A regularization approach to common correlated effects estimation

Juodis, A.

DOI
10.1002/jae.2899

Publication date
2022

Document Version
Final published version

Published in
Journal of Applied Econometrics

Received: 22 February 2021 | Revised: 22 October 2021 | Accepted: 24 October 2021

DOI: 10.1002/jae.2899

WILEY Journal of APPLIED ECONOMETRICS

RESEARCH ARTICLE

A regularization approach to common correlated effects estimation

Artūras Juodis

Amsterdam School of Economics, University of Amsterdam and Tinbergen Institute, Amsterdam, The Netherlands

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

Funding information
Netherlands Organisation for Scientific Research (NWO), Grant/Award Number: 451-17-002

Summary

Cross-section average-augmented panel regressions introduced by Pesaran (2006) have been a popular empirical tool to estimate panel data models with common factors. However, the corresponding common correlated effects (CCEs) estimator can be sensitive to the number of cross-section averages used and/or the static factor representation for observables. In this paper, we show that most of the corresponding problems documented in the literature can be solved once cross-section averages are appropriately regularized, thus extending the applicability of the CCE setup. As the standard plug-in variance estimators are not able to account for all sources of estimation uncertainty, we suggest the use of cross-section bootstrap to construct confidence intervals. The proposed procedure is illustrated both using real and simulated data.

KEYWORDS

common correlated effects, factor models, incidental parameters problem, regularization

1 | INTRODUCTION

Standard panel data models typically include additive unit- and time-specific fixed effects to account for unobserved characteristics. Over the past five decades such an additive error component structure has been a dominant empirical strategy in panel data studies. However, while additive models might be justified in some applications, some economic models predict that common shocks should enter the model multiplicatively instead; see, for example, Juodis and Kučinskas (2019) and Cesa-Bianchi et al. (2020).

One of the most popular estimation approaches to factor-augmented regression models is the common correlated effects (CCE) approach of Pesaran (2006), which uses the cross-section averages of observed variables as proxies/estimates for unobserved factors. The main reason for the popularity of this approach is its simplicity (as the estimator has a closed form solution), extendibility of the approach to non-linear and non-stationary models (e.g., Boneva & Linton, 2017; Kapetanios et al., 2011, respectively)[1] and good documented Monte Carlo performance; see, for example, Pesaran (2006) and Westerlund and Urbain (2015).

The performance of the CCE estimator against the quasi maximum likelihood (QML) principal components (PCs) estimator of Bai (2009) has been mostly documented using stylized setups where the regressors have a factor structure, and all factors can be estimated by cross-section averages of some observables; see, for example, Chudik et al. (2011),

[1]Cross-section average-augmented models were also used in the construction of the unit-root tests; see, for example, Pesaran (2007), Reese and Westerlund (2016), Juodis and Westerlund (2019), and Norkutė and Westerlund (2021).

Chudik and Pesaran (2015), and Westerlund and Urbain (2015). Alternatively, if some of the factors in the equation of interest cannot be estimated, then additional restrictions on the correlation structure of factor loadings need to be imposed; see, for exmaple, Westerlund and Urbain (2013). However, irrespective of the setup, the crucial idea in Pesaran (2006) is that adding additional observables to estimate common factors bears no costs, at least asymptotically.

Only recently it has been recognized by Karabıyık et al. (2017) that inclusion of too many factor proxies has a non-trivial bias effect even asymptotically. Their result was derived under the assumption that cross-section averages can consistently estimate all strong factors present in the regressors. The recent work of Juodis et al. (2021) deviates from this assumption and considers a setup where only a subset of factors are estimable by cross-section averages. They show that deviations of this type are substantial enough to reduce the convergence rate, and change the asymptotic distribution of the CCE estimator. As a result, including more cross-section averages can actually be harmful in very general setups. In this paper, we show how this issue can be almost completely eliminated, thus naturally extending the applicability of the CCE approach beyond the setup of Pesaran (2006).

Motivation. This research is motivated by two empirical problems associated with the cross-section average-augmented procedures:

1. First of all, the well documented unsatisfactory statistical properties of the pooled CCE estimator with more observables than factors (e.g., in Juodis et al., 2021; Karabıyık et al., 2017) call for new methods to be considered to address the underlying shortcomings of the CCE estimator in the linear model.
2. Second, there has been a growing interest in the application of cross-section average-augmented models in non-linear panel data models, for example, binary choice (Boneva & Linton, 2017), quantile (Harding & Lamarche, 2014; Harding et al., 2020), count (Desbordes & Eberhardt, 2019), and non-linear mean (Hacioglu Hoke & Kapetanios, 2021) models. Some of these setups implicitly assume that there are as many unobserved factors as cross-section averages. Otherwise, some of the regularity conditions cannot be justified.[2] Thus, any technical difficulties associated with the empirically relevant setup with more factor proxies than the underlying factors are ignored.

This paper. We introduce the notion of regularized cross-section averages, and the related regularized CCE estimator. As a basis of our procedure we use the Singular Value Decomposition (SVD) to remove the asymptotically redundant singular values of appropriately normalized cross-section averages. This regularization ensures tractability of the asymptotic distribution for the resulting class of least squares estimators, even allowing for some unproxied factors in regressors. In addition, we argue that despite the normality of the pooled or mean-group estimators, the standard plug-in estimators for the variance–covariance matrices are not consistent. We recommend cross-section (pairs) bootstrap based inference procedure. To select the number of factors we use eigenvalues based selection criterion, as in Ahn and Horenstein (2013). The resulting regularized CCE (rCCE) estimator extends the applicability of the CCE procedure to a more general class of linear models than originally suggested by Pesaran (2006). Our results are established both under the fixed T and the large T asymptotic approximations.

In our empirical illustration, we re-evaluate the results from the recent studies in Voigtländer (2014) and Yin et al. (2021), and investigate the causes of the historically increasing wage inequality between high-skilled and low-skilled workers in the US manufacturing industries. Our procedure provides strong evidence that, irrespective of the setup considered, the number of the underlying factors is small in comparison with the total number of cross-section averages. This fact confirms the necessity of the "more observables than factors" setup in Pesaran (2006) and the need for regularization even for a setup with as many as three cross-sectional averages.

The remainder of this paper is organized as follows. Section 2 introduces a linear panel model with common factors. Section 3 develops the regularization approach and heuristically discusses the main results of this paper for models with homogeneous and heterogeneous coefficients. All formal asymptotic results are discussed in Section 4. Section 5 reports a Monte Carlo study to assess the finite sample performance of the proposed class of estimators. Section 6 presents an empirical application. Finally, Section 7 concludes. Additional (more technical) discussions and proofs are relegated to the supporting information.

Notation. The generic constant δ is used to denote a small positive real number. For a generic matrix $\boldsymbol{A}$, vec($\boldsymbol{A}$) denotes the vertical column stacking operator, $\otimes$ denotes the Kronecker product, $tr(\boldsymbol{A})$ denotes the trace operator, and $\|\boldsymbol{A}\| =$

[2]For example, the standard assumption that the underlying parameter space (including individual-level heterogeneity) is compact (e.g., Assumption C1 in Boneva & Linton, 2017 cannot be satisfied for the parameter space of the rotated factor loadings, if there are more factors than observables).

$\sqrt{tr(\boldsymbol{A}'\boldsymbol{A})}$ the Frobenius norm. For any $T \times L$ matrix $\boldsymbol{A}$ with full column rank its orthogonal projection matrix $\boldsymbol{M_A}$ is defined as $\boldsymbol{M_A} = \mathbf{I}_T - \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}$. $E_i[\cdot]$ denotes unit level expectations conditional on all unit specific, time-invariant stochastic variables. Finally, all random variables are defined on a common probability space $(\Omega, \mathcal{A}, P)$.

2 | THE COMMON CORRELATED EFFECTS (CCE) SETUP

2.1 | The Model

In this paper, we consider the scalar panel data variable $y_{i,t}$, observed for $t = 1, \dots, T$ time periods and $i = 1, \dots, N$ cross-section units. The data generating process (DGP) of the stacked $[T \times 1]$ vector $\boldsymbol{y}_i = (y_{i,1}, \dots, y_{i,T})'$ is given by

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

where $\boldsymbol{X}_i = (\boldsymbol{x}_{i,1}, \dots, \boldsymbol{x}_{i,T})'$ is a $[T \times K]$ matrix of covariates, $\boldsymbol{\beta}_i$ is a $[K \times 1]$ vector of corresponding (individual-specific) parameters, $\boldsymbol{F} = (\boldsymbol{f}_1, \dots, \boldsymbol{f}_T)'$ is a $[T \times R]$ matrix of unobserved common factors, λ_i is a $[R \times 1]$ vector of factor loadings, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,T})'$ is a $[T \times 1]$ vector of idiosyncratic errors. If $\boldsymbol{X}_i$ is allowed to be correlated with $\boldsymbol{F}\boldsymbol{\lambda}_i$ (e.g., through the individual specific fixed-effects), then the standard pooled OLS and two-way fixed effects estimators are inconsistent; see, for example, Juodis (2020) and Sarafidis and Wansbeek (2021).

If $\boldsymbol{F}$ was an observed matrix and $\boldsymbol{\beta}_i = \boldsymbol{\beta}$, then the parameter of interest would be consistently estimable (for T large) using the pooled least squares (fixed effects) estimator of the form

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{M_F}\boldsymbol{X}_i\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{M_F}\boldsymbol{y}_i\right). \tag{2}$$

In most cases, however, $\boldsymbol{F}$ is unobserved. To circumvent this problem, we follow the suggestion of Pesaran (2006) (see also Bai & Li, 2014; Li et al., 2020) and assume that covariates $\boldsymbol{X}_i$ are linear in factors

$$\boldsymbol{X}_i = \boldsymbol{F}\Lambda_i + \boldsymbol{F}_\perp\Lambda_{i,\perp} + \boldsymbol{V}_i, \tag{3}$$

where $\boldsymbol{F}_\perp = (\boldsymbol{f}_{1,\perp}, \dots, \boldsymbol{f}_{T,\perp})'$ is a $[T \times R_\perp]$ factor matrix, while $\Lambda_{i,\perp}$ is the corresponding $[R_\perp \times K]$ matrix of factor loadings. Finally, $\boldsymbol{V}_i = (\boldsymbol{v}_{i,1}, \dots, \boldsymbol{v}_{i,T})'$ is a $[T \times K]$ matrix of idiosyncratic errors. Unlike the majority of the follow up literature to Pesaran (2006), we explicitly assume that K regressors are driven by more factors than the composite error term of the variable of interest $\boldsymbol{y}_i$, that is, $R_\perp \geq 0$. This distinction between the factors in Equations (1) and (3) is essential, as we will explicitly consider the setup where all factors in $\boldsymbol{F}$ can be consistently estimated (up to a rotation), while this is not the case for $\boldsymbol{F}_\perp$.

Finally, we assume that to estimate $\boldsymbol{F}$ factors in $\boldsymbol{y}_i$ the researcher uses K_z observed variables $\boldsymbol{Z}_i$, that are also assumed to be linear in factors

$$\boldsymbol{Z}_i = \boldsymbol{F}\boldsymbol{C}_i + \boldsymbol{F}_\perp\boldsymbol{C}_{i,\perp} + \boldsymbol{U}_i. \tag{4}$$

Here, analogously to Equation (3), $\boldsymbol{C}_{i,\perp}$ is a $[R_\perp \times K_z]$ matrix of factor loadings and $\boldsymbol{U}_i = (\boldsymbol{u}_{i,1}, \dots, \boldsymbol{u}_{i,T})'$ is a $[T \times K_z]$ matrix of idiosyncratic errors. This formulation is quite general, as it captures among others the original model of Pesaran (2006), where $\boldsymbol{Z}_i = (\boldsymbol{y}_i, \boldsymbol{X}_i)$, as well as the setups of Pesaran et al. (2013) and Karabıyık et al. (2019), where additional variables (not included among $\boldsymbol{X}_i$) are also included in the definition of $\boldsymbol{Z}_i$.

Pesaran (2006) proposed the common correlated effects pooled (CCEP) estimator with $\boldsymbol{F}$ replaced by the estimator $\widehat{\boldsymbol{F}} = \overline{\boldsymbol{Z}}$ in Equation (2), that is,

$$\hat{\boldsymbol{\beta}}_{CCEP} = \left(\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{X}_i\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{y}_i\right). \tag{5}$$

As it is well acknowledged in the literature, the so-called "rank condition":

$$\text{rk}(E[(\boldsymbol{C}_i, \boldsymbol{C}_{i,\perp})]) = R + R_\perp \leq K_z \tag{6}$$

should hold for this estimator to have a well behaved normal limiting distribution; see, for example, Westerlund and Urbain (2013). In contrast to Pesaran (2006), we do not assume that Equation (6) holds. Instead, motivated by the fact that $\boldsymbol{y}_i$ is the only variable of interest, while the set of covariates $\boldsymbol{X}_i$ is usually much larger, we assume that at least all factors in $\boldsymbol{F}$ can be consistently estimated from cross-section averages of $\boldsymbol{Z}_i$. Following Juodis et al. (2021) we assume that $\boldsymbol{F}_\perp$ are potentially inestimable from the cross-section averages. For specific DGPs that motivate this choice, we refer to Juodis et al. (2021). See also Li et al. (2020) and Norkutė et al. (2021) for an alternative motivation for a setup with more factors in $\boldsymbol{X}_i$ than in $\boldsymbol{y}_i$.

