Federal Reserve Bank of Dallas

Analysis of Multiple Long-Run Relations in Panel Data Models

Alexander Chudik, M. Hashem Pesaran and Ron P. Smith

Working Paper 2523 **June 2025 (Revised September 2025)**

Research Department

https://doi.org/10.24149/wp2523r2

Analysis of Multiple Long-Run Relations in Panel Data Models[*]

Alexander Chudik[†], M. Hashem Pesaran[‡] and Ron P. Smith[§]

First version: June 2, 2025
This version: September 11, 2025

Abstract

The literature on panel cointegration is extensive but does not cover data sets where the cross section dimension, n, is larger than the time series dimension T. This paper proposes a novel methodology that filters out the short run dynamics using sub-sample time averages as deviations from their full-sample counterpart, and estimates the number of long-run relations and their coefficients using eigenvalues and eigenvectors of the pooled covariance matrix of these sub-sample deviations. We refer to this procedure as pooled minimum eigenvalue (PME). We show that the PME estimator is consistent and asymptotically normal as n and $T \rightarrow \infty$ jointly, such that $T \approx n^d$, with $d > 0$ for consistency and $d > 1/2$ for asymptotic normality. Extensive Monte Carlo studies show that the number of long-run relations can be estimated with high precision, and the PME estimators have good size and power properties. The utility of our approach is illustrated by micro and macro applications using Compustat and Penn World Tables.

Keywords: Multiple long-run relations, Pooled Minimum Eigenvalue (PME) estimator, eigenvalue thresholding, panel data, cointegration, interactive time effects, financial ratios, Penn World Table

JEL Classification: C13, C23, C33, G30

---

[*]The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas or the Federal Reserve System. We gratefully acknowledge our use of computational resources provided by the Big-Tex High Performance Computing Group at the Federal Reserve Bank of Dallas.

[†]Alexander Chudik, Federal Reserve Bank of Dallas. alexander.chudik@dal.frb.org.

[‡]M. Hashem Pesaran, University of Cambridge, UK and University of Southern California, USA. mhp1@cam.ac.uk.

[§]Ron P. Smith, Birkbeck, University of London, United Kingdom, r.smith@bbk.ac.uk.

1 Introduction

This paper provides a new methodology for the analysis of multiple long-run relations in panel data models where the cross section dimension, n, is large relative to the time series dimension, T. While there is an extensive literature that considers multiple long-run (cointegrating) relations for time series models, for panel data models with large n researchers have mainly focussed on a single long-run relation with known long-run causal links. The panel literature that does consider multiple long-run relations assumes n is fixed as $T \to \infty$, or adopt sequential asymptotics whereby $T \to \infty$ first followed by $n \to \infty$, effectively requiring T to be large relative to n and do not cover many applications of interest in economics and finance that involve many cross section units, such as firms and countries, observed over relatively short time spans. One example is empirical corporate finance, which investigates the stability of long-run relations, including financial ratios, using accounting data, such as Compustat, where thousands of firms are observed over relatively few time periods. Coles and Li (2023) provide examples from a number of sub-fields of corporate finance, including: propensity to pay dividends, leverage, investment policy, and firm performance. Another example is cross country empirical growth studies that use data sets such as the Penn World Tables that provide annual data on a range of macro variables for as many as $n = 183$ countries over different time periods, with a maximum time span of $T = 70$. For both panel data sets one would expect multiple long-run relations between the variables, some of which may be the mean reverting ratios discussed in the macroeconomic and finance literatures.

This paper proposes an estimation and testing strategy that applies to panel data models with n possibly much larger than T. We consider an $m \times 1$ vector $\mathbf{w}_{it}$ for units $i = 1, 2, ..., n$ over the time periods $t = 1, 2, ..., T$, with an unknown number, $r_0 \in \{0, 1, ..., m-1\}$, of linear combinations that are stationary. We refer to such linear combinations as long-run relations. If it is known that all elements of $\mathbf{w}_{it}$ are $I(1)$ then the stationary relations can be viewed as cointegrating relations. The focus of our analysis is to estimate r_0, the number of common long-run relations, and their coefficients, when $r_0 \geq 1$. We filter out the short run dynamics by means of q (≥ 2) non-overlapping sub-sample time averages, $\bar{\mathbf{w}}_{i\ell}$, $\ell = 1, 2, ..., q$, as deviations from their full-sample counterpart, $\bar{\mathbf{w}}_{i\circ}$, namely $\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}$, and then construct a pooled sample covariance matrix of these deviations which we denote by $\mathbf{Q}_{\bar{w}\bar{w}}$. The number of long-run relations and their coefficients are estimated using the eigenvalues and eigenvectors of $\mathbf{Q}_{\bar{w}\bar{w}}$. We refer to this procedure as pooled minimum eigenvalue (PME), and note that it is simple to implement, extends readily to unbalanced panels, and is shown to be robust to stationary interactive time effects. It is semi-parametric since it does not require modelling the short run dynamics and applies to general linear process, thus allowing for moving average processes and is not confined to vector autoregressions (VAR). Most importantly, the PME approach does not require knowing long-run causal linkages that might exist amongst the variables under consideration. To our knowledge, no other panel estimation procedure exists for such a setting.

