Uniform Theory for CCE under Heterogeneous Slopes and General Unknown Factors

Stauskas, Ovidijus

Working Paper 2021:9

Department of Economics
School of Economics and Management

Uniform Theory for CCE Under Heterogeneous Slopes and General Unknown Factors

Ovidijus Stauskas

May 2021

Uniform Theory for CCE Under Heterogeneous Slopes and General Unknown Factors

Ovidijus Stauskas*

May 18, 2021

Abstract

A recent study proposed by Westerlund (CCE in Panels with General Unknown Factors, *Econometrics Journal*, **21**, 264-276, 2018) showed that a very popular Common Correlated Effects (CCE) estimator is significantly more applicable than it was thought before. Contrary to the usual stationarity assumption, common factors can in fact be much more general and not only unit root. This also helps to alleviate the uncertainty over deterministic model components since they can be treated as unknown, similarly to unobserved stochastic factors. While very promising, these theoretical results concern only the pooled (CCEP) version of the estimator for the homogeneous parameters, which does no take heterogeneous effects into account. Therefore, it is natural to generalize these findings to the case of unit-specific slopes. It is especially interesting, because many previous studies on heterogeneous slopes did not rigorously account for the usual situation when the factors are proxied by more explanatory variables than needed. As a result, the current setup introduces more uniformity to the CCE theory. We demonstrate that save for some regularity conditions, CCEP and the mean group (CCEMG) estimators are asymptotically normal and unbiased under heterogeneous slopes and general unknown factors.

1 Introduction

Consider the following panel data model often used in previous research (see e.g. Pesaran, 2006, or Karabiyik et al., 2017):

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{F}\gamma_i + \boldsymbol{\varepsilon}_i \tag{1.1}$$

*Corresponding author: Department of Economics, Lund University. Building: EC1. Room: 275. Email address. ovidijus.stuskas@nek.lu.se. I am grateful to Ignace De Vos for comments and discussions that helped to significantly improve this paper.

$$\mathbf{X}_i = \mathbf{F}\boldsymbol{\Gamma}_i + \mathbf{V}_i, \tag{1.2}$$

where $\mathbf{y}_i = [y_{i,1}, \ldots, y_{i,T}]' \in \mathbb{R}^{T\times 1}$ for $i = 1, \ldots, N$, $\mathbf{X}_i = [\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,T}]' \in \mathbb{R}^{T\times k}$ is the matrix of explanatory variables, $\mathbf{V}_i = [\mathbf{v}_{i,1}, \ldots, \mathbf{v}_{i,T}]' \in \mathbb{R}^{T\times k}$ and $\boldsymbol{\varepsilon}_i = [\varepsilon_{i,1}, \ldots, \varepsilon_{i,T}]' \in \mathbb{R}^{T\times}$ are matrix and vector of idiosyncratic errors and $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_T]'$ is the matrix of unobservable common factors. Also, $\boldsymbol{\Gamma}_i \in \mathbb{R}^{m\times k}$ and $\gamma_i \in \mathbb{R}^{m\times 1}$ are individual-specific factor loadings and $\boldsymbol{\beta}_i \in \mathbb{R}^{k+1}$ is the parameter vector of interest.

This is a setup of the so-called 'interactive effects' model, which helps to flexibly take unobserved heterogeneity in $\mathbf{y}_i$ into account. While the standard 'fixed effects' models subsume heterogeneity into an additive unit– and time–specific constants, the current setup is much more general. This is because time– and unit–specific effects enter in a multiplicative way, and this allows for many time–specific (observed and unobserved) factors to which individuals respond differently through the loadings.[1] Moreover, because the factors can be stochastic, this gives a way to model cross-section dependence.

Clearly, examination of (1.1)–(1.2) reveals that estimation of $\boldsymbol{\beta}_i$ is problematic due to unobserved $\mathbf{F}$. Using the structure of (1.1)–(1.2), we obtain a system

$$\mathbf{Z}_i = \mathbf{F}\mathbf{C}_i + \mathbf{U}_i, \tag{1.3}$$

where $\mathbf{Z}_i = [\mathbf{y}_i, \mathbf{X}_i] \in \mathbb{R}^{T\times(k+1)}$, $\mathbf{C}_i = \widetilde{\mathbf{C}}_i\mathbf{B}_i = [\gamma_i, \boldsymbol{\Gamma}_i]\mathbf{B}_i = [\boldsymbol{\Gamma}_i\boldsymbol{\beta}_i + \gamma_i, \boldsymbol{\Gamma}_i] \in \mathbb{R}^{m\times(k+1)}$ and $\mathbf{U}_i = \widetilde{\mathbf{U}}_i\mathbf{B}_i = [\boldsymbol{\varepsilon}_i, \mathbf{V}_i]\mathbf{B}_i = [\mathbf{V}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \mathbf{V}_i] \in \mathbb{R}^{T\times(k+1)}$, where

$$\mathbf{B}_i = \begin{bmatrix} 1 & \mathbf{0}_{1\times k} \\ \boldsymbol{\beta}_i & \mathbf{I}_k \end{bmatrix}.$$

This is known as static factor model, where following common correlated effects (CCE) procedure by Pesaran (2006), the estimator of $\mathbf{F}$ is given by

$$\widehat{\mathbf{F}} = \overline{\mathbf{Z}} = \mathbf{F}\overline{\mathbf{C}} + \overline{\mathbf{U}}, \tag{1.4}$$

where $\overline{\mathbf{A}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{A}_i$ is the cross-section average of arbitrary matrices $\mathbf{A}_i$ for $i = 1, \ldots, N$. It is known that as $N \to \infty$, $\overline{\mathbf{U}} \to_p \mathbf{0}_{T\times(k+1)}$ under various empirically relevant assumptions (see e.g. Pesaran and Tosetti, 2011), where '$\to_p$' represents convergence in probability. Hence, $\widehat{\mathbf{F}}$ is consistent for the space spanned by $\mathbf{F}$, which is sufficient to control their effect.

[1] In fact, letting $\mathbf{F} = [\mathbf{1}_T, \boldsymbol{\theta}]'$ and $[a_i, 1]'$, such that $\mathbf{1}_T$ is a $T\times 1$ vector of ones and $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_T]'$ is a vector with time–specific parameters, we obtain $\mathbf{F}\boldsymbol{\gamma}_i = a_i\mathbf{1}_T + \boldsymbol{\theta}$, thus fixed effects framework is a special case of the interactive effects.

