FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

Bias-corrected Common Correlated Effects Pooled estimation in homogeneous dynamic panels

Ignace De Vos

Gerdie Everaert

April 2016

2016/920

D/2016/7012/04

Bias-corrected Common Correlated Effects Pooled estimation in homogeneous dynamic panels

Ignace De Vos*
SHERPPA, UGent
Ghent, Belgium

Gerdie Everaert
SHERPPA, UGent
Ghent, Belgium

April 2016

Abstract

This paper extends the Common Correlated Effects Pooled (CCEP) estimator designed by Pesaran (2006) to dynamic homogeneous models. For static panels, this estimator is consistent as the number of cross-sections (N) goes to infinity irrespectively of the time series dimension (T). However, it suffers from a large bias in dynamic models when T is fixed (Everaert and De Groote, 2016). We develop a bias-corrected CCEP estimator based on an asymptotic bias expression that is valid for a multi-factor error structure provided that a sufficient number of cross-sectional averages, and lags thereof, are added to the model. We show that the resulting CCEPbc estimator is consistent as N tends to infinity, both for T fixed or T growing large, and derive its limiting distribution. Monte Carlo experiments show that our bias correction performs very well. It is nearly unbiased, even when T and/or N are small, and hence offers a strong improvement over the severely biased CCEP estimator. CCEPbc is also found to be superior to alternative bias correction methods available in the literature in terms of bias, variance and inference.
JEL-code: C23, C13, C15
Keywords: Dynamic panel data, bias, bias-correction, common correlated effects, unobserved common factors, cross-section dependence, lagged dependent variable

1 Introduction

In an influential paper, Pesaran (2006) proposed the Common Correlated Effects (CCE) approach to estimate panels with a multi-factor error structure. The procedure relies on augmenting the model with the cross-sectional averages of the observed variables such that asymptotically (as the cross-sectional dimension $N \to \infty$) the effect of unobserved common factors is eliminated. This may be critical in many practical settings as common factors in the error term result in

*Correspondence to: Ignace De Vos, Faculty of Economics and Business Administration, Ghent University, Sint-Pietersplein 6, 9000 Gent, Belgium. E-mail address: ignace.devos@ugent.be

misleading inference and may even induce endogeneity when these factors are correlated with the explanatory variables. Both the mean group (CCEMG) and the pooled (CCEP) version of the CCE approach are shown to be 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 a variety of more general settings such as unit roots in the factors (Kapetanios et al., 2011), spatial correlation in the idiosyncratic errors (Pesaran and Tosetti, 2011) or, under certain conditions, an infinite number of factors (Chudik et al., 2011) and it can easily be adjusted to account for endogenous regressors (Harding and Lamarche, 2011). The straightforwardness of the CCE approach in combination with its robustness seem to be very appealing to practitioners and has led to numerous applications in a variety of empirical settings, especially in a macroeconomic context.

The CCE approach is well developed in the static model but was originally not intended for use in dynamic settings. Dynamic models are however commonplace in economics where variables are often slow to react to changes in their determinants and hence display considerable persistence over time. Typically a lagged dependent variable is added to the specification as an easy way to account for these dynamics. However, this leads to new econometric challenges. A well-known result from Nickell (1981) is that in dynamic panel data regressions the Fixed Effects (FE) estimator is inconsistent as $N \to \infty$ but T is fixed. The presence of cross-sectional dependence further complicates matters. Everaert and De Groote (2016) show that for a temporally dependent factor the FE estimator is inconsistent even as both $N, T \to \infty$ while the CCEP estimator is inconsistent as $N \to \infty$ with T fixed. Moreover, the asymptotic bias of the CCEP estimator is shown to be considerably more sizable compared to the standard dynamic panel data bias of the FE estimator under cross-sectional independence. Especially in highly persistent panels, a notable bias remains even for moderately large T up to 50. The Monte Carlo simulations in Everaert and De Groote further show that the small sample properties of the CCEP estimator are not very sensitive to the size of N. Hence, in dynamic panels it is mainly the time series dimension that should be sufficiently large to allow for reliable CCEP estimation and inference. Similar results are obtained by Chudik and Pesaran (2015) for the CCEMG estimator in dynamic panels. In an attempt to correct for its fixed T bias, they suggest the recursive mean adjustment of So and Shin (1999) or the half-panel jackknife of Dhaene and Jochmans (2014). Although these approaches succeed in mitigating the bias, they are unable to fully resolve the issue for small values of T.

Most of the recent literature on estimating homogenous dynamic models with a common factor error structure takes the *micro panel* perspective where T is small and fixed relative to N which is large. The unobserved common factors can then be treated as unknown parameters but the factor loadings are interactive individual effects that induce an incidental parameter problem since their number grows with N. Using a Generalized Method of Moments (GMM) framework, several contributions transform the model with quasi-differencing techniques to remove the individual effects (Holtz-Eakin et al., 1988; Nauges and Thomas, 2003; Ahn et al., 2013) whereas Robertson and Sarafidis (2015) avoid incidental parameters by estimating loading covariances in stead of the loadings themselves. Using a Maximum Likelihood (ML) framework, a similar idea is used by Bai (2013) who further controls for correlation between the regressors and the factor loadings by using Chamberlain (1982) type projections. Hayakawa et al. (2014) suggest a transformed ML estimator applied to the dynamic panel data model after first-differencing. In contrast to the ML estimator on the levels suggested by Bai (2013), this allows the individual fixed effects to be correlated over cross-sections. Treating the factors as period-specific parameters is convenient when T is small but is troublesome when taking a *macro panel* perspective where T is moderate and may even be larger than N. As here the number of factor-related parameters grows in both N and T, an incidental parameters problem is present in both dimensions. Although Bai (2013) shows that despite the incidental parameter problem his ML estimator is still correctly centered when $N, T \to \infty$ it cannot be implemented in the presence of exogenous regressors when $T \geq N$ due to oversaturation.

In this paper we focus on the homogeneous dynamic panel model and derive a bias-corrected CCEP estimator (referred to as CCEPbc) based on a generalization of the analytical bias expression in Everaert and De Groote (2016) to settings allowing for multiple common factors and exogenous variables. Our CCEPbc estimator is mainly intended for macro panels where T is typically neither small nor large and often larger than N. However, because we don't rely on large T approximations of the asymptotic bias we expect our correction method to be accurate also in small T settings. In the spirit of the relative simplicity of the CCE approach the proposed bias correction is straightforward to implement in practice. We further show that in a homogenous dynamic model, the CCE approach is valid provided that the number of observables that are affected by the unobserved common factors is at least as large as the number of factors (rank condition) and a sufficient number of lagged cross-sectional averages is included. In contrast to the heterogeneous dynamic model in Chudik and Pesaran (2015), which requires the number of lags to grow with T, the requirement in a homogenous dynamic model is that the number of lagged cross-sectional averages for each of the observables matches with the autoregressive order of their data generating process. As these orders are typically unknown, a practical solution is to let the number of lags grow with T as in Chudik and Pesaran.

Most closely related to our paper is the least squares with interactive fixed effects (referred to as FLS) estimator and its bias-corrected version (FLSbc) of Moon and Weidner (2014). The FLS estimator was originally proposed by Bai (2009) for large N and T panels with strictly exogenous variables and strong[1] unobserved common factors. It is essentially a LS method that concentrates out the common factors by taking out a known number of principal components. Deciding on the number of factors can be difficult in practice and will lead to bias in case this number is set too low. Additionally, the possibility of fixed (time-invariant) effects further complicates the estimation procedure since these constants need to be regarded as an additional (restricted)

[1] In the sense that norms of the factors and their loadings grow at rates $\sqrt{T}$ and $\sqrt{N}$ respectively. This is required in order for the factors to be picked up as the main principal components. See Moon and Weidner (2014) for more details or Chudik et al. (2011) for a discussion on weak and strong factors.

factor. By relying on cross-sectional averages the CCE method avoids direct assumptions on the number of factors but in return requires assumptions on the matrix of factor loadings. The fixed effects however do not require special treatment since they are removed from the model by demeaning. Bai (2009) shows that the FLS estimator is consistent when $N, T \to \infty$ but may have an asymptotic bias for fixed T. Moon and Weidner (2014) generalize his approach to the dynamic panel data case and show that the presence of predetermined regressors induces an additional asymptotic bias for fixed T. They further provide a consistent estimator for the asymptotic bias that can be used to construct a bias-corrected FLS estimator.

Monte Carlo simulations show that our CCEPbc estimator provides considerable improvements (in terms of both bias and variance) over the original CCEP estimator and is virtually unbiased in all settings even for $T = 10$. Moreover, our CCEPbc estimator is found to be superior to both (i) alternative bias-corrected CCEP estimators using the jackknife or recursive mean adjustments as suggested by Chudik and Pesaran for the CCEMG estimator and (ii) the FLSbc estimator of Moon and Weidner (2014). Using bootstrap inference, the CCEPbc estimator has an actual size which is close to the desired nominal level even when T is small.

The remainder of this paper is structured as follows. Section 2 outlines the model and assumptions. In section 3 we extend the CCEP estimator to homogeneous dynamic panel models and derive expressions for its finite T inconsistency and use this in section 4 to construct a bias-corrected CCEP estimator. Monte Carlo simulation results are presented in Section 5. Section 6 concludes. Mathematical proofs are presented in Appendix A. Appendix B contains additional small sample simulation results for the experiments reported in Section 5.

Before proceeding we introduce some notation that will be used throughout the paper: For a $T \times k$ matrix $\mathbf{A}$, $\widetilde{\mathbf{A}}$ represents that matrix in deviation of its column means. The $T \times T$ matrix $\mathbf{H}_{\mathrm{A}} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$ is the projection matrix onto the space spanned by the columns of $\mathbf{A}$, with its orthogonal complement given by $\mathbf{M}_{\mathrm{A}} = \mathbf{I}_T - \mathbf{H}_{\mathrm{A}}$ and $(\mathbf{A}'\mathbf{A})^{-}$ the Moore-Penrose inverse of $\mathbf{A}'\mathbf{A}$. We will further use $\mathbf{H}_{\mathrm{A}}(i,j)$ to denote the (i,j)-th element in $\mathbf{H}_{\mathrm{A}}$ and $\mathrm{tr}(\mathbf{H}_{\mathrm{A}})$ for its trace. 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. $(N,T)_{\mathrm{seq.}} \to \infty$ denotes the sequential limit where first N and then T tends to infinity.

2 Model and assumptions

Consider the following first-order dynamic panel data model

$$\begin{aligned} y_{it} &= \alpha_i + \rho y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}, & (1)\\ u_{it} &= \boldsymbol{\gamma}_i'\mathbf{f}_t + \varepsilon_{it}, & (2) \end{aligned}$$

for $i = 1, ..., N$ and $t = 1, ..., T$ 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 u_{it} a multi-factor error term that is composed of a $(m \times 1)$ 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 homogeneous over cross-sections and bounded by a finite constant. For notational convenience we assume y_{i0} known.

Following Pesaran et al. (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 drive y_{it}. The individual-specific variables and covariates 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

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

where $\mathbf{c}_{\mathrm{z},i}$ is a $k \times 1$ column vector of unobserved individual effects, p denotes the autoregressive order of $\mathbf{z}_{it}$[2], $\boldsymbol{\lambda}_l$ is a $k \times k$ matrix of coefficients corresponding to lags $l = 1, ..., p$ of $\mathbf{z}_{it}$, $\boldsymbol{\Gamma}_i$ is a $m \times k$ matrix of factor loadings and $\mathbf{v}_{it}$ a $k \times 1$ column vector of idiosyncratic errors.

We make the following assumptions:

Assumption 1. (Idiosyncratic errors) The individual-specific error terms ε_{it} and $\mathbf{v}_{jt}$ are zero mean variables distributed independently across i and t and of each other for all i, j and t with finite moments up to the fourth order. In particular,

$$\varepsilon_{it} \sim \mathrm{IID}(0, \sigma^2_{\varepsilon,i}),$$
$$\mathbf{v}_{it} \sim \mathrm{IID}(0, \boldsymbol{\Omega}_{\mathrm{v},i}),$$

with both $\sigma^2_{\varepsilon,i}$ and $\boldsymbol{\Omega}_{\mathrm{v},i}$ bounded positive definite and $\mathrm{plim}_{N\to\infty}(1/N)\sum_{i=1}^{N}\sigma^2_{\varepsilon,i} = \sigma^2_{\varepsilon}$.

REMARK 1. Assumption 1 allows the individual-specific error terms ε_{it} to be either homoscedastic with common variance $\sigma^2_{\varepsilon,i} = \sigma^2_{\varepsilon}$ for all i, or cross-sectionally heteroscedastic with the condition that $\mathrm{plim}_{N\to\infty}(1/N)\sum_{i=1}^{N}\sigma^2_{\varepsilon,i} = \sigma^2_{\varepsilon}$. The latter implies that the cross-section specific variances $\sigma^2_{\varepsilon,i}$ are random draws distributed around a common mean σ^2_{ε}.

Assumption 2. (Common factors) The m common factors $\mathbf{f}_t$ are covariance stationary with absolute summable autocovariances and bounded fourth moments and are distributed independently of the individual-specific errors ε_{it} and $\mathbf{v}_{it}$ and of the factor loadings $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Gamma}_i$.

Assumption 3. (Factor loadings) The individual-specific factor loadings $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Gamma}_i$ are distributed independently across i and of the individual-specific errors ε_{it} and $\mathbf{v}_{jt}$ and of the common factors

[2] 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.

$\mathbf{f}_t$ for all i, j and t with finite fixed means $\boldsymbol{\gamma}$ and $\boldsymbol{\Gamma}$, respectively, and bounded second moments. In particular, we specify for the factor loadings in the process of y_{it},

$$\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_i, \qquad \boldsymbol{\eta}_i \sim \text{IID}(\mathbf{0}, \boldsymbol{\Omega}_\eta), \tag{4}$$

with $\boldsymbol{\Omega}_\eta$ a $m \times m$ bounded positive definite matrix.

Assumption 4. (Rank condition) The $(1+k) \times m$ matrix $\mathbf{C} = (\boldsymbol{\gamma}, \boldsymbol{\Gamma})'$ is of full column rank such that $\text{rank}(\mathbf{C}) = m \leq k+1$.

Assumption 5. (Stationarity) $|\rho| < 1$ and the 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, stacking the model in equation (1) over time we obtain

$$\mathbf{y}_i \;=\; \alpha_i \boldsymbol{\iota}_T + \rho \mathbf{y}_{i,-1} + \mathbf{X}_i \boldsymbol{\beta} + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \tag{5}$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, $\mathbf{y}_{i,-1} = (y_{i0}, y_{i1}, \ldots, y_{i,T-1})'$, $\boldsymbol{\iota}_T$ is a $(T \times 1)$ column vector of ones, $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT})'$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_T)'$ and $\boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}_{i1}, \ldots, \boldsymbol{\varepsilon}_{iT})'$. Similarly specify $\mathbf{G}_i = (\mathbf{g}_{i1}, \ldots, \mathbf{g}_{iT})'$ and $\mathbf{Z}_i = (\mathbf{X}_i, \mathbf{G}_i)$. Further stacking over individuals we have

$$\mathbf{y} \;=\; (\mathbf{I}_N \otimes \boldsymbol{\iota}_T)\,\boldsymbol{\alpha} + \rho \mathbf{y}_{-1} + \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_N \otimes \mathbf{F})\,\boldsymbol{\Lambda} + \boldsymbol{\varepsilon}, \tag{6}$$

with $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_N')'$, $\mathbf{X} = \left(\mathbf{X}_1', \ldots, \mathbf{X}_N'\right)'$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_N)'$, $\boldsymbol{\Lambda} = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2', ..., \boldsymbol{\gamma}_N')'$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2', ..., \boldsymbol{\varepsilon}_N')'$.

Based on Assumption 5 the expression in equation (6) can be inverted to obtain

$$\mathbf{y} \;=\; (\mathbf{I}_N \otimes \boldsymbol{\iota}_T)\,\boldsymbol{\alpha}^+ + \mathbf{X}^+ \boldsymbol{\beta} + \mathbb{F}^+ \boldsymbol{\Lambda} + \boldsymbol{\varepsilon}^+, \tag{7}$$

with $\mathbb{F}^+ = \left(\mathbf{I}_N \otimes \mathbf{F}^+\right)$ and variables with a $+$ superscript defined as $\mathbf{X}^+ = (1-\rho L)^{-1}\mathbf{X}$. The infinite sums $\boldsymbol{\varepsilon}^+$, $\mathbf{X}^+$ and $\mathbb{F}^+$ will be of particular interest for deriving the asymptotic bias expression of the CCEP estimator.

3 CCEP estimation in dynamic panels

Pesaran (2006) developed the CCE approach in a model with strictly exogenous regressors and showed that the differential effects of unobserved common factors can be eliminated as $N \to \infty$ by augmenting the model with the cross-sectional averages of the observed dependent and explanatory variables. The common factors will be effectively eliminated from the model when the rank condition that the number of observed variables that are related to the common factors is at least as large as the number of factors is satisfied. However, the consistency of the CCE estimators is not affected by a rank deficiency problem in the original Pesaran (2006) setting[3].

[3] In a model with strictly exogenous regressors, the required conditions for this result to hold are that the factor loadings $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\gamma}_i$ satisfy a random coefficient assumption and are independently distributed from each other and over individuals (see also Westerlund and Urbain, 2013; Karabiyik et al., 2014).

Chudik and Pesaran (2015) extend the CCE approach to heterogenous panel data models with a lagged dependent variable and/or weakly exogenous regressors. They show that eliminating the unobserved factors requires augmenting the model with an infinite number of lagged cross-sectional averages. This can be adequately approximated by letting their number grow with T, but at a slower rate. They show that the resulting CCE mean group estimator is consistent when both N and T tend to infinity and the rank condition is satisfied. Only when the factors are serially uncorrelated, the rank condition is not required for the consistency of their CCEMG. In this section, we extend the CCE methodology to homogeneous dynamic panels. We first review whether the cross-sectional averages of the observed data still serve as suitable proxies for the unobserved common factors. We next show that, in contrast to the static case, the CCEP estimator is inconsistent when $N \to \infty$ and T fixed and derive asymptotic bias expressions that will be used in Section 4 to construct bias-corrected CCEP estimators.

3.1 Cross-sectional averages as proxies for the unobserved common factors

Rewriting equations (1)-(3) as

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

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

$$\begin{aligned} \rho(L)\,\bar{\mathrm{y}}_t &= \bar{\alpha} + \bar{\mathbf{x}}_t'\boldsymbol{\beta} + \boldsymbol{\gamma}'\mathbf{f}_t + O_p\left(N^{-1/2}\right), & (8) \\ \boldsymbol{\lambda}(L)\,\bar{\mathbf{z}}_t &= \bar{\mathbf{c}}_z + \boldsymbol{\Gamma}'\mathbf{f}_t + O_p\left(N^{-1/2}\right), & (9) \end{aligned}$$

with the affix notation on $\bar{\mathrm{y}}_t$ used to denote the cross-sectional average of 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

$$\mathbf{f}_t = \left(\mathbf{C}'\mathbf{C}\right)^{-1}\mathbf{C}'\left(\begin{bmatrix} \rho(L) & -\boldsymbol{\beta}^{*\prime} \\ 0 & \boldsymbol{\lambda}(L) \end{bmatrix}\begin{bmatrix} \bar{\mathrm{y}}_t \\ \bar{\mathbf{z}}_t \end{bmatrix} - \begin{bmatrix} \bar{\alpha} \\ \bar{\mathbf{c}}_z \end{bmatrix}\right) + O_p\left(N^{-1/2}\right), \tag{10}$$

with $\boldsymbol{\beta}^* = (\boldsymbol{\beta}', \mathbf{0}_{k_g \times 1}')'$. Equation (10) shows that as $N \to \infty$ the factors can be effectively approximated by the cross-section averages of y_{it} and $\mathbf{z}_{it}$ as well as a finite number of lags thereof 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 with heterogeneous coefficients an infinite number of lags is required to eliminate the factors from the model.

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

$$\bar{y}_t = \bar{\alpha} + \bar{\gamma}\mathrm{f}_t + \bar{\varepsilon}_t + \rho\left(\frac{\bar{\alpha}}{1-\rho} + \bar{\gamma}\mathrm{f}^+_{t-1} + \bar{\varepsilon}^+_{t-1}\right), \tag{11}$$

$$= \frac{\bar{\alpha}}{1-\rho} + \bar{\gamma}\left[\mathrm{f}_t + \rho\mathrm{f}^+_{t-1}\right] + O_p\left(N^{-1/2}\right), \tag{12}$$

so that it is not only a function of the factors at time t, a constant and an $O_p\left(N^{-1/2}\right)$ term but also of the past realizations of the factors through $\mathrm{f}^+_{t-1} = \sum_{l=0}^{\infty}\rho^l\mathrm{f}_{t-l-1}$. Solving the contemporaneous factor f_t from (12) would therefore still depend on the unobservable f^+_{t-1} so a proxy can not be constructed from it. However, noting that the term inside the brackets of (11) equals $\bar{y}_{t-1}$, subtracting $\rho\bar{y}_{t-1}$ from (11) yields

$$\bar{y}_t - \rho\bar{y}_{t-1} = \bar{\alpha} + \bar{\gamma}\mathrm{f}_t + \bar{\varepsilon}_t, \tag{13}$$

so that the past factor realizations are cut out and this equation can be solved for f_t in function of observables and $O_p\left(N^{-1/2}\right)$ terms. 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 Pesaran (2006) where absence of dynamics implies $\rho = 0$ in equation (11) so that cross-sectional averages do not contain the past factors and lags are not required to separate them from f_t.