Denoting the first r_0 eigenvectors of $\mathbf{Q}_{\bar{w}\bar{w}}$ by $\hat{\boldsymbol{\beta}}_{j0}$, for $j = 1, 2, ..., r_0$, we then consider structural estimation of the long-run relations assuming they are subject to $r_0 \times r_0$ exact identifying restrictions. Assuming r_0 is known, we derive the asymptotic distribution of the exactly identified long-run relations and propose consistent estimators for their covariance matrices that does not require estimation of the dynamics of individual $\mathbf{w}_{it}$ processes; thus allowing us to test restrictions on the elements of the exactly identified long-run relations with relatively short T. The identified long-run relations are shown to be consistent and asymptotically normally distributed as n and $T \to \infty$ jointly such that $T \approx n^d$. For consistency only $d > 0$ is required, but for asymptotic normality a faster relative rate of $d > 1/2$ is required. Many panel time series estimation and inference procedures require $d > 1$ $(n/T \to 0)$. Our requirement $d > 1/2$ indicates that the procedure should work well with large n and moderate T allowing one to estimate the coefficients of multiple long-run relations in such cases, without estimating short-run dynamics.

We propose to estimate r_0 by the number of eigenvalues of $\mathbf{Q}_{\bar{w}\bar{w}}$ that fall below a given threshold $C_T = CT^{-\delta}$, for some $C > 0$ and $\delta > 0$. One could use cross validation procedures to set C and δ, but based on extensive Monte Carlo experiments we have found that setting $C = 1$ works well *if* we base our selection procedure on the eigenvalues of the correlation matrix, $\mathbf{R}_{\bar{w}\bar{w}} = [diag\left(\mathbf{Q}_{\bar{w}\bar{w}}\right)]^{-1/2} \mathbf{Q}_{\bar{w}\bar{w}} [diag\left(\mathbf{Q}_{\bar{w}\bar{w}}\right)]^{-1/2}$. Such an estimator can be written conveniently as $\tilde{r} = \sum_{j=1}^{m} \mathcal{I}\left(\tilde{\lambda}_j < T^{-\delta}\right)$, where $\tilde{\lambda}_j$, $j = 1, 2, ..., m$ are the eigenvalues of $\mathbf{R}_{\bar{w}\bar{w}}$, and $\mathcal{I}(\mathcal{A}) = 1$ if $\mathcal{A}$ is true and zero otherwise. In the Monte Carlo experiments and the empirical applications we report results for $\delta = (1/4, 1/2)$, and find that overall setting $\delta = 1/4$ works well.

Monte Carlo experiments show near-perfect performance of $\tilde{r}$ as an estimator of r_0, when $\delta = 1/4$, for all $n = 50$, 500, 1,000, 3,000 and $T = 20$, 50, 100 sample size combinations and across a large number of VAR and VARMA data generating processes, with and without interactive time effects, non-Gaussian errors, generalized autoregressive conditional heteroskedasticity (GARCH), threshold autoregressions (TAR), and for different patterns of long-run causal ordering. $\tilde{r}$ performs almost equally well for the smallest sample sizes of $T = 20$ and $n = 50$, as well as for the largest $T = 100$ and $n = 3,000$. As an alternative approach we considered Johansen's trace tests applied to each cross section unit separately, and then estimated r_0 by the simple average of these individual estimates. This was done purely for comparison since to the best of our knowledge there are no other methods that apply to panels in the literature. We found that this average type estimator performed reasonably well when T was large, but still fell short as compared to the thresholding estimator.

The finite sample performance of PME estimator of the coefficients of the long-run relations is found to be satisfactory with inference based on PME estimator with $q = 2$ (sub-sample time averages) generally more accurate in terms of empirical size of the tests, compared with $q = 4$, in line with intuition that suggest a larger choice of q is likely to result in a larger finite-sample bias. We also considered a simple VAR(1) design with $m = 2$, $r_0 = 1$ and one-way long-run causality to see how PME performs compared to the many single equation estimators proposed in the literature (and cited below). We found that in this simple case the PME estimator is less efficient in terms of root mean square errors only when $T = 100$. However, PME with $q = 2$ proved to be less biased and performed much better in terms of size than the single equation approaches for all sample size combinations.