This paper. The results in this paper are built upon a specific deviation from the usual “rank condition”. In particular, instead of Equation (6) we assume that

$$\mathrm{rk}(E[\boldsymbol{C}_i]) = R, \text{ and } \mathrm{rk}(E[\boldsymbol{C}_{i,\perp}]) = 0. \tag{7}$$

Under this assumption the cross-sectional averages of $\boldsymbol{Z}_i$ are absolutely uninformative about $\boldsymbol{F}_\perp$.

Remark 1. In this paper, we assume that the number of estimable factors is always R. Alternatively, a setup where the number of factors estimated by $\overline{\boldsymbol{Z}}$ (call it R_z) is larger than R can be allowed, as for $R_z > R$ we can always augment the vector λ_i with $R_z - R$ rows of zeros. See the supporting information for a setup to motivate the decomposition into $\boldsymbol{F}$ and $\boldsymbol{F}_\perp$ factors.

2.2 | The problem of too many cross-section averages

As shown in Theorem 3.1. of Juodis et al. (2021), the presence of $\boldsymbol{F}_\perp$ in $\boldsymbol{Z}_i$ and/or $\boldsymbol{X}_i$ implies that for $K_z > R$ the CCEP estimator can have a non-standard asymptotic distribution. In particular, for a special case of $\boldsymbol{Z}_i = (\boldsymbol{y}_i, \boldsymbol{X}_i)$ and $N \approx T$, the aforementioned paper shows that

$$\sqrt{NT}(\hat{\boldsymbol{\beta}}_{CCEP} - \boldsymbol{\beta}_0) = \Sigma_X^{-1}\left(\boldsymbol{b}_0 + \sqrt{\frac{N}{T}}\boldsymbol{b}_1 + \sqrt{\frac{T}{N}}\boldsymbol{b}_2 + \sqrt{T}\boldsymbol{\xi}\right) + o_P(1). \tag{8}$$

Here $\boldsymbol{b}_0$ is the mean-zero asymptotically normal variance component which is present even if the factor component $\boldsymbol{F}\boldsymbol{\lambda}_i$ is known. The first bias term, $\boldsymbol{b}_1$, is the “Nickell bias” associated with models containing weakly exogenous regressors that are estimated using the “fixed effects” approach. The other terms, $\boldsymbol{b}_2$ and $\boldsymbol{\xi}$, originate from the estimation error when factors $\boldsymbol{F}$ are replaced by the corresponding estimates $\widehat{\boldsymbol{F}} = \overline{\boldsymbol{Z}}$. In particular, here $\boldsymbol{b}_2$ is the incidental parameter bias term originating from estimating $\mathcal{O}(T)$ elements in $\widehat{\boldsymbol{F}}$.

The most challenging component in the above decomposition is $\boldsymbol{\xi}$. In particular, $\boldsymbol{\xi} = \mathcal{O}_P(1)$ is a non-linear function of sample averages of $\boldsymbol{C}_{i,\perp}$, that is, the factor loadings of factors that are not estimable by $\overline{\boldsymbol{Z}}$. Hence, the asymptotic distribution of $\hat{\boldsymbol{\beta}}_{CCEP}$ is not (mixed-) normal. Furthermore, as $\boldsymbol{C}_{i,\perp}$ are inestimable from the residuals, the asymptotic distribution of $\hat{\boldsymbol{\beta}}_{CCEP}$ cannot be easily replicated/simulated using, for example, bootstrap or re-sampling. Finally, the asymptotic behavior of the inverse term Σ_X^{-1} (the distributional limit of the inverse term in Equation (5)) is also non-standard. This limit is generally stochastic and correlated in a non-linear way with the stochastic components in $\boldsymbol{\xi}$.[3]

Heterogeneous coefficients. The common way to model heterogeneous coefficients is the so-called random coefficients model, that is,

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}_0 + \tilde{\boldsymbol{\beta}}_i, \ \tilde{\boldsymbol{\beta}}_i \sim IID(\boldsymbol{0}, \Sigma_{\boldsymbol{\beta}}). \tag{9}$$

If the joint distribution of the factor loadings and $\tilde{\boldsymbol{\beta}}_i$ is unrestricted, then pooled estimators are generally inconsistent. However, we can use the CCE principle to estimate $\boldsymbol{\beta}_i$ unit-by-unit, that is,

$$\hat{\boldsymbol{\beta}}_{CCE,i} = \left(\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{X}_i\right)^{-1}\left(\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{M}_{\widehat{\boldsymbol{F}}}\boldsymbol{y}_i\right), \ i = 1, \ldots, N. \tag{10}$$

[3]We note that these components (i.e., $\boldsymbol{b}_2$, $\boldsymbol{\xi}$, and Σ_X) are non-standard only when $K_z > R$.

Under the assumptions of Pesaran (2006) with $R_\perp = 0$ this estimator can be expanded as

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_{CCE,i} - \boldsymbol{\beta}_i) = \Sigma_{X,i}^{-1}\boldsymbol{b}_{0,i} + o_P(1). \tag{11}$$

Here $\boldsymbol{b}_{0,i}$ is an asymptotically normal random variable, while $\Sigma_{X,i}^{-1}$ is the unit-specific limit of the corresponding term in Equation (10).

These results explicitly build upon the fact that $R_\perp = 0$. If this condition is violated, then one can show[4] that the presence of $\boldsymbol{F}_\perp$ leads to a non-trivial contribution to the distribution of the unit-by-unit estimator:

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_{CCE,i} - \boldsymbol{\beta}_i) = \Sigma_{X,i}^{-1}\left(\boldsymbol{b}_{0,i} + \sqrt{\frac{T}{N}}\xi_i\right) + o_P(1). \tag{12}$$

Here, as with pooled estimator, $\Sigma_{X,i}$ is a random matrix, while $\boldsymbol{\xi}_i = \mathcal{O}_P(1)$ is a non-linear function of random variables generated by cross-section averages of $\boldsymbol{C}_{i,\perp}$. Thus, unlike $\boldsymbol{b}_{0,i}$ that are (conditionally on factors) independent for all pairs (i, j), the residual components $\boldsymbol{\xi}_i$ are strongly dependent between all units $\boldsymbol{i}$.

The individual specific coefficients are generally not of primal interest in empirical research. Instead, it is common to consider the sample average of the estimated coefficients $\hat{\boldsymbol{\beta}}_{CCE,i}$, that is, the so-called mean-group CCE estimator (CCE-MG):

$$\hat{\boldsymbol{\beta}}_{CCEMG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_{CCE,i}. \tag{13}$$

In our setup with $R_\perp > 0$, this estimator can be expanded

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{CCEMG} - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\boldsymbol{\beta}}_i + \underbrace{\frac{1}{N}\sum_{i=1}^{N}\Sigma_{X,i}^{-1}\xi_i}_{=\xi_{MG}} + o_P(1). \tag{14}$$

The additional component $\xi_{MG} = \mathcal{O}_P(1)$ is a non-linear function of asymptotically normal random variables. Hence, all the negative properties associated with the pooled estimator also extend to mean-group estimator.

3 | DISCUSSION OF THE MAIN RESULT

In this section, we introduce the regularized CCE estimator and summarize the main theoretical results of this approach. All formal statements are relegated to Section 4.

3.1 | Regularized estimation

The presence of non-standard (non-normal) components ξ and Σ_X^{-1} makes valid inference on $\boldsymbol{\beta}_0$ using the CCE methodology practically impossible. In what follows, we show that a simple modification of the CCE procedure is sufficient to solve these problems. To be specific, we suggest the following procedure to construct the regularized version of the CCEP and CCE-MG estimators, for a given choice of $\boldsymbol{Z}_i$ (that might differ from the standard option of $\boldsymbol{Z}_i = (\boldsymbol{y}_i, \boldsymbol{X}_i)$).

1. Construct $\hat{\Sigma} = (NT)^{-1}\sum_{i=1}^{N}(\boldsymbol{Z}_i - \overline{\boldsymbol{Z}})'(\boldsymbol{Z}_i - \overline{\boldsymbol{Z}})$.
2. Construct the normalized factor proxies $\hat{\boldsymbol{F}} = \overline{\boldsymbol{Z}}\left(\hat{\Sigma}^{-1/2}\right)'$.[5]
3. Use the Eigenvalue Ratio (ER) approach of Ahn and Horenstein (2013) to estimate R:

$$\hat{R} = \text{argmax}_{r\in\{1,\ldots,r_{\max}\}}ER(r);\ ER(r) = \frac{\hat{v}_r}{\hat{v}_{r+1}}, \tag{15}$$

where $\hat{v}_r$ is the r^{th} largest eigenvalue of $T^{-1}\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}$. Set $r_{\max} = K_z - 1$.[6]

[4]Follows immediately from Theorem 3.1 in Juodis et al. (2021).

[5]Here we follow the convention to denote by $\boldsymbol{A}^{-1/2}$ the inverse of a square root of symmetric positive definite matrix $\boldsymbol{A}$.

[6]See Remark 6 in Section 3.3 for implementation with $r_{\max} = K_z$.

4. Construct $\widehat{\boldsymbol{F}}_r = \sqrt{T}\widehat{\boldsymbol{U}}_{\hat{R}}$, where $\widehat{\boldsymbol{U}}_{\hat{R}}$ are the associated eigenvectors corresponding to the first $\hat{R}$ largest eigenvalues of $T^{-1}\widehat{\boldsymbol{F}}\widehat{\boldsymbol{F}}'$.
5. Use $\widehat{\boldsymbol{F}}_r$ to define the *regularized* CCEP/CCE-MG estimator:

$$\hat{\boldsymbol{\beta}}_{rCCEP} = \left(\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{M}_{\widehat{\boldsymbol{F}}_r}\boldsymbol{X}_i\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{M}_{\widehat{\boldsymbol{F}}_r}\boldsymbol{y}_i\right), \tag{16}$$

or

$$\hat{\boldsymbol{\beta}}_{rCCEMG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_{rCCE,i} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{M}_{\widehat{\boldsymbol{F}}_r}\boldsymbol{X}_i\right)^{-1}\left(\frac{1}{T}\boldsymbol{X}_i'\boldsymbol{M}_{\widehat{\boldsymbol{F}}_r}\boldsymbol{y}_i\right). \tag{17}$$

In Section 4 we show that rCCE estimators have asymptotically normal distribution over a wider class of DGPs than the original CCE estimators.

In order to appreciate how large the gains from regularization can be, we consider a simplified version of Equations (1) and (3) with one regressor (i.e., $K_z = 2$) and $R = R_{\perp} = 1$. The finite sample results for the pooled estimators are illustrated graphically in Figure A1.

From the figure we see that due to the non-linearities associated with $\boldsymbol{\xi}$ and Σ_X^{-1}, the standard CCEP estimator has a bimodal distribution, invalidating normal approximation. The regularized version, on the other hand, is uni-modal and resembles the normal distribution well. Finally, in this example, in almost 99% of the cases the ER criterion selects the correct number of estimable actors, $R = 1$.

Remark 2. We use the singular value based regularization idea similar to the one recently used by Juodis and Sarafidis (2022) in the context of fixed T factor-augmented panel data models with endogenous regressors. The main difference is that in this context we use the least squares principles for estimation, whereas they use the GMM estimator of Robertson and Sarafidis (2015). Moreover, in this paper we mostly focus on the case where jointly $N, T \to \infty$.

The proposed regularization procedure (and the estimator of R) is closely related to the literature on rank tests; see, for example, Robin and Smith (2000) and Kleibergen and Paap (2006). In particular, normalization by $\left(\widehat{\Sigma}^{-1/2}\right)'$ relates to the common practice in that literature to consider rank statistics over appropriately rotated quantities. Normalization by $\left(\widehat{\Sigma}^{-1/2}\right)'$ ensures that eigenvalues and eigenvectors are invariant to any non-singular column transformations of $\boldsymbol{Z}_i$. This way the estimator $\widehat{\boldsymbol{F}}_r$ is invariant to $\boldsymbol{\beta}$ when $\boldsymbol{Z}_i = (\boldsymbol{y}_i, \boldsymbol{X}_i)$, that is, the property shared by the original CCE estimator.