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, it is reasonable to expect that p is small and as more time series observations become available the factor approximation should not suffer from including too many lags $p_T > p$ of $\bar{\mathbf{z}}_t$. Hence, in practice it may be convenient to choose $p_T = T^{1/3}$ as in Chudik and Pesaran (2015) to make the CCEP estimator robust to misspecification of p while ensuring that the number of lags does not increase too fast in T such that sufficient degrees of freedom are available for consistent CCEP estimation.

In order for the suggested cross-sectional averages to be able to approximate the factors, the $(1+k) \times m$ matrix $\mathbf{C} = (\boldsymbol{\gamma}, \boldsymbol{\Gamma})'$ is required to be of full column rank, regardless of whether the regressors are strictly or weakly exogenous and of whether the parameters are homogenous or heterogeneous. If this is not the case, the $\mathbf{C}'\mathbf{C}$ matrix is not invertible implying that the cross-sectional averages are insufficient to approximate the factors. In practice, the rank condition translates into the requirement of having at least as much observables holding some linearly independent information about the unobserved factors as there are factors in the model. In the absence of sufficient covariates (k_x) within the model the researcher can add cross-section averages of an additional k_g variables $(\mathbf{g}_{it})$ from outside the model. Provided that these covariates hold new information about the factors $\mathbf{f}_t$ (i.e. their mean loading does not tend to zero and is sufficiently different from that of other variables) they can be used to make $\mathbf{C}$ a full column rank matrix. Even when $\mathbf{C}$ is already of full rank it may be advantageous to use extra variables because the additional information should improve the approximation of the factors. Hence, the strength of CCE estimation lies in the fact that the actual number of factors does not need to be specified as long as there are a sufficient number of variables available. Moreover, the approach is flexible in the sense that additional covariates can be used to approximate the factors without the need to alter the specification of interest.

3.2 Dynamic CCEP estimator

Considering the discussion in Subsection 3.1, the orthogonal projection matrix $\mathbf{M}_{\mathrm{Q}}$ is constructed setting $\mathbf{Q} = \left(\boldsymbol{\iota}_T, \bar{\mathbf{y}}, \bar{\mathbf{y}}_{-1}, \bar{\mathbf{Z}}, \bar{\mathbf{Z}}_{-1}, \ldots, \bar{\mathbf{Z}}_{-p_T}\right)$. With pooling weights set to N^{-1} the dynamic CCEP estimator for $\boldsymbol{\delta} = (\rho, \boldsymbol{\beta}')'$ is given by

$$\widehat{\boldsymbol{\delta}} = \begin{pmatrix} \widehat{\rho} \\ \widehat{\boldsymbol{\beta}} \end{pmatrix} = \left(\sum_{i=1}^{N} \mathbf{w}_i' \mathbf{M}_{\mathrm{Q}} \mathbf{w}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{w}_i' \mathbf{M}_{\mathrm{Q}} \mathbf{y}_i, \tag{14}$$

where $\mathbf{w}_i = \left(\mathbf{y}_{i,-1}, \mathbf{X}_i\right)$. Using equation (6), this estimator can conveniently be split up into separate expressions for ρ and $\boldsymbol{\beta}$

$$\widehat{\rho} = \left(\mathbf{y}_{-1}' \mathbb{M}_{\mathrm{XQ}} \mathbf{y}_{-1}\right)^{-1} \mathbf{y}_{-1}' \mathbb{M}_{\mathrm{XQ}} \mathbf{y}, \tag{15}$$

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}' \mathbb{M}_{\mathrm{Q}} \mathbf{X}\right)^{-1} \mathbf{X}' \mathbb{M}_{\mathrm{Q}} \left(\mathbf{y} - \widehat{\rho} \mathbf{y}_{-1}\right), \tag{16}$$

with $\mathbb{M}_{\mathrm{XQ}} = \mathbb{M}_{\mathrm{x}} \mathbb{M}_{\mathrm{Q}}$, $\mathbb{M}_{\mathrm{Q}} = \mathbf{I}_N \otimes \mathbf{M}_{\mathrm{Q}}$ and $\mathbb{M}_{\mathrm{x}} = \mathbf{I}_{NT} - \mathbb{M}_{\mathrm{Q}} \mathbf{X} \left(\mathbf{X}' \mathbb{M}_{\mathrm{Q}} \mathbf{X}\right)^{-1} \mathbf{X}' \mathbb{M}_{\mathrm{Q}}$. Since the asymptotic bias of the dynamic CCEP estimator will depend on the matrix of cross-section averages and the resulting projection matrix $\mathbf{H}_{\mathrm{Q}}$, we also need to analyze its asymptotic properties. To this end, we use Assumption 5 to invert equations (8) and (9) and express the cross-section averages in $\mathbf{Q}$ as

$$\bar{y}_t = \left[\bar{\alpha}^{+} + \rho(L)^{-1} \boldsymbol{\beta}^{*'} \boldsymbol{\lambda}(L)^{-1} \bar{\mathbf{c}}_{\mathrm{z}}\right] + \rho(L)^{-1} \left[\bar{\boldsymbol{\gamma}}' + \boldsymbol{\beta}^{*'} \boldsymbol{\lambda}(L)^{-1} \bar{\boldsymbol{\Gamma}}'\right] \mathbf{f}_t + O_p\left(N^{-1/2}\right),$$
$$\bar{\mathbf{z}}_t = \boldsymbol{\lambda}(L)^{-1} \bar{\mathbf{c}}_{\mathrm{z}} + \boldsymbol{\lambda}(L)^{-1} \bar{\boldsymbol{\Gamma}}' \mathbf{f}_t + O_p\left(N^{-1/2}\right),$$

so that

$$\operatorname*{plim}_{N \to \infty} \bar{y}_t = \bar{\alpha}^* + \bar{\boldsymbol{\gamma}}^{*'} \mathbf{f}_t = \mathbf{f}_{\mathrm{y},t}^*, \tag{17}$$

$$\operatorname*{plim}_{N \to \infty} \bar{\mathbf{z}}_t = \bar{\mathbf{c}}_{\mathrm{z}}^* + \bar{\boldsymbol{\Gamma}}^{*'} \mathbf{f}_t = \mathbf{f}_{\mathrm{z},t}^*, \tag{18}$$

where $\bar{\alpha}^* = \bar{\alpha}^{+} + \rho(L)^{-1} \boldsymbol{\beta}^{*'} \bar{\mathbf{c}}_{\mathrm{z}}^*$, $\bar{\mathbf{c}}_{\mathrm{z}}^* = \boldsymbol{\lambda}(L)^{-1} \bar{\mathbf{c}}_{\mathrm{z}}$, $\bar{\boldsymbol{\gamma}}^* = \rho(L)^{-1} \left[\bar{\boldsymbol{\gamma}} + \bar{\boldsymbol{\Gamma}}^* \boldsymbol{\beta}^*\right]$ and $\bar{\boldsymbol{\Gamma}}^* = \bar{\boldsymbol{\Gamma}} \boldsymbol{\lambda}(L)'^{-1}$. Therefore, as $N \to \infty$ we obtain $\mathbf{Q} \to \mathbf{Q}_{\mathrm{f}}$ with $\mathbf{Q}_{\mathrm{f}} = (\boldsymbol{\iota}_T, \mathbf{F}_{\mathrm{y}}^*, \mathbf{F}_{\mathrm{y},-1}^*, \mathbf{F}_{\mathrm{z}}^*, \mathbf{F}_{\mathrm{z},-1}^*, \ldots, \mathbf{F}_{\mathrm{z},-p_T}^*)$ and $\mathbf{F}_{\mathrm{y}}^* = (\mathbf{f}_{\mathrm{y},1}^*, \ldots, \mathbf{f}_{\mathrm{y},T}^*)'$, $\mathbf{F}_{\mathrm{z}}^* = (\mathbf{f}_{\mathrm{z},1}^*, \ldots, \mathbf{f}_{\mathrm{z},T}^*)'$ and their lags defined similarly.

Karabiyik et al. (2016) build on the work of Andrews (1987) to show that despite the convergence of the matrix of cross-section averages $\mathbf{Q} \to \mathbf{Q}_{\mathrm{f}}$ as $N \to \infty$, it is not necessarily the case that also $\mathbf{H}_{\mathrm{Q}} \to \mathbf{H}_{\mathrm{Q}_{\mathrm{f}}}$ since the g-inverses $(\mathbf{Q}'\mathbf{Q})^{-}$ and $(\mathbf{Q}_{\mathrm{f}}'\mathbf{Q}_{\mathrm{f}})^{-}$ may not converge. The latter can occur in case $m < 1 + k$ because the matrix $\mathbf{Q}'\mathbf{Q}$ is then of greater rank than its limit $\mathbf{Q}_{\mathrm{f}}'\mathbf{Q}_{\mathrm{f}}$, which leads to violation of the requirement that their ranks converge almost surely.[4] Karabiyik et al. (2016) show specifically in the static context that $\mathbf{H}_{\mathrm{Q}} \nrightarrow \mathbf{H}_{\mathrm{Q_f}}$ unless $m = 1 + k$, which also holds in our setting. We therefore define the limit of the projection matrix $\mathbf{H}_{\mathrm{Q}}$ as

$$\mathbf{H}_{\mathrm{Q}}^{\mathrm{L}} = \operatorname*{plim}_{N\to\infty} \mathbf{Q}\left(\mathbf{Q}'\mathbf{Q}\right)^{-}\mathbf{Q}', \tag{19}$$

and note that (19) simplifies to $\mathbf{H}_{\mathrm{Q}}^{\mathrm{L}} = \mathbf{Q}_{\mathrm{f}}\left(\mathbf{Q}_{\mathrm{f}}'\mathbf{Q}_{\mathrm{f}}\right)^{-}\mathbf{Q}_{\mathrm{f}}'$ in case $m = 1 + k$.

The following theorem provides the asymptotic bias of the dynamic CCEP estimator for $N \to \infty$ and T fixed.

Theorem 1. *Consider the dynamic panel data model in equations (1)-(3). In this model the CCEP estimator is inconsistent as $N \to \infty$ and T fixed with its asymptotic bias given by*

(a) *Under Assumptions 1-5* and with $p_T \geq p$,

$$\operatorname*{plim}_{N\to\infty}(\widehat{\rho} - \rho) = -\frac{1}{T}\frac{\sigma_\varepsilon^2}{\sigma_{\breve{y}_{-1}}^2}\Upsilon(\rho, \mathbf{H}_{\mathrm{Q}}^{\mathrm{L}}), \tag{20}$$

$$\operatorname*{plim}_{N\to\infty}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = -\boldsymbol{\zeta}\operatorname*{plim}_{N\to\infty}(\widehat{\rho} - \rho), \tag{21}$$

with

- $\boldsymbol{\zeta} = \operatorname{plim}_{N\to\infty}\left(\mathbf{X}'\mathbb{M}_{\mathrm{Q}}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbb{M}_{\mathrm{Q}}\mathbf{y}_{-1}$
- $\sigma_{\breve{y}_{-1}}^2 = \operatorname{plim}_{N\to\infty}(NT)^{-1}\mathbf{y}_{-1}'\mathbb{M}_{\mathrm{XQ}}\mathbf{y}_{-1}$
- $\Upsilon(\rho, \mathbf{H}_{\mathrm{Q}}^{\mathrm{L}}) = \sum_{t=1}^{T-1}\rho^{t-1}\sum_{s=t+1}^{T}\mathbf{H}_{\mathrm{Q}}^{\mathrm{L}}(s, s-t)$

Hence, from (20) and (21) the following asymptotic expression holds for $\widehat{\boldsymbol{\delta}}$

$$\operatorname*{plim}_{N\to\infty}\widehat{\boldsymbol{\delta}} = \mathbf{m}(\boldsymbol{\delta}) = \boldsymbol{\delta} + \begin{pmatrix}-1\\ \boldsymbol{\zeta}\end{pmatrix}\frac{1}{T}\frac{\sigma_\varepsilon^2}{\sigma_{\breve{y}_{-1}}^2}\Upsilon(\rho, \mathbf{H}_{\mathrm{Q}}^{\mathrm{L}}). \tag{22}$$

(b) *Under Assumptions 1-5, with $p_T \geq p$ and making explicit use of $|\rho| < 1$, the expression for the asymptotic bias of $\widehat{\rho}$ can be particularized as*

$$\operatorname*{plim}_{N\to\infty}(\widehat{\rho} - \rho) = \frac{-\left[A(\rho, T) + D(\rho, \mathbf{H}_{\tilde{\mathrm{Q}}}^{\mathrm{L}})\right]}{\left[B(\rho, T) - E(\rho, \mathbf{H}_{\tilde{\mathrm{Q}}}^{\mathrm{L}}) + TC\right]} = \psi(\rho, T, \mathbf{H}_{\tilde{\mathrm{Q}}}^{\mathrm{L}}, C), \tag{23}$$

with

- $A(\rho, T) = \frac{1}{1-\rho}\left(1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right)$
- $B(\rho, T) = \frac{T}{1-\rho^2}\left(1 - \frac{1}{T}\frac{1+\rho}{1-\rho} - \frac{2\rho}{T^2}\frac{1-\rho^T}{(1-\rho)^2}\right)$
- $D(\rho, \mathbf{H}_{\tilde{\mathrm{Q}}}^{\mathrm{L}}) = \sum_{t=1}^{T-1}\rho^{t-1}\sum_{s=t+1}^{T}\mathbf{H}_{\tilde{\mathrm{Q}}}^{\mathrm{L}}(s, s-t)$

[4] Theorem 2 in Andrews (1987) states that if it holds for a sequence $\mathbf{A}_n$ of random matrices that $\mathbf{A}_n \xrightarrow{p} \mathbf{A}$, then $(\mathbf{A}_n)^- \xrightarrow{p} \mathbf{A}^-$ iff $\mathrm{P}\left[\mathrm{rk}(\mathbf{A}_n) = \mathrm{rk}(\mathbf{A})\right] \xrightarrow{n\to\infty} 1$. In our setting this requires that $\mathrm{P}\left[\mathrm{rk}(\mathbf{Q}'\mathbf{Q}) = \mathrm{rk}(\mathbf{Q}_{\mathrm{f}}'\mathbf{Q}_{\mathrm{f}})\right] \xrightarrow{N\to\infty} 1$, which Karabiyik et al. (2016) show is not true unless $m = 1 + k$.

- $E(\rho, \mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}}) = \frac{1}{1-\rho^2}\left[\mathrm{tr}(\mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}}) + 2\rho D(\rho, \mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}})\right]$

where $\mathbf{H}_{\widetilde{\mathrm{Q}}}$ *is set up in function of* $\widetilde{\mathbf{Q}}$ *, the matrix of cross-section averages in deviation of their column mean, and*

$$C = \operatorname*{plim}_{N\to\infty} \frac{\boldsymbol{\beta}'\boldsymbol{\Omega}_{\check{\mathrm{x}}}\boldsymbol{\beta} + \boldsymbol{\Lambda}'\boldsymbol{\Omega}_{\check{\mathrm{f}}}\boldsymbol{\Lambda} + 2\boldsymbol{\beta}'\boldsymbol{\Omega}_{\check{\mathrm{x}},\check{\mathrm{f}}}\boldsymbol{\Lambda}}{\sigma^2_{\varepsilon}}, \tag{24}$$

with $\boldsymbol{\Omega}_{\check{\mathrm{x}}} = (NT)^{-1}\mathbf{X}^{+'}_{-1}\mathbb{M}_{\mathrm{XQ}}\mathbf{X}^{+}_{-1}$, $\boldsymbol{\Omega}_{\check{\mathrm{f}}} = (NT)^{-1}\mathbb{F}^{+'}_{-1}\mathbb{M}_{\mathrm{XQ}}\mathbb{F}^{+}_{-1}$, $\boldsymbol{\Omega}_{\check{\mathrm{x}},\check{\mathrm{f}}} = (NT)^{-1}\mathbf{X}^{+'}_{-1}\mathbb{M}_{\mathrm{XQ}}\mathbb{F}^{+}_{-1}$.

Theorem 1 extends the results in Everaert and De Groote (2016), who consider a model with one common factor and no additional covariates, to a model with multiple factors and additional exogenous regressors. Equation (21) shows that the asymptotic bias of $\widehat{\boldsymbol{\beta}}$ is a fraction $-\boldsymbol{\zeta}$ of the asymptotic bias of $\widehat{\rho}$, with $\boldsymbol{\zeta}$ being the asymptotic (as $N \to \infty$) CCEP estimates in a model regressing $\mathbf{y}_{i,-1}$ on $\mathbf{X}_i$. The expression in equation (20) shows that for fixed T the asymptotic bias of $\widehat{\rho}$ depends on a term $-(\sigma^2_{\varepsilon}/T)\Upsilon(.)$, which captures the covariance between the defactored lagged dependent variable $\check{\mathbf{y}}_{-1} = \mathbb{M}_{\mathrm{XQ}}\mathbf{y}_{-1}$ and the error term $\boldsymbol{\varepsilon}$, scaled by the variance $\sigma^2_{\check{\mathrm{y}}_{-1}}$ of $\check{\mathbf{y}}_{-1}$. The asymptotic bias expression for $\widehat{\rho}$ in equation (20) will be convenient to devise a generally applicable bias-corrected CCEP estimator in Section 4.1 below, but is difficult to evaluate. Equation (23) therefore presents a more explicit bias expression that will be used in Section 4.2 to derive two restricted bias-corrected CCEP estimators which will be valid in the single factor case in a model with and without additional covariates respectively.

Equation (23) shows that the inconsistency as $N \to \infty$ of the CCEP estimator for ρ stems from three different sources. The first source is the standard FE dynamic panel data bias as calculated by Nickell (1981) and represented by the terms $A(\rho, T)$ and $B(\rho, T)$ for a dynamic model without additional covariates and common factors. This bias is induced by the within transformation (through including $\boldsymbol{\iota}_T$ in the transformation matrix $\mathbf{M}_{\mathrm{Q}}$) which implies negative correlation between the within transformed error term and the lagged dependent variable.

Second, the presence of exogenous variables and common factors reduce the magnitude of the asymptotic bias to the extent that these variables increase the variance $\sigma^2_{\check{\mathrm{y}}_{-1}}$ of the defactored $\mathbf{y}_{-1}$ relative to σ^2_{ε}. This is captured by the term $\sigma^2_{\varepsilon} \big/ \sigma^2_{\check{\mathrm{y}}_{-1}}$ in equation (20) and by the positive variance term C in the denominator of equation (23). Although the contemporaneous effect of the covariates and the common factors (when the rank condition holds) is eliminated by defactoring $\mathbf{y}$, they still imply a higher signal in the lagged dependent variable as this is affected by lagged values of $\mathbf{X}$ and $\mathbf{F}$. Note that the C-term in equation (23) can be worked out further if we are willing to make additional assumptions about the data generating processes of $\mathbf{X}$ and $\mathbf{F}$. However, the resulting bias expression will depend on unknown parameters in these data generating processes that are difficult to estimate in practice and hence will be of little use when constructing a bias-corrected estimator below. We can nevertheless observe that the first and the last term in the numerator of C drop out in a model with no exogenous variables ($\boldsymbol{\beta} = \mathbf{0}$) and likewise the second and last term in case of a single factor ($\boldsymbol{\Omega}_{\check{\mathrm{f}}} = \boldsymbol{\Omega}_{\check{\mathrm{x}},\check{\mathrm{f}}} = \mathbf{0}$ because $\mathbb{M}_{\mathrm{Q}}\mathbb{F}^{+}_{-1} = \mathbf{0}$

in this case). When $m > 1$, $\mathbb{M}_{\mathrm{Q}}\mathbb{F}^{+}_{-1}$ is nonzero regardless of the rank condition. This follows from the fact that the orthogonalization matrix $\mathbf{M}_{\mathrm{Q}}$ is able to wipe out contemporaneous factor representations ($\mathbf{F}$) but not their accumulated manifestations through time in $\mathbf{y}_{-1}$ ($\mathbf{F}^{+}_{-1}$), unless in the special case where $m = 1$.[5] As such, an increase in the importance of the factors will only reduce the asymptotic bias of the CCEP estimator when more than one factor is present.

