Bias-Corrected Common Correlated Effects Pooled Estimation in Dynamic Panels

De Vos, Ignace; Everaert, Gerdie

Published in:
Journal of Business & Economic Statistics

DOI:
10.1080/07350015.2019.1654879

2021

Document Version:
Publisher's PDF, also known as Version of record

Supplementary materials for this article are available online. Please go to http://tandfonline.com/r/JBES

OPEN ACCESS

Check for updates

Bias-Corrected Common Correlated Effects Pooled Estimation in Dynamic Panels

Ignace De Vos

Department of Economics, Lund University, Tycho Brahes väg 1, P.O. Box 7082, S-220 07, Department of Economics, Ghent University, Sint-Pietersplein 6, 9000 Ghent, Belgium (*ignace.de_vos@nek.lu.se*)

Gerdie Everaert

Department of Economics, Ghent University, Sint-Pietersplein 6, 9000 Ghent, Belgium (*gerdie.everaert@ugent.be*)

This article extends the common correlated effects pooled (CCEP) estimator to homogenous dynamic panels. In this setting, CCEP suffers from a large bias when the time span (T) of the dataset is fixed. We develop a bias-corrected CCEP estimator that is consistent as the number of cross-sectional units (N) tends to infinity, for T fixed or growing large, provided that the specification is augmented with a sufficient number of cross-sectional averages, and lags thereof. Monte Carlo experiments show that the correction offers strong improvements in terms of bias and variance. We apply our approach to estimate the dynamic impact of temperature shocks on aggregate output growth.

KEY WORDS: Common correlated effects; Dynamic panel bias; Factor augmented regression; Multifactor error structure.

1. INTRODUCTION

Error cross-sectional dependence is one of the major themes in recent panel data econometrics. It is well documented that neglecting such dependencies can distort inference or even lead to inconsistent estimates (see Andrews 2005; Sarafidis and Robertson 2009; Sarafidis and Wansbeek 2012 for details). One of the leading approaches to model cross-sectional dependence is by assuming a multifactor error structure, in which cross-section units are simultaneously influenced by a limited number of unobserved common factors, to which they can respond with different intensities. The common factors may reflect business cycle fluctuations, technological progress, risk and liquidity premia or other global trends and shocks that affect all cross-sectional units in the panel with a potentially differential impact across units arising from differences in institutions, absorptive capacity, technological rigidities, innate ability, preferences, risk aversion, social background, etc. (see, e.g., Ahn, Lee, and Schmidt 2001; Moon and Perron 2007; Eberhardt and Teal 2011; Sarafidis and Wansbeek 2012). Not accounting for unobserved global variables or shocks results in inconsistent estimates when the omitted factors are correlated with the included regressors.

A popular estimation technique for panel data models with a multifactor error structure is the common correlated effects (CCE) estimator introduced by Pesaran (2006). This consists of augmenting the model with the cross-sectional averages (CSA) of the observed variables such that asymptotically—as the cross-sectional dimension $N \to \infty$—the effect of the common factors is eliminated. Both a mean group and a pooled version are suggested, depending on whether the slope coefficients are assumed to be heterogeneous (variable) or homogenous (constant) over cross-sectional units. Under the more general assumption of slope heterogeneity, the mean group (CCEMG) estimator is calculated as the average of the individual CCE slope coefficient estimates. The pooled (CCEP) estimator yields efficiency gains when the slope coefficients are homogenous over cross-sectional units. Under the appropriate set of assumptions, the CCEMG and the CCEP estimators are consistent as $N \to \infty$ for either the time series dimension T fixed or $T \to \infty$. Building on the results in Pesaran (2006), the CCE approach is shown to be robust to various generalizations (see, e.g., Chudik, Pesaran, and Tosetti 2011; Harding and Lamarche 2011; Kapetanios, Pesaran, and Yamagata 2011; Pesaran and Tosetti 2011). The computational straightforwardness of the CCE approach in combination with its robustness has led to numerous applications in many areas of economics and beyond.

The CCE methodology is well developed in the static model but was originally not intended for use in dynamic settings. Dynamic models are, however, common in practice since many (economic) variables tend to react slowly to changes in their determinants and hence display considerable persistence over time. Typically a lagged dependent variable is added to the empirical specification to account for these dynamics. However, this has important consequences for the properties of the CCE estimators. A first complication arises in the approximation

Journal of Business & Economic Statistics
XXXX, 2019, Vol. 00, No. 0
DOI: 10.1080/07350015.2019.1654879
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/jbes.

of the common factors. Chudik and Pesaran (2015) showed that the combination of dynamics and coefficient heterogeneity requires that an infinite number of lagged CSA should be added to the model to eliminate the factors. As this is not feasible in finite samples, they suggest to let the number of CSA grow with T. The implications of dynamics for approximating the common factors in models with homogenous slope coefficients have not yet been studied. Second, Everaert and De Groote (2016) show that in a dynamic setting the CCEP estimator is inconsistent as $N \to \infty$ with T fixed, and that its asymptotic bias tends to be much larger than the standard dynamic panel data bias (Nickell 1981) of the FE estimator in the absence of common factors. Especially when persistence is high, the CCEP estimator remains notably biased even for a moderately long time dimension T up to 50 periods. Monte Carlo simulations further show that the small sample properties of the CCEP estimator are not very sensitive to the size of N. Similar results were obtained by Chudik and Pesaran (2015) for the CCEMG estimator. Hence, in dynamic panels it is mainly the time series dimension that should be sufficiently large to allow for reliable CCE estimation and inference. This is problematic especially for estimating micro-level dynamic models where N tends to be large and T is typically (very) small, for instance when estimating dynamic employment equations with firm-level data (see, e.g., Carlsson, Eriksson, and Gottfries 2013; Eriksson and Stadin 2017), but also in macro-level panels, where although T tends to be larger than or similar to N the available time span is in many cases smaller than what is needed to make the bias negligibly small. In an attempt to reduce the small T bias, Chudik and Pesaran (2015) suggested the recursive mean adjustment of So and Shin (1999) or the split-panel jackknife of Dhaene and Jochmans (2015). Although these approaches succeed in mitigating the bias, they are unable to resolve the issue for short-T panels. Despite these two important complications, the CCE approach is increasingly used to estimate dynamic panel data models with common factors in a variety of empirical settings, including—among many others—development economics (Temple and Van de Sijpe 2017); economic growth (Minniti and Venturini 2017); international economics (Wu and Wu 2018); the economics of inequality (Madsena, Minnitib, and Venturini 2018); environmental economics (Tao 2018).

This article considers the CCEP approach to estimate a homogenous dynamic panel data model. We first show that, in contrast to the heterogeneous slope model, only a finite number of lagged CSA are required to eliminate the factors from the error terms. We next remove the finite T bias of the CCEP estimator by deriving a bias-corrected alternative (referred to as CCEPbc) based on large N analytical bias expressions allowing for multiple common factors and exogenous variables. We show that, when correctly specified, the resulting estimator is consistent as $N \to \infty$ with T fixed or $T \to \infty$. Monte Carlo simulations show that CCEPbc provides considerable improvements (both in terms of bias and variance) over the original CCEP estimator and is practically unbiased in all of the considered settings. Moreover, CCEPbc is found to outperform both (i) alternative bias-adjusted CCEP estimators and (ii) the bias-corrected least squares with interactive fixed effects estimator of Moon and Weidner (2017), which is the main alternative to the CCEP methodology in dynamic panels. We further find that a (bootstrap) hypothesis test based on the CCEPbc estimator has an actual size close to the desired nominal level, even when T is small.

The remainder of this article is structured as follows. Section 2 outlines the model and assumptions. In Section 3, we extend the CCEP estimator to homogenous dynamic panel data models and derive an expression for its finite T inconsistency that will be used in Section 4 to construct a bias-corrected CCEP estimator. Monte Carlo simulation results are presented in Section 5. In Section 6, we use our CCEPbc approach to estimate the dynamic impact of temperature shocks on aggregate output growth in a panel of 125 countries. Section 7 concludes. Proofs and additional results are collected in an online supplement.

Before proceeding we introduce some notation that will be used throughout the article. For a $T \times c$ matrix $\mathbf{A}$, $\|\mathbf{A}\| = [tr(\mathbf{A}\mathbf{A}')]^{1/2}$ denotes the Euclidian (Frobenius) matrix norm, $tr(\cdot)$ the trace, $rk(\cdot)$ the rank, $\text{vec}(\cdot)$ is the vectorization operator, and $(\mathbf{A}'\mathbf{A})^{\dagger}$ is the Moore–Penrose pseudoinverse of $\mathbf{A}'\mathbf{A}$. A $-p$ subscript corresponds to the p-period lag of the respective variable or matrix so that $\mathbf{A}_{-p} = L^p\mathbf{A}$, where L is the lag operator.

2. MODEL AND ASSUMPTIONS

Consider the following first-order dynamic panel data model

$$y_{it} = \alpha_i + \rho y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + e_{it}, \tag{1}$$

$$e_{it} = \boldsymbol{\gamma}_i'\mathbf{f}_t + \varepsilon_{it}, \tag{2}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$ and where y_{it} is the observation on the dependent variable for unit i at time t, α_i is an unobserved individual effect, $\mathbf{x}_{it}$ an individual-specific $k_x \times 1$ column vector of strictly exogenous regressors, and e_{it} a multifactor error term that is composed of an $m \times 1$ vector of unobserved common factors $\mathbf{f}_t$ with heterogeneous factor loadings $\boldsymbol{\gamma}_i$ and an idiosyncratic error term ε_{it}. The unknown parameters ρ and $\boldsymbol{\beta}$ are assumed to be homogenous over cross-sectional units i and bounded by a finite constant. For notational convenience we assume y_{i0} known.

Following Pesaran, Smith, and Yamagata (2013), we also exploit information regarding the unobserved common factors that is shared by variables other than y_{it} and $\mathbf{x}_{it}$. To this end consider a $k_g \times 1$ vector of individual-specific strictly exogenous covariates $\mathbf{g}_{it}$ that have no effect on the dependent variable y_{it} but that are driven by the same factors $\mathbf{f}_t$ that affect y_{it}. The individual-specific covariates and other variables are collected in the $k \times 1$ column vector $\mathbf{z}_{it} = [\mathbf{x}_{it}', \mathbf{g}_{it}']'$, with $k = k_x + k_g$, and are assumed to be generated as

$$\mathbf{z}_{it} = \begin{bmatrix} \mathbf{x}_{it} \\ \mathbf{g}_{it} \end{bmatrix} = \mathbf{c}_{z,i} + \sum_{l=1}^{p} \boldsymbol{\lambda}_l \mathbf{z}_{i,t-l} + \boldsymbol{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{it}, \tag{3}$$

where $\mathbf{c}_{z,i}$ is a $k \times 1$ column vector of unobserved individual effects, p denotes the autoregressive order of $\mathbf{z}_{it}$, $\boldsymbol{\lambda}_l$ is a $k \times k$ matrix of coefficients corresponding to lags $l = 1, \ldots, p$ of $\mathbf{z}_{it}$, $\boldsymbol{\Gamma}_i$ is a $m \times k$ matrix of factor loadings, and $\mathbf{v}_{it}$ a $k \times 1$ vector of idiosyncratic errors. The assumption that p is equal for all variables in $\mathbf{z}_{it}$ is for notational convenience only and can easily be relaxed within the current notation by interpreting p as the maximum lag length and setting some of the parameters in $\boldsymbol{\lambda}_l$ equal to zero.

