Long-Run Effects in Large Heterogenous Panel Data Models with Cross-Sectionally Correlated Errors[*]

Alexander Chudik
Federal Reserve Bank of Dallas

Kamiar Mohaddes
Griton College and University of Cambridge

M. Hashem Pesaran
USC Dornsife INET, University of Southern California and Trinity College

Mehdi Raissi
International Monetary Fund

January 2015

Abstract

This paper develops a cross-sectionally augmented distributed lag (CS-DL) approach to the estimation of long-run effects in large dynamic heterogeneous panel data models with cross-sectionally dependent errors. The asymptotic distribution of the CS-DL estimator is derived under coefficient heterogeneity in the case where the time dimension (T) and the cross-section dimension (N) are both large. The CS-DL approach is compared with more standard panel data estimators that are based on autoregressive distributed lag (ARDL) specifications. It is shown that unlike the ARDL type estimator, the CS-DL estimator is robust to misspecification of dynamics and error serial correlation. The theoretical results are illustrated with small sample evidence obtained by means of Monte Carlo simulations, which suggest that the performance of the CS-DL approach is often superior to the alternative panel ARDL estimates particularly when T is not too large and lies in the range of $30 \leq T < 100$.

JEL codes: C23

1 Introduction

Estimation of long-run effects, or level relationships, is of great importance in economics. The concept of "long-run relations" is typically associated with the steady-state solution of a structural macroeconomic model. Often the same long-run relations can also be obtained from arbitrage conditions within and across markets. As a result, many long-run relationships in economics are free of particular model assumptions; examples being purchasing power parity, uncovered interest parity and the Fisher inflation parity. Other long-run relations, such as those between macroeconomic aggregates like consumption and income, output and investment, and technological progress and real wages, are less grounded in arbitrage and hence are more controversial, but still form a major part of what is generally agreed-upon in empirical macroeconomic modelling. This is in contrast to the analysis of short-run effects, which are model specific and subject to identification problems.

This paper is concerned with the estimation and inference of long-run effects using panel data models where the time dimension (T) and the cross-section dimension (N) are both large. Such panels are becoming increasingly available and cover countries, counties, regions, industries and firms, and typically feature dynamics in the form of lagged dependent variables, slope heterogeneity (at least in the case of short-run coefficients), as well as cross-sectionally dependent innovations. These three key features complicate estimation and inference.

Earlier literature on the estimation of long-run effects using panel data, including the pooled mean group approach (Pesaran, Shin, and Smith 1999), the panel dynamic OLS approach (Mark and Sul 2003) and the panel fully modified OLS approach (Pedroni 2001), allows for lagged dependent variables and heterogeneity of short-run dynamics, but it does not allow for error cross-section dependence. Wrongly assuming that errors are independently distributed leads to incorrect inference and in some cases inconsistent estimates, depending on the nature of error cross-section dependence. For example, when cross-section dependence is due to the presence of unobserved common factors, parameter inconsistency arises if the factors and the regressors are correlated.

The problem of error cross-section dependence has been addressed in the literature primarily in the context of panel data models without lagged dependent variables. See, for example, the common correlated effects (CCE) approach of Pesaran (2006), the interactive fixed effects estimator (IFE) of Bai (2009), or the quasi-maximum likelihood estimator (QMLE) of Moon and Weidner (2010). A survey of the recent literature is provided by Chudik and Pesaran (2014b). Two exceptions are Song (2013) who extends Bai's approach to allow for coefficient heterogeneity, and Chudik and Pesaran (2014a), who extend the CCE

approach to allow for weakly exogenous regressors (including lagged dependent variables). Both approaches rely on the estimation of unit-specific ARDL specifications, appropriately augmented with cross-section averages to filter out the effects of the unobserved common factors, from which long-run effects can be indirectly estimated. We refer to this approach as cross-sectionally augmented ARDL or CS-ARDL in short. The main drawback of computing the long-run coefficients from CS-ARDL specifications is that due to the inclusion of lagged dependent variables in the regressions a relatively large time dimension is required for satisfactory small sample performance, especially if the sum of the AR coefficients in the ARDL specifications are close to one. In the case of heterogenous slope specifications the CS-ARDL estimates of the long-run coefficients could also be sensitive to outlier estimates of the long-run effects for individual cross-section units.

This paper makes a theoretical contribution to the econometric analysis of the long run by proposing a new approach to the estimation of the long-run coefficients in dynamic heterogeneous panels with cross-sectionally dependent errors. The approach is based on a distributed lag representation that does not feature lags of the dependent variable, and allows for residual factor error structure and weak cross-section dependence of idiosyncratic errors. Similar to CCE estimators proposed by Pesaran (2006), we appropriately augment the individual regressions by cross-section averages to deal with the effects of common factors. We derive the asymptotic distribution of the proposed cross-sectionally augmented distributed lag (or CS-DL in short) mean group and pooled estimators under the coefficient heterogeneity and large time and cross-section dimensions. We also investigate consequences of various departures from our maintained assumptions by means of Monte Carlo experiments, including a unit root in factors and/or in regressors, homogeneity of coefficients or breaks in error processes. We also investigate whether the imposition of CS-DL estimates of long-run coefficients can improve the estimation of the short-run coefficients.

The main advantage of the proposed CS-DL approach is that its small sample performance is often better compared to estimating unit-specific CS-ARDL specifications, under a variety of settings investigated in the Monte Carlo experiments when T is moderately large $(30 \leq T < 100)$. Furthermore, the imposition of CS-DL estimates of long-run coefficients can substantially improve the estimates of short-run coefficients when T is moderately large. However, the CS-DL approach should be seen as complementary and not as superior to the CS-ARDL approach. The main drawback is that, unlike the panel CS-ARDL approach, the CS-DL approach does not allow for feedback effects from the dependent variable onto the regressors. However, a careful investigation of the size of the small sample bias emanating from the presence of such feedback effects suggests that the CS-DL approach can still outperform the CS-ARDL approach when T is moderately large. The relative merits of different

approaches are carefully documented in the paper, and our main conclusion is that the CS-DL approach is a valuable complementary method for estimating long-run effects in panels where the time dimension is moderately large.

The remainder of the paper is organized as follows. We begin with the definition of long-run coefficients and discuss their estimation in Section 2. The next section introduces the CS-DL approach to the estimation of long-run relationships. Section 4 investigates the small sample performance of the CS-DL approach and compares it with the performance of the CS-ARDL approach by means of Monte Carlo experiments. The last section concludes. Mathematical derivations are relegated to the Appendix.

A brief word on notation: All vectors are column vectors represented by bold lower case letters and matrices are represented by bold capital letters. $\|\mathbf{A}\| = \sqrt{\varrho(\mathbf{A}'\mathbf{A})}$ is the spectral norm of $\mathbf{A}$, $\varrho(\mathbf{A})$ is the spectral radius of $\mathbf{A}$.[1] $a_n = O(b_n)$ denotes the deterministic sequence $\{a_n\}$ is at most of order b_n. Convergence in probability and convergence in distribution are denoted by $\xrightarrow{p}$ and $\xrightarrow{d}$, respectively. $(N, T) \xrightarrow{j} \infty$ denotes joint asymptotics in N and T, with N and $T \to \infty$, in no particular order. We use K to denote a positive fixed constant that does not vary with N or T.

2 Estimation of long-run or level relationships in economics

The estimation of long-run relations can be carried out with or without constraining the short-run dynamics. In this section, we focus on the estimation of long-run relations without restricting the short-run dynamics and assuming that there exists a single long-run relationship between the dependent variable, y_t, and a set of regressors.[2] For illustrative purposes, suppose that there is one regressor x_t and suppose that $\mathbf{z}_t = (y_t, x_t)'$ is jointly determined by the following vector autoregression of order 1, VAR(1),

$$\mathbf{z}_t = \mathbf{\Phi}\mathbf{z}_{t-1} + \mathbf{e}_t, \tag{1}$$

where $\mathbf{\Phi} = (\phi_{ij})$ is a 2×2 matrix of unknown parameters, and $\mathbf{e}_t = (e_{yt}, e_{xt})'$ is a 2-dimensional vector of reduced form errors. Denoting the covariance of e_{yt} and e_{xt} by

---

[1]Note that if $\mathbf{x}$ is a vector, then $\|\mathbf{x}\| = \sqrt{\varrho(\mathbf{x}'\mathbf{x})} = \sqrt{\mathbf{x}'\mathbf{x}}$ corresponds to the Euclidean length of the vector $\mathbf{x}$.

[2]The problem of estimation and inference in the case of multiple long-run relations is further complicated by the identification problem and simultaneous determination of variables. The case of multiple long-run relations is discussed for example in Pesaran (1997).

$\omega Var(e_{xt})$, we can write

$$e_{yt} = E(e_{yt}|e_{xt}) + u_t = \omega e_{xt} + u_t, \tag{2}$$

where by construction u_t is uncorrelated with e_{xt}, namely $E(u_t|e_{xt}) = 0$. Substituting (2) for e_{yt}, the equation for the dependent variable y_t in (1) is

$$y_t = \phi_{11} y_{t-1} + \phi_{12} x_{t-1} + \omega e_{xt} + u_t. \tag{3}$$

Using the equation for the regressor x_t in (1), we obtain the following expression for e_{xt}

$$e_{xt} = x_t - \phi_{21} y_{t-1} - \phi_{22} x_{t-1},$$

and substituting this expression for e_{xt} back in (3) yields the following *conditional* model for y_t,

$$y_t = \varphi y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + u_t, \tag{4}$$

where

$$\varphi = \phi_{11} - \omega\phi_{21}, \ \beta_0 = \omega, \ \beta_1 = \phi_{12} - \omega\phi_{22}. \tag{5}$$

Note that u_t is uncorrelated with the regressor x_t and its lag by construction. (4) is ARDL(1,1) representation of y_t conditional on x_t, and the short-run coefficients φ, β_0, and β_1 can be directly estimated from (4) by least squares. Model (4) can also be written in the form of the error-correction model,

$$\Delta y_t = -(1 - \varphi)(y_{t-1} - \theta x_{t-1}) + \beta_0 \Delta x_t + u_t,$$

or as the following level relationship

$$y_t = \theta x_t + \alpha(L)\Delta x_t + \tilde{u}_t, \tag{6}$$

where $\tilde{u}_t = (1 - \varphi L)^{-1} u_t$, $\alpha(L) = \sum_{\ell=0}^{\infty} \alpha_\ell L^\ell$, $\alpha_\ell = \sum_{s=\ell+1}^{\infty} \delta_s$, for $\ell = 0, 1, 2, ...$, and

$$\delta(L) = \sum_{\ell=0}^{\infty} \delta_\ell L^\ell = (1 - \varphi L)^{-1}(\beta_0 + \beta_1 L).$$

The level coefficient, θ, is defined by

$$\theta = \frac{\beta_0 + \beta_1}{1 - \varphi}.$$

Note that if $\mathbf{z}_t$ is integrated of order one ($I(1)$ for short) then $(1, -\theta)'$ is the cointegrating vector and the level relation (6) is also cointegrating.

The level coefficient θ can still be motivated as the long-run outcome of a counterfactual exercise even if $\mathbf{z}_t$ is stationary . One possible counterfactual is to consider the effects of a permanent shock to the x_t process on y_t in the long run. Let

$$g_{yt} = \lim_{s \to \infty} E\left(y_{t+s} - \mu_{y,t+s} \middle| \mathcal{I}_{t-1}, e_{x,t+h} = \sigma_x, \text{ for } h = 0, 1, 2, ...\right),$$

and similarly

$$g_{xt} = \lim_{s \to \infty} E\left(x_{t+s} - \mu_{x,t+s} \middle| \mathcal{I}_{t-1}, e_{x,t+h} = \sigma_x, \text{ for } h = 0, 1, 2, ...\right),$$

where μ_{yt} and μ_{xt}, respectively, are the deterministic components of y_t and x_t (in the current illustrative example deterministic components are zero) and $\mathcal{I}_t$ is the set containing all information up to the period t. Using (1) and noting that $E\left(e_{yt} \middle| e_{xt}\right) = \omega e_{xt}$, we obtain $g_{yt} = g_y$, $g_{xt} = g_x$,[3]

$$\mathbf{g} = \begin{pmatrix} g_y \\ g_x \end{pmatrix} = (\mathbf{I}_2 - \mathbf{\Phi})^{-1} \begin{pmatrix} \omega \\ 1 \end{pmatrix} \sigma_x = \begin{pmatrix} -\frac{\omega + \phi_{12} - \omega\phi_{22}}{\phi_{11} + \phi_{22} - \phi_{11}\phi_{22} + \phi_{12}\phi_{21} - 1} \\ -\frac{\omega\phi_{21} - \phi_{11} + 1}{\phi_{11} + \phi_{22} - \phi_{11}\phi_{22} + \phi_{12}\phi_{21} - 1} \end{pmatrix} \sigma_x,$$

and

$$\frac{g_y}{g_x} = \frac{\omega + \phi_{12} - \omega\phi_{22}}{1 - (\phi_{11} - \omega\phi_{21})},$$

which upon using (5), yields, $g_y = \theta g_x$, namely the long-run impact of a permanent change in the mean of x on y is given by θ. Note that only in the special case when the reduced form errors are uncorrelated ($\omega = 0$), is the short-run coefficient β_0 in the ARDL model (4) equal to 0 and the long-run coefficient θ reduces to $\phi_{12}/(1 - \phi_{11})$. But, in general, when $\omega \neq 0$, the short-run coefficient β_0 is non-zero and contemporaneous values of the regressor should not be excluded from (4). In the stationary case with regressors not strictly exogenous, θ depends also on the parameters of the x_t process and the estimation of θ should therefore be based on (4).

An alternative way to show that θ is equal to the ratio g_y/g_x is to consider the ARDL representation (4) for the future period $t + s$, given the information at time $t - 1$. We first note that

$$y_{t+s} = \varphi y_{t+s-1} + \beta_0 x_{t+s} + \beta_1 x_{t+s-1} + u_{t+s},$$

and after taking the conditional expectation with respect to $\{\mathcal{I}_{t-1}, e_{x,t+h} = \sigma_x, \text{ for } h = 0, 1, 2, ...\}$,

---

[3]Note that, in the stationary case, $\sum_{\ell=0}^{\infty} \mathbf{\Phi}^\ell = (\mathbf{I} - \mathbf{\Phi})^{-1}$.

taking limits as $s \to \infty$, and noting that in the stationary case $g_{yt} = g_y$ and $g_{xt} = g_x$, we obtain

$$g_y = \varphi g_y + \beta_0 g_x + \beta_1 g_x,$$

and hence

$$\frac{g_y}{g_x} = \frac{\beta_0 + \beta_1}{1 - \varphi} = \theta,$$

as desired.

Regardless of whether the variables are integrated of order one or integrated of order zero or whether the regressors are exogenous or not, the level coefficient θ is well defined and can be consistently estimated. The rates of convergence and the asymptotic distributions of the ARDL estimates of θ are established in Pesaran and Shin (1999). See, in particular, their Theorem 3.3.

2.1 Two approaches to the estimation of long-run effects

Consider now the problem of estimation of long-run effects in heterogeneous dynamic panels with a multi-factor error structure. Let y_{it} be the dependent variable of the ith cross-section unit, $\mathbf{x}_{it}$ be the $k \times 1$ vector of unit-specific regressors, and consider the following panel $ARDL(p_{yi}, p_{xi})$ specification,

$$y_{it} = \sum_{\ell=1}^{p_{yi}} \varphi_{i\ell} y_{i,t-\ell} + \sum_{\ell=0}^{p_{xi}} \boldsymbol{\beta}'_{i\ell} \mathbf{x}_{i,t-\ell} + u_{it}, \tag{7}$$

$$u_{it} = \boldsymbol{\gamma}'_i \mathbf{f}_t + \varepsilon_{it}, \tag{8}$$

for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$, where $\mathbf{f}_t$ is an $m \times 1$ vector of unobserved common factors, and p_{yi} and p_{xi} are the lag orders chosen to be sufficiently long so that u_{it} is a serially uncorrelated process across all i. The vector of long-run coefficients is then given by

$$\boldsymbol{\theta}_i = \frac{\sum_{\ell=0}^{p_{xi}} \boldsymbol{\beta}_{i\ell}}{1 - \sum_{\ell=1}^{p_{yi}} \varphi_{i\ell}}. \tag{9}$$

There are two approaches to estimating the long-run coefficients. One approach, already considered in the literature, is to estimate the individual short-run coefficients $\{\varphi_{i\ell}\}$ and $\{\boldsymbol{\beta}_{i\ell}\}$ in the ARDL relation, (7), and then compute the estimates of long-run effects using formula (9) with the short-run coefficients replaced by their estimates $\{\hat{\varphi}_{i\ell}\}$ and $\{\hat{\boldsymbol{\beta}}_{i\ell}\}$. We shall refer to this approach as the "ARDL approach to the estimation of long-run effects". The advantage of this approach is that the estimates of short-run coefficients are also obtained.

But when the focus is on the long-run coefficients, $\boldsymbol{\theta}_i$ can be estimated *directly* without first estimating the short run coefficients. This is possible by observing that the ARDL model, (7), can be written as

$$y_{it} = \boldsymbol{\theta}_i \mathbf{x}_{it} + \boldsymbol{\alpha}_i'(L) \Delta \mathbf{x}_{it} + \tilde{u}_{it}, \tag{10}$$

where $\tilde{u}_{it} = \varphi(L)^{-1} u_{it}$, $\varphi_i(L) = 1 - \sum_{\ell=1}^{p_{yi}} \varphi_{i\ell} L^\ell$, $\boldsymbol{\theta}_i = \boldsymbol{\delta}_i(1)$, $\boldsymbol{\delta}_i(L) = \varphi_i^{-1}(L) \boldsymbol{\beta}_i(L) = \sum_{\ell=0}^{\infty} \boldsymbol{\delta}_{i\ell} L^\ell$, $\boldsymbol{\beta}_i(L) = \sum_{\ell=0}^{p_{xi}} \boldsymbol{\beta}_{i\ell} L^\ell$, and $\boldsymbol{\alpha}_i(L) = \sum_{\ell=0}^{\infty} \sum_{s=\ell+1}^{\infty} \boldsymbol{\delta}_s L^\ell$. We shall refer to the direct estimation of $\boldsymbol{\theta}_i$ based on the distributed lag representation (10) as the "distributed lag (DL) approach to the estimation of long-run effects". Under the usual assumptions on the roots of $\varphi_i(L)$ falling strictly outside the unit circle, the coefficients of $\boldsymbol{\alpha}_i(L)$ are exponentially decaying; and it is possible to show that, in the absence of feedback effects from lagged values of y_{it} onto the regressors $\mathbf{x}_{it}$, a consistent estimate of $\boldsymbol{\theta}_i$ can be obtained directly based on the least squares regression of y_{it} on $\mathbf{x}_{it}$ and $\{\Delta \mathbf{x}_{it-\ell}\}_{\ell=0}^{p}$, where the truncation lag order p is chosen appropriately as an increasing function of the sample size. But, when the feedback effects from the lagged values of the dependent variable to the regressors are present, $\tilde{u}_{it}$ will be correlated with $\mathbf{x}_{it}$ and the DL approach would no longer be consistent. Note that strict exogeneity is, however, not necessarily required for the consistency of the DL approach, since arbitrary correlations amongst the individual reduced form innovations in $\mathbf{e}_t$ are still allowed. After the individual estimates $\hat{\boldsymbol{\theta}}_i$ are obtained, either using ARDL or DL approach, they can then be averaged across i to obtain a consistent estimate of the average long-run effects, given by $\bar{\hat{\boldsymbol{\theta}}} = N^{-1} \Sigma_i^N \hat{\boldsymbol{\theta}}_i$.

2.2 Pros and cons of the two approaches to the estimation of long-run effects

Consider first the ARDL approach, where the estimates of long-run effects are computed based on the estimates of the short-run coefficients in (7). In the case where the unobserved common factors are serially uncorrelated and are also uncorrelated with the regressors, the long-run coefficients can be estimated consistently from the Ordinary Least Squares (OLS) estimates of the short-run coefficients, irrespective of whether the regressors are strictly exogenous or jointly determined with y_{it}, in the sense that $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it}')'$ follows a VAR model. The long-run estimates are also consistent irrespective of whether the underlying variables are $I(0)$ or $I(1)$. These robustness properties are clearly important in empirical research. However, the ARDL approach has also a number of drawbacks. The sampling uncertainty could be large especially when the speed of convergence towards the long-run relation is rather slow and the time dimension is not sufficiently long. This is readily apparent from (9) since even a small change to $1 - \sum_{\ell=1}^{p_{yi}} \hat{\varphi}_{i\ell}$ could have large impact on the estimates of

$\boldsymbol{\theta}_i$ when $\sum_{\ell=1}^{p_{yi}} \hat{\varphi}_{i\ell}$ is close to unity. In this respect, a correct specification of lag orders could be quite important for the performance of the ARDL estimates of $\boldsymbol{\theta}_i$. Underestimating the lag orders leads to inconsistent estimates, whilst overestimating the lag orders could result in loss of efficiency and low power when the ARDL long-run estimates are used for inference.

In the more general case when the unobserved common factors are correlated with the regressors then LS estimation of the ARDL model is no longer consistent and the effects of unobserved common factors need to be taken into account. There are so far two possible estimators developed in the literature for this case:[4] a principal-components based approach by Song (2013) who extends the interactive fixed effects estimator of Bai (2009) to the dynamic heterogeneous panels, and the dynamic common correlated effects mean group estimator suggested by Chudik and Pesaran (2014a). A recent overview of these methods is provided in Chudik and Pesaran (2014b). These estimators have (so far) been proposed only for stationary panels, and are subject to the small T bias of the ARDL approach discussed above. Bias correction techniques can also be used, but overall they do not seem to be effective when the speed of adjustment to the steady state is slow.[5]

The main merit of the DL approach proposed in this paper is its robustness along a number of important dimensions, and the fact that it tends to exhibit better small sample performance as compared to the panel ARDL estimates when the time dimension T is not very large. Specifically, (i) it is robust to the possibility of unit roots in regressors and/or factors, (ii) it is applicable irrespective of whether the short and/or long-run coefficients are heterogenous or homogeneous, (iii) it is robust to an arbitrary degree of serial correlation in ε_{it} and $\mathbf{f}_t$,[6] (vi) it does not require knowledge of the number of unobserved common factors under certain conditions, and (v) it continues to be valid under weak cross-section dependence in the idiosyncratic errors, ε_{it}. These robustness properties are very important considerations in applied research. In addition, the CS-DL approach does not require specifying the individual lag orders, p_{yi} and p_{xi}, and is robust to possible breaks in ε_{it}. The main drawback of the CS-DL approach, however, is that $\tilde{u}_{it} = \varphi(L)^{-1} u_{it}$ will be correlated with $\mathbf{x}_{it}$ when there are feedback effects from lagged values of y_{it} onto the regressors, $\mathbf{x}_{it}$. This correlation in turn introduces a bias even when N and T sufficiently large, and therefore the CS-DL estimation of the long-run effects is consistent only in the case when the feedback effects (or reverse causality) are not present. The second drawback is that the small sample performance is very good only when the eigenvalues of $\varphi(L)$ are not close to the unit circle. We will provide small

---

[4]Related is also the quasi maximum likelihood estimator for dynamic panels by Moon and Weidner (2010), but this estimators has been developed only for panels with homogeneous slope coefficients.