Estimation of R from $\overline{\boldsymbol{Z}}$ is fairly straightforward as compared to the usual principal components/factor estimation of R (and $R_{\perp}$). As at least one of the dimensions of $\overline{\boldsymbol{Z}}$ is fixed, consistent estimation of R can be done for T fixed/large, with only $N \to \infty$. In contrast, R (and $R_{\perp}$) in the principal components (PC) literature is estimated from the sample variance–covariance matrices, for example, $N^{-1}\sum_{i=1}^{N}\boldsymbol{y}_i\boldsymbol{y}_i'$ this requires that both $N \to \infty$ and $T \to \infty$, making it a more demanding task for moderate sized panels.[7]

Finally, as with any CCE procedure, our proposed regularized CCE estimator is less sensitive to weak factors (see, e.g., Chudik et al., 2011) than PC estimators, studied in, for example, Westerlund and Urbain (2015) and Li et al. (2020). On the other hand, similar to the standard CCE estimator, the rCCE estimator requires a restriction on the rank condition as in Equation (7).

3.2 | Inference

Below we summarize the inferential strategy for the pooled estimator. The same procedure can be directly applied to the regularized CCE-MG estimator after ignoring Step 2. It is important to notice that, the commonly used plug-in

[7] See, for example, Breitung and Hansen (2021), for some recent simulation results.

variance–covariance matrix estimator of the form

$$\hat{\Sigma}_{rCCEMG} = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\beta}}_{rCCE,i} - \hat{\boldsymbol{\beta}}_{rCCEMG})(\hat{\boldsymbol{\beta}}_{rCCE,i} - \hat{\boldsymbol{\beta}}_{rCCEMG})', \tag{18}$$

can be inconsistent for the variance–covariance matrix of $\hat{\boldsymbol{\beta}}_{rCCEMG}$. In particular, this estimator fails to account for the estimation uncertainty coming from the regularized analog of $\boldsymbol{\xi}_{MG}$. For this reason, we suggest that the proposed bootstrap procedure should be used irrespective of the degree of heterogeneity in $\boldsymbol{\beta}_i$.

1. Obtain $\hat{\boldsymbol{\beta}}_{rCCEP}$ and $\hat{R}$ using the procedure outlined above.
2. Construct the bias-corrected estimator $\tilde{\boldsymbol{\beta}}_{rCCEP}$ by removing the $\boldsymbol{b}_2$ bias first (using the analytical formulae in the supporting information). If the Half-Panel jackknife correction of Dhaene and Jochmans (2015) for the "Nickell" bias is also needed, then the $\boldsymbol{b}_2$ component should be removed for all three estimators (i.e., for the full sample and the two half-sample $\hat{\boldsymbol{\beta}}_{rCCEP}$ estimates separately).[8] For all sub-sample estimators take $\hat{R}$ as estimated in Step 1.
3. Let $\boldsymbol{Q}_i = (\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i)$, $i = 1, \ldots, N$. The cross-section bootstrap randomly draws $\boldsymbol{Q}_1^*, \ldots, \boldsymbol{Q}_N^*$ from $\{\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_N\}$ with replacement. Use $\{\boldsymbol{Q}_1^*, \ldots, \boldsymbol{Q}_N^*\}$ to construct $\tilde{\boldsymbol{\beta}}^*_{rCCEP,b}$, taking $\hat{R}$ as in Step 1.
4. Repeat Step 3 B times and collect all estimators $\{\tilde{\boldsymbol{\beta}}^*_{rCCEP,b}\}_{b=1}^{B}$.
5. Construct percentile bootstrap confidence intervals using $\{\tilde{\boldsymbol{\beta}}^*_{rCCEP,b}\}_{b=1}^{B}$.

When all regressors are strictly exogenous the Half-Panel jackknife correction in Step 2 is not necessary. $\boldsymbol{b}_2$ bias term, on the other hand, is generally non-zero for all CCEP estimators (even if $R_{\perp} = 0$); see, for example, Westerlund and Urbain (2015), Westerlund (2018), and Karabıyık et al. (2019). Thus, this is not a by-product of the regularization procedure suggested in this paper. As the proposed procedure mostly addresses the way factor proxies $\hat{\boldsymbol{F}}_r$ are constructed (and corresponding sampling uncertainty), and can be used for any model that uses $\overline{\boldsymbol{Z}}$ for factor proxies, for example, Focused Information Criterion based model averaging of Yin et al. (2021), the gravity model of Desbordes and Eberhardt (2019), or the discrete choice model of Boneva and Linton (2017).

Remark 3 (Bootstrap)**.** The bootstrap procedure largely follows the algorithm of Galvao and Kato (2014), where percentile bootstrap confidence intervals are constructed for de-biased estimators. Finite sample evidences provided in this paper indicate that the proposed bootstrap procedure has good size control and power for reasonable values of B, for example, $B = 199$. While we do not formally prove consistency of the outlined bootstrap procedure, we conjecture that (point-wise) consistency of this procedure can be established using the proof strategy of Galvao and Kato (2014), Westerlund et al. (2019), and De Vos and Stauskas (2021). Here point-wise consistency is defined with respect to the two DGPs indexed by the nuisance parameters $\boldsymbol{C}_{i,\perp}$ mentioned in Section 2. For more details, see the supporting information.

Remark 4 (Fixed T Results)**.** While the main focus of this paper is on the large N, T asymptotic distribution of the rCCE estimator, the CCEP estimator can be consistent for T fixed; see, for example, Su and Jin (2012) and Westerlund et al. (2019).[9] In Section 4.4 we also study asymptotic results of the rCCEP estimator when the time-series dimension is assumed to be fixed.

The proposed procedure is only expected to perform well if the number of factors R can be estimated precisely. Given that consistency of the regularization procedure only required that N is large, while T can be fixed/large, one can expect good performance of this procedure already for cases where one of the dimensions is moderate in size. This intuition is confirmed using simulated data, where the correct selection rates exceed 90% already for $(N, T) = (20{,}20)$. Consistency of the selection procedure is proved in Section 4.2 (for T large) and Section 4.4 (for T fixed).

[8] Alternatively, use double jackknife or other variants proposed in Cruz-Gonzalez et al. (2017).

[9] However, only the paper of Westerlund et al. (2019) rigorously proves consistency and (mixed-) asymptotic normality of the CCEP estimator for the setup with $K_z > R$ and $R_{\perp} = 0$.

Remark 5. While the suggested procedure is informative about R, it is completely silent about $R_\perp$. Thus, even if for a given dataset we observe that $\hat{R} < K_z$, this should not be interpreted as any evidence for the non-standard distribution of the regular CCE estimator (as $R_\perp = 0$ case cannot be ruled out).

3.3 | Implementation

In this section, we provide additional comments regarding the implementability of the procedure discussed in this paper. Additional technical notes are relegated to the supporting information.

Remark 6 (Maximum Number of Factors). To allow for the possibility that $r_{\max} = K_z$, we use the dummy-eigenvalue idea put forward in Juodis and Sarafidis (2022): in the second step we augment $\boldsymbol{Z}_i$ with an additional column that has zero expectations by construction. Our preferred choice is to augment the original factor proxies $\widehat{\boldsymbol{F}}$, with a column $\hat{\boldsymbol{f}}_p = \widehat{\boldsymbol{F}}_p \boldsymbol{\iota}_{K_z}/K_z$. Here the "perturbed" factor proxies $\widehat{\boldsymbol{F}}_p$ are constructed as $\widehat{\boldsymbol{F}}$ upon replacing $\boldsymbol{Z}_i$ by $\boldsymbol{Z}_{i,(p)} = \boldsymbol{Z}_i \omega_i$, where ω_i follows the Rademacher $\{-1; 1\}$ distribution.

Remark 7 (Observed Factors). It is a common practice to include $\boldsymbol{\iota}_T$ as constant factor in the model, that is, the standard time-invariant fixed effect. To accommodate such possibility, the regularized factor proxies should be calculated on the residualized data (with respect to observed factors). This modification plays no major role for the main result of this paper, except for the "Nickell bias" term which should be appropriately adjusted.

Remark 8 (Minimum Number of Factors). In this paper, we explicitly assume that $R > 0$. If this assumption is expected to be violated, then it can be tested using any (appropriately adjusted) standard rank test, for example, Kleibergen and Paap (2006). However, we believe that $R > 0$ assumption is highly plausible for most applications, as evidently supported by the common practice of including time-specific (intercepts) effects in empirical models.

Remark 9 (On Bootstrap Resampling). Westerlund et al. (2019) also mention the use of cross-section bootstrap in the context of CCEP estimator. However, while they suggest the fixed $\widehat{\boldsymbol{F}}^{(b)} = \widehat{\boldsymbol{F}}$ bootstrap procedure (see their supporting information), we follow Goncalves and Perron (2014) and re-estimate $\widehat{\boldsymbol{F}}^{(b)}$ in every bootstrap replication. This way, we appropriately account for factor estimation uncertainty that can be non-neglile for $R_\perp > 0$. In contrast, the setup of Westerlund et al. (2019) is for $R_\perp = 0$, thus the additional re-estimation step is redundant.

4 | LARGE SAMPLE RESULTS

4.1 | Assumptions

In what follows we discuss a set of sufficient conditions used throughout this paper. These assumptions are mostly inspired by those of Pesaran (2006) and Juodis et al. (2021), but are appropriately modified for the purpose of this paper.

For the DGP in Equations (1), (3), and (4), denote by $\boldsymbol{e}_{i,t} = (\varepsilon_{i,t}, \boldsymbol{v}'_{i,t}, \boldsymbol{u}'_{i,t})'$ the full vector of the idiosyncratic components. Furthermore, denote by $\boldsymbol{d}_t = (\boldsymbol{f}'_t, \boldsymbol{f}'_{t,\perp})'$ the full vector of unobserved factors. It is useful to introduce the following stationary autocovariance functions for any lag value h:

$$\Gamma_{\boldsymbol{d}}(h) = E[(\boldsymbol{d}_t - E[\boldsymbol{d}_t])(\boldsymbol{d}_{t-h} - E[\boldsymbol{d}_t])'] = \begin{pmatrix} \Gamma(h) & \Gamma_\perp(h) \\ \Gamma_\perp(-h)' & \Gamma_{\perp,\perp}(h) \end{pmatrix}, \tag{19}$$

$$\Gamma_{i,\boldsymbol{e}}(h) = E_i[(\boldsymbol{e}_{i,t} - E_i[\boldsymbol{e}_{i,t}])(\boldsymbol{e}_{i,t-h} - E_i[\boldsymbol{e}_{i,t}])'] = \begin{pmatrix} \Gamma_{i,\varepsilon,\varepsilon}(h) & \Gamma_{i,\varepsilon,\boldsymbol{v}}(h) & \Gamma_{i,\varepsilon,\boldsymbol{u}}(h) \\ \Gamma_{i,\varepsilon,\boldsymbol{v}}(-h)' & \Gamma_{i,\boldsymbol{v},\boldsymbol{v}}(h) & \Gamma_{i,\boldsymbol{v},\boldsymbol{u}}(h) \\ \Gamma_{i,\varepsilon,\boldsymbol{u}}(-h)' & \Gamma_{i,\boldsymbol{v},\boldsymbol{u}}(-h)' & \Gamma_{i,\boldsymbol{u},\boldsymbol{u}}(h) \end{pmatrix}. \tag{20}$$

Here $E_i[\cdot]$ denotes the expectation operator conditional on all i unit-specific time-invariant random variables. Given the conditioning argument, the moment restrictions in the next assumption should hold almost surely. Let ℓ be some positive integer to be specified later.

Assumption 4.1 (Errors)**.** (a) (i) $\boldsymbol{e}_{i,t}$ is a covariance stationary process independent across i; (ii) $\boldsymbol{e}_{i,t}$ and $\boldsymbol{e}_{j,s}$ are independent for all t, s and $i \neq j$; (b) $E_i[\boldsymbol{e}_{i,t}] = \mathbf{0}$ and $E_i[\|\boldsymbol{e}_{i,t}\|^{\ell+\delta}] < \infty$; (c) (i) $\boldsymbol{e}_{i,t}$ admits factorization $\boldsymbol{e}_{i,t} = \Omega_i \tilde{\boldsymbol{e}}_{i,t}$ with: $\Gamma_{i,\boldsymbol{e}}(h) = \Omega_i \Gamma_{\tilde{\boldsymbol{e}}}(h) \Omega_i'$; (ii) The sequence $\{\Gamma_{\tilde{\boldsymbol{e}}}(h)\}_{h=-\infty}^{\infty}$ is absolutely summable; (d) (i) $\Gamma_{i,\varepsilon,\boldsymbol{v}}(h) = \mathbf{0}_K'$ for $h \geq 0$. $\Gamma_{i,\varepsilon,\varepsilon}(h) = 0$ for $h \neq 0$; (ii) $\Gamma_{i,\boldsymbol{v},\boldsymbol{v}}(0)$ and $\Gamma_{i,\boldsymbol{u},\boldsymbol{u}}(0)$ are positive definite matrices for every i.

