z}}$. The distribution of $\widehat{\boldsymbol{\beta}}_A$ is then given by

$$\begin{aligned} P(\sqrt{NT}(\widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}) \leq \delta) = P(\sqrt{NT}(\widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}) \leq \delta | \varrho(\overline{\mathbf{Z}}_A) = m) \\ + P(\sqrt{NT}(\widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}) \leq \delta | \varrho(\overline{\mathbf{Z}}_A) \neq m)(1 - P(\varrho(\overline{\mathbf{Z}}_A) = m)) \end{aligned}$$

where $\varrho(\overline{\mathbf{Z}}_A)$ denotes the rank of the loading matrix implied by $\overline{\mathbf{Z}}_A$, such that $P(\varrho(\overline{\mathbf{Z}}_A) = m)$ is the probability that $\widehat{RC} = 0$ when the RC was not satisfied (the augmentation sequence is kicked into gear) and an RC restoring set of augmentations was selected from $\overline{\mathbf{Z}}_+$ by the IC criterion. Since the classifier correctly evaluates the rank condition as $(N, T) \to \infty$ by Proposition 2, and hence leads to the IC selection step in stage (3) and beyond of Algorithm 1 if $\varrho(\overline{\mathbf{Z}}) \neq m$, we have by the consistency of (21) for selecting the correct rank-restoring averages, established in Karabiyik et al. (2019), that $P(\varrho(\overline{\mathbf{Z}}_A) = m) \to 1$ as $(N, T) \to \infty$. Hence, asymptotically

$$P(\sqrt{NT}(\widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}) \leq \delta) = P(\sqrt{NT}(\widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}) \leq \delta | \varrho(\overline{\mathbf{Z}}_A) = m),$$

such that the distribution of the augmented estimator asymptotically equals that of the CCE estimator when the rank condition is satisfied. As is well known from the CCE literature, this asymptotic distribution is independent of the specific choice of CSA provided that $T/N \to 0$ (see e.g. Theorem 1 in Karabiyik et al. (2019)). Hence, the distribution of $\sqrt{NT}(\widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta})$ is asymptotically unaffected by pre-testing and augmentations, and inference can proceed as for the original CCE approach, with rank condition satisfied, provided $T/N \to 0$.