[5]Chudik and Pesaran (2014a) consider the application of two bias correction procedures to dynamic CCE type estimators, but find that they do not fully eliminate the bias.

[6]Note that $\boldsymbol{\theta}_i$ is identified even when ε_{it} is serially correlated.

sample evidence on the two approaches by means of Monte Carlo experiments in Section 4.

3 Cross-sectionally augmented distributed lag (CS-DL) approach to estimation of mean long-run coefficients

3.1 The ARDL panel data model

To simplify the exposition we consider the panel ARDL data model (7) with $p_{yi} = 1$ and $p_{xi} = 0$,

$$y_{it} = \varphi_i y_{i,t-1} + \boldsymbol{\beta}'_i \mathbf{x}_{it} + \boldsymbol{\gamma}'_i \mathbf{f}_t + \varepsilon_{it}. \tag{11}$$

To allow for correlation between the m unobserved factors, $\mathbf{f}_t$, and the k observed regressors, $\mathbf{x}_{it}$, we assume that the latter is generated according to the following factor model

$$\mathbf{x}_{it} = \boldsymbol{\Gamma}'_i \mathbf{f}_t + \mathbf{v}_{it}, \tag{12}$$

for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$, where $\boldsymbol{\Gamma}_i$ is $m \times k$ matrix of factor loadings, and $\mathbf{v}_{it}$ are the idiosyncratic components of $\mathbf{x}_{it}$ which are assumed to be distributed independently of the idiosyncratic errors, ε_{it}. The panel data model (11) and (12) is identical to the model considered by Pesaran (2006), with the exception that the lagged dependent variable is included in (11). We have also omitted observed common effects and deterministics (such as intercepts and time trends) from (11) to simplify the exposition. Introducing these terms and additional lags of the dependent variable and regressors is relatively straightforward.

We are interested in the estimation of the mean long-run coefficients $\boldsymbol{\theta} = E(\boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i$, $i = 1, 2, ..., N$ are the cross-section specific long-run coefficients defined by (9), which for $p_{yi} = 1$ and $p_{xi} = 0$ reduces to

$$\boldsymbol{\theta}_i = \frac{\boldsymbol{\beta}_i}{1 - \varphi_i}. \tag{13}$$

We postulate the following assumptions.

Assumption 1 *(Individual Specific Errors) Individual specific errors ε_{it} and $\mathbf{v}_{jt'}$ are independently distributed for all i, j, t and t'. ε_{it} follows a linear stationary process with absolute summable autocovariances (uniformly in i),*

$$\varepsilon_{it} = \sum_{\ell=0}^{\infty} \alpha_{\varepsilon i \ell} \zeta_{i,t-\ell}, \tag{14}$$

for $i = 1, 2, ..., N$, where the vector of innovations $\boldsymbol{\zeta}_t = (\zeta_{1t}, \zeta_{2t}, ..., \zeta_{Nt})'$ is spatially correlated

according to

$$\boldsymbol{\zeta}_t = \mathbf{R}\boldsymbol{\varsigma}_t,$$

in which the elements of $\boldsymbol{\varsigma}_t$ are independently and identically distributed (IID) with zero means, unit variances and finite fourth-order cumulants. Matrix $\mathbf{R}$ has bounded row and column matrix norms, namely $\|\mathbf{R}\|_\infty < K$ and $\|\mathbf{R}\|_1 < K$. In particular,

$$Var\left(\varepsilon_{it}\right) = \sum_{\ell=0}^{\infty} \alpha_{\varepsilon i\ell}^2 \sigma_{\zeta i}^2 = \sigma_i^2 \leq K < \infty, \tag{15}$$

for $i = 1, 2, ..., N$, where $\sigma_{\zeta i}^2 = Var\left(\zeta_{it}\right)$. $\mathbf{v}_{it}$ follows a linear stationary process with absolute summable autocovariances uniformly in i,

$$\mathbf{v}_{it} = \sum_{\ell=0}^{\infty} \mathbf{S}_{i\ell}\boldsymbol{\nu}_{i,t-\ell}, \tag{16}$$

for $i = 1, 2, ..., N$, where $\boldsymbol{\nu}_{it}$ is a $k \times 1$ vector of IID random variables, with mean zero, variance matrix $\mathbf{I}_k$, and finite fourth-order cumulants. In particular,

$$\|Var\left(\mathbf{v}_{it}\right)\| = \left\|\sum_{\ell=0}^{\infty} \mathbf{S}_{i\ell}\mathbf{S}_{i\ell}'\right\| = \|\boldsymbol{\Sigma}_i\| \leq K < \infty, \tag{17}$$

for $i = 1, 2, ..., N$, where $\|\mathbf{A}\|$ denotes the spectral norm of matrix $\mathbf{A}$.

Assumption 2 *(Common Factors) The $m \times 1$ vector of unobserved common factors, $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{mt})$, is covariance stationary with absolute summable autocovariances, distributed independently of $\varsigma_{it'}$ and $\mathbf{v}_{it'}$ for all i, t and t'. Fourth moments of $f_{\ell t}$, for $\ell = 1, 2, ..., m$, are bounded.*

Assumption 3 *(Factor Loadings) The factor loadings, $\boldsymbol{\gamma}_i$, and $\boldsymbol{\Gamma}_i$, are independently and identically distributed across i, and of the common factors $\mathbf{f}_t$, for all i and t, with fixed means $\boldsymbol{\gamma}$ and $\boldsymbol{\Gamma}$, respectively, and bounded second moments. In particular,*

$$\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_{\gamma i}, \; \boldsymbol{\eta}_{\gamma i} \sim IID\left(\underset{m \times 1}{\mathbf{0}}, \boldsymbol{\Omega}_\gamma\right), \; \text{for } i = 1, 2, ..., N,$$

and

$$vec\left(\boldsymbol{\Gamma}_i\right) = vec\left(\boldsymbol{\Gamma}\right) + \boldsymbol{\eta}_{\Gamma i}, \; \boldsymbol{\eta}_{\Gamma i} \sim IID\left(\underset{km \times 1}{\mathbf{0}}, \boldsymbol{\Omega}_\Gamma\right), \; \text{for } i = 1, 2, ..., N,$$

where $\boldsymbol{\Omega}_\gamma$ and $\boldsymbol{\Omega}_\Gamma$ are $m \times m$ and $k\,m \times k\,m$ symmetric nonnegative definite matrices, $\|\boldsymbol{\gamma}\| < K$, $\|\boldsymbol{\Omega}_\gamma\| < K$, $\|\boldsymbol{\Gamma}\| < K$, and $\|\boldsymbol{\Omega}_\Gamma\| < K$.

Assumption 4 *(Coefficients) The long-run coefficients, $\boldsymbol{\theta}_i$, defined in (13), follow the random coefficient model*

$$\boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{v}_i, \; \boldsymbol{v}_i \sim IID\left(\underset{k\times 1}{\mathbf{0}}, \boldsymbol{\Omega}_\theta\right), \; \text{for } i = 1, 2, ..., N, \tag{18}$$

where $\|\boldsymbol{\theta}\| < K$, $\|\boldsymbol{\Omega}_\theta\| < K$, $\boldsymbol{\Omega}_\theta$ is $k \times k$ symmetric nonnegative definite matrix, and the random deviations $\boldsymbol{v}_i$ are independently distributed of $\boldsymbol{\gamma}_j$, $\boldsymbol{\Gamma}_j$, ς_{jt}, $\mathbf{v}_{jt}$, and $\mathbf{f}_t$ for all i,j, and t. The coefficients of the lagged dependent variable, φ_i, are distributed with a support strictly inside the unit circle.

The polynomial $1 - \varphi_i L$ is invertible under Assumption 4, and multiplying (11) by $(1 - \varphi_i L)^{-1}$ we obtain

$$
\begin{aligned}
y_{it} &= (1 - \varphi_i L)^{-1} \boldsymbol{\beta}_i' \mathbf{x}_{it} + (1 - \varphi_i L)^{-1} \boldsymbol{\gamma}_i' \mathbf{f}_t + (1 - \varphi_i L)^{-1} \varepsilon_{it} \\
&= \boldsymbol{\theta}_i \mathbf{x}_{it} - \boldsymbol{\alpha}_i'(L) \Delta \mathbf{x}_{it} + \boldsymbol{\gamma}_i' \tilde{\mathbf{f}}_{it} + \tilde{\varepsilon}_{it}, \text{ for } i = 1, 2, ..., N,
\end{aligned} \tag{19}
$$

where $\Delta \mathbf{x}_{it} = \mathbf{x}_{it} - \mathbf{x}_{i,t-1}$, $\boldsymbol{\alpha}_i(L) = \sum_{\ell=0}^{\infty} \varphi_i^{\ell+1} (1 - \varphi_i)^{-1} \boldsymbol{\beta}_i L^\ell$, $\tilde{\mathbf{f}}_{it} = (1 - \varphi_i L)^{-1} \mathbf{f}_t$ and $\tilde{\varepsilon}_{it} = (1 - \varphi_i L)^{-1} \varepsilon_{it}$. The distributed lag specification in (19) does not include lagged values of the dependent variable, and as a result the CCE estimation procedure can be applied to (19) directly. The level regression of y_{it} on $\mathbf{x}_{it}$ is estimated by augmenting the individual regressions by differences of unit specific regressors $\mathbf{x}_{it}$ and their lags, in addition to the augmentation by the cross-section averages that take care of the effects of unobserved common factors. The CCE procedure continues to be applicable despite the fact that the errors, $\tilde{\varepsilon}_{it}$, are serially correlated. (see Pesaran (2006)).

Let $\mathbf{w} = (w_1, w_2, ..., w_N)'$ be an $N \times 1$ vector of weights that satisfies the following 'granularity' conditions

$$\|\mathbf{w}\| = O\left(N^{-\frac{1}{2}}\right), \tag{20}$$

$$\frac{w_i}{\|\mathbf{w}\|} = O\left(N^{-\frac{1}{2}}\right) \text{ uniformly in } i, \tag{21}$$

and the normalization condition

$$\sum_{i=1}^{N} w_i = 1. \tag{22}$$

Define the cross-section averages $\bar{\mathbf{z}}_{wt} = (y_{wt}, \bar{\mathbf{x}}_{wt}')' = \sum_{i=1}^{N} w_i \mathbf{z}_{it}$, and consider augmenting the regressions of y_{it} on $\mathbf{x}_{it}$ and the current and lagged values of $\Delta \mathbf{x}_{it}$, with the following set of cross-section averages, $\mathcal{S}_{Npt} = \bar{\mathbf{z}}_{wt} \cup \{\Delta \bar{\mathbf{x}}_{w,t-\ell}\}_{\ell=0}^{p}$. Cross-section averages approximate

the unobserved common factors arbitrarily well if

$$\vartheta_{fNp} = \mathbf{f}_t - E\left(\mathbf{f}_t | \mathcal{S}_{Npt}\right) \xrightarrow{p} \mathbf{0}, \tag{23}$$

uniformly in t, as N and $p \xrightarrow{j} \infty$. Sufficient conditions for result (23) to hold are given by Assumptions 1-4 and if the rank condition $rank\left(\mathbf{\Gamma}\right) = m$ holds. Different sets of cross section-averages could also be considered. For example, if the set of cross-section averages is defined as $\mathcal{S}_{Np_{\bar{z}}t} = \{\bar{\mathbf{z}}_{wt-\ell}\}_{\ell=0}^{p_{\bar{z}}}$, then the sufficient condition for (23) to hold under Assumption 1-4 would be the usual rank condition $rank\left(\mathbf{C}\right) = m$, where $\mathbf{C} = (\boldsymbol{\gamma}, \mathbf{\Gamma})$. Using covariates to enlarge the set of cross-section averages could also be considered, as in Chudik and Pesaran (2014a). Theses rank conditions can be relaxed in the case where $\boldsymbol{\gamma}_i$ and $\mathbf{\Gamma}_i$ are independently distributed.[7] In this case, the asymptotic variance of the CCE estimator does depend on the rank condition, nevertheless the CS-DL estimators are consistent and the proposed non-parametric estimators of the covariance matrix of the CS-DL estimators given below continue to be valid regardless of whether the rank condition holds.

More formally, let $\mathbf{y}_i = (y_{i,p+1}, y_{i,p+2}, ..., y_{i,T})'$, $\mathbf{X}_i = \left(\mathbf{x}_{i,p+1}, \mathbf{x}_{i,p+2}, ..., \mathbf{x}_{i,T}\right)'$, $\bar{\mathbf{Z}}_w = (\bar{\mathbf{z}}_{w,p+1}, \bar{\mathbf{z}}_{w,p+2}, ..., \bar{\mathbf{z}}_{w,T}$

$$\underset{(T-p)\times pk}{\Delta \mathbf{X}_{ip}} = \begin{pmatrix} \Delta \mathbf{x}'_{i,p+1} & \Delta \mathbf{x}'_{i,p} & \cdots & \Delta \mathbf{x}'_{i2} \\ \Delta \mathbf{x}'_{i,p+2} & \Delta \mathbf{x}'_{i,p+1} & \cdots & \Delta \mathbf{x}'_{i3} \\ \vdots & \vdots & & \vdots \\ \Delta \mathbf{x}'_{iT} & \Delta \mathbf{x}'_{i,T-1} & \cdots & \Delta \mathbf{x}'_{i,T-p+1} \end{pmatrix},$$

$\Delta \bar{\mathbf{X}}_{wp} = \sum_{i=1}^{N} w_i \Delta \mathbf{X}_{ip}$, $\mathbf{Q}_{wi} = \left(\bar{\mathbf{Z}}_w, \Delta \bar{\mathbf{X}}_{wp}, \Delta \mathbf{X}_{ip}\right)$, and define the projection matrix

$$\mathbf{M}_{qi} = \mathbf{I}_{T-p} - \mathbf{Q}_{wi}\left(\mathbf{Q}'_{wi}\mathbf{Q}_{wi}\right)^{+}\mathbf{Q}'_{wi}, \tag{24}$$

for $i = 1, 2, ..., N$, where $p = p\left(T\right)$ is a chosen non-decreasing truncation lag function such that $0 \leq p < T$, and $\mathbf{A}^{+}$ is the Moore-Penrose pseudoinverse of $\mathbf{A}$. We use the Moore-Penrose pseudoinverse as opposed to standard inverse in (24) because the column vectors of $\mathbf{Q}_{wi}$ could be asymptotically (as $N \to \infty$) linearly dependent.

The CS-DL mean group estimator of the long-run coefficients is given by

$$\widehat{\boldsymbol{\theta}}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\theta}}_i, \tag{25}$$

---

[7]Correlation of $\boldsymbol{\gamma}_i$ and $\mathbf{\Gamma}_i$ could introduce a bias in the rank deficient case, as noted by Sarafidis and Wansbeek (2012).

where

$$\widehat{\boldsymbol{\theta}}_i = \left(\mathbf{X}_i'\mathbf{M}_{qi}\mathbf{X}_i'\right)^{-1}\mathbf{X}_i'\mathbf{M}_{qi}\mathbf{y}_i. \tag{26}$$

The CS-DL pooled estimator of the long-run coefficients is

$$\widehat{\boldsymbol{\theta}}_P = \left(\sum_{i=1}^{N} w_i\mathbf{X}_i'\mathbf{M}_{qi}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N} w_i\mathbf{X}_i'\mathbf{M}_{qi}\mathbf{y}_i. \tag{27}$$

Estimators $\widehat{\boldsymbol{\theta}}_{MG}$ and $\widehat{\boldsymbol{\theta}}_P$ differ from the mean group and pooled CCE estimator developed in Pesaran (2006), which only allows for the inclusion of a fixed number of regressors, whilst the CS-DL type estimators include p_T lags of $\Delta\mathbf{x}_{it}$ and their cross-section averages, where p_T increases with T, albeit at a slower rate.

In addition to Assumptions 1-4 above, we shall also require the following assumption to hold. Assumption 5 below ensures that $\widehat{\boldsymbol{\theta}}_{MG}$ and $\widehat{\boldsymbol{\theta}}_P$ and their asymptotic distributions are well defined.

Assumption 5 *(a) The matrix $\lim_{N,T,p\xrightarrow{j}\infty}\sum_{i=1}^{N} w_i\boldsymbol{\Sigma}_i = \boldsymbol{\Psi}^*$ exists and is nonsingular, and $\sup_{i,p}\left\|\boldsymbol{\Sigma}_i^{-1}\right\| < K$, where $\boldsymbol{\Sigma}_i = p\lim T^{-1}\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{X}_i$, and $\mathbf{M}_{hi}$ is defined in (A.3).*

(b) Denote the t-th row of matrix $\widetilde{\mathbf{X}}_i = \mathbf{M}_{hi}\mathbf{X}_i$ by $\widetilde{\mathbf{x}}_{it}' = (\widetilde{x}_{i1t}, \widetilde{x}_{i2t}, ..., \widetilde{x}_{ikt})$. The individual elements of $\widetilde{\mathbf{x}}_{it}$ have uniformly bounded fourth moments, namely there exists a positive constant $K < \infty$ such that $E\left(\widetilde{x}_{ist}^4\right) < K$, for any $t = 1, 2, ..., T$, $i = 1, 2, ..., N$ and $s = 1, 2, ..., k$.

(c) There exists T_0 such that for all $T \geq T_0$, $\left(\sum_{i=1}^{N} w_i\mathbf{X}_i'\mathbf{M}_{qi}\mathbf{X}_i/T\right)^{-1}$ exists.

(d) There exists N_0,T_0 and $p_0 = p(T_0)$ such that for all $N \geq N_0$, $T \geq T_0$ and $p(T) \geq p(T_0)$, the $k \times k$ matrices $(\mathbf{X}_i'\mathbf{M}_{qi}\mathbf{X}_i/T)^{-1}$ exist for all i, uniformly.

Our main findings are summarized in the following theorems.

Theorem 1 (Asymptotic distribution of $\widehat{\boldsymbol{\theta}}_{MG}$) *Suppose y_{it}, for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ is given by the panel data model (11)-(12), Assumptions 1-5 hold, and $(N, T, p(T)) \xrightarrow{j} \infty$ such that $\sqrt{N}p(T)\rho^p \to 0$, for any constant $0 < \rho < 1$ and $p(T)^3/T \to \varkappa$, $0 < \varkappa < \infty$. Then, if $rank(\boldsymbol{\Gamma}) = m$ we have*

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{MG} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(0, \boldsymbol{\Omega}_\theta\right), \tag{28}$$

where $\boldsymbol{\Omega}_\theta = Var(\boldsymbol{\theta}_i)$ and $\widehat{\boldsymbol{\theta}}_{MG}$ is given by (25). If $rank(\boldsymbol{\Gamma}) \neq m$ and $\boldsymbol{\gamma}_i$ is independently distributed of $\boldsymbol{\Gamma}_i$, we have

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{MG} - \boldsymbol{\theta}\right) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{MG}), \tag{29}$$

where

$$\boldsymbol{\Sigma}_{MG} = \boldsymbol{\Omega}_\theta + \lim_{p,N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i^{-1}\mathbf{Q}_{if}\boldsymbol{\Omega}_\gamma\mathbf{Q}_{if}'\boldsymbol{\Sigma}_i^{-1}\right], \tag{30}$$

in which $\boldsymbol{\Omega}_\gamma = Var(\boldsymbol{\gamma}_i)$, $\boldsymbol{\Sigma}_i = p\lim_{T\to\infty} T^{-1}\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{X}_i$ and $\mathbf{Q}_{if} = p\lim_{T\to\infty} T^{-1}\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}$. In both cases, the asymptotic variance of $\widehat{\boldsymbol{\theta}}_{MG}$ can be consistently estimated nonparametrically by

$$\widehat{\boldsymbol{\Sigma}}_{MG} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}}_{MG}\right)\left(\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}}_{MG}\right)'. \tag{31}$$

Theorem 2 (Asymptotic distribution of $\widehat{\boldsymbol{\theta}}_P$) *Suppose y_{it}, for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ are generated by the panel data model (11)-(12), Assumptions 1-5 hold, and $(N, T, p(T)) \xrightarrow{j} \infty$, such that $\sqrt{N}p(T)\rho^p \to 0$, for any constant $0 < \rho < 1$ and $p(T)^3/T \to \varkappa$, $0 < \varkappa < \infty$. Then, if $\boldsymbol{\gamma}_i$ is independently distributed of $\boldsymbol{\Gamma}_i$, we have*

$$\left(\sum_{i=1}^{N} w_i^2\right)^{-1/2}\left(\widehat{\boldsymbol{\theta}}_P - \boldsymbol{\theta}\right) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_P), \tag{32}$$

where $\widehat{\boldsymbol{\theta}}_P$ is given by (27),

$$\boldsymbol{\Sigma}_P = \boldsymbol{\Psi}^{*-1}\mathbf{R}^*\boldsymbol{\Psi}^{*-1}, \ \boldsymbol{\Psi}^* = \lim_{N\to\infty}\sum_{i=1}^{N} w_i\boldsymbol{\Sigma}_i, \tag{33}$$

$$\mathbf{R}^* = \mathbf{R}_\theta^* + \mathbf{R}_\gamma^*, \ \mathbf{R}_\theta^* = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\widetilde{w}_i^2\boldsymbol{\Sigma}_i\boldsymbol{\Omega}_\theta\boldsymbol{\Sigma}_i, \ \mathbf{R}_\gamma^* = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\widetilde{w}_i^2\mathbf{Q}_{if}\boldsymbol{\Omega}_\gamma\mathbf{Q}_{if}',$$

$\boldsymbol{\Omega}_\theta = Var(\boldsymbol{\theta}_i)$, $\boldsymbol{\Omega}_\gamma = Var(\boldsymbol{\gamma}_i)$, $\boldsymbol{\Sigma}_i = p\lim T^{-1}\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{X}_i$, $\mathbf{Q}_{if} = p\lim T^{-1}\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}$, and $\widetilde{w}_i = \sqrt{N}w_i\left(\sum_{i=1}^{N} w_i^2\right)^{-1/2}$. If $rank(\boldsymbol{\Gamma}) = m$, then $\boldsymbol{\gamma}_i$ is no longer required to be independently distributed of $\boldsymbol{\Gamma}_i$ and (32) continues to hold with $\boldsymbol{\Sigma}_P = \boldsymbol{\Psi}^{*-1}\mathbf{R}_\theta^*\boldsymbol{\Psi}^{*-1}$. In both cases, $\boldsymbol{\Sigma}_P$ can be consistently estimated by $\widehat{\boldsymbol{\Sigma}}_P$ defined by equation (A.25) in the Appendix.