Assumption 4.1 is fairly standard in the panel data literature with weakly exogenous regressors and is inspired by those in Pesaran (2006) and Juodis et al. (2021). For example, similarly to Moon and Weidner (2017) we allow the regressors $\boldsymbol{X}_i$ to be weakly exogenous, as $\Gamma_{i,\varepsilon,\boldsymbol{v}}(h)$ is unrestricted for $h < 0$. To allow for this possibility and, at the same time, to rule out any endogeneity concerns, we assume that $\varepsilon_{i,t}$ are serially uncorrelated in (d).[10]

Assumption 4.2 (Factors)**.** (a) (i) $\boldsymbol{d}_t$ is a covariance stationary process with $E[\boldsymbol{d}_t] = \mathbf{0}_{R+R_\perp}$ and $\Sigma_{\boldsymbol{d}} = \Gamma_{\boldsymbol{d}}(0)$ a positive definite matrix; (ii) $E[\|\boldsymbol{d}_t\|^{r+\delta}] < \infty$; (b) The sequence $\{\Gamma_{\boldsymbol{d}}(h)\}_{h=-\infty}^{\infty}$ is absolutely summable; (c) $\Gamma_\perp(0) = \mathbf{O}_{R\times R_\perp}$; (d) $\boldsymbol{e}_{i,t}$ and $\boldsymbol{d}_s$ are mutually independent for all i, t and s.

The restriction on $E[\boldsymbol{d}_t] = \mathbf{0}$ is without loss of generality, as long as the model contains time-invariant fixed effects, for example, as in our empirical application. Finally, as it is argued by Juodis and Reese (2021),[11] we can without loss of generality assume that $\Gamma_\perp(0) = \mathbf{O}_{R\times R_\perp}$.

For the next assumption, definite the following stacked vector

$$\boldsymbol{h}_i = (\lambda_i', \mathrm{vec}(\Omega_i)', \mathrm{vec}(\boldsymbol{C}_i)', \mathrm{vec}(\boldsymbol{C}_{i,\perp})', \mathrm{vec}(\Lambda_i)', \mathrm{vec}(\Lambda_{i,\perp})')'. \tag{21}$$

Assumption 4.3 (Unit Heterogeneity)**.** (a) The random vector $\boldsymbol{h}_i$ is independent and identically distributed over i with $E[\|\boldsymbol{h}_i\|^{\ell+\delta}] < \infty$; (b) $E[\boldsymbol{C}_i] = \boldsymbol{C}$ such that $\mathrm{rk}(\boldsymbol{C}) = R$ with $0 < R < K_z$ while $E[\boldsymbol{C}_{i,\perp}] = \mathbf{O}$; (c) $\boldsymbol{h}_i$ and $\boldsymbol{d}_t$ are mutually independent for all i and t; (d) $\boldsymbol{h}_i$ and $\tilde{\boldsymbol{e}}_{j,s}$ are mutually independent for all i, j, t, s.

Notice that this assumption leaves the dependence structure between the individual elements of $\boldsymbol{h}_i$ completely unrestricted. In this regard, we follow the conventional "fixed-effects" approach (even if $\boldsymbol{h}_i$ is assumed to be a random vector). Assumption 4.3 (b) is the relevant rank condition for this paper which together with Assumption 4.1 directly implies that:

$$\overline{\boldsymbol{Z}}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1} \xrightarrow{p} \boldsymbol{F}, \tag{22}$$

as $N \to \infty$ for any fixed T.

Assumption 4.4 (Eigenvalues)**.** The eigenvalues of the $[R \times R]$ matrix $\boldsymbol{C}E[\widehat{\Sigma}]^{-1}\boldsymbol{C}'\Gamma(0)$ are distinct.

Assumption 4.4 is analogous to the eigenvalue condition in Bai (2003) (Assumption G), which ensures the existence of a well defined asymptotic rotation matrix for $\tilde{\boldsymbol{F}}_r$.[12] We note that Assumption 4.4 rules out any forms of "weak" factors from the model; see, for example, Chudik et al. (2011). For those setups the matrix $\boldsymbol{C}$ cannot be assumed to be a fixed constant, and instead should be modeled as a drifting sequence.

Assumption 4.5 (Asymptotics)**.** $N/T \to c$ as N, $T \to \infty$ jointly with $c \in (0, \infty)$.

Assumption 4.5 bounds the relative expansion rate of N and T such that $c \neq 0$ and $c^{-1} \neq 0$; see, for example, Fernández-Val and Weidner (2016) and Juodis et al. (2021).

4.2 | Pooled estimator

Our first result formally establishes asymptotic validity of the ER procedure.

[10] Once appropriately extended our results can be also used to derive asymptotic properties of the CCE-GMM procedure of Everaert and Pozzi (2014).

[11] See Lemma S.18 in the supporting information of that paper.

[12] In this paper, we will not study in detail the properties of the factor estimates themselves, only their contribution to the asymptotic distribution of the rCCE estimators.

Proposition 1. *Under Assumptions 4.1-4.5 with $\ell \geq 4$:*

$$\lim_{N,T\to\infty} P(\hat{R} = R) = 1. \tag{23}$$

Using this result, we can treat the true number of factors R as given in the remainder of this section. In Section 4.4 we also show that a similar conclusion also holds for T fixed (under a slightly modified set of regularity conditions).

The next theorem fully characterizes the asymptotic properties of the pooled regularized CCE estimator. The results are provided in the "semi-asymptotic" form (i.e., without taking probability limits in N). This approach makes the direct comparison with the results in, for example, Westerlund and Urbain (2015) and Juodis et al. (2021) relatively straightforward. Finally, we denote by $\hat{\Sigma}_{\boldsymbol{F}_\perp} = T^{-1}\boldsymbol{F}_\perp'\boldsymbol{F}_\perp$, $\boldsymbol{G} = \overline{\boldsymbol{C}}$, $\boldsymbol{G}_\perp = \overline{\boldsymbol{C}_\perp}$, and $\boldsymbol{V}_{i,+} = \boldsymbol{V}_i + \boldsymbol{F}_\perp\Lambda_{i,\perp}$.

Theorem 1. *Under Assumptions 4.1–4.5 with $\ell \geq 4$:*

$$\sqrt{NT}(\hat{\boldsymbol{\beta}}_{rCCEP} - \boldsymbol{\beta}_0) = \Sigma_{X,r}^{-1}\left(\boldsymbol{b}_0 + \sqrt{\frac{N}{T}}\boldsymbol{b}_1 + \sqrt{\frac{T}{N}}\boldsymbol{b}_{2,r} + \sqrt{T}\boldsymbol{\xi}_r\right) + o_P(1), \tag{24}$$

where:

$$\Sigma_{X,r} = E[\boldsymbol{v}_{i,t}\boldsymbol{v}_{i,t}'] + E[\Lambda_{i,\perp}'\Gamma_{\perp,\perp}(0)\Lambda_{i,\perp}] + o_P(1), \tag{25}$$

$$\boldsymbol{b}_0 = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{V}_{i,+}'\boldsymbol{\varepsilon}_i + \left(\frac{1}{N}\sum_{i=1}^{N}\lambda_i'\otimes\Lambda_{i,\perp}'\right)\left(\boldsymbol{S}'\otimes\mathbf{I}_{R_\perp}\right)\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}vec(\boldsymbol{F}_\perp'\boldsymbol{U}_i), \tag{26}$$

$$\boldsymbol{b}_1 = -\sum_{h=1}^{\infty}E[\Gamma_{i,\varepsilon,\boldsymbol{v}}(-h)']tr[\Gamma(h)\Gamma(0)^{-1}], \tag{27}$$

$$\begin{aligned}\boldsymbol{b}_{2,r} = {} & \frac{1}{N}\sum_{i=1}^{N}\Lambda_i'\boldsymbol{S}'\left(N\boldsymbol{G}_\perp'\hat{\Sigma}_{\boldsymbol{F}_\perp}\boldsymbol{G}_\perp + \frac{1}{N}\sum_{i=1}^{N}E_i[\boldsymbol{u}_{i,t}\boldsymbol{u}_{i,t}']\right)\boldsymbol{S}\lambda_i \\ & + \frac{1}{N}\sum_{i=1}^{N}\Lambda_i'\boldsymbol{S}'E_i[\boldsymbol{u}_{i,t}\varepsilon_{i,t}] + \frac{1}{N}\sum_{i=1}^{N}E_i[\boldsymbol{v}_{i,t}\boldsymbol{u}_{i,t}']\boldsymbol{S}\lambda_i,\end{aligned} \tag{28}$$

$$\boldsymbol{\xi}_r = \left(\frac{1}{N}\sum_{i=1}^{N}\lambda_i'\otimes\Lambda_{i,\perp}'\right)\left(\boldsymbol{S}'\otimes\hat{\Sigma}_{\boldsymbol{F}_\perp}\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}vec\left(\boldsymbol{C}_{i,\perp}\right), \tag{29}$$

$$\boldsymbol{S} = -\hat{\Sigma}^{-1}\boldsymbol{G}'(\boldsymbol{G}\hat{\Sigma}^{-1}\boldsymbol{G}')^{-1}. \tag{30}$$

Here all terms are $\mathcal{O}_P(1)$.

Below we briefly explain the properties of all stochastic components presented in Theorem 1. The variance component $\boldsymbol{b}_0$ consists of two terms. The first part is identical to that of the infeasible (oracle) estimator where $\boldsymbol{F}$ is known, while the second part is the variance contribution due to the estimation error originating from $\boldsymbol{U}_i$ (the idiosyncratic part of $\boldsymbol{Z}_i$). Using Theorem 3.2 of Hall and Heyde (1980) it is a straightforward to show that $\boldsymbol{b}_0$ has normal limiting distribution. The first bias term, $\boldsymbol{b}_1$, is the "Nickell" bias term, only depends on $\boldsymbol{F}$ factors, thus is the same as for the infeasible estimator with $\boldsymbol{F}$ known (and also the standard CCEP estimator).

The bias term $\boldsymbol{b}_{2,r}$ consists of three components that are functions of the unit level time invariant parameters from Assumption 4.3. The first component is non-zero as long as the second cross-moment between Λ_i and λ_i is non-negligible. Furthermore, this component is stochastic in the limit $N \to \infty$ (as it is a quadratic form in $\sqrt{N}\boldsymbol{G}_\perp$). The second (third) component is non-negligible only if Λ_i (λ_i) is uncorrelated with certain elements of Ω_i, and the corresponding factor loadings are mean-zero. As we discuss in the supporting information, most of the elements in $\boldsymbol{b}_{2,r}$ can be consistently estimated using the plug-in principle. Thus, the analytical bias-correction suggested in Section 2 is feasible.

Finally, unlike the non-regularized counterpart, $\boldsymbol{\xi}_r$ is a linear function of $\overline{\boldsymbol{C}}_\perp$. This component is present due to the estimation error stemming from the common part of $\boldsymbol{Z}_i$, that is, $\boldsymbol{F}_\perp\boldsymbol{C}_{i,\perp}$. As for each pair (i,j) these contributions are strongly correlated, cross-section dependence induced by this term is such that the convergence rate of the rCCEP is primarily determined by $\boldsymbol{\xi}_r$, and not by $\boldsymbol{b}_0$. While the regularization by itself is not sufficient to recover the parametric $\sqrt{NT}$ convergence rate of the pooled estimator, this component is at least asymptotically normal, in contrast with the non-regularized CCEP estimator.

Remark 10. Notice that for $K_z = R$ and $R_\perp = 0$ (i.e., no regularization is required), the scaling matrix $\boldsymbol{S}$ reduces to $\boldsymbol{S} = -\boldsymbol{G}^{-1}$, which is exactly the result in Westerlund and Urbain (2015) and Karabıyık et al. (2019). Finally, for $\hat{\Sigma} = \mathbf{I}_{K_z}$ we obtain the result of Westerlund and Urbain (2015) for $K_z > R$ (later shown in Karabıyık et al. (2017) to be incorrect).

Regularization effectively mitigates most of the non-desirable features of the standard CCEP estimator for $R_\perp > 0$. However, the method cannot solve all issues originating from the fact that $\overline{\boldsymbol{Z}}$ are uninformative about $\boldsymbol{F}_\perp$. The presence of both $\boldsymbol{b}_0$ and ξ_r greatly complicates any attempts to perform inference which is uniform over the parameters' space indexed by the covariance structures of $\boldsymbol{F}_\perp$ and $\boldsymbol{h}_i$. The proposed bootstrap procedure can be shown to be asymptotically valid when either (for all elements point-wise) $E[(\lambda_i \otimes \Lambda_{i,\perp})] = \mathcal{O}(N^\kappa)$ for $\kappa \in (-1/2; 0]$ or $\sum_{i=1}^N (\lambda_i \otimes \Lambda_{i,\perp}) = o_P(N)$, that is, when all loadings either strongly correlated or are weak.[13] Hence, the bootstrap procedure can be shown to be point-wise consistent for the setups of Juodis et al. (2021) and Pesaran (2006), respectively. On the other hand, for some of the "intermediate" cases, where some of the factor loadings are strong but mutually uncorrelated, thus $\sqrt{N}\xi_r = \mathcal{O}_P(1)$,[14] this procedure cannot be justified for all diagonal sequences of (N, T). Nevertheless, the bootstrap based inference is expected to be well behaved, provided that in those cases N is sufficiently larger than T. A more detailed discussion regarding these intermediate cases is relegated to the supporting information.