As a result, we have two versions of CCE estimator, namely pooled (CCEP) and mean group (CCEMG):

$$\widehat{\boldsymbol{\beta}}_P = \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M}_{\widehat{F}} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M}_{\widehat{F}} \mathbf{y}_i, \tag{1.5}$$

$$\widehat{\boldsymbol{\beta}}_{MG} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{\beta}}_i = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{X}_i' \mathbf{M}_{\widehat{F}} \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{M}_{\widehat{F}} \mathbf{y}_i, \tag{1.6}$$

where $\mathbf{M}_A = \mathbf{I}_T - \mathbf{P}_A = \mathbf{I}_T - \mathbf{A}(\mathbf{A}'\mathbf{A})^+ \mathbf{A}'$ is a projection matrix for an arbitrary T-rowed matrix $\mathbf{A}$ and $(\mathbf{A}'\mathbf{A})^+$ is the Moore-Penrose (MP) inverse of $\mathbf{A}'\mathbf{A}$. The estimator in (1.5) assumes that $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all $i = 1, \ldots, N$, while the one in (1.6) takes potential parameter heterogeneity into account. As was alluded in Pesaran (2006) and further extensively explored in Westerlund and Urbain (2015), CCEP estimator is asymptotically biased if $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all $i = 1, \ldots, N$ is true, unless $TN^{-1} \to 0$. In fact, due to its complicated nature, this bias has attracted much attention within theoretical CCE literature (see e.g. Karabiyik et al., 2017, De Vos and Everaert, 2021, or De Vos and Stauskas, 2021).

On the other hand, if $\boldsymbol{\beta}_i$ is unit–specific, such that $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{v}_i$, where $\boldsymbol{v}_i$ is mean-zero random variable, both (1.5) and (1.6) are asymptotically normal and unbiased with no restrictions on N, T expansion rate. The reason for such elegant property is the fact that the stochastic component $\boldsymbol{v}_i$ dominates the asymptotic distribution. The parameter heterogeneity assumption is very popular in applied literature, where individual-specific economic relationships are the focus, while it is important to control for possible unobserved heterogeneity and cross-section dependence. For instance, in macroeconomic setting this need arises in estimation of cross-country growth regressions (see e.g. Eberhardt and Teal, 2011, or Eberhardt and Presbitero, 2015) or analysis of real estate price elasticity (see e.g. Holly et al., 2010). An example from microeconomics is the average treatment effect (see Petrova and Westerlund, 2020).

The aim of the current study is to consider CCEP and CCEMG estimators and their convenient properties under $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{v}_i$ in the light of new theoretical findings that previously concerned mostly homogeneous $\boldsymbol{\beta}$. Firstly, we focus on $\mathbf{F}$, which was assumed to be stationary in the previous bulk of literature. In particular, $T^{-1}\mathbf{F}'\mathbf{F}$ is usually assumed to have a constant positive definite limit. Although Kapetanios et al. (2011) allow unit root factors, such setup is rather specific and it ignores other empirically relevant scenarios. Instead, we adopt the framework employed by Westerlund (2018) in case of homogeneous $\boldsymbol{\beta}$. It allows many different candidates for $\mathbf{F}$ without any knowledge thereof, save for some regularity conditions which ensure that $\mathbf{D}_{T,F}^{-1}\mathbf{F}'\mathbf{F}\mathbf{D}_{T,F}^{-1}$ has (almost surely) positive definite limit as $T \to \infty$, where the matrix $\mathbf{D}_{T,F}$ will be described below together with our assumptions. Examples of such factors include (mixtures of) polynomial trends of finite order, stochastic series of finite integration order, (near) unit root series or stationary series with absolute summable autocovariances. This approach is especially convenient, as it alleviates uncertainty over the deterministic model components, since $\mathbf{F}$ can absorb the time specific components a researcher is unsure of.

Secondly, we derive the asymptotic distributions following Karabiyik et al. (2017), who showed that the asymptotic behavior of $(\mathbf{D}_{T,\widehat{F}}^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}}\mathbf{D}_{T,\widehat{F}}^{-1})^{+}$ differs depending on whether $m = k+1$ or $m < k+1$. In the latter case, given that $\boldsymbol{\beta}$ is homogeneous, we obtain extra non-estimable bias components as a price for the asymptotic boundedness of the second moment matrix. If, however, the slopes are heterogeneous, it turns out that the analysis is very different from the standard steps in Pesaran (2006) and that the higher moment conditions of the idiosyncratic components $\mathbf{v}_{i,t}$ and $\varepsilon_{i,t}$ are necessary to derive the asymptotic distributions. To our knowledge, De Vos and Stauskas (2021) were the first to address this previously unnoticed issue for CCEP and CCEMG. Therefore, it is important to understand how this new approach squares with non-stationary factors. As a result, we show that under individual specific coefficients both CCEP and CCEMG estimators are still asymptotically normal and unbiased given the general unknown factors. This further demonstrates a wide range of applicability of the CCE type estimators and provides a more uniform theoretical framework for the heterogeneous slope analysis.

2 Assumptions

We state and discuss assumptions under which we work in this paper. Throughout the paper we use the following notation: $\text{rk}(\mathbf{A})$ represents the rank of an arbitrary matrix $\mathbf{A}$, while $\text{vec}(\mathbf{A})$ vectorizes $\mathbf{A}$ by putting its columns on top of each other and $\otimes$ stands for Kronecker product. Moreover, $\|\mathbf{A}\| = \sqrt{\text{tr}(\mathbf{A}'\mathbf{A})}$ is the Frobenius (Euclidean) norm and $\text{tr}(\mathbf{A})$ is the trace, while '$\to_d$' and '$\Rightarrow$' stand for convergence in distribution and weak convergence, respectively. Next, $\mathbf{x}_n = O_p(a_n)$ means that a random vector sequence $\mathbf{x}_n$ is at most of order a_n in probability, where a_n is some deterministic sequence, while $\mathbf{x}_n = o_p(a_n)$ means it is of smaller order in probability than a_n. Finally, $M < \infty$ is some generic positive constant, which is not necessarily always of the same value when applied to different statements.