Theorems 1-2 establish asymptotic distributions of $\widehat{\boldsymbol{\theta}}_{MG}$ and $\widehat{\boldsymbol{\theta}}_P$ under slope heterogeneity. These theorems distinguish between cases where the rank condition that ensures (23) is satisfied or not. Under the former, unobserved common factors can be approximated

by cross-section averages when N is large and regardless of whether $\boldsymbol{\gamma}_i$ is correlated with $\boldsymbol{\Gamma}_i$, $\widehat{\boldsymbol{\theta}}_{MG}$ and $\widehat{\boldsymbol{\theta}}_P$ are consistent and asymptotically normal. In the latter case, where the unobserved common factors cannot be approximated by cross-section averages when N is large, then so long as $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Gamma}_i$ are independently distributed, both $\widehat{\boldsymbol{\theta}}_{MG}$ and $\widehat{\boldsymbol{\theta}}_P$ continue to be consistent and asymptotically normal, but the asymptotic variance depends also on unobserved common factors and their loadings. In both (full rank or rank deficient) cases, the asymptotic variance of the CS-DL estimators can be estimated consistently using the same non-parametric formulae as in the full rank case.

There are several departures from the assumptions of these theorems that might be of interest in applied work, such as the consequences of breaks in the error processes, ε_{it}, possibility of unit roots in factors and/or regressor specific components, and situations where some or all coefficients are homogeneous over the cross-section units. These theoretical extensions are outside the scope of the present paper but we investigate the robustness of the proposed CS-DL estimator to such departures by means of Monte Carlo simulations in the next section.

4   Monte Carlo experiments

This section investigates small sample properties of the CS-DL estimators and compares them with the estimates obtained from the panel ARDL approach using the dynamic CCEMG estimator of the short-run coefficients advanced in Chudik and Pesaran (2014a), which we denote by CS-ARDL. First, we present results from the baseline experiments with heterogeneous slopes (long- and short-run coefficients), and then we document small sample performance of the alternative estimators under various deviations from the baseline experiments, including robustness of the estimators to the introduction of unit roots in the regressors or factors, possible breaks in the idiosyncratic error processes, and the consequences of feedback effects from lagged values of y_{it} onto $\mathbf{x}_{it}$. Second, we investigate whether it is possible to improve on the estimation of short-run coefficients, provided the model is correctly specified, by imposing CS-DL estimates of the long-run coefficients.

We start with a brief summary of the estimation methods and a description of the data generating processes (DGP). Then we present findings on the estimation of the mean long-run coefficient and on the extent to which estimates of the short-run coefficients can be improved by using the CS-DL estimators of the long-run effects.

4.1 Estimation methods

The CS-DL estimators are based on the following auxiliary regressions:

$$y_{it} = c_{yi} + \boldsymbol{\theta}'_i \mathbf{x}_{it} + \sum_{\ell=0}^{p-1} \delta_{i\ell} \Delta x_{i,t-\ell} + \sum_{\ell=0}^{p_{\bar{y}}} \omega_{y,i\ell} \bar{y}_{t-\ell} + \sum_{\ell=0}^{p_{\bar{x}}} \boldsymbol{\omega}'_{x,i\ell} \bar{\mathbf{x}}_{t-\ell} + e_{it}, \tag{34}$$

where $\bar{\mathbf{x}}_t = N^{-1} \sum_{i=1}^N \mathbf{x}_{it}$, $\bar{y}_t = N^{-1} \sum_{i=1}^N y_{it}$, $p_{\bar{x}}$ is set equal to the integer part of $T^{1/3}$, denoted as $\left[T^{1/3} \right]$, $p = p_{\bar{x}}$ and $p_{\bar{y}}$ is set to 0. We consider both CS-DL mean group and pooled estimators based on (34).

The CS-ARDL estimator is based on the following regressions:

$$y_{it} = c^*_{yi} + \sum_{\ell=1}^{p_y} \varphi_{i\ell} y_{i,t-\ell} + \sum_{\ell=0}^{p_x} \boldsymbol{\beta}'_{i\ell} \mathbf{x}_{i,t-\ell} + \sum_{\ell=0}^{p_{\bar{z}}} \boldsymbol{\psi}'_{i\ell} \bar{\mathbf{z}}_{t-\ell} + e^*_{it}, \tag{35}$$

where $\bar{\mathbf{z}}_t = (\bar{y}_t, \bar{\mathbf{x}}'_t)'$, $p_{\bar{z}} = \left[T^{1/3} \right]$ and two options for the remaining lag orders are considered: ARDL(2,1) specification, $p_y = 2$ and $p_x = 1$, and ARDL(1,0) specification, $p_y = 1$ and $p_x = 0$. The CS-ARDL estimates of individual mean level coefficient are then given by

$$\hat{\boldsymbol{\theta}}_{CS-ARDL,i} = \frac{\sum_{\ell=0}^{p_x} \hat{\boldsymbol{\beta}}_{i\ell}}{1 - \sum_{\ell=1}^{p_y} \hat{\varphi}_{i\ell}}, \tag{36}$$

where the estimates of short-run coefficients $(\hat{\varphi}_{i\ell}, \hat{\boldsymbol{\beta}}_{i\ell})$ are based on (35). The mean long-run effects are estimated as $N^{-1} \sum_{i=1}^N \hat{\boldsymbol{\theta}}_{CS-ARDL,i}$ and the inference is based on the usual non-parametric estimator of asymptotic variance of the mean group estimator.

4.2 Data generating process

The dependent variable and regressors are generated using the following ARDL(2,1) panel data model with factor error structure,

$$y_{it} = c_{yi} + \varphi_{i1} y_{i,t-1} + \varphi_{i2} y_{i,t-2} + \beta_{i0} x_{it} + \beta_{i1} x_{i,t-1} + u_{it}, \ u_{it} = \boldsymbol{\gamma}'_i \mathbf{f}_t + \varepsilon_{it}, \tag{37}$$

and

$$x_{it} = c_{xi} + \kappa_{yi} y_{i,t-1} + \boldsymbol{\gamma}'_{xi} \mathbf{f}_t + v_{it}. \tag{38}$$

We generate $y_{it}, \mathbf{x}_{it}$ for $i = 1, 2, ..., N$, and $t = -99, ..., 0, 1, 2, ..., T$ with the starting values $y_{i,-101} = y_{i,-100} = 0$, and discard the first 100 observations ($t = -99, -48, ..., 0$) to reduce the effects of the initial values on the outcomes. The fixed effects are generated as $c_{iy} \sim IIDN(1,1)$, and $c_{xi} = c_{yi} + \varsigma_{c_x i}$, where $\varsigma_{c_x i} \sim IIDN(0,1)$, thus allowing for dependence

between x_{it} and c_{yi}.

We consider three cases depending on the heterogeneity/homogeneity of the slopes:

- (heterogeneous slopes - baseline) $\varphi_{i1} = (1 + \varkappa_{\varphi i}) \eta_{\varphi i}$, $\varphi_{i2} = -\varkappa_{\varphi i} \eta_{\varphi i}$, $\varkappa_{\varphi i} \sim IIDU(0.2, 0.3)$, $\eta_{\varphi i} \sim IIDU(0, \varphi_{\max})$. The long-run coefficients are generated as $\theta_i \sim IIDN(1, 0.2^2)$ and the regression coefficient are generated as $\beta_{i0} = \varkappa_{\beta i} \eta_{\beta i}$, $\beta_{i1} = (1 - \varkappa_{\beta i}) \eta_{\beta i}$, where $\eta_{\beta i} = \theta_i / (1 - \varphi_{i1} - \varphi_{i2})$ and $\varkappa_{\beta i} \sim IIDU(0, 1)$.

- (homogeneous long-run, heterogenous short-run slopes) $\theta_i = 1$ for all i and the remaining coefficients $(\varphi_{i1}, \varphi_{i2}, \beta_{i0}, \beta_{i1})$ are generated as in the previous fully heterogeneous case.

- (homogeneous long- and short-run slopes) $\varphi_{i1} = 1.15\varphi_{\max}/2$, $\varphi_{i2} = -0.15\varphi_{\max}/2$, $\theta_i = 1$, and $\beta_{i0} = \beta_{i1} = 0.5/(1 - \varphi_{\max}/2)$.

We also consider the case of ARDL(1,0) panel model by setting $\varkappa_{\varphi i} = 0$ and $\varkappa_{\beta i} = 1$ for all i, which gives $\varphi_{i2} = \beta_{i1} = 0$ for all i. We consider three values for $\varphi_{\max} = 0.6$, 0.8 or 0.9.

The unobserved common factors in $\mathbf{f}_t$ and the unit-specific components, v_{it}, are generated as independent AR(1) processes:

$$f_{t\ell} = \rho_{f\ell} f_{t-1,\ell} + \varsigma_{ft\ell}, \quad \varsigma_{ft\ell} \sim IIDN\left(0, \sigma^2_{\varsigma f\ell}\right), \tag{39}$$

$$v_{it} = \rho_{xi} v_{i,t-1} + \nu_{it}, \quad \varsigma_{xit} \sim IIDN\left(0, \sigma^2_{\nu i}\right), \tag{40}$$

for $i = 1, 2, ..., N$, $\ell = 1, 2, .., m$, and for $t = -99, ..., 0, 1, 2, ..., T$ with the starting values $f_{\ell, -100} = 0$, and $v_{i, -100} = 0$. The first 100 time observations ($t = -99, -48, ..., 0$) are discarded. We consider three possibilities for the AR(1) coefficients $\rho_{f\ell}$ and ρ_{xi}:

- (stationary baseline) $\rho_{xi} \sim IIDU[0.0.95]$, $\sigma^2_{\nu i} = 1 - \rho^2_{xi}$ for all i; $\rho_{f\ell} = 0.6$, and $\sigma^2_{\varsigma f\ell} = 1 - \rho^2_{f\ell}$ for $\ell = 1, 2, ..., m$.

- (nonstationary factors) $\rho_{xi} \sim IIDU[0.0.95]$, $\sigma^2_{\nu i} = 1 - \rho^2_{xi}$ for all i; and $\rho_{f\ell} = 1$, $\sigma^2_{\varsigma f\ell} = 0.1^2$ for $\ell = 1, 2, ..., m$.

- (nonstationary regressors and stationary factors) $\rho_{xi} = 1$, $\sigma^2_{\nu i} = 0.1^2$ for all i; and $\rho_{f\ell} = 0.6$, $\sigma^2_{\varsigma f\ell} = 1 - \rho^2_{f\ell}$, for $\ell = 1, 2, ..., m$.

We consider also two options for the feedback coefficients κ_{yi}: no feedback effects, $\kappa_{yi} = 0$ for all i, and with feedback effects, $\kappa_{yi} \sim IIDU(0, 0.2)$.

Factor loadings are generated as

$$\gamma_{i\ell} \sim IIDN\left(\gamma_\ell, 0.2^2\right) \text{ and } \gamma_{xi\ell} \sim IIDN\left(\gamma_{x\ell}, 0.2^2\right),$$

for $\ell = 1, 2, .., m$, and $i = 1, 2, ..., N$. Also, without loss of generality, the means of factor loadings are calibrated so that $Var(\gamma_i' \mathbf{f}_t) = Var(\gamma_{xi}' \mathbf{f}_t) = 1$ in the stationary case. We set $\gamma_\ell = \sqrt{b_\gamma}$, and $\gamma_{x\ell} = \sqrt{\ell b_x}$, for $\ell = 1, 2, ..., m$, where $b_\gamma = 1/m - 0.2^2$, and $b_x = 2/\left[m\left(m+1\right)\right] - 2/\left(m+1\right)0.2^2$. This ensures that the contribution of the unobserved factors to the variance of y_{it} does not rise with m in the stationary case. We consider $m = 2$ or 3 unobserved common factors.

Finally, the idiosyncratic errors, ε_{it}, are generated to be heteroskedastic, weakly cross-sectionally dependent and serially correlated. Specifically,

$$\varepsilon_{it} = \rho_{\varepsilon i}\varepsilon_{i,t-1} + \zeta_{it}, \tag{41}$$

where $\boldsymbol{\zeta}_t = (\zeta_{1t}, \zeta_{2t}, ..., \zeta_{Nt})'$ are generated using the following spatial autoregressive model (SAR),

$$\boldsymbol{\zeta}_t = a_\epsilon \mathbf{S}_\epsilon \boldsymbol{\zeta}_t + \boldsymbol{\varsigma}_t, \tag{42}$$

in which the elements of $\boldsymbol{\varsigma}_t$ are drawn as $IIDN\left[0, \frac{1}{2}\sigma_i^2\left(1 - \rho_{\varepsilon i}^2\right)\right]$, with σ_i^2 obtained as independent draws from $\chi^2(2)$ distribution,

$$\mathbf{S}_\epsilon = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & & 0 \\ 0 & \frac{1}{2} & 0 & \ddots & & \vdots \\ 0 & 0 & \ddots & \ddots & \frac{1}{2} & 0 \\ \vdots & & & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix},$$

and the spatial autoregressive parameter is set to $a_\epsilon = 0.6$. Note that $\{\epsilon_{it}\}$ is cross-sectionally weakly dependent for $|a_\epsilon| < 1$. We consider $\rho_{\varepsilon i} = 0$ for all i or $\rho_{\varepsilon i} \sim IIDU\left(0, 0.8\right)$. We also consider the possibility of breaks in ε_{it} by generating for each i random break points $b_i \in \{1, 2, ..T\}$ and

$$\begin{aligned} \varepsilon_{it} &= \rho_{\varepsilon i}^a \varepsilon_{i,t-1} + \zeta_{it}, \text{ for } t = 1, 2, ..., b_i \\ \varepsilon_{it} &= \rho_{\varepsilon i}^b \varepsilon_{i,t-1} + \zeta_{it}, \text{ for } t = b_i + 1, b_i + 2, ..., T, \end{aligned}$$

where $\rho_{\varepsilon i}^a, \rho_{\varepsilon i}^b \sim IIDU\left(0, 0.8\right)$, and $\boldsymbol{\zeta}_t = (\zeta_{1t}, \zeta_{2t}, ..., \zeta_{Nt})'$ is generated using SAR model (42)

with $\varsigma_{it} \sim IIDN \left[0, \frac{1}{2}\sigma_i^2 \left(1 - \rho_{\varepsilon i}^{a2}\right)\right]$.

The above DGP is more general than the other DGPs used in MC experiments in the literature and allows the factors and regressors to be correlated and persistent. The above DGPs also include models with unit roots, breaks in the error processes, and allows for correlated fixed effects. To summarize, we consider the following cases:

1. (3 options for heterogeneity of coefficients) heterogeneous baseline, homogeneous long-run with heterogeneous short-run, and both long-and short-run homogeneous,

2. (2 options for lags) ARDL(2,1) baseline, and ARDL(1,0) model where $\varkappa_{\varphi i} = 0$ and $\varkappa_{\beta i} = 1$ for all i, which gives $\varphi_{i2} = \beta_{i1} = 0$ for all i.

3. (3 options for $\varphi_{\max}$) $\varphi_{\max} = 0.6$ (baseline), 0.8, or 0.9

4. (3 options for the persistence of factors and regressors) stationary baseline, I(1) factors, or I(1) regressor specific components v_{it},

5. (2 options for the number of factors) full rank case baseline $m = 2$, or rank deficient case $m = 3$,

6. (3 options for the persistence of idiosyncratic errors) serially uncorrelated baseline $\rho_{\varepsilon i} = 0$, $\rho_{\varepsilon i} \sim IIDU(0, 0.8)$, or breaks in the error process.

7. (2 options for feedback effects) $\kappa_{yi} = 0$ for all i (baseline), or $\kappa_{yi} \sim IIDU(0, 0.2)$.

Due to the large number of possible cases (648 in total), we only consider baseline experiments and various departures from the baseline. We consider the following combinations of sample sizes: $N, T \in \{30, 50, 100, 150, 200\}$, and set the number of replications to $R = 2,000$, in the case of all experiments.

4.3 Monte Carlo findings on the estimation of mean long-run coefficients

The results for the baseline DGP are summarized in Table 1. This table shows that both CS-DL estimators (MG and pooled) perform well in the baseline experiments. This table also shows that the CS-ARDL approach does not perform well when T is not large (<100) due to the small sample problems arising when $\sum_{\ell=1}^{p_y} \hat{\varphi}_{i\ell}$ is close to unity. Also, CS-ARDL estimates that are based on misspecified lag orders are inconsistent, as to be expected. In

contrast, the consistency of the CS-DL estimators does not depend on knowing the correct lag specifications of the underlying ARDL model.

Next, we investigate robustness of the results to different assumptions regarding slope heterogeneity. Table 2 presents findings for the experiment that depart from the baseline DGP by assuming homogeneous long-run slopes, while allowing the short-run slopes to be heterogeneous. Table 3 gives the results when both long- and short-run slopes are homogeneous. These results show that the CS-DL estimators continue to have good size and power properties in all cases.

Experiments based on the ARDL(1,0) specification (as the DGP) are summarized in Table 4. CS-DL estimators continue to perform well, showing their robustness to the underlying ARDL specification.

The effects of increasing the value of $\varphi_{\max}$ on the properties of the various estimators are summarized in Tables 5 (for $\varphi_{\max} = 0.8$) and 6 (for $\varphi_{\max} = 0.9$). Small sample performance of the CS-DL estimators deteriorates as $\varphi_{\max}$ moves closer to unity, as to be expected. Tables 5-6 show that the performance deteriorates substantially for values of $\varphi_{\max}$ close to unity, due to the bias that results from the truncation of lags for the first differences of regressors. It can take a large lag order for the truncation bias to be negligible when the largest eigenvalue of the dynamic specification (given by the lags of the dependent variable) is close to one. We see quite a substantial bias when $\varphi_{\max} = 0.9$. Therefore, it is important that the CS-DL approach is used when the speed of convergence towards equilibrium is not too slow and/or T is sufficiently large so that biases arising from the approximation of dynamics by distributed lag functions can be controlled.

The robustness of the results to the number of unobserved factors (m) is investigated in Table 7. This table provides a summary in the case of $m = 3$ factors, which represents the rank deficient case. It is interesting to note that despite the failure of the rank condition, the CS-DL estimators continue to perform well (the results are almost unchanged as compared with those in Table 1), while the CS-ARDL estimates are affected by two types of biases (the time series bias and the bias due to rank deficiency) that operate in opposite directions.

Consider now the robustness of the results to the presence of unit roots in the unobserved factors (Table 8) or in the regressors (Table 9). As can be seen the CS-DL estimators continue to perform well when factors contain unit roots. Table 9, on the other hand, shows large RMSE and low power for $T = 30$ and 50, when the idiosyncratic errors have unit roots. But, interestingly enough, the reported size is correct and biases are very small for all sample sizes.

The robustness of the CS-DL estimators to the patterns of residual serial correlation is investigated in Table 10, whilst Table 11 present results on the robustness of CS-DL

estimators to possible breaks in the error processes. As can be seen, and as predicted by the theory, the CS-DL estimators are robust to both of these departures from the baseline scenario, whereas the CS-ARDL approach is not. Recall, that CS-ARDL approach requires that the lag orders are correctly specified, and does not allow for residual serial correlation and/or breaks in the error processes, whilst CS-DL does.

Last but not least, the consequences of feedback effects from y_{it} to the regressors, x_{it}, is documented in Table 12. This table shows that the CS-ARDL approach is consistent regardless of the feedback effects, provided that the lag orders are correctly specified, again as predicted by the theory. But a satisfactory performance (in terms of bias and size of the test) for the CS-ARDL approach requires T to be sufficiently large. On the other hand, in the presence of feedbacks, the CS-DL estimators are inconsistent and show positive bias even for T sufficiently large. But the bias due to feedback effects seem to be quite small; between -0.02 and 0.06, and the CS-DL estimators tend to outperform the CS-ARDL estimators when $T < 100$, even when the underlying ARDL model is correctly specified.

4.4 Monte Carlo findings on the improvement in estimation of short-run coefficients

As a final exercise, we consider if it is possible to improve on the estimation of short-run coefficients by imposing the CS-DL estimates of the long run, before estimating the short-run coefficients. We consider the experiment that departs from the baseline model by assuming a homogeneous long-run coefficient, whilst all the short-run slopes are allowed to be heterogeneous, and use the ARDL(1,0) as the data generating process. More specifically, we impose the CS-DL pooled estimator of the long-run coefficient, $\hat{\theta}_P$, when estimating the short-run coefficients using the CS-ARDL approach. In particular, we estimate the following unit-specific regressions,

$$\Delta y_{it} = c_{yi}^* + \lambda_i \left(y_{i,t-1} - \hat{\theta}_P x_{it} \right) + \sum_{\ell=0}^{p_{\bar{z}}} \boldsymbol{\delta}_{i\ell}' \bar{\mathbf{z}}_{t-\ell} + \varepsilon_{it}^*, \tag{43}$$

for $i = 1, 2, ..., N$, and the resulting mean group estimator of $E(\varphi_{i1}) = 1 + E(\lambda_i)$ is denoted by

$$\tilde{\varphi}_{1,MG} = \frac{1}{N} \sum_{i=1}^{N} \tilde{\varphi}_{i1}, \ \tilde{\varphi}_{i1} = 1 - \tilde{\lambda}_i,$$

where $\tilde{\lambda}_i$ is the least square estimate of λ_i based on (43). The results of these experiments are summarized in Table 13. Imposing the CS-DL pooled estimator of the long-run coefficient

improves the small sample properties of the short-run estimates substantially, about 80-90% reduction of the difference between the RMSE of the infeasible CS-ARDL estimator and the RMSE of the unconstrained estimator, when $T = 30$.

5   Concluding remarks

Panel data estimation of long-run effects is an important task in economics. This often requires a large time dimension for a panel data model featuring slope heterogeneity, lagged dependent variables, and cross-sectionally correlated innovations. This paper proposes a cross-sectionally augmented distributed lag (CS-DL) approach to the estimation of long-run effects as a complementary method to cross-sectionally augmented ARDL specifications. Based on a series of Monte Carlo simulations, we show the robustness of panel CS-ARDL estimates to endogeneity problem. We also show that the CS-DL estimators are robust to residual serial correlation, breaks in error processes and dynamic misspecifications. However, unlike the CS-ARDL approach, the CS-DL procedure could be subject to simultaneity bias. Nevertheless, the extensive Monte Carlo experiments reported in the paper suggest that the endogeneity bias of the CS-DL approach is more than compensated for by its better small sample performance as compared to the CS-ARDL procedure when the time dimension is not very large. CS-ARDL seems to dominate CS-DL only if the time dimension is sufficiently large and the underlying ARDL model is correctly specified.

Table 1: Monte Carlo Estimates of Bias, RMSE, Size and Power for Estimation of LR Coefficient (θ) in the Baseline Experiment

DGP is an ARDL(2,1) model with heterogeneous coefficients, $\varphi_{\max} = 0.6$, stationary regressors, $m = 2$ factors, no feedback effects and $\rho_{\varepsilon i} = 0$.