To illustrate the utility of PME procedure we present one micro and one macro application. The micro application considers a number of key financial variables (in logs) and investigates if they are cointegrated, and whether financial ratios can be regarded as stationary variables. To this end we used accounting data for individual firms from CRSP/Compustat on their book value (BV), market value (MV), short-term debt (SD), long-term debt (LD), total assets (TA) and total debt outstanding (DO). The panels involving these variables are unbalanced and cover the period 1950−2021. We consider firms with at least 20 years of data, with n varying between about $1,000$ and $2,500$. The variables are grouped into three sets, where we have prior expectations about possible cointegration and identification. The first set considered has just two variables: the logarithm of total debt outstanding and logarithm of total assets: $\{DO_{it},\ TA_{it}\}$. The ratio of total debt outstanding to total assets is often used as a measure of leverage, which suggests a single hypothesized long-run relation. The other two variable sets are: the logarithms of short and long term debt and total assets, $\{SD_{it},\ LD_{it},\ TA_{it}\}$; and the logarithms of total debt outstanding, book value and market value, $\{DO_{it},\ BV_{it},\ MV_{it}\}$. We expect two hypothesized long-run relations in these sets with three variables. For each set of variables we provide estimates for the full sample 1950-2021 as well as for a shorter sample that ends in 2010. The estimates provide strong evidence of one long-run relation when we consider two variables, and, with one exception, two long-run relations when we consider panels with three variables. In the case of panels with $m = 2$, we illustrate that the PME estimates are invariant to normalization, which is in contrast to the estimates obtained using panel regressions that depend on which way the regression is run. For the relation between logarithms of debt and total assets, we find the estimates of the long-run coefficients are close to one in all cases, ranging from 1.113 to 1.143, and precisely estimated. In the case of panels with $m = 3$ we find $\tilde{r} = 2$, and the null hypothesis that long-run coefficients are equal to unity is not rejected in about a quarter of the panel estimates. These results provide partial support for use of logarithm of financial ratio in corporate finance. In cases where the use of log ratio is not supported, one could use the PME estimates of long-run relations in second stage regressions on stationary variables that also include short run dynamics as well as other stationary variables.

The macro application investigates long-run relations using unbalanced cross country macroeconomic time series data from the Penn World Tables, featuring up to $n = 177$ countries over the years $1950 - 2019$. This dataset has a much smaller cross-section dimension and a larger average time dimension compared with the micro application. We focus on four key macro variables: per capita real merchandise exports (ex_{it}) and imports (im_{it}), real labour productivity per hour worked ($prod_{it}$), and real wages per hour worked ($wage_{it}$). The choice of these variables was motivated by two widely maintained hypotheses. Firstly, real wages and productivity should balance for steady state growth to be feasible. Secondly export and imports should balance for international solvency, though the constraint may not be binding for reserve-currency countries such as the US. These hypotheses are largely confirmed for emerging economies, and, with notable departures from unit long-run elasticities, also for advanced economies. In addition, when we consider all the four variables together we uncover cross country evidence on the long-run relation between exports and productivity without making any assumption about the direction of causality between these variables.

Related literature: We first discuss the literature for a single time series process, which could be viewed as the $p \times 1$ ($p = m\ n$) stacked vector, $\mathbf{w}_t = (\mathbf{w}'_{1t}, \mathbf{w}'_{2t}, ..., \mathbf{w}'_{nt})'$. Our approach is related to that of Phillips and Ouliaris (1988, 1990) in that they also start from general linear processes. They propose testing the null of no cointegration using the smallest eigenvalues of the spectral density of $\Delta \mathbf{w}_t$ evaluated at zero frequency. However, it is difficult to obtain reasonably precise estimates of the spectral density, particularly in the presence of high persistence in first differences. Attempting to eliminate the effects of the short run dynamics by using time averages of sub-samples of time series data is also widely used. Müller and Watson (2018) consider using sub-sample averages to estimate the long-run relation between two variables (y_t and x_t). Their estimated long-run coefficient from regression of sub-sample averages of y_t on those of x_t is not the same as the reciprocal of the estimate that will be obtained from the reverses regression. A panel version of their procedure can be considered, but will be subject to the same limitations, namely it can handle only one long-run relation and will require knowing the direction of long-run causality. In not requiring any assumptions regarding the direction of long-run causality our approach is comparable to the maximum likelihood approach pioneered by Johansen (1988, 1991) that allows for multiple long-run relations without assuming any long-run causal orderings of the variables, but assumes a $VAR(s)$ specification in $\mathbf{w}_t$ where p and s are fixed (and quite small) relative to T. Onatski and Wang (2018, 2019) investigate the asymptotic properties of Johansen test when $\mathbf{w}_t$ follows VAR(1) but allow $p, T \to \infty$, such that $p/T \to c \in (0, 1]$. They provide theoretical arguments why Johansen's test of cointegration rank is likely to be severely over-sized even if p takes moderate values. Extensions to higher order VARs are provided by (Bykhovskaya and Gorin 2022). This is a promising approach which is yet to be fully developed for the analysis of multiple cointegrations across many units, which is the primary focus of this paper. Since the ordering of the variables in the VAR does not affect the Johansen's tests of the cointegration rank, without further restrictions the use of high-dimensional VARs in $\mathbf{w}_t$ does not distinguish between cointegration across units as compared to cointegration between the variables specific to the cross section units. Also, the condition $p/T = nm/T \to c \in (0, 1]$ is unlikely to be met when $m > 1$ and n is of the same order of magnitude as T.