We make the following assumptions:

Assumption 1 (Idiosyncratic errors). The ε_{it} and $\mathbf{v}_{it}$ are iid across i and t with $E(\varepsilon_{it}\mathbf{v}_{js}) = \mathbf{0}_{k\times 1}$, $E(\varepsilon_{it}^4) < \infty$, and $E(\|\mathbf{v}_{it}\|^4) < \infty$ for all i, j, t, and s. In particular,

$$\varepsilon_{it} \sim \text{iid}(0, \sigma_\varepsilon^2), \quad \mathbf{v}_{it} \sim \text{iid}(\mathbf{0}_{k\times 1}, \boldsymbol{\Omega}_{\text{v}}),$$

with $\sigma_\varepsilon^2 > 0$ and $\boldsymbol{\Omega}_{\text{v}}$ a positive definite $k \times k$ matrix.

Assumption 2 (Common factors). The $\mathbf{f}_t$ are covariance stationary with absolute summable autocovariances, $E(\|\mathbf{f}_t\|^4) < \infty$ and they are distributed independently of ε_{is}, $\mathbf{v}_{is}$, $\boldsymbol{\gamma}_i$, and $\boldsymbol{\Gamma}_i$ for all i, t, and s.

Assumption 3 (Factor loadings). The $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Gamma}_i$ are iid across i, independent of ε_{jt}, $\mathbf{v}_{jt}$, and $\mathbf{f}_t$ for all i, j, and t, with $E(\|\boldsymbol{\gamma}_i\|^4) < \infty$ and $E(\|\boldsymbol{\Gamma}_i\|^4) < \infty$. In particular,

$$\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_i, \quad \boldsymbol{\eta}_i \sim \text{iid}(\mathbf{0}_{m\times 1}, \boldsymbol{\Omega}_\eta), \tag{4}$$

$$\boldsymbol{\Gamma}_i = \boldsymbol{\Gamma} + \boldsymbol{\nu}_i, \quad \text{vec}(\boldsymbol{\nu}_i) \sim \text{iid}(\mathbf{0}_{mk\times 1}, \boldsymbol{\Omega}_\nu), \tag{5}$$

where $E(\|\boldsymbol{\eta}_i' \otimes \boldsymbol{\nu}_i'\|) \geq 0$ and $\boldsymbol{\Omega}_\eta$, $\boldsymbol{\Omega}_\nu$ are bounded $m \times m$ and $km \times km$ matrices, respectively.

Assumption 4 (Rank condition). The $(1 + k) \times m$ matrix $\mathbf{C} = [\boldsymbol{\gamma}, \boldsymbol{\Gamma}]'$ has $rk(\mathbf{C}) = m$.

Assumption 5 (Stationarity). $|\rho| < 1$ and the elements in $\boldsymbol{\lambda}_l$ are such that $\boldsymbol{\lambda}(L) = \mathbf{I}_k - \sum_{l=1}^{p} \boldsymbol{\lambda}_l L^l$ is invertible. The process of y_{it} was initiated in the infinite past.

For future reference, we let $k_w = 1 + k_x$ and stack the model in Equations (1) and (2) over time as

$$\mathbf{y}_i = \alpha_i \boldsymbol{\iota}_T + \mathbf{w}_i \boldsymbol{\delta} + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \tag{6}$$

where $\boldsymbol{\delta} = [\rho, \boldsymbol{\beta}']'$ and $\mathbf{w}_i = [\mathbf{y}_{i,-1}, \mathbf{X}_i]$ are $k_w \times 1$ and $T \times k_w$, and $\mathbf{X}_i = [\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}]'$, $\mathbf{y}_i = [\text{y}_{i1}, \ldots, \text{y}_{iT}]'$, $\mathbf{y}_{i,-1} = [\text{y}_{i0}, \ldots, \text{y}_{i,T-1}]'$, $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_T]'$, $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1}, \ldots, \varepsilon_{iT}]'$, and $\boldsymbol{\iota}_T$ is a $T \times 1$ column vector of ones. Similarly specify $\mathbf{G}_i = [\mathbf{g}_{i1}, \ldots, \mathbf{g}_{iT}]'$ and $\mathbf{Z}_i = [\mathbf{X}_i, \mathbf{G}_i]$.

3. CCEP ESTIMATION IN DYNAMIC PANELS

Pesaran (2006) developed the CCE approach in a static model with strictly exogenous regressors and showed that under Assumption 4 the differential effects of the unobserved factors can be eliminated as $N \to \infty$ by augmenting the model with the CSA of the observables. In this section, we first review whether the CSA still serve as suitable proxies for the factors in homogenous dynamic panels. We next show that, in contrast to the static case, the CCEP estimator is inconsistent when $N \to \infty$ and T fixed by deriving its bias expression that will be used in Section 4 to construct a bias-corrected CCEP estimator.

3.1. Cross-Sectional Averages as Proxies for the Common Factors

Rewriting Equations (1)–(3) as

$$\rho(L)\,\text{y}_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \boldsymbol{\gamma}_i'\mathbf{f}_t + \varepsilon_{it},$$
$$\boldsymbol{\lambda}(L)\,\mathbf{z}_{it} = \mathbf{c}_{z,i} + \boldsymbol{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{it},$$

where $\rho(L) = 1 - \rho L$ and $\boldsymbol{\lambda}(L) = \mathbf{I}_k - \sum_{l=1}^{p} \boldsymbol{\lambda}_l L^l$, and taking CSA yields

$$\rho(L)\,\bar{\text{y}}_t = \bar{\alpha} + \bar{\mathbf{x}}_t'\boldsymbol{\beta} + \boldsymbol{\gamma}'\mathbf{f}_t + O_p(N^{-1/2}), \tag{7}$$

$$\boldsymbol{\lambda}(L)\,\bar{\mathbf{z}}_t = \bar{\mathbf{c}}_z + \boldsymbol{\Gamma}'\mathbf{f}_t + O_p(N^{-1/2}), \tag{8}$$

with the affix notation on $\bar{\text{y}}_t$ used to denote the CSA $\bar{\text{y}}_t = \frac{1}{N}\sum_{i=1}^{N} \text{y}_{it}$ and similarly for all other series. Under Assumption 4 that $\mathbf{C}$ has full column rank, we can solve for $\mathbf{f}_t$ to obtain

$$\mathbf{f}_t = \left(\mathbf{C}'\mathbf{C}\right)^{-1}\mathbf{C}'\left(\begin{bmatrix}\rho(L) & -(\boldsymbol{\beta}^*)' \\ 0 & \boldsymbol{\lambda}(L)\end{bmatrix}\begin{bmatrix}\bar{\text{y}}_t \\ \bar{\mathbf{z}}_t\end{bmatrix} - \begin{bmatrix}\bar{\alpha} \\ \bar{\mathbf{c}}_z\end{bmatrix}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{9}$$

with $\boldsymbol{\beta}^* = [\boldsymbol{\beta}', \mathbf{0}_{k_g\times 1}']'$. Equation (9) shows that as $N \to \infty$ the factors can be approximated by the CSA of y_{it} and $\mathbf{z}_{it}$ as well as a finite number of their lags determined by the orders of the polynomials $\rho(L)$ and $\boldsymbol{\lambda}(L)$. This result differs from the heterogeneous dynamic model considered by Chudik and Pesaran (2015) who find that an infinite number of lags is required in this case.

The intuition behind the above result is that in the presence of dynamics the lags are needed to separate the contemporaneous factor from its past realizations within the CSA. This is necessary to approximate $\mathbf{f}_t$ in function of observables as $N \to \infty$. To see this, consider the simple case of models (1) and (2) with one factor and $\boldsymbol{\beta} = \mathbf{0}$. The CSA of y_{it} can then be written as

$$\bar{\text{y}}_t = \bar{\alpha} + \bar{\gamma}\text{f}_t + \bar{\varepsilon}_t + \rho\left(\frac{\bar{\alpha}}{1-\rho} + \bar{\gamma}\text{f}_{t-1}^{+} + \bar{\varepsilon}_{t-1}^{+}\right), \tag{10}$$

$$= \frac{\bar{\alpha}}{1-\rho} + \gamma\left[\text{f}_t + \rho\text{f}_{t-1}^{+}\right] + O_p(N^{-1/2}), \tag{11}$$

so that it is not only a function of the factors at time t, a constant and an $O_p(N^{-1/2})$ term, but also of the past realizations of the factors through $\text{f}_{t-1}^{+} = \sum_{l=0}^{\infty} \rho^l \text{f}_{t-l-1}$. Solving the contemporaneous factor f_t from (11) would therefore still depend on the unobservable f_{t-1}^{+} so a proxy cannot be constructed from it. However, noting that the term inside the brackets of (10) equals $\bar{\text{y}}_{t-1}$, subtracting $\rho\bar{\text{y}}_{t-1}$ from (10) yields

$$\bar{\text{y}}_t - \rho\bar{\text{y}}_{t-1} = \bar{\alpha} + \gamma\text{f}_t + O_p(N^{-1/2}), \tag{12}$$

so that the past factor realizations are cut out and this equation can be solved for f_t as a function of observables, estimable parameters and an $O_p(N^{-1/2})$ term. The combination of observables can then be used to project out the factors at time t as $N \to \infty$. A similar reasoning holds for $\mathbf{z}_{it}$ as well. This clearly illustrates the difference with the static case in Pesaran (2006), where the absence of dynamics implies that $\rho = 0$ so that the CSA do not contain the past factors and, hence, lags are not required to separate them from f_t.

Remark 1. The requirement that we have to know the order of $\boldsymbol{\lambda}(L)$ may be unfortunate in practice as p is typically unknown (and may also differ over variables included in $\mathbf{z}_{it}$). Decisions on p imply assumptions about the autoregressive order of $\mathbf{z}_{it}$ that may be hard to verify since the observed persistence in $\mathbf{z}_{it}$ may stem from serially correlated factors $\mathbf{f}_t$ or from $\boldsymbol{\lambda}(L) \neq \mathbf{I}_k$. However, as more time series observations become available the factor approximation should not suffer from including too many lags $p^* > p$ of $\bar{\mathbf{z}}_t$. Hence, in practice it may be convenient to choose $p^* = \lfloor T^{1/3} \rceil$ as in Chudik and Pesaran (2015), with $\lfloor x \rceil$ denoting the integer part of x, to make the CCEP estimator robust to misspecification of p while ensuring that the number of lags does not increase too fast in T and sufficient degrees of freedom are available.