The third source is the orthogonalization on the cross-sectional averages. This results in additional terms in both the numerator and the denominator of equation (23). The term $D(\rho, \widetilde{\mathbf{Q}}_{\mathrm{f}})$ in the numerator[6] is stochastic as for fixed T the matrix $\mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}}$ depends, through the cross-sectional averages, on the particular realization of the factors $\mathbf{F}$. This randomness will be adequately dealt with in our bias correction procedure (as discussed below) but complicates a general evaluation of the bias as its magnitude and even its sign may vary depending on ρ and the specific data. In general, we expect $D(.)$ to be negative[7] but smaller in magnitude[8] than $A(\rho, T)$, which is a positive deterministic quantity. Hence, the additional orthogonalization on the cross-sectional averages tends to reduce the absolute value of the numerator in equation (23). Note that this part of the bias is not driven by the strength of the cross-sectional dependence (as the contemporaneous impact of the common factors is effectively eliminated from the model by the CCE approach) but stems from the fact that we include the cross-sectional averages as additional variables in the transformation matrix $\mathbf{M}_{\mathrm{Q}}$. The mitigating effect in the numerator of the bias expression is likely to be counteracted by the extra terms in the denominator of equation (23). This is proportional to the variance (conditional on $\mathbf{X}$) that remains in the lagged dependent variable $\mathbf{y}_{-1}$ after the within transformation and the orthogonalization on the cross-sectional averages. The latter implies that more of the variation in $\mathbf{y}_{-1}$ will be wiped out compared to the within transformation, as shown by the $-E\,(.)$ term which is always negative. Moreover the C-term represents what remains of the exogenous regressor and common factor variation in the defactored $\check{\mathbf{y}}_{-1}$ (expressed relative to σ^2_ε) and it too will be smaller as a result of the orthogonalization on the cross-section averages. Hence, when the cross-section averages cut out a relatively large amount of variation the denominator of equation (23) may decrease faster than the reduction in the numerator and result in a larger bias. For a given number of common factors, increasing the number of cross-section

[5]The reasoning behind this result is that in the single factor case $\mathbf{F}^{+}_{-1}$ can be written in function of a constant and $\bar{\mathbf{y}}_{-1}$ so that the inclusion of this cross-section average in $\mathbf{M}_{\mathrm{Q}}$ suffices to cut out $\mathbf{F}^{+}_{-1}$ (see Everaert and De Groote, 2016). However, in multifactor settings multiple variables are required to cut $\mathbf{F}^{+}_{-1}$ and it can be shown that next to $\bar{\mathbf{y}}_{-1}$ also $\bar{\mathbf{z}}_{t-1}$ and an infinite number of its lags are needed. Therefore, in finite T scenarios the lag length of $\bar{\mathbf{z}}_{t-1}$ may not be sufficient to remove $\mathbf{F}^{+}_{-1}$.

[6]Note that this term was overlooked by Phillips and Sul (2007) in their footnote 4 where the authors suggest that the numerator converges to $-\sigma^2_\varepsilon A(\rho, T)$.

[7]This reasoning follows from the fact that $\mathrm{tr}(\mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}})/2 = (2 + k(1 + p_T))/2 = -\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}}(s, s-t)$. Since $\mathrm{tr}(\mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}})/2 > 0$ and $D(.) = \sum_{t=1}^{T-1}\rho^{t-1}\sum_{s=t+1}^{T}\mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}}(s, s-t)$ is the same sum (without the minus sign) reweighted in function of ρ, uniformly positive weights ($\rho > 0$) are likely to result in a negative $D(.)$. However, one can not uniformly claim $D(.) < 0$ since individual elements in $\mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}}$ can be positive, and in case ρ is small these may be the leading elements in the sum. Moreover, with $\rho < 0$ the sign of the respective elements change.

[8]We have the numerical property that $\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\mathbf{H}^{\mathrm{L}}_{\mathrm{Q}}(s, s-t) = \left(T - \mathrm{tr}(\mathbf{H}^{\mathrm{L}}_{\mathrm{Q}})\right)/2 = (T - 3 - k(1 + p_T))/2 > 0$ and since $A(.) + D(.) = \sum_{t=1}^{T-1}\rho^{t-1}\sum_{s=t+1}^{T}\mathbf{H}^{\mathrm{L}}_{\mathrm{Q}}(s, s-t)$ it is to be expected that for $\rho > 0$ the numerator is negative.

averages used by the CCEP estimator is likely to increase its asymptotic bias. This is confirmed by the Monte Carlo simulations of section 5.

4 Bias-corrected dynamic CCEP estimator

In this section we develop a bias-corrected CCEP estimator based on the large N asymptotic bias expressions derived in Section 3. We refrain from deriving and using large T expansions of these bias expressions for two reasons. First, these expansions may imply a considerable loss in accuracy when T is (very) small. Second, they are a function of unknown parameters related to the unobserved factors (see Everaert and De Groote, 2016, equation (30), Theorem 1.a) which are not only difficult to obtain in practice but would also introduce additional uncertainty into the procedure. Using the large N bias expressions does not require these parameters and holds for any T.

Equation (23) in Theorem 1 can in principle be used for constructing a bias-corrected CCEP estimator but the presence of the C-term in the denominator is troublesome since it contains multiple unknown quantities. More specifically, C depends on the unobserved loadings and variance-covariance matrices of the infinite sums $\mathbf{X}^{+}_{-1}$ and $\mathbf{F}^{+}_{-1}$ that are difficult to estimate unless in some specific settings (see Section 4.2). Equation (20) in Theorem 1 is more convenient since the variance $\sigma^2_{\check{y}_{-1}}$ of the defactored lagged dependent variable can be replaced by its sample analog $\widehat{\sigma}^2_{\check{y}_{-1}} = (NT)^{-1}\mathbf{y}'_{-1}\mathbb{M}_{XQ}\mathbf{y}_{-1}$. Similarly, $\mathbf{H}^{\mathrm{L}}_{\mathrm{Q}}$ is replaced by the known projection matrix of cross-section averages $\mathbf{H}_{\mathrm{Q}}$ and the asymptotic coefficient vector $\boldsymbol{\zeta}$ by its sample analog $\widehat{\boldsymbol{\zeta}}$ obtained by regressing $\mathbf{y}_{i,-1}$ on $\mathbf{X}_i$ using the CCEP estimator. In Section 4.1 we therefore use equation (20) to devise a generally applicable bias-corrected CCEP estimator. In Section 4.3 we show that the bias-corrected CCEP estimators are consistent for $N \to \infty$ and T fixed.

4.1 A generally applicable correction: CCEPbc

Using the asymptotic bias equations in (20) and (21), a bias-corrected CCEP estimator $\widehat{\boldsymbol{\delta}}_{bc} = (\widehat{\rho}_{bc}, \widehat{\boldsymbol{\beta}}'_{bc})'$ for $\boldsymbol{\delta}$ can be obtained by solving the following system of equations for $\boldsymbol{\delta}$

$$\widehat{\boldsymbol{\delta}} = \mathbf{m}^*(\boldsymbol{\delta}) = \boldsymbol{\delta} + \begin{pmatrix} -1 \\ \widehat{\boldsymbol{\zeta}} \end{pmatrix} \frac{1}{T} \frac{\widehat{\sigma}^2_{\varepsilon}(\boldsymbol{\delta})}{\widehat{\sigma}^2_{\check{y}_{-1}}} \Upsilon(\rho, \mathbf{H}_{\mathrm{Q}}), \tag{25}$$

where the unknown parameters $\mathbf{H}^{\mathrm{L}}_{\mathrm{Q}}$, $\sigma^2_{\check{y}_{-1}}$ and $\boldsymbol{\zeta}$ have been replaced by their sample analogs $\mathbf{H}_{\mathrm{Q}}$, $\widehat{\sigma}^2_{\check{y}_{-1}}$ and $\widehat{\boldsymbol{\zeta}}$ and the unknown variance σ^2_{ε} by its infeasible estimator

$$\widehat{\sigma}^2_{\varepsilon}(\boldsymbol{\delta}) = \frac{\sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{w}_i\boldsymbol{\delta})' \mathbf{M}_{\mathrm{Q}} (\mathbf{y}_i - \mathbf{w}_i\boldsymbol{\delta})}{N(T - 3 - k(1 + p_T)) - (1 + k_x)}. \tag{26}$$

The error variance estimator based on the uncorrected CCEP's error terms is inconsistent for finite T, but by defining $\widehat{\sigma}^2_{\varepsilon}(\boldsymbol{\delta})$ as a function of the parameters of interest, we construct and use a bias-corrected estimator for σ_ε^2 as well. This estimation problem is non-linear in $\boldsymbol{\delta}$ but it is easily managed by standard numerical optimizers. The solution $\widehat{\boldsymbol{\delta}}_{bc}$ is in effect the vector of parameters that follows from inverting $\mathbf{m}^*(\boldsymbol{\delta})$, the feasible alternative to the asymptotic bias expression in equation (22). The CCEPbc estimator can thus be expressed as $\widehat{\boldsymbol{\delta}}_{bc} = \mathbf{m}^{*-1}(\widehat{\boldsymbol{\delta}})$. Note that the bias correction approach presented here resembles the one proposed by Bun and Carree (2005) to obtain a bias-corrected FE estimator in dynamic panel models without cross-sectional dependence.

4.2 Restricted corrections for models with a single factor

The procedure outlined above is a generally applicable method in the sense that it does not require the number of factors to be known. However, in some specific cases restricted forms of the bias-corrected CCEP estimator are possible. These may be more efficient as a result of the imposed restrictions.

First, in a model without covariates ($\boldsymbol{\beta} = \mathbf{0}$) and a single common factor ($m = 1$) the problem simplifies considerably since the C-term, which makes bias correction from equation (23) inconvenient, is zero for $N \to \infty$. This further implies that the bias expression for ρ no longer depends on σ_ε^2 such that ρ is the only unknown parameter in equation (23). A first restricted bias-corrected CCEP estimator $\widehat{\boldsymbol{\delta}}_{bcr1}$ can thus be obtained by solving the following equation for ρ, where $\mathbf{H}_{\widetilde{\mathrm{Q}}}$ is again used as the sample analog of $\mathbf{H}_{\widetilde{\mathrm{Q}}}^{\mathrm{L}}$

$$\widehat{\rho} = \rho + \psi\left(\rho, T, \mathbf{H}_{\widetilde{\mathrm{Q}}}, 0\right). \tag{27}$$

Second, adding exogenous regressors implies that $C \neq 0$ but if the assumption of a single factor is maintained we obtain the relatively simple form

$$C = \operatorname*{plim}_{N\to\infty} \frac{\boldsymbol{\beta}'\boldsymbol{\Omega}_{\check{\mathrm{x}}}\boldsymbol{\beta}}{\sigma_\varepsilon^2}, \tag{28}$$

which through $\boldsymbol{\Omega}_{\check{\mathrm{x}}}$ also depends on the unknown parameter ρ and on the infinite sum of explanatory variables $\mathbf{X}_{-1}^{+} = \sum_{l=0}^{\infty} \rho^l \mathbf{X}_{-1-l}$. In a finite sample, the latter can be approximated by the truncated sum $\widehat{\mathbf{X}_{-1}^{+}} = \left(\widehat{\mathbf{X}^{+}}_{1,-1}', ..., \widehat{\mathbf{X}^{+}}_{N,-1}'\right)'$ where $\widehat{\mathbf{X}^{+}}_{i,-1} = \mathbf{J}^{-1}\mathbf{X}_{i,-1}$, and $\mathbf{J}$ is a $T \times T$ matrix with ones on the main diagonal and $-\rho$ on the first subdiagonal. The variance-covariance matrix is then estimated as $\widehat{\boldsymbol{\Omega}}_{\check{\mathrm{x}}}(\rho) = (NT)^{-1}\widehat{\mathbf{X}_{-1}^{+}}' \mathbb{M}_{\mathrm{XQ}} \widehat{\mathbf{X}_{-1}^{+}}$. Further substituting $\widehat{\sigma}_\varepsilon^2(\boldsymbol{\delta})$ as defined in (26) for σ_ε^2, the infeasible estimator for C is given by

$$\widehat{C}\left(\boldsymbol{\delta}\right) = \frac{\boldsymbol{\beta}'\widehat{\boldsymbol{\Omega}}_{\check{\mathrm{x}}}(\rho)\boldsymbol{\beta}}{\widehat{\sigma}_\varepsilon^2(\boldsymbol{\delta})}, \tag{29}$$

which is, conditionally on the unknown parameters ρ and $\boldsymbol{\beta}$, a function of the observed data only. Hence, a second bias-corrected CCEP estimator $\widehat{\boldsymbol{\delta}}_{bcr2}$ can be obtained by solving the following system of equations for $\boldsymbol{\delta}$

$$\widehat{\boldsymbol{\delta}} = \boldsymbol{\delta} + \begin{pmatrix} 1 \\ -\widehat{\boldsymbol{\zeta}} \end{pmatrix} \psi(\rho, T, \mathbf{H}_{\widetilde{\mathrm{Q}}}, \widehat{C}\left(\boldsymbol{\delta}\right)). \tag{30}$$

The resulting correction should perform well when the single factor assumption is true and the inaccuracies by the lag truncation in the approximation of $\mathbf{X}_{-1}^{+}$ are not too severe. Note that the truncation implies that the estimator is inconsistent for finite T, but in practice the bias may be negligible (depending on the size of ρ). In case more than one factor is present, equation (29) may also be a poor approximation of C and lead to additional bias, especially when the factors have a large overall influence on the model (relative to σ_{ε}^{2}).

4.3 Asymptotic distribution and consistency

The CCEPbc presented in Section 4.1 is a function of the original CCEP estimator $\widehat{\boldsymbol{\delta}}$ and follows from inverting its asymptotic bias expression in equation (22) of Theorem 1. In case σ_{ε}^{2}, $\sigma_{\check{\mathbf{y}}_{-1}}^{2}$, $\mathbf{H}_{\mathrm{Q}}^{\mathrm{L}}$ and $\boldsymbol{\zeta}$ are known quantities, equation (22) can be used directly and the CCEPbc estimator is given by $\widehat{\boldsymbol{\delta}}_{bc} = \mathbf{m}^{-1}(\widehat{\boldsymbol{\delta}})$. We then have

$$\operatorname*{plim}_{N\to\infty} \widehat{\boldsymbol{\delta}}_{bc} = \operatorname*{plim}_{N\to\infty} \mathbf{m}^{-1}(\widehat{\boldsymbol{\delta}}) = \mathbf{m}^{-1}\left(\operatorname*{plim}_{N\to\infty} \widehat{\boldsymbol{\delta}}\right) = \mathbf{m}^{-1}\left(\mathbf{m}\left(\boldsymbol{\delta}\right)\right) = \boldsymbol{\delta}, \tag{31}$$

and $\widehat{\boldsymbol{\delta}}_{bc}$ is a consistent estimator for $\boldsymbol{\delta}$ as $N \to \infty$ and T fixed. In practice $\mathbf{m}(.)$ is unknown and replaced by its feasible version $\mathbf{m}^{*}(.)$ introduced in Section 4.1, with the unknown quantities replaced by their respective estimators. In particular, $\sigma_{\check{\mathbf{y}}_{-1}}^{2}$, $\mathbf{H}_{\mathrm{Q}}^{\mathrm{L}}$ and $\boldsymbol{\zeta}$ are represented by their sample analogs and σ_{ε}^{2} is replaced by the infeasible estimator in (26) for which we find, when evaluated at the population value $\boldsymbol{\delta}$,

$$\begin{aligned}
\operatorname*{plim}_{N\to\infty} \widehat{\sigma}_{\varepsilon}^{2}\left(\boldsymbol{\delta}\right) &= \operatorname*{plim}_{N\to\infty} \frac{\sum_{i=1}^{N}\left(\mathbf{y}_i - \mathbf{w}_i\boldsymbol{\delta}\right)'\mathbf{M}_{\mathrm{Q}}\left(\mathbf{y}_i - \mathbf{w}_i\boldsymbol{\delta}\right)}{N(T-3-k(1+p_T)) - (1+k_x)} = \operatorname*{plim}_{N\to\infty} \frac{\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i'\mathbf{M}_{\mathrm{Q}}\boldsymbol{\varepsilon}_i}{T-3-k(1+p_T) - \frac{1}{N}(1+k_x)}, \\
&= \sigma_{\varepsilon}^{2}\frac{T - \operatorname{tr}\left(\mathbf{H}_{\mathrm{Q}}^{\mathrm{L}}\right)}{T-3-k(1+p_T)} = \sigma_{\varepsilon}^{2}\frac{T-3-k(1+p_T)}{T-3-k(1+p_T)} = \sigma_{\varepsilon}^{2},
\end{aligned} \tag{32}$$

whereas for $\boldsymbol{\delta}^{0} \neq \boldsymbol{\delta}$,

$$\operatorname*{plim}_{N\to\infty} \widehat{\sigma}_{\varepsilon}^{2}\left(\boldsymbol{\delta}^{0}\right) \neq \sigma_{\varepsilon}^{2}. \tag{33}$$

Hence, noting that evaluating $\widehat{\boldsymbol{\delta}}_{bc} = \mathbf{m}^{*-1}(\widehat{\boldsymbol{\delta}})$ is equivalent to finding the vector $\widehat{\boldsymbol{\delta}}_{bc}$ that satisfies

$$\left(\widehat{\boldsymbol{\delta}} - \widehat{\boldsymbol{\delta}}_{bc}\right) - \begin{pmatrix} -1 \\ \widehat{\boldsymbol{\zeta}} \end{pmatrix} \frac{1}{T}\frac{\widehat{\sigma}_{\varepsilon}^{2}(\widehat{\boldsymbol{\delta}}_{bc})}{\widehat{\sigma}_{\check{\mathbf{y}}_{-1}}^{2}} \Upsilon(\widehat{\rho}_{bc}, \mathbf{H}_{\mathrm{Q}}) = \mathbf{0},$$

we have, taking probability limits as $N \to \infty$,

$$\left(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}}_{bc}\right) + \frac{1}{T}\frac{1}{\sigma_{\check{\mathbf{y}}_{-1}}^{2}} \begin{pmatrix} -1 \\ \boldsymbol{\zeta} \end{pmatrix} \left[\sigma_{\varepsilon}^{2}\Upsilon(\rho, \mathbf{H}_{\mathrm{Q}}^{\mathrm{L}}) - \operatorname*{plim}_{N\to\infty} \widehat{\sigma}_{\varepsilon}^{2}(\widehat{\boldsymbol{\delta}}_{bc})\Upsilon(\widehat{\rho}_{bc}, \mathbf{H}_{\mathrm{Q}})\right] = \mathbf{0}, \tag{34}$$

so that by (32) and (33) the vector $\widehat{\boldsymbol{\delta}}_{bc} = \boldsymbol{\delta}$ is the only solution as $N \to \infty$. Consequently, replacing the unknown (σ_{ε}^{2}, $\sigma_{\check{\mathbf{y}}_{-1}}^{2}$, $\mathbf{H}_{\mathrm{Q}}^{\mathrm{L}}$, $\boldsymbol{\zeta}$) by their respective estimators maintains the consistency of $\widehat{\boldsymbol{\delta}}_{bc}$ for fixed T and $N \to \infty$. A similar argument can be made for the restricted correction of section 4.2

in an AR(1) model without exogenous predictors and a single factor (CCEPbcr1), but not in the presence of exogenous regressors (CCEPbcr2) or multiple factors.

In order to obtain the asymptotic distribution of CCEPbc we will assume that the original CCEP estimator is inconsistent but, conditional on $\mathbf{F}$, normally distributed as $N \to \infty$ and T fixed with variance-covariance matrix $\mathbf{V}$. Using the functional relationship between CCEP and the corrected estimator we then have, by the delta method

$$\sqrt{N}\left(\widehat{\boldsymbol{\delta}}_{bc} - \boldsymbol{\delta}\right) \xrightarrow[N\to\infty]{d} N\left(\mathbf{0}, \mathbf{LVL}'\right), \tag{35}$$

where $\mathbf{L}$ is the $(1+k_x) \times (1+k_x)$ matrix of first order partial derivatives of $\mathbf{m}^{*-1}(.)$. The finite T variance $\mathbf{V}$ of the CCEP estimator has a complicated form and is difficult to evaluate so that the corresponding variance-covariance matrix $\mathbf{LVL}'$ of the CCEPbc estimator is hard to calculate as well. The bootstrap provides a straightforward alternative that allows us to avoid this complexity. To this end, we follow Kapetanios (2008) and obtain bootstrap samples by resampling whole cross-sections with replacement from the original dataset. This resampling scheme is valid for $N \to \infty$ and preserves both the dynamics and the assumed factor structure in the data. The distribution of the corrected estimator is then simulated by applying CCEPbc to each of the J bootstrap datasets to obtain the coefficient vectors $\widehat{\boldsymbol{\delta}}^{b}_{bc,j}$ that correspond to samples $j = 1, \ldots, J$. The bootstrapped variance-covariance matrix of the CCEPbc estimator can then be calculated as

$$\widehat{\mathbf{V}}_{bc} = \lim_{J\to\infty} \frac{1}{J-1} \sum_{j=1}^{J} \left(\widehat{\boldsymbol{\delta}}^{b}_{bc,j} - \bar{\widehat{\boldsymbol{\delta}}}^{b}_{bc,.}\right)\left(\widehat{\boldsymbol{\delta}}^{b}_{bc,j} - \bar{\widehat{\boldsymbol{\delta}}}^{b}_{bc,.}\right)', \tag{36}$$

where $\bar{\widehat{\boldsymbol{\delta}}}^{b}_{bc,.}$ represents the average of CCEPbc estimates over the J samples. The resulting standard errors can be used to perform inference for the CCEPbc estimator based on the normal distribution. Alternatively, the obtained bootstrap distribution can also be used to construct bootstrap confidence intervals and avoid the finite T normality assumption of the original CCEP estimator.