Turning to the panel cointegration literature, most studies consider $I(1)$ variables with a single cointegrating vector where the direction of long-run causality is known. These estimators are typically generalizations of the time series procedures such as the panel Fully Modified OLS of Pedroni (1996, 2001a, 2001b), the Pooled Mean Group (PMG) estimator of Pesaran, Shin, and Smith (1999),or the panel Dynamic OLS of Mark and Sul (2003). There are panel generalizations of Johansen's approach, such as Groen and Kleibergen (2003), and Larsson and Lyhagen (2007), which can be used to test for the number of cointegrating relations and estimate their parameters. These are based on a vector error correction model, VECM, which can deal with multiple cointegrating vectors, but require T to be large relative to n. In their applications Larsson and Lyhagen (2007) have $m = 3$, $n = 4$. Breitung (2005) proposes a systems estimator, but that requires that every cross section unit cointegrate. Chudik, Pesaran, and Smith (2023b) suggest system pooled mean group estimator for a single common long-run relation coefficient θ, that can handle any long-run causal ordering and allow some units to fail to cointegrate, but again requires T to be large relative to n.

A large number of other topics have been examined within the context of a panel with a single cointegrating relation. These include: estimation with I(1) latent factors: Bai, Kao, and Ng (2009) and Kapetanios, Pesaran, and Yamagata (2011); structural breaks: Banerjee and i Silvestre (2024) and Ditzen, Karavias, and Westerlund (2025); and non-linear effects: de Jong and Wagner (2025). Further details can be found in the surveys by Breitung and Pesaran (2008) and Choi (2015) that also cover testing for cointegration using residuals (Westerlund, 2005), and second generation panel unit root tests allowing for cross section dependence (Pesaran, 2007). As this brief overview indicates, none of the methods advanced in the literature consider multiple long-run relations when $n >> T$.

Outline of the paper: The rest of the paper is set out as follows: Section 2 sets out the panel data model and introduces the PME estimator. Section 3 introduces the assumptions and discusses the identification conditions. Section 4 gives a formal description of the PME estimator and some of its asymptotic properties. Section 5 considers identification and derives the asymptotic distribution of the exactly identified PME estimator. Section 6 shows how r_0 can be estimated by eigenvalue thresholding. Section 7 allows for interactive time effects. Section 8 discusses the choice of the number of sub-sample averages, q, and how to set the parameters of the thresholding estimator of r_0. Section 9 provides Monte Carlo evidence on the small sample properties of the PME estimators of r_0 and $\boldsymbol{\beta}_{j0}$, $j = 1, 2, ..., r_0$. Section 10 discusses the empirical applications, and Section 11 provides some concluding remarks. The proofs of the propositions and theorems are provided in an appendix, with related lemmas given in a supplement. This supplement also includes sub-sections on extensions of PME to panels with interactive time effects, on how to implement the proposed estimator for unbalanced panels, and gives details of the data generating processes used in the Monte Carlo experiments, plus additional information on data sources and the construction of the variables used in the empirical applications.

Notations: Matrices are denoted by bold upper case letters and vectors are denoted by bold lower case letters. All vectors are column vectors. $\|\mathbf{x}\|$ denotes the Euclidean norm of a vector $\mathbf{x}$. $rank(\mathbf{A})$ denotes the column rank of $\mathbf{A}$. $vec(\mathbf{A})$ denotes vectorization of $\mathbf{A}$. $tr(\mathbf{A})$ denotes the trace of a square matrix $\mathbf{A}$. Eigenvalues of $m \times m$ symmetric positive semi-definite real matrix $\mathbf{A}$ sorted in ascending order are $0 \leq \lambda_1(\mathbf{A}) \leq \lambda_2(\mathbf{A}) \leq ... \leq \lambda_m(\mathbf{A})$. $\|\mathbf{A}\|$ is the spectral norm of $\mathbf{A}$. Small and large finite positive constants that do not depend on sample sizes n and T are denoted by ϵ and K, respectively. These constants can take different values at different instances in the paper. $T_n \approx n^d$ if there exist $n_0 \geq 1$ and positive constants ϵ and K, such that $\inf_{n \geq n_0} \left(T_n/n^d\right) \geq \epsilon$ and $\sup_{n \geq n_0} \left(T_n/n^d\right) \leq K$. For simplicity of exposition we omit subscript n and write $T \approx n^d$. Convergence in probability and distribution are denoted by $\rightarrow_p$ and $\rightarrow_d$, respectively. In this paper $o_p(1)$ is short for sequence of random variables, random vectors or random matrices that converge to zero in probability as $n \rightarrow \infty$ for all values of $d > 1/2$. $\mathbf{A}_n = O_p(1)$ if sequence $\mathbf{A}_n$ is bounded in probability. $a_n = O(b_n)$ denotes the deterministic sequence $\{a_n\}$ is at most of order b_n. Equivalence of asymptotic distributions is denoted by $\overset{a}{\sim}$.