3.2. Dynamic CCEP Estimator

In light of the discussion in the previous section, construct the following $T \times c$ matrix $\mathbf{Q} = [\boldsymbol{\iota}_T, \bar{\mathbf{y}}, \bar{\mathbf{y}}_{-1}, \bar{\mathbf{Z}}, \ldots, \bar{\mathbf{Z}}_{-p^*}]$ and augment the model in Equation (6) as

$$\mathbf{y}_i = \mathbf{w}_i \boldsymbol{\delta} + \mathbf{Q}\boldsymbol{\kappa}_i + \mathbf{e}_i, \tag{13}$$

where the CSA in $\mathbf{Q}$ serve to control for the common factors absorbed in the error terms $\mathbf{e}_i$, and $\boldsymbol{\kappa}_i$ are parameters to be estimated along with the slope coefficients of interest $\boldsymbol{\delta} = [\rho, \boldsymbol{\beta}']'$. Assuming that $T \geq k_w + c$ (estimability) and setting pooling weights to N^{-1}, the dynamic CCEP estimator for $\boldsymbol{\delta}$ in Equation (13) is

$$\widehat{\boldsymbol{\delta}} = \left(\sum_{i=1}^{N} \mathbf{w}_i' \mathbf{M} \mathbf{w}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{w}_i' \mathbf{M} \mathbf{y}_i, \tag{14}$$

where $\mathbf{M} = \mathbf{I}_T - \mathbf{H}$ and $\mathbf{H} = \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{\dagger}\mathbf{Q}'$ is the projection on $\mathbf{Q}$.

The dynamic CCEP estimator in Equation (14) controls, as $N \to \infty$, for the unobserved factors provided that the rank condition (Assumption 4) holds and the model is augmented with a sufficient number ($p^* \geq p$) of lagged CSA. However, despite controlling for the factors, the inclusion of the CSA induces a new finite T bias term. The following theorem provides an analytical expression of the asymptotic bias of the dynamic CCEP estimator for $N \to \infty$ and T fixed conditional on the factors and CSA.

Theorem 1. Suppose that $p^* \geq p$ and Assumptions 1–5 hold, and let $\mathcal{C}$ be the σ-algebra generated by the common factors and $[\bar{\mathbf{y}}, \bar{\mathbf{Z}}, \ldots, \bar{\mathbf{y}}_{-p^*}, \bar{\mathbf{Z}}_{-p^*}]$. The CCEP estimator in Equation (14) is inconsistent as $N \to \infty$ and T fixed with its asymptotic bias conditional on $\mathcal{C}$ given by

$$\underset{N\to\infty}{\text{plim}}\, \widehat{\boldsymbol{\delta}} = \mathbf{m}(\boldsymbol{\delta}) = \boldsymbol{\delta} - \frac{\sigma_\varepsilon^2}{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{\upsilon}(\rho, \mathbf{H}), \tag{15}$$

with $\boldsymbol{\upsilon}(\rho, \mathbf{H}) = \upsilon(\rho, \mathbf{H})\mathbf{q}_1$, $\mathbf{q}_1 = [1, \mathbf{0}_{1\times k_x}]'$, $\upsilon(\rho, \mathbf{H}) = \sum_{t=1}^{T-1} \rho^{t-1} \sum_{s=t+1}^{T} h_{s,s-t}$, and $h_{s,s-t}$ is the element on row s and column $s-t$ of $\mathbf{H}$. $\boldsymbol{\Sigma} = \lim_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \mathbf{w}_i' \mathbf{M} \mathbf{w}_i$ is given in (C-7) of the online supplement. Letting $\mathbf{S}_x = [\mathbf{0}_{k_x \times 1}, \mathbf{I}_{k_x}]'$, Equation (15) can be decomposed as

$$\underset{N\to\infty}{\text{plim}}\, (\widehat{\rho} - \rho) = -\frac{1}{T} \frac{\sigma_\varepsilon^2}{\sigma_{\check{y}_{-1}}^2} \upsilon(\rho, \mathbf{H}), \tag{16}$$

$$\underset{N\to\infty}{\text{plim}}\, \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = -\boldsymbol{\zeta} \underset{N\to\infty}{\text{plim}}\, (\widehat{\rho} - \rho), \tag{17}$$

with $\sigma_{\check{y}_{-1}}^2 = \text{plim}_{N\to\infty} \frac{\check{\mathbf{y}}_{-1}'\check{\mathbf{y}}_{-1}}{NT}$, $\boldsymbol{\zeta} = (\mathbf{S}_x' \boldsymbol{\Sigma} \mathbf{S}_x)^{-1} \mathbf{S}_x' \boldsymbol{\Sigma} \mathbf{q}_1$ and where $\check{\mathbf{y}}_{-1} = \mathbb{M}_x [\mathbf{y}_{1,-1}', \ldots, \mathbf{y}_{N,-1}']'$, $\mathbb{M}_x = \mathbb{M} - \check{\mathbf{X}}(\check{\mathbf{X}}'\check{\mathbf{X}})^{-1}\check{\mathbf{X}}'$, $\check{\mathbf{X}} = \mathbb{M}[\mathbf{X}_1', \ldots, \mathbf{X}_N']'$, and $\mathbb{M} = \mathbf{I}_N \otimes \mathbf{M}$.

Theorem 1 extends the results in Everaert and De Groote (2016), who consider a model with one factor and no additional covariates, to a model with multiple factors and exogenous regressors. Also in this more general setting, the asymptotic bias of the CCEP estimator is not caused by the factor structure but it is induced by projecting the data on $\mathbf{Q}$ as this induces weak endogeneity in the transformed lagged dependent variable and hence inconsistency of the autoregressive parameter $\widehat{\rho}$ for finite T, as shown in Equation (16). Concerning the coefficients of the exogenous regressors, Equation (17) reveals that the bias of $\widehat{\boldsymbol{\beta}}$ is a fraction $-\boldsymbol{\zeta}$ of the bias of $\widehat{\rho}$, with $\boldsymbol{\zeta}$ being the CCEP estimates (as $N \to \infty$) when regressing $\mathbf{y}_{i,-1}$ on $\mathbf{X}_i$. As such, the direction of the distortion in $\widehat{\boldsymbol{\beta}}$ is determined by the correlation between $\mathbf{y}_{i,-1}$ and $\mathbf{X}_i$ given by $\boldsymbol{\zeta}$, but it is the inconsistency of $\widehat{\rho}$ that creates bias for the entire coefficient vector. The inconsistency in $\widehat{\rho}$ is therefore the principal driver of the overall bias, and we study it in more detail in Section A.2 of the online supplement. The most important conclusions of that analysis are:

- The asymptotic bias is expected to be negative for $\rho > 0$.
- The asymptotic bias is a stochastic variable because it depends on the projection matrix $\mathbf{H}$, which is a random matrix even as $N \to \infty$.
- The absolute value of the asymptotic bias is, ceteris paribus, increasing in the persistence ρ and the number of CSA (columns of $\mathbf{Q}$), and decreasing in T and in the importance of the factors when there is more than one factor.

The key practical implication of Theorem 1 is that there is a trade-off associated with augmenting the model with the CSA. On the one hand, to control for the unobserved common factors, a sufficient number of CSA should be included such that the rank condition is satisfied and $p^* > p$. Simulation evidence further suggests that even when the rank condition is satisfied, using additional CSA improves factor approximation in finite N samples (see Section 5). On the other hand, in finite T settings, the augmentation generates a bias term that increases in magnitude with the number of CSA. As such, whereas adding CSA is beneficial to treat the common factor problem, it can simultaneously be detrimental for the finite T properties of the CCEP estimator. Our objective in the next section is to resolve this trade-off by removing the bias induced by projecting out the CSA.

4. BIAS-CORRECTED DYNAMIC CCEP

In what follows we develop a bias-corrected CCEP estimator based on the analytical bias expression for $N \to \infty$ and T fixed presented in Equation (15) of Theorem 1, and derive its asymptotic distribution.

4.1. Bias Correction Procedure

The CCEPbc estimator $\widehat{\boldsymbol{\delta}}_{bc}$ can be obtained as the vector $\boldsymbol{\delta}_0$ that satisfies

$$\widehat{\boldsymbol{\delta}} - \widehat{\mathbf{m}}(\boldsymbol{\delta}_0) = \mathbf{0}_{k_w \times 1}, \tag{18}$$

with $\widehat{\mathbf{m}}(\cdot)$ the feasible version of the asymptotic bias expression in Equation (15),

$$\widehat{\mathbf{m}}(\boldsymbol{\delta}_0) = \boldsymbol{\delta}_0 - T^{-1}\widehat{\sigma}^2_\varepsilon(\boldsymbol{\delta}_0)\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\upsilon}(\rho_0, \mathbf{H}), \tag{19}$$

where $\boldsymbol{\Sigma}$ is replaced by its sample analog $\widehat{\boldsymbol{\Sigma}} = \frac{1}{NT}\sum_{i=1}^N \mathbf{w}_i'\mathbf{M}\mathbf{w}_i$ and the unknown variance σ^2_ε is substituted by the function

$$\widehat{\sigma}^2_\varepsilon(\boldsymbol{\delta}_0) = \frac{1}{N(T-c)}\sum_{i=1}^N \left\|\mathbf{M}\left(\mathbf{y}_i - \mathbf{w}_i\boldsymbol{\delta}_0\right)\right\|^2. \tag{20}$$

The traditional estimator for σ^2_ε based on the uncorrected CCEP error terms $\widehat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{w}_i\widehat{\boldsymbol{\delta}}$ is inconsistent for finite T due to the inconsistency of $\widehat{\boldsymbol{\delta}}$, but by constructing $\widehat{\sigma}^2_\varepsilon(\cdot)$ as a function of the parameters of interest, solving (18) implies that we use a bias-adjusted estimator for σ^2_ε as well. In summary, the CCEPbc estimator is

$$\widehat{\boldsymbol{\delta}}_{bc} = \underset{\boldsymbol{\delta}_0 \in \chi}{\arg\min}\ \frac{1}{2}\left\|\widehat{\boldsymbol{\delta}} - \widehat{\mathbf{m}}(\boldsymbol{\delta}_0)\right\|^2, \tag{21}$$

with $\chi \subseteq \mathbb{R}^{k_w}$. This optimization problem is easily managed by standard numerical solvers and requires very little additional programming besides computing the CCEP estimates $\widehat{\boldsymbol{\delta}}$. The solution $\widehat{\boldsymbol{\delta}}_{bc}$ is equivalent to the vector of parameters that follows from inverting $\widehat{\boldsymbol{\delta}} = \widehat{\mathbf{m}}(\boldsymbol{\delta})$ so that we can alternatively write the CCEPbc estimator as $\widehat{\boldsymbol{\delta}}_{bc} = \widehat{\mathbf{m}}^{-1}(\widehat{\boldsymbol{\delta}})$. Notice how Equation (21) implies that the bias adjustment can be seen as a minimum distance estimator, or a GMM approach that employs the bias-corrected orthogonality conditions in (18) to estimate the population parameters. To make this point explicit, straightforward manipulations in (18) give

$$\begin{aligned}\widehat{\boldsymbol{\delta}} - \widehat{\mathbf{m}}(\boldsymbol{\delta}_0) &= (\boldsymbol{\delta} - \boldsymbol{\delta}_0) \\ &\quad + \widehat{\boldsymbol{\Sigma}}^{-1}\left[\frac{1}{NT}\sum_{i=1}^N \mathbf{w}_i'\mathbf{M}(\mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i) + \mathbf{b}_\varepsilon(\boldsymbol{\delta}_0)\right] \\ &= \mathbf{0}_{k_w \times 1},\end{aligned} \tag{22}$$

with $\mathbf{b}_\varepsilon(\boldsymbol{\delta}_0) = T^{-1}\widehat{\sigma}^2_\varepsilon(\boldsymbol{\delta}_0)\boldsymbol{\upsilon}(\rho_0, \mathbf{H})$. This shows that the moment conditions underlying CCEPbc in Equation (21) are identical to those of the CCEP estimator, except for the $\mathbf{b}_\varepsilon(\boldsymbol{\delta}_0)$ term which corrects for the finite T bias. In this sense, our approach is similar in spirit to ideas presented in Chudik and Pesaran (2017). Also note that Bun and Carree (2005) use a similar approach to obtain a bias-corrected FE estimator in dynamic panel data models without common factors.