4.3 | Mean-group estimator

In this section, we formally prove the consistency and asymptotic normality of the regularized CCE-MG estimator introduced in Equation (17). This estimator is informative about the population mean when the underlying unit-specific coefficients satisfy the random-coefficients setup in Equation (9). In what follows, we impose additional regularity conditions on $\tilde{\boldsymbol{\beta}}_i$ (and all other unit-specific characteristics).

Assumption 4.6 (Bounded Heterogeneity)**.** (a) The vector $\boldsymbol{h}_i^a = (\boldsymbol{h}_i', \tilde{\boldsymbol{\beta}}_i')'$ is independent and identically distributed over i; (b) $\text{rk}[E_i[\boldsymbol{v}_{i,t}\boldsymbol{v}_{i,t}']] = K$ a.s. for each i; (c) $\|\text{vec}(\Omega_i)\| < \Delta$, $\|\text{vec}(\Lambda_i)\| < \Delta$ and $\|\text{vec}(\Lambda_{i,\perp})\| < \Delta$ for each i.

This assumption is more restrictive than usually considered in the CCE literature, as we assume that some unit-specific variables uniformly bounded.[15] Among other things, this assumption is sufficient to claim that $\sup_i \|\boldsymbol{h}_i^a\| = \mathcal{O}_P(1)$, which greatly simplifies derivations for all Mean-Group estimator. It is important to note that we do not impose any restrictions on the covariance structure between the individual elements of $\boldsymbol{h}_i^a$.

Assumption 4.7 (Moments)**.** (a) For $\boldsymbol{D} = (\boldsymbol{F}, \boldsymbol{F}_\perp)$ and $\boldsymbol{E}_i = (\boldsymbol{e}_i, \boldsymbol{V}_i, \boldsymbol{U}_i) = \tilde{\boldsymbol{E}}_i\Omega_i'$ we have $E\left[\|T^{-1/2}\tilde{\boldsymbol{E}}_i'\boldsymbol{D}\|^4\right] < \Delta$ for all i; (b) $E\left[\|T^{-1/2}\sum_t \left(\tilde{\boldsymbol{e}}_{i,t}\tilde{\boldsymbol{e}}_{i,t}' - E\left[\tilde{\boldsymbol{e}}_{i,t}\tilde{\boldsymbol{e}}_{i,t}'\right]\right)\|^4\right] < \Delta$ for all i; (c) $E\left[\|(NT)^{-1/2}\sum_j\sum_t \left(\tilde{\boldsymbol{e}}_{i,t}\tilde{\boldsymbol{e}}_{j,t}' - E\left[\tilde{\boldsymbol{e}}_{i,t}\tilde{\boldsymbol{e}}_{j,t}'\right]\right)\|^4\right] < \Delta$ for all i.

Assumption 4.7 is an adapted version of regularity conditions in Norkutė et al. (2021). We use these high-level conditions to bound uniformly (over i) all estimation errors originating in $\hat{\boldsymbol{\beta}}_{rCCE,i} - \boldsymbol{\beta}_i$. Equipped with these additional assumptions we are able to formulate the next result.

Theorem 2. *Under Assumptions 4.1-4.7 with $\ell \geq 8$:*

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{rCCEMG} - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{\boldsymbol{\beta}}_i + \frac{1}{N}\sum_{i=1}^N \Sigma_{X,r,i}^{-1}\xi_{r,i} + o_P(1), \tag{31}$$

where

$$\Sigma_{X,r,i} = E_i[\boldsymbol{v}_{i,t}\boldsymbol{v}_{i,t}'] + \Lambda_{i,\perp}'\Gamma_{\perp,\perp}(0)\Lambda_{i,\perp}, \tag{32}$$

$$\xi_{r,i} = \left(\lambda_i' \otimes \Lambda_{i,\perp}'\right)\left(\boldsymbol{S}' \otimes \hat{\Sigma}_{\boldsymbol{F}_\perp}\right)\frac{1}{\sqrt{N}}\sum_{i=1}^N vec\left(\boldsymbol{C}_{i,\perp}\right). \tag{33}$$

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

[13] The standard inference procedures (including bootstrap) are applicable for the original CCEP estimator only in the latter case.

[14] Moreover, the stochastic part of $\boldsymbol{b}_{2,r}$ can be of the same order as that of $\boldsymbol{b}_0$ or ξ_r, see the supporting information for further discussion.

[15] This restriction can be motivated by similar conditions specified in, for example, Fernández-Val and Weidner (2016) and Juodis (2020).

From Theorem 2 we can conclude that the rCCE-MG estimator is asymptotically normal with the variance–covariance matrix determined by the composite vector $\boldsymbol{\psi}_i = (\boldsymbol{\beta}_i', \text{vec}(\boldsymbol{C}_{i,\perp})')'$. Hence, for $R_\perp > 0$ sampling uncertainty associated with $\overline{\boldsymbol{Z}}$ has a non-negligible variance effect, contrasting with the setup of Pesaran (2006). As a result, standard normal/chi-square inference based using the variance–covariance matrix estimator in Equation (18) can be both under- and over-sized (see the corresponding section in supporting information for an intuitive explanation). For this reason, cross-sectional bootstrap (with re-estimated $\overline{\boldsymbol{Z}}$ and $\widehat{\boldsymbol{F}}_r$ in every replication) is more appropriate.

4.4 | Fixed T Theory

In this section, we discuss the properties of the pooled rCCE estimator in fixed T panels (as in Westerlund et al., 2019). In particular, we re-establish the main conclusions from Proposition 1 and Theorem 1 under modified assumptions discussed in the supporting information.

At first, we prove validity of the ER criterion for fixed T.

Proposition 2. *Under Assumptions S.1.1-S.1.4 for T fixed:*

$$\lim_{N\to\infty} P(\hat{R} = R) = 1. \tag{34}$$

Hence, subject to some additional (mild) regularity conditions the R can be consistently estimated even for T fixed. Our next result characterizes the asymptotic properties of the pooled regularized CCE estimator for T fixed, where all regressors are strictly exogenous.

Theorem 3. *Under Assumptions S.1.1-S.1.4 for T fixed:*

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{rCCEP} - \boldsymbol{\beta}_0) = \Sigma_{X,r}^{-1}\left(\boldsymbol{b}_0 + \boldsymbol{\xi}_r\right) + o_P(1), \tag{35}$$

where:

$$\Sigma_{X,r} = E[\boldsymbol{V}_i'\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{V}_i|\boldsymbol{F}] + E[\Lambda_{i,\perp}'\widehat{\Sigma}_{\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{F}_\perp}\Lambda_{i,\perp}|\boldsymbol{F},\boldsymbol{F}_\perp] + o_P(1), \tag{36}$$

$$\boldsymbol{b}_0 = \frac{1}{T}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{V}_{i,+}'\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{\varepsilon}_i + \left(\frac{1}{N}\sum_{i=1}^{N}\lambda_i'\otimes\Lambda_{i,\perp}'\right)\left(\boldsymbol{S}'\otimes\mathbf{I}_{R_\perp}\right)\frac{1}{T}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}vec(\boldsymbol{F}_\perp'\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{U}_i), \tag{37}$$

$$\boldsymbol{\xi}_r = \left(\frac{1}{N}\sum_{i=1}^{N}\lambda_i'\otimes\Lambda_{i,\perp}'\right)\left(\boldsymbol{S}'\otimes\widehat{\Sigma}_{\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{F}_\perp}\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}vec\left(\boldsymbol{C}_{i,\perp}\right). \tag{38}$$

Here $\boldsymbol{S}$ *is defined as in Theorem 1 and* $\widehat{\Sigma}_{\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{F}_\perp} = T^{-1}\boldsymbol{F}_\perp'\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{F}_\perp$. *All other terms are* $\mathcal{O}_P(1)$.

Due to the presence of stochastic common factors ($\boldsymbol{F}$ and $\boldsymbol{F}_\perp$) in all leading components of Theorem 3, the asymptotic distribution of the rCCEP estimator is only mixed-normal for T fixed. The distributional result of this type is common for all fixed T estimators with common factors; see, for example, Westerlund et al. (2019) and Juodis and Sarafidis (2022). When $R_\perp = 0$ and $R = K_z$ our result coincides with that of Westerlund et al. (2019) derived for the original CCEP estimator. On the other hand, in comparison to CCEP for $R < K_z$ (with $R_\perp = 0$) the rCCEP estimator follows a more conventional asymptotic distribution.

As compared to Theorem 1, the distinguishing feature of the main result of Theorem 3 is the absence of $\boldsymbol{b}_1$ and $\boldsymbol{b}_{2,r}$ components. While the former is missing by assumption (strict exogeneity of regressors), the latter is negligible in the T fixed setup. The fact that $\boldsymbol{b}_{2,r}$ is asymptotically negligible has some non-trivial benefits as the asymptotic distribution is always (mixed-) normal. As such, the cross-sectional bootstrap (as implemented in Section 3) is always valid, irrespective of the correlation structure between the factor loadings.

Remark 11. If at least one of the regressors is only weakly exogenous, the rCCEP estimator is no longer consistent for T fixed. In that case similarly to Everaert and De Vos (2021), one can consider analytical bias-correction (for a certain type of the DGPs of $\boldsymbol{X}_i$). For more general DGPs, where analytical correction is impossible, the linear GMM estimators of (Juodis & Sarafidis, 2021, 2022) are more suitable alternative to the bias-corrected (r)CCEP estimator. See Juodis and Sarafidis (2018) for a more comprehensive review of the GMM procedures (for fixed T panels).

5 | SIMULATION STUDY

In this section, simulation experiments are carried out to assess the finite sample performance of the proposed regularized CCE procedure. The details of the DGP are provided in Section 5.1, followed by the summary of the results for the homogeneous and the heterogeneous coefficients setups.

5.1 | The setup

We restrict our attention to a simplified DGP of Section 2 with $K = 1$ and two factors:

$$\boldsymbol{y}_i = \boldsymbol{x}_i\beta_i + \boldsymbol{f}\lambda_i + \boldsymbol{\varepsilon}_i, \tag{39}$$

$$\boldsymbol{x}_i = \boldsymbol{f}_{\perp}\gamma_{i,\perp} + \boldsymbol{f}\lambda_i + \boldsymbol{v}_i. \tag{40}$$

We focus on the situations where the vector $\boldsymbol{f}$ can be always approximated by the cross-section averages $\overline{\boldsymbol{Z}} = (\bar{\boldsymbol{y}}, \bar{\boldsymbol{x}})$ (i.e., $K_z = 2$), while this is not always the case for $\boldsymbol{f}_{\perp}$.[16] To be specific, we assume that:[17]

$$\begin{pmatrix} \lambda_i \\ \gamma_{i,\perp} \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ \gamma_{\perp} \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right). \tag{41}$$

As $\boldsymbol{f}_{\perp}$ does not enter the equation for $\boldsymbol{y}_i$ directly, $\text{rk}[E[\boldsymbol{C}_i]] = 1 + \text{rk}[\gamma_{\perp}]$. Thus, for $\gamma_{\perp} = 0$ we consider the "more observables than factors" setup, while for $\gamma_{\perp} = 1$ we have as many observables as the underlying (identifiable) factors.[18]

The common factors $\boldsymbol{d}_t = (f_t, f_{t,\perp})'$ are drawn from the multivariate normal distribution with an identity variance–covariance matrix and $\Gamma_{\boldsymbol{d}}(h) = \mathbf{O}_{2\times 2}$ for all $h \neq 0$. The DGP for $\boldsymbol{e}_{i,t} = (\varepsilon_{i,t}, v_{i,t})'$, has an identical second moment structure.

Finally, for the homogeneous coefficients setup we set $\beta_i = 0$ for all i. The setup where all coefficients are set to zero, is without loss of generality as all quantities are invariant to non-singular transformations of $\boldsymbol{Z}_i$. In the heterogeneous coefficients setup we follow Pesaran (2006) and set $\beta_i \sim N(0, 0.04)$.[19]

In total, we consider nine combinations of (N, T), where every dimension takes values in the set $\{20, 50, 100\}$. Thus, the total of 18 different simulation designs is considered, as we also vary $\gamma_{\perp} \in \{0, 1\}$. For each design, the number of replications is set to $M = 4000$. Finally, all stochastic quantities are drawn in every replication.

Computational notes. We implement our regularization procedure using the five steps algorithm described in Section 3. To accommodate the possibility that $K_z = R$, we use the dummy column idea from Remark 6. $B = 199$ bootstrap samples (i.e., 200 including the original sample) are used to calculate the rejection frequencies with the 5% nominal level. For all estimators we report summary statistics for their scaled and centered versions—$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. We report the mean bias and the RMSE, and the empirical rejection frequencies for three null hypotheses: $\beta_0 = \{-0.1; 0; 0.1\}$. Here at $\beta_0 = 0$, we

[16] Here in order to shrink the nuisance parameter space we assume that the factor loadings on $\boldsymbol{f}$ in both equations are perfectly correlated.

[17] Note that in Figure A1, the correlation coefficient between the factor loadings is higher, and is set to ≈ 0.625.