(N,T)	Bias (×100)					RMSE (×100)					Size (5% level, $H_0:\theta=1$)					Power (5% level, $H_1:\theta=1.2$)				
	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
CS-DL mean group ($p_{\bar{y}}=0$, $p_{\bar{x}}=\left[T^{1/3}\right]$ and $p=p_{\bar{x}}-1$)																				
30	-0.65	-0.49	0.04	-0.11	-0.15	16.88	11.24	8.55	7.34	6.44	6.35	6.15	7.75	5.80	6.40	28.30	50.20	70.30	80.10	87.15
50	-1.12	-1.00	-0.34	-0.12	-0.03	13.19	8.83	6.33	5.82	4.92	5.90	6.15	5.25	6.45	5.20	39.45	70.20	89.10	93.45	97.80
100	-1.32	-0.92	-0.09	-0.11	0.15	9.66	6.25	4.49	4.03	3.56	5.95	6.30	5.55	4.50	5.45	62.95	92.45	99.50	99.75	100.00
150	-1.19	-0.96	-0.11	0.16	-0.05	7.91	5.24	3.78	3.38	2.94	5.90	6.50	5.90	6.25	5.75	79.45	98.20	99.85	100.00	100.00
200	-1.06	-0.75	-0.24	-0.07	0.03	6.70	4.38	3.17	2.86	2.47	5.60	6.00	4.85	5.10	4.50	88.65	99.80	100.00	100.00	100.00
CS-DL pooled ($p_{\bar{y}}=0$, $p_{\bar{x}}=\left[T^{1/3}\right]$ and $p=p_{\bar{x}}-1$)																				
30	-0.40	-0.19	0.16	-0.08	-0.04	15.31	10.44	8.12	7.10	6.41	6.95	6.60	7.20	6.95	6.50	33.75	53.65	73.15	82.45	88.25
50	-0.93	-1.00	-0.29	-0.16	-0.06	11.47	8.31	6.08	5.55	4.80	6.00	6.70	5.70	6.00	4.60	47.05	75.50	91.40	94.80	97.80
100	-1.09	-0.83	-0.13	-0.10	0.14	8.17	5.88	4.26	3.90	3.53	5.85	6.35	5.20	4.60	5.80	74.20	95.10	99.75	99.90	99.95
150	-1.02	-0.72	-0.09	0.11	-0.02	6.83	4.82	3.55	3.28	2.87	5.90	6.10	5.80	5.95	5.55	87.60	98.95	99.95	100.00	100.00
200	-0.81	-0.68	-0.22	-0.06	0.03	5.89	4.11	3.04	2.74	2.43	5.25	5.70	5.20	5.15	5.05	94.55	99.95	100.00	100.00	100.00
Based on **CS-ARDL estimates of short-run coefficients** (ARDL(2,1) specifications with $p_{\bar{z}}=\left[T^{1/3}\right]$).																				
30	-10.40	-3.87	-1.73	-1.38	-0.88	361.68	20.62	7.59	6.24	5.54	10.75	9.80	9.45	8.25	7.50	39.95	66.05	86.05	94.10	96.00
50	3.72	-3.96	-2.12	-1.30	-1.00	182.36	9.56	5.83	4.89	4.32	10.45	12.75	8.10	7.65	6.55	45.40	81.30	98.00	98.90	99.70
100	17.57	-4.03	-2.02	-1.39	-0.83	966.97	7.31	4.34	3.58	3.16	12.90	13.90	10.05	8.45	8.05	61.30	96.40	100.00	100.00	100.00
150	-9.46	-3.93	-2.03	-1.20	-1.07	159.90	6.46	3.84	3.03	2.68	13.65	18.45	12.05	9.45	8.95	67.60	99.50	100.00	100.00	100.00
200	11.29	-3.97	-2.15	-1.42	-1.01	678.37	6.66	3.52	2.79	2.31	13.90	20.10	13.10	11.05	7.70	71.00	99.40	100.00	100.00	100.00
Based on **CS-ARDL estimates of short-run coefficients** (ARDL(1,0) specifications with $p_{\bar{z}}=\left[T^{1/3}\right]$).																				
30	-23.29	-26.91	-23.58	-22.66	-22.01	274.46	28.91	24.83	23.77	22.99	58.05	75.70	85.90	90.40	91.95	86.10	98.00	99.90	100.00	100.00
50	-27.78	-27.49	-23.95	-22.72	-22.28	109.97	28.68	24.70	23.36	22.83	73.20	90.75	97.25	98.60	99.25	92.00	99.80	100.00	100.00	100.00
100	-31.85	-27.64	-24.18	-22.94	-22.21	62.82	28.23	24.54	23.26	22.50	87.75	99.45	100.00	100.00	100.00	96.10	100.00	100.00	100.00	100.00
150	-30.11	-27.60	-24.01	-22.82	-22.34	81.52	28.02	24.26	23.05	22.53	93.35	99.90	100.00	100.00	100.00	97.20	100.00	100.00	100.00	100.00
200	-31.20	-27.73	-24.20	-22.96	-22.40	50.42	28.04	24.39	23.12	22.55	94.25	100.00	100.00	100.00	100.00	97.30	100.00	100.00	100.00	100.00

Notes: The dependent variable and regressors are generated according to (37)-(38) with correlated fixed effects, and with cross-sectionally weakly dependent and serially correlated heteroskedastic idiosyncratic innovations generated according to (41)-(42) with $a_\varepsilon = 0.6$. The knowledge of lag orders is not used in the estimation stage and the integer part of $T^{1/3}$ gives 3, 3, 4, 5 and 5 for $T = 30, 50, 100, 150$ and 200, respectively.

Table 2: Monte Carlo Estimates of Bias, RMSE, Size and Power for Estimation of Long-Run Coefficient (θ) in the Case of the Homogeneous Long-Run

DGP is an ARDL(2,1) model with homogeneous long-run coefficients, heterogeneous short-run coefficients, $\varphi_{max} = 0.6$, stationary regressors, $m = 2$ factors, no feedback effects and $\rho_{\varepsilon i} = 0$.

(N,T)	Bias (×100)					RMSE (×100)					Size (5% level, $H_0 : \theta = 1$)					Power (5% level, $H_1 : \theta = 1.2$)				
	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
CS-DL mean group ($p_{\bar{y}} = 0$, $p_{\bar{x}} = \lceil T^{1/3} \rceil$ and $p = p_{\bar{x}} - 1$)																				
30	-1.59	-0.63	0.00	-0.05	0.01	16.55	10.27	7.40	6.21	5.31	6.00	5.30	6.40	5.25	5.85	30.85	55.85	78.45	88.30	95.05
50	-1.49	-0.95	-0.34	-0.02	-0.11	12.92	8.04	5.89	4.97	4.22	5.25	5.85	6.55	5.60	6.30	41.75	74.65	93.15	97.30	99.55
100	-0.96	-0.55	-0.20	0.02	0.02	9.16	5.76	4.09	3.53	3.02	5.10	5.70	5.40	5.10	5.70	64.90	94.40	99.75	100.00	100.00
150	-1.06	-0.83	-0.18	0.00	-0.10	7.63	4.94	3.35	2.88	2.41	5.45	6.50	5.45	5.35	5.35	79.55	98.70	100.00	100.00	100.00
200	-1.10	-0.77	0.02	-0.02	-0.08	6.47	4.22	2.87	2.52	2.10	4.60	5.95	5.20	5.30	5.00	89.70	99.90	100.00	100.00	100.00
CS-DL pooled ($p_{\bar{y}} = 0$, $p_{\bar{x}} = \lceil T^{1/3} \rceil$ and $p = p_{\bar{x}} - 1$)																				
30	-0.97	-0.42	-0.04	-0.08	0.04	14.33	9.31	6.81	5.83	4.92	7.00	5.55	6.75	5.50	5.95	36.45	62.00	84.30	91.90	97.00
50	-1.01	-0.83	-0.29	-0.02	-0.09	11.07	7.38	5.40	4.59	3.91	5.90	5.20	6.20	5.30	6.85	49.75	80.45	95.50	98.15	99.75
100	-0.72	-0.51	-0.16	0.00	0.05	7.87	5.32	3.76	3.21	2.76	5.30	5.70	5.45	5.60	5.80	76.25	96.60	99.85	100.00	100.00
150	-0.80	-0.61	-0.14	0.05	-0.08	6.55	4.36	3.06	2.59	2.24	6.00	6.05	5.40	5.05	5.15	89.55	99.80	100.00	100.00	100.00
200	-0.95	-0.67	0.04	-0.03	-0.07	5.49	3.83	2.59	2.29	1.91	5.00	6.35	5.10	5.60	5.25	96.15	100.00	100.00	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(2,1) specifications with $p_{\bar{z}} = \lceil T^{1/3} \rceil$).																				
30	-0.25	-3.59	-1.80	-1.24	-0.80	284.59	10.67	6.61	5.00	4.10	10.90	9.90	10.00	8.30	7.90	41.75	72.45	92.20	98.40	99.65
50	-8.02	-3.87	-2.03	-1.39	-1.03	85.61	8.89	5.21	4.08	3.37	10.80	11.10	9.65	9.75	8.55	50.70	85.65	99.35	99.90	100.00
100	-3.69	-3.87	-2.17	-1.38	-0.95	154.40	6.94	4.03	3.01	2.45	12.25	14.85	12.45	10.05	9.10	62.05	97.65	100.00	100.00	100.00
150	-0.25	-3.94	-2.11	-1.37	-1.06	90.79	6.27	3.51	2.60	2.12	14.05	18.10	14.80	12.30	11.75	67.95	99.50	100.00	100.00	100.00
200	-1.53	-3.98	-2.06	-1.37	-1.05	320.07	5.75	3.16	2.39	1.92	15.25	22.55	16.25	14.25	12.15	71.90	99.70	100.00	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(1,0) specifications with $p_{\bar{z}} = \lceil T^{1/3} \rceil$).																				
30	-30.79	-27.20	-23.78	-22.80	-22.12	55.20	28.90	24.89	23.63	22.85	62.55	81.15	91.10	96.25	98.30	87.00	98.90	100.00	100.00	100.00
50	-32.12	-27.63	-24.23	-23.14	-22.57	47.11	28.66	24.88	23.66	23.00	76.85	94.25	99.05	99.75	100.00	94.25	99.90	100.00	100.00	100.00
100	-34.77	-27.78	-24.41	-23.21	-22.57	252.97	28.33	24.72	23.46	22.79	89.85	99.60	99.95	100.00	100.00	96.65	100.00	100.00	100.00	100.00
150	-32.53	-27.88	-24.35	-23.16	-22.62	36.50	28.27	24.56	23.33	22.76	92.55	99.90	100.00	100.00	100.00	96.90	100.00	100.00	100.00	100.00
200	-33.96	-27.90	-24.38	-23.16	-22.63	53.98	28.17	24.54	23.28	22.75	94.35	99.85	100.00	100.00	100.00	97.05	99.95	100.00	100.00	100.00

Notes: The dependent variable and regressors are generated according to (37)-(38) with correlated fixed effects, and with cross-sectionally weakly dependent and serially correlated heteroskedastic idiosyncratic innovations generated according to (41)-(42) with $a_\varepsilon = 0.6$. The knowledge of lag orders is not used in the estimation stage and the integer part of $T^{1/3}$ gives 3, 3, 4, 5 and 5 for $T = 30, 50, 100, 150$ and 200, respectively.

Table 3: Monte Carlo Estimates of Bias, RMSE, Size and Power for Estimation of LR Coefficient (θ) in the Case of the Homogeneous Short-Run

DGP is an ARDL(2,1) model with homogeneous short-run coefficients, $\varphi_{\max} = 0.6$, stationary regressors, $m = 2$ factors, no feedback effects and $\rho_{\varepsilon i} = 0$.

(N,T)	Bias (×100)					Root Mean Square Errors (×100)					Size (5% level, $H_0 : \theta = 1$)					Power (5% level, $H_1 : \theta = 1.2$)				
	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
CS-DL mean group ($p_{\tilde{y}} = 0$, $p_{\tilde{x}} = [T^{1/3}]$ and $p = p_{\tilde{x}} - 1$)																				
30	-2.20	-1.52	-0.09	-0.34	-0.35	19.01	12.19	8.55	7.37	6.14	7.25	6.00	6.40	5.75	5.70	27.80	49.20	69.05	80.30	90.00
50	-2.23	-1.69	-0.51	0.17	0.05	14.93	9.35	6.85	5.79	4.84	6.10	5.45	6.30	6.55	5.60	35.95	67.20	86.15	92.40	97.25
100	-2.24	-1.84	-0.36	-0.21	0.06	10.55	6.93	4.68	4.01	3.33	6.10	7.15	5.45	5.10	4.90	59.10	89.95	98.75	99.90	100.00
150	-1.98	-1.99	-0.47	-0.11	-0.03	8.79	5.82	3.91	3.35	2.66	6.30	7.30	6.50	5.35	4.20	75.35	98.00	99.95	100.00	100.00
200	-2.22	-1.86	-0.35	-0.20	-0.01	7.94	4.96	3.38	2.85	2.38	6.70	6.90	5.25	4.75	4.80	84.40	99.65	100.00	100.00	100.00
CS-DL pooled ($p_{\tilde{y}} = 0$, $p_{\tilde{x}} = [T^{1/3}]$ and $p = p_{\tilde{x}} - 1$)																				
30	-1.94	-1.39	-0.03	-0.35	-0.44	16.68	10.97	7.95	6.83	5.81	7.05	6.00	6.55	6.20	5.85	32.40	54.20	74.35	84.85	93.15
50	-1.96	-1.45	-0.40	0.16	0.02	12.88	8.70	6.29	5.36	4.43	6.80	6.65	6.25	6.60	5.45	44.55	72.80	89.75	95.50	98.75
100	-2.00	-1.66	-0.31	-0.16	0.04	9.10	6.34	4.37	3.70	3.07	6.25	6.10	6.20	5.35	5.25	70.55	93.50	99.30	99.95	100.00
150	-1.68	-1.62	-0.43	-0.08	-0.04	7.61	5.22	3.57	3.05	2.48	6.40	7.10	6.00	4.95	4.05	84.15	99.25	100.00	100.00	100.00
200	-1.94	-1.61	-0.31	-0.19	-0.04	6.76	4.50	3.13	2.59	2.20	6.95	6.55	5.05	3.95	4.45	92.70	99.75	100.00	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(2,1) specifications with $p_{\tilde{z}} = [T^{1/3}]$).																				
30	-14.08	-3.81	-1.92	-1.59	-1.26	310.27	12.81	7.66	6.02	5.04	10.50	10.45	9.20	9.40	9.35	32.50	61.25	84.10	95.30	98.25
50	-4.56	-4.15	-2.14	-1.35	-1.00	242.69	10.59	6.20	4.72	3.98	10.15	11.85	9.80	8.75	8.95	37.90	76.00	96.15	99.50	99.75
100	2.62	-4.32	-2.28	-1.51	-1.11	203.52	8.24	4.61	3.48	2.88	10.70	14.15	11.60	9.35	9.25	47.05	93.10	99.85	100.00	100.00
150	-3.39	-4.50	-2.35	-1.56	-1.12	163.77	7.29	4.09	3.03	2.43	9.90	18.25	14.30	12.10	9.25	51.30	98.35	100.00	100.00	100.00
200	-13.55	-4.32	-2.31	-1.64	-1.18	298.99	6.58	3.71	2.77	2.22	11.40	21.30	15.50	14.05	11.80	56.00	99.45	100.00	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(1,0) specifications with $p_{\tilde{z}} = [T^{1/3}]$).																				
30	-14.02	-10.86	-6.61	-5.78	-5.30	53.22	15.60	9.81	8.21	7.30	24.60	25.10	22.30	22.70	23.35	58.60	82.10	94.65	98.45	99.90
50	-15.83	-11.12	-7.07	-5.88	-5.26	67.84	14.14	9.05	7.41	6.59	30.70	34.25	30.60	31.40	33.65	71.95	93.60	99.90	99.95	100.00
100	-8.04	-11.36	-7.41	-6.14	-5.54	407.90	12.88	8.38	6.88	6.20	41.55	53.15	51.55	52.70	56.60	83.55	99.55	100.00	100.00	100.00
150	-15.24	-11.63	-7.51	-6.24	-5.54	43.70	12.68	8.17	6.77	5.99	50.75	68.95	68.25	70.15	72.20	88.85	100.00	100.00	100.00	100.00
200	-14.75	-11.53	-7.54	-6.33	-5.62	37.89	12.32	8.03	6.70	5.94	56.55	79.15	79.75	82.75	84.60	90.60	99.95	100.00	100.00	100.00

Notes: The dependent variable and regressors are generated according to (37)-(38) with correlated fixed effects, and with cross-sectionally weakly dependent and serially correlated heteroskedastic idiosyncratic innovations generated according to (41)-(42) with $a_\varepsilon = 0.6$. The knowledge of lag orders is not used in the estimation stage and the integer part of $T^{1/3}$ gives 3, 3, 4, 5 and 5 for $T = 30, 50, 100, 150$ and 200, respectively.

Table 4: Monte Carlo Estimates of Bias, RMSE, Size and Power for Estimation of LR Coefficient (θ) in the Case of ARDL(1,0) Model

DGP is an ARDL(1,0) model with heterogeneous coefficients, $\varphi_{max} = 0.6$, stationary regressors, $m = 2$ factors, no feedback effects and $\rho_{\varepsilon i} = 0$.

(N,T)	Bias (×100)					RMSE (×100)					Size (5% level, $H_0: \theta = 1$)					Power (5% level, $H_1: \theta = 1.2$)				
	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
CS-DL mean group ($p_{\bar{y}} = 0$, $p_{\bar{x}} = \lceil T^{1/3} \rceil$ and $p = p_{\bar{x}} - 1$)																				
30	-2.81	-2.34	-1.01	-0.45	-0.51	16.57	10.87	7.93	7.39	6.40	7.00	6.50	5.60	6.80	6.35	34.90	59.65	77.10	82.55	88.80
50	-2.77	-2.29	-0.96	-0.53	-0.57	13.13	8.64	6.25	5.48	5.04	7.30	6.50	6.25	5.45	6.40	47.70	77.20	92.30	95.45	98.20
100	-2.34	-2.30	-0.99	-0.44	-0.57	8.99	6.39	4.51	3.97	3.57	5.90	7.15	6.05	4.90	5.75	71.20	96.00	99.65	99.90	100.00
150	-2.50	-2.18	-1.04	-0.52	-0.52	7.94	5.21	3.71	3.21	2.95	7.10	7.35	6.10	4.90	5.80	85.10	99.45	100.00	100.00	100.00
200	-2.95	-2.36	-1.08	-0.51	-0.57	7.05	4.73	3.37	2.88	2.57	8.25	8.55	7.70	6.50	5.70	94.55	99.90	100.00	100.00	100.00
CS-DL pooled ($p_{\bar{y}} = 0$, $p_{\bar{x}} = \lceil T^{1/3} \rceil$ and $p = p_{\bar{x}} - 1$)																				
30	-2.50	-2.03	-0.98	-0.38	-0.55	14.54	10.27	7.74	7.23	6.34	6.60	7.20	6.10	6.65	5.90	39.80	62.10	79.85	82.80	89.35
50	-2.58	-2.10	-0.81	-0.51	-0.52	11.58	8.08	6.09	5.31	4.95	6.85	6.50	6.15	5.75	5.60	55.75	81.10	93.30	96.45	98.40
100	-2.08	-2.14	-0.88	-0.37	-0.54	7.97	6.04	4.31	3.89	3.51	5.70	7.25	6.15	5.40	6.10	80.65	96.75	99.80	99.95	100.00
150	-2.40	-1.92	-0.92	-0.49	-0.46	6.89	4.88	3.54	3.10	2.90	7.25	7.10	5.80	5.10	5.55	93.30	99.70	100.00	100.00	100.00
200	-2.69	-2.14	-0.96	-0.44	-0.49	6.22	4.39	3.22	2.75	2.53	8.25	7.75	6.75	6.05	5.60	98.45	100.00	100.00	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(2,1) specifications with $p_{\bar{z}} = \lceil T^{1/3} \rceil$).																				
30	-20.46	-3.19	-1.79	-1.05	-0.78	1278.76	10.93	7.21	6.12	5.43	9.95	9.10	7.80	7.70	6.80	39.75	66.70	87.80	93.50	96.25
50	-9.43	-3.23	-1.59	-1.12	-0.88	356.48	8.81	5.55	4.70	4.34	8.95	9.90	7.20	7.25	6.60	48.55	81.60	97.10	99.15	99.70
100	-2.25	-3.30	-1.72	-1.13	-0.96	99.43	6.69	4.25	3.42	3.11	9.05	12.20	9.20	6.95	7.60	58.00	97.15	99.95	100.00	100.00
150	-34.29	-2.95	-1.83	-1.15	-0.87	819.84	5.85	3.63	2.94	2.65	11.15	11.85	10.50	8.60	8.60	65.35	98.90	100.00	100.00	100.00
200	-2.09	-3.32	-1.76	-1.18	-0.93	101.18	6.98	3.22	2.59	2.30	11.70	16.50	10.80	8.95	7.70	70.55	99.45	100.00	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(1,0) specifications with $p_{\bar{z}} = \lceil T^{1/3} \rceil$).																				
30	-6.54	-4.05	-1.86	-0.96	-0.64	19.35	9.58	6.56	5.50	4.98	13.80	12.15	9.75	8.60	6.90	62.55	82.50	93.95	96.85	98.10
50	-6.10	-4.52	-1.89	-1.25	-0.89	26.10	8.02	5.05	4.35	4.04	17.55	14.40	9.00	7.40	6.65	76.65	95.00	99.25	99.75	99.85
100	-18.34	-4.40	-2.12	-1.38	-1.04	537.53	6.50	4.06	3.21	2.91	23.30	19.50	12.10	8.55	7.80	90.55	99.85	100.00	100.00	100.00
150	-11.14	-4.26	-2.27	-1.43	-1.01	129.28	5.75	3.58	2.83	2.51	32.05	23.20	15.65	11.10	9.65	95.65	99.90	100.00	100.00	100.00
200	-7.45	-4.62	-2.21	-1.46	-1.11	11.91	5.73	3.25	2.59	2.24	36.55	31.35	17.45	12.40	9.95	97.30	100.00	100.00	100.00	100.00

Notes: The dependent variable and regressors are generated according to (37)-(38) with correlated fixed effects, and with cross-sectionally weakly dependent and serially correlated heteroskedastic idiosyncratic innovations generated according to (41)-(42) with $a_\varepsilon = 0.6$. The knowledge of lag orders is not used in the estimation stage and the integer part of $T^{1/3}$ gives 3, 3, 4, 5 and 5 for $T = 30, 50, 100, 150$ and 200, respectively.

Table 5: Monte Carlo Estimates of Bias, RMSE, Size and Power for Estimation of LR Coefficient (θ) in the Case of $\phi_{\max} = 0.8$

DGP is an ARDL(2,1) model with heterogeneous coefficients, $\varphi_{\max} = 0.8$, stationary regressors, $m = 2$ factors, no feedback effects and $\rho_{\varepsilon i} = 0$.