2 Preliminaries

Consider the following general linear model for $\mathbf{w}_{it}$

$$\mathbf{w}_{it} = \mathbf{a}_i + \mathbf{G}_i \mathbf{f}_t + \mathbf{C}_i \mathbf{s}_{it} + \mathbf{v}_{it}, \text{ for } t = 1, 2, ..., T; i = 1, 2, ..., n, \tag{1}$$

where $\mathbf{w}_{it}$ is an $m \times 1$ vector of outcomes, $\mathbf{a}_i$ is $m \times 1$ vector of fixed effects, $\mathbf{f}_t$ is a vector of stationary latent factors with associated loading matrices, $\mathbf{G}_i$. $\mathbf{s}_{it}$ is the partial sum process defined by

$$\mathbf{s}_{it} = \mathbf{u}_{i1} + \mathbf{u}_{i2} + ... + \mathbf{u}_{it}, \text{ for } t \geq 1, \text{ and } \mathbf{s}_{it} = \mathbf{0}, \text{ for } t < 1, \tag{2}$$

$\mathbf{u}_{it}$ is independently distributed over i and t with mean zero and the $m \times m$ positive definite matrix, $\mathbf{\Sigma}_i$, $\mathbf{C}_i$ is an $m \times m$ matrix of fixed coefficients, and $\mathbf{v}_{it} = \mathbf{C}_i^*(L)\mathbf{u}_{it}$, where $\mathbf{C}_i^*(L) = \sum_{\ell=0}^{\infty} \mathbf{C}_{i\ell}^* L^{\ell}$. This model covers many specifications of interest such as vector autoregressions, error correction models, as well as first-differenced stationary models. It allows for interactive time effects which reduce to time effects under the so-called parallel trends assumption, namely setting $\mathbf{G}_i = \mathbf{G}$ for all i. In stacked form the model for all n units can be written as $\mathbf{w}_t = \mathbf{a} + \mathbf{G}\mathbf{f}_t + \mathbf{C}\mathbf{s}_t + \mathbf{C}^*(L)\mathbf{u}_t$, where $\mathbf{w}_t = (\mathbf{w}_{1t}', \mathbf{w}_{2t}', ..., \mathbf{w}_{nt}')'$, $\mathbf{a} = (\mathbf{a}_1', \mathbf{a}_2', ..., \mathbf{a}_n')'$, $\mathbf{G} = (\mathbf{G}_1', \mathbf{G}_2', ..., \mathbf{G}_n')'$, and $\mathbf{u}_t = (\mathbf{u}_{1t}', \mathbf{u}_{2t}', ..., \mathbf{u}_{nt}')'$. Under our specification $\mathbf{C}$ and $\mathbf{C}_i^*(L)$ are assumed to be block-diagonal matrices with $\mathbf{C}_i$ and $\mathbf{C}_i^*(L)$ as their i^{th} block, respectively. Such restrictions seem inevitable when n is large relative to T, and seems plausible considering that we allow for cross-sectional dependence through the common factors, $\mathbf{f}_t$.

Assuming that $\mathbf{C}_i$ has rank $m - r_0 > 0$ for all i, we are interested in estimating r_0, and the associated stationary linear combinations defined by $\boldsymbol{\beta}_{j0}'\mathbf{w}_{it}$, $j = 1, 2, ..., r_0$, where $\mathbf{B}_0 = \left(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, ..., \boldsymbol{\beta}_{r_0 0}\right)$ is the $m \times r_0$ matrix of long-run relations that are common across all i, and satisfies $\mathbf{B}_0'\mathbf{C}_i = 0$. We also consider estimation of long-run relations subject to the exactly identifying restrictions that are motivated by the theory. The estimator we propose involves splitting the data for each unit into $q \geq 2$ sub-samples; taking time averages of these sub-samples and forming a pooled demeaned covariance matrix we label $\mathbf{Q}_{\bar{w}\bar{w}}$. The eigenvalues of this matrix allow us to estimate r_0, and the eigenvectors corresponding to the first r_0 eigenvalues provide estimates of $\boldsymbol{\beta}_{j0}$. But to simplify the exposition and focus on the main contribution of the paper, initially we abstract from the interactive time effects, but return to this complication in Section 7, where we show that our analysis remains valid so long as the latent factors are stationary.