Remark 2. The CCEPbc estimator outlined above is a generally applicable method in the sense that it does not require the number of factors to be known. In the single factor setting, Equations (16) and (17) can be simplified to obtain more efficient restricted bias corrections. We present two alternative restricted CCEPbc estimators in Section A.3 of the online supplement.

4.2. Asymptotic Properties and Inference

The CCEPbc estimator presented in Equation (21) builds on the orthogonality conditions in Equation (22) to estimate the population parameters of interest. We show in Theorem 3 of the online supplement that these moment conditions are satisfied as $N \to \infty$ at $\boldsymbol{\delta}_0 = \boldsymbol{\delta}$ and that the CCEPbc estimator is thus consistent as $N \to \infty$ and T fixed

$$\widehat{\boldsymbol{\delta}}_{bc} - \boldsymbol{\delta} \longrightarrow^p \mathbf{0}_{k_w \times 1}. \tag{23}$$

As such, in dynamic models the proposed correction restores the large N finite T consistency of the CCEP estimator established by Pesaran (2006) in a static setting.

The finite T distribution of the CCEPbc estimator is generally intractable due to the presence of nuisance parameters, unless one makes the very stringent assumption that $m = 1 + k$ (see, e.g., Karabiyik, Reese, and Westerlund 2017 for more details). In the next theorem, we establish asymptotic normality for the CCEPbc estimator in the general $m \leq 1 + k$ case letting $(N, T) \to \infty$.

Theorem 2. Let Assumptions 1–5 hold and suppose that $p^* \geq p$ and $\chi \subseteq \mathbb{R}^{k_w}$ is compact such that $|\rho_0| < 1$ with $\boldsymbol{\delta} \in \chi$. Then, as $(N, T) \to \infty$ it holds that $\widehat{\boldsymbol{\delta}}_{bc} \longrightarrow^p \boldsymbol{\delta}$, and provided $T/N \to 0$

$$\sqrt{NT}(\widehat{\boldsymbol{\delta}}_{bc} - \boldsymbol{\delta}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_{k_w \times 1}, \dot{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Phi}\dot{\boldsymbol{\Sigma}}^{-1}\right), \tag{24}$$

with $\dot{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Phi}$ defined in Equations (D-30) and (D-53) of the online supplement, respectively.

Theorem 2 establishes that the CCEPbc estimator is asymptotically normally distributed as $(N, T) \to \infty$ and that it enables unbiased inference provided $T/N \to 0$. This requirement on the relative growth rate of N and T stems from estimating the factors with the CSA and it is identical to what Pesaran (2006) and Karabiyik, Reese, and Westerlund (2017) require for unbiased inference with CCEP in the static model.

The asymptotic variance in Equation (24) can consistently, as $(N, T) \to \infty$, be estimated by

$$\widehat{\boldsymbol{\Omega}} = (\widehat{\boldsymbol{\Delta}}'\widehat{\boldsymbol{\Delta}})^{-1}\widehat{\boldsymbol{\Delta}}'\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Delta}}(\widehat{\boldsymbol{\Delta}}'\widehat{\boldsymbol{\Delta}})^{-1}, \tag{25}$$

where $\widehat{\boldsymbol{\Delta}} = \mathbf{J}_a(\widehat{\boldsymbol{\delta}}_{bc})$ and $\mathbf{J}_a(\cdot)$ is the Jacobian presented in Equation (A-2), and with $\widehat{\boldsymbol{\Phi}} = \frac{1}{NT}\sum_{i=1}^N \widehat{\mathbf{q}}_i\widehat{\mathbf{q}}_i'$, $\widehat{\mathbf{q}}_i = \mathbf{w}_i'\mathbf{M}\widehat{\mathbf{e}}_i + \widehat{\sigma}^2_\varepsilon(\widehat{\boldsymbol{\delta}}_{bc})\boldsymbol{\upsilon}(\widehat{\rho}_{bc}, \mathbf{H})$ and $\widehat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{w}_i\widehat{\boldsymbol{\delta}}_{bc}$.

In practice, in particular in settings when T is not large, the bootstrap provides a convenient alternative to (25) for estimating the finite sample variance of $\widehat{\boldsymbol{\delta}}_{bc}$. To that end, we follow Kapetanios (2008) and obtain bootstrap samples by resampling whole cross-sectional units with replacement from the original dataset. In particular, let $\mathcal{B}^0 = [\mathbf{a}_1, \ldots, \mathbf{a}_N]$ be the original dataset, with $\mathbf{a}_i = [\mathbf{d}_{i,-p^*}, \ldots, \mathbf{d}_{iT}]'$ and $\mathbf{d}_{it} = [\mathrm{y}_{it}, \mathbf{z}_{it}']'$. Bootstrap sample $j = 1, \ldots, J$ is generated by drawing N indices with replacement from $(1, \ldots, N)$, and collecting the $\mathbf{a}_i$ corresponding to these indices in $\mathcal{B}^j$. This resampling scheme is valid as $N \to \infty$ and preserves both the dynamics and the assumed factor structure in the data. The distribution of $\widehat{\boldsymbol{\delta}}_{bc}$ is then simulated by applying CCEPbc to each of the J bootstrap datasets $[\mathcal{B}^1, \ldots, \mathcal{B}^J]$ to obtain the corresponding coefficient vectors $[\widehat{\boldsymbol{\delta}}^b_{bc,1}, \ldots, \widehat{\boldsymbol{\delta}}^b_{bc,J}]$. Inference can then be made using the bootstrapped variance-covariance matrix

$$\widehat{\boldsymbol{\Omega}}_b = \lim_{J\to\infty} \frac{1}{J-1} \sum_{j=1}^{J} \left(\widehat{\boldsymbol{\delta}}^b_{bc,j} - \bar{\boldsymbol{\delta}}^b_{bc}\right)\left(\widehat{\boldsymbol{\delta}}^b_{bc,j} - \bar{\boldsymbol{\delta}}^b_{bc}\right)', \quad (26)$$

with $\bar{\boldsymbol{\delta}}^b_{bc} = \frac{1}{J}\sum_{j=1}^{J} \widehat{\boldsymbol{\delta}}^b_{bc,j}$ the average of the estimates over the J samples.

Remark 3. Lemmas 14 and 15 in the online supplement show that, in contrast to the CCEPbc estimator, the asymptotic distribution of the uncorrected CCEP estimator in Equation (14) features bias terms unless both $N/T \to 0$ (due to the finite T bias in Theorem 1) and $T/N \to 0$ (due to estimation of the factors). As this is clearly a contradiction, bias correction is crucial for reliable inference with the CCEP approach despite that the estimator is consistent as $(N, T) \to \infty$ in the dynamic model.

5. MONTE CARLO SIMULATION

In this section, we use Monte Carlo simulations to investigate the small sample properties of our bias-corrected CCEP estimator and compare its performance to the original CCEP estimator and a number of alternative methods proposed in the literature.

5.1. Design

We generate data for y_{it} and $\mathbf{z}_{it}$ according to the model in Equations (1)–(3) assuming a single explanatory variable x_{it} ($k_x = 1$) and one additional variable g_{it} ($k_g = 1$) that has no impact on y_{it} but provides additional information about the common factors. We set $\beta = 1 - \rho$ to normalize the long-run impact of x_{it} to one and assume $\boldsymbol{\lambda}(L) = (1 - \lambda L)\mathbf{I}_2$ which restricts the autoregressive order of x_{it} and g_{it} to be at most one ($p = 1$). This implies that the one period lagged CSA $\bar{\mathrm{x}}_{t-1}$ should be added to the CCE orthogonalization matrix in settings where $\lambda \neq 0$ (and preferably also $\bar{\mathrm{g}}_{t-1}$ when g_{it} is used as an additional variable).

The m common factors are generated as

$$\mathrm{f}_{jt} = \theta \mathrm{f}_{j,t-1} + \mu_{jt},$$

with $\mu_{jt} \sim N\left(0, (1-\theta^2)/m\right)$ for every $j = 1, \ldots, m$. The reason for dividing the variance by m is to prevent the factors from dominating the model as their number m rises. We will conduct experiments with $m = 1$ and $m = 2$.

The fixed effects are generated as $\alpha_i \sim N\left(0, \sigma^2_\alpha\right)$ and $\mathbf{c}_{z,i} \sim N\left(\mathbf{0}, \sigma^2_c \mathbf{I}_2\right)$ and the idiosyncratic errors as $\varepsilon_{it} \sim N\left(0, 1-\rho^2\right)$ and $\mathbf{v}_{it} \sim N\left(\mathbf{0}, (1-\lambda^2)\mathbf{I}_2\right)$. The variance parameters σ^2_α and σ^2_c are set such that the contributions of the fixed effects to the variance of y_{it} and $\mathbf{z}_{it}$ equal that of their respective idiosyncratic innovations (ε_{it} and $\mathbf{v}_{it}$). The factor loadings in the data generating process (DGP) of y_{it}, x_{it}, and g_{it} are generated as

$$\mathbf{C}_i = \begin{bmatrix} \boldsymbol{\gamma}_i' \\ \boldsymbol{\Gamma}^{x\prime}_i \\ \boldsymbol{\Gamma}^{g\prime}_i \end{bmatrix} = \begin{bmatrix} \gamma_{1,i} & \gamma_{2,i} \\ \Gamma^x_{1,i} & \Gamma^x_{2,i} \\ \Gamma^g_{1,i} & \Gamma^g_{2,i} \end{bmatrix} \sim \mathrm{IIDU} \begin{bmatrix} [0, \gamma_u] & [0, \gamma_u - 3/5] \\ [0, 1] & [0, 0.2] \\ [-0.6, 0] & [-1.4, 0] \end{bmatrix},$$

when $m = 2$ or with the second column set to zero in case $m = 1$. The upper bound γ_u is calibrated such that the relative importance of the factors and the idiosyncratic errors in the total variance of y_{it}, denoted RI, is either 1 or 3. RI $= 1$ corresponds to cases where the factors have a normal influence on y_{it} whereas RI $= 3$ is a scenario where the factors are very influential. The specific values for the upper and lower bounds of the uniform distributions for the loadings in $\mathbf{C}_i$ are sufficiently different to ensure that the rank condition is satisfied and that the full set of CSA contains enough independent information about the common factors.

Experiments are conducted for combinations of the following parameter values: $\rho \in \{0.4; 0.8\}$, RI $\in \{1; 3\}$, and $\lambda \in \{0; 0.6\}$. The autoregressive parameter θ in the DGP of the factors is set to 0.6 in all experiments to account for the fact that factors are often persistent in practice. We consider $\rho = 0.8$, $\lambda = 0$, $m = 1$, and RI $= 1$ our baseline scenario. This is a challenging setting for our bias-correction procedure as the large autoregressive parameter ρ will result in a considerable bias for the CCEP estimator. We generate datasets with $N = (25, 50, 100, 500, 1000, 5000)$ and $T = (10, 15, 20, 30, 50, 100)$. As such, next to a typical macro panel dimension (N small and T small to moderate) we also consider a more micro panel perspective (N large and T small). To conserve space we will report only a few relevant combinations of N and T in each table.