(N,T)	Bias ($\times 100$)					Root Mean Square Errors ($\times 100$)					Size (5% level, $H_0 : \theta = 1$)					Power (5% level, $H_1 : \theta = 1.2$)				
	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
CS-DL mean group ($p_{\bar{y}} = 0$, $p_{\bar{x}} = \left[T^{1/3}\right]$ and $p = p_{\bar{x}} - 1$)																				
30	-5.95	-5.68	-2.87	-1.19	-1.86	21.83	14.77	10.87	9.28	8.23	7.70	8.65	7.45	5.75	6.90	28.35	51.30	63.45	66.20	79.25
50	-6.32	-5.87	-2.92	-1.59	-1.70	17.52	12.12	8.42	7.36	6.68	7.90	9.45	6.90	6.60	6.50	40.50	70.75	82.50	86.35	92.55
100	-6.47	-5.47	-3.03	-1.81	-1.67	13.09	9.40	6.46	5.24	4.83	9.35	12.35	8.85	5.85	7.70	65.10	91.90	98.25	98.95	99.70
150	-6.24	-5.60	-2.95	-1.65	-1.70	11.21	8.34	5.55	4.40	4.07	10.55	13.85	9.40	6.25	8.50	80.95	98.45	99.95	99.85	99.95
200	-6.32	-5.68	-3.08	-1.66	-1.68	10.26	7.94	5.05	4.00	3.55	12.30	18.90	10.70	8.05	8.10	90.40	99.65	100.00	100.00	100.00
CS-DL pooled ($p_{\bar{y}} = 0$, $p_{\bar{x}} = \left[T^{1/3}\right]$ and $p = p_{\bar{x}} - 1$)																				
30	-5.46	-5.46	-2.49	-1.12	-1.73	19.66	13.97	10.21	8.88	7.85	7.20	7.75	7.45	5.90	6.50	32.70	55.50	65.80	69.15	81.85
50	-5.54	-5.31	-2.80	-1.29	-1.52	15.73	11.23	8.01	7.04	6.45	8.00	8.75	6.45	6.35	6.70	44.60	73.95	84.45	87.80	93.65
100	-6.04	-5.03	-2.78	-1.64	-1.48	11.81	8.76	6.07	4.98	4.68	10.20	10.65	8.65	6.20	7.95	72.05	94.20	98.50	99.10	99.70
150	-5.61	-5.12	-2.63	-1.46	-1.54	10.06	7.77	5.20	4.19	3.86	9.95	13.25	10.20	6.90	7.70	86.00	98.80	99.95	99.85	99.95
200	-5.78	-5.05	-2.70	-1.47	-1.56	9.20	7.19	4.72	3.79	3.41	12.45	17.45	10.35	7.05	7.55	94.60	99.75	100.00	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(2,1) specifications with $p_{\bar{z}} = \left[T^{1/3}\right]$).																				
30	-12.89	-5.14	-2.82	-1.50	-1.28	786.15	19.34	9.43	7.65	6.69	10.85	11.55	8.90	7.35	7.60	35.00	55.00	74.90	82.50	90.00
50	-2.09	-8.46	-2.72	-1.69	-1.29	394.84	187.62	7.57	6.04	5.40	11.55	10.80	9.15	8.00	7.45	37.90	68.50	89.35	95.00	97.90
100	-30.77	-5.24	-2.87	-2.08	-1.45	768.23	19.56	5.70	4.55	4.04	11.50	14.75	10.10	9.15	9.00	44.10	85.50	98.85	99.75	100.00
150	-15.09	-4.79	-2.98	-1.96	-1.47	375.14	18.01	5.25	3.85	3.37	12.30	16.90	14.00	10.65	9.30	46.45	92.40	99.50	100.00	100.00
200	-1.15	-6.97	-3.07	-1.96	-1.47	229.52	69.18	4.68	3.59	2.98	12.30	22.20	15.75	12.30	9.85	49.15	95.30	100.00	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(1,0) specifications with $p_{\bar{z}} = \left[T^{1/3}\right]$).																				
30	-26.96	-31.57	-23.02	-21.56	-20.97	145.97	203.84	24.91	23.16	22.29	48.25	60.65	70.50	74.85	79.35	72.25	89.15	98.15	99.60	99.70
50	-25.05	-27.49	-23.46	-21.76	-21.08	424.65	29.74	24.99	22.68	21.90	57.20	76.25	87.55	91.00	94.00	77.80	96.45	99.70	100.00	100.00
100	-35.41	-27.96	-23.42	-22.14	-21.19	148.21	32.71	24.01	22.60	21.62	68.30	92.05	98.20	99.70	99.75	84.25	98.30	100.00	100.00	100.00
150	-96.09	-27.09	-23.65	-22.06	-21.17	2622.39	41.48	24.07	22.37	21.46	72.70	96.05	99.85	99.95	100.00	85.90	99.10	100.00	100.00	100.00
200	-30.44	-28.11	-23.68	-21.92	-21.27	248.48	30.67	24.00	22.18	21.48	73.70	97.70	100.00	100.00	100.00	85.25	98.95	100.00	100.00	100.00

Notes: The dependent variable and regressors are generated according to (37)-(38) with correlated fixed effects, and with cross-sectionally weakly dependent and serially correlated heteroskedastic idiosyncratic innovations generated according to (41)-(42) with $a_\varepsilon = 0.6$. The knowledge of lag orders is not used in the estimation stage and the integer part of $T^{1/3}$ gives 3, 3, 4, 5 and 5 for $T = 30, 50, 100, 150$ and 200, respectively.

Table 6: Monte Carlo Estimates of Bias, RMSE, Size and Power for Estimation of LR Coefficient (θ) in the Case of $\phi_{\max} = 0.9$

DGP is an ARDL(2,1) model with heterogeneous coefficients, $\varphi_{\max} = 0.9$, stationary regressors, $m = 2$ factors, no feedback effects and $\rho_{\varepsilon i} = 0$.

(N,T)	Bias (×100)					Root Mean Square Errors (×100)					Size (5% level, $H_0 : \theta = 1$)					Power (5% level, $H_1 : \theta = 1.2$)				
	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
CS-DL mean group ($p_{\bar{y}} = 0$, $p_{\bar{x}} = \lceil T^{1/3}\rceil$ and $p = p_{\bar{x}} - 1$)																				
30	-12.05	-11.37	-6.94	-4.56	-4.89	27.31	19.55	14.78	12.60	11.66	9.00	9.85	9.75	8.05	9.45	31.80	53.15	61.35	63.65	73.00
50	-11.64	-10.52	-6.96	-4.69	-4.71	22.01	16.59	12.23	10.16	9.21	9.15	13.45	11.15	7.85	10.40	41.40	69.35	80.35	79.80	89.90
100	-12.19	-10.74	-6.77	-4.77	-4.63	18.30	14.13	9.85	7.95	7.19	16.35	21.95	16.95	11.60	13.30	67.65	92.05	97.05	96.85	99.45
150	-11.60	-10.76	-6.67	-4.79	-4.63	15.71	13.28	8.88	7.07	6.48	19.90	31.35	22.20	15.05	17.95	84.40	98.20	99.70	99.45	100.00
200	-11.87	-10.66	-6.58	-4.88	-4.77	15.17	12.52	8.22	6.67	6.15	24.00	38.55	27.40	19.25	22.05	92.45	99.80	99.95	100.00	100.00
CS-DL pooled ($p_{\bar{y}} = 0$, $p_{\bar{x}} = \lceil T^{1/3}\rceil$ and $p = p_{\bar{x}} - 1$)																				
30	-11.26	-10.51	-6.36	-4.21	-4.53	24.79	18.53	14.13	12.10	11.24	9.00	10.00	9.75	8.25	9.35	34.00	55.35	63.80	65.80	74.20
50	-10.81	-9.80	-6.44	-4.41	-4.45	20.53	15.76	11.48	9.76	8.83	9.60	13.35	10.95	8.35	9.35	47.60	71.95	81.60	82.50	91.20
100	-11.12	-9.97	-6.24	-4.39	-4.29	16.72	13.31	9.26	7.47	6.85	15.10	20.80	15.20	11.10	13.00	73.60	92.65	97.50	97.55	99.20
150	-10.72	-10.13	-6.22	-4.53	-4.33	14.51	12.54	8.31	6.71	6.13	19.60	30.10	20.00	14.40	17.05	88.05	98.60	99.65	99.70	99.85
200	-10.97	-9.94	-6.18	-4.53	-4.41	14.03	11.76	7.82	6.30	5.81	25.35	35.70	26.10	17.80	22.55	94.70	99.80	99.90	99.95	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(2,1) specifications with $p_z = \lceil T^{1/3}\rceil$).																				
30	-21.58	-6.64	-3.38	-1.70	-1.36	438.72	121.05	12.25	9.96	8.51	11.25	11.05	8.35	8.00	7.25	29.00	46.80	61.95	69.25	76.05
50	4.94	-5.49	-3.31	-2.27	-1.61	922.24	34.12	9.95	7.80	6.81	10.25	10.70	8.70	7.55	7.60	32.30	57.05	75.60	85.50	90.60
100	8.35	-3.40	-3.87	-2.43	-1.82	770.17	112.28	7.72	5.92	5.01	10.95	13.90	11.85	9.55	8.00	35.55	71.20	94.40	97.10	99.25
150	-22.01	-7.03	-3.45	-2.61	-1.84	513.85	91.47	6.40	5.07	4.24	10.65	16.85	11.95	11.30	8.75	35.85	78.90	97.40	99.75	99.80
200	41.07	-4.61	-3.61	-2.70	-1.92	2063.94	127.98	5.90	4.57	3.80	12.45	17.90	14.55	12.30	10.25	38.60	82.55	99.35	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(1,0) specifications with $p_z = \lceil T^{1/3}\rceil$).																				
30	21.99	-27.42	-22.99	-20.75	-20.04	1921.14	70.72	26.25	23.39	22.18	40.55	51.10	57.40	60.25	63.65	60.40	80.05	90.60	94.50	97.40
50	-40.53	-25.17	-23.32	-21.48	-20.71	238.17	219.25	25.35	22.96	21.99	46.45	61.40	71.85	78.45	82.15	65.90	87.55	96.65	98.90	99.65
100	-93.36	-34.65	-23.62	-21.97	-21.09	3030.40	242.83	27.18	22.76	21.72	54.20	75.05	92.25	93.50	97.30	72.30	91.90	99.55	99.80	100.00
150	-33.97	-33.48	-23.56	-21.88	-20.96	472.17	329.85	24.24	22.39	21.39	55.15	81.65	96.60	99.15	99.60	71.60	93.60	99.85	100.00	99.95
200	-37.78	-14.74	-23.75	-22.07	-21.10	247.78	701.23	24.30	22.42	21.41	59.35	86.65	98.80	99.90	100.00	73.45	94.25	100.00	100.00	100.00

Notes: The dependent variable and regressors are generated according to (37)-(38) with correlated fixed effects, and with cross-sectionally weakly dependent and serially correlated heteroskedastic idiosyncratic innovations generated according to (41)-(42) with $a_\varepsilon = 0.6$. The knowledge of lag orders is not used in the estimation stage and the integer part of $T^{1/3}$ gives 3, 3, 4, 5 and 5 for $T = 30, 50, 100, 150$ and 200, respectively.

Table 7: Monte Carlo Estimates of Bias, RMSE, Size and Power for Estimation of LR Coefficient (θ) in the Case of $m = 3$ Factors

DGP is an ARDL(2,1) model with heterogeneous coefficients, $\varphi_{\max} = 0.6$, stationary regressors, $m = 3$ factors, no feedback effects and $\rho_{\varepsilon i} = 0$.

(N,T)	Bias (×100)					Root Mean Square Errors (×100)					Size (5% level, $H_0 : \theta = 1$)					Power (5% level, $H_1 : \theta = 1.2$)				
	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
CS-DL mean group ($p_{\bar{y}} = 0$, $p_{\bar{x}} = [T^{1/3}]$ and $p = p_{\bar{x}} - 1$)																				
30	-1.85	-1.02	0.00	-0.01	-0.02	17.23	11.03	8.20	7.45	6.63	6.70	6.15	6.90	5.75	6.40	29.80	51.70	68.05	78.05	85.25
50	-0.79	-0.63	-0.24	0.03	0.10	12.90	8.79	6.68	5.77	5.04	4.90	6.55	6.35	6.20	5.05	38.65	67.25	85.85	92.70	96.60
100	-1.00	-0.94	-0.23	0.09	-0.15	9.57	6.23	4.73	4.12	3.72	5.70	5.40	5.55	5.20	5.35	61.05	92.80	99.10	99.60	99.95
150	-1.19	-0.88	-0.02	0.00	-0.06	7.77	5.03	3.73	3.34	2.92	6.05	4.85	5.15	5.40	4.60	78.45	98.30	99.90	100.00	100.00
200	-0.99	-0.78	-0.03	-0.03	0.09	6.62	4.50	3.23	2.88	2.61	5.00	5.60	5.20	4.85	4.90	89.15	99.60	100.00	100.00	100.00
CS-DL pooled ($p_{\bar{y}} = 0$, $p_{\bar{x}} = [T^{1/3}]$ and $p = p_{\bar{x}} - 1$)																				
30	-1.59	-0.79	0.08	-0.04	0.01	15.07	10.40	7.96	7.15	6.57	7.05	6.40	7.05	5.95	6.45	34.10	54.25	72.20	80.00	86.40
50	-0.91	-0.57	-0.17	0.03	0.11	11.51	8.27	6.44	5.57	4.93	5.65	6.50	6.55	5.50	5.00	47.50	71.95	88.25	93.90	97.65
100	-1.01	-0.88	-0.21	0.10	-0.16	8.33	5.78	4.51	3.97	3.58	5.55	5.85	5.50	5.35	5.20	72.30	94.90	99.55	99.85	100.00
150	-0.90	-0.72	-0.04	0.00	-0.07	6.78	4.73	3.61	3.28	2.88	6.20	5.15	5.70	5.35	4.80	87.25	99.10	100.00	100.00	100.00
200	-0.96	-0.63	-0.05	0.01	0.08	5.79	4.19	3.08	2.81	2.55	5.25	5.60	5.10	5.40	5.60	94.80	99.85	100.00	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(2,1) specifications with $p_{\bar{z}} = [T^{1/3}]$).																				
30	-5.34	-2.54	-0.89	-0.10	0.34	144.86	11.33	7.34	6.29	5.65	9.10	8.80	7.80	6.70	7.15	37.10	61.60	83.70	88.95	93.35
50	-5.60	-2.52	-1.03	-0.20	0.23	157.20	9.36	5.91	4.86	4.28	9.20	9.80	8.10	6.20	5.85	44.65	76.50	94.60	98.05	99.40
100	-0.04	-3.29	-1.21	-0.23	-0.07	83.58	10.20	4.28	3.43	3.10	9.65	11.05	7.85	6.65	6.25	56.30	95.20	100.00	100.00	100.00
150	-4.99	-2.93	-0.94	-0.44	0.01	94.15	7.97	3.48	2.82	2.54	11.80	13.50	8.20	4.80	5.40	64.25	98.65	99.95	100.00	100.00
200	7.81	1.08	-1.01	-0.36	0.08	383.41	178.92	3.08	2.51	2.23	12.10	15.05	8.15	6.75	6.55	67.70	99.60	100.00	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(1,0) specifications with $p_{\bar{z}} = [T^{1/3}]$).																				
30	-24.61	-25.74	-22.19	-20.87	-20.48	271.79	27.83	23.55	22.09	21.54	57.05	70.00	81.45	84.80	86.65	84.95	97.50	99.85	99.95	100.00
50	-29.47	-25.95	-22.39	-21.20	-20.52	39.80	27.22	23.27	21.90	21.18	70.30	88.00	94.45	97.15	98.05	91.90	99.65	100.00	100.00	100.00
100	-30.85	-26.44	-22.73	-21.26	-20.94	37.04	27.17	23.16	21.63	21.28	85.75	98.60	100.00	99.95	100.00	96.20	99.95	100.00	100.00	100.00
150	-31.83	-26.42	-22.46	-21.61	-20.86	48.65	26.97	22.76	21.85	21.08	89.95	99.95	99.95	100.00	100.00	96.05	99.95	100.00	100.00	100.00
200	-29.13	-26.22	-22.55	-21.46	-20.80	108.14	26.59	22.77	21.66	20.97	91.85	99.85	100.00	100.00	100.00	96.20	100.00	100.00	100.00	100.00

Notes: The dependent variable and regressors are generated according to (37)-(38) with correlated fixed effects, and with cross-sectionally weakly dependent and serially correlated heteroskedastic idiosyncratic innovations generated according to (41)-(42) with $a_\varepsilon = 0.6$. The knowledge of lag orders is not used in the estimation stage and the integer part of $T^{1/3}$ gives 3, 3, 4, 5 and 5 for $T = 30, 50, 100, 150$ and 200, respectively.

29

Table 8: Monte Carlo Estimates of Bias, RMSE, Size and Power for Estimation of LR Coefficient (θ) in the Case of Unit Roots in Factors

DGP is an ARDL(2,1) model with heterogeneous coefficients, $\varphi_{\max} = 0.6$, unit roots in factors, $m = 2$ factors, no feedback effects and $\rho_{\varepsilon i} = 0$.

(N,T)	Bias (×100)					Root Mean Square Errors (×100)					Size (5% level, $H_0: \theta = 1$)					Power (5% level, $H_1: \theta = 1.2$)				
	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
CS-DL mean group ($p_{\bar{y}} = 0$, $p_{\bar{x}} = \lceil T^{1/3}\rceil$ and $p = p_{\bar{x}} - 1$)																				
30	-1.04	-1.04	-0.14	0.06	0.05	16.26	11.33	8.28	7.27	6.55	5.65	6.85	6.60	5.95	6.55	28.50	53.25	71.45	78.30	86.40
50	-0.84	-0.84	-0.30	0.20	-0.16	12.76	8.51	6.56	5.79	5.18	5.15	4.85	5.35	6.40	5.70	38.25	70.50	87.65	92.80	97.10
100	-1.42	-0.99	-0.04	0.03	-0.14	9.37	6.29	4.55	4.05	3.62	5.30	5.50	5.10	5.55	5.40	63.55	92.60	98.75	99.80	100.00
150	-1.15	-0.91	-0.14	-0.08	0.01	7.87	5.26	3.73	3.31	2.91	5.90	6.40	4.75	4.95	4.90	78.30	98.10	99.95	100.00	100.00
200	-1.14	-0.79	-0.21	-0.03	0.03	6.79	4.43	3.24	2.90	2.50	5.50	4.95	5.20	5.10	4.95	88.00	99.65	100.00	100.00	100.00
CS-DL pooled ($p_{\bar{y}} = 0$, $p_{\bar{x}} = \lceil T^{1/3}\rceil$ and $p = p_{\bar{x}} - 1$)																				
30	-0.69	-0.78	-0.15	0.09	0.04	14.53	10.66	8.01	7.04	6.47	6.25	6.75	6.80	6.15	7.20	33.10	57.20	72.95	80.50	88.00
50	-0.69	-0.81	-0.27	0.19	-0.24	11.42	7.97	6.17	5.62	5.06	5.05	5.25	5.80	6.40	5.60	45.45	74.10	90.60	94.15	97.55
100	-1.17	-0.80	-0.05	0.02	-0.09	8.25	5.77	4.38	4.03	3.60	5.40	5.55	5.80	5.65	5.60	72.60	95.00	100.00	99.90	100.00
150	-0.94	-0.82	-0.14	-0.05	0.03	6.98	4.88	3.55	3.20	2.85	6.05	5.90	5.70	5.15	5.40	86.00	98.90	100.00	100.00	100.00
200	-0.98	-0.70	-0.20	-0.01	0.01	5.96	4.15	3.16	2.83	2.48	5.60	5.45	5.60	5.40	5.20	94.05	99.90	100.00	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(2,1) specifications with $p_{\bar{z}} = \lceil T^{1/3}\rceil$).																				
30	14.28	-3.42	-1.78	-0.89	-0.48	760.36	14.43	7.34	6.04	5.47	9.40	10.10	8.35	6.75	7.20	38.75	66.45	88.85	93.30	96.40
50	-3.16	-3.92	-2.23	-1.19	-0.94	106.12	9.21	6.14	4.87	4.52	11.35	10.70	10.25	7.65	8.40	49.10	83.25	96.90	99.25	99.55
100	-9.23	-5.09	-2.62	-1.63	-1.20	293.70	7.95	4.67	3.75	3.31	15.00	18.15	12.30	9.90	8.40	65.15	97.95	100.00	100.00	100.00
150	-12.59	-5.44	-2.95	-1.97	-1.39	395.28	7.37	4.33	3.35	2.78	18.40	25.50	18.25	13.50	8.80	73.35	99.55	100.00	100.00	100.00
200	-5.66	-5.90	-3.07	-2.04	-1.47	91.39	7.55	4.15	3.13	2.56	21.95	34.35	22.30	16.80	11.60	77.50	99.90	100.00	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(1,0) specifications with $p_{\bar{z}} = \lceil T^{1/3}\rceil$).																				
30	-28.65	-26.01	-22.78	-21.75	-21.38	43.69	28.01	24.09	22.81	22.34	54.90	72.60	84.75	89.45	91.55	84.60	97.80	99.90	100.00	100.00
50	-30.99	-26.05	-23.52	-21.99	-21.78	80.99	27.34	24.34	22.65	22.38	67.60	87.65	96.40	98.10	99.00	89.65	99.60	100.00	100.00	100.00
100	-32.82	-27.17	-23.82	-22.65	-21.99	69.50	27.81	24.20	22.98	22.28	86.30	99.45	100.00	100.00	100.00	95.80	99.95	100.00	100.00	100.00
150	-26.37	-27.61	-24.12	-22.92	-22.31	175.08	28.01	24.37	23.13	22.50	92.10	99.95	100.00	100.00	100.00	96.95	100.00	100.00	100.00	100.00
200	-32.57	-27.95	-24.27	-22.94	-22.32	41.83	28.25	24.46	23.11	22.46	93.85	100.00	100.00	100.00	100.00	96.45	100.00	100.00	100.00	100.00

Notes: The dependent variable and regressors are generated according to (37)-(38) with correlated fixed effects, and with cross-sectionally weakly dependent and serially correlated heteroskedastic idiosyncratic innovations generated according to (41)-(42) with $a_\varepsilon = 0.6$. The knowledge of lag orders is not used in the estimation stage and the integer part of $T^{1/3}$ gives 3, 3, 4, 5 and 5 for $T = 30, 50, 100, 150$ and 200, respectively.

Table 9: Monte Carlo Estimates of Bias, RMSE, Size and Power for Estimation of LR Coefficient (θ) in the Case of Unit Roots in Regressor Specific Components

DGP is an ARDL(2,1) model with heterogeneous coefficients, $\varphi_{max} = 0.6$, unit roots in v_{it}, $m = 2$ factors, no feedback effects and $\rho_{\varepsilon i} = 0$.