It is possible to allow for non-linear features, such as GARCH and threshold autoregressions, so long as the effects of shocks to $\mathbf{u}_{it}$ decay exponentially fast. But to keep the theoretical analyses relatively simple, we only consider the robustness of our estimation and testing strategies to such non-linear effects using Monte Carlo experiments.

3 Assumptions and identification conditions

We directly work with (1) and make the following assumptions:

Assumption 1 *The error terms,* $\mathbf{u}_{it}$*, are distributed independently over* $i = 1, 2, ..., n$ *and* $t = 1, 2, ..., T$ *with* $E(\mathbf{u}_{it}) = \mathbf{0}$*, and the covariance matrix* $E(\mathbf{u}_{it}\mathbf{u}'_{it}) = \mathbf{\Sigma}_i$*, where* $\mathbf{\Sigma}_i$ *is a positive definite matrix,* $\inf_i \boldsymbol{\lambda}_1(\mathbf{\Sigma}_i) > \epsilon$*,* $\sup_i \boldsymbol{\lambda}_m(\mathbf{\Sigma}_i) < K$*, and* $\sup_{it} E\|\mathbf{u}_{it}\|^{4+\epsilon} < K$*, for some* $\epsilon > 0$*.*

Assumption 2 *The coefficient matrices* $\mathbf{C}_i$ *and* $\mathbf{C}^*_{i\ell}$*, are non-stochastic constants such that* $\sup_i \|\mathbf{C}_i\| < K$ *and* $\sup_i \|\mathbf{C}^*_{i\ell}\| < K\rho^\ell$*, where* ρ *lies in the range* $0 < \rho < 1$*. The* $m \times m$ *matrix* $\mathbf{C}_i$ *has rank* $m - r_0$*, for* $i = 1, 2, ..., n$*.*

Assumption 3 *(a) Let*

$$\mathbf{\Psi}_n = n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{\Sigma}_i\mathbf{C}'_i, \text{ and } \mathbf{\Psi} = lim_{n\to\infty}\mathbf{\Psi}_n. \tag{3}$$

Then there exists n_0 *such that for all* $n > n_0$*,* $rank(\mathbf{\Psi}_n) = rank(\mathbf{\Psi}) = m - r_0 > 0$*. (b) The orthonormalized eigenvectors associated with the* r_0 *zero eigenvalues of* $\left(\frac{q-1}{6q}\right)\mathbf{\Psi}_n$ *are denoted by* $\boldsymbol{\beta}_{j0}$*, for* $j = 1, 2, ..., r_0$*, and the orthonormalized eigenvectors associated with the ordered non-zero eigenvalues of* $\left(\frac{q-1}{6q}\right)\mathbf{\Psi}_n$*, namely* $\lambda_{r_0+1} \leq \lambda_{r_0+2} \leq ... \leq \lambda_m$*, by* $\boldsymbol{\beta}_j$*, for* $r_0 + 1, r_0 + 2, ..., m$*. Specifically*

$$\left(\frac{q-1}{6q}\right)\mathbf{\Psi}_n\boldsymbol{\beta}_{j0} = 0, \text{ for } j = 1, 2, ..., r_0, \tag{4}$$

and

$$\left(\frac{q-1}{6q}\right)\mathbf{\Psi}_n\boldsymbol{\beta}_j = \lambda_j\boldsymbol{\beta}_j, \text{ for } j = r_0 + 1, r_0 + 2, ..., m. \tag{5}$$

Remark 1 *Under Assumptions 1-2* $\{\mathbf{v}_{it}\}$ *has absolute summable autocovariances,* $\sum_{h=0}^{\infty}\sup_i \|\mathbf{\Gamma}_i(h)\| < K$*, where* $\mathbf{\Gamma}_i(h) = E\left(\mathbf{v}_{it}\mathbf{v}'_{i,t-h}\right)$ *is the autocovariance function of* $\mathbf{v}_{it}$*. See Lemma 1 in the supplement, Section S1.*

Remark 2 *Under Assumptions 3,* $\boldsymbol{\beta}'_{j0}\mathbf{\Psi}_n\boldsymbol{\beta}_{j0} = 0$*, for* $j = 1, 2, ..., r_0$ *and* $\lambda_j = \left(\frac{q-1}{6q}\right)\boldsymbol{\beta}'_j\mathbf{\Psi}_n\boldsymbol{\beta}_j > 0$*, for* $j = r_0 + 1, r_0 + 2, ..., m$*. Nonzero eigenvalues* λ_j*, for* $j = r_0+1, r_0+2, ..., m$*, and the corresponding eigenvectors* $\boldsymbol{\beta}_j$*, for* $j = r_0+1, r_0+2, ..., m$*, depend on* n*, but to simplify the notations we avoid using the subscript* n*.*