We now consider samples where both N and T may be large by letting $(N, T)_{\text{seq.}} \to \infty$ and find from (22) and (34) that both $\widehat{\boldsymbol{\delta}} = \mathbf{m}(\boldsymbol{\delta}) \to \boldsymbol{\delta}$ and $\widehat{\boldsymbol{\delta}}_{bc} \to \boldsymbol{\delta}$. Accordingly, CCEPbc is also consistent for both N and T large and is asymptotically equivalent to CCEP,

$$\widehat{\boldsymbol{\delta}} - \widehat{\boldsymbol{\delta}}_{bc} \xrightarrow{p} \mathbf{0} \qquad \text{for} \qquad (N, T)_{\text{seq.}} \to \infty. \tag{37}$$

Given this equivalence, the normality result in Pesaran (2006) for CCEP implies that the corrected CCEPbc is also normally distributed in large N and T samples with asymptotic variance equal to that of the original CCEP estimator. For sufficiently large T the variance estimator in Pesaran (2006) for CCEP therefore also applies to $\widehat{\boldsymbol{\delta}}_{bc}$ in our setting.

In the discussion so far we have assumed that Assumption 4, the rank condition, was satisfied. This is a crucial requirement in fixed T settings since the consistency of the CCEPbc estimator for $N \to \infty$ and T fixed relies on the ability of the cross-sectional averages of the observables to effectively approximate the unobserved factors as $N \to \infty$. When the rank condition is not satisfied or an insufficient number of lagged cross-sectional averages is included, the common factors will not be eliminated from the error terms. The part of $\boldsymbol{\gamma}_i'\mathbf{f}_t$ that remains in the errors will then be correlated with the lagged dependent variable since the latter is also a function of $\boldsymbol{\gamma}_i$. This results in an additional bias term that is not accounted for by the bias-correction and is $O_p\left(T^{-1/2}\right)$ at best (for serially uncorrelated factors) but $O_p(1)$ in general. Hence, the rank condition cannot be relaxed for fixed T. When both $N, T \to \infty$ the rank condition can be relaxed for both CCEP and CCEPbc when the common factors are serially uncorrelated and the loadings $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Gamma}_i$ follow a random coefficient assumption and are independent of eachother and over individuals.[9]

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 against 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 the explanatory variable to one. We further 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 cross-sectional average $\bar{\mathrm{x}}_{t-1}$ (and preferably also $\bar{\mathrm{g}}_{t-1}$ when g_{it} is used as an additional variable) should be added to the CCE orthogonalization matrix in settings where $\lambda \neq 0$.

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 $j = 1, ..., m$ where we divide the variance by m in order 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 $\alpha_i \sim N\left(0, \sigma_\alpha^2\right)$ and $\mathbf{c}_{z,i} \sim N\left(\mathbf{0}, \sigma_c^2 \mathbf{I}_2\right)$ and the idiosyncratic errors as $\varepsilon_{it} \sim N\left(0, 1-\rho^2\right)$ and $\mathbf{v}_{it} \sim N\left(0, (1-\lambda^2)\mathbf{I}_2\right)$. The variance parameters σ_α^2 and σ_c^2 are set such that the contributions of the fixed effects to the variance of y_{it} and $\mathbf{z}_{it}$ equal that of their

[9]See Chudik and Pesaran (2015) for a similar result for the CCEMG estimator in dynamic models and Westerlund and Urbain (2013) for a discussion on CCEP estimation in static models when the rank condition does not hold and loadings are correlated.

respective idiosyncratic innovations (ε_{it} and $\mathbf{v}_{it}$). The factor loadings in the DGPs of y_{it}, x_{it} and g_{it} are generated as

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

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 cross-section averages 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 $RI = 1$ to be 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 $N = (25, 100, 500)$ and $T = (10, 20, 30, 50, 100)$.[10] 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).

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 mean bias, standard error (SE) and its mean estimate ($\widehat{\text{SE}}$), root mean squared error (rmse) and actual size. The SE of the coefficient estimates is computed as their standard deviation over the Monte Carlo iterations. Although analytical variance expressions are available for some estimators, to make fair comparisons possible estimated standard errors ($\widehat{\text{SE}}$) are obtained using a bootstrap approach for each of the considered estimators. Following Kapetanios (2008) we resample cross-sectional units as a whole. The advantage of this scheme is that it preserves both the persistence and the cross-sectional dependence in the data.[11] The reported estimated standard errors are based on 150 bootstrap samples. The reported actual size is the false rejection probability of a t-test at the 5% nominal significance level.

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

[10] We take $T = 10$ to be our smallest time series dimension to have sufficient degrees of freedom to calculate CCEP estimates.

[11] Note that this resampling scheme is not valid in the case of local cross-sectional dependence but is appropriate in the presence of the assumed factor structure which induces global cross-sectional dependence which is symmetric across all panel units.

and alternative setups for the bias corrections. Since differences between estimators are more outspoken for large N we mostly report tables for $N = 500$ in the main text. Small N versions ($N = 25$) for most tables are provided in Appendix B.

5.2 Baseline results

We start our discussion with a comparison of the performance of our generally applicable CCEPbc estimator (introduced in section 4.1) 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 include two alternative bias-corrected CCEP estimators as direct comparisons to our approach, i.e. the recursive mean adjustment (denoted CCEPrm) based on So and Shin (1999) and the half-panel jackknife correction (denoted CCEPjk) based on Dhaene and Jochmans (2014). 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 (2014) 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 should be the optimal choice based on the simulation results of Moon and Weidner for high persistence settings).[12]

The results in Table 1 show that the original CCEP estimator has a severe negative small T bias for ρ but is more or less unbiased for β. When $T = 10$, the bias for ρ 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 over alternative values of ρ (see Table B-1 in Appendix B) 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 randomness of the bias of the CCEP estimator contributes to its variance. The combination of bias removal and variance reduction implies that the rmse of the CCEPbc estimator is always much lower compared to 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 bias-corrected version is correctly

[12] We use the LS_factor.m routine by M. Weidner obtained from his website http://www.ucl.ac.uk/~uctpmw0/.

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

			ρ					β				
N	T		bias	SE	$\widehat{\text{SE}}$	rmse	size	bias	SE	$\widehat{\text{SE}}$	rmse	size
25	10	CCEP	-0.392	0.143	0.107	0.418	0.89	-0.033	0.048	0.051	0.058	0.09
		CCEPbc	0.007	0.150	0.145	0.150	0.06	0.000	0.052	0.056	0.052	0.04
		CCEPjk	0.040	0.356	0.312	0.359	0.08	0.017	0.094	0.113	0.096	0.02
		CCEPrm	-0.294	0.170	0.122	0.339	0.60	0.007	0.052	0.054	0.053	0.04
		FLSbc	-0.246	0.125	0.141	0.276	0.37	-0.015	0.049	0.054	0.051	0.04
25	20	CCEP	-0.178	0.060	0.048	0.188	0.93	-0.011	0.031	0.031	0.033	0.07
		CCEPbc	0.006	0.063	0.061	0.063	0.08	0.000	0.031	0.032	0.031	0.05
		CCEPjk	0.037	0.118	0.088	0.124	0.16	0.011	0.040	0.041	0.041	0.05
		CCEPrm	-0.147	0.073	0.060	0.165	0.65	0.014	0.034	0.034	0.037	0.07
		FLSbc	-0.062	0.064	0.073	0.089	0.04	0.002	0.035	0.039	0.035	0.03
25	30	CCEP	-0.111	0.038	0.033	0.118	0.89	-0.006	0.023	0.024	0.024	0.06
		CCEPbc	0.001	0.037	0.038	0.037	0.05	-0.000	0.024	0.025	0.024	0.05
		CCEPjk	0.029	0.069	0.053	0.075	0.16	0.007	0.028	0.029	0.029	0.06
		CCEPrm	-0.097	0.050	0.040	0.109	0.63	0.012	0.026	0.026	0.029	0.07
		FLSbc	-0.025	0.048	0.052	0.054	0.04	0.001	0.028	0.034	0.028	0.03
25	50	CCEP	-0.062	0.023	0.021	0.067	0.81	-0.002	0.018	0.018	0.018	0.06
		CCEPbc	-0.001	0.023	0.022	0.023	0.06	-0.000	0.018	0.018	0.018	0.05
		CCEPjk	0.014	0.035	0.029	0.037	0.11	0.003	0.020	0.020	0.020	0.06
		CCEPrm	-0.053	0.031	0.026	0.062	0.51	0.008	0.021	0.020	0.022	0.09
		FLSbc	-0.010	0.031	0.038	0.033	0.04	0.001	0.022	0.029	0.022	0.02
100	10	CCEP	-0.398	0.124	0.066	0.417	1.00	-0.033	0.026	0.025	0.042	0.29
		CCEPbc	0.014	0.097	0.093	0.098	0.08	0.001	0.026	0.028	0.026	0.04
		CCEPjk	0.057	0.319	0.223	0.324	0.19	0.019	0.053	0.055	0.057	0.05
		CCEPrm	-0.249	0.152	0.080	0.292	0.71	0.006	0.031	0.027	0.031	0.09
		FLSbc	-0.262	0.070	0.051	0.271	0.97	-0.020	0.026	0.023	0.033	0.19
100	20	CCEP	-0.179	0.048	0.026	0.185	1.00	-0.010	0.015	0.015	0.018	0.11
		CCEPbc	0.002	0.032	0.032	0.032	0.04	0.001	0.015	0.016	0.015	0.05
		CCEPjk	0.042	0.102	0.053	0.110	0.38	0.013	0.021	0.020	0.025	0.10
		CCEPrm	-0.123	0.066	0.037	0.140	0.79	0.013	0.018	0.017	0.022	0.12
		FLSbc	-0.078	0.032	0.029	0.084	0.73	-0.002	0.016	0.017	0.016	0.05
100	30	CCEP	-0.112	0.026	0.017	0.115	1.00	-0.005	0.012	0.012	0.013	0.07
		CCEPbc	-0.000	0.018	0.018	0.018	0.05	-0.000	0.012	0.012	0.012	0.04
		CCEPjk	0.031	0.055	0.029	0.063	0.38	0.008	0.014	0.014	0.016	0.08
		CCEPrm	-0.077	0.041	0.024	0.087	0.76	0.011	0.014	0.013	0.018	0.13
		FLSbc	-0.040	0.018	0.020	0.043	0.50	-0.001	0.012	0.013	0.012	0.03
500	10	CCEP	-0.403	0.120	0.047	0.420	1.00	-0.034	0.016	0.012	0.037	0.72
		CCEPbc	0.008	0.056	0.056	0.056	0.06	0.001	0.012	0.013	0.012	0.04
		CCEPjk	0.051	0.312	0.190	0.316	0.27	0.019	0.040	0.030	0.044	0.18
		CCEPrm	-0.214	0.139	0.045	0.255	0.87	0.004	0.020	0.012	0.021	0.24
		FLSbc	-0.264	0.058	0.021	0.270	0.99	-0.022	0.016	0.010	0.027	0.62
500	20	CCEP	-0.184	0.044	0.015	0.189	1.00	-0.011	0.007	0.007	0.013	0.40
		CCEPbc	0.001	0.014	0.014	0.014	0.04	-0.000	0.007	0.007	0.007	0.04
		CCEPjk	0.036	0.102	0.035	0.108	0.56	0.012	0.013	0.009	0.017	0.28
		CCEPrm	-0.097	0.059	0.021	0.113	0.93	0.010	0.010	0.008	0.014	0.27
		FLSbc	-0.079	0.025	0.011	0.083	0.99	-0.004	0.008	0.007	0.009	0.15

Note:

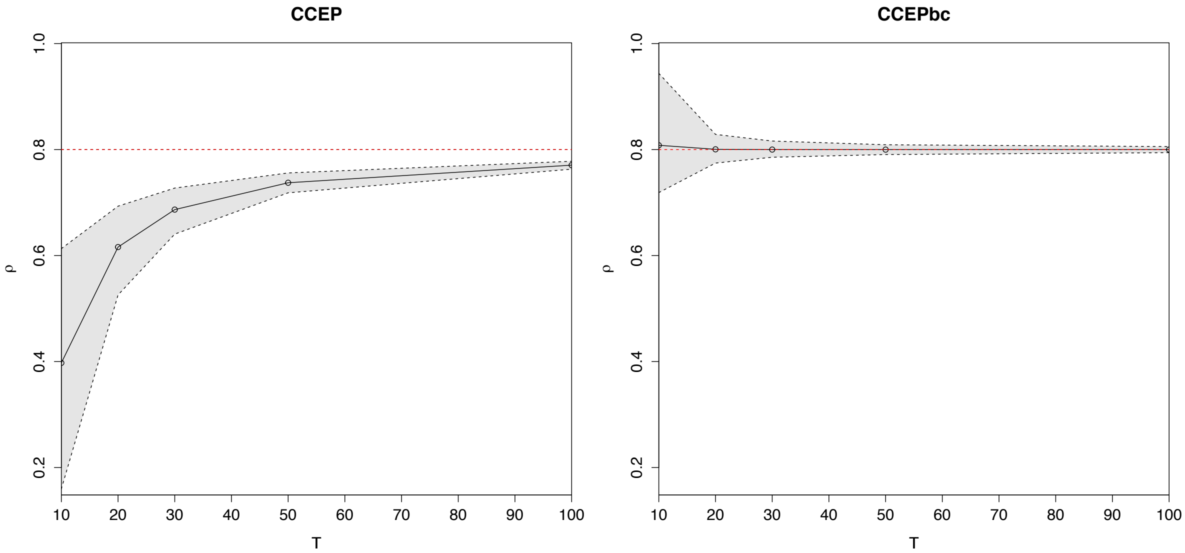
(i) Reported are simulation results for estimating ρ and β in the baseline case ($\rho = 0.8$, $\beta = 0.2$, $\lambda = 0$, $m = 1$). The factor has a contribution to the total variance of the dependent variable that is equal to that of the idiosyncratic errors ($RI = 1$).

(ii) CCEP is the Pooled CCE estimator and CCEPbc its bias-corrected version. CCEPjk and CCEPrm are the (correctly specified) jacknife and recursive mean CCEP corrections respectively. FLSbc is the bias-corrected least squares with interactive effects estimator supplied with the correct number of factors $(m+1)$. All CCEP-type estimators make no use of the cross-section average of g_{it} and include no lags of $\bar{\mathbf{x}}_t$.

(iii) All test sizes use estimated standard errors based on 150 bootstrap samples. $\widehat{\text{SE}}$ reports the mean of the bootstrapped standard errors.

centered apart from a small bias when $T = 10$. CCEPbc also offers substantial improvements regarding inference. In contrast to the CCEP estimator, 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 sensitive to the size of N. As such, it is even applicable in a sample as small as $N = 25$ and $T = 10$.

Figure 1: Monte Carlo results for ρ : comparison of CCEP and CCEPbc estimators over T



Notes:

(i) Reported are simulation results for estimating ρ in the baseline case when $N = 500$ (see notes Table 1).

(ii) Solid black lines indicate the mean estimate and the shaded area is the middle 95% interval of the simulated distribution. Dotted red lines indicate the population parameter value ($\rho = 0.8$).

The alternative bias-corrected CCEP estimators offer some alleviation of the bias but are much less effective compared to CCEPbc. Both the CCEPrm and FLSbc estimator still have a considerable negative small T bias for ρ, while the CCEPjk estimator is able to remove most of the bias but only at the cost of a much larger variance. Accordingly, these alternatives have a much larger rmse compared to CCEPbc and even for larger T the latter remains superior due the more effective bias removal. Similar results are obtained in the low persistence scenario (see Table B-1), but differences between estimators are smaller since there is less bias to correct for.

Additionally we find that the bootstrap standard errors are fairly accurate for the CCEPbc estimator but that they considerably underestimate the true standard errors for the CCEP estimator and the alternative bias corrections CCEPrm and CCEPjk. As mentioned before, the bias of the CCEP estimator is stochastic due to its dependence on the unobserved factors and this contributes to its variance. The bootstrap standard errors are unable to account for this since the randomness of the unobserved factors cannot be incorporated in the resampling scheme. To the extent that a bias remains this will also distort the bootstrapped standard errors for CCEPrm and CCEPjk, whereas CCEPbc is unaffected due to its unbiasedness.

5.3 Number of factors and their strength

The purpose of this section is to analyze the performance of the CCEPbc estimator 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 B-2 of appendix B. Like before the CCEP estimator and its bias corrections do not use the cross-section average of g_{it} when approximating the factors. To analyze the impact of adding $\bar{g}_t$ we now also include CCEP variants that do use g_{it} and denote them using the (+g) suffix.

The results in Table 2 show that the performance of the CCEP estimator 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 estimator. Table 3 further summarizes the behavior of CCEPbc for various sizes of N and T with two very strong factors. Although the small T bias disappears as N grows larger, it results in distorted inference especially when T and/or N are large. The explanation for this slight increase in bias is that even though the rank condition is exactly satisfied (2 variables 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 the CCEP estimator will have an additional finite N bias term which is not taken into account by our CCEPbc estimator.

Although the small T bias in the presence of two strong factors should disappear as N increases further, 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 $\bar{g}_t$ 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 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 with that of CCEP(+g) in Table 2 also confirms our theoretical finding that adding more cross-sectional averages to the orthogonalization matrix increases time series bias. Fortunately, the CCEPbc estimator is in both cases equally 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 that more influential factors (i.e. increasing RI from 1 to 3) does not change the bias in the one factor case (upper panel) but it will reduce the time series bias when more than one factor is present (lower panel).[13]

[13]Note that analyzing the behavior of CCEP(+g) is more fair than using the standard CCEP as poor factor approximation will lead to a positive bias term in our design. Hence, since CCEP does not deal with multiple strong factors as effectively as CCEP(+g) does, its negative time series bias is to some extent offset by a more positive factor bias when we increase RI.

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

	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	−0.403	0.420	1.00	−0.184	0.189	1.00	−0.113	0.115	1.00	−0.404	0.422	1.00	−0.184	0.189	1.00	−0.113	0.116	1.00
CCEPbc	0.008	0.056	0.06	0.001	0.014	0.04	0.000	0.008	0.04	0.009	0.057	0.06	0.000	0.014	0.04	0.000	0.008	0.04
CCEPbcr2	−0.022	0.034	0.19	−0.009	0.014	0.16	−0.005	0.008	0.10	−0.022	0.034	0.19	−0.009	0.014	0.17	−0.005	0.008	0.10
CCEPjk	0.051	0.316	0.27	0.036	0.108	0.56	0.028	0.059	0.63	0.051	0.313	0.26	0.036	0.108	0.57	0.028	0.059	0.64
CCEP(+g)	−0.410	0.431	1.00	−0.185	0.190	1.00	−0.113	0.116	1.00	−0.410	0.431	1.00	−0.185	0.190	1.00	−0.113	0.116	1.00
CCEPbc(+g)	0.010	0.066	0.06	0.001	0.014	0.02	0.000	0.008	0.04	0.010	0.064	0.06	0.001	0.014	0.03	−0.000	0.008	0.04
CCEPbcr2(+g)	−0.023	0.036	0.14	−0.009	0.014	0.15	−0.005	0.009	0.09	−0.023	0.036	0.15	−0.009	0.014	0.15	−0.005	0.009	0.10
CCEPjk(+g)	-	-	-	0.040	0.116	0.41	0.029	0.061	0.54	-	-	-	0.041	0.116	0.42	0.029	0.061	0.55
FLSbc	−0.264	0.270	0.99	−0.079	0.083	0.99	−0.039	0.041	1.00	−0.236	0.248	0.99	−0.069	0.073	1.00	−0.036	0.038	1.00
	two factors																	
	$RI = 1$									$RI = 3$								
	$T = 10$			$T = 20$			$T = 30$			$T = 10$			$T = 20$			$T = 30$		
CCEP	−0.415	0.432	1.00	−0.192	0.197	1.00	−0.118	0.121	1.00	−0.406	0.424	1.00	−0.183	0.188	1.00	−0.110	0.112	1.00
CCEPbc	0.010	0.060	0.06	0.001	0.014	0.04	0.000	0.008	0.04	0.017	0.070	0.10	0.007	0.020	0.08	0.006	0.013	0.12
CCEPbcr2	−0.021	0.033	0.18	−0.009	0.014	0.16	−0.004	0.008	0.09	0.009	0.045	0.21	0.010	0.021	0.27	0.010	0.017	0.32
CCEPjk	0.042	0.310	0.25	0.031	0.108	0.54	0.028	0.059	0.62	0.056	0.315	0.31	0.047	0.116	0.59	0.036	0.063	0.62
CCEP(+g)	−0.453	0.473	1.00	−0.210	0.216	1.00	−0.130	0.132	1.00	−0.420	0.441	1.00	−0.191	0.196	1.00	−0.116	0.118	1.00
CCEPbc(+g)	0.011	0.070	0.07	0.001	0.016	0.04	0.000	0.009	0.04	0.011	0.068	0.06	0.001	0.014	0.04	0.001	0.008	0.04
CCEPbcr2(+g)	−0.022	0.036	0.15	−0.009	0.015	0.16	−0.005	0.009	0.09	0.012	0.047	0.17	0.009	0.020	0.24	0.008	0.014	0.28
CCEPjk(+g)	-	-	-	0.036	0.121	0.49	0.031	0.066	0.59	-	-	-	0.050	0.124	0.51	0.034	0.064	0.61
FLSbc	−0.524	0.526	1.00	−0.197	0.200	1.00	−0.097	0.099	1.00	−0.499	0.504	1.00	−0.165	0.171	0.99	−0.067	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. CCEPbcr2 represents the restricted bias-correction and CCEPjk the jacknife corrected CCEP. FLSbc is the bias-corrected least squares with interactive effects estimator supplied with the correct number of factors $(m+1)$. CCEP-type estimators with suffix '(+g)' indicate that the cross-section average of g_{it} was included in the orthogonalization matrix. All CCEP variants use no lags of $\bar{x}_t$ and $\bar{g}_t$.