We initialize $\mathrm{y}_{i,-50}$, $\mathbf{z}_{i,-50}$, and $\mathrm{f}_{j,-50}$ at zero and discard the first 50 observations to neutralize initial conditions. We generate 2000 datasets for each combination of N and T and calculate performance measures including median bias, root mean squared error (RMSE), and actual size. Although analytical variance expressions are available for some estimators, to make fair comparisons we obtain standard errors using a bootstrap approach for each of the considered estimators. Following Kapetanios (2008), we resample cross-sectional units as a whole as described in Section 4.2. The advantage of this scheme is that it preserves both the persistence and the cross-sectional dependence in the data and is valid even when T is small. We calculate actual test size using bootstrap standard errors based on 150 bootstrap samples. The reported actual size is the false rejection probability of a t-test at the 5% nominal significance level. Results for the CCEPbc estimator with standard errors estimated using (25) are available upon request.

We summarize and discuss our main findings below. We start with some baseline results for estimating ρ and β using various estimators and sample sizes. Next, we focus on a number of interesting aspects with respect to estimating ρ by considering changes to the baseline design and alternative setups for the bias corrections. Since differences between estimators are more pronounced for large N we mostly report tables for $N = 500$ in the main text. Small N versions ($N = 25$) are provided in Section E of the online supplement, while in Section F we fix $T = 10$ and plot the behavior of CCEPbc as N grows very large to assess its behavior as $N \to \infty$.

5.2. Baseline Results

We start our discussion with a comparison of the performance of our CCEPbc estimator to various alternative estimators in the baseline scenario where $\rho = 0.8$, $\lambda = 0$, $m = 1$, and RI $= 1$. The CCEP estimator is included as the benchmark estimator. Inspired by Chudik and Pesaran (2015), we also consider two alternative bias-corrected CCEP estimators as direct comparisons to our approach, that is, the recursive mean adjustment (denoted CCEPrm) proposed by So and Shin (1999) and the split-panel jackknife correction (denoted CCEPjk) of Dhaene and Jochmans (2015). We find that CCEPrm provides no improvement over CCEP in any scenario so we exclude it from the tables. In our baseline scenario, the CCEP estimator and the various bias corrections thereof make no use of the additional g_{it} variable or lags of the exogenous variables (which is in line with $\lambda = 0$) in the orthogonalization matrix. Finally, we consider Moon and Weidner's (2017) bias-corrected version of the least squares with interactive effects estimator of Bai (2009). This estimator (denoted FLSbc) is implemented selecting the correct number of factors (2 in our baseline scenario due to the presence of fixed effects) and a bandwidth for the bias correction equal to 4 (which is the optimal choice based on the simulation results of Moon and Weidner for high persistence settings).

The results in Table 1 show that the original CCEP estimator has a severe negative small T bias for ρ of which a fraction is carried over to the estimates for β. When $T = 10$, the bias for $\widehat{\rho}$ amounts to -0.4, while the more moderate time series dimensions of $T = 20$ and $T = 30$ still result in biases of -0.18 and -0.11, respectively. Even for $T = 50$, the bias of -0.06 should not be neglected as this implies seriously distorted inference. Figure 1 further visualizes this in a setting with $N = 500$ and shows that even for $T = 100$ the CCEP estimator will suffer from some bias and hence unreliable inference. Although the CCE approach relies on $N \to \infty$, the results show that biases are more or less stable over alternative values of N. Experiments for $\rho = 0.4$ (see Table E-1 in the online supplement) confirm that the absolute value of the bias of the CCEP estimator is increasing in ρ.

The main takeaway from Table 1 is that our bias-corrected CCEP estimator is (nearly) unbiased in all of the considered sample sizes and hence offers a strong improvement over the original CCEP estimator. Interestingly, CCEPbc also provides a considerable variance reduction whenever $N > 25$. This is due to the fact that the bias of the CCEP estimator is stochastic, as discussed in Theorem 1, which contributes to its variance. The combination of bias removal and variance reduction implies that the RMSE of the CCEPbc estimator is always much lower than that of the CCEP estimator, even for moderately large T. The behavior of CCEPbc for $N = 500$ and varying T is also visualized in Figure 1, showing that in contrast to the CCEP estimator our corrected version is correctly centered. In Figure F-1 of the online supplement, we set $T = 10$ and let the cross-section size N grow large to illustrate the behavior of CCEPbc as $N \to \infty$ and T fixed. The plot reveals that the corrected estimator is indeed consistent as $N \to \infty$, which is clearly not the case for the uncorrected estimator. CCEPbc also offers substantial improvements regarding inference. In contrast to the CCEP, its actual size is always close to the nominal 5% level. As all of these findings hold for each of the considered sample sizes, the CCEPbc is not only an appropriate small T estimator but should also be preferred over CCEP for larger values of T. Moreover, Table 1 shows that the performance of the CCEPbc estimator is not too sensitive to the size of N. As such, it is even applicable in a sample as small as $N = 25$ and $T = 10$.

The alternative bias-adjusted estimators offer some alleviation of the bias but appear less effective compared to CCEPbc. The FLSbc still has a considerable negative small T bias for ρ, while the CCEPjk is able to remove a lot of bias but at the cost of a much larger variance. Accordingly, these alternatives have a much larger RMSE compared to CCEPbc, which should be preferred even for larger T due the more effective correction. Since the bias for $\widehat{\beta}$ is a fraction of that for $\widehat{\rho}$, also $\widehat{\beta}$ is not correctly centered for the alternative estimators and the test size for this coefficient is generally distorted, whereas in the case of CCEPbc it is at the desired 5% level. Similar results are obtained in the low persistence scenario (see Table E-1 in the online supplement), but differences between estimators are smaller since there is less bias to correct for.

5.3. Number of Factors and Their Strength

In this section, we analyze the performance of CCEPbc when varying the number of factors (m is 1 and 2) and their strength (RI is 1 and 3). Table 2 reports simulation results for $N = 500$. Small N results are provided in Table E-2 of the online supplement. Next to the CCEP estimator and its bias corrections that do not use the CSA of g_{it} when approximating the factors, we now also include CCEP variants that do use g_{it} and denote them by adding the (+g) suffix.

The results in Table 2 show that the performance of CCEP and of its bias corrections is not very sensitive to the number of factors or their strength. Only when we drive up the factor strength in the presence of two factors (see the lower right panel of Table 2), we note a slight increase in the bias of our CCEPbc approach. Table 3 further summarizes the behavior of CCEPbc for various sizes of N and T with two strong factors. The top panel reveals that even though the small T bias clearly decreases as N grows, it results in distorted inference unless N is much larger than T. The explanation for this finding is that even though the rank condition is exactly satisfied (2 observables for 2 factors) the information in $\bar{y}_t$ and $\bar{x}_t$ may not be sufficiently distinct to effectively remove two strong factors in finite N settings. In this case, CCEP will have an additional finite N bias term which is not taken into account by our CCEPbc estimator.

Although the remaining bias in the presence of two strong factors disappears as N increases further (see Figure F-4 in the online supplement), we find that the inclusion of $\bar{g}_t$ is a highly effective solution in finite samples. The additional information on the factors that is added through including $\bar{g}_t$

Table 1. Monte Carlo results for ρ and β: baseline design

Estimator	(N,T)	Bias 10	Bias 20	Bias 30	Bias 50	RMSE 10	RMSE 20	RMSE 30	RMSE 50	Size 10	Size 20	Size 30	Size 50
Results for $\widehat{\rho}$													
CCEP	25	−0.385	−0.176	−0.109	−0.061	0.417	0.188	0.118	0.067	0.90	0.92	0.90	0.81
	100	−0.391	−0.176	−0.112	−0.062	0.417	0.185	0.115	0.064	1.00	1.00	1.00	1.00
	500	−0.397	−0.183	−0.113	−0.062	0.421	0.189	0.116	0.063	1.00	1.00	1.00	1.00
	5000	−0.396	−0.179	−0.111	−0.062	0.417	0.186	0.114	0.063	1.00	1.00	1.00	1.00
CCEPbc	25	−0.004	0.000	0.000	0.000	0.151	0.064	0.038	0.022	0.06	0.08	0.06	0.06
	100	−0.003	0.001	0.000	−0.001	0.100	0.031	0.017	0.011	0.08	0.04	0.04	0.06
	500	0.000	0.001	0.000	0.000	0.057	0.014	0.008	0.005	0.06	0.04	0.05	0.04
	5000	0.000	0.000	0.000	0.000	0.015	0.004	0.002	0.002	0.02	0.05	0.05	0.05
CCEPjk	25	0.027	0.037	0.028	0.014	0.358	0.124	0.074	0.037	0.40	0.31	0.30	0.25
	100	0.044	0.045	0.032	0.015	0.325	0.110	0.063	0.028	0.51	0.53	0.55	0.46
	500	0.031	0.036	0.031	0.016	0.315	0.108	0.058	0.025	0.63	0.72	0.78	0.73
	5000	0.045	0.040	0.035	0.016	0.312	0.105	0.059	0.024	0.66	0.84	0.92	0.90
FLSbc	25	−0.261	−0.067	−0.029	−0.012	0.276	0.089	0.054	0.033	0.37	0.04	0.04	0.04
	100	−0.271	−0.076	−0.038	−0.019	0.271	0.084	0.043	0.022	0.97	0.72	0.49	0.27
	500	−0.280	−0.079	−0.038	−0.018	0.270	0.083	0.041	0.020	0.99	0.99	1.00	0.98
	5000	−0.283	−0.077	−0.037	−0.018	0.270	0.081	0.040	0.019	1.00	1.00	1.00	1.00
Results for $\widehat{\beta}$													
CCEP	25	−0.033	−0.011	−0.006	−0.002	0.058	0.033	0.024	0.018	0.09	0.07	0.06	0.06
	100	−0.033	−0.010	−0.005	−0.002	0.042	0.018	0.013	0.009	0.29	0.11	0.07	0.06
	500	−0.033	−0.011	−0.005	−0.002	0.038	0.014	0.008	0.004	0.73	0.40	0.18	0.09
	5000	−0.033	−0.010	−0.005	−0.002	0.036	0.012	0.005	0.002	0.97	0.94	0.76	0.33
CCEPbc	25	−0.001	0.000	−0.001	0.000	0.052	0.031	0.024	0.018	0.04	0.05	0.07	0.06
	100	0.001	0.000	0.000	0.000	0.026	0.015	0.012	0.009	0.05	0.05	0.05	0.06
	500	0.000	0.000	0.000	0.000	0.012	0.007	0.005	0.004	0.04	0.05	0.05	0.05
	5000	0.000	0.000	0.000	0.000	0.004	0.002	0.002	0.001	0.04	0.06	0.04	0.06
CCEPjk	25	0.014	0.012	0.006	0.003	0.097	0.041	0.029	0.020	0.25	0.13	0.11	0.08
	100	0.018	0.013	0.008	0.003	0.056	0.025	0.016	0.011	0.33	0.22	0.14	0.10
	500	0.019	0.011	0.008	0.003	0.044	0.017	0.011	0.006	0.52	0.43	0.35	0.17
	5000	0.019	0.012	0.008	0.004	0.041	0.016	0.009	0.004	0.69	0.81	0.84	0.70
FLSbc	25	−0.016	0.001	0.000	0.001	0.051	0.036	0.029	0.022	0.04	0.03	0.03	0.02
	100	−0.021	−0.003	−0.002	0.000	0.032	0.016	0.012	0.009	0.18	0.05	0.03	0.03
	500	−0.023	−0.004	−0.002	−0.001	0.027	0.009	0.006	0.004	0.62	0.15	0.08	0.06
	5000	−0.022	−0.004	−0.002	−0.001	0.025	0.006	0.003	0.001	0.89	0.54	0.25	0.09

NOTES: (i) Reported are simulation results for estimating ρ and β in the baseline case ($\rho = 0.8$, $\beta = 0.2$, $\lambda = 0$, $m = 1$). The factor has a contribution to the variance of the dependent variable that is equal to that of the idiosyncratic errors (RI = 1). (ii) CCEPbc is the bias-corrected CCEP estimator. CCEPjk is the jackknife CCEP correction and FLSbc is the bias-adjusted least squares with interactive effects estimator supplied with the correct number of factors ($m+1$). CCEP estimators do not use $\bar{g}_t$ and include no lags of $\bar{x}_t$. (iii) The size column reports actual test size for t-tests based on bootstrap standard errors estimated with 150 bootstrap samples.

yields a notable improvement in the finite N performance of the CCEPbc approach in the lower right panel of Table 2. This is further demonstrated in the lower panel of Table 3 which shows that the CCEPbc(+g) estimator suffers less bias compared to CCEPbc and has an adequate actual size for all combinations of N and T.