Assumption 1 *(Idiosyncratic errors) $\varepsilon_{i,t}$ and $\mathbf{v}_{i,t}$ are stationary and independent across i with absolute-summable autocovariances, $\mathbb{E}(\varepsilon_{i,t}) = 0$, $\mathbb{E}(\mathbf{v}_{i,t}) = \mathbf{0}_{k\times 1}$, $\sigma_i^2 = \mathbb{E}(\varepsilon_{i,t}^2)$, $\mathbf{\Sigma}_{\mathbf{V}_i} = \mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}_{i,t}')$, $\mathbf{\Omega}_i = \mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')$, with $\mathbf{\Omega}_i, \mathbf{\Sigma}_{\mathbf{V}_i}$ positive definite and $\mathbb{E}(\varepsilon_{i,t}^6) < M$, $\mathbb{E}(\|\mathbf{v}_{i,t}\|^8) < M$ for all i and t. Moreover, $\frac{1}{N}\sum_{i=1}^N \sigma_i^2 \to \sigma^2 < M$ and $\frac{1}{N}\sum_{i=1}^N \mathbf{\Sigma}_{\mathbf{V}_i} \to \mathbf{\Sigma}_{\mathbf{V}}$ with $\|\mathbf{\Sigma}_{\mathbf{V}}\| < M$ as $N \to \infty$, and we define $\mathbf{\Sigma}_{\mathbf{U}_i} = \mathbb{E}(\mathbf{u}_{i,t}\mathbf{u}_{i,t}') = \mathbf{B}'\mathbb{E}(\widetilde{\mathbf{u}}_{i,t}\widetilde{\mathbf{u}}_{i,t}')\mathbf{B} = \mathbf{B}'\mathbf{\Sigma}_{\widetilde{\mathbf{U}}_i}\mathbf{B}$ and $\frac{1}{N}\sum_{i=1}^N \mathbf{\Sigma}_{\mathbf{U}_i} \to \mathbf{B}'\mathbf{\Sigma}_{\widetilde{\mathbf{U}}}\mathbf{B} = \mathbf{\Sigma}_{\mathbf{U}}$ positive definite, where $\mathbf{\Sigma}_{\widetilde{\mathbf{U}}_i} = [[\sigma_i^2, \mathbf{0}_{1\times k}]', [\mathbf{0}_{k\times 1}, \mathbf{\Sigma}_{\mathbf{V}_i}]']$ and $\mathbf{\Sigma}_{\widetilde{\mathbf{U}}} = [[\sigma^2, \mathbf{0}_{1\times k}]', [\mathbf{0}_{k\times 1}, \mathbf{\Sigma}_{\mathbf{V}}]']$ and $\mathbf{B}$ comes from $\mathbf{B}_i = \mathbf{B} + \widetilde{\mathbf{B}}_i$, such that $\widetilde{\mathbf{B}}_i = [[0, \boldsymbol{v}_i]', [\mathbf{0}_{k\times 1}, \mathbf{0}_{k\times k}]']$.*

Assumption 2 *(Common factors) Consider the $m \times m$ matrix $\mathbf{D}_{T,F} = \mathrm{diag}(T^{p_1}, ..., T^{p_m})$ with $p_j \geq 1/2$ for all j. Given this matrix, the following holds:*

(i) *$T^\kappa \cdot \mathbb{E}(\|\mathbf{D}_{T,F}^{-1}\mathbf{F}'\mathbf{F}\mathbf{D}_{T,F}^{-1} - \mathbf{\Sigma}_F\|^2) \leq M$ for some $\kappa > 0$ and some $m \times m$ matrix $\mathbf{\Sigma}_F$, which is such that $\mathbb{E}(\|\mathbf{\Sigma}_F\|^2) \leq M$ and $\mathbb{P}(\mathrm{rk}(\mathbf{\Sigma}_F) = m) = 1$.*

(ii) *$N \cdot \mathbb{E}(\|\mathbf{D}_{T,F}^{-1}\mathbf{F}'\overline{\mathbf{V}}\|^2) \leq M$ and $\mathbb{E}(\|\mathbf{D}_{T,F}^{-1}\mathbf{F}'\mathbf{V}_i\|^2) \leq M$ for all i. The same is true when $\overline{\mathbf{V}}$ and $\mathbf{V}_i$ are replaced by $\overline{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\varepsilon}_i$, respectively.*

Assumption 3 *(Factor loadings) The $\mathbf{C}_i$ are generated according to*

$$\mathbf{C}_i = \widetilde{\mathbf{C}}_i\mathbf{B}_i = (\widetilde{\mathbf{C}} + \widetilde{\boldsymbol{\eta}}_i)\mathbf{B}_i = \mathbf{C} + \boldsymbol{\eta}_i, \qquad vec(\widetilde{\boldsymbol{\eta}}_i) \sim IID(\mathbf{0}_{m(1+k)}, \mathbf{\Omega}_{\widetilde{\eta}}), \tag{2.1}$$

where $\widetilde{\mathbf{C}} = \mathbb{E}(\widetilde{\mathbf{C}}_i) = [\boldsymbol{\gamma}, \mathbf{\Gamma}]$, $\mathbf{\Omega}_{\widetilde{\eta}} = \mathbb{E}(vec(\widetilde{\boldsymbol{\eta}})vec(\widetilde{\boldsymbol{\eta}})')$ positive definite and $\left\|\widetilde{\mathbf{C}}\right\|, \|\mathbf{\Omega}_{\widetilde{\eta}}\| < M$.

Assumption 4 *(Rank condition) If $m < k+1$, then $\overline{\mathbf{C}} = [\overline{\mathbf{C}}_m, \overline{\mathbf{C}}_{-m}]$, where $\overline{\mathbf{C}}_m$ and $\overline{\mathbf{C}}_{-m}$ are $m \times m$ and $m \times (k+1-m)$, respectively, whereas if $m = k+1$, then $\overline{\mathbf{C}} = \overline{\mathbf{C}}_m$. In both cases, $\mathrm{rk}(\overline{\mathbf{C}}_m) = m$ and $\|\overline{\mathbf{C}}\| \leq M$.*

Assumption 5 *(Independence) $\mathbf{f}_t, \varepsilon_{i,s}, \mathbf{v}_{j,l}, \widetilde{\boldsymbol{\eta}}_n$ are mutually independent for all i, j, n, t, s, l.*

Assumption 6 *(Slope heterogeneity) The heterogeneous slope coefficients follow*

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{v}_i, \qquad \boldsymbol{v}_i \sim IID(\mathbf{0}_{k\times 1}, \mathbf{\Omega}_{\boldsymbol{v}_i})$$

with $\mathbf{\Omega}_{\boldsymbol{v}_i}$ a finite nonnegative definite $k \times k$ matrix such that $\frac{1}{N}\sum_{i=1}^N \mathbf{\Omega}_{\boldsymbol{v}_i} \to \mathbf{\Omega}_{\boldsymbol{v}}$ and the $\boldsymbol{v}_i$ are independent of $\mathbf{f}_t, \varepsilon_{i,s}, \mathbf{v}_{j,l}, \widetilde{\boldsymbol{\eta}}_n$ for all i, j, n, t, s, l and $\mathbb{E}(\|\boldsymbol{v}_i\|^6) < M$.