(iii) All test sizes use estimated standard errors based on 150 bootstrap samples.

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	0.026	0.027	0.019	0.017	0.016	0.08	0.14	0.09	0.16	0.26
50	0.034	0.020	0.017	0.015	0.014	0.11	0.11	0.11	0.21	0.36
100	0.030	0.017	0.014	0.012	0.012	0.13	0.09	0.13	0.24	0.42
500	0.017	0.007	0.006	0.005	0.005	0.10	0.08	0.12	0.16	0.28
					CCEPbc(+g)					
25	0.020	0.015	0.008	0.006	0.006	0.05	0.09	0.05	0.06	0.08
50	0.023	0.008	0.004	0.004	0.003	0.07	0.06	0.04	0.06	0.07
100	0.025	0.005	0.002	0.002	0.002	0.08	0.04	0.05	0.06	0.06
500	0.011	0.001	0.001	0.000	0.000	0.06	0.04	0.04	0.05	0.04

Notes:

(i) 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 ($RI = 3$). We display results for estimating ρ.

(ii) CCEPbc is the unrestricted bias-corrected CCEP estimator. Estimators followed by '(+g)' indicate that the cross-section average of g_{it} was included in the orthogonalization matrix. All CCEP variants use no lags of $\bar{x}_t$ and $\bar{g}_t$.

(iii) All test sizes use estimated standard errors based on 150 bootstrap samples.

5.4 Restricted bias-corrections and rank condition

In this section we compare the performance of the unrestricted bias correction CCEPbc to that of the restricted versions CCEPbcr1 and CCEPbcr2 derived in Section 4.2 and analyze the implications of rank deficiency. As before we also report results for variants that add the additional cross-sectional average to the orthogonalization matrix.

Table 2 can be used to compare the performance of CCEPbc to CCEPbcr2 for a setting with one and two common factors. The distinction between these scenarios is of interest since CCEPbcr2 is derived under the assumption that only one factor is present whereas CCEPbc is applicable irrespectively of the number of factors (provided that the rank condition is satisfied). In general, we find that CCEPbcr2 is a fairly accurate bias correction method, even in the case of two factors. Comparing the unrestricted and restricted version shows some tradeoff between bias and variance, though. CCEPbc dominates in terms of bias correction but has a downside that the estimator $\widehat{\sigma}^2_{\breve{y}_{-1}}$ used in the denominator of equation (25) introduces uncertainty in small samples. CCEPbcr2 has a smaller variance as it imposes a specific form for the denominator in equation (30) but is slightly less effective as a bias correction method because of the truncation error made in the estimation of C and the resulting finite T inconsistency. Because this small bias is offset by the lower variance (on a rmse basis) in small samples (also see Table B-2 for $N = 25$), CCEPbcr2 may still be an interesting alternative to CCEPbc. As N grows large, however, this relative efficiency only compensates for bias when the single factor assumption is true (see upper panel of Table 2) or when the factors are not too strong in case $m > 1$ (see lower left panel of Table 2). Moreover, as a result of the small inconsistency for finite T, CCEPbcr2 has a size distortion especially when N is large. For the unrestricted version, inference is reliable in all settings (although this may require adding $\bar{g}_t$), but at the cost of a higher variance.

We next analyze the performance of the restricted correction CCEPbcr1, which was derived under the assumption of a single common factor and no covariates. Under these assumptions CCEPbcr1 is consistent for $N \to \infty$. Table 4 reports simulation results for a pure AR(1) model ($\beta = 0$) with one or two strong factors for $N = 500$ (see Table B-3 for $N = 25$). Note that as there is only one variable in this model, the rank condition is no longer satisfied in the two factor setting. Because x_{it} is excluded as a regressor but still holds information about the common factors, the (+xg) suffix now denotes CCEP-type estimators that use both the cross-sectional averages $\bar{x}_t$ and $\bar{g}_t$ in the orthogonalization matrix. The results confirm that in the single factor setting (see upper panel Table 4) the restricted correction is perfectly centered for all T with an actual size very close or even at the 5% nominal level. As expected, however, CCEPbcr1 is less effective in the multifactor case (see lower panel Table 4), i.e. it has a small positive bias resulting in a considerable size distortion. This is due to the assumption of a single factor that is no longer satisfied and hence using the information in $\bar{x}_t$ and $\bar{g}_t$ to satisfy the rank condition does not solve the problem.

The results in Table 4 show that also the unrestricted CCEPbc estimator suffers from a bias and size distortion in the setting with two common factors. As this is due to the rank deficiency, which can be solved by adding additional cross-sectional averages, the CCEPbc(+xg) again has satisfactory finite sample properties. Note that due to the dominance of the factors in the pure AR(1) model (3/5 of the variation in $\mathbf{y}$ is due to the factors) the inclusion of additional covariates for orthogonalization is even more important than in the model including x_{it} as a covariate. Although the performance of the CCEPbc estimator improves with the inclusion of $\bar{g}_t$ (+g), it is only satisfactory when both $\bar{x}_t$ and $\bar{g}_t$ are used. In line with the results for the model with exogenous regressors this suggests that even though an exactly satisfied rank condition may be sufficient for consistent estimation, it may be preferable to include more cross-sectional averages than required for the rank condition to hold.

5.5 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 cross-sectional averages to adequately capture the common factors. Table 5 reports the main results in a setting where factors are very strong ($RI = 3$) and $N = 500$. Results for $N = 25$ are reported in Table B-4. 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 $\mathbf{M}_Q$. The suffix notation $\_p_T$ is used to indicate the inclusion of $p_T = T^{1/3}$

Table 4: Monte Carlo results for ρ : pure AR(1) model 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	−0.437	0.452	1.00	−0.203	0.208	1.00	−0.125	0.128	1.00	−0.070	0.070	1.00
CCEPbc	0.009	0.056	0.12	0.001	0.015	0.05	−0.000	0.008	0.05	−0.000	0.005	0.05
CCEPbcr1	0.000	0.026	0.05	−0.000	0.011	0.05	−0.000	0.008	0.06	−0.000	0.005	0.06
CCEP(+g)	−0.440	0.459	1.00	−0.203	0.209	1.00	−0.126	0.128	1.00	−0.069	0.070	1.00
CCEPbc(+g)	0.008	0.061	0.08	0.001	0.016	0.04	−0.000	0.009	0.04	−0.000	0.005	0.05
CCEPbcr1(+g)	−0.000	0.028	0.04	0.000	0.012	0.05	−0.000	0.008	0.05	−0.000	0.005	0.05
CCEP(+xg)	−0.446	0.468	1.00	−0.204	0.210	1.00	−0.126	0.128	1.00	−0.070	0.070	1.00
CCEPbc(+xg)	0.007	0.066	0.05	0.001	0.017	0.03	−0.000	0.009	0.04	−0.000	0.005	0.05
CCEPbcr1(+xg)	−0.000	0.031	0.03	0.000	0.012	0.05	−0.000	0.008	0.05	−0.000	0.005	0.05
	two factors											
	$T = 10$			$T = 20$			$T = 30$			$T = 50$		
CCEP	−0.400	0.420	1.00	−0.167	0.177	0.99	−0.090	0.098	0.98	−0.038	0.044	0.91
CCEPbc	0.041	0.103	0.39	0.042	0.068	0.54	0.037	0.053	0.69	0.032	0.039	0.85
CCEPbcr1	0.049	0.089	0.41	0.039	0.056	0.62	0.036	0.048	0.74	0.032	0.038	0.87
CCEP(+g)	−0.429	0.448	1.00	−0.192	0.198	1.00	−0.115	0.118	1.00	−0.061	0.062	1.00
CCEPbc(+g)	0.020	0.078	0.15	0.009	0.025	0.07	0.007	0.015	0.13	0.006	0.010	0.17
CCEPbcr1(+g)	0.066	0.102	0.41	0.036	0.048	0.59	0.026	0.032	0.65	0.016	0.019	0.67
CCEP(+xg)	−0.442	0.464	1.00	−0.199	0.205	1.00	−0.121	0.123	1.00	−0.066	0.067	1.00
CCEPbc(+xg)	0.011	0.072	0.07	0.002	0.016	0.04	0.001	0.009	0.05	0.001	0.005	0.05
CCEPbcr1(+xg)	0.071	0.107	0.39	0.036	0.046	0.62	0.023	0.029	0.65	0.013	0.015	0.65

Notes:

(i) Data for this experiment are generated using a pure AR(1) model with $\rho = 0.8$, $\beta = 0$ and $\lambda = 0$. 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. CCEPbcr1 represents the restricted bias-correction in the AR(1) model. Estimators followed by '(+g)' also use the cross-section average of g_{it} in the orthogonalization matrix whereas '(+xg)' indicates that both $\bar{x}_t$ and $\bar{g}_t$ are included. All CCEP variants use no lags of $\bar{x}_t$ and $\bar{g}_t$.

(iii) All test sizes use estimated standard errors based on 150 bootstrap samples.

lags while $\_p_0$ denotes the misspecified variant without lags of $\bar{\mathbf{Z}}$. We report results for CCEP-type estimators that add the cross-sectional averages 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 an insufficient number of degrees of freedom (because of the larger number of cross-sectional averages 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

Table 5: 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)	−0.590	0.610	0.99	−0.254	0.262	1.00	−0.147	0.150	1.00	−0.077	0.078	1.00
CCEP_p_1(+g)	−0.683	0.713	0.95	−0.272	0.280	1.00	−0.153	0.157	1.00	−0.078	0.079	1.00
CCEP_p_T(+g)	-	-	-	−0.337	0.350	0.99	−0.204	0.210	1.00	−0.091	0.093	1.00
CCEPbc_p_0(+g)	0.015	0.090	0.05	0.000	0.017	0.03	−0.001	0.009	0.03	−0.000	0.005	0.04
CCEPbc_p_1(+g)	0.028	0.141	0.03	0.001	0.021	0.02	−0.000	0.009	0.03	0.000	0.005	0.03
CCEPbc_p_T(+g)	-	-	-	0.008	0.046	0.04	−0.000	0.013	0.03	−0.000	0.006	0.03
CCEPjk_p_1(+g)	-	-	-	0.141	0.227	0.27	0.091	0.124	0.43	0.039	0.050	0.58
FLSbc	−0.248	0.257	0.98	−0.062	0.066	0.99	−0.031	0.032	0.99	−0.015	0.015	0.95
	two factors											
	$T = 10$			$T = 20$			$T = 30$			$T = 50$		
CCEP_p_0(+g)	−0.643	0.661	1.00	−0.289	0.297	1.00	−0.171	0.174	1.00	−0.091	0.092	1.00
CCEP_p_1(+g)	−0.766	0.794	0.99	−0.319	0.328	1.00	−0.179	0.183	1.00	−0.089	0.090	1.00
CCEP_p_T(+g)	-	-	-	−0.397	0.410	1.00	−0.245	0.252	1.00	−0.105	0.107	1.00
CCEPbc_p_0(+g)	−0.021	0.097	0.14	−0.019	0.027	0.23	−0.014	0.017	0.35	−0.009	0.011	0.44
CCEPbc_p_1(+g)	0.032	0.152	0.06	0.002	0.027	0.03	−0.000	0.010	0.04	−0.000	0.005	0.05
CCEPbc_p_T(+g)	-	-	-	0.011	0.057	0.06	−0.000	0.014	0.03	−0.000	0.006	0.05
CCEPjk_p_1(+g)	-	-	-	0.137	0.232	0.48	0.109	0.146	0.71	0.051	0.062	0.81
FLSbc	−0.515	0.519	1.00	−0.167	0.172	1.00	−0.070	0.073	0.99	−0.022	0.023	0.97

Notes:

(i) 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 $T^{1/3}$ lags of $\bar{x}_t$ and $\bar{g}_t$ in the orthogonalization matrix.

(iii) All test sizes use estimated standard errors based on 150 bootstrap samples.

the single factor case but that it is not sufficient to control for multiple strong factors without lags of $\bar{\mathrm{x}}_t$ (and $\bar{\mathrm{g}}_t$). The correctly specified CCEPbc_p_1 estimator instead performs much better, with an adequate size for all values of T. This confirms our theoretical result that the approximation of the factors requires the number of lagged cross-sectional averages 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 let the number of lags $p_T = T^{1/3}$ grow with T as a precaution against misspecification. As this implies orthogonalization on a large number of cross-section averages, the resulting bias of the uncorrected CCEP_p_T estimator is very large. CCEPbc_p_T is however highly effective in removing this bias and has an adequate size. The price paid for this robustness is that the larger number of cross-sectional averages translates in a substantially higher variance compared to the correctly specified CCEPbc_p_1. As expected, this difference disappears as T grows. Results are highly similar for small N (with marginally larger biases) but whenever some bias remains this has a much smaller impact on inference.

6 Conclusion

In this article we have extended the CCEP estimator designed by Pesaran (2006) to dynamic homogeneous panel data models and developed a bias-corrected version that eliminates its finite T bias. We first showed that in homogeneous dynamic panels, the unobserved common factors can be effectively approximated by cross-sectional averages of the observed data provided that a sufficient number of observables is available (rank condition) and an appropriate number of lagged cross-sectional averages 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 showed that the resulting CCEPbc estimator is consistent as $N \to \infty$, both for T fixed or $T \to \infty$, and derived its limiting distribution. Our CCEPbc estimator is applicable in a model with multiple common factors and exogenous covariates but we additionally derived restricted versions for the case of a single common factor in a model with or without covariates.

Extensive Monte Carlo experiments showed that, when appropriately specified, our CCEPbc estimator performs very well and is superior to the original CCEP estimator and to alternative bias corrections available in the literature. More specifically, CCEPbc was found to be nearly unbiased across all of the sample sizes and designs we have 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 moderately 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 the CCEPbc estimator was found to be close to the 5% nominal level. The Monte Carlo simulations further showed that it is important to include a sufficient number of cross-sectional averages 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 cross-sectional averages of additional observables when these hold information about highly influential common factors. Second, the simulation results confirm our theoretical finding that lagged cross-sectional averages 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. Finally, it was shown that in a model with a single common factor, restricted CCEPbc versions can yield a further improvement in terms of efficiency. This is especially the case in a model without covariates.

7 Acknowledgements

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. Ignace De Vos and Gerdie Everaert further acknowledge financial support from the National Bank of Belgium.

References

Ahn, S. C., Lee, Y. H., and Schmidt, P. (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics*, 174(1):1–14.

Andrews, D. W. K. (1987). Asymptotic results for generalized wald tests. *Econometric Theory*, 3(3):348–358.

Bai, J. (2009). Panel Data Models with Interactive Fixed Effects. *Econometrica*, 77(4):1229–1279.

Bai, J. (2013). Likelihood approach to dynamic panel models with interactive effects. MPRA Paper 50267, University Library of Munich, Germany.

Bun, M. and Carree, M. (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business and Economic Statistics*, 23(2):200–210.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, 18(1):5 – 46.

Chudik, A., Pesaran, M., and Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal*, 14(1):C45–C90.

Chudik, A. and Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, 188(2):393 – 420. Heterogeneity in Panel Data and in Nonparametric Analysis in honor of Professor Cheng Hsiao.

Dhaene, G. and Jochmans, K. (2014). Split-Panel Jackknife Estimation of Fixed-Effect Models. Sciences Po Economics Discussion Papers 2014-03, Sciences Po Departement of Economics.

Everaert, G. and De Groote, T. (2016). Common correlated effects estimation of dynamic panels with cross-sectional dependence. *Econometric Reviews*, 35(3):428–463.

Harding, M. and Lamarche, C. (2011). Least squares estimation of a panel data model with multifactor error structure and endogenous covariates. *Economics Letters*, 111(3):197–199.

Hayakawa, K., Pesaran, H., and Smith, V. (2014). Transformed Maximum Likelihood Estimation of Short Dynamic Panel Data Models with interactive effects. Cambridge Working Papers in Economics 1412, Faculty of Economics, University of Cambridge.

Holtz-Eakin, D., Newey, W., and Rosen, H. (1988). Estimating vector autoregressions with panel data. *Econometrica*, 56(6):1371–1395.

Kapetanios, G. (2008). A bootstrap procedure for panel data sets with many cross-sectional units. *Econometrics Journal*, 11(2):377–395.

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

Karabiyik, H., Reese, S., and Westerlund, J. (2016). On the Role of the Rank Condition in CCE Estimation of Factor-Augmented Panel Regressions. Unpublished manuscript.

Karabiyik, H., Urbain, J., and Westerlund, J. (2014). CCE estimation of factor-augmented regression models with more factors than observables. Research Memorandum 007, Maastricht University, Graduate School of Business and Economics (GSBE).

Moon, H. R. and Weidner, M. (2014). Dynamic linear panel regression models with interactive fixed effects. CeMMAP working papers CWP47/14, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Nauges, C. and Thomas, A. (2003). Consistent estimation of dynamic panel data models with time-varying individual effects. *Annales d'Economie et de Statistique*, (70):53–75.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, 49(6):1417–1426.

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

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.

Pesaran, M. and Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161(2):182–202.

Phillips, P. C. B. and Sul, D. (2007). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics*, 137(1):162–188.

Robertson, D. and Sarafidis, V. (2015). IV estimation of panels with factor residuals. *Journal of Econometrics*, 185(2):526–541.

So, B. and Shin, D. (1999). Recursive mean adjustment in time-series inferences. *Statistics & Probability Letters*, 43(1):65 – 73.

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

Appendices

Appendix A Proofs

Definitions

We first introduce some notation that will be used later on. We restrict the notation to settings with $p = 1$ but generalizations follow straightforwardly. With $p = 1$ model (1)-(3) can be written in VAR(1) form

$$\begin{bmatrix} 1 & -\boldsymbol{\beta}^{*'} \\ \mathbf{0}_{k\times 1} & \mathbf{I}_k \end{bmatrix} \begin{bmatrix} y_{it} \\ \mathbf{z}_{it} \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \mathbf{c}_{z,i} \end{bmatrix} + \begin{bmatrix} \rho & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \boldsymbol{\lambda} \end{bmatrix} \begin{bmatrix} y_{it-1} \\ \mathbf{z}_{it-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\gamma}_i' \\ \boldsymbol{\Gamma}_i' \end{bmatrix} \mathbf{f}_t + \begin{bmatrix} \varepsilon_{it} \\ \mathbf{v}_{it} \end{bmatrix},$$

with $\boldsymbol{\beta}^* = (\boldsymbol{\beta}', \mathbf{0}_{1\times k_g})'$ and the associated more compact form

$$\mathbf{A}_0 \mathbf{d}_{it} = \mathbf{c}_{d,i} + \boldsymbol{\Theta} L \mathbf{d}_{it} + \mathbf{C}_i \mathbf{f}_t + \mathbf{u}_{it},$$

where $\mathbf{A}_0 = \begin{bmatrix} 1 & -\boldsymbol{\beta}^{*'} \\ \mathbf{0}_{k\times 1} & \mathbf{I}_k \end{bmatrix}$, $\mathbf{c}_{d,i} = (\alpha_i, \mathbf{c}_{z,i}')'$, $\mathbf{d}_{it} = (y_{it}, \mathbf{z}_{it}')'$, $\boldsymbol{\Theta} = \begin{bmatrix} \rho & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \boldsymbol{\lambda} \end{bmatrix}$, $\mathbf{C}_i = (\boldsymbol{\gamma}_i, \boldsymbol{\Gamma}_i)'$ and $\mathbf{u}_{it} = (\varepsilon_{it}, \mathbf{v}_{it}')'$. Since $\mathbf{A}_0$ is invertible,

$$\mathbf{d}_{it} = \mathbf{A}_0^{-1}\mathbf{c}_{d,i} + \mathbf{A}_0^{-1}\boldsymbol{\Theta} L \mathbf{d}_{it} + \mathbf{A}_0^{-1}\mathbf{C}_i \mathbf{f}_t + \mathbf{A}_0^{-1}\mathbf{u}_{it},$$

which can be rewritten further as

$$\begin{aligned} (\mathbf{I}_{1+k} - \boldsymbol{\Theta}^* L)\mathbf{d}_{it} &= \mathbf{c}_{d,i}^* + \mathbf{C}_i^* \mathbf{f}_t + \mathbf{u}_{it}^*, \\ \boldsymbol{\Theta}(L)\mathbf{d}_{it} &= \mathbf{c}_{d,i}^* + \mathbf{C}_i^* \mathbf{f}_t + \mathbf{u}_{it}^*, \end{aligned}$$

where the terms with an asterisk are defined as $\boldsymbol{\Theta}^* = \mathbf{A}_0^{-1}\boldsymbol{\Theta}$ and with $\boldsymbol{\Theta}(L) = \mathbf{I}_{1+k} - \boldsymbol{\Theta}^* L$. Then, as $\boldsymbol{\Theta}(L)$ is invertible by Assumption 5 we obtain the reduced form

$$\begin{aligned} \boldsymbol{\Theta}(L)\mathbf{d}_{it} &= \mathbf{c}_{d,i}^* + \mathbf{C}_i^* \mathbf{f}_t + \mathbf{u}_{it}^*, \\ \mathbf{d}_{it} &= \boldsymbol{\Theta}(L)^{-1}\mathbf{c}_{d,i}^* + \boldsymbol{\Theta}(L)^{-1}\mathbf{C}_i^* \mathbf{f}_t + \boldsymbol{\Theta}(L)^{-1}\mathbf{u}_{it}^*, \end{aligned}$$

with its cross-section average given by

$$\begin{aligned} \bar{\mathbf{d}}_t &= \boldsymbol{\Theta}(L)^{-1}\bar{\mathbf{c}}_d^* + \boldsymbol{\Theta}(L)^{-1}\bar{\mathbf{C}}^* \mathbf{f}_t + \boldsymbol{\Theta}(L)^{-1}\bar{\mathbf{u}}_t^*, \\ \bar{\mathbf{d}}_t &= \ddot{\mathbf{c}}_d + \ddot{\mathbf{C}}\mathbf{f}_t + \ddot{\mathbf{u}}_t, \end{aligned} \quad \text{(A-1)}$$

where $\ddot{\mathbf{C}} = \boldsymbol{\Theta}(L)^{-1}\bar{\mathbf{C}}^*$ and similarly for other variables in (A-1). Using the notation above and setting $\mathbf{D} = (\bar{\mathbf{d}}_1, ..., \bar{\mathbf{d}}_T)'$, the matrix of cross-section averages $\mathbf{Q}$ is

$$\mathbf{Q} = (\boldsymbol{\iota}_T, \mathbf{D}, \ldots, \mathbf{D}_{-p_T}) = \begin{pmatrix} 1 & \bar{\mathbf{d}}_1' & \bar{\mathbf{d}}_0' & \ldots & \bar{\mathbf{d}}_{1-p_T}' \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \bar{\mathbf{d}}_T' & \bar{\mathbf{d}}_{T-1}' & \ldots & \bar{\mathbf{d}}_{T-p_T}' \end{pmatrix}.$$