The above discussion shows that additional covariates can have a beneficial effect on CCE-type estimators when factors are very influential in the model, even in cases where the rank condition is already satisfied. However, comparing the bias of the CCEP estimator to that of CCEP(+g) in Table 2 also confirms our theoretical finding that adding more CSA to the orthogonalization matrix increases the bias of the uncorrected CCEP estimator. Fortunately, the CCEPbc adjustment is effective in removing this bias. For less influential factors (RI = 1) the only downside is a relative loss in efficiency compared to not using g_{it}. Finally, comparing CCEP(+g) over different factor strengths confirms our claim (see discussion in Theorem 1) that more influential factors (i.e., increasing RI from 1 to 3) do not change the bias in the one factor case (upper

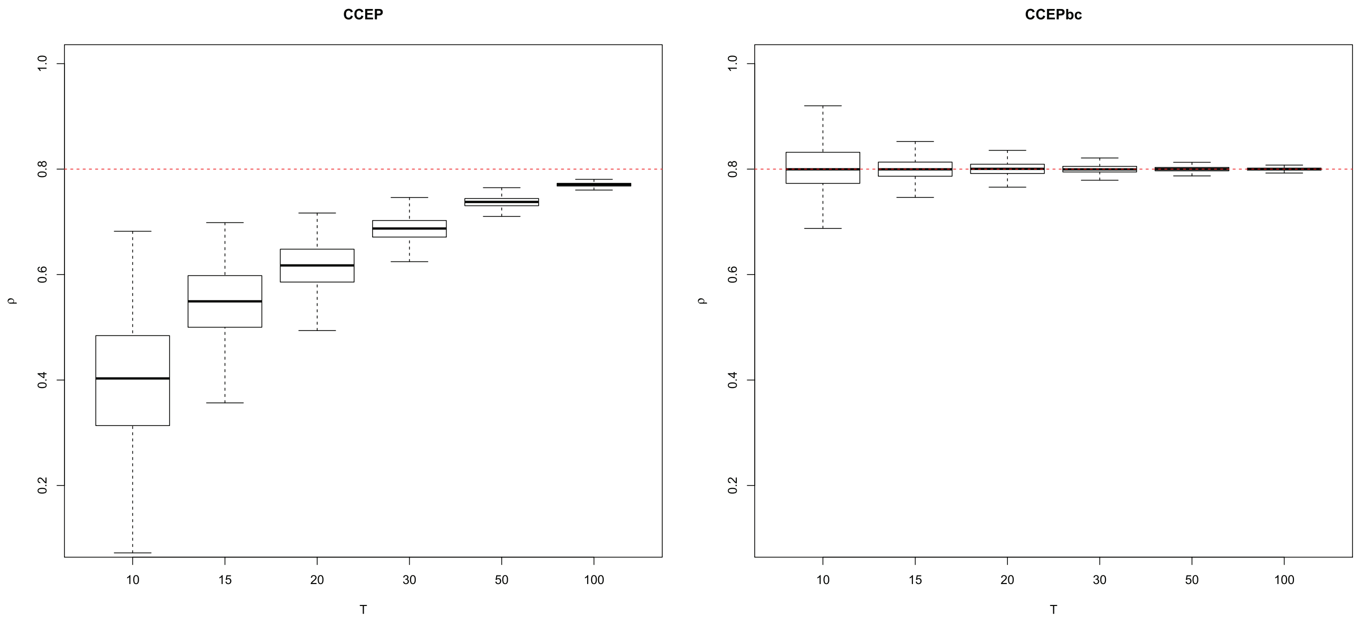

Figure 1. Monte Carlo results for ρ: comparison of CCEP and CCEPbc over T for $N = 500$. NOTES: Reported are simulation results for estimating ρ in the baseline case when $N = 500$ (see notes in Table 1). Dotted red lines indicate the population parameter value ($\rho = 0.8$). The boxplot "whiskers" extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box.

Table 2. Monte Carlo results for ρ: number and strength of factors ($N = 500$)

	Bias	RMSE	Size	Bias	RMSE	Size	Bias	RMSE	Size	Bias	RMSE	Size	Bias	RMSE	Size	Bias	RMSE	Size
	One factor																	
	RI = 1									RI = 3								
	$T = 10$			$T = 20$			$T = 30$			$T = 10$			$T = 20$			$T = 30$		
CCEP	−0.397	0.421	1.00	−0.183	0.189	1.00	−0.113	0.116	1.00	−0.399	0.424	1.00	−0.182	0.189	1.00	−0.112	0.116	1.00
CCEPbc	0.000	0.057	0.06	0.001	0.014	0.04	0.000	0.008	0.05	0.001	0.055	0.06	0.000	0.014	0.04	0.000	0.008	0.04
CCEPjk	0.031	0.315	0.63	0.036	0.108	0.72	0.031	0.058	0.78	0.037	0.312	0.63	0.040	0.108	0.73	0.031	0.059	0.79
CCEP(+g)	−0.406	0.431	1.00	−0.183	0.191	1.00	−0.113	0.116	1.00	−0.408	0.433	1.00	−0.183	0.190	1.00	−0.112	0.116	1.00
CCEPbc(+g)	0.000	0.066	0.06	0.000	0.014	0.02	0.000	0.008	0.04	0.000	0.064	0.06	0.000	0.014	0.03	0.000	0.008	0.04
CCEPjk(+g)	–	–	–	0.041	0.116	0.41	0.033	0.060	0.53	–	–	–	0.044	0.116	0.42	0.033	0.061	0.56
FLSbc	−0.280	0.270	0.99	−0.079	0.083	0.99	−0.038	0.041	1.00	−0.259	0.248	0.99	−0.064	0.072	1.00	−0.035	0.038	1.00
	Two factors																	
	RI = 1									RI = 3								
	$T = 10$			$T = 20$			$T = 30$			$T = 10$			$T = 20$			$T = 30$		
CCEP	−0.408	0.431	1.00	−0.190	0.197	1.00	−0.117	0.120	1.00	−0.402	0.425	1.00	−0.181	0.188	1.00	−0.109	0.112	1.00
CCEPbc	0.000	0.058	0.06	0.000	0.014	0.04	0.000	0.008	0.04	0.006	0.070	0.10	0.006	0.020	0.08	0.005	0.012	0.12
CCEPjk	0.035	0.311	0.62	0.034	0.107	0.72	0.031	0.059	0.79	0.034	0.316	0.64	0.045	0.116	0.73	0.037	0.063	0.76
CCEP(+g)	−0.447	0.471	1.00	−0.209	0.216	1.00	−0.128	0.131	1.00	−0.416	0.443	1.00	−0.190	0.196	1.00	−0.115	0.118	1.00
CCEPbc(+g)	0.000	0.070	0.08	0.000	0.016	0.03	0.000	0.008	0.05	0.000	0.067	0.06	0.001	0.014	0.04	0.000	0.008	0.05
CCEPjk(+g)	–	–	–	0.036	0.122	0.50	0.034	0.067	0.60	–	–	–	0.049	0.124	0.52	0.035	0.064	0.60
FLSbc	−0.532	0.525	1.00	−0.204	0.199	1.00	−0.097	0.098	1.00	−0.518	0.504	1.00	−0.170	0.170	1.00	−0.066	0.071	0.98

NOTES: (i) Data for this experiment are generated with $\rho = 0.8$, $\beta = 0.2$, and $\lambda = 0$. RI $= (1, 3)$ represents factors that have a contribution to the total variance of the dependent variable that is equal to, or, respectively, 3 times that of the idiosyncratic errors. We display results for estimating ρ with $N = 500$. (ii) CCEP is the pooled CCE estimator and CCEPbc its bias-corrected version. CCEPjk represents the jackknife corrected CCEP and FLSbc is the bias-adjusted least squares with interactive effects estimator supplied with the correct number of factors ($m + 1$). CCEP-type estimators with suffix "(+g)" indicate that $\bar{g}_t$ was included in the orthogonalization matrix. No lags of $\bar{x}_t$ and $\bar{g}_t$ are employed. (iii) The reported actual test size (size) is for a t-test using bootstrap standard errors based on 150 samples.

panel) but it will reduce the bias when more than one factor is present (lower panel).

5.4. Dynamics in $\mathbf{z}_{it}$

In this section, we allow for dynamics in $\mathbf{z}_{it}$ (setting $\lambda = 0.6$) to analyze the importance of including lagged CSA to adequately capture the common factors. Table 4 reports the main results in a setting where factors are strong (RI $= 3$) and $N = 500$. Results for $N = 25$ are reported in Table E-3 of the online supplement. We let CCEP and CCEPbc with suffix notation $\_p_1$ denote the estimators that are correctly specified with one lag of $\bar{\mathbf{Z}} = [\bar{\mathbf{X}}, \bar{\mathbf{G}}]$ added to the orthogonal projection matrix $\mathbf{M}$. The suffix notation $\_p_T$ is used

Table 3. Monte Carlo results for ρ: CCEPbc estimators with two highly influential factors

	Bias					Size				
(N, T)	10	20	30	50	100	10	20	30	50	100
					CCEPbc					
25	0.014	0.019	0.017	0.017	0.016	0.08	0.12	0.09	0.16	0.27
100	0.012	0.013	0.012	0.012	0.011	0.13	0.10	0.14	0.24	0.42
500	0.006	0.006	0.005	0.005	0.004	0.10	0.08	0.12	0.17	0.28
5000	0.001	0.001	0.001	0.001	0.001	0.04	0.06	0.06	0.06	0.08
					CCEPbc(+g)					
25	0.006	0.008	0.006	0.006	0.006	0.05	0.08	0.05	0.06	0.09
100	0.004	0.003	0.002	0.002	0.002	0.07	0.04	0.05	0.06	0.06
500	0.000	0.001	0.000	0.001	0.000	0.06	0.04	0.05	0.05	0.05
5000	0.000	0.000	0.000	0.000	0.000	0.02	0.05	0.05	0.05	0.05

NOTES: (i) Reported are simulation results for estimation and inference on the ρ coefficient. Data for this experiment are generated with $\rho = 0.8$, $\beta = 0.2$, $m = 2$, and $\lambda = 0$. Factors have a contribution to the total variance of the dependent variable that is 3 times that of the idiosyncratic errors (RI = 3). (ii) CCEPbc is the unrestricted corrected CCEP estimator. The "(+g)" indicates that $\bar{g}_t$ was included in the orthogonalization matrix. No lags of $\bar{x}_t$ and $\bar{g}_t$ are used. (iii) The test size (size) is for a t-test using bootstrap standard errors based on 150 samples.