(N,T)	Bias (×100)					Root Mean Square Errors (×100)					Size (5% level, $H_0: \theta = 1$)					Power (5% level, $H_1: \theta = 1.2$)				
	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
CS-DL mean group ($p_{\bar{y}} = 0$, $p_{\bar{x}} = [T^{1/3}]$ and $p = p_{\bar{x}} - 1$)																				
30	-2.99	0.37	0.25	0.33	-0.13	65.72	32.60	15.74	10.52	8.06	5.70	5.00	5.45	5.70	5.65	7.85	11.45	29.45	52.30	72.85
50	-0.20	0.66	-0.40	0.24	0.20	51.19	26.47	12.29	8.26	6.42	5.45	6.40	6.00	5.25	5.30	8.30	14.55	42.20	69.35	87.30
100	-0.76	-0.50	-0.09	0.11	0.09	37.01	18.87	9.00	5.88	4.44	4.70	5.65	6.00	5.00	4.45	9.40	22.00	64.90	91.60	99.05
150	-0.85	-0.17	-0.25	-0.06	0.08	30.36	15.18	7.17	4.76	3.72	4.55	5.35	5.80	4.90	5.45	10.75	27.15	81.85	98.60	99.90
200	0.15	-0.14	-0.18	-0.17	0.18	27.24	13.26	6.28	4.21	3.15	5.90	5.55	5.40	6.00	5.00	12.25	34.05	89.65	99.60	100.00
CS-DL pooled ($p_{\bar{y}} = 0$, $p_{\bar{x}} = [T^{1/3}]$ and $p = p_{\bar{x}} - 1$)																				
30	-1.75	0.62	-0.04	-0.01	-0.06	47.00	24.71	12.38	9.06	7.13	5.20	5.60	6.50	7.40	6.70	8.95	14.60	40.85	64.45	78.95
50	-0.67	0.37	0.00	-0.01	0.02	36.30	19.19	9.39	6.88	5.64	4.95	5.40	5.70	6.45	6.30	10.40	20.25	57.30	82.95	93.15
100	-0.92	-0.95	-0.24	-0.01	-0.01	27.06	13.97	6.90	4.75	4.01	6.15	5.90	5.85	4.95	5.10	14.50	35.45	84.50	98.25	99.50
150	-1.09	-0.34	-0.08	-0.13	0.13	21.26	11.24	5.47	3.98	3.31	4.70	5.55	4.75	5.45	4.85	16.95	46.00	95.00	99.75	100.00
200	-0.05	-0.11	-0.03	-0.10	0.07	18.94	9.81	4.74	3.45	2.76	4.95	5.45	5.00	5.05	5.20	20.10	55.20	98.25	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(2,1) specifications with $p_{\bar{z}} = [T^{1/3}]$).																				
30	11.59	-2.76	-0.64	-0.09	-0.27	1487.96	52.46	15.63	10.51	8.10	5.00	5.80	5.60	5.70	6.00	7.30	12.85	31.35	55.35	73.00
50	-28.86	-0.82	-0.92	-0.12	-0.02	1128.77	27.42	12.19	8.26	6.34	4.60	6.10	6.70	5.80	5.60	7.95	14.95	44.25	70.55	88.30
100	-6.61	-1.86	-0.68	-0.24	-0.12	361.66	19.50	9.00	5.77	4.47	4.70	5.10	6.35	5.25	5.80	8.75	22.25	67.60	93.20	99.20
150	-61.57	-1.90	-0.82	-0.28	-0.11	2620.58	16.18	7.17	4.67	3.67	4.05	5.40	5.50	4.95	5.80	10.05	30.85	83.80	99.00	99.95
200	2.04	-1.54	-0.68	-0.41	0.00	336.21	14.05	6.24	4.21	3.12	5.05	5.25	5.45	5.55	4.90	11.10	36.55	91.45	99.70	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(1,0) specifications with $p_{\bar{z}} = [T^{1/3}]$).																				
30	-14.85	-9.27	-4.73	-2.76	-2.30	146.76	36.98	17.15	11.44	8.94	6.65	6.45	6.85	7.05	7.30	12.20	17.30	38.80	58.70	76.15
50	-18.57	-8.15	-4.38	-2.42	-1.79	136.51	28.29	13.16	8.92	6.79	6.80	7.15	7.35	6.70	6.15	12.90	22.80	53.80	76.50	91.50
100	-16.57	-8.00	-3.42	-2.03	-1.39	72.70	20.94	9.73	6.19	4.75	6.60	7.90	7.75	6.75	7.15	16.30	33.75	76.10	95.75	99.40
150	-12.87	-7.13	-3.19	-1.80	-1.24	85.37	17.40	7.87	5.12	3.91	8.00	7.75	7.90	6.95	7.00	19.85	43.15	89.60	99.30	100.00
200	-16.82	-6.73	-2.89	-1.82	-1.04	346.63	15.13	6.99	4.62	3.33	8.45	7.75	8.35	7.75	6.15	23.20	51.75	95.40	99.90	100.00

Notes: The dependent variable and regressors are generated according to (37)-(38) with correlated fixed effects, and with cross-sectionally weakly dependent and serially correlated heteroskedastic idiosyncratic innovations generated according to (41)-(42) with $a_\varepsilon = 0.6$. The knowledge of lag orders is not used in the estimation stage and the integer part of $T^{1/3}$ gives 3, 3, 4, 5 and 5 for $T = 30, 50, 100, 150$ and 200, respectively.

Table 10: Monte Carlo Estimates of Bias, RMSE, Size and Power for Estimation of LR Coefficient (θ) in the Case of Serially Correlated Idiosyncratic Errors

DGP is an ARDL(2,1) model with heterogeneous coefficients, $\varphi_{\max} = 0.6$, stationary regressors, $m = 2$ factors, no feedback effects and $\rho_{\varepsilon i} \sim IIDU\ (0, 0.8)$.

(N,T)	Bias (×100)					Root Mean Square Errors (×100)					Size (5% level, $H_0:\theta=1$)					Power (5% level, $H_1:\theta=1.2$)				
	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
CS-DL mean group ($p_{\bar y}=0$, $p_{\bar x} = \lceil T^{1/3}\rceil$ and $p = p_{\bar x} - 1$)																				
30	-1.11	-0.45	-0.21	0.01	0.12	21.91	14.75	11.44	10.33	8.89	7.05	6.00	5.40	6.25	5.55	20.20	31.40	47.00	53.70	64.15
50	-0.29	-1.04	0.08	-0.15	-0.04	16.85	11.71	8.88	7.89	6.96	5.25	6.20	5.95	5.00	5.95	24.05	46.65	64.10	71.85	82.20
100	-1.22	-0.75	-0.11	0.03	0.08	12.21	8.14	6.04	5.55	4.97	4.80	5.05	4.85	4.45	5.90	43.90	72.35	90.05	94.50	97.70
150	-1.21	-1.04	-0.07	0.09	-0.03	10.05	6.90	5.22	4.65	4.16	5.05	5.40	5.55	5.35	6.65	58.00	86.40	96.80	99.15	99.70
200	-1.28	-0.80	0.05	0.01	-0.08	8.62	5.89	4.56	3.99	3.45	5.40	5.25	6.25	5.45	4.15	70.00	94.45	99.05	99.95	100.00
CS-DL pooled ($p_{\bar y}=0$, $p_{\bar x} = \lceil T^{1/3}\rceil$ and $p = p_{\bar x} - 1$)																				
30	-0.86	-0.35	-0.06	-0.12	0.22	20.29	14.11	11.03	10.01	8.71	6.85	6.30	6.25	6.20	6.30	24.05	33.80	49.75	58.05	67.25
50	-0.35	-0.97	0.10	-0.11	-0.05	15.46	11.22	8.51	7.57	6.73	5.70	5.95	5.10	5.60	5.55	28.65	50.00	67.10	75.40	83.80
100	-1.11	-0.57	-0.11	0.08	0.15	11.05	7.75	5.89	5.28	4.75	5.30	4.50	4.65	5.00	4.55	49.65	75.05	91.10	95.30	98.35
150	-1.00	-0.77	-0.07	0.02	-0.05	9.06	6.50	5.02	4.43	4.01	5.20	5.45	5.55	4.90	6.00	66.10	90.00	97.60	99.45	99.75
200	-1.01	-0.52	0.00	0.04	-0.05	7.93	5.59	4.35	3.82	3.36	5.95	5.30	5.70	4.60	4.75	76.15	95.40	99.60	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(2,1) specifications with $p_{\bar z} = \lceil T^{1/3}\rceil$).																				
30	-15.73	22.36	16.11	16.50	17.10	1331.93	276.49	20.13	19.44	19.45	6.85	14.05	31.85	40.85	50.60	9.65	10.65	11.80	11.75	11.80
50	-1.80	13.61	16.51	16.65	17.00	568.06	81.68	19.19	18.46	18.44	7.50	20.60	46.55	61.45	71.70	8.95	11.60	11.20	12.80	11.60
100	43.49	16.66	15.91	16.56	17.04	1184.81	34.38	17.18	17.47	17.81	10.55	36.15	72.45	88.25	94.65	7.80	10.90	12.10	12.80	14.00
150	21.78	16.67	16.16	16.54	16.87	382.97	38.39	17.12	17.17	17.40	11.05	47.10	88.60	96.60	99.05	7.10	14.00	16.20	15.60	15.90
200	15.42	-93.00	16.23	16.65	16.86	323.97	4968.11	16.94	17.14	17.26	13.40	57.50	95.35	99.35	99.90	5.85	13.25	17.45	17.80	17.90
Based on CS-ARDL estimates of short-run coefficients (ARDL(1,0) specifications with $p_{\bar z} = \lceil T^{1/3}\rceil$).																				
30	16.40	-6.69	16.36	17.54	18.91	697.94	880.16	23.81	22.65	23.23	6.75	7.70	17.15	24.65	30.80	17.05	15.95	12.15	10.70	9.60
50	-20.75	12.88	15.81	17.78	18.31	2302.74	61.28	23.24	20.87	21.06	4.45	9.05	23.80	38.50	44.45	15.95	16.20	11.95	8.50	8.80
100	48.11	12.90	15.75	18.12	18.87	1567.80	164.94	41.04	22.77	20.28	5.10	11.70	46.05	66.15	75.90	17.95	17.55	11.35	8.55	6.65
150	26.90	12.79	16.79	17.82	18.81	1050.80	491.93	18.79	18.97	19.79	3.85	16.40	63.75	83.05	91.10	15.30	18.55	11.85	8.90	8.05
200	0.27	-25.36	16.89	16.07	18.58	354.88	1004.96	19.27	84.64	19.27	3.65	20.55	79.10	92.90	96.70	16.75	17.85	11.20	9.20	7.35

Notes: The dependent variable and regressors are generated according to (37)-(38) with correlated fixed effects, and with cross-sectionally weakly dependent and serially correlated heteroskedastic idiosyncratic innovations generated according to (41)-(42) with $a_{\varepsilon} = 0.6$. The knowledge of lag orders is not used in the estimation stage and the integer part of $T^{1/3}$ gives 3, 3, 4, 5 and 5 for $T = 30, 50, 100, 150$ and 200, respectively.

Table 11: Monte Carlo Estimates of Bias, RMSE, Size and Power for Estimation of LR Coefficient (θ) in the Case of Breaks in Errors

DGP is an ARDL(2,1) model with heterogeneous coefficients, $\varphi_{\max} = 0.6$, stationary regressors, $m = 2$ factors, no feedback effects and breaks in errors.

(N,T)	Bias (×100)					Root Mean Square Errors (×100)					Size (5% level, $H_0 : \theta = 1$)					Power (5% level, $H_1 : \theta = 1.2$)				
	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
CS-DL mean group ($p_{\bar{y}} = 0$, $p_{\bar{x}} = \lceil T^{1/3}\rceil$ and $p = p_{\bar{x}} - 1$)																				
30	-1.72	-1.26	0.09	0.11	0.25	18.91	12.45	9.88	9.03	7.86	7.20	6.45	6.60	6.05	6.15	25.55	44.75	59.60	64.90	72.45
50	-0.83	-0.82	-0.08	0.09	0.05	14.33	10.13	7.55	6.75	6.27	4.90	6.25	5.75	4.80	5.85	33.70	58.60	78.05	83.05	87.85
100	-1.00	-0.91	-0.08	0.10	-0.06	10.50	7.00	5.23	5.00	4.37	5.30	5.15	4.75	5.70	4.95	54.90	85.70	95.85	96.95	99.20
150	-0.88	-0.90	-0.10	0.05	-0.05	8.57	5.62	4.43	4.01	3.63	6.00	5.35	6.05	4.50	4.75	71.40	95.80	99.25	99.70	100.00
200	-0.63	-0.85	-0.29	0.00	-0.01	7.47	4.92	3.72	3.36	3.12	5.50	5.00	4.90	4.55	4.70	80.85	98.15	100.00	100.00	100.00
CS-DL pooled ($p_{\bar{y}} = 0$, $p_{\bar{x}} = \lceil T^{1/3}\rceil$ and $p = p_{\bar{x}} - 1$)																				
30	-1.57	-1.05	0.01	0.15	0.10	16.82	12.04	9.49	8.78	7.74	7.10	7.10	6.60	6.70	6.80	30.20	48.15	61.50	66.10	74.70
50	-1.13	-0.72	-0.04	0.10	0.06	12.71	9.34	7.25	6.48	6.09	4.80	6.20	5.70	5.05	5.75	41.60	63.65	80.55	84.65	89.50
100	-0.90	-0.78	-0.15	0.03	-0.02	9.11	6.53	5.03	4.80	4.27	5.55	5.45	4.80	5.65	5.50	63.95	89.15	96.85	98.20	99.55
150	-0.83	-0.79	-0.03	0.09	-0.03	7.40	5.37	4.23	3.86	3.52	5.45	5.75	5.55	4.70	4.75	81.20	96.60	99.40	99.85	99.95
200	-0.65	-0.78	-0.27	-0.01	-0.05	6.61	4.71	3.59	3.25	3.01	5.20	4.75	4.90	4.50	4.60	87.80	99.35	99.95	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(2,1) specifications with $p_{\bar{z}} = \lceil T^{1/3}\rceil$).																				
30	6.24	6.67	9.59	11.95	14.00	143.70	16.06	14.00	14.80	16.05	5.15	7.90	20.45	33.40	47.30	19.30	28.25	32.35	24.85	19.80
50	10.20	7.71	9.03	11.55	13.95	111.76	19.37	11.73	13.27	15.25	5.70	11.20	24.55	46.10	68.65	21.30	33.45	41.95	32.90	22.50
100	30.94	7.07	9.19	11.58	13.88	514.17	26.38	10.71	12.48	14.59	6.45	13.35	44.50	75.60	91.90	24.45	50.40	57.90	48.75	35.30
150	24.76	7.07	8.80	11.62	13.67	876.89	15.90	9.83	12.24	14.13	7.50	18.80	58.30	89.95	98.20	25.80	57.75	74.45	62.40	47.05
200	1.98	6.22	8.85	11.51	13.76	403.92	28.52	9.58	11.97	14.12	7.80	22.55	70.15	96.35	99.70	28.30	66.50	84.55	73.25	54.60
Based on CS-ARDL estimates of short-run coefficients (ARDL(1,0) specifications with $p_{\bar{z}} = \lceil T^{1/3}\rceil$).																				
30	-28.03	-10.64	-5.13	-1.40	2.17	411.57	21.22	12.54	10.55	10.34	25.20	21.75	13.05	8.55	6.90	54.20	64.00	67.10	60.45	52.10
50	-15.90	-9.72	-5.33	-1.90	2.08	78.66	20.38	11.70	8.89	8.24	28.55	23.30	14.90	9.95	6.65	59.20	75.90	82.00	76.55	65.05
100	-30.60	-11.17	-5.39	-1.63	2.16	353.77	28.48	9.12	5.93	6.09	34.05	33.95	20.45	8.00	6.95	69.50	89.50	95.20	94.20	86.30
150	-15.01	38168	-5.79	-1.55	2.13	497.45	17476	7.85	5.01	5.18	38.45	41.60	26.80	8.50	7.75	74.20	93.75	98.75	97.80	94.45
200	-14.40	-11.21	-5.69	-1.72	2.24	123.28	26.47	7.33	4.61	4.59	41.95	49.25	30.75	9.95	8.20	76.90	95.45	99.60	99.55	98.75

Notes: The dependent variable and regressors are generated according to (37)-(38) with correlated fixed effects, and with cross-sectionally weakly dependent and serially correlated heteroskedastic idiosyncratic innovations generated according to (41)-(42) with $a_{\varepsilon} = 0.6$. The knowledge of lag orders is not used in the estimation stage and the integer part of $T^{1/3}$ gives 3, 3, 4, 5 and 5 for $T = 30, 50, 100, 150$ and 200, respectively.

Table 12: Monte Carlo Estimates of Bias, RMSE, Size and Power for Estimation of LR Coefficient (θ) in the Case of Feedback Effects

DGP is an ARDL(2,1) model with heterogeneous coefficients, $\varphi_{\max} = 0.6$, stationary regressors, $m = 2$ factors,
$\kappa_{yi} \sim IIDU(0, 0.2)$, and $\rho_{\varepsilon i} = 0$.

(N,T)	Bias (×100)					RMSE (×100)					Size (5% level, $H_0 : \theta = 1$)					Power (5% level, $H_1 : \theta = 1.2$)				
	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
CS-DL mean group ($p_{\bar{y}} = 0$, $p_{\bar{x}} = [T^{1/3}]$ and $p = p_{\bar{x}} - 1$)																				
30	-1.51	2.38	5.10	5.74	5.84	16.20	10.51	9.35	9.16	8.54	6.00	6.60	12.50	15.70	18.20	32.55	43.95	53.10	57.00	65.75
50	-1.45	2.38	5.11	5.48	6.06	12.54	8.54	7.91	7.68	7.79	6.50	8.10	14.45	19.00	25.55	45.25	62.35	70.55	77.40	81.15
100	-0.97	2.85	5.16	5.61	6.05	8.88	6.34	6.70	6.85	6.93	5.80	7.90	23.30	31.85	43.05	69.00	85.35	92.80	95.60	97.85
150	-1.23	2.64	5.05	5.68	6.15	7.50	5.41	6.12	6.45	6.76	5.80	8.70	30.90	46.05	59.55	84.60	95.60	98.90	99.50	99.90
200	-1.46	2.55	4.91	5.61	6.03	6.57	4.72	5.75	6.26	6.50	6.30	9.90	37.55	55.20	71.25	92.90	98.75	99.90	99.95	100.00
CS-DL pooled ($p_{\bar{y}} = 0$, $p_{\bar{x}} = [T^{1/3}]$ and $p = p_{\bar{x}} - 1$)																				
30	2.28	4.73	6.80	7.10	7.00	14.57	10.87	10.37	10.04	9.43	7.25	8.90	17.20	19.90	21.25	30.75	40.75	45.90	50.95	57.55
50	2.26	4.83	6.77	6.89	7.40	11.41	9.10	8.96	8.75	8.94	6.55	10.70	21.25	26.60	33.25	41.50	54.60	61.40	69.95	72.05
100	2.96	5.30	6.89	7.10	7.42	8.53	7.65	8.07	8.11	8.17	7.20	16.60	37.05	47.05	58.10	61.00	75.60	85.40	89.95	92.90
150	2.70	5.07	6.72	7.20	7.50	7.21	6.83	7.55	7.84	8.03	8.05	20.65	50.65	63.80	75.60	76.60	90.55	95.50	97.85	98.55
200	2.47	5.10	6.68	7.15	7.36	6.21	6.40	7.30	7.68	7.78	7.60	25.45	62.00	75.80	85.45	87.80	95.70	99.05	99.50	99.80
Based on CS-ARDL estimates of short-run coefficients (ARDL(2,1) specifications with $p_z = [T^{1/3}]$).																				
30	-49.31	-6.64	-2.84	-1.50	-1.23	1613.71	11.90	7.13	6.12	5.29	17.15	14.35	8.40	8.70	7.50	54.75	80.05	93.55	95.05	97.95
50	15.94	-7.02	-3.05	-2.03	-1.28	1208.10	10.96	6.01	4.85	4.27	19.45	19.25	11.10	8.95	7.75	65.05	93.30	99.15	99.90	99.70
100	-17.83	-6.54	-3.06	-2.00	-1.50	120.30	9.22	4.83	3.81	3.25	26.80	26.90	15.40	12.55	10.15	78.50	99.35	100.00	100.00	100.00
150	-14.20	-6.84	-3.24	-2.06	-1.45	127.42	8.28	4.40	3.27	2.74	32.55	35.75	21.15	13.70	9.90	82.20	99.90	100.00	100.00	100.00
200	-17.89	-7.08	-3.39	-2.15	-1.57	283.24	8.22	4.28	3.15	2.56	37.60	47.30	27.80	19.30	12.55	83.90	99.85	100.00	100.00	100.00
Based on CS-ARDL estimates of short-run coefficients (ARDL(1,0) specifications with $p_z = [T^{1/3}]$).																				
30	-37.31	-29.77	-24.72	-23.08	-22.33	236.77	31.59	25.86	24.09	23.27	74.60	86.65	91.90	93.30	94.25	93.00	99.70	100.00	100.00	100.00
50	-39.45	-29.96	-25.01	-23.50	-22.67	100.20	30.98	25.63	24.07	23.19	88.15	97.15	99.00	99.55	99.80	96.60	100.00	100.00	100.00	100.00
100	-38.09	-30.16	-25.06	-23.56	-22.75	60.47	30.71	25.39	23.86	23.02	94.90	99.80	100.00	100.00	100.00	97.70	99.95	100.00	100.00	100.00
150	-37.23	-30.28	-25.21	-23.61	-22.85	52.96	30.76	25.43	23.80	23.03	95.80	99.85	100.00	100.00	100.00	97.80	99.90	100.00	100.00	100.00
200	-39.43	-30.46	-25.40	-23.73	-22.89	54.51	30.72	25.56	23.88	23.03	95.80	99.90	100.00	100.00	100.00	97.85	99.95	100.00	100.00	100.00

Notes: The dependent variable and regressors are generated according to (37)-(38) with correlated fixed effects, and with cross-sectionally weakly dependent and serially correlated heteroskedastic idiosyncratic innovations generated according to (41)-(42) with $a_\varepsilon = 0.6$. The knowledge of lag orders is not used in the estimation stage and the integer part of $T^{1/3}$ gives 3, 3, 4, 5 and 5 for $T = 30, 50, 100, 150$ and 200, respectively.

Table 13: Monte Carlo Estimates of Bias, RMSE, Size and Power for Estimation of $\varphi_1 = E(\varphi_{i1})$

DGP is an ARDL(1,0) model with homogeneous long-run coefficients, heterogeneous short-run coefficients, $\varphi_{\max} = 0.6$, stationary regressors, $m = 2$ factors, no feedback effects and $\rho_{\varepsilon i} = 0$.