Here, Assumption 2 comes from Westerlund (2018) and it helps to fully generalize the factor structure. It is formulated as a high level moment condition and it avoids imposing a specific form on $\mathbf{F}$. This is convenient, because it allows to carry out the distribution analysis under many possible scenarios simultaneously as $\mathbf{D}_{T,F}^{-1}\mathbf{F}'\mathbf{F}\mathbf{D}_{T,F}^{-1}$ is an important driver of the distribution. To illustrate (i) and (ii) in Assumption 2, one of empirically relevant scenarios is the polynomial trend of a finite order m. In particular, let $\mathbf{f}_t = \mathbf{t}_t = [1, t, t^2, \ldots, t^{m-1}]' \in \mathbb{R}^{m\times 1}$, and $\mathbf{r}_t \in \mathbb{R}^{p\times 1}$ be zero mean white noise with $\mathbb{E}(\mathbf{r}_t\mathbf{r}_t') = \mathbf{I}_p$ for simplicity. Then, by defining $\mathbf{D}_{T,F} = \text{diag}[T^{1/2}, T^{3/2}, \ldots, T^{(m+1)/2}]$, we obtain

$$\text{(i) } \mathbf{D}_{T,F}^{-1}\mathbf{F}'\mathbf{F}\mathbf{D}_{T,F}^{-1} = \mathbf{D}_{T,F}^{-1}\sum_{t=1}^{T}\mathbf{t}_t\mathbf{t}_t'\mathbf{D}_{T,F}^{-1} \to \int_{s=0}^{1}\mathbf{s}_{1,m}\mathbf{s}_{1,m}'ds,$$

$$\text{(ii) } \mathbf{D}_{T,F}^{-1}\mathbf{F}'\mathbf{R} = \mathbf{D}_{T,F}^{-1}\sum_{t=1}^{T}\mathbf{t}_t\mathbf{r}_t' \Rightarrow \int_{s=0}^{1}\mathbf{s}_{1,m}d\mathbf{W}_{\mathbf{r}}(s)'$$

as $T \to \infty$, where $\mathbf{W}_{\mathbf{r}}(s)$ is the p–variate Wiener process generated by $\{\mathbf{r}_s\}_{s=1}^{t}$, and the typical elements in (i) and (ii) are $\int_{s=0}^{1} s^j s^k ds$ for $j, k = 0, \ldots, m-1$ or $\int_{s=0}^{1} s^j dW_{r,k}(s)$ for $j = 0, \ldots, m-1$ and $k = 1, \ldots, p$, respectively. Also, the matrix in (i) does not have to be deterministic. Following Phillips (1987), we can define $\mathbf{f}_t = \mathbf{h}_t = \mathbf{\Theta}(\mathbf{c})\mathbf{h}_{t-1} + \mathbf{w}_t \in \mathbb{R}^{m\times 1}$, which is a vector autoregressive process constituted of heterogeneous near unit root coordinates, such that $\mathbf{\Theta}(\mathbf{c}) = \text{diag}\left[\theta(c_1),\ \ \theta(c_2),\ \ \ldots,\ \ \theta(c_m)\right]$. Here, $\theta(c_l) = \exp(c_l/T) = 1 + c_l/T + o(1)$ for $l = 1, \ldots, m$ with $c_l < 0$ ($\mathbf{c} \to \mathbf{0}_m$ gives a simple unit root process) and $\mathbf{h}_0 = \mathbf{0}_m$. Also, $\mathbf{w}_t$ is independent and identically distributed (IID) with $\mathbb{E}(\mathbf{w}_t) = \mathbf{0}_m$ and $\mathbb{E}(\mathbf{w}_t\mathbf{w}_t') = \mathbf{I}_m$. Letting $\mathbf{D}_{T,F} = T\mathbf{I}_m$, we obtain as $T \to \infty$

$$\text{(i) } \mathbf{D}_{T,F}^{-1}\mathbf{F}'\mathbf{F}\mathbf{D}_{T,F}^{-1} = \mathbf{D}_{T,F}^{-1}\sum_{t=2}^{T}\mathbf{h}_t\mathbf{h}_t'\mathbf{D}_{T,F}^{-1} \Rightarrow \int_{s=0}^{1}\mathbf{J}_{\mathbf{c}}(s)\mathbf{J}_{\mathbf{c}}(s)'ds,$$

$$\text{(i) } \mathbf{D}_{T,F}^{-1}\mathbf{F}'\mathbf{R} = \mathbf{D}_{T,F}^{-1}\sum_{t=2}^{T}\mathbf{h}_t\mathbf{r}_t' \Rightarrow \int_{s=0}^{1}\mathbf{J}_{\mathbf{c}}(s)d\mathbf{W}_{\mathbf{r}}(s)',$$

where $\mathbf{J}_{\mathbf{c}}(s) = \int_{u=0}^{s}\text{diag}\left[\exp[c_1(s-u)], \ldots, \exp[c_m(s-u)]\right]d\mathbf{W}_{\mathbf{w}}(u) \in \mathbb{R}^{m\times 1}$ is the Ornstein-Uhlenbeck process and $\mathbf{W}_{\mathbf{w}}(u)$ is the m–variate Wiener process generated by $\{\mathbf{w}_s\}_{s=1}^{t}$. In the Supplementary material, we provide more complicated examples of such limiting matrices. Note that despite only weak convergence in such cases, by employing arguments from Park and Phillips (2001), the sample space can be enlarged such that $\mathbf{D}_{T,F}^{-1}\mathbf{F}'\mathbf{F}\mathbf{D}_{T,F}^{-1} \to \mathbf{\Sigma}_F$ almost surely.

The rest of the assumptions are similar to the ones in Pesaran (2006) or Karabiyik et al.