Table 4. Monte Carlo results for ρ: dynamics in $\mathbf{z}_{it}$ with strong factors ($N = 500$)

	Bias	RMSE	Size	Bias	RMSE	Size	Bias	RMSE	Size	Bias	RMSE	Size
							One factor					
		$T = 10$			$T = 20$			$T = 30$			$T = 50$	
CCEP_p_0(+g)	−0.600	0.610	0.99	−0.253	0.261	1.00	−0.146	0.150	1.00	−0.076	0.078	1.00
CCEP_p_1(+g)	−0.685	0.713	0.95	−0.271	0.280	1.00	−0.152	0.157	1.00	−0.078	0.079	1.00
CCEP_p_T(+g)	–	–	–	−0.336	0.349	0.99	−0.203	0.210	1.00	−0.091	0.093	1.00
CCEPbc_p_0(+g)	0.000	0.090	0.05	0.000	0.017	0.03	−0.001	0.009	0.04	0.000	0.005	0.04
CCEPbc_p_1(+g)	−0.001	0.140	0.03	0.001	0.020	0.02	0.000	0.009	0.03	0.000	0.005	0.04
CCEPbc_p_T(+g)	–	–	–	0.001	0.029	0.02	−0.001	0.013	0.03	0.000	0.006	0.03
CCEPjk_p_1(+g)	–	–	–	0.140	0.227	0.27	0.088	0.124	0.43	0.039	0.050	0.58
FLSbc	−0.269	0.257	0.98	−0.058	0.065	0.99	−0.029	0.032	0.99	−0.014	0.015	0.95
							Two factors					
		$T = 10$			$T = 20$			$T = 30$			$T = 50$	
CCEP_p_0(+g)	−0.655	0.661	1.00	−0.290	0.296	1.00	−0.170	0.174	1.00	−0.091	0.092	1.00
CCEP_p_1(+g)	−0.779	0.794	0.99	−0.320	0.328	1.00	−0.179	0.183	1.00	−0.089	0.090	1.00
CCEP_p_T(+g)	–	–	–	−0.400	0.409	1.00	−0.246	0.251	1.00	−0.105	0.107	1.00
CCEPbc_p_0(+g)	−0.034	0.096	0.13	−0.018	0.028	0.24	−0.013	0.017	0.35	−0.009	0.011	0.45
CCEPbc_p_1(+g)	0.007	0.155	0.06	0.000	0.021	0.03	−0.001	0.010	0.04	0.000	0.005	0.05
CCEPbc_p_T(+g)	–	–	–	0.001	0.033	0.03	−0.001	0.014	0.04	0.000	0.006	0.05
CCEPjk_p_1(+g)	–	–	–	0.134	0.233	0.48	0.106	0.146	0.71	0.050	0.062	0.81
FLSbc	−0.528	0.519	1.00	−0.174	0.172	1.00	−0.069	0.073	0.99	−0.021	0.023	0.97

NOTES: (i) Reported are simulation results for estimating the ρ coefficient. Data for this experiment are generated with $\rho = 0.8$, $\beta = 0.2$, and $\lambda = 0.6$. The contribution of the factors to the total variance of the dependent variable is 3 times that of the idiosyncratic errors (RI = 3). We display results for estimating ρ with $N = 500$. (ii) CCEP is the pooled CCE estimator and CCEPbc its unrestricted bias-correction. CCEPjk represents the jacknife corrected CCEP and FLSbc is the bias-corrected least squares with interactive effects estimator supplied with the correct number of factors $(m + 1)$. All CCEP estimators additionally include $\bar{g}_t$ to project out the factors. CCEP estimators with a p_0, p_1, or p_T suffix, respectively, include no, one or $\lfloor T^{1/3} \rfloor$ lags of $\bar{x}_t$ and $\bar{g}_t$ in the orthogonalization matrix. (iii) The reported test size (size) is for a t-test using bootstrap standard errors based on 150 samples.

to indicate the inclusion of $p_T = \lfloor T^{1/3} \rfloor$ lags while _p_0 denotes the misspecified variant without lags of $\bar{\mathbf{Z}}$. We report results for CCEP-type estimators that add the CSA of g_{it} to avoid that the results are driven by using an insufficient number of covariates to proxy for the common factors. The correctly specified FLSbc and jackknife correction are included as alternative estimators. Note that some estimators cannot be implemented when $T = 10$ due to insufficient degrees of freedom (because of the larger number of CSA used for orthogonalization).

The simulation results for the misspecified CCEPbc_p_0 estimator reveal that it performs well when $m = 1$ but that it is not correctly centered when $m = 2$, despite the use of $\bar{g}_t$. Especially when T is large, the bias that remains in the latter case results in large size distortions. This suggests that the lag of $\bar{y}_t$ holds enough information to deal with the unobserved components in the single factor case but that it is not sufficient to control for multiple strong factors without lags of $\bar{x}_t$ (and $\bar{g}_t$). The correctly specified CCEPbc_p_1 estimator instead performs much better, with an adequate size for all values of T. This confirms that the approximation of the factors requires the number of lagged CSA to be equal to the AR lag order (p) of the exogenous variables. When p is unknown, we have suggested to follow the approach of Chudik and Pesaran (2015) and specify the number of lags as $p^* = \lfloor T^{1/3} \rceil$ to let them grow with T as a precaution against misspecification. As this implies orthogonalization on a large number of CSA, the resulting bias of the uncorrected CCEP_p_T estimator is very large. CCEPbc_p_T is however highly effective in removing the distortions and has an adequate size. The price paid for this robustness is that the larger number of CSA translates in a substantially higher variance compared to the correctly specified CCEPbc_p_1. As expected, this difference disappears as T grows. Results for small N (see Table E-3 in the online supplement) are highly similar (with marginally larger biases) but whenever bias remains it has a much smaller impact on inference.

6. TEMPERATURE SHOCKS AND ECONOMIC GROWTH

In this section, we apply our bias-corrected CCEP estimator to identify the dynamic effects of temperature shocks on aggregate output growth. In line with the recent literature (see, e.g., Dell, Jones, and Olken 2012; Colacito, Hoffmann, and Phan 2018) we consider the benchmark dynamic model

$$g_{it} = \alpha_i + \rho g_{i,t-1} + \beta_1 T_{it} + \beta_2 T_{i,t-1} + u_{it}, \qquad (27)$$

where g_{it} is per-capita real output growth and T_{it} is temperature. The lagged dependent variable $g_{i,t-1}$ is included to capture output growth persistence, while lagged temperature $T_{i,t-1}$ is added to discriminate between permanent and transitory output effects. The contemporaneous impact of a transitory 1°C rise in temperature on output growth is measured by β_1. If $\beta_2 = -\beta_1$, the impact on output growth is reversed in the next period (or periods if $\rho > 0$) such that the level of output (eventually) bounces back, that is, the cumulative growth effect $(\beta_1 + \beta_2)/(1 - \rho) = 0$. There is no (complete) reversal if $\beta_2 \neq -\beta_1$, which implies that the level of output is permanently affected by a transitory temperature shock. Typically no additional variables are included because most economic and political variables are potentially affected by weather variables, such that including them as controls implies that the estimates do not capture all relevant channels through which weather affects the economy.

The strategy in the recent climate-economy literature (see Dell, Jones, and Olken 2014 for an overview) is to exploit random variation in weather events over time within countries to identify its causal effects. Country fixed effects α_i are included to isolate weather effects from time-invariant characteristics, while time fixed effects (possibly interacted with region dummies) are added to neutralize common shocks. In panels with a relatively short time span, the latter avoids that the estimates pick up spurious correlation between global trends in weather and growth. However, time fixed effects impose a homogenous reaction (within regions) to common shocks. We allow for a more general heterogeneous response by letting u_{it} take the multifactor structure specified in Equation (2).

Data are taken from Dell, Jones, and Olken (2012), who have collected yearly output growth and annual average temperatures for an unbalanced panel of 125 countries over the period 1961–2003. We follow their approach of allowing the temperature effects to be different for "rich" and "poor" countries (defined as having above, respectively, below-median PPP-adjusted per capita GDP). We further split the time dimension into two subperiods since weather effects may have become either larger (due to intensification) or smaller (due to adaptation) in recent years (Dell, Jones, and Olken 2014). This results in a balanced sample of 93 countries over the period 1962–1982 and 118 countries over the period 1983–2003. As this makes the time series dimension relatively short ($T = 21$), at least much smaller than N, this is the ideal setting to illustrate our CCEPbc estimator.

Estimation results are presented in Table 5. Beginning with the left panel for the first part of the sample, the FE estimates in column (1) confirm the finding of Dell, Jones, and Olken (2012) that temperature shocks have a significantly negative effect on output growth only in poor countries, where a transitory 1°C rise in temperature reduces output growth in the same year by about 2 percentage points. Moreover, output does not significantly bounce back in the year after the shock, resulting in a 1.67% permanent decrease in output. The CCEP estimates in column (2) show a highly similar contemporaneous impact, but the coefficient on $T_{i,t-1}$ increases substantially, even to the extent that temperature shocks only have a temporary impact on output. The bounce-back effect is, however, only significant at the 10% level of significance. Theorem 1 implies that the CCEP estimates of ρ and β_2 are expected to be downward biased, while β_1 should be unbiased. This is because $g_{i,t-1}$ is not correlated with future temperature shocks (which show no significant persistence) and negatively correlated with current shocks, such that the CCEP estimates $\boldsymbol{\zeta}$ of $g_{i,t-1}$ on T_{it} and $T_{i,t-1}$ in Equation (17) are expected to show a zero value for T_{it} and a negative value for $T_{i,t-1}$. The CCEPbc estimation results reported in column (3) indeed show an upward adjustment of the coefficients on $g_{i,t-1}$ and $T_{i,t-1}$. In particular, the coefficient on $T_{i,t-1}$ turns significant at the 5% level of significance, reinforcing the finding that temperature shocks only have a transitory impact on output.