(N,T)	Bias (×100)					Root Mean Square Errors (×100)					Size (5% level, $H_0: \varphi_1 = 0.3$)					Power (5% level, $H_1: \varphi_1 = 0.4$)				
	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
Imposing CS-DL pooled estimate of long-run coefficient																				
30	-8.38	-4.35	-1.90	-1.01	-0.86	10.26	6.22	4.18	3.68	3.64	51.45	27.60	12.50	9.10	9.35	96.25	95.00	93.45	90.70	90.80
50	-8.89	-4.78	-2.13	-1.36	-0.98	10.00	5.88	3.55	3.13	2.86	70.70	42.25	16.15	12.35	9.25	99.50	99.70	99.60	98.85	99.10
100	-9.27	-4.92	-2.30	-1.49	-1.15	9.85	5.51	3.10	2.44	2.24	91.55	64.60	24.90	15.75	11.70	100.00	100.00	100.00	100.00	100.00
150	-9.36	-5.10	-2.40	-1.48	-1.16	9.79	5.48	2.92	2.18	1.97	98.15	82.40	36.15	19.75	14.55	100.00	100.00	100.00	100.00	100.00
200	-9.37	-5.09	-2.36	-1.55	-1.10	9.72	5.40	2.76	2.08	1.74	99.15	89.95	45.75	24.10	16.00	100.00	100.00	100.00	100.00	100.00
Infeasible estimator: Imposing knowledge of long run coefficients																				
30	-7.95	-4.01	-1.77	-0.94	-0.78	9.48	5.73	4.00	3.53	3.53	48.40	23.30	11.55	8.25	8.50	97.25	95.85	93.10	91.65	91.15
50	-8.42	-4.41	-1.98	-1.29	-0.88	9.31	5.43	3.38	3.01	2.77	70.60	36.55	14.45	11.00	8.00	99.80	99.85	99.65	98.90	99.20
100	-8.73	-4.50	-2.13	-1.44	-1.07	9.21	5.05	2.91	2.36	2.15	93.00	60.65	22.30	14.25	10.85	100.00	100.00	100.00	100.00	100.00
150	-8.77	-4.67	-2.21	-1.41	-1.08	9.13	5.03	2.71	2.10	1.88	98.20	77.70	31.50	18.25	13.00	100.00	100.00	100.00	100.00	100.00
200	-8.83	-4.67	-2.18	-1.48	-1.01	9.12	4.95	2.57	2.00	1.65	99.30	86.65	38.80	22.30	14.10	100.00	100.00	100.00	100.00	100.00
Unconstrained CS-ARDL approach																				
30	-12.76	-6.34	-2.78	-1.58	-1.20	13.99	7.73	4.63	3.83	3.70	74.60	42.30	16.65	10.75	10.50	99.30	98.15	95.80	93.50	92.40
50	-13.34	-6.79	-3.04	-1.96	-1.37	14.11	7.59	4.14	3.40	3.01	92.00	63.95	24.05	14.95	10.55	99.95	99.95	99.80	99.30	99.60
100	-13.71	-6.96	-3.22	-2.13	-1.57	14.12	7.38	3.82	2.87	2.47	99.45	88.25	40.65	23.45	15.25	100.00	100.00	100.00	100.00	100.00
150	-13.77	-7.14	-3.30	-2.09	-1.59	14.10	7.41	3.68	2.63	2.23	99.85	96.95	57.80	30.35	21.40	100.00	100.00	100.00	100.00	100.00
200	-13.83	-7.13	-3.27	-2.17	-1.52	14.10	7.35	3.56	2.56	2.02	100.00	99.55	67.90	40.00	23.70	100.00	100.00	100.00	100.00	100.00

Notes: The dependent variable and regressors are generated according to (37)-(38) with correlated fixed effects, and with cross-sectionally weakly dependent and serially correlated heteroskedastic idiosyncratic innovations generated according to (41)-(42) with $a_\varepsilon = 0.6$.

A Appendix

We start by briefly summarizing the notations used in the paper, and introduce new notations which will prove useful in the proofs provided below. We use $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}'\mathbf{b}$ to denote the inner product (corresponding to the Euclidean norm) of vectors $\mathbf{a}$ and $\mathbf{b}$. $\|\mathbf{A}\|_1 \equiv \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}|$, and $\|\mathbf{A}\|_\infty \equiv \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|$ denote the maximum absolute column and row sum norms of $\mathbf{A} \in \mathbb{M}^{n \times n}$, respectively, where $\mathbb{M}^{n \times n}$ is the space of real-valued $n \times n$ matrices. $\|\mathbf{A}\| = \sqrt{\varrho\left(\mathbf{A}'\mathbf{A}\right)}$ is the spectral norm of $\mathbf{A}$, $\varrho\left(\mathbf{A}\right)$ is the spectral radius of $\mathbf{A}$, $Col\left(\mathbf{A}\right)$ denotes the space spanned by the column vectors of $\mathbf{A}$, and $\mathbf{A}^+$ is the Moore-Penrose pseudoinverse of $\mathbf{A}$. Note that $\|\mathbf{a}\| = \sqrt{\varrho\left(\mathbf{a}'\mathbf{a}\right)} = \sqrt{\mathbf{a}'\mathbf{a}}$ corresponds to the Euclidean length of vector $\mathbf{a}$.

Let $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$, $\bar{\mathbf{z}}_{wt} = (\bar{y}_{wt}, \bar{\mathbf{x}}'_{wt})' = \sum_{i=1}^{N} w_i \mathbf{z}_{it}$, $\Delta = (1-L)$, L is the lag operator,

$$
\underset{T-p \times 1}{\mathbf{y}_i} = \begin{pmatrix} y_{i,p+1} \\ y_{i,p+2} \\ \vdots \\ y_{iT} \end{pmatrix}, \quad \underset{T-p \times k}{\mathbf{X}_i} = \begin{pmatrix} \mathbf{x}'_{i,p+1} \\ \mathbf{x}'_{i,p+2} \\ \vdots \\ \mathbf{x}'_{iT} \end{pmatrix}, \quad \underset{(T-p) \times pk}{\Delta \mathbf{X}_{ip}} = \begin{pmatrix} \Delta\mathbf{x}'_{i,p+1} & \Delta\mathbf{x}'_{i,p} & \cdots & \Delta\mathbf{x}'_{i2} \\ \Delta\mathbf{x}'_{i,p+2} & \Delta\mathbf{x}'_{i,p+1} & \cdots & \Delta\mathbf{x}'_{i3} \\ \vdots & \vdots & & \vdots \\ \Delta\mathbf{x}'_{iT} & \Delta\mathbf{x}'_{i,T-1} & \cdots & \Delta\mathbf{x}'_{i,T-p+1} \end{pmatrix},
$$

$$
\underset{(T-p) \times k+1}{\bar{\mathbf{Z}}_w} = \begin{pmatrix} \bar{\mathbf{z}}'_{i,p+1} \\ \bar{\mathbf{z}}'_{i,p+2} \\ \vdots \\ \bar{\mathbf{z}}'_{iT} \end{pmatrix}, \quad \underset{(T-p) \times pk}{\Delta \bar{\mathbf{X}}_{wp}} = \begin{pmatrix} \Delta\bar{\mathbf{x}}'_{w,p+1} & \Delta\bar{\mathbf{x}}'_{w,p} & \cdots & \Delta\bar{\mathbf{x}}'_{w2} \\ \Delta\bar{\mathbf{x}}'_{w,p+2} & \Delta\bar{\mathbf{x}}'_{w,p+1} & \cdots & \Delta\bar{\mathbf{x}}'_{w3} \\ \vdots & \vdots & & \vdots \\ \Delta\bar{\mathbf{x}}'_{wT} & \Delta\bar{\mathbf{x}}'_{w,T-1} & \cdots & \Delta\bar{\mathbf{x}}'_{w,T-p+1} \end{pmatrix}, \quad \underset{T-p \times k}{\mathbf{V}_i} = \begin{pmatrix} \mathbf{v}'_{i,p+1} \\ \mathbf{v}'_{i,p+2} \\ \vdots \\ \mathbf{v}'_{iT} \end{pmatrix}
$$

$\mathbf{Q}_{wi} = (\mathbf{Q}_w, \Delta\mathbf{X}_{ip})$, $\mathbf{Q}_w = (\bar{\mathbf{Z}}_w, \Delta\bar{\mathbf{X}}_{wp})$,

$$
\mathbf{M}_{qi} = \mathbf{I}_{T-p} - \mathbf{Q}_{wi}\left(\mathbf{Q}'_{wi}\mathbf{Q}_{wi}\right)^+ \mathbf{Q}'_{wi}, \tag{A.1}
$$

$\boldsymbol{\gamma}_{ip} = (\boldsymbol{\gamma}'_i, \varphi_i\boldsymbol{\gamma}'_i, ..., \varphi_i^p\boldsymbol{\gamma}'_i)'$,

$$
\underset{T-p \times mp}{\mathbf{F}_p} = \left(\mathbf{F}_{(0)}, \mathbf{F}_{(1)}, ..., \mathbf{F}_{(p)}\right), \quad \underset{T-p \times m}{\mathbf{F}_{(\ell)}} \begin{pmatrix} \mathbf{f}'_{p+1-\ell} \\ \mathbf{f}'_{p+2-\ell} \\ \vdots \\ \mathbf{f}'_{T-\ell} \end{pmatrix}, \text{ for } \ell = 0, 1, 2, ..., p, \text{ and } \boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{i,p+1} \\ \varepsilon_{i,p+2} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}. \tag{A.2}
$$

Using the above notations, the model for the dependent variable can be written as

$$
\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\theta}_i + \Delta\mathbf{X}_{ip}\boldsymbol{\alpha}_{ip} + \mathbf{F}_p\boldsymbol{\gamma}_{ip} + \boldsymbol{\vartheta}_i + \boldsymbol{\varepsilon}_i,
$$

for $i = 1, 2, ..., N$, where $\boldsymbol{\alpha}_{ip}$ is a $pk \times 1$ vector containing the first p coefficients vectors of the

polynomial $\boldsymbol{\alpha}_i\left(L\right)$ stacked into one single column vector, $\boldsymbol{\vartheta}_i=\left(\vartheta_{i,p+1},\vartheta_{i,p+1},...,\vartheta_{i,T}\right)'$, and

$$\vartheta_{it}=\sum_{\ell=p+1}^{\infty}\varphi_i^{\ell+1}\left(\boldsymbol{\beta}_i'\Delta\mathbf{x}_{i,t-\ell+1}+\gamma_if_{t-\ell}\right),$$

for $i=1,2,...,N$ and $t=p+1,p+2,...,T$. The model for regressors can be written as

$$\mathbf{X}_i=\mathbf{F}_{(0)}\boldsymbol{\Gamma}_i+\mathbf{V}_i,$$

for $i=1,2,...,N$.

Define also the following projection matrix

$$\underset{T-p\times T-p}{\mathbf{M}_{hi}}=\mathbf{I}_{T-p}-\mathbf{H}_{wi}\left(\mathbf{H}_{wi}'\mathbf{H}_{wi}\right)^{+}\mathbf{H}_{wi}', \tag{A.3}$$

in which

$$\underset{T-p\times k(p+2)+1}{\mathbf{H}_{wi}}=\left(\mathbf{H}_w,\Delta\mathbf{X}_{ip}\right),\quad\underset{T-p\times k(p+1)+1}{\mathbf{H}_w}=\begin{pmatrix}\mathbf{h}_{wp,p+1}'\\\mathbf{h}_{w,p+2}'\\\vdots\\\mathbf{h}_{wp,T}'\end{pmatrix},$$

and

$$\underset{k(p+1)+1\times1}{\mathbf{h}_{wpt}}=\begin{pmatrix}\bar{\boldsymbol{\theta}}_w'\bar{\boldsymbol{\Gamma}}_w'-\boldsymbol{\alpha}_w'\left(L\right)\bar{\boldsymbol{\Gamma}}_w'+\boldsymbol{\gamma}_w'\left(L\right)\\\bar{\boldsymbol{\Gamma}}_w'\\\left(1-L\right)\bar{\boldsymbol{\Gamma}}_w'\\L\left(1-L\right)\bar{\boldsymbol{\Gamma}}_w'\\\vdots\\L^{p-1}\left(1-L\right)\bar{\boldsymbol{\Gamma}}_w'\end{pmatrix}\mathbf{f}_t,$$

where

$$\bar{\boldsymbol{\theta}}_w=\sum_{i=1}^{N}w_i\boldsymbol{\theta}_i,\ \bar{\boldsymbol{\Gamma}}_w=\sum_{i=1}^{N}w_i\boldsymbol{\Gamma}_i,\ \boldsymbol{\alpha}_w\left(L\right)=\sum_{i=1}^{N}w_i\boldsymbol{\alpha}_i\left(L\right),\ \boldsymbol{\gamma}_w\left(L\right)=\sum_{i=1}^{N}w_i\boldsymbol{\gamma}_i\left(L\right),$$

and $\boldsymbol{\gamma}_i\left(L\right)=\sum_{\ell=0}^{\infty}\varphi_i^p\boldsymbol{\gamma}_iL^p$.

A.1 Proofs of Theorems

Proof of Theorem 1. We have

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{MG}-\boldsymbol{\theta}\right)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{v}_{\theta i}+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\boldsymbol{\Psi}}_{iT}^{-1}\frac{\mathbf{X}_i'\mathbf{M}_{qi}\mathbf{F}_p\boldsymbol{\gamma}_{ip}}{T}+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\boldsymbol{\Psi}}_{iT}^{-1}\frac{\mathbf{X}_i'\mathbf{M}_{qi}\boldsymbol{\vartheta}_i}{T}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\boldsymbol{\Psi}}_{iT}^{-1}\frac{\mathbf{X}_i'\mathbf{M}_{qi}\boldsymbol{\varepsilon}_i}{T} \tag{A.4}$$

where $\widehat{\boldsymbol{\Psi}}_{iT} = T^{-1}\mathbf{X}'_i\mathbf{M}_{qi}\mathbf{X}_i$,

$$\underset{T-p\times m(p+1)}{\mathbf{F}_p} = \begin{pmatrix} \mathbf{f}'_{p+1} & \mathbf{f}'_p & \cdots & \mathbf{f}'_1 \\ \vdots & \vdots & & \vdots \\ \mathbf{f}'_T & \mathbf{f}'_{T-1} & \cdots & \mathbf{f}'_{T-p} \end{pmatrix},$$

$\boldsymbol{\gamma}_{ip} = (\boldsymbol{\gamma}'_i, \varphi_i\boldsymbol{\gamma}'_i, ..., \varphi_i^p\boldsymbol{\gamma}'_i)'$, $\boldsymbol{\vartheta}_i = (\vartheta_{i,p+1}, \vartheta_{i,p+1}, ..., \vartheta_{i,T})'$, and

$$\vartheta_{it} = \sum_{\ell=p+1}^{\infty} \varphi_i^{\ell+1}\left(\boldsymbol{\beta}'_i\Delta\mathbf{x}_{i,t-\ell+1} + \gamma_i f_{t-\ell}\right).$$

Consider the asymptotics $(N, T, p) \xrightarrow{j} \infty$ such that $\sqrt{N}p\rho^p \to 0$, for any constant $0 < \rho < 1$ and $p^3/T \to \varkappa$, $0 < \varkappa < \infty$. In what follows we establish convergence of the individual terms on the right side of (A.4).

It follows from (A.26) of Lemma A.1 and (A.27) of Lemma A.2 that

$$\widehat{\boldsymbol{\Psi}}_{\Xi,iT} - \boldsymbol{\Sigma}_i = o_p\left(N^{-1/2}\right) \text{ uniformly in } i. \tag{A.5}$$

(A.5), (A.28) of Lemma A.2, and (A.30) of Lemma A.3 imply

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\boldsymbol{\Psi}}_{iT}^{-1}\frac{\mathbf{X}'_i\mathbf{M}_{qi}\boldsymbol{\varepsilon}_i}{T} \xrightarrow{p} \underset{k\times 1}{\mathbf{0}}. \tag{A.6}$$

Consider now the second term on the right side of (A.4), which involves common factors and their loadings. In the previous literature on CCE estimators, Pesaran (2006) established the asymptotic results for the term involving factors and their loadings in the expression for his CCEMG estimator by focusing on the properties of the matrix (using Pesaran (2006)'s notations) $\mathbf{X}'_i\bar{\mathbf{M}}_w\mathbf{F}/T$, see equation (40) in Pesaran (2006), in the full rank case, and by exploring the relation (still using Pesaran (2006)'s notations) $\mathbf{M}_q\mathbf{F}\bar{\mathbf{C}}_w = \mathbf{0}$, see p. 979 of Pesaran (2006), in the rank deficient case. But unlike in the set-up of Pesaran (2006), the dimension of $\mathbf{X}'_i\mathbf{M}_{qi}\mathbf{F}_p/T$ in this paper increases with the sample size, and furthermore $\mathbf{M}_{hi}\mathbf{F}_p\bar{\boldsymbol{\gamma}}_{wp}$ is not necessarily zero since $\mathbf{F}_p\bar{\boldsymbol{\gamma}}_{wp}$ (due to the truncation lag p) does not necessarily belong to the linear space spanned by the column vectors of $\mathbf{H}_{wi}$. We therefore focus on the elements of the vector $\mathbf{X}'_i\mathbf{M}_{qi}\mathbf{F}_p\boldsymbol{\gamma}_{ip}/T$ below, which has fixed (finite) dimensions, and we also take advantage of the exponential decay of certain coefficients below. Using (A.5), boundedness of $\boldsymbol{\Sigma}_i^{-1}$ (by Assumption 5), and the result (A.29) of Lemma A.2 we obtain

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\mathbf{X}'_i\mathbf{M}_{qi}\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}'_i\mathbf{M}_{qi}\mathbf{F}_p}{T}\boldsymbol{\gamma}_{ip} - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\mathbf{X}'_i\mathbf{M}_{hi}\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}'_i\mathbf{M}_{hi}\mathbf{F}_p}{T}\boldsymbol{\gamma}_{ip} \xrightarrow{p} \underset{k\times 1}{\mathbf{0}}.$$

Vector $\boldsymbol{\gamma}_{ip}$ can be written as $\boldsymbol{\gamma}_{ip} = \left(\overline{\boldsymbol{\gamma}}_{wp} - \overline{\boldsymbol{\eta}}_{\gamma wp}\right) + \boldsymbol{\eta}_{\gamma ip}$, and

$$T^{-1}\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}_p\boldsymbol{\gamma}_{ip} = T^{-1}\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}_p\overline{\boldsymbol{\gamma}}_{wp} + T^{-1}\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}_p\left(\boldsymbol{\eta}_{\gamma ip} - \overline{\boldsymbol{\eta}}_{\gamma wp}\right).$$

Note again that $\mathbf{F}_p\overline{\boldsymbol{\gamma}}_{wp}$ does not necessarily belong to the linear space spanned by the column vectors of $\mathbf{H}_{wi}$ due to the truncation lag p. But Assumption 4 constrains the support of φ_i to fall strictly within the unit circle, which implies that there exists a positive constant $\rho < 1$ such that $|\varphi_i| < \rho < 1$ for all possible realizations of the random variable φ_i. Therefore, under Assumptions 3-4, the coefficients in the polynomials $\boldsymbol{\alpha}_w(L) = \sum_{i=1}^N w_i \boldsymbol{\alpha}_i(L)$ and $\boldsymbol{\gamma}_w(L) = \sum_{i=1}^N w_i \boldsymbol{\gamma}_i(L)$, where $\boldsymbol{\alpha}_i(L) = \sum_{\ell=0}^\infty \varphi_i^{\ell+1}(1-\varphi_i)^{-1}\boldsymbol{\beta}_i L^\ell$ and $\boldsymbol{\gamma}_i(L) = \sum_{\ell=0}^\infty \varphi_i^p \boldsymbol{\gamma}_i L^p$, decay exponentially to zero[8] and we have

$$\overline{\boldsymbol{\gamma}}_w'(L,p)\mathbf{f}_t - E\left[\overline{\boldsymbol{\gamma}}_w'(L,p)\mathbf{f}_t \,|\mathbf{h}_{wpt}\right] = O_p\left(\rho^p\right), \tag{A.7}$$

uniformly in t, where $\overline{\boldsymbol{\gamma}}_w(L,p) = \sum_{\ell=0}^p \sum_{i=1}^N w_i \varphi_i^\ell \boldsymbol{\gamma}_i L^\ell$ is the truncated polynomial of $\overline{\boldsymbol{\gamma}}_w(L)$ featuring only orders up to L^p. Using the properties of orthogonal projectors, we obtain[9]

$$\left\|\mathbf{M}_{hi}\mathbf{F}_p\overline{\boldsymbol{\gamma}}_{wp}\right\| \leq \left\|\mathbf{F}_p\overline{\boldsymbol{\gamma}}_{wp} - \mathbf{H}_{wi}\mathbf{c}\right\|, \tag{A.8}$$

for any $k(p+1)+1 \times 1$ vector $\mathbf{c}$. Let $\mathbf{c}$ be defined by $E\left[\overline{\boldsymbol{\gamma}}_w'(L,p)\mathbf{f}_t\,|\mathbf{h}_{wpt}\right] = \mathbf{c}'\mathbf{h}_{wpt}$. Then it follows from (A.7) that the individual elements of $T-p \times 1$ vector $\left(\mathbf{F}_p\overline{\boldsymbol{\gamma}}_{wp} - \mathbf{H}_{wi}\mathbf{c}\right)$ are uniformly $O_p\left(\rho^p\right)$ and using (A.8) we have

$$\left\|\mathbf{M}_{hi}\mathbf{F}_p\overline{\boldsymbol{\gamma}}_{wp}\right\| = O_p\left[(T-p)^{1/2}\rho^p\right].$$

Now using Cauchy-Schwarz inequality, we obtain[10]

$$T^{-1}\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}_p\overline{\boldsymbol{\gamma}}_{wp} = O_p\left(\rho^p\right). \tag{A.9}$$

Noting that $\sqrt{N}\rho^p \to 0$, and using (A.5) and boundedness of $\boldsymbol{\Sigma}_i^{-1}$ (by Assumption 5) we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \left(\frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}_p}{T}\overline{\boldsymbol{\gamma}}_{wp} \xrightarrow{p} \mathbf{0},$$

---

[8]See Pesaran and Chudik (2014) for a related discussion.

[9]We use the following property. Let $\mathbf{A}$ be a $s_1 \times s_2$ dimensional matrix, $s_1 > s_2$, and let $\mathbf{M}_A = \mathbf{I}_{s_1} - \mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^+\mathbf{A}'$ be the corresponding orthogonal projector that projects on orthogonal complement of the space spanned by the column vectors of $\mathbf{A}$. Then for any $s_1 \times 1$ dimensional vector $\mathbf{x}$ and any $s_2 \times 1$ dimensional vector $\mathbf{c}$, $\left\|\mathbf{M}_A\mathbf{x}\right\| \leq \left\|\mathbf{x} - \mathbf{A}\mathbf{c}\right\|$.