(2017). One assumption which is different between those two important studies and is used here is the random factor loadings. As Westerlund and Urbain (2013) demonstrate, under the random loadings it is possible to obtain consistent estimate of $\boldsymbol{\beta}_i$ even if $m > k+1$ provided that γ_i and $\boldsymbol{\Gamma}_i$ are independent. However, this is unlikely in practice and we acknowledge this by letting $\boldsymbol{\Omega}_{\widetilde{\eta}}$ be only positive definite and not block-diagonal. A more significant difference is reflected in higher moment conditions in Assumption 1 and Assumption 6. The intuition for this requirement is as follows. Provided that the condition $m \leq k+1$ is taken into account properly, some terms that used to be treated as negligible in, for example, the standard analysis in equation (56) and further in Pesaran (2006) are now at most $O_p(\sqrt{N}T^{-1})$. Therefore, in order to derive the asymptotic distributions without any restrictions on N, T expansion rate, the price of higher moment requirements comes into play. Interestingly, these assumptions are closer to the principal components setup of Bai and Ng (2002).

Remark 1. Assumption 4 makes it more difficult to put many unknown deterministic components in $\mathbf{F}$. However, letting $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2]$, where $\mathbf{F}_1 \in \mathbb{R}^{T\times m_1}$, $\mathbf{F}_2 \in \mathbb{R}^{T\times(m-m_1)}$ and it represents $m - m_1$ known factors, then they can be estimated unrestrictedly. An example of this is individual fixed effects that are common in many empirical applications.

Remark 2. Note that the cross-section independence of $\widetilde{\mathbf{U}}_i$ comes purely for convenience. We can relax this along the lines of Pesaran and Tosetti (2011) by requiring that $\widetilde{\mathbf{u}}_t = (\mathbf{M} \otimes \mathbf{I}_m)\boldsymbol{\nu}_t$, where $\widetilde{\mathbf{u}}_t \in \mathbb{R}^{N(k+1)\times 1}$ is a cross-section stack of $\widetilde{\mathbf{u}}_{i,t}$ and $\boldsymbol{\nu}_t$ obeys Assumption 1. Here, $\mathbf{M}$ is an $N \times N$ 'network matrix' with bounded row and column norms.

3 Asymptotic Results

3.1 CCEMG Results

In this section we derive the asymptotic distribution based on (1.6). As suggested by the previous literature, it must be driven by the coefficient heterogeneity component.

Theorem 1. *Under Assumptions 1–6, as $N, T \to \infty$ unrestrictedly,*

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{MG} - \boldsymbol{\beta}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{v}_i + o_p(1) \to_d \mathcal{N}\left(\mathbf{0}_k, \boldsymbol{\Omega}_{\boldsymbol{v}}\right),$$

where $\boldsymbol{\Omega}_{\boldsymbol{v}} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(\boldsymbol{v}_i\boldsymbol{v}_i')$.

As can be seen, indeed the distribution is driven by the heterogeneity component of the coefficients and the distribution is bias-free. We retain the $\sqrt{N}$ rate of consistency. These findings coincide with the ones in Pesaran (2006) or Kapetanios et al. (2011), where the analysis was done either under stationary or pure unit root $\mathbf{F}$, respectively. We can see that the result remains robust to a way more general factor structure and to the correctly accounted case of $m \leq k+1$. Also, the results stay unchanged if $m = k+1$. Therefore, this result is reassuring for practical purposes. It implies that usage of excess $k+1-m$ cross-section averages to control for common factors does not induce a non-estimable bias term, which should be dealt with either by imposing restrictions on N, T expansion rate or employing bootstrap methods (see e.g. De Vos and Stauskas, 2021).

3.2 CCEP Results

It is clearly possible to pool information from the heterogeneous coefficients by using CCEP estimator. Again, the previous literature suggests that the asymptotic distribution should be driven by $\boldsymbol{v}_i$. Therefore, we derive the asymptotic distribution based on (1.5).

Theorem 2. *Under Assumptions 1–6, as $N, T \to \infty$ unrestrictedly,*

$$\begin{aligned}\sqrt{N}(\widehat{\boldsymbol{\beta}}_P - \boldsymbol{\beta}) &= \boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(T^{-1}\mathbf{V}_i'\mathbf{V}_i\right)\boldsymbol{v}_i + o_p(1)\\ &\to_d \mathcal{N}\left(\mathbf{0}_k, \boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\boldsymbol{\Psi}_v\Sigma_{\mathbf{V}}^{-1}\right),\end{aligned}$$

where $\boldsymbol{\Psi}_{\boldsymbol{v}} = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\left(T^{-1}\mathbf{V}_i'\mathbf{V}_i\right)\boldsymbol{\Omega}_{\boldsymbol{v}_i}\left(T^{-1}\mathbf{V}_i'\mathbf{V}_i\right)\right]$.

This result is in nature similar to the one in Theorem 2. Again, given Assumption 5, the coefficient heterogeneity component $\boldsymbol{v}_i$ dominates the asymptotic distribution because it is independent from $\mathbf{V}_i$. The estimator is $\sqrt{N}$–consistent under general unknown factors. Even under pooled estimator and potentially non-stationary factors the asymptotic distribution is correctly centered at zero and the result is robust to $m \leq k+1$ condition. Clearly, under $m = k+1$ this result remains unchanged, similarly to the result in Theorem 1.

3.3 Inference

In order to use the results from Theorem 1 and 2 for statistical inference, we need to estimate $\mathbf{\Omega}_v$ and $\mathbf{\Sigma}_{\mathbf{V}}^{-1}\mathbf{\Psi}_v\Sigma_{\mathbf{V}}^{-1}$ consistently. For this, we analyze two versions of robust variance matrix estimator for CCEP and CCEMG estimators proposed by Pesaran (2006). In particular,

$$\widehat{\mathbf{\Theta}} = \left(\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i\right)^{-1}\left(\frac{1}{N(N-1)}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i(\widehat{\boldsymbol{\beta}}_i-\widehat{\boldsymbol{\beta}}_{MG})(\widehat{\boldsymbol{\beta}}_i-\widehat{\boldsymbol{\beta}}_{MG})'\widehat{\mathbf{Q}}_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{Q}}_i\right)^{-1}, \tag{3.1}$$

where $\widehat{\mathbf{Q}}_i = T^{-1}\mathbf{X}_i'\mathbf{M}_{\widehat{F}}\mathbf{X}_i$ and

$$\widehat{\mathbf{\Omega}}_v = \frac{1}{N(N-1)}\sum_{i=1}^{N}(\widehat{\boldsymbol{\beta}}_i-\widehat{\boldsymbol{\beta}}_{MG})(\widehat{\boldsymbol{\beta}}_i-\widehat{\boldsymbol{\beta}}_{MG})'. \tag{3.2}$$

Both (3.1) and (3.2) show a good small sample performance in case of stationary factors. Theorem 3 establishes the consistency results under general unknown factors.