[18] In the supporting information we consider a variation of this setup where the factor loadings λ_i and $\gamma_{i,\perp}$ are assumed to be uncorrelated.

[19] Note that this DGP as well as the DGP for factor loadings violates Assumption 4.6. However, that assumption is only sufficient and not necessary, the all conclusions of Theorem 2 are expected to hold for random variables with thin tails.

TABLE 1 Eigenvalue Ratio (ER) based estimates of R

		$\#\hat{R}=1$	
N	T	$\gamma_\perp=0$	$\gamma_\perp=1$
20	20	0.9363	0.0525
20	50	0.9540	0.0198
20	100	0.9578	0.0160
50	20	0.9893	0.0033
50	50	0.9940	0.0000
50	100	0.9945	0.0000
100	20	0.9990	0.0000
100	50	0.9988	0.0000
100	100	0.9985	0.0000

Note. Homogeneous setup. The ER statistic is implemented as in Equation (15) using the normalized cross-section averages, and the dummy-column as in Remark 6. For $\gamma_\perp=0$, the true number of factors is $R=1$, while for $\gamma_\perp=1$ it is $R=2$. $M=4000$.

report the size, while at the other two values we report (non-adjusted) power. We note that for $\gamma_\perp=0$ all pooled estimators are $\sqrt{N}-$consistent, while for $\gamma_\perp=1$ they are $\sqrt{NT}-$consistent. Furthermore, as the regressor is strictly exogenous, the rCCEP estimator is fixed T consistent in this setup.

Remark 12. In the supporting information, we analyze several extensions of the basic Monte Carlo design provided above. In particular, we consider the setups with (a) uncorrelated factors loadings λ_i and $\gamma_{i,\perp}$; (b) heteroscedastic error terms; (c) weakly identified R from the ER criterion. We also compare the properties of the pooled bias-corrected regularized CCE estimator with those of the bias-corrected Interactive FE (IFE) estimator of Bai (2009). The IFE estimator is a popular alternative to the CCE estimator for the setups where $N\approx T$. Overall, we find that the IFE estimator dominates the rCCE approach when the true number of factors is weakly identified, while the rCCE is the preferable approach for the setups with strong heteroscedasticity

5.2 | Results: Pooled estimator

In this section analyse the properties of the following three estimators: the standard CCE ("CCE"), the regularized CCE ("rCCE"), and the bias-corrected regularized CCE ("rCCE-BC"). We use the suggested ER statistic to estimate the number of factors R.

From Table 1 we see that one can precisely estimate the true number of factors (either 1 or 2) already for very limited sample sizes. As the procedure is consistent for T fixed, the length of the individual time-series plays almost no role for the selection precision, while there is a clear benefit of larger N.

The estimation results are provided in Table 2. Below we briefly summarize the main findings.

- The CCE estimator is substantially biased for $\gamma_\perp=0$. This is especially pronounced for smaller values of N. However, even when the bias is negligible, the bootstrap based inference procedure does not control size. Given the non-normality of the asymptotic distribution, this result is expected.
- For $\gamma_\perp=0$ the rCCE estimator is well behaved in finite samples. The magnitude of the $\boldsymbol{b}_{2,r}$ bias-term implied by the DGP is rather negligible. Consequently, the bootstrap based testing procedure is well-sized. The power is non-negligible already for small values of (N,T), even if the corresponding power curves can be asymmetric for those combinations.
- The effect of bias-correction is mostly negligible for $\gamma_\perp=0$. For a few combinations of (N,T) the analytical bias-correction shifts the bootstrap distribution leading to an oversized test. In terms of the RMSE, the bias-correction is marginally beneficial. These results are mostly driven by the slow $\sqrt{N}$ convergence rate of the rCCE estimator, where bias terms are expected to be of lower order than the variance.
- For $\gamma_\perp=1$, the CCE estimator is consistent and asymptotically normal. Moreover, because of the estimation uncertainty associated with $\hat{R}$, for small values of N it dominates the rCCE estimator. However, already for $N=50$, any discrep-

ancies between the two estimators quickly disappear. Bootstrap inference is substantially distorted (for the estimator without bias-correction), due to the non-negligible bias term $\boldsymbol{b}_2 = \boldsymbol{b}_{2,r}$. Size distortions generally increase in T/N.

- For $\gamma_\perp = 1$ bias-correction is instrumental for size-correct inference. At the expense of a marginal increase in the variance, the bias-corrected estimator based bootstrap confidence intervals provide rejection rates close to the nominal 5% level. At the same time, as can be expected, bias-correction symmetrizes the power curves around the true value.

5.3 | Results: Heterogeneous estimator

In this section, we analyse the properties of the CCE-MG and the rCCE-MG estimators in the heterogeneous coefficients setup. As before, we estimate the number of factors using the ER statistic, see Table 3 for the corresponding results. As can be expected, the performance of the ER statistic is only marginally affected by the additional variation in β_i.

Below we briefly discuss the estimation results from Table 4.

- For $\gamma_\perp = 0$ the CCE-MG estimator is dominated by the rCCE-MG both in terms of the bias and the RMSE. Furthermore, with regularization the rejection frequencies under the null hypothesis are much closer to the nominal 5% level, irrespective of N and T values. This is mostly explained by the near-unbiasedness of the rCCE-MG estimator. The discrepancy between the two estimators is especially pronounced when $T \gg N$, as in those cases the presence of $\boldsymbol{\xi}$ (or $\boldsymbol{\xi}_r$) has a non-negligible effect on the asymptotic properties.
- For $\gamma_\perp = 1$ both estimation procedures have almost identical statistical properties. Both MG procedures are substantially biased and badly sized for $N = 20$.The situation improves substantially only when both N and T increase, highlighting the need for larger datasets in order to minimize the estimation uncertainty associated with unit-specific estimation of the regression coefficients.

Overall, we document that gains from regularization (both in terms of the estimation precision, and in terms of better sized inference procedures) greatly outweigh any uncertainty resulting from the need to estimate the number of factors. This is especially true for the pooled estimator.

6 | EMPIRICAL ILLUSTRATION

In this section, we illustrate the suggested regularization procedure using the empirical dataset of Voigtländer (2014). It was recently used by Yin et al. (2021) to illustrate the CCE approach in the context of model averaging. The dataset is available at the *Journal of Business and Economic Statistics* website, and is a slightly adjusted (to ensure balancedness) version of the original data of Voigtländer (2014) spanning 313 sectors of the US economy over the period of 1958 – 2005 ($T = 48$).

6.1 | The model

In what follows we use the empirical strategy of Yin et al. (2021) to investigate the potential causes of the historically increasing wage inequality between high-skilled and low-skilled workers in the US manufacturing industries. The empirical specification they consider is a linear regression model of the form:

$$\ln\left(\frac{w_{L.i,t}}{w_{H.i,t}}\right) = \alpha_i + \beta_{1i}\ln(\sigma_{i,t}) + \beta_{2i}\ln\left(\frac{H_{i,t}}{L_{i,t}}\right) + \boldsymbol{\theta}_i'\boldsymbol{z}_{i,t} + \boldsymbol{\lambda}_i'\boldsymbol{f}_t + \varepsilon_{i,t}, \tag{42}$$

where the variable $\frac{w_{L.i,t}}{w_{H.i,t}}$ is the relative wage of low-skilled workers to high skilled workers. The regressors of interest are (a) $\sigma_{i,t}$, the input skill intensity measure; (b) $\frac{H_{i,t}}{L_{i,t}}$, the ratio of high and low skilled workers in the sector i. $\sigma_{i,t}$ is included to proxy the effects of inter-sectoral technology skill complementarity (ITSC) and is constructed by Voigtländer (2014) using the weighted average of white-collar workers in other industries than i, with weights calculated using the Input-Output expenditures.

The remaining variables captured in $\boldsymbol{z}_{i,t}$ are control variables, for example, the capital equipment per worker and R&D intensity, among others. Among other models, Yin et al. (2021) estimated the "full model" in Equation (42), as well as

TABLE 2 Estimation and inference

Design			*CCE*					*rCCE*					*rCCE − BC*				
$\gamma_\perp$	*N*	*T*	Bias	RMSE	Power−	Size	Power+	Bias	RMSE	Power−	Size	Power+	Bias	RMSE	Power−	Size	Power+
0	20	20	0.10	0.25	0.82	0.12	0.10	0.03	0.24	0.55	0.07	0.32	0.02	0.23	0.58	0.05	0.31
0	20	50	0.10	0.20	0.99	0.19	0.16	0.02	0.18	0.74	0.06	0.43	0.02	0.16	0.78	0.04	0.44
0	20	100	0.10	0.18	1	0.24	0.18	0.02	0.17	0.81	0.06	0.47	0.01	0.14	0.28	0.10	0.52
0	50	20	0.06	0.22	0.98	0.08	0.58	0.01	0.20	0.94	0.05	0.80	0.01	0.19	0.83	0.06	0.85
0	50	50	0.06	0.16	1	0.13	0.74	0.01	0.15	0.99	0.05	0.89	0.01	0.14	0.96	0.06	0.91
0	50	100	0.06	0.14	1	0.16	0.78	0.01	0.12	1	0.05	0.90	0.01	0.11	0.98	0.07	0.92
0	100	20	0.04	0.21	1	0.06	0.96	0.00	0.19	1	0.06	0.99	0.00	0.19	1	0.07	0.99
0	100	50	0.04	0.15	1	0.09	0.99	0.00	0.14	1	0.05	0.99	0.00	0.14	1	0.06	1
0	100	100	0.04	0.13	1	0.10	1	0.01	0.12	1	0.05	1	0.01	0.11	1	0.07	1
1	20	20	0.12	0.29	0.72	0.12	0.11	0.11	0.33	0.69	0.13	0.12	0.02	0.31	0.56	0.07	0.15
1	20	50	0.12	0.21	0.97	0.18	0.17	0.12	0.25	0.95	0.18	0.18	0.02	0.22	0.90	0.09	0.24
1	20	100	0.11	0.18	1	0.22	0.23	0.11	0.21	0.99	0.22	0.23	0.01	0.16	0.98	0.09	0.32
1	50	20	0.07	0.25	0.94	0.10	0.54	0.07	0.25	0.93	0.10	0.54	0.00	0.25	0.89	0.07	0.64
1	50	50	0.07	0.17	1	0.15	0.80	0.07	0.17	1	0.15	0.80	0.00	0.15	1	0.08	0.92
1	50	100	0.07	0.13	1	0.19	0.89	0.07	0.13	1	0.19	0.89	0.00	0.10	1	0.08	0.97
1	100	20	0.05	0.25	1	0.08	0.95	0.05	0.25	1	0.08	0.95	0.00	0.24	0.99	0.06	0.97
1	100	50	0.05	0.15	1	0.10	1	0.05	0.15	1	0.10	1	0.00	0.15	1	0.07	1
1	100	100	0.05	0.11	1	0.15	1	0.05	0.11	1	0.15	1	0.00	0.10	1	0.08	1

Note: Homogeneous setup. Here “Bias” is the average of any scaled estimator over $M = 4000$ replications. “RMSE” is the corresponding root-mean-square error of any scaled estimator. *CCE* is the pooled CCE estimator; *rCCE* is the pooled regularized CCE estimator; *rCCE − BC* is the *rCCE* estimator with analytical bias correction as in the supporting information. “Power−” and “Power+” are respectively the rejection frequencies for the null hypothesis of $\beta_0 = -0.1$ and $\beta_0 = 0.1$. “Size” corresponds to the null hypothesis $\beta_0 = 0$.

TABLE 3 Eigenvalue Ratio based estimates of R

		$\#\hat{R}=1$	
N	T	$\gamma_\perp=0$	$\gamma_\perp=1$
20	20	0.9265	0.0725
20	50	0.9593	0.0280
20	100	0.9620	0.0255
50	20	0.9880	0.0050
50	50	0.9925	0.0000
50	100	0.9948	0.0003
100	20	0.9985	0.0000
100	50	0.9985	0.0000
100	100	0.9988	0.0000

Note: Heterogeneous setup. See Table 1 for more details.