[10]$\langle\mathbf{a},\mathbf{b}\rangle \leq \|\mathbf{a}\|\|\mathbf{b}\|$. We set $\mathbf{a} = T^{-1}\mathbf{X}_i$, and $\mathbf{b} = \mathbf{M}_h\mathbf{F}_p\overline{\boldsymbol{\gamma}}_{pw}$, where $\|\mathbf{a}\| = O_p\left[(T-p)^{-1/2}\right]$, and $\|\mathbf{b}\| = O_p\left[(T-p)^{1/2}\rho^p\right]$.

and it then follows that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \widehat{\boldsymbol{\Psi}}_{\Xi,iT}^{-1} \frac{\mathbf{X}_i'\mathbf{M}_{qi}\mathbf{F}_p}{T} \boldsymbol{\gamma}_{ip} - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{X}_i}{T} \right)^{-1} \frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}_p}{T} \left(\boldsymbol{\eta}_{\gamma ip} - \overline{\boldsymbol{\eta}}_{\gamma wp} \right) \xrightarrow{p} \underset{k\times 1}{\mathbf{0}}. \quad \text{(A.10)}$$

Now consider the term $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{X}_i}{T} \right)^{-1} \frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}_p}{T} \overline{\boldsymbol{\eta}}_{\gamma wp}$. Let us denote individual columns of $\mathbf{F}_p$ as $\mathbf{f}_{p,[j]}$, for $j = 1, 2, ..., mp$, and individual elements of $\overline{\boldsymbol{\eta}}_{\gamma wp}$ and $\overline{\boldsymbol{\gamma}}_{wp}$ as $\overline{\eta}_{\gamma wp,j}$ and $\overline{\gamma}_{wp,j}$, respectively, for $j = 1, 2, ..., mp$. $\mathbf{F}_p \overline{\boldsymbol{\eta}}_{\gamma wp}$ thus can be written as $\sum_{j=1}^{mp} \mathbf{f}_{p,[j]} \overline{\eta}_{\gamma wp,j}$. Let

$$\pi_j = \frac{\overline{\eta}_{\gamma wp,j}}{\gamma_{p,j} + \overline{\eta}_{\gamma wp,j}},$$

where $\gamma_{p,j}$ is the j-th element of the vector $E\left(\boldsymbol{\gamma}_{ip}\right)$. Note that $p\lim_{N\to\infty} \pi_j = 1$ if $\gamma_{p,j} = 0$ and $p\lim_{N\to\infty} \pi_j = 0$ if $\gamma_{p,j} \neq 0$. Expression $\mathbf{F}_p \overline{\boldsymbol{\eta}}_{ywp}$ can now be written as $\mathbf{F}_p \overline{\boldsymbol{\eta}}_{ywp} = \sum_{j=1}^{mp} \mathbf{f}_{p,[j]} \overline{\gamma}_{wp,j} \pi_j$ and

$$\frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}_p}{T} \overline{\boldsymbol{\eta}}_{ywp} = \sum_{j=1}^{mp} \frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{f}_{p,[j]}}{T} \overline{\gamma}_{wp,j} \pi_j.$$

Using the same arguments as in the derivation of (A.9), we obtain $\frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{f}_{p,[j]}}{T} \overline{\gamma}_{wp,j} = O_p\left(\rho^p\right)$ and using the properties of π_j we have

$$\sum_{j=1}^{mp} \frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{f}_{p,[j]}}{T} \overline{\gamma}_{wp,j} \pi_j = O_p\left(p\rho^p\right).$$

But $\sqrt{N}p\rho^p \to 0$ and therefore

$$\sqrt{N} \frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}_p}{T} \overline{\boldsymbol{\eta}}_{ywp} \xrightarrow{p} \underset{k\times 1}{\mathbf{0}}. \quad \text{(A.11)}$$

Using this result in (A.10) together with (A.5) and the boundedness of $\left\| \boldsymbol{\Sigma}_i^{-1} \right\|$ we obtain

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \widehat{\boldsymbol{\Psi}}_{\Xi,iT}^{-1} \frac{\mathbf{X}_i'\mathbf{M}_{qi}\mathbf{F}_p}{T} \boldsymbol{\gamma}_i - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{X}_i}{T} \right)^{-1} \frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}_p}{T} \boldsymbol{\eta}_{\gamma ip} \xrightarrow{p} \underset{k\times 1}{\mathbf{0}}. \quad \text{(A.12)}$$

Consider now the third term on the right side of (A.4). Let $\tilde{\mathbf{x}}_{it}$ denote the column $(t - p)$ of the matrix $\mathbf{X}_i'\mathbf{M}_{qi}$, for $t = p + 1, p + 2, ..., T$. We have $\tilde{\mathbf{x}}_{it} = O_p(1)$ uniformly in i, $\widehat{\boldsymbol{\Psi}}_{iT}^{-1} = O_p(1)$ uniformly in i, and

$$E\left|\sqrt{N}\vartheta_{it}\right| \leq \sqrt{N} \sum_{\ell=p+1}^{\infty} |\varphi_i|^{\ell+1} E\left|\boldsymbol{\beta}_i'\Delta\mathbf{x}_{i,t-\ell+1} + \gamma_i f_{t-\ell}\right| < K\sqrt{N}\rho^p, \quad \text{(A.13)}$$

uniformly in i and t. It follows that $E\left|\sqrt{N}\vartheta_{it}\right| \xrightarrow{p} 0$ as $\sqrt{N}\rho^p \to 0$,

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}\vartheta_{it} \xrightarrow{p} 0 \text{ uniformly in } i, \tag{A.14}$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{\Psi}}_{iT}^{-1}\left(\frac{\mathbf{X}_i'\overline{\mathbf{M}}_{wi}\left(\sqrt{N}\vartheta_i\right)}{T}\right) \xrightarrow{p} \underset{k\times1}{\mathbf{0}}. \tag{A.15}$$

Using (A.6), (A.12) and (A.15) in (A.4), we obtain

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{MG} - \boldsymbol{\theta}\right) \overset{d}{\sim} \vartheta_{\theta i}, \text{,}$$

where

$$\vartheta_{\theta i} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{v}_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}_p}{T}\boldsymbol{\eta}_{\gamma ip}, \tag{A.16}$$

and recall that $\boldsymbol{v}_i$ and $\boldsymbol{\eta}_{\gamma ip}$ are independently distributed across i. It now follows that when $\boldsymbol{\eta}_{\gamma i}$ is independently distributed from $\boldsymbol{\Gamma}_i$ and regardless whether the rank condition holds, $\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{MG} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(\underset{k\times1}{\mathbf{0}}, \boldsymbol{\Sigma}_{MG}\right)$, where

$$\boldsymbol{\Sigma}_{MG} = \boldsymbol{\Omega}_\theta + \lim_{p,N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{i\xi}^{-1}\mathbf{Q}_{if}\boldsymbol{\Omega}_\gamma\mathbf{Q}_{if}'\boldsymbol{\Sigma}_{i\xi}^{-1}\right], \tag{A.17}$$

in which $\boldsymbol{\Omega}_\theta = Var\left(\boldsymbol{\theta}_i\right)$, $\boldsymbol{\Omega}_\gamma = Var\left(\boldsymbol{\gamma}_i\right)$, and $\boldsymbol{\Sigma}_i = p\lim T^{-1}\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{X}_i$ and $\mathbf{Q}_{if} = p\lim T^{-1}\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}$. When the rank condition stated in assumptions of Theorem 1 holds then $\mathbf{Q}_{if} = \underset{k\times m}{\mathbf{0}}$, and therefore even if $\boldsymbol{\eta}_{\gamma i}$ is correlated with $\boldsymbol{\Gamma}_i$, $\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{MG} - \boldsymbol{\theta}\right) \overset{d}{\sim} \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{v}_i$. Consistency of the nonparametric estimator can be established in the same way as in Chudik and Pesaran (2014a). ∎

Proof of Theorem 2. Consider

$$\left(\sum_{i=1}^{N}w_i^2\right)^{-1/2}\left(\widehat{\boldsymbol{\theta}}_P - \boldsymbol{\theta}\right) = \left(\sum_{i=1}^{N}w_i\frac{\mathbf{X}_i'\mathbf{M}_{qi}\mathbf{X}_i}{T}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{w}_i\frac{\mathbf{X}_i'\mathbf{M}_{qi}\left(\mathbf{X}_i\mathbf{v}_i + \mathbf{F}_p\boldsymbol{\gamma}_{ip} + \vartheta_i + \boldsymbol{\varepsilon}_i\right)}{T}, \tag{A.18}$$

where ϑ_i is defined below (A.4), $\widetilde{w}_i = \sqrt{N}w_i\left(\sum_{i=1}^{N}w_i^2\right)^{-1/2}$, and, by granularity conditions (20)-(21) there exists a constant $K < \infty$ (independent of i and N), such that

$$|\widetilde{w}_i| = \left|\sqrt{N}w_i\left(\sum_{i=1}^{N}w_i^2\right)^{-1/2}\right| < K. \tag{A.19}$$

We focus on the individual terms on the right side of (A.18) below and assume that $(N,T,p) \xrightarrow{j} \infty$

such that $\sqrt{N}p\rho^p \to 0$ for any constant $0 < \rho < 1$ and $p^3/T \to \varkappa$, $0 < \varkappa < \infty$.

Using results (A.26) of Lemma A.1 we have

$$\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i'\mathbf{M}_{qi}\mathbf{X}_i}{T} - \sum_{i=1}^{N} w_i \mathbf{\Sigma}_{iq} \overset{p}{\to} \underset{k \times 1}{\mathbf{0}},$$

for any weights $\{w_i\}$ satisfying granularity conditions (20)-(21). The limit $\lim_{N\to\infty} \sum_{i=1}^{N} w_i \mathbf{\Sigma}_{iq} = \mathbf{\Psi}^*$ exists by Assumption 5 and furthermore, by the same assumption, $\mathbf{\Psi}^*$ is nonsingular. It therefore follows that

$$\left(\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i'\mathbf{M}_{qi}\mathbf{X}_i}{T} \right)^{-1} \overset{p}{\to} \mathbf{\Psi}^{*-1}. \tag{A.20}$$

Noting that $\boldsymbol{\gamma}_{ip}$ can be written as $\boldsymbol{\gamma}_{ip} = \overline{\boldsymbol{\gamma}}_{wp} + \boldsymbol{\eta}_{ip} - \overline{\boldsymbol{\eta}}_{wp}$, and using (A.9), (A.11), (A.19) and $\sqrt{N}\rho^p \to 0$ we obtain[11]

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \widetilde{w}_i \frac{\mathbf{X}_i'\mathbf{M}_{qi}\mathbf{F}_p}{T}\boldsymbol{\gamma}_{ip} - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \widetilde{w}_i \frac{\mathbf{X}_i'\mathbf{M}_{qi}\mathbf{F}_p}{T}\boldsymbol{\eta}_{ip} \overset{p}{\to} \underset{k \times 1}{\mathbf{0}}. \tag{A.21}$$

(A.14) and (A.19) imply

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \widetilde{w}_i \frac{\mathbf{X}_i'\mathbf{M}_{qi}\boldsymbol{\vartheta}_i}{T} \overset{p}{\to} \underset{k \times 1}{\mathbf{0}}. \tag{A.22}$$

Result (A.28) of Lemma A.2 and result (A.30) of Lemma A.3 establish

$$\sqrt{N}\frac{\mathbf{X}_i'\mathbf{M}_{qi}\boldsymbol{\varepsilon}_i}{T} \overset{p}{\to} \underset{k \times 1}{\mathbf{0}} \text{ uniformly in } i,$$

and therefore (noting that $\widetilde{w}_i$ is uniformly bounded in i, see (A.19)),

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \widetilde{w}_i \frac{\mathbf{X}_i'\mathbf{M}_{qi}\boldsymbol{\varepsilon}_i}{T} = \frac{1}{N} \sum_{i=1}^{N} \widetilde{w}_i \left(\sqrt{N}\frac{\mathbf{X}_i'\mathbf{M}_{qi}\boldsymbol{\varepsilon}_i}{T} \right) \overset{p}{\to} \underset{k \times 1}{\mathbf{0}}. \tag{A.23}$$

Using (A.20), (A.21), (A.22), (A.23) and result (A.27) of Lemma A.2 in (A.18), we obtain

$$\left(\sum_{i=1}^{N} w_i^2 \right)^{-1/2} \left(\widehat{\boldsymbol{\theta}}_P - \boldsymbol{\theta} \right) \overset{d}{\sim} \mathbf{\Psi}^{*-1}\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \widetilde{w}_i \frac{\mathbf{X}_i'\mathbf{M}_{hi}\left(\mathbf{X}_i\mathbf{v}_i + \mathbf{F}_p\boldsymbol{\eta}_{ip} \right)}{T}.$$

Assumption 5 is sufficient for the bounded second moments of $\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{X}_i/T$ and $\mathbf{X}_i'\mathbf{M}_{hi}\mathbf{F}_p/T$. In particular, condition $E\left(\widetilde{x}_{ist}^4 \right) < K$, for $s = 1, 2, .., k$, is sufficient for the bounded second moment

---

[11](A.21) can also be established by noting that the column vectors of $\mathbf{X}_w = \sum_{i=1}^{N} w_i\mathbf{X}_i$ are included in $\mathbf{Q}_{wi}$ and therefore $\mathbf{X}_w'\mathbf{M}_{qi} = \mathbf{0}$.

of $\mathbf{X}'_i\mathbf{M}_{hi}\mathbf{X}_i/T$. To see this note that

$$\frac{\mathbf{X}'_i\mathbf{M}_{hi}\mathbf{X}_i}{T} = \frac{1}{T}\sum_{t=1}^{T}\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}'_{it},$$

and, by Minkowski's inequality,

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\widetilde{x}_{ist}\widetilde{x}'_{ipt}\right\|_{L_2} \le \frac{1}{T}\sum_{t=1}^{T}\left\|\widetilde{x}_{ist}\widetilde{x}'_{ipt}\right\|_{L_2},$$

for any $s, p = 1, 2, .., k$. But by Cauchy-Schwarz inequality, we have $E\left(\widetilde{x}^2_{ist}\widetilde{x}^2_{ipt}\right) \le \left[E\left(\widetilde{x}^4_{ist}\right)E\left(\widetilde{x}^4_{ipt}\right)\right]^{1/2}$, and therefore bounded fourth moments of the elements of $\widetilde{\mathbf{x}}_{it}$ are sufficient for the existence of an upper bound for the second moments of $\mathbf{X}'_i\mathbf{M}_{hi}\mathbf{X}_i/T$. Similar arguments can be used to establish that $\mathbf{X}'_i\mathbf{M}_{hi}\mathbf{F}_p/T$ has bounded second moments. Note also that $\mathbf{v}_i$ and $\boldsymbol{\eta}_{ip}$ are independently distributed across i; and, independently distributed of $\mathbf{M}_{hi}$, $\mathbf{F}_p$ and, assuming that $\boldsymbol{\gamma}_i$ is independently distributed of $\boldsymbol{\Gamma}_i$, also $\mathbf{X}_i$. It therefore follows, using similar arguments as in Lemma 4 of Pesaran (2006), that

$$\left(\sum_{i=1}^{N}w_i^2\right)^{-1/2}\left(\widehat{\boldsymbol{\theta}}_P - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_P\right),$$

where

$$\boldsymbol{\Sigma}_P = \boldsymbol{\Psi}^{*-1}\mathbf{R}^*\boldsymbol{\Psi}^{*-1}, \tag{A.24}$$

in which

$$\boldsymbol{\Psi}^* = \lim_{N\to\infty}\sum_{i=1}^{N}w_i\boldsymbol{\Sigma}_i, \ \mathbf{R}^* = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\widetilde{w}_i^2\left(\boldsymbol{\Sigma}_i\boldsymbol{\Omega}_\theta\boldsymbol{\Sigma}_i + \mathbf{Q}_{if}\boldsymbol{\Omega}_\gamma\mathbf{Q}'_{if}\right),$$

$\boldsymbol{\Omega}_\theta = Var\left(\boldsymbol{\theta}_i\right)$, $\boldsymbol{\Omega}_\gamma = Var\left(\boldsymbol{\gamma}_i\right)$, $\boldsymbol{\Sigma}_i = p\lim T^{-1}\mathbf{X}'_i\mathbf{M}_{hi}\mathbf{X}_i$ and $\mathbf{Q}_{if} = p\lim T^{-1}\mathbf{X}'_i\mathbf{M}_{hi}\mathbf{F}$. $\boldsymbol{\Sigma}_P$ can be estimated as

$$\widehat{\boldsymbol{\Sigma}}_P = \left(\sum_{i=1}^{N}w_i^2\right)\widehat{\boldsymbol{\Psi}}^{*-1}\widehat{\mathbf{R}}^*\widehat{\boldsymbol{\Psi}}^{*-1}, \tag{A.25}$$

where

$$\widehat{\boldsymbol{\Psi}}^* = \sum_{i=1}^{N}w_i\left(\frac{\mathbf{X}'_i\mathbf{M}_{qi}\mathbf{X}_i}{T}\right),$$

and

$$\widehat{\mathbf{R}}^* = \frac{1}{N-1}\sum_{i=1}^{N}\widetilde{w}_i^2\left(\frac{\mathbf{X}'_i\mathbf{M}_{qi}\mathbf{X}_i}{T}\right)\left(\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}}_{MG}\right)\left(\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}}_{MG}\right)'\left(\frac{\mathbf{X}'_i\bar{\mathbf{M}}_{wi}\mathbf{X}_i}{T}\right).$$

When the rank condition holds, then column vectors of $\mathbf{F}_p$ belong to the space spanned by the column vectors of $\mathbf{H}_w$, and therefore regardless whether $\boldsymbol{\eta}_{\gamma i}$ is correlated with $\boldsymbol{\Gamma}_i$ or not, $\left(\sum_{i=1}^{N}w_i^2\right)^{-1/2}\left(\widehat{\boldsymbol{\theta}}_P - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_P\right)$ in the full rank case with $\boldsymbol{\Sigma}_P$ reduced to $\boldsymbol{\Psi}^{*-1}\mathbf{R}^*_\theta\boldsymbol{\Psi}^{*-1}$ and

$\mathbf{Q}_{if} = \underset{k \times m}{\mathbf{0}}$. Consistency of $\widehat{\boldsymbol{\Sigma}}_P$ can be established using similar arguments as in Pesaran (2006). ∎

A.2 Lemmas

Lemma A.1 *Suppose Assumptions 1- 5 hold and $(N, T, p) \xrightarrow{j} \infty$ such that $p^3/T \to \varkappa$, $0 < \varkappa < \infty$. Then,*

$$\frac{\mathbf{X}_i' \mathbf{M}_{hi} \mathbf{X}_i}{T} \xrightarrow{p} \boldsymbol{\Sigma}_i, \text{ uniformly in } i. \tag{A.26}$$

Proof. Let $\boldsymbol{\xi}_{hit}'$ denote the individual rows of $\mathbf{M}_{hi}\mathbf{X}_i$ so that

$$\frac{\mathbf{X}_i' \mathbf{M}_{hi} \mathbf{X}_i}{T} = \frac{T - p}{T} \frac{1}{T - p} \sum_{t=p+1}^{T} \boldsymbol{\xi}_{hit} \boldsymbol{\xi}_{hit}'.$$

Ergodicity in mean of $\boldsymbol{\xi}_{hit}$ has been established in Chudik and Pesaran, (2014a, Lemma A3). This completes the proof of (A.26). ∎

Lemma A.2 *Suppose Assumptions 1- 5 hold and $(N, T, p) \xrightarrow{j} \infty$ such that $p^3/T \to \varkappa$, $0 < \varkappa < \infty$. Then,*

$$\sqrt{N} \frac{\mathbf{X}_i' \mathbf{M}_{qi} \mathbf{X}_i}{T} - \sqrt{N} \frac{\mathbf{X}_i' \mathbf{M}_{hi} \mathbf{X}_i}{T} \xrightarrow{p} \underset{k \times k}{\mathbf{0}}, \text{ uniformly in } i. \tag{A.27}$$

$$\sqrt{N} \frac{\mathbf{X}_i' \mathbf{M}_{qi} \boldsymbol{\varepsilon}_i}{T} - \sqrt{N} \frac{\mathbf{X}_i' \mathbf{M}_{hi} \boldsymbol{\varepsilon}_i}{T} \xrightarrow{p} \underset{k \times 1}{\mathbf{0}}, \text{ uniformly in } i. \tag{A.28}$$

$$\left\| \sqrt{N} \frac{\mathbf{X}_i' \mathbf{M}_{qi} \mathbf{F}_p}{T} - \sqrt{N} \frac{\mathbf{X}_i' \mathbf{M}_{hi} \mathbf{F}_p}{T} \right\|_1 \xrightarrow{p} 0, \text{ uniformly in } i. \tag{A.29}$$

Proof. Results (A.27) and (A.28) can be established in the same way as Chudik and Pesaran, (2014a, results A.21 and A.22 of Lemma A6). Consider now (A.29). $\mathbf{F}_p$ can be written as $\mathbf{F}_p = \left[\mathbf{F}_{(0)}, \mathbf{F}_{(1)}, ..., \mathbf{F}_{(p)} \right]$, where $\mathbf{F}_{(\ell)} = (\mathbf{f}_{p+1-\ell}, \mathbf{f}_{p+2-\ell}, ..., \mathbf{f}_{T-\ell})'$ for $\ell = 0, 1, 2, ..., p$. Using the same arguments as in Chudik and Pesaran, (2014a, results A.23 of Lemma A6), it can be shown that

$$\sqrt{N} \frac{\mathbf{X}_i' \mathbf{M}_{qi} \mathbf{F}_{(\ell)}}{T} - \sqrt{N} \frac{\mathbf{X}_i' \mathbf{M}_{hi} \mathbf{F}_{(\ell)}}{T} \xrightarrow{p} \underset{k \times m}{\mathbf{0}},$$

uniformly in i and ℓ. This is sufficient for (A.29) to hold. ∎

Lemma A.3 *Suppose Assumptions 1- 5 hold and $(N, T, p) \xrightarrow{j} \infty$ such that $p^3/T \to \varkappa$, $0 < \varkappa < \infty$. Then,*

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\mathbf{X}_i' \mathbf{M}_{hi} \boldsymbol{\varepsilon}_i}{T} \xrightarrow{p} \underset{k \times 1}{\mathbf{0}}, \text{ uniformly in } i. \tag{A.30}$$

Proof. Results (A.27) can be established in the same way as Chudik and Pesaran, (2014a, results A.26). ∎

44

References

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

Chudik, A. and M. H. Pesaran (2014a). Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors. *forthcoming in Journal of Econometrics*.

Chudik, A. and M. H. Pesaran (2014b). Large Panel Data Models with Cross-Sectional Dependence: A Survey. In B. Baltagi (Ed.), *Forthcoming in The Oxford Handbook on Panel Data*. Oxford University Press.

Mark, N. C. and D. Sul (2003). Cointegration Vector Estimation by Panel DOLS and Long-run Money Demand. *Oxford Bulletin of Economics and Statistics 65*(5), 655–680.

Moon, H. R. and M. Weidner (2010). Dynamic Linear Panel Regression Models with Interactive Fixed Effects. Mimeo, July 2010.

Pedroni, P. (2001). Fully Modified OLS for Heterogeneous Cointegrated Panels. *Advances in Econometrics 15*, 93–130.

Pesaran, M. H. (1997). The Role of Economic Theory in Modelling the Long Run. *Economic Journal 107*, 178–191.

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

Pesaran, M. H. and A. Chudik (2014). Aggregation in Large Dynamic Panels. *Journal of Econometrics 178*, 273–285.

Pesaran, M. H. and Y. Shin (1999). An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis. In S. Strom (Ed.), *Econometrics and Economic Theory in 20th Century: The Ragnar Frisch Centennial Symposium*, Chapter 11, pp. 371–413. Cambridge: Cambridge University Press.

Pesaran, M. H., Y. Shin, and R. P. Smith (1999). Pooled Mean Group Estimation of Dynamic Heterogeneous Panels. *Journal of the American Statistical Association 94*, 621–634.

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

Song, M. (2013). Asymptotic Theory for Dynamic Heterogeneous Panels with Cross-Sectional Dependence and Its Applications. Mimeo, January 2013.