Theorem 3. *Under Assumptions 1–6, as $N, T \to \infty$ unrestrictedly,*

$$N\widehat{\mathbf{\Theta}} \to_p \mathbf{\Sigma}_{\mathbf{V}}^{-1}\mathbf{\Psi}_v\Sigma_{\mathbf{V}}^{-1}, \quad \textit{and} \quad N\widehat{\mathbf{\Omega}}_v \to_p \mathbf{\Omega}_v.$$

The result in Theorem 3 implies that

$$t_{\widehat{\beta}_{MG}} = \frac{\mathbf{a}_j'(\widehat{\boldsymbol{\beta}}_{MG}-\boldsymbol{\beta})}{\sqrt{\mathbf{a}_j'\widehat{\mathbf{\Omega}}_v\mathbf{a}_j}} \to_d \mathcal{N}(0,1),\;\; t_{\widehat{\beta}_P} = \frac{\mathbf{a}_j'(\widehat{\boldsymbol{\beta}}_P-\boldsymbol{\beta})}{\sqrt{\mathbf{a}_j'\widehat{\mathbf{\Theta}}\mathbf{a}_j}} \to_d \mathcal{N}(0,1) \tag{3.3}$$

as $N, T \to \infty$, where $\mathbf{a}_j$ is a vector of zeros with 1 in the j–th coordinate for $j = 1, \ldots, k$. We will further evaluate these theoretical predictions regarding the variance estimators and test statistics in Monte Carlo simulations.

Remark 3. The results in Theorem 3 can be generalized to weak cross-section dependence–robust covariance matrices from Pesaran and Tosetti (2011).

4 Monte Carlo Study

In this section we carry out a small-scale Monte Carlo exercise to evaluate the performance of CCEP and CCEMG when factors are non-stationary. We will consider four scenarios, where T1 and T2 reflect deterministic trends, while R1 and R2 describe the effect of stochastic trends. In particular,

T1. $\mathbf{f}_t = [1, t]'$, which corresponds to incidental trend setting.

T2. $\mathbf{f}_t = [1, t, t^2]'$, which augments the trend further.

R1. $\mathbf{f}_t = [1, g_t]'$, where $g_t = g_{t-1} + u_t$, such that $u_t \sim \mathcal{N}(0, 1)$.

R2. $\mathbf{f}_t = [1, t, g_t]'$, which effectively makes $\mathbf{y}_i$ a unit root process with an incidental drift.

These versions of $\mathbf{f}_t$ are transparent, yet more complex than in Westerlund (2018), because we combine deterministic and stochastic factors. Further, model error components are generated as follows:

$$\mathbf{v}_{i,t} = \rho_{\mathbf{v}} \mathbf{v}_{i,t-1} + \sqrt{1 - \rho_{\mathbf{v}}^2}\mathbf{e}_{i,t}, \quad \mathbf{e}_{i,t} \sim \mathcal{N}(\mathbf{0}_{k\times 1}, \sigma_{\mathbf{e},i}^2 \mathbf{I}_k) \tag{4.1}$$

$$\varepsilon_{i,t} = \rho_{\varepsilon} \varepsilon_{i,t-1} + \sqrt{1 - \rho_{\varepsilon}^2}\xi_{i,t}, \quad \xi_{i,t} \sim \mathcal{N}(0, \sigma_{\xi,i}^2) \tag{4.2}$$

$$\widetilde{\mathbf{C}}_i = \widetilde{\mathbf{C}} + \widetilde{\boldsymbol{\eta}}_i \mathbf{1}_{k+1}', \quad \widetilde{\boldsymbol{\eta}}_i \sim \mathcal{N}(\mathbf{0}_{m\times 1}, \sigma_{\boldsymbol{\eta}}^2 \mathbf{I}_m), \tag{4.3}$$

where $\rho_{\mathbf{v}} = \rho_{\varepsilon} = 0.8$ to introduce a degree of persistence. Also, $\sigma_{\mathbf{e},i}^2 = \sigma_{\mathbf{e}}^2 + (w_{\mathbf{e},i} - 1)$, $\sigma_{\xi,i}^2 = \sigma_{\xi}^2 + (w_{\xi,i} - 1)$, where $w_{\mathbf{e},i}$ and $w_{\xi,i} \sim \chi^2(1)$. We set $\sigma_{\mathbf{e}}^2 = \sigma_{\xi}^2 = 2$ and $\sigma_{\boldsymbol{\eta}}^2 = 1$. Here, $\widetilde{\mathbf{C}}$ is a constant matrix, such that $\mathrm{rk}(\widetilde{\mathbf{C}}) = m$ and thus $\mathrm{rk}(\overline{\mathbf{C}}) = m$ almost surely. Note that (4.3) gives an extreme case because $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\gamma}_i$ are perfectly correlated within an individual. This setup is more complex than in Westerlund (2018) due to dynamics and heterogeneity in the errors components, similarly to Kapetanios et al. (2011). Finally,

$$\boldsymbol{\beta}_i = \mathbf{1}_k + \boldsymbol{v}_i, \quad \boldsymbol{v}_i \sim \mathcal{N}(\mathbf{0}_{k\times 1}, \sigma_v^2 \mathbf{I}_k) \tag{4.4}$$

where $\sigma_v^2 = 0.02$ is directly taken from Pesaran and Tosetti (2011) as it ensures that the parameters are not too far from the mean. We set $k = 3$, therefore we maintain $m < k + 1$ for all the factor scenarios, which is an important aspect of our theory. We run 1000 simulation rounds for $N, T \in \{25, 50, 100, 200, 500\}$ and allow for 20 burn-in values in order to eliminate the effect of initial values. We report 5% empirical size, bias and root mean-squared error (RMSE). In the Supplement, we provide additional simulations, where we increase $\sigma_{\boldsymbol{\eta}}^2$ and σ_v^2 and also simulate $\boldsymbol{v}_i$ from a different distribution.

We begin with Table 1 and Table 2, which involve deterministic components only. We see that the bias and RMSE generally decline with N and T as expected and the size does not suffer from significant distortions. Slight jumps in bias occur when T starts dominating N. The t–test becomes a little oversized when, again, $T >> N$ and N is rather small. This is expected, because approximation of the factor space is not ideal and this is especially visible for T2 (in Table 2), where the quadratic trend dominates. In general, both CCEP and CCEMG perform similarly for the same N, T combination, and for larger N, T values the empirical size hovers around the nominal 5% level. These results are strikingly different from Monte Carlo results in Westerlund (2018) for deterministic trends, where $\boldsymbol{\beta}$ is homogeneous and huge size distortions occur.