Also, defining $\mathbf{Q}_{\mathrm{f}} = \check{\mathbf{F}}\ddot{\mathbf{P}}$ with

$$\check{\mathbf{F}} = (\boldsymbol{\iota}_T, \mathbf{F}, \ldots, \mathbf{F}_{-p_T}) = \begin{pmatrix} 1 & \mathbf{f}_1^{'} & \mathbf{f}_0^{'} & \ldots & \mathbf{f}_{1-p_T}^{'} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \mathbf{f}_T^{'} & \mathbf{f}_{T-1}^{'} & \ldots & \mathbf{f}_{T-p_T}^{'} \end{pmatrix}, \tag{A-2}$$

and

$$\ddot{\mathbf{P}} = \begin{pmatrix} 1 & \ddot{\mathbf{c}}_{\mathrm{d}}^{'} & \ddot{\mathbf{c}}_{\mathrm{d}}^{'} & \ldots & \ddot{\mathbf{c}}_{\mathrm{d}}^{'} \\ \mathbf{0}_{m\times 1} & \ddot{\mathbf{C}}^{'} & \mathbf{0}_{m\times(1+k)} & \ldots & \mathbf{0}_{m\times(1+k)} \\ \mathbf{0}_{m\times 1} & \mathbf{0}_{m\times(1+k)} & \ddot{\mathbf{C}}^{'} & & \mathbf{0}_{m\times(1+k)} \\ \vdots & \vdots & & \ddots & \vdots \\ \mathbf{0}_{m\times 1} & \mathbf{0}_{m\times(1+k)} & \mathbf{0}_{m\times(1+k)} & \ldots & \ddot{\mathbf{C}}^{'} \end{pmatrix}, \tag{A-3}$$

the matrix of cross-section averages can be decomposed into

$$\mathbf{Q} = \mathbf{Q}_{\mathrm{f}} + \ddot{\mathbf{U}} = \check{\mathbf{F}}\ddot{\mathbf{P}} + \ddot{\mathbf{U}}, \tag{A-4}$$

where

$$\ddot{\mathbf{U}} = \begin{pmatrix} 0 & \ddot{\mathbf{u}}_1^{'} & \ddot{\mathbf{u}}_0^{'} & \ldots & \ddot{\mathbf{u}}_{1-p_T}^{'} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ddot{\mathbf{u}}_T^{'} & \ddot{\mathbf{u}}_{T-1}^{'} & \ldots & \ddot{\mathbf{u}}_{T-p_T}^{'} \end{pmatrix}. \tag{A-5}$$

Next, turning to the explanatory variables in the model

$$\begin{aligned} y_{it-1} &= \left[\alpha_i^{+} + \rho(L)^{-1}\boldsymbol{\beta}^{*'}\mathbf{c}_{z,i}^{+,\lambda}\right] + \rho(L)^{-1}\left[\boldsymbol{\gamma}_i' + \boldsymbol{\beta}^{*'}\boldsymbol{\lambda}(L)^{-1}\boldsymbol{\Gamma}_i'\right]\mathbf{f}_{t-1} + \left[\varepsilon_{it-1}^{+} + \rho(L)^{-1}\boldsymbol{\beta}^{*'}\mathbf{v}_{it-1}^{+,\lambda}\right], \\ &= \alpha_i^{*} + \boldsymbol{\gamma}_i^{*'}\mathbf{f}_{t-1} + \vartheta_{it-1}, \end{aligned} \tag{A-6}$$

where $\alpha_i^{*} = \alpha_i^{+} + \rho(L)^{-1}\boldsymbol{\beta}^{*'}\mathbf{c}_{z,i}^{+,\lambda}$, $\boldsymbol{\gamma}_i^{*} = \rho(L)^{-1}\left[\boldsymbol{\gamma}_i + \boldsymbol{\Gamma}_i\boldsymbol{\lambda}(L)^{-1'}\boldsymbol{\beta}^{*}\right]$, $\vartheta_{it-1} = \varepsilon_{it-1}^{+} + \rho(L)^{-1}\boldsymbol{\beta}^{*'}\mathbf{v}_{it-1}^{+,\lambda}$ and all variables with a $^{+,\lambda}$ superscript are defined as $\mathbf{c}_{z,i}^{+,\lambda} = \boldsymbol{\lambda}(L)^{-1}\mathbf{c}_{z,i}$. Similarly for the vector of exogenous variables

$$\mathbf{z}_{it} = \mathbf{c}_{z,i}^{+,\lambda} + \boldsymbol{\lambda}(L)^{-1}\boldsymbol{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{it}^{+,\lambda}, \tag{A-7}$$

so that with $\mathbf{F}^{\mathrm{d}} = (\boldsymbol{\iota}_T, \mathbf{F}, \mathbf{F}_{-1})$ and $\boldsymbol{\Pi}_i^{y} = (\alpha_i^{*}, \mathbf{0}_{m\times 1}', \boldsymbol{\gamma}_i^{*'})'$ we have the following shorthand for $\mathbf{y}_{i,-1}$,

$$\mathbf{y}_{i,-1} \;=\; \begin{pmatrix} 1 & \mathbf{f}_1^{'} & \mathbf{f}_0^{'} \\ \vdots & \vdots & \vdots \\ 1 & \mathbf{f}_T^{'} & \mathbf{f}_{T-1}^{'} \end{pmatrix}\begin{pmatrix} \alpha_i^{*} \\ \mathbf{0}_{m\times 1} \\ \boldsymbol{\gamma}_i^{*} \end{pmatrix} + \begin{pmatrix} \vartheta_{i0} \\ \vdots \\ \vartheta_{iT-1} \end{pmatrix} = \mathbf{F}^{\mathrm{d}}\boldsymbol{\Pi}_i^{y} + \boldsymbol{\vartheta}_{i,-1}, \tag{A-8}$$

where $\boldsymbol{\vartheta}_{i,-1} = (\vartheta_{i0}, ..., \vartheta_{iT-1})'$ collects the reduced form errors. In similar fashion for $\mathbf{Z}_i$

$$\mathbf{Z}_i \;=\; \mathbf{F}^{\mathrm{d}}\begin{pmatrix} \mathbf{c}_{z,i}^{+,\lambda'} \\ \boldsymbol{\Gamma}_i\boldsymbol{\lambda}^{'}(L)^{-1} \\ \mathbf{0}_{m\times k} \end{pmatrix} + \begin{pmatrix} \mathbf{v}_{i1}^{+,\lambda'} \\ \vdots \\ \mathbf{v}_{iT}^{+,\lambda'} \end{pmatrix} = \mathbf{F}^{\mathrm{d}}\boldsymbol{\Pi}_i^{z} + \mathbf{V}_i^{+,\lambda}. \tag{A-9}$$

with $\mathbf{\Pi}_i^z = (\mathbf{c}_{z,i}^{+,\lambda}, \boldsymbol{\lambda}(L)^{-1}\mathbf{\Gamma}_i', \mathbf{0}_{k\times m})'$ and $\mathbf{V}_i^{+,\lambda} = (\mathbf{v}_{i1}^{+,\lambda}, ..., \mathbf{v}_{iT}^{+,\lambda})'$. Finally, when the rank condition is satisfied (Assumtion 4) the factors can be written as

$$\mathbf{F} \;=\; \mathbf{Q}_{\mathrm{F}}\bar{\mathbf{P}}^* - \bar{\mathbf{U}}^*\bar{\mathbf{C}}^*(\bar{\mathbf{C}}^{*'}\bar{\mathbf{C}}^*)^{-1}, \tag{A-10}$$

with, for $p=1$, $\mathbf{Q}_{\mathrm{F}} = (\boldsymbol{\iota}_T, \mathbf{D}, \mathbf{D}_{-1})$, $\bar{\mathbf{P}}^* = (-\bar{\mathbf{c}}_{\mathrm{d}}^*, \mathbf{I}_{1+k}, -\mathbf{\Theta}^*)'\bar{\mathbf{C}}^*(\bar{\mathbf{C}}^{*'}\bar{\mathbf{C}}^*)^{-1}$ and $\bar{\mathbf{U}}^* = (\bar{\mathbf{u}}_1^*, ..., \bar{\mathbf{u}}_T^*)'$.

Lemmas

Lemma 1. Under Assumptions 1-5 and with $p_T \geq p$ it holds that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{w}_i'\mathbf{M}_{\mathrm{Q}}\mathbf{F}}{T}\boldsymbol{\gamma}_i \to \mathbf{0} \qquad \text{as} \qquad N\to\infty. \tag{A-11}$$

Proof: From equation (4) in Assumption 3 we have $\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_i$, with $\boldsymbol{\gamma}$ a bounded constant and $\boldsymbol{\eta}_i \sim \mathrm{IID}\left(\mathbf{0}, \mathbf{\Omega}_\eta\right)$. It follows that

$$\mathbf{\Delta}_{\mathrm{F}} \;=\; \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{w}_i'\mathbf{M}_{\mathrm{Q}}\mathbf{F}}{T}\boldsymbol{\gamma}_i = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{w}_i'\mathbf{M}_{\mathrm{Q}}\mathbf{F}}{T}\boldsymbol{\eta}_i,$$

since $N^{-1}\sum_{i=1}^{N}\mathbf{w}_i'\mathbf{M}_{\mathrm{Q}}\mathbf{F}\boldsymbol{\gamma} = \bar{\mathbf{w}}'\mathbf{M}_{\mathrm{Q}}\mathbf{F}\boldsymbol{\gamma} = \mathbf{0}$ because $\mathbf{M}_{\mathrm{Q}}\bar{\mathbf{w}} = \mathbf{0}$. Then, by (A-10)

$$\mathbf{\Delta}_{\mathrm{F}} \;=\; -\frac{1}{NT}\sum_{i=1}^{N}\mathbf{w}_i'\mathbf{M}_{\mathrm{Q}}\bar{\mathbf{U}}^*\bar{\mathbf{C}}^*(\bar{\mathbf{C}}^{*'}\bar{\mathbf{C}}^*)^{-1}\boldsymbol{\eta}_i, \tag{A-12}$$

because the rank condition holds by Assumption 4 and $\mathbf{M}_{\mathrm{Q}}\mathbf{Q}_{\mathrm{F}}\bar{\mathbf{P}}^* = \mathbf{0}$ when $\mathbf{Q}_{\mathrm{F}} \subseteq \mathbf{Q}$, which shows that next to the rank condition it is also required that $p_T \geq p$. We split (A-12) as $\mathbf{\Delta}_{\mathrm{F}} = (\mathbf{\Delta}_{\mathrm{F,y}}, \mathbf{\Delta}_{\mathrm{F,x}}')'$

$$\mathbf{\Delta}_{\mathrm{F,y}} \;=\; -\frac{1}{NT}\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{M}_{\mathrm{Q}}\bar{\mathbf{U}}^*\bar{\mathbf{C}}^*(\bar{\mathbf{C}}^{*'}\bar{\mathbf{C}}^*)^{-1}\boldsymbol{\eta}_i,$$
$$\mathbf{\Delta}_{\mathrm{F,x}} \;=\; -\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_{i,-1}'\mathbf{M}_{\mathrm{Q}}\bar{\mathbf{U}}^*\bar{\mathbf{C}}^*(\bar{\mathbf{C}}^{*'}\bar{\mathbf{C}}^*)^{-1}\boldsymbol{\eta}_i,$$

where $\bar{\mathbf{C}}^*(\bar{\mathbf{C}}^{*'}\bar{\mathbf{C}}^*)^{-1}$ is a constant $(1+k)\times m$ matrix over all $i = 1, \ldots, N$. Then, setting

$$\boldsymbol{\eta}_i^* = \bar{\mathbf{C}}^*(\bar{\mathbf{C}}^{*'}\bar{\mathbf{C}}^*)^{-1}\boldsymbol{\eta}_i, \tag{A-13}$$

the equations for $\mathbf{\Delta}_{\mathrm{F,y}}$ and $\mathbf{\Delta}_{\mathrm{F,x}}$ can be rewritten further as

$$\mathbf{\Delta}_{\mathrm{F,y}} \;=\; -\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{y}_{i,-1}'\bar{\mathbf{U}}^*}{T}\boldsymbol{\eta}_i^* + \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{y}_{i,-1}'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\bar{\mathbf{U}}^*}{T}\boldsymbol{\eta}_i^*, \tag{A-14}$$
$$\mathbf{\Delta}_{\mathrm{F,x}} \;=\; -\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{X}_i'\bar{\mathbf{U}}^*}{T}\boldsymbol{\eta}_i^* + \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{X}_i'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\bar{\mathbf{U}}^*}{T}\boldsymbol{\eta}_i^*. \tag{A-15}$$

Turning to the first term of $\mathbf{\Delta}_{\mathrm{F,y}}$ in equation (A-14), it holds that

$$\frac{\mathbf{y}_{i,-1}'\bar{\mathbf{U}}^*}{T}\boldsymbol{\eta}_i^* = \mathbf{\Pi}_i^{y'}\frac{\mathbf{F}^{\mathrm{d}'}\bar{\mathbf{U}}^*}{T}\boldsymbol{\eta}_i^* + \frac{\boldsymbol{\vartheta}_{i,-1}'\bar{\mathbf{U}}^*}{T}\boldsymbol{\eta}_i^* = \mathbf{\Pi}_i^{y'}\frac{\mathbf{F}^{\mathrm{d}'}\bar{\mathbf{U}}^*}{T}\boldsymbol{\eta}_i^* + O_p\left(\frac{1}{\sqrt{NT}}\right).$$

However, note that $\mathbf{F}^{\mathrm{d}}\mathbf{\Pi}_i^y$ is not independent of $\boldsymbol{\eta}_i^*$. To see this, in the simplified case with $\alpha_i = 0$ and $\boldsymbol{\beta}^* = \mathbf{0}$, then $\mathbf{\Pi}_i^{y'}\mathbf{F}^{\mathrm{d}'} = \boldsymbol{\gamma}_i'\mathbf{F}_{-1}^{+'} = (\boldsymbol{\gamma}+\boldsymbol{\eta}_i)'\mathbf{F}_{-1}^{+'}$ and

$$\mathbf{\Pi}_i^{y'}\frac{\mathbf{F}^{\mathrm{d}'}\bar{\mathbf{U}}^*}{T}\boldsymbol{\eta}_i = \boldsymbol{\gamma}'\frac{\mathbf{F}_{-1}^{+'}\bar{\mathbf{U}}^*}{T}\boldsymbol{\eta}_i + \boldsymbol{\eta}_i'\frac{\mathbf{F}_{-1}^{+'}\bar{\mathbf{U}}^*}{T}\boldsymbol{\eta}_i,$$

where $\mathbf{F}^{+'}_{-1}\bar{\mathbf{U}}^{*}/T = O_p\left(\frac{1}{\sqrt{NT}}\right)$. Hence, in finite T settings $\sqrt{N}\mathbf{F}^{+'}_{-1}\bar{\mathbf{U}}^{*}/T = O_p(1)$ and

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{\Pi}_i^{y'}\frac{\mathbf{F}^{d'}\bar{\mathbf{U}}^{*}}{T}\boldsymbol{\eta}_i = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\gamma}'\frac{\sqrt{N}\mathbf{F}^{+'}_{-1}\bar{\mathbf{U}}^{*}}{T}\boldsymbol{\eta}_i + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}_i'\frac{\sqrt{N}\mathbf{F}^{+'}_{-1}\bar{\mathbf{U}}^{*}}{T}\boldsymbol{\eta}_i.$$

It is then easily seen that the first term tends to zero as $N \to \infty$ but the second term is $O_p(1)$ since the loadings are correlated. For example, simplifying further to $m = 1$, $\boldsymbol{\eta}_i$ becomes a scalar and

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{\Pi}_i^{y'}\frac{\mathbf{F}^{d'}\bar{\mathbf{U}}^{*}}{T}\boldsymbol{\eta}_i = \frac{1}{N}\sum_{i=1}^{N}\gamma\eta_i\frac{\sqrt{N}\mathbf{F}^{+'}_{-1}\bar{\mathbf{U}}^{*}}{T} + \frac{1}{N}\sum_{i=1}^{N}\eta_i^2\frac{\sqrt{N}\mathbf{F}^{+'}_{-1}\bar{\mathbf{U}}^{*}}{T}.$$

Then, since $\sqrt{N}\mathbf{F}^{+'}_{-1}\bar{\mathbf{U}}^{*}/T$ is independent of η_i and $\frac{1}{N}\sum_{i=1}^{N}\gamma\eta_i \to 0$ and $\frac{1}{N}\sum_{i=1}^{N}\eta_i^2 \to \sigma_\eta^2 = \mathbf{\Omega}_\eta$, it is clear that the second term will not disappear since $0 < \sigma_\eta^2 < \infty$ by Assumption 3. Consequently,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbf{y}_{i,-1}'\bar{\mathbf{U}}^{*}}{T}\boldsymbol{\eta}_i^{*} = O_p(1) \qquad \text{as} \qquad N \to \infty,$$

which implies that the finite T distribution of dynamic CCEP depends on the unobserved loadings. However,

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{y}_{i,-1}'\bar{\mathbf{U}}^{*}}{T}\boldsymbol{\eta}_i^{*} \to \mathbf{0} \qquad \text{as} \qquad N \to \infty.$$

Turning to the second term in (A-14) we note that by a straightforward extension of the arguments in Pesaran (2006) we obtain for finite T that

$$\frac{\mathbf{y}_{i,-1}'\mathbf{Q}}{T} = \frac{\mathbf{y}_{i,-1}'\check{\mathbf{F}}\ddot{\mathbf{P}}}{T} + \frac{\mathbf{y}_{i,-1}'\ddot{\mathbf{U}}}{T} = \frac{\mathbf{y}_{i,-1}'\check{\mathbf{F}}\ddot{\mathbf{P}}}{T} + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{A-16}$$

$$\frac{\mathbf{Q}'\mathbf{Q}}{T} = \ddot{\mathbf{P}}'\frac{\check{\mathbf{F}}'\check{\mathbf{F}}}{T}\ddot{\mathbf{P}} + \ddot{\mathbf{P}}'\frac{\check{\mathbf{F}}'\ddot{\mathbf{U}}}{T} + \frac{\ddot{\mathbf{U}}'\check{\mathbf{F}}}{T}\ddot{\mathbf{P}} + \frac{\ddot{\mathbf{U}}'\ddot{\mathbf{U}}}{T} = \ddot{\mathbf{P}}'\frac{\check{\mathbf{F}}'\check{\mathbf{F}}}{T}\ddot{\mathbf{P}} + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{A-17}$$

$$\frac{\mathbf{Q}'\bar{\mathbf{U}}^{*}}{T} = \ddot{\mathbf{P}}'\frac{\check{\mathbf{F}}'\bar{\mathbf{U}}^{*}}{T} + \frac{\ddot{\mathbf{U}}'\bar{\mathbf{U}}^{*}}{T} = \ddot{\mathbf{P}}'\frac{\check{\mathbf{F}}'\bar{\mathbf{U}}^{*}}{T} + O_p\left(\frac{1}{N}\right). \tag{A-18}$$