TABLE 4 Estimation and inference

Design			$CCE-MG$					$rCCE-MG$				
$\gamma_\perp$	N	T	Bias	RMSE	Power−	Size	Power+	Bias	RMSE	Power−	Size	Power+
0	20	20	0.10	0.33	0.53	0.09	0.08	0.02	0.31	0.32	0.06	0.22
0	20	50	0.10	0.28	0.72	0.14	0.10	0.02	0.27	0.39	0.07	0.29
0	20	100	0.10	0.26	0.78	0.16	0.12	0.02	0.25	0.42	0.07	0.33
0	50	20	0.06	0.31	0.79	0.07	0.37	0.01	0.30	0.67	0.06	0.58
0	50	50	0.06	0.26	0.93	0.09	0.52	0.01	0.25	0.81	0.06	0.71
0	50	100	0.06	0.24	0.95	0.10	0.57	0.01	0.23	0.85	0.06	0.76
0	100	20	0.03	0.30	0.95	0.06	0.81	0.00	0.29	0.93	0.06	0.90
0	100	50	0.04	0.25	0.99	0.07	0.93	0.00	0.25	0.98	0.06	0.97
0	100	100	0.04	0.23	1	0.08	0.96	0.00	0.23	0.99	0.06	0.98
1	20	20	0.11	0.36	0.48	0.10	0.08	0.11	0.38	0.46	0.10	0.09
1	20	50	0.12	0.29	0.69	0.13	0.11	0.11	0.30	0.68	0.13	0.12
1	20	100	0.11	0.27	0.77	0.15	0.13	0.11	0.28	0.76	0.15	0.14
1	50	20	0.08	0.33	0.74	0.08	0.36	0.08	0.34	0.74	0.08	0.36
1	50	50	0.08	0.26	0.91	0.10	0.51	0.08	0.26	0.91	0.10	0.51
1	50	100	0.07	0.24	0.95	0.11	0.59	0.07	0.24	0.95	0.11	0.59
1	100	20	0.05	0.33	0.92	0.07	0.77	0.05	0.33	0.92	0.07	0.77
1	100	50	0.05	0.25	0.99	0.08	0.93	0.05	0.25	0.99	0.08	0.93
1	100	100	0.05	0.23	1	0.09	0.97	0.05	0.23	1	0.09	0.97

Note: Heterogeneous setup. $CCE-MG$ is the mean-group CCE estimator; $rCCE-MG$ is the mean-group regularized CCE estimator. See Table 2 for more details.

the "narrow model" without any controls using the CCE-MG estimator. Motivated by their setup we estimate these two specifications also using the regularized versions of the CCEP and the CCE-MG estimators.[20] We use demeaned data (over the time dimension) to filter out the time invariant unit specific effects, α_i from Equation (42).

6.2 | The number of factors

At first, we provide estimates the number of factors for the cross-section averages used in both the narrow and the full models.

Our benchmark approach uses the ER criterion with normalized factors augmented by the dummy column, labeled as $\hat{\boldsymbol{F}}+$. As a robustness check, we also include results without the dummy column ($\hat{\boldsymbol{F}}$), as well as non-normalized cross-section average $\overline{\boldsymbol{Z}}$. From Table 5 we observe strong evidence for a model with a single factor, that is, $R=1$. Thus, our procedure that the number of the underlying factors that drive the cross-section averages is small in comparison with the number of cross-section averages included. The only deviation occurs for simple cross-section averages $\overline{\boldsymbol{Z}}$, where the

[20]Note that Yin et al. (2021) do not motivate the use of the heterogeneous coefficients setup in this empirical specification. The original studies of Ciccone and Peri (2005) and Voigtländer (2014), on the other hand, use pooled estimators.

TABLE 5 The number of factors estimates for the full and narrow models

	Full ($K_z = 9$)			Narrow ($K_z = 3$)		
Criterion	$\hat{F}+$	$\hat{F}$	$\bar{Z}$	$\hat{F}+$	$\hat{F}$	$\bar{Z}$
ER	1	1	1	1	1	2*
GR	2	2	2	1	—	—

Note: Here $\hat{\boldsymbol{F}}+$ denotes normalized factor proxies with an additional dummy column included; $\hat{\boldsymbol{F}}$ normalized factor proxies; $\bar{\boldsymbol{Z}}$ cross-section averages without any normalizations imposed. "ER" and "GR" are correspondingly the Eigenvalue Ratio and the Growth Ratio criteria of Ahn and Horenstein (2013). $*$ indicates that this is the maximum number of factors allowed by this criterion; – indicates that for these entries criterion automatically predicts only one value 1.

TABLE 6 Estimation results using the mean-group CCE

	Narrow		Full	
	$rCCE-MG$	$CCE-MG$	$rCCE-MG$	$CCE-MG$
$\ln(\sigma_{i,t})$	−0.54	−0.59	−0.99	−0.71
	(−1.10;0.07)	(−1.01;−0.14)	(−1.40;-0.60)	(−1.20;-0.26)
$\ln\left(\frac{H_{i,t}}{L_{i,t}}\right)$	0.35	0.35	0.42	0.45
	(0.33;0.38)	(0.32;0.38)	(0.39;0.44)	(0.42;0.49)
$k_{i,t}^{equip}$			−0.78	−1.64
			(−1.49;0.53)	(−2.87;−0.42)
$(OCAM/K)_{i,t}$			−1.87	−3.29
			(−3.24;−0.42)	(−7.13;0.81)
$(HT/K-OCAM/K)_{i,t}$			2.20	2.28
			(0.68;3.55)	(0.02;4.90)
$RD_{lag.i,t}$			1.11	1.44
			(−0.46;2.76)	(−0.37;3.05)
$OS_{i,t}^{narr}$			−1.26	−1.20
			(−2.95;−0.18)	(−2.41;−0.15)
$(OS^{broad}-OS^{narr})_{i,t}$			0.18	0.20
			(−0.36;0.77)	(−0.74;0.80)

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

ER criterion selects a marginally larger number of factors, that is, $R = 2$. However, as we argued above, we should prefer factor selection procedures based on $\hat{\boldsymbol{F}}$ as they are invariant to non-singular transformations to $\boldsymbol{Z}_i$.

Besides the eigenvalue ratio criterion, we also report the Growth Ratio (GR) criterion of Ahn and Horenstein (2013). While we do not discuss GR criterion in our Monte Carlo Study, Juodis and Sarafidis (2022) provide some evidence for superior properties of this approach if the number of proxies is substantially larger than the number of the underlying factors. In contrast to ER based estimates of R, the GR approach suggests that $R = 2$ in the "Full" model. [21] In the supporting information we also report rCCEP and rCCE-MG estimates for $R = 2$.

6.3 | Estimation results

At first, in Table 6 we discuss estimation results for the Mean Group Estimators. This way we can easily compare our results with those of Yin et al. (2021). The results for the pooled estimators are presented in Table 7.

Mean group estimation. From Table 6 we see that regularization has no major effect on the estimation results in the "Narrow" model. This is not very surprising, as the single regularized factor proxy $\hat{\boldsymbol{F}}_r$ is extracted from 3 cross-section aver-

[21] As we use only one dummy column in $\hat{F}+$, the GR approach excludes the possibility that $R = K_z$.

TABLE 7 Estimation results based on the Pooled CCE

	Narrow			Full		
	rCCE	*rCCE*$_{BC}$	*CCE*	*rCCE*	*rCCE*$_{BC}$	*CCE*
$\ln(\sigma_{i,t})$	−0.62	−0.63	−0.61	−0.84	−0.84	−0.52
	(−1.03;−0.21)	(−1.04;-0.23)	(−1.01;−0.22)	(−1.21;−0.49)	(−1.21;−0.49)	(−0.93;-0.19)
$\ln\left(\frac{H_{i,t}}{L_{i,t}}\right)$	0.36	0.36	0.40	0.37	0.37	0.48
	(0.29;0.41)	(0.29;0.41)	(0.33;0.47)	(0.30;0.43)	(0.30;0.43)	(0.41;0.54)
$k_{i,t}^{equip}$				−0.06	−0.06	−0.19
				(−0.31;0.22)	(−0.31;0.22)	(−0.42;0.09)
$(OCAM/K)_{i,t}$				−0.36	−0.36	1.23
				(−0.81;0.18)	(−0.81;0.18)	(−0.58;2.83)
$(HT/K - OCAM/K)_{i,t}$				0.86	0.86	0.25
				(0.17;1.49)	(0.17;1.49)	(−0.38;1.16)
$RD_{lag.i,t}$				−0.25	−0.24	0.20
				(−0.71;0.26)	(−0.70;0.26)	(−0.36;0.71)
$OS_{i,t}^{narr}$				−0.14	−0.14	−0.08
				(−0.29;−0.01)	(−0.29;−0.01)	(−0.24;0.06)
$(OS^{broad} - OS^{narr})_{i,t}$				−0.13	−0.13	−0.09
				(−0.28;0.04)	(−0.28;0.04)	(−0.28;0.10)

Note: See Table 6 for a detailed description of the regressors.

ages, where each individual entry follows a similar trending pattern over time (see figures in the supporting information). The only real change is the inclusion of 0 in the confidence interval for $\ln(\sigma_{i,t})$ for the regularized estimator.

For the "Full" model the results for regularized and non-regularized versions of the mean-group estimators are also comparable for the main regressor of interest $\ln(\frac{H_{i,t}}{L_{i,t}})$. This provides sufficient evidence to conclude that high-skilled and low-skilled labor are substitutes. On the other hand, the coefficient for $\ln(\sigma_{i,t})$ changes from −0.71 to −0.99, which is a substantial decrease of the average elasticity coefficient. In particular, the standard CCE-MG coefficient is almost outside of the confidence interval of the rCCE-MG estimator. Nevertheless, as for both estimators the coefficient is negative we can at least confirm the results in Yin et al. (2021) that ITSC (summarized by $\ln(\sigma_{i,t})$) substantially increases inequality.

As for control variables, the only two exceptions where the two methods produce different results are $k_{i,t}^{equip}$ and $(OCAM/K)_{i,t}$. The corresponding estimates are substantially smaller (in absolute value) for the regularized approach. Furthermore, for $k_{i,t}^{equip}$ the reported confidence interval includes 0, while this is not the case without regularization. On the other hand, the opposite conclusion can be drawn for $(OCAM/K)_{i,t}$. Also, one can observe substantially narrower confidence intervals for these coefficients after regularization.

Overall, the effect of regularization is mostly visible in (slightly) narrower confidence intervals for all coefficients. This can be partially attributed to the fact that for each sector i the ordinary CCE-MG approach requires estimation of 18 unit specific coefficients, while that number is only 11 after regularization. This is a substantial reduction given that $T = 48$.

Pooled estimators. Moving toward the pooled estimator in Table 7, we observe that for the "Narrow model" the results are almost identical as those obtained using the MG approach. In particular, regularization has no visible effect on the estimated coefficients. Furthermore, bias-correction has no visible effects on coefficient estimates. This conclusion is not too surprising, given that N is substantially larger for this dataset, thus the magnitude of the overall bias is expected to be small.

Comparing the point estimates for the main regressors of interest, we find the most substantial coefficient change is visible for $\ln(\sigma_{i,t})$. This mirrors exactly the situation we observed also for the MG estimator. In particular, the coefficient for $\ln(\sigma_{i,t})$ changes from −0.52 to −0.84 when going from the standard CCE estimator to its regularized version. Moreover, a substantial decrease in the elasticity coefficient for $\ln\left(\frac{H_{i,t}}{L_{i,t}}\right)$ is also noticeable (in contrast to the MG setup), when comparing the results of the CCE and rCCE estimators.

Alternative methods. In the remainder of this section we briefly compare the (pooled) empirical results based on the rCCE estimator with those of the IFE/PC estimator Bai (2009). We follow the suggestion of Petrova and Westerlund (2020) and implement this estimator after double de-meaning the data (i.e., after the two-way fixed effects transformation).[22]

[22]We use the rCCE estimator to initialize the non-linear estimation procedure. As the results for the bias-corrected IFE estimator are nearly identical to those without bias-correction, we report the latter.

TABLE 8 Estimation results based on the IFE Estimator of Bai (2009)

	Narrow		Full	
	$R = 1$	$R = 2$	$R = 1$	$R = 2$
$\ln(\sigma_{i,t})$	−0.85	−0.65	−0.87	−0.65
	(−1.24;−0.52)	(−1.11;−0.36)	(−1.25;−0.54)	(−1.08;−0.32)
$\ln\left(\frac{H_{i,t}}{L_{i,t}}\right)$	0.35	0.40	0.35	0.40
	(0.29;0.41)	(0.33;0.47)	(0.29;0.41)	(0.33;0.47)
$k^{equip}_{i,t}$			−0.17	−0.07
			(−0.49;0.13)	(−0.49;0.21)
$(OCAM/K)_{i,t}$			−0.06	0.48
			(−0.74;0.63)	(−0.53;1.30)
$(HT/K - OCAM/K)_{i,t}$			0.56	0.21
			(−0.20;1.26)	(−0.36;1.07)
$RD_{lag.i,t}$			−0.10	−0.17
			(−0.65;0.42)	(−0.61;0.23)
$OS^{narr}_{i,t}$			−0.07	−0.06
			(−0.24;0.09)	(−0.22;0.07)
$(OS^{broad} - OS^{narr})_{i,t}$			−0.08	−0.08
			(−0.27;0.09)	(−0.28;0.14)

Note: The iterative optimization procedure is initialized using the rCCEP estimator. All variables are double de-meaned prior to estimation as in Petrova and Westerlund (2020). See Table 6 for a detailed description of the regressors.

Estimation results for the model with one and two factors (i.e., $R = 1$ and $R = 2$) are summarized in Table 8.[23] The optimal number of factors estimated using the Eigenvalue Ratio (ER) criterion of Ahn and Horenstein (2013), as well as using the alternative implementation suggested by Chen et al. (2021), is always 1. In this regard, the results are comparable with those in Table 5. Overall, the estimated coefficients both in the "Narrow" model as well as the "Full" model are either directly comparable in their magnitude to those of the rCCE estimator, or (at least) fall within the corresponding bootstrap CI of the rCCE estimator.