Going to Table 3 and Table 4, a very important observation is that the overall results are very similar to the ones in Table 1 and Table 2. Again, we see a similar performance of both CCEP and CCEMG under R1 and R2. That is, the size distortions occur when T heavily dominates N. Again, the performance of CCEP and CCEMG is similar even when the most significant distortions occur, which happens in case of R2 – the most complex scenario combining both deterministic and stochastic trends. Overall, for large N, T with no restrictions, the size is close to the nominal 5% level.

The patterns reported in Table 1 – Table 4 resemble the results in De Vos and Stauskas (2021), where different stress tests for CCEP and CCEMG estimators under heterogeneous slopes were evaluated under purely stationary factors. This suggests that our theoretical predictions are borne out well and the nature of the factors essentially does not matter as long as the regularity conditions are satisfied. Clearly, while both estimators are theoretically unbiased, some remainder exists for the finite N and T. Hence, the similarity between t-test performance for the stationary and non-stationary cases suggests that the inference under trending factors for finite N and T could potentially be enhanced via bootsrapping procedures, as was demonstrated in this other paper for stationary factors only.

5 Conclusions

In this study we discussed the effect of non-stationary factors in case of a popular CCE estimator. Specifically, we extended the framework of general unknown factors from Westerlund (2018) to heterogeneous slopes. As our theorems showed and simulations confirmed, irrespective of the nature of the common factors, both CCEP and CCEMG are asymptotically normal and unbiased. The results hold even if we over-control for the number of factors, therefore this study further stresses flexibility and wide applicability of the CCE procedure.

References

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

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

De Vos, I. and Stauskas, O. (2021). Bootstrap improved inference for factor-augmented regressions with cce. *Submitted Manuscript*.

Eberhardt, M. and Presbitero, A. F. (2015). Public debt and growth: Heterogeneity and non-linearity. *Journal of international Economics*, 97(1):45–58.

Eberhardt, M. and Teal, F. (2011). Econometrics for grumblers: a new look at the literature on cross-country growth empirics. *Journal of Economic Surveys*, 25(1):109–155.

Holly, S., Pesaran, M. H., and Yamagata, T. (2010). A spatio-temporal model of house prices in the usa. *Journal of Econometrics*, 158(1):160–173.

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

Karabiyik, H., Reese, S., and Westerlund, J. (2017). On the role of the rank condition in CCE estimation of factor-augmented panel regressions. *Journal of Econometrics*, 197(1):60 – 64.

Park, J. Y. and Phillips, P. C. (2001). Nonlinear regressions with integrated time series. *Econometrica*, 69(1):117–161.

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

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

Petrova, Y. and Westerlund, J. (2020). Fixed effects demeaning in the presence of interactive effects in treatment effects regressions and elsewhere. *Journal of Applied Econometrics*, 35(7):960–964.

Phillips, P. C. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3):535–547.

Westerlund, J. (2018). Cce in panels with general unknown factors. *The Econometrics Journal*, 21(3):264–276.

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

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

Table 1: Simulation results for T1

		CCEP			CCEMG		
N	T	5% size	Bias $\times$100	RMSE$\times$100	5% size	Bias $\times$100	RMSE$\times$100
25	25	6.9	0.989	8.204	6.3	1.050	8.565
25	50	7.7	1.149	6.830	7.4	1.296	6.832
25	100	8.5	1.422	5.586	7.8	1.462	5.422
25	200	9.6	1.114	4.629	8.3	1.367	4.362
25	500	9.1	1.049	3.936	9.0	1.205	3.611
50	25	6.2	0.109	5.996	6.0	0.457	6.144
50	50	6.5	0.383	4.984	6.7	0.683	4.922
50	100	7.3	0.687	4.003	6.8	0.951	3.935
50	200	6.2	0.732	3.335	6.4	0.836	3.155
50	500	7.7	0.800	2.831	7.0	0.802	2.559
100	25	6.1	0.445	4.129	5.7	0.678	4.192
100	50	5.4	0.183	3.378	5.5	0.397	3.357
100	100	5.4	0.289	2.644	5.3	0.295	2.671
100	200	5.5	0.433	2.323	6.1	0.516	2.204
100	500	6.7	0.266	2.006	5.8	0.385	1.783
200	25	5.6	0.142	2.950	6.0	0.182	3.163
200	50	5.2	0.275	2.381	5.8	0.350	2.460
200	100	5.5	0.119	1.966	5.7	0.148	1.938
200	200	4.7	0.135	1.675	5.1	0.218	1.585
200	500	5.2	0.064	1.412	5.5	0.131	1.238
500	25	6.0	0.097	1.933	6.2	0.123	1.994
500	50	4.0	0.166	1.505	5.0	0.119	1.525
500	100	4.6	0.146	1.272	4.3	0.111	1.207
500	200	4.5	−0.020	1.011	4.0	0.010	0.968
500	500	5.8	0.019	0.912	4.4	0.052	0.793

Notes: "T1" refers to the case when $\mathbf{f}_t = [1, t]'$. Also $\{\rho_{\mathbf{v}}, \rho_\varepsilon, \sigma^2_{\mathbf{e}}, \sigma^2_{\xi}, \sigma^2_{\boldsymbol{\eta}}, \sigma^2_{v}\} = \{0.8, 0.8, 2, 2, 1, 0.02\}$.