Remark 3 *Under part (b) of Assumption 3* $\mathbf{\Psi}_n$ *and* $\mathbf{\Psi}$ *can be written as* $\mathbf{\Psi}_n = \mathbf{P}_n'\mathbf{P}_n$, *and* $\mathbf{\Psi} = \mathbf{P}'\mathbf{P}$, *where* $\mathbf{P}_n$ *and* $\mathbf{P}$ *are* $(m - r_0) \times m$ *full rank matrices, with rank* $(\mathbf{P}_n) =$ *rank* $(\mathbf{P}) = m - r_0$, *for all* $n > n_0$.

Remark 4 *Condition rank* $(\mathbf{C}_i) = m - r_0$ *for all* $i = 1, 2, ..., n$ *in Assumption 2 can be relaxed to allow for no stochastic trends for some cross section units, so long as the rank condition rank* $(\mathbf{\Psi}_n) =$ *rank* $(\mathbf{\Psi}) = m - r_0$ *holds. Specifically, suppose that* $\mathbf{C}_i = \mathbf{0}$ *for* $i = 1, 2, ..., n_1$, *but the cointegration rank condition holds for the remaining units. Then*

$$\mathbf{\Psi}_n = n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{\Sigma}_i\mathbf{C}_i' = (1-\pi_n)\left(\frac{1}{n-n_1}\sum_{i=n_1+1}^{n}\mathbf{C}_i\mathbf{\Sigma}_i\mathbf{C}_i'\right),$$

where $\pi_n = n_1/n$ *is the proportion of units without stochastic trends. Then the rank requirement continues to hold if* $\pi_n > 0$, *and* $\frac{1}{n-n_1}\sum_{i=n_1+1}^{n}\mathbf{C}_i\mathbf{\Sigma}_i\mathbf{C}_i'$ *tends to a matrix having rank* $m - r_0$. *But for clarity of exposition we maintain Assumption 2 without loss of generality.*

4 Estimation of long-run relations

4.1 Introducing sub-sample time averages

We base our estimation procedure on non-overlapping sub-sample time averages of $\mathbf{w}_{it}$. For the ease of exposition, suppose the panel data under consideration is balanced, T is divisible by q, and consider q (≥ 2) non-overlapping time averages of equal length T_q defined by

$$\bar{\mathbf{w}}_{i\ell} = \frac{1}{T_q}\sum_{t=(\ell-1)T_q+1}^{\ell T_q}\mathbf{w}_{it}, \text{ for } \ell = 1, 2, ..., q, \tag{6}$$

where $T_q = T/q$. To simplify the exposition we abstract from interactive effects and apply the above time average operator to (1) to obtain[1]

$$\bar{\mathbf{w}}_{i\ell} = \mathbf{a}_i + \mathbf{C}_i\bar{\mathbf{s}}_{i\ell} + \bar{\mathbf{v}}_{i\ell}, \text{ for } \ell = 1, 2, ..., q, \tag{7}$$

[1]Sections S2 and S3 of the supplement consider unbalanced panels and models with interactive effects.

where $\bar{\mathbf{v}}_{i\ell} = T_q^{-1} \sum_{t=(\ell-1)T_q+1}^{\ell T_q} \mathbf{v}_{it}$ and $\bar{\mathbf{s}}_{i\ell} = T_q^{-1} \sum_{t=(\ell-1)T_q+1}^{\ell T_q} \mathbf{s}_{it}$. We now use standard de-meaning procedure and eliminate $\mathbf{a}_i$ from (7) to obtain

$$\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ} = \mathbf{C}_i \left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ} \right) + \left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ} \right), \text{ for } \ell = 1, 2, ..., q, \tag{8}$$

where $\bar{\mathbf{w}}_{i\circ} = q^{-1} \sum_{\ell=1}^{q} \bar{\mathbf{w}}_{i\ell}$, and similarly $\bar{\mathbf{s}}_{i\circ} = q^{-1} \sum_{\ell=1}^{q} \bar{\mathbf{s}}_{i\ell}$ and $\bar{\mathbf{v}}_{i\circ} = q^{-1} \sum_{\ell=1}^{q} \bar{\mathbf{v}}_{i\ell}$. Consider now the $m \times m$ pooled sample covariance matrix

$$\mathbf{Q}_{\bar{w}\bar{w}} = n^{-1} \sum_{i=1}^{n} \mathbf{Q}_{\bar{w}_i \bar{w}_i} \tag{9}$$

where

$$\mathbf{Q}_{\bar{w}_i \bar{w}_i} = T^{-1} q^{-1} \sum_{\ell=1}^{q} \left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ} \right) \left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ} \right)'. \tag{10}$$