Then, assuming for notational convenience that $m = 1 + k$,

$$\frac{\mathbf{y}_{i,-1}'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\bar{\mathbf{U}}^{*}}{T} = \left(\frac{\mathbf{y}_{i,-1}'\check{\mathbf{F}}\ddot{\mathbf{P}}}{T} + R\right)\left(\ddot{\mathbf{P}}'\frac{\check{\mathbf{F}}'\check{\mathbf{F}}}{T}\ddot{\mathbf{P}} + R\right)^{-1}\left(\ddot{\mathbf{P}}'\frac{\check{\mathbf{F}}'\bar{\mathbf{U}}^{*}}{T} + O_p\left(\frac{1}{N}\right)\right),$$

with $R = O_p\left(\frac{1}{\sqrt{N}}\right)$ and $\ddot{\mathbf{P}}'\frac{\check{\mathbf{F}}'\bar{\mathbf{U}}^{*}}{T} = O_p\left(\frac{1}{\sqrt{N}}\right)$. Hence,

$$\frac{\sqrt{N}\mathbf{Q}'\bar{\mathbf{U}}^{*}}{T} = \frac{\sqrt{N}\ddot{\mathbf{P}}'\check{\mathbf{F}}'\bar{\mathbf{U}}^{*}}{T} + O_p\left(\frac{1}{\sqrt{N}}\right) = c + O_p\left(\frac{1}{\sqrt{N}}\right),$$

with $c = \sqrt{N}\ddot{\mathbf{P}}'\check{\mathbf{F}}'\bar{\mathbf{U}}^{*}/T = O_p(1)$. Similarly setting $a = \mathbf{y}_{i,-1}'\check{\mathbf{F}}\ddot{\mathbf{P}}/T = O_p(1)$ and $b = \ddot{\mathbf{P}}'\check{\mathbf{F}}'\check{\mathbf{F}}\ddot{\mathbf{P}}/T = O_p(1)$ yields,

$$\frac{\mathbf{y}_{i,-1}'\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\sqrt{N}\mathbf{Q}'\bar{\mathbf{U}}^{*}}{T} = (a + R)(b + R)^{-1}(c + R) = ab^{-1}c + R,$$

$$= \frac{\mathbf{y}_{i,-1}^{'}\check{\mathbf{F}}\ddot{\mathbf{P}}}{T}\left(\ddot{\mathbf{P}}'\frac{\check{\mathbf{F}}'\check{\mathbf{F}}}{T}\ddot{\mathbf{P}}\right)^{-1}\frac{\sqrt{N}\ddot{\mathbf{P}}'\check{\mathbf{F}}'\bar{\mathbf{U}}^*}{T} + O_p\left(\frac{1}{\sqrt{N}}\right). \quad \text{(A-19)}$$

As a result, since

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbf{y}_{i,-1}^{'}\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\bar{\mathbf{U}}^*}{T} = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{y}_{i,-1}^{'}\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\sqrt{N}\mathbf{Q}'\bar{\mathbf{U}}^*}{T},$$

the following holds for finite T

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbf{y}_{i,-1}^{'}\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\bar{\mathbf{U}}^*}{T}\boldsymbol{\eta}_i^* = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{y}_{i,-1}^{'}\mathbf{Q}_{\mathrm{f}}}{T}\left(\frac{\mathbf{Q}_{\mathrm{f}}'\mathbf{Q}_{\mathrm{f}}}{T}\right)^{-1}\frac{\sqrt{N}\mathbf{Q}_{\mathrm{f}}'\bar{\mathbf{U}}^*}{T}\boldsymbol{\eta}_i^* + O_p\left(\frac{1}{\sqrt{N}}\right),$$

where the first term is $O_p(1)$ by similar arguments as above. However, without the stabilizing factor $\sqrt{N}$

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{y}_{i,-1}^{'}\mathbf{Q}}{T}\left(\frac{\mathbf{Q}'\mathbf{Q}}{T}\right)^{-1}\frac{\mathbf{Q}'\bar{\mathbf{U}}^*}{T}\boldsymbol{\eta}_i^* \to \mathbf{0} \qquad \text{as} \qquad N \to \infty.$$

The analysis is similar for $\boldsymbol{\Delta}_{\mathrm{F,x}}$ in (A-15) and the result of Lemma 1 follows. For $m < 1 + k$ a normalization similar to that of Karabiyik et al. (2016) can be employed but the final result will not change.

Theorems

Proof of Theorem 1a. The probability limit as $N \to \infty$ of the CCEP estimator in equations (15) and (16) is given by

$$\operatorname*{plim}_{N\to\infty}(\widehat{\rho} - \rho) = \left(\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\mathbf{y}_{-1}'\mathbb{M}_{\mathrm{XQ}}\mathbf{y}_{-1}\right)^{-1}\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\mathbf{y}_{-1}'\mathbb{M}_{\mathrm{XQ}}\boldsymbol{\varepsilon}, \quad \text{(A-20)}$$

$$\operatorname*{plim}_{N\to\infty}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = -\operatorname*{plim}_{N\to\infty}\left(\mathbf{X}'\mathbb{M}_{\mathrm{Q}}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbb{M}_{\mathrm{Q}}\mathbf{y}_{-1}\operatorname*{plim}_{N\to\infty}(\widehat{\rho} - \rho), \quad \text{(A-21)}$$

where we used the strict exogeneity of $\mathbf{X}$ such that $\operatorname{plim}_{N\to\infty}(1/NT)\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathrm{Q}}\boldsymbol{\varepsilon}_i = \mathbf{0}$ and Lemma 1, which implies that $(1/NT)\mathbf{y}_{-1}'\mathbb{M}_{\mathrm{Q}}(\mathbf{I}_N \otimes \mathbf{F})\boldsymbol{\Lambda} = (1/NT)\mathbf{y}_{-1}'\left(\mathbf{I}_N \otimes \mathbf{M}_{\mathrm{Q}}\mathbf{F}\right)\boldsymbol{\Lambda} = 0$ and $(1/NT)\mathbf{X}'\mathbb{M}_{\mathrm{Q}}(\mathbf{I}_N \otimes \mathbf{F})\boldsymbol{\Lambda} = \mathbf{0}$ as $N \to \infty$. The asymptotic bias for $\widehat{\boldsymbol{\beta}}$ in (A-21) is as stated in equation (21).

Using the strict exogeneity of $\mathbf{X}$, the numerator of equation (A-20) is given by

$$\begin{aligned}\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\mathbf{y}_{-1}'\mathbb{M}_{\mathrm{XQ}}\boldsymbol{\varepsilon} &= \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\mathbf{y}_{-1}'\mathbb{M}_{\mathrm{Q}}\boldsymbol{\varepsilon} - \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\mathbf{y}_{-1}'\mathbb{M}_{\mathrm{Q}}\mathbf{X}\left(\mathbf{X}'\mathbb{M}_{\mathrm{Q}}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbb{M}_{\mathrm{Q}}\boldsymbol{\varepsilon}, \\ &= \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\mathbf{y}_{-1}'\mathbb{M}_{\mathrm{Q}}\boldsymbol{\varepsilon}. \qquad \text{(A-22)}\end{aligned}$$

Under Assumptions (1)-(4), we have

$$\begin{aligned}\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\mathbf{y}_{-1}'\mathbb{M}_{\mathrm{Q}}\boldsymbol{\varepsilon} &= \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\left[\boldsymbol{\varepsilon}_i - \mathbf{Q}\left(\mathbf{Q}'\mathbf{Q}\right)^{-}\mathbf{Q}'\boldsymbol{\varepsilon}_i\right], \\ &= -\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{Q}\left(\mathbf{Q}'\mathbf{Q}\right)^{-}\mathbf{Q}'\boldsymbol{\varepsilon}_i,\end{aligned}$$

$$= -\frac{\sigma_\varepsilon^2}{T}\Upsilon(\rho, \mathbf{H}_{\mathrm{Q}}^{\mathrm{L}}), \tag{A-23}$$

with $\Upsilon(\rho, \mathbf{H}_{\mathrm{Q}}^{\mathrm{L}}) = \sum_{t=1}^{T-1} \rho^{t-1} \sum_{s=t+1}^{T} \mathbf{H}_{\mathrm{Q}}^{\mathrm{L}}(s, s-t)$. Note that for sake of notational convenience, the proofs here and below are derived under homoskedastic errors ($\sigma_{\varepsilon,i}^2 = \sigma_\varepsilon^2$) but can be straightforwardly extended to allow for cross-sectional heteroskedasticity.

Defining $\sigma_{\check{y}_{-1}}^2 = \mathrm{plim}_{N\to\infty}(NT)^{-1}\mathbf{y}_{-1}'\mathbb{M}_{XQ}\mathbf{y}_{-1}$ as the variance of the orthogonal projection of $\mathbf{y}_{-1}$ on $\mathbf{X}$ and $\mathbf{Q}$ and substituting it together with (A-23) in (A-20) yields

$$\operatorname*{plim}_{N\to\infty}(\widehat{\rho} - \rho) = -\frac{1}{T}\frac{\sigma_\varepsilon^2}{\sigma_{\check{y}_{-1}}^2}\Upsilon(\rho, \mathbf{H}_{\mathrm{Q}}^{\mathrm{L}}), \tag{A-24}$$

as stated in Theorem 1a. □

Proof of Theorem 1b. Using equation (6), we can write $\mathbf{y}_{-1}$ as

$$\begin{aligned}
\mathbf{y}_{-1} &= (\mathbf{I}_N \otimes \boldsymbol{\iota}_T)\boldsymbol{\alpha} + \rho\mathbf{y}_{-2} + \mathbf{X}_{-1}\boldsymbol{\beta} + (\mathbf{I}_N \otimes \mathbf{F}_{-1})\boldsymbol{\Lambda} + \boldsymbol{\varepsilon}_{-1}, \\
&= (\mathbf{I}_N \otimes \boldsymbol{\iota}_T)\rho(L)^{-1}\boldsymbol{\alpha} + \rho(L)^{-1}\mathbf{X}_{-1}\boldsymbol{\beta} + (\mathbf{I}_N \otimes \rho(L)^{-1}\mathbf{F}_{-1})\boldsymbol{\Lambda} + \rho(L)^{-1}\boldsymbol{\varepsilon}_{-1}, \\
&= (\mathbf{I}_N \otimes \boldsymbol{\iota}_T)\boldsymbol{\alpha}^{+} + \mathbf{X}_{-1}^{+}\boldsymbol{\beta} + \left(\mathbf{I}_N \otimes \mathbf{F}_{-1}^{+}\right)\boldsymbol{\Lambda} + \boldsymbol{\varepsilon}_{-1}^{+},
\end{aligned}$$

by Assumption 5 such that

$$\mathbb{M}_{\mathrm{Q}}\mathbf{y}_{-1} = \mathbb{M}_{\mathrm{Q}}\mathbf{X}_{-1}^{+}\boldsymbol{\beta} + \left(\mathbf{I}_N \otimes \mathbf{M}_{\mathrm{Q}}\mathbf{F}_{-1}^{+}\right)\boldsymbol{\Lambda} + \mathbb{M}_{\mathrm{Q}}\boldsymbol{\varepsilon}_{-1}^{+}. \tag{A-25}$$

Consequently, from the strict exogeneity of $\mathbf{X}$ and the independence of $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ (Assumption 2), we have

$$\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\mathbf{y}_{-1}'\mathbb{M}_{\mathrm{Q}}\boldsymbol{\varepsilon} = \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\boldsymbol{\varepsilon}_{-1}^{+\prime}\mathbb{M}_{\mathrm{Q}}\boldsymbol{\varepsilon} = \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_{i,-1}^{+\prime}\mathbf{M}_{\mathrm{Q}}\boldsymbol{\varepsilon}_i. \tag{A-26}$$

Using equation (A-22) and defining $\widetilde{\mathbf{Q}} = \left(\widetilde{\bar{\mathbf{y}}}, \widetilde{\bar{\mathbf{y}}}_{-1}, \widetilde{\bar{\mathbf{Z}}}, ..., \widetilde{\bar{\mathbf{Z}}}_{-p_T}\right)$, with the affix $\tilde{}$ denoting demeaned variables (deviations from time means), the numerator of equation (A-20) is given by

$$\begin{aligned}
\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_{i,-1}^{+\prime}\mathbf{M}_{\mathrm{Q}}\boldsymbol{\varepsilon}_i &= \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_{i,-1}^{+\prime}\left[(\boldsymbol{\varepsilon}_i - \bar{\boldsymbol{\varepsilon}}_i) - \widetilde{\mathbf{Q}}\left(\widetilde{\mathbf{Q}}'\widetilde{\mathbf{Q}}\right)^{-}\widetilde{\mathbf{Q}}'(\boldsymbol{\varepsilon}_i - \bar{\boldsymbol{\varepsilon}}_i)\right] \\
&= -\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_{i,-1}^{+\prime}\bar{\boldsymbol{\varepsilon}}_i - \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_{i,-1}^{+\prime}\widetilde{\mathbf{Q}}\left(\widetilde{\mathbf{Q}}'\widetilde{\mathbf{Q}}\right)^{-}\widetilde{\mathbf{Q}}'\boldsymbol{\varepsilon}_i \\
&= -\frac{\sigma_\varepsilon^2}{T}A(\rho, T) - \frac{\sigma_\varepsilon^2}{T}\sum_{t=1}^{T-1}\rho^{t-1}\sum_{s=t+1}^{T}\mathbf{H}_{\tilde{\mathrm{Q}}}^{\mathrm{L}}(s, s-t) \\
&= -\frac{\sigma_\varepsilon^2}{T}A(\rho, T) - \frac{\sigma_\varepsilon^2}{T}D(\rho, \mathbf{H}_{\tilde{\mathrm{Q}}}^{\mathrm{L}})
\end{aligned} \tag{A-27}$$

with $A(\rho, T) = \frac{1}{1-\rho}\left(1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right)$, $D(\rho, \widetilde{\mathbf{Q}}_{\mathrm{f}}) = \sum_{t=1}^{T-1}\rho^{t-1}\sum_{s=t+1}^{T}\mathbf{H}_{\tilde{\mathrm{Q}}}^{\mathrm{L}}(s, s-t)$ and $\mathbf{H}_{\tilde{\mathrm{Q}}}^{\mathrm{L}}$ the limit of the projection matrix $\mathbf{H}_{\tilde{\mathrm{Q}}}$, i.e. $\mathbf{H}_{\tilde{\mathrm{Q}}}^{\mathrm{L}} = \mathrm{plim}_{N\to\infty}\widetilde{\mathbf{Q}}\left(\widetilde{\mathbf{Q}}'\widetilde{\mathbf{Q}}\right)^{-}\widetilde{\mathbf{Q}}'$.

Turning to the denominator of equation (A-20), using (A-25) yields

$$\mathbf{y}_{-1}'\mathbb{M}_{XQ}\mathbf{y}_{-1} = \left(\mathbb{M}_{\mathrm{Q}}\mathbf{X}_{-1}^{+}\boldsymbol{\beta} + \left(\mathbf{I}_N \otimes \mathbf{M}_{\mathrm{Q}}\mathbf{F}_{-1}^{+}\right)\boldsymbol{\Lambda} + \mathbb{M}_{\mathrm{Q}}\boldsymbol{\varepsilon}_{-1}^{+}\right)'\mathbb{M}_{\mathrm{x}}\left(\mathbb{M}_{\mathrm{Q}}\mathbf{X}_{-1}^{+}\boldsymbol{\beta} + \left(\mathbf{I}_N \otimes \mathbf{M}_{\mathrm{Q}}\mathbf{F}_{-1}^{+}\right)\boldsymbol{\Lambda} + \mathbb{M}_{\mathrm{Q}}\boldsymbol{\varepsilon}_{-1}^{+}\right)$$

$$= \boldsymbol{\beta}'\mathbf{X}_{-1}^{+'}\mathbb{M}_{XQ}\mathbf{X}_{-1}^{+}\boldsymbol{\beta} + \boldsymbol{\Lambda}'\left(\mathbf{I}_N \otimes \mathbf{F}_{-1}^{+'}\right)\mathbb{M}_{XQ}\left(\mathbf{I}_N \otimes \mathbf{F}_{-1}^{+}\right)\boldsymbol{\Lambda} + \boldsymbol{\varepsilon}_{-1}^{+'}\mathbb{M}_{XQ}\boldsymbol{\varepsilon}_{-1}^{+}$$
$$+2\boldsymbol{\beta}'\mathbf{X}_{-1}^{+'}\mathbb{M}_{XQ}\boldsymbol{\varepsilon}_{-1}^{+} + 2\boldsymbol{\beta}'\mathbf{X}_{-1}^{+'}\mathbb{M}_{XQ}\left(\mathbf{I}_N \otimes \mathbf{F}_{-1}^{+}\right)\boldsymbol{\Lambda} + 2\boldsymbol{\Lambda}'\left(\mathbf{I}_N \otimes \mathbf{F}_{-1}^{+'}\right)\mathbb{M}_{XQ}\boldsymbol{\varepsilon}_{-1}^{+}.$$

Defining

$$C^{+} = \boldsymbol{\beta}'\mathbf{X}_{-1}^{+'}\mathbb{M}_{XQ}\mathbf{X}_{-1}^{+}\boldsymbol{\beta} + \boldsymbol{\Lambda}'\left(\mathbf{I}_N \otimes \mathbf{F}_{-1}^{+'}\right)\mathbb{M}_{XQ}\left(\mathbf{I}_N \otimes \mathbf{F}_{-1}^{+}\right)\boldsymbol{\Lambda} + 2\boldsymbol{\beta}'\mathbf{X}_{-1}^{+'}\mathbb{M}_{XQ}\boldsymbol{\varepsilon}_{-1}^{+}$$
$$+2\boldsymbol{\beta}'\mathbf{X}_{-1}^{+'}\mathbb{M}_{XQ}\left(\mathbf{I}_N \otimes \mathbf{F}_{-1}^{+}\right)\boldsymbol{\Lambda} + 2\boldsymbol{\Lambda}'\left(\mathbf{I}_N \otimes \mathbf{F}_{-1}^{+'}\right)\mathbb{M}_{XQ}\boldsymbol{\varepsilon}_{-1}^{+},$$

and taking its probability limit

$$\operatorname*{plim}_{N\to\infty}\frac{1}{NT}C^{+} = \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\boldsymbol{\beta}'\mathbf{X}_{-1}^{+'}\mathbb{M}_{XQ}\mathbf{X}_{-1}^{+}\boldsymbol{\beta} + \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\boldsymbol{\Lambda}'\left(\mathbf{I}_N \otimes \mathbf{F}_{-1}^{+'}\right)\mathbb{M}_{XQ}\left(\mathbf{I}_N \otimes \mathbf{F}_{-1}^{+}\right)\boldsymbol{\Lambda}$$
$$+\operatorname*{plim}_{N\to\infty}\frac{1}{NT}2\boldsymbol{\beta}'\mathbf{X}_{-1}^{+'}\mathbb{M}_{XQ}\left(\mathbf{I}_N \otimes \mathbf{F}_{-1}^{+}\right)\boldsymbol{\Lambda}, \quad \text{(A-28)}$$

where we made use of

$$\operatorname*{plim}_{N\to\infty}\frac{1}{NT}2\boldsymbol{\Lambda}'\left(\mathbf{I}_N \otimes \mathbf{F}_{-1}^{+'}\right)\mathbb{M}_{XQ}\boldsymbol{\varepsilon}_{-1}^{+} = 0,$$
$$\operatorname*{plim}_{N\to\infty}\frac{1}{NT}2\boldsymbol{\beta}'\mathbf{X}_{-1}^{+'}\mathbb{M}_{XQ}\boldsymbol{\varepsilon}_{-1}^{+} = 0,$$

from Assumptions 2 and 3 and the exogeneity of $\mathbf{X}$. Hence

$$\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\mathbf{y}_{-1}'\mathbb{M}_{XQ}\mathbf{y}_{-1} = \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\boldsymbol{\varepsilon}_{-1}^{+'}\mathbb{M}_{XQ}\boldsymbol{\varepsilon}_{-1}^{+} + \operatorname*{plim}_{N\to\infty}\frac{1}{NT}C^{+},$$
$$= \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\boldsymbol{\varepsilon}_{-1}^{+'}\mathbb{M}_{Q}\boldsymbol{\varepsilon}_{-1}^{+} - \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\boldsymbol{\varepsilon}_{-1}^{+'}\mathbb{M}_{Q}\mathbf{X}\left(\mathbf{X}'\mathbb{M}_{Q}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbb{M}_{Q}\boldsymbol{\varepsilon}_{-1}^{+}$$
$$+\operatorname*{plim}_{N\to\infty}\frac{1}{NT}C^{+},$$
$$= \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_{i,-1}^{+'}\mathbf{M}_{Q}\boldsymbol{\varepsilon}_{i,-1}^{+} + \operatorname*{plim}_{N\to\infty}\frac{1}{NT}C^{+}. \quad \text{(A-29)}$$

Focusing on the first term and using earlier definitions

$$\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_{i,-1}^{+'}\mathbf{M}_{Q}\boldsymbol{\varepsilon}_{i,-1}^{+} = \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_{i,-1}^{+'}\left[\left(\boldsymbol{\varepsilon}_{i,-1}^{+} - \bar{\boldsymbol{\varepsilon}}_{i,-1}^{+}\right) - \widetilde{\mathbf{Q}}\left(\widetilde{\mathbf{Q}}'\widetilde{\mathbf{Q}}\right)^{-}\widetilde{\mathbf{Q}}'\left(\boldsymbol{\varepsilon}_{i,-1}^{+} - \bar{\boldsymbol{\varepsilon}}_{i,-1}^{+}\right)\right],$$
$$= \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\varepsilon_{i,t-1}^{+} - \bar{\varepsilon}_{i,-1}^{+}\right)^2 - \operatorname*{plim}_{N\to\infty}\frac{1}{NT}\boldsymbol{\varepsilon}_{i,-1}^{+'}\mathbf{H}_{\widetilde{Q}}\boldsymbol{\varepsilon}_{i,-1}^{+},$$
$$= \frac{\sigma_\varepsilon^2}{T}B(\rho,T) - \frac{\sigma_\varepsilon^2}{1-\rho^2}\frac{1}{T}\left[2 + k(1+p_T) + 2\rho\sum_{t=1}^{T-1}\rho^{t-1}\sum_{s=t+1}^{T}\mathbf{H}_{\widetilde{Q}}^{L}(s,s-t)\right],$$
$$= \frac{\sigma_\varepsilon^2}{T}\left(B(\rho,T) - \frac{1}{1-\rho^2}\left[\mathrm{tr}(\mathbf{H}_{\widetilde{Q}}^{L}) + 2\rho D(\rho,\mathbf{H}_{\widetilde{Q}}^{L})\right]\right), \quad \text{(A-30)}$$

where $B(\rho,T) = \frac{T}{1-\rho^2}\left(1 - \frac{1}{T}\frac{1+\rho}{1-\rho} - \frac{2\rho}{T^2}\frac{1-\rho^T}{(1-\rho)^2}\right)$.