Turning to the results for the second part of the sample reported in columns (4)–(6), temperature shocks again only have a negative impact in poor countries, but this now turns out to be more moderate. The contemporaneous impact decreases from roughly −2 to around −1.2, suggesting that there may be some adaptation in more recent years. Note that the FE estimates now reveal a significant bounce-back effect while this is not the case for the CCEP and CCEPbc estimators. However, the three estimators agree that there is no significant permanent impact of temperature shocks on output. Concerning our bias-correction method, it is interesting to note that the CCEP estimate for ρ reported in column (5) is only 0.07 and not significant, while its bias-corrected estimate in column (6) is

Table 5. Temperature shocks and economic growth

	$N = 93$, 1962–1982			$N = 118$, 1983–2003		
	FE	CCEP	CCEPbc	FE	CCEP	CCEPbc
	(1)	(2)	(3)	(4)	(5)	(6)
$g_{i,t-1}$	0.17	0.15	0.24	0.18	0.07	0.22
	(0.07)**	(0.08)*	(0.08)***	(0.08)**	(0.06)	(0.07)***
Rich countries						
T_{it}	0.17	0.47	0.48	0.13	0.47	0.44
	(0.45)	(0.53)	(0.52)	(0.20)	(0.39)	(0.37)
$T_{i,t-1}$	−0.30	−0.35	−0.39	0.44	0.09	0.08
	(0.33)	(0.55)	(0.54)	(0.22)**	(0.34)	(0.32)
Poor countries						
T_{it}	−2.08	−1.94	−1.93	−1.26	−1.11	−1.24
	(0.57)***	(0.79)**	(0.80)**	(0.45)***	(0.66)*	(0.68)*
$T_{i,t-1}$	0.69	1.76	1.84	0.83	0.30	0.57
	(0.69)	(0.91)*	(0.92)**	(0.33)**	(0.66)	(0.68)
Implied cumulative growth effects	$(\beta_1 + \beta_2)/(1 - \rho)$					
Rich countries	−0.16	0.14	0.12	0.70	0.60	0.66
	(0.83)	(0.98)	(1.05)	(0.37)*	(0.64)	(0.68)
Poor countries	−1.67	−0.21	−0.12	−0.53	−0.87	−0.87
	(0.75)**	(1.17)	(1.27)	(0.44)	(0.85)	(0.94)

NOTES: The dependent variable g_{it} is the growth rate of per-capita real GDP, T_{it} is the average annual temperature. Both subsamples include a balanced sample of countries. The FE specifications include country, region × year, and poor × year fixed effects (see Dell, Jones, and Olken 2012 for region compositions). Rich and poor are defined as countries having above, respectively, below-median PPP-adjusted per capita GDP in the first year of the sample. The CCEP estimators use the contemporaneous and one-year lagged CSA of g_{it}, rich × T_{it}, and poor × T_{it}. Bootstrapped standard deviations are reported in brackets. ***/**/* denote significance at the 1%/5%/10% level, respectively.

0.22 and highly significant. Moreover, the coefficient on $T_{i,t-1}$ roughly doubles when bias-correcting the CCEP estimator, but given the relatively large standard error it is not significant.

7. CONCLUSION

In this article, we extend the CCEP estimator designed by Pesaran (2006) to dynamic homogenous panel data models and develop a bias-corrected version that eliminates its finite T bias. We first show that in homogenous dynamic panels, the unobserved common factors can be effectively approximated by CSA of the observed data provided that a sufficient number of observables is available (rank condition) and an appropriate number of lagged CSA is added to the model. This number of lags should coincide with the autoregressive order of the observed data. We next derived the asymptotic bias expression for $N \to \infty$ of the CCEP estimator and used this to devise a bias-corrected estimator. We show that the resulting CCEPbc estimator is consistent as $N \to \infty$, both for T fixed or $T \to \infty$.

Extensive Monte Carlo experiments show that, when appropriately specified, CCEPbc performs very well and is superior to the original CCEP estimator and to alternative corrections available in the literature. More specifically, CCEPbc is found to be nearly unbiased across all of the sample sizes and designs we considered. Hence, it offers a strong improvement over the severely biased CCEP estimator. This is especially the case when T is small but even holds true for large T. Interestingly, CCEPbc also provides a notable variance reduction compared to the original CCEP estimator. This is due to the fact that the stochastic bias of the latter also drives up its variance. Moreover, using bootstrapped standard errors, the actual size of CCEPbc was found to be close to the 5% nominal level. The Monte Carlo simulations further show that it is important to include a sufficient number of CSA of observables in the model. First, the number of observables is important to satisfy the rank condition, but even when this already holds it is beneficial in terms of bias correction and inference to add CSA of additional observables when these hold information about highly influential common factors. Second, the simulation results confirm our theoretical finding that lagged CSA should be added to the model in line with the autoregressive order of the observables. In case the autoregressive order is unknown, letting the number of lags grow with T was found to be a robust approach.

SUPPLEMENTARY MATERIALS

The supplemental material includes an appendix with (i) derivations, lemmas and proofs, (ii) a detailed analysis of the asymptotic bias expression presented in Theorem 1, (iii) two restricted CCEPbc estimators and (iv) additional Monte Carlo simulation results.

ACKNOWLEDGMENTS

The authors are grateful to the editor, associate editor, and two anonymous referees for their comments that improved the article. In addition, we thank Joakim Westerlund, Alexander Chudik, Kazuhiko Hayakawa, Stijn Vansteelandt, and Bart Cockx for their insightful comments and suggestions as well as participants of the 22nd Panel Data Conference, the 3rd Conference of the International Association for Applied Econometrics, and the 2016 Asian and European Summer Meetings of the Econometric Society. This article also benefited from seminar presentations at the Erasmus School of Economics, the UNSW Business School, the ANU Research School of Economics, and the Monash Business School.

FUNDING

The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the Flemish Supercomputer Center, funded by Ghent University; the Hercules Foundation; and the Economy, Science, and Innovation Department of the Flemish Government. Ignace De Vos gratefully acknowledges financial support from the Ghent University BOF research fund and the Research Foundation Flanders (FWO). Ignace De Vos and Gerdie Everaert further acknowledge financial support from the National Bank of Belgium.

[Received January 2018. Revised March 2019.]

REFERENCES

Ahn, S. C., Lee, Y. H., and Schmidt, P. (2001), "GMM Estimation of Linear Panel Data Models With Time-Varying Individual Effects," *Journal of Econometrics*, 101, 219–255. [1]

Andrews, D. W. K. (2005), "Cross-Section Regression With Common Shocks," *Econometrica*, 73, 1551–1585. [1]

Bai, J. (2009), "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279. [7]

Bun, M., and Carree, M. (2005), "Bias-Corrected Estimation in Dynamic Panel Data Models," *Journal of Business and Economic Statistics*, 23, 200–210. [5]

Carlsson, M., Eriksson, S., and Gottfries, N. (2013), "Product Market Imperfections and Employment Dynamics," *Oxford Economic Papers*, 65, 447–470. [2]

Chudik, A., Pesaran, M., and Tosetti, E. (2011), "Weak and Strong Cross-Section Dependence and Estimation of Large Panels," *The Econometrics Journal*, 14, C45–C90. [1]

Chudik, A., and Pesaran, M. H. (2015), "Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models With Weakly Exogenous Regressors," *Journal of Econometrics*, 188, 393–420. [2,3,4,7,11]

——— (2017), "A Bias-Corrected Method of Moments Approach to Estimation of Dynamic Short-T Panels," USC-INET Research Paper 17-26, USC Dornsife Institute for New Economic Thinking. [5]

Colacito, R., Hoffmann, B., and Phan, T. (2018), "Temperatures and Growth: A Panel Analysis of the United States," Working Paper 18-09, Federal Reserve Bank of Richmond. [11]

Dell, M., Jones, B., and Olken, B. (2012), "Temperature Shocks and Economic Growth: Evidence From the Last Half Century," *American Economic Journal: Macroeconomics*, 4, 66–95. [11,12]

——— (2014), "What Do We Learn From the Weather? The New Climate-Economy Literature," *Journal of Economic Literature*, 52, 740–798. [11]

Dhaene, G., and Jochmans, K. (2015), "Split-Panel Jackknife Estimation of Fixed-Effect Models," *Review of Economic Studies*, 82, 991–1030. [2,7]

Eberhardt, M., and Teal, F. (2011), "Econometrics for Grumblers: A New Look at the Literature on Cross-Country Growth Empirics," *Journal of Economic Surveys*, 25, 109–155. [1]

Eriksson, S., and Stadin, K. (2017), "What Are the Determinants of Hiring? The Importance of Product Market Demand and Search Frictions," *Applied Economics*, 49, 5144–5165. [2]

Everaert, G., and De Groote, T. (2016), "Common Correlated Effects Estimation of Dynamic Panels With Cross-Sectional Dependence," *Econometric Reviews*, 35, 428–463. [2,4]

Harding, M., and Lamarche, C. (2011), "Least Squares Estimation of a Panel Data Model With Multifactor Error Structure and Endogenous Covariates," *Economics Letters*, 111, 197–199. [1]

Kapetanios, G. (2008), "A Bootstrap Procedure for Panel Data Sets With Many Cross-Sectional Units," *Econometrics Journal*, 11, 377–395. [5,6]

Kapetanios, G., Pesaran, M., and Yamagata, T. (2011), "Panels With Non-Stationary Multifactor Error Structures," *Journal of Econometrics*, 160, 326–348. [1]

Karabiyik, H., Reese, S., and Westerlund, J. (2017), "On the Role of the Rank Condition in CCE Estimation of Factor-Augmented Panel Regressions," *Journal of Econometrics*, 197, 60–64. [5]

Madsena, J., Minnitib, A., and Venturini, F. (2018), "Assessing Pikettys Second Law of Capitalism," *Oxford Economic Papers*, 70, 1–21. [2]

Minniti, A., and Venturini, F. (2017), "The Long-Run Growth Effects of R&D Policy," *Research Policy*, 46, 316–326. [2]

Moon, H. R. M., and Perron, B. (2007), "An Empirical Analysis of Nonstationarity in a Panel of Interest Rates With Factors," *Journal of Applied Econometrics*, 22, 383–400. [1]

Moon, H. R., and Weidner, M. (2017), "Dynamic Linear Panel Regression Models With Interactive Fixed Effects," *Econometric Theory*, 33, 158–195. [2,7]

Nickell, S. (1981), "Biases in Dynamic Models With Fixed Effects," *Econometrica*, 49, 1417–1426. [2]

Pesaran, M. (2006), "Estimation and Inference in Large Heterogeneous Panels With a Multifactor Error Structure," *Econometrica*, 74, 967–1012. [1,3,5,12]

Pesaran, M., Smith, L., and Yamagata, T. (2013), "Panel Unit Root Tests in the Presence of a Multifactor Error Structure," *Journal of Econometrics*, 175, 94–115. [2]

Pesaran, M., and Tosetti, E. (2011), "Large Panels With Common Factors and Spatial Correlation," *Journal of Econometrics*, 161, 182–202. [1]

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

Sarafidis, V., and Wansbeek, T. (2012), "Cross-Sectional Dependence in Panel Data Analysis," *Econometric Reviews*, 31, 483–531. [1]

So, B., and Shin, D. (1999), "Recursive Mean Adjustment in Time-Series Inferences," *Statistics & Probability Letters*, 43, 65–73. [2,7]

Tao, X. (2018), "Investigating Environmental Kuznets Curve in China—Aggregation Bias and Policy Implications," *Energy Policy*, 114, 315–322. [2]

Temple, J., and Van de Sijpe, N. (2017), "Foreign Aid and Domestic Absorption," *Journal of International Economics*, 108, 431–443. [2]

Wu, J.-W., and Wu, J.-L. (2018), "Does the Launch of the Euro Hinder the Current Account Adjustment of the Eurozone?," *Economic Inquiry*, 56, 1116–1135. [2]