7 | CONCLUDING REMARKS

In this paper, we develop a simple method to estimate factor-augmented regressions using regularized cross-section averages of the observed data. The novelty of our approach is the regularization step that uses normalized cross-section averages of the data. This step is crucial to ensure that the proposed estimator is invariant to non-singular transformations of the data, while the regularization step ensures that all factor estimates are asymptotically well-behaved.

The proposed procedure is intuitive to use and is easy to implement numerically by practitioners using any software that has in built functions for the singular value decomposition, for example, STATA, Eviews, Python, or R. Thus we suggest that the proposed procedure to be (at least) used as a part of the sensitivity analysis for any model estimated using the CCE approach.

The resulting regularized Common Correlated Effects (rCCE) estimator shares most of the advantages of the original CCE estimator of Pesaran (2006). At the same time, regularization safeguards against the potential problems when there are more observables than the underlying factors. The applicability of pooled and mean-group versions of the regularized estimator is illustrated using the dataset of Voigtländer (2014) and Yin et al. (2021).

There are several open problems that remain to be investigated, that we only briefly mention in this paper. In particular, while we partially motivated this paper by non-linear (generalized linear) cross-section augmented-models as in Boneva and Linton (2017) or Desbordes and Eberhardt (2019), any rigorous results for those models are yet to be formalized. We leave these, and other related questions for future research.

ACKNOWLEDGEMENTS

I would like to thank the Co-Editor Edward Vytlacil and three anonymous referees for numerous suggestions that greatly improved this paper. I would also like to thank Tom Boot, Peter Egger, Frank Kleibergen, and Vasilis Sarafidis for con-

[23]Estimation results for $R = 3, \dots, 5$ are comparable to those of $R = 2$, thus omitted.

structive comments and suggestions. A substantial part of this paper was written while I was affiliated with the University of Groningen. Financial support from the Netherlands Organisation for Scientific Research (NWO) under research grant number 451-17-002 is gratefully acknowledged.

DATA AVAILABILITY STATEMENT

The dataset used in the empirical section of this paper is available in the JAE Data Archive http://qed.econ.queensu.ca/jae/datasets/juodis001/.

OPEN RESEARCH BADGES

This article has been awarded Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. Data is available at http://qed.econ.queensu.ca/jae/datasets/juodis001/.

REFERENCES

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

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, *71*(1), 135–171. https://doi.org/10.1111/1468-0262.00392

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, *77*(4), 1229–1279. https://doi.org/10.3982/ECTA6135

Bai, J., & Li, K. (2014). Theory and methods of panel data models with interactive effects. *Annals of Statistics*, *42*(1), 142–170. https://doi.org/10.1214/13-AOS1183

Boneva, L., & Linton, O. (2017). A discrete choice model for large heterogeneous panels with interactive fixed effects with an application to the determinants of corporate bond issuance. *Journal of Applied Econometrics*, *32*(7), 1226–1243. https://doi.org/10.1002/jae.2438

Breitung, J., & Hansen, P. (2021). Alternative estimation approaches for the factor augmented panel data model with small t. *Empirical Economics*, *60*(1), 327–351.

Cesa-Bianchi, A., Pesaran, M. H., & Rebucci, A. (2020). Uncertainty and economic activity: A multicountry perspective. *The Review of Financial Studies*, *33*(8), 3393–3445. https://doi.org/10.1093/rfs/hhz098

Chen, M., Fernández-Val, I., & Weidner, M. (2021). Nonlinear factor models for network and panel data. *Journal of Econometrics*, *220*(2), 296–324.

Chudik, A., & 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.

Chudik, A., Pesaran, M. H., & Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal*, *14*(1), 45–90. https://doi.org/10.1111/j.1368-423X.2010.00330.x

Ciccone, A., & Peri, G. (2005). Long-run substitutability between more and less educated workers: Evidence from U.S. states, 1950–1990. *The Review of Economics and Statistics*, *87*(4), 652–663. https://doi.org/10.1162/003465305775098233

Cruz-Gonzalez, M., Fernández-Val, I., & Weidner, M. (2017). Bias corrections for probit and logit models with two-way fixed effects. *The Stata Journal*, *17*(3), 517–545.

De Vos, I., & Stauskas, O. (2021). Bootstrap improved inference for factor-augmented regressions with cce. mimeo.

Desbordes, R., & Eberhardt, M. (2019). Gravity. GEP WP 2019/02.

Dhaene, G., & Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies*, *82*(3), 991–1030.

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

Everaert, G., & Pozzi, L. (2014). The predictability of aggregate consumption growth in oecd countries: A panel data analysis. *Journal of Applied Econometrics*, *29*(3), 431–453. https://doi.org/10.1002/jae.2338

Fernández-Val, I., & Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, *192*(1), 291–312.

Galvao, A. F., & Kato, K. (2014). Estimation and inference for linear panel data models under misspecification when both n and t are large. *Journal of Business & Economic Statistics*, *32*(2), 285–309. https://doi.org/10.1080/07350015.2013.875473

Goncalves, S., & Perron, B. (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics*, *182*(1), 156–173.

Hacioglu Hoke, S., & Kapetanios, G. (2021). Common correlated effect cross-sectional dependence corrections for nonlinear conditional mean panel models. *Journal of Applied Econometrics*, *36*(1), 125–150. https://doi.org/10.1002/jae.2799

Hall, P., & Heyde, C. C. (1980). *Martingale limit theory and its application*, Probability and Mathematical Statistics: Academic Press.

Harding, M., & Lamarche, C. (2014). Estimating and testing a quantile regression model with interactive effects. *Journal of Econometrics*, *178, Part 1*, 101–113. Annals Issue: Misspecification Test Methods in Econometrics.

Harding, M., Lamarche, C., & Pesaran, M. H. (2020). Common correlated effects estimation of heterogeneous dynamic panel quantile regression models. *Journal of Applied Econometrics*, *35*(3), 294–314. https://doi.org/10.1002/jae.2753

Juodis, A. (2020). This shock is different: Estimation and inference in misspecified two-way fixed effects regressions. mimeo.

Juodis, A., Karabıyık, H., & Westerlund, J. (2021). On the robustness of the pooled cce estimator. *Journal of Econometrics*, *220*(2), 325–348.

Juodis, A., & Kučinskas, S. (2019). Quantifying noise. mimeo.

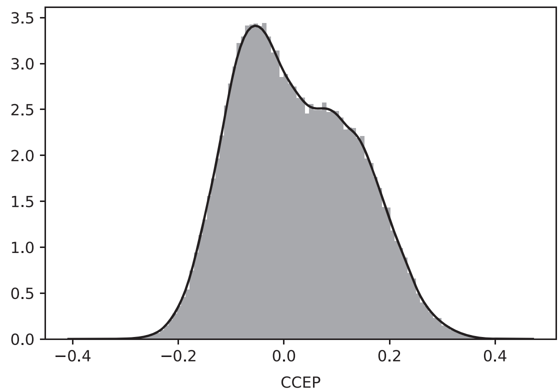
Juodis, A., & Reese, S. (2021). The incidental parameters problem in testing for remaining cross-section correlation. *Journal of Business and Economic Statistics*, (forthcoming).
Juodis, A., & Sarafidis, V. (2018). Fixed t dynamic panel data estimators with multi-factor errors. *Econometric Reviews*, *37*(8), 893–929.
Juodis, A., & Sarafidis, V. (2022). A linear estimator for factor-augmented fixed-t panels with endogenous regressors. *Journal of Business and Economic Statistics*, *40*(1), 1–15. https://doi.org/10.1080/07350015.2020.1766469
Juodis, A., & Sarafidis, V. (2021). An incidental parameters free inference approach for panels with common factors. *Journal of Econometrics*, (forthcoming).
Juodis, A., & Westerlund, J. (2019). Optimal panel unit root testing with covariates. *The Econometrics Journal*, *22*(1), 55–72.
Kapetanios, G., Pesaran, M. H., & Yamagata, T. (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics*, *160*(2), 326–348.
Karabıyık, H., Reese, S., & 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.
Karabıyık, H., Urbain, J.-P., & Westerlund, J. (2019). Cce estimation of factor-augmented regression models with more factors than observables. *Journal of Applied Econometrics*, *34*(2), 268–284. https://doi.org/10.1002/jae.2661
Kleibergen, F. R., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*.
Li, K., Cui, G., & Lu, L. (2020). Efficient estimation of heterogeneous coefficients in panel data models with common shocks. *Journal of Econometrics*, *216*(2), 327–353.
Moon, H. R., & Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, *33*(1), 158–195. https://doi.org/10.3982/ECTA9382
Norkutė, M., Sarafidis, V., Yamagata, T., & Cui, G. (2021). Instrumental variable estimation of dynamic linear panel data models with defactored regressors and a multifactor error structure. *Journal of Econometrics*, *220*(2), 416–446.
Norkutė, M., & Westerlund, J. (2021). The factor analytical approach in near unit root interactive effects panels. *Journal of Econometrics*, *221*(2), 569–590.
Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, *74*(4), 967–1012. https://doi.org/10.1111/j.1468-0262.2006.00692.x
Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, *22*(2), 265–312. https://doi.org/10.1002/jae.951
Pesaran, M. H., Smith, L. V., & Yamagata, T. (2013). Panel unit root tests in the presence of a multifactor error structure. *Journal of Econometrics*, *175*(2), 94–115.
Petrova, Y., & Westerlund, J. (2020). Fixed effects demeaning in the presence of interactive effects in treatment effects regressions and elsewhere. *Journal of Applied Econometrics*, *35*(7), 960–964.
Reese, S., & Westerlund, J. (2016). Panicca: Panic on cross-section averages. *Journal of Applied Econometrics*, *31*(6), 961–981. https://doi.org/10.1002/jae.2487
Robertson, D., & Sarafidis, V. (2015). Iv estimation of panels with factor residuals. *Journal of Econometrics*, *185*(2), 526–541.
Robin, J. M., & Smith, R. J. (2000). Tests of rank. *Econometric Theory*, *16*, 151–175.
Sarafidis, V., & Wansbeek, T. (2021). Celebrating 40 years of panel data analysis: Past, present and future. *Journal of Econometrics*, *220*(2), 215–226. Annals Issue: Celebrating 40 Years of Panel Data Analysis: Past, Present and Future.
Su, L., & Jin, S. (2012). Sieve estimation of panel data models with cross-section dependence. *Journal of Econometrics*, *169*(1), 34–47.
Voigtländer, N. (2014). Skill bias magnified: Intersectoral linkages and white-collar labor demand in u.s. manufacturing. *The Review of Economics and Statistics*, *96*(3), 495–513. https://doi.org/10.1162/REST_a_00390
Westerlund, J. (2018). Cce in panels with general unknown factors. *The Econometrics Journal*, *21*(3), 264–276. https://doi.org/10.1111/ectj.12110
Westerlund, J., Petrova, Y., & Norkutė, M. (2019). Cce in fixed-t panels. *Journal of Applied Econometrics*, *34*(5), 746–761. https://doi.org/10.1002/jae.2707
Westerlund, J., & Urbain, J.-P. (2013). On the estimation and inference in factor-augmented panel regressions with correlated loadings. *Economics Letters*, *119*(3), 247–250.
Westerlund, J., & Urbain, J.-P. (2015). Cross-sectional averages versus principal components. *Journal of Econometrics*, *185*(2), 372–377.
Yin, S.-Y., Liu, C.-A., & Lin, C.-C. (2021). Focused information criterion and model averaging for large panels with a multifactor error structure. *Journal of Business & Economic Statistics*, *39*(1), 54–68. https://doi.org/10.1080/07350015.2019.1623044

SUPPORTING INFORMATION

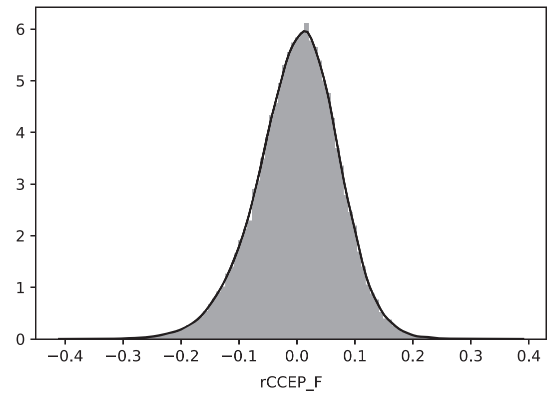
Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Juodis A. (2022). A regularization approach to common correlated effects estimation. *Journal of Applied Econometrics*, *37*(4), 788–810. https://doi.org/10.1002/jae.2899

APPENDIX A



(a) Standard CCEP.



(b) Regularized (feasible) CCEP.

FIGURE A1 Finite sample distribution of the scaled and centered common correlated effects pooled (CCEP) and rCCEP estimators using $B = 10^5$ Monte Carlo replications. $N = T = 200$. For more details regarding the DGP, see Section 5