Table 2: Simulation results for T2

		CCEP			CCEMG		
N	T	5% size	Bias ×100	RMSE ×100	5% size	Bias ×100	RMSE ×100
25	25	8.3	2.086	8.123	6.7	2.333	8.136
25	50	9.4	2.045	7.012	8.7	2.439	6.827
25	100	8.5	1.780	5.582	10.0	2.149	5.506
25	200	11.8	2.090	4.776	9.9	2.331	4.471
25	500	11.4	1.717	3.900	11.2	1.953	3.560
50	25	6.7	0.749	5.711	6.1	0.995	5.937
50	50	7.7	0.852	4.851	5.4	1.097	4.654
50	100	7.6	1.113	3.931	7.3	1.282	3.795
50	200	7.9	1.003	3.352	8.1	1.235	3.224
50	500	7.5	0.985	2.708	7.6	1.205	2.426
100	25	6.1	0.488	3.991	5.8	0.775	4.212
100	50	6.1	0.412	3.419	6.1	0.504	3.387
100	100	6.5	0.383	2.803	6.0	0.595	2.678
100	200	6.0	0.519	2.279	6.2	0.628	2.184
100	500	7.7	0.588	2.004	8.5	0.737	1.825
200	25	5.7	0.341	2.859	6.9	0.396	2.898
200	50	4.1	0.289	2.259	4.7	0.372	2.295
200	100	5.0	0.318	1.910	5.4	0.380	1.905
200	200	5.8	0.219	1.640	6.5	0.325	1.595
200	500	5.0	0.318	1.391	5.5	0.368	1.260
500	25	4.0	0.124	1.707	4.2	0.042	1.751
500	50	6.2	0.089	1.564	5.8	0.096	1.535
500	100	4.7	0.143	1.206	4.9	0.194	1.172
500	200	5.4	0.102	1.022	5.5	0.129	0.980
500	500	4.8	0.095	0.888	5.8	0.122	0.806

Notes: "T2" refers to the case when $\mathbf{f}_t = [1, t, t^2]'$. Also, $\{\rho_{\mathbf{v}}, \rho_{\varepsilon}, \sigma^2_{\mathbf{e}}, \sigma^2_{\zeta}, \sigma^2_{\boldsymbol{\eta}}, \sigma^2_{\boldsymbol{v}}\} = \{0.8, 0.8, 2, 2, 1, 0.02\}$.

Table 3: Simulation results for R1

		CCEP			CCEMG		
N	T	5% size	Bias ×100	RMSE ×100	5% size	Bias ×100	RMSE ×100
25	25	8.6	1.468	8.645	8.8	1.614	9.002
25	50	8.1	1.396	7.100	7.8	1.526	7.161
25	100	8.2	1.136	5.492	7.3	1.455	5.500
25	200	8.9	1.163	4.610	8.2	1.226	4.443
25	500	8.1	1.160	3.838	8.3	1.310	3.586
50	25	5.5	0.641	6.123	6.0	0.762	6.250
50	50	6.9	0.377	4.915	7.4	0.524	5.067
50	100	7.4	0.534	3.955	5.3	0.668	3.831
50	200	7.5	0.602	3.306	7.0	0.745	3.111
50	500	8.4	0.631	2.902	7.7	0.783	2.569
100	25	5.6	0.195	4.352	6.5	0.302	4.368
100	50	5.7	0.244	3.468	5.1	0.184	3.320
100	100	6.4	0.216	2.797	5.9	0.259	2.792
100	200	6.3	0.178	2.359	6.5	0.320	2.236
100	500	5.5	0.313	1.927	5.5	0.400	1.710
200	25	4.7	−0.012	2.902	4.6	0.065	2.952
200	50	5.8	0.287	2.459	5.8	0.258	2.443
200	100	5.0	0.126	1.965	4.6	0.112	1.927
200	200	5.7	0.201	1.650	5.3	0.214	1.548
200	500	5.3	0.214	1.384	6.6	0.209	1.278
500	25	5.1	0.037	1.865	3.3	0.074	1.854
500	50	5.1	−0.001	1.536	5.1	0.085	1.550
500	100	5.2	0.097	1.247	5.9	0.081	1.242
500	200	5.3	0.072	1.048	5.0	0.076	0.964
500	500	5.1	0.082	0.911	5.3	0.055	0.818

Notes: "R1" refers to the case when $\mathbf{f}_t = [1, g_t]'$, where $g_t = g_{t-1} + u_t$, such that $u_t \sim \mathcal{N}(0,1)$. Also, $\{\rho_{\mathbf{v}}, \rho_\varepsilon, \sigma^2_{\mathbf{e}}, \sigma^2_{\xi}, \sigma^2_{\boldsymbol{\eta}}, \sigma^2_{\boldsymbol{v}}\} = \{0.8, 0.8, 2, 2, 1, 0.02\}$.

Table 4: Simulation results for R2

		CCEP			CCEMG		
N	T	5% size	Bias $\times$100	RMSE $\times$100	5% size	Bias $\times$100	RMSE $\times$100
25	25	7.5	1.858	7.982	7.7	1.928	8.362
25	50	9.4	2.199	6.832	8.2	2.410	6.769
25	100	10.5	2.236	5.546	9.7	2.502	5.336
25	200	8.7	1.820	4.646	9.4	2.144	4.387
25	500	12.5	1.966	3.933	12.3	2.201	3.633
50	25	4.8	0.876	5.604	5.4	0.934	5.764
50	50	5.9	1.246	4.659	5.7	1.405	4.559
50	100	7.2	0.861	3.896	7.3	1.138	3.846
50	200	8.3	0.987	3.365	9.1	1.212	3.153
50	500	9.0	1.020	2.879	9.0	1.270	2.597
100	25	6.4	0.639	4.158	6.8	0.826	4.462
100	50	6.3	0.364	3.462	5.4	0.445	3.350
100	100	6.7	0.573	2.802	7.9	0.726	2.838
100	200	6.8	0.592	2.293	6.0	0.708	2.175
100	500	5.4	0.494	1.935	7.4	0.643	1.792
200	25	5.0	0.358	2.845	4.0	0.533	2.795
200	50	5.4	0.351	2.366	4.9	0.283	2.344
200	100	4.3	0.278	1.915	5.0	0.418	1.911
200	200	4.8	0.266	1.654	5.7	0.334	1.588
200	500	5.4	0.224	1.437	5.7	0.316	1.276
500	25	5.1	0.092	1.854	5.6	0.167	1.898
500	50	5.4	0.214	1.493	5.9	0.215	1.507
500	100	4.8	0.085	1.252	5.0	0.127	1.211
500	200	5.1	0.135	1.022	4.7	0.140	0.973
500	500	5.8	0.126	0.906	4.9	0.173	0.813

Notes: "R2" refers to the case when $\mathbf{f}_t = [1, t, g_t]'$, where $g_t = g_{t-1} + u_t$, such that $u_t \sim \mathcal{N}(0,1)$. Also, $\{\rho_{\mathbf{v}}, \rho_\varepsilon, \sigma^2_{\mathbf{e}}, \sigma^2_{\xi}, \sigma^2_{\boldsymbol{\eta}}, \sigma^2_{\boldsymbol{v}}\} = \{0.8, 0.8, 2, 2, 1, 0.02\}$.