Combining (A-27), (A-28) and (A-30)

$$\operatorname*{plim}_{N\to\infty}(\widehat{\rho} - \rho) = \frac{-\frac{\sigma_\varepsilon^2}{T}\left(A(\rho,T) + D(\rho,\mathbf{H}_{\widetilde{Q}}^{L})\right)}{\frac{\sigma_\varepsilon^2}{T}\left(B(\rho,T) - \frac{1}{1-\rho^2}\left[\mathrm{tr}(\mathbf{H}_{\widetilde{Q}}^{L}) + 2\rho D(\rho,\mathbf{H}_{\widetilde{Q}}^{L})\right]\right) + \operatorname{plim}_{N\to\infty}\frac{1}{NT}C^{+}},$$

$$= \frac{-A(\rho,T) - D(\rho,\mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}})}{B(\rho,T) - \frac{1}{1-\rho^2}\left[\mathrm{tr}(\mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}}) + 2\rho D(\rho,\mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}})\right] + \mathrm{plim}_{N\to\infty}\frac{1}{N\sigma_\varepsilon^2}C^+}, \quad \text{(A-31)}$$

which we reformulate in Theorem 1.b to

$$\underset{N\to\infty}{\mathrm{plim}}\,(\widehat{\rho}-\rho) = \frac{-\left[A(\rho,T) + D(\rho,\mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}})\right]}{\left[B(\rho,T) - E(\rho,\mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}}) + TC\right]} \quad \text{(A-32)}$$

where $E(\rho,\widetilde{\mathbf{Q}}_{\mathrm{f}}) = \frac{1}{1-\rho^2}\left[\mathrm{tr}(\mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}}) + 2\rho D(\rho,\mathbf{H}^{\mathrm{L}}_{\widetilde{\mathrm{Q}}})\right]$ and

$$\begin{aligned}
C &= \underset{N\to\infty}{\mathrm{plim}}\frac{1}{NT\sigma_\varepsilon^2}\left(\boldsymbol{\beta}'\mathbf{X}^{+'}_{-1}\mathbb{M}_{\mathrm{XQ}}\mathbf{X}^{+}_{-1}\boldsymbol{\beta} + \boldsymbol{\Lambda}'\left(\mathbf{I}_N\otimes\mathbf{F}^{+'}_{-1}\right)\mathbb{M}_{\mathrm{XQ}}\left(\mathbf{I}_N\otimes\mathbf{F}^{+}_{-1}\right)\boldsymbol{\Lambda} + 2\boldsymbol{\beta}'\mathbf{X}^{+'}_{-1}\mathbb{M}_{\mathrm{XQ}}\left(\mathbf{I}_N\otimes\mathbf{F}^{+}_{-1}\right)\boldsymbol{\Lambda}\right)\\
&= \underset{N\to\infty}{\mathrm{plim}}\frac{1}{NT\sigma_\varepsilon^2}\left(\boldsymbol{\beta}'\mathbf{X}^{+'}_{-1}\mathbb{M}_{\mathrm{XQ}}\mathbf{X}^{+}_{-1}\boldsymbol{\beta} + \boldsymbol{\Lambda}'\mathbb{F}^{+'}_{-1}\mathbb{M}_{\mathrm{XQ}}\mathbb{F}^{+}_{-1}\boldsymbol{\Lambda} + 2\boldsymbol{\beta}'\mathbf{X}^{+'}_{-1}\mathbb{M}_{\mathrm{XQ}}\mathbb{F}^{+}_{-1}\boldsymbol{\Lambda}\right)\\
&= \underset{N\to\infty}{\mathrm{plim}}\frac{\boldsymbol{\beta}'\boldsymbol{\Omega}_{\check{\mathrm{x}}}\boldsymbol{\beta} + \boldsymbol{\Lambda}'\boldsymbol{\Omega}_{\check{\mathrm{f}}}\boldsymbol{\Lambda} + 2\boldsymbol{\beta}'\boldsymbol{\Omega}_{\check{\mathrm{x}},\check{\mathrm{f}}}\boldsymbol{\Lambda}}{\sigma_\varepsilon^2}
\end{aligned}$$

with $\boldsymbol{\Omega}_{\check{\mathrm{x}}} = (NT)^{-1}\mathbf{X}^{+'}_{-1}\mathbb{M}_{\mathrm{XQ}}\mathbf{X}^{+}_{-1}$, $\boldsymbol{\Omega}_{\check{\mathrm{f}}} = (NT)^{-1}\mathbb{F}^{+'}_{-1}\mathbb{M}_{\mathrm{XQ}}\mathbb{F}^{+}_{-1}$, $\boldsymbol{\Omega}_{\check{\mathrm{x}},\check{\mathrm{f}}} = (NT)^{-1}\mathbf{X}^{+'}_{-1}\mathbb{M}_{\mathrm{XQ}}\mathbb{F}^{+}_{-1}$ and $\mathbb{F}^{+}_{-1} = \left(\mathbf{I}_N\otimes\mathbf{F}^{+}_{-1}\right)$. □

Appendix B Additional simulation tables

Table B-1: Monte Carlo results for ρ and β : baseline design with $\rho = 0.4$

			ρ					β				
N	T		bias	SE	$\widehat{\text{SE}}$	rmse	size	bias	SE	$\widehat{\text{SE}}$	rmse	size
25	10	CCEP	-0.200	0.097	0.087	0.222	0.61	-0.033	0.079	0.082	0.086	0.07
		CCEPbc	0.005	0.094	0.100	0.094	0.03	0.001	0.080	0.084	0.080	0.04
		CCEPjk	0.074	0.256	0.253	0.267	0.05	0.088	0.163	0.193	0.185	0.04
		CCEPrm	-0.155	0.095	0.085	0.182	0.45	-0.007	0.072	0.076	0.072	0.04
		FLSbc	-0.084	0.099	0.107	0.130	0.10	0.003	0.084	0.094	0.085	0.04
25	20	CCEP	-0.092	0.046	0.043	0.102	0.55	-0.008	0.047	0.048	0.048	0.06
		CCEPbc	0.001	0.045	0.045	0.045	0.06	-0.000	0.047	0.048	0.047	0.06
		CCEPjk	0.015	0.067	0.064	0.069	0.07	0.016	0.057	0.061	0.060	0.05
		CCEPrm	-0.066	0.053	0.049	0.085	0.30	0.003	0.049	0.049	0.049	0.06
		FLSbc	-0.018	0.049	0.060	0.052	0.03	0.011	0.055	0.064	0.057	0.04
25	30	CCEP	-0.058	0.033	0.033	0.067	0.43	-0.004	0.036	0.037	0.036	0.06
		CCEPbc	0.001	0.033	0.033	0.033	0.07	-0.001	0.036	0.037	0.036	0.06
		CCEPjk	0.009	0.041	0.041	0.042	0.06	0.007	0.041	0.042	0.041	0.06
		CCEPrm	-0.030	0.042	0.039	0.051	0.14	0.002	0.039	0.039	0.039	0.06
		FLSbc	-0.007	0.036	0.046	0.037	0.02	0.007	0.043	0.054	0.043	0.03
25	50	CCEP	-0.035	0.024	0.024	0.042	0.32	-0.001	0.028	0.028	0.028	0.06
		CCEPbc	0.000	0.024	0.024	0.024	0.07	-0.000	0.028	0.028	0.028	0.06
		CCEPjk	0.003	0.028	0.027	0.028	0.07	0.002	0.030	0.030	0.030	0.05
		CCEPrm	0.003	0.034	0.031	0.035	0.09	-0.001	0.033	0.032	0.033	0.07
		FLSbc	-0.003	0.026	0.035	0.026	0.02	0.004	0.032	0.045	0.032	0.01
100	10	CCEP	-0.203	0.074	0.051	0.216	0.92	-0.035	0.042	0.041	0.055	0.15
		CCEPbc	0.002	0.043	0.047	0.043	0.03	0.000	0.039	0.041	0.039	0.04
		CCEPjk	0.082	0.218	0.176	0.233	0.14	0.085	0.103	0.103	0.134	0.12
		CCEPrm	-0.140	0.079	0.054	0.160	0.65	-0.005	0.038	0.038	0.039	0.06
		FLSbc	-0.103	0.048	0.046	0.114	0.60	-0.017	0.041	0.041	0.044	0.08
100	20	CCEP	-0.094	0.027	0.023	0.098	0.97	-0.008	0.023	0.024	0.025	0.06
		CCEPbc	-0.000	0.022	0.022	0.022	0.05	0.000	0.023	0.024	0.023	0.05
		CCEPjk	0.016	0.047	0.036	0.049	0.16	0.019	0.030	0.030	0.035	0.10
		CCEPrm	-0.058	0.038	0.029	0.069	0.50	0.004	0.024	0.024	0.024	0.06
		FLSbc	-0.026	0.023	0.023	0.034	0.20	-0.001	0.024	0.026	0.024	0.04
100	30	CCEP	-0.061	0.018	0.017	0.064	0.94	-0.004	0.018	0.018	0.018	0.06
		CCEPbc	-0.000	0.016	0.017	0.016	0.06	-0.000	0.018	0.018	0.018	0.06
		CCEPjk	0.008	0.024	0.021	0.025	0.09	0.008	0.020	0.021	0.021	0.06
		CCEPrm	-0.023	0.028	0.023	0.036	0.22	0.003	0.020	0.020	0.020	0.05
		FLSbc	-0.012	0.016	0.017	0.020	0.10	-0.000	0.018	0.020	0.018	0.04
500	10	CCEP	-0.203	0.066	0.036	0.213	0.99	-0.034	0.024	0.020	0.042	0.40
		CCEPbc	0.001	0.019	0.021	0.020	0.04	0.000	0.017	0.018	0.017	0.04
		CCEPjk	0.088	0.210	0.150	0.228	0.21	0.087	0.087	0.070	0.123	0.28
		CCEPrm	-0.118	0.075	0.035	0.140	0.74	-0.003	0.020	0.018	0.020	0.08
		FLSbc	-0.106	0.034	0.019	0.112	0.96	-0.022	0.019	0.016	0.029	0.32
500	20	CCEP	-0.095	0.020	0.013	0.097	1.00	-0.009	0.011	0.011	0.014	0.13
		CCEPbc	0.000	0.010	0.010	0.010	0.04	-0.000	0.010	0.010	0.010	0.05
		CCEPjk	0.015	0.040	0.023	0.043	0.31	0.018	0.017	0.014	0.025	0.27
		CCEPrm	-0.041	0.035	0.018	0.053	0.53	0.003	0.011	0.011	0.012	0.06
		FLSbc	-0.027	0.012	0.010	0.029	0.73	-0.003	0.010	0.010	0.011	0.06

Note: See Table 1 but with $\rho = 0.4$ and $\beta = 0.6$.

Table B-2: Monte Carlo results for ρ : number and strength of factors and restricted bias corrections ($N = 25$)

	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	−0.392	0.418	0.92	−0.178	0.188	0.89	−0.111	0.118	0.84	−0.399	0.424	0.92	−0.181	0.191	0.90	−0.113	0.120	0.84
CCEPbc	0.007	0.150	0.06	0.006	0.063	0.08	0.001	0.037	0.05	0.009	0.153	0.07	0.005	0.064	0.08	0.000	0.038	0.05
CCEPbcr2	−0.032	0.108	0.06	−0.012	0.046	0.06	−0.006	0.032	0.07	−0.032	0.110	0.06	−0.012	0.046	0.07	−0.007	0.033	0.07
CCEPjk	0.040	0.359	0.08	0.037	0.124	0.16	0.029	0.075	0.16	0.036	0.350	0.09	0.039	0.124	0.19	0.030	0.076	0.18
CCEP(+g)	−0.395	0.424	0.89	−0.180	0.190	0.89	−0.112	0.119	0.83	−0.402	0.431	0.90	−0.182	0.193	0.89	−0.114	0.121	0.84
CCEPbc(+g)	0.008	0.163	0.04	0.006	0.066	0.08	0.001	0.039	0.06	0.009	0.166	0.04	0.006	0.066	0.08	0.000	0.039	0.06
CCEPbcr2(+g)	−0.030	0.116	0.04	−0.012	0.047	0.06	−0.007	0.033	0.07	−0.030	0.119	0.04	−0.013	0.047	0.06	−0.007	0.033	0.07
CCEPjk(+g)	-	-	-	0.042	0.132	0.13	0.029	0.077	0.14	-	-	-	0.044	0.133	0.15	0.030	0.078	0.16
FLSbc	−0.246	0.276	0.37	−0.062	0.089	0.04	−0.025	0.054	0.04	−0.223	0.260	0.32	−0.053	0.084	0.03	−0.023	0.053	0.05
	two factors																	
	$RI = 1$									$RI = 3$								
	$T = 10$			$T = 20$			$T = 30$			$T = 10$			$T = 20$			$T = 30$		
CCEP	−0.393	0.418	0.92	−0.178	0.188	0.89	−0.112	0.119	0.83	−0.385	0.412	0.90	−0.162	0.175	0.82	−0.096	0.105	0.71
CCEPbc	0.009	0.155	0.06	0.006	0.063	0.08	0.000	0.038	0.05	0.026	0.157	0.08	0.027	0.074	0.14	0.019	0.046	0.09
CCEPbcr2	−0.031	0.107	0.05	−0.012	0.045	0.06	−0.007	0.033	0.07	−0.006	0.109	0.06	0.010	0.050	0.10	0.012	0.037	0.13
CCEPjk	0.038	0.356	0.08	0.037	0.123	0.17	0.027	0.074	0.15	0.069	0.358	0.10	0.057	0.139	0.23	0.047	0.086	0.23
CCEP(+g)	−0.431	0.460	0.93	−0.198	0.208	0.91	−0.125	0.132	0.87	−0.410	0.440	0.90	−0.180	0.191	0.88	−0.110	0.117	0.80
CCEPbc(+g)	0.003	0.165	0.04	0.007	0.068	0.08	0.001	0.041	0.06	0.020	0.171	0.05	0.015	0.068	0.09	0.008	0.040	0.05
CCEPbcr2(+g)	−0.031	0.120	0.04	−0.012	0.047	0.05	−0.007	0.034	0.07	−0.001	0.119	0.05	0.007	0.047	0.08	0.008	0.034	0.09
CCEPjk(+g)	-	-	-	0.034	0.135	0.13	0.028	0.080	0.15	-	-	-	0.049	0.141	0.16	0.039	0.083	0.17
FLSbc	−0.515	0.530	0.85	−0.156	0.183	0.22	−0.061	0.088	0.07	−0.490	0.510	0.81	−0.134	0.164	0.14	−0.051	0.079	0.05

Note: see Table 2 but with $N = 25$.

Table B-3: Monte Carlo results for ρ : pure AR(1) model with strong factors ($N = 25$)

	bias	*rmse*	*size*	*bias*	*rmse*	*size*	*bias*	*rmse*	*size*	*bias*	*rmse*	*size*
	one factor											
	$T = 10$			$T = 20$			$T = 30$			$T = 50$		
CCEP	−0.429	0.452	0.96	−0.199	0.209	0.97	−0.125	0.132	0.93	−0.070	0.075	0.86
CCEPbc	−0.008	0.143	0.11	0.005	0.065	0.10	−0.000	0.040	0.07	−0.001	0.024	0.07
CCEPbcr1	−0.003	0.118	0.08	−0.001	0.050	0.06	−0.002	0.035	0.07	−0.001	0.023	0.08
CCEP(+g)	−0.434	0.460	0.93	−0.200	0.210	0.95	−0.126	0.133	0.93	−0.070	0.075	0.85
CCEPbc(+g)	−0.011	0.154	0.07	0.005	0.068	0.09	−0.000	0.041	0.06	−0.001	0.024	0.06
CCEPbcr1(+g)	−0.001	0.125	0.06	−0.000	0.051	0.06	−0.001	0.035	0.07	−0.001	0.023	0.07
CCEP(+xg)	−0.441	0.472	0.87	−0.199	0.210	0.94	−0.126	0.133	0.91	−0.070	0.075	0.84
CCEPbc(+xg)	−0.016	0.168	0.05	0.005	0.070	0.08	0.000	0.043	0.07	−0.001	0.024	0.06
CCEPbcr1(+xg)	−0.002	0.138	0.05	−0.000	0.053	0.05	−0.001	0.036	0.07	−0.001	0.023	0.07
	two factors											
	$T = 10$			$T = 20$			$T = 30$			$T = 50$		
CCEP	−0.399	0.425	0.93	−0.164	0.179	0.85	−0.092	0.104	0.72	−0.038	0.049	0.47
CCEPbc	0.021	0.148	0.14	0.042	0.087	0.24	0.036	0.062	0.20	0.031	0.044	0.31
CCEPbcr1	0.044	0.143	0.13	0.038	0.075	0.19	0.033	0.055	0.25	0.031	0.042	0.34
CCEP(+g)	−0.406	0.436	0.89	−0.170	0.184	0.86	−0.098	0.108	0.75	−0.044	0.053	0.52
CCEPbc(+g)	0.016	0.162	0.09	0.035	0.084	0.19	0.028	0.055	0.14	0.024	0.037	0.23
CCEPbcr1(+g)	0.047	0.150	0.10	0.038	0.075	0.17	0.030	0.052	0.23	0.027	0.039	0.29
CCEP(+xg)	−0.432	0.463	0.86	−0.186	0.198	0.90	−0.112	0.121	0.83	−0.058	0.064	0.70
CCEPbc(+xg)	0.008	0.170	0.05	0.020	0.075	0.11	0.012	0.044	0.06	0.009	0.025	0.07
CCEPbcr1(+xg)	0.057	0.161	0.08	0.036	0.071	0.14	0.025	0.046	0.16	0.017	0.029	0.16

Note: see Table 4 but with $N = 25$.

Table B-4: Monte Carlo results for ρ : dynamics in $\mathbf{z}_{it}$ with strong factors ($N = 25$)

	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)	−0.525	0.554	0.89	−0.236	0.248	0.95	−0.142	0.149	0.94	−0.077	0.081	0.91
CCEP_p_1(+g)	−0.650	0.693	0.80	−0.265	0.278	0.93	−0.151	0.159	0.93	−0.079	0.083	0.90
CCEP_p_T(+g)	-	-	-	−0.326	0.344	0.89	−0.200	0.212	0.89	−0.091	0.096	0.88
CCEPbc_p_0(+g)	0.010	0.197	0.04	0.004	0.075	0.06	−0.002	0.040	0.04	−0.003	0.023	0.05
CCEPbc_p_1(+g)	−0.003	0.254	0.02	0.009	0.088	0.06	0.001	0.044	0.04	−0.001	0.023	0.04
CCEPbc_p_T(+g)	-	-	-	0.011	0.109	0.04	0.003	0.062	0.05	−0.001	0.026	0.04
CCEPjk_p_1(+g)	-	-	-	0.129	0.242	0.14	0.089	0.136	0.20	0.037	0.058	0.17
FLSbc	−0.240	0.269	0.46	−0.052	0.079	0.05	−0.021	0.047	0.04	−0.009	0.028	0.03
	two factors											
	$T = 10$			$T = 20$			$T = 30$			$T = 50$		
CCEP_p_0(+g)	−0.563	0.591	0.93	−0.258	0.269	0.97	−0.158	0.165	0.97	−0.086	0.090	0.95
CCEP_p_1(+g)	−0.712	0.751	0.86	−0.296	0.310	0.96	−0.170	0.178	0.96	−0.087	0.091	0.93
CCEP_p_T(+g)	-	-	-	−0.365	0.383	0.94	−0.227	0.239	0.94	−0.101	0.106	0.94
CCEPbc_p_0(+g)	0.000	0.201	0.04	−0.006	0.074	0.07	−0.009	0.043	0.06	−0.008	0.024	0.06
CCEPbc_p_1(+g)	0.001	0.268	0.03	0.010	0.092	0.08	0.002	0.047	0.04	−0.001	0.023	0.04
CCEPbc_p_T(+g)	-	-	-	0.015	0.113	0.05	0.006	0.069	0.07	−0.001	0.026	0.04
CCEPjk_p_1(+g)	-	-	-	0.121	0.239	0.14	0.097	0.148	0.24	0.044	0.064	0.22
FLSbc	−0.510	0.526	0.89	−0.136	0.163	0.19	−0.047	0.072	0.06	−0.011	0.033	0.03

Note: see Table 5 but with $N = 25$.