The limiting value of $\mathbf{Q}_{\bar{w}\bar{w}}$, as $n, T \to \infty$, plays a critical role in our approach to estimation of long-run relations. Using (8) in (10) we first note that

$$\mathbf{Q}_{\bar{w}_i \bar{w}_i} = \mathbf{C}_i \mathbf{Q}_{\bar{s}_i \bar{s}_i} \mathbf{C}_i' + \mathbf{C}_i \mathbf{Q}_{\bar{s}_i \bar{v}_i} + \mathbf{Q}_{\bar{v}_i \bar{s}_i}' \mathbf{C}_i' + \mathbf{Q}_{\bar{v}_i \bar{v}_i}, \tag{11}$$

where $T\mathbf{Q}_{\bar{s}_i \bar{s}_i} = q^{-1} \sum_{\ell=1}^{q} \left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ} \right) \left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ} \right)'$, $T\mathbf{Q}_{\bar{s}_i \bar{v}_i} = q^{-1} \sum_{\ell=1}^{q} \left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ} \right) \left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ} \right)' = T\mathbf{Q}_{\bar{v}_i \bar{s}_i}'$, and $T\mathbf{Q}_{\bar{v}_i \bar{v}_i} = q^{-1} \sum_{\ell=1}^{q} \left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ} \right) \left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ} \right)'$. Since $\{\mathbf{v}_{it}\}$ is covariance stationary with absolute summable autocovariances and $\{\mathbf{s}_{it}\}$ is a partial sum process it then follows that $\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ} = O_p\left(T^{-1/2}\right)$, and $\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ} = O_p(T^{1/2})$. Moreover, as established in Lemma 3,

$$\sup_i E \left\| \mathbf{Q}_{\bar{v}_i \bar{v}_i} \right\| = O\left(T^{-2}\right), \text{ and } \sup_i E \left\| \mathbf{Q}_{\bar{s}_i \bar{v}_i} \right\| = O\left(T^{-1}\right). \tag{12}$$

4.2 Pooled minimum eigenvalue (PME) estimator

Our proposed estimation procedure is based on eigenvalues and eigenvectors of $\mathbf{Q}_{\bar{w}\bar{w}}$, defined by (9). Averaging $\mathbf{Q}_{\bar{w}_i \bar{w}_i}$ in (11) over all cross section units now yields:

$$\mathbf{Q}_{\bar{w}\bar{w}} = n^{-1} \sum_{i=1}^{n} \mathbf{C}_i \mathbf{Q}_{\bar{s}_i \bar{s}_i} \mathbf{C}_i' + n^{-1} \sum_{i=1}^{n} \mathbf{C}_i \mathbf{Q}_{\bar{s}_i \bar{v}_i} + n^{-1} \sum_{i=1}^{n} \mathbf{Q}_{\bar{v}_i \bar{s}_i}' \mathbf{C}_i' + n^{-1} \sum_{i=1}^{n} \mathbf{Q}_{\bar{v}_i \bar{v}_i}. \tag{13}$$

The pooled minimum eigenvalue (PME) estimator of $\boldsymbol{\beta}_{j0}$, $j = 1, 2, ..., r_0$, is given by the j^{th} orthonormalized eigenvector of $\mathbf{Q}_{\bar{w}\bar{w}}$, $\hat{\boldsymbol{\beta}}_j$, associated with its r_0 smallest eigenvalues, $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq ... \leq \hat{\lambda}_{r_0}$. Specifically, for $j = 1, 2, ..., m$, $\mathbf{Q}_{\bar{w}\bar{w}}\hat{\boldsymbol{\beta}}_j = \hat{\lambda}_j\hat{\boldsymbol{\beta}}_j$, such that $\hat{\boldsymbol{\beta}}_j'\hat{\boldsymbol{\beta}}_j = 1$, and $\hat{\lambda}_j = \hat{\boldsymbol{\beta}}_j'\mathbf{Q}_{\bar{w}\bar{w}}\hat{\boldsymbol{\beta}}_j$. In matrix notations we have

$$\hat{\mathbf{B}} = \left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, ..., \hat{\boldsymbol{\beta}}_m\right), \; \left\|\hat{\mathbf{B}}\right\| = 1, \tag{14}$$

and the PME estimator of $\mathbf{B}_0$ is given by $\hat{\mathbf{B}}_0 = \left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, ..., \hat{\boldsymbol{\beta}}_{r_0}\right)$.

4.3 Consistency of the PME estimator

Under Assumption 1, $\mathbf{Q}_{\bar{s}_i\bar{s}_i}$ has the following exact moment (established in Lemma 2)

$$E\left(\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right) = \frac{(q-1)}{6}\left(\frac{1}{q} + \frac{1}{T^2}\right)\boldsymbol{\Sigma}_i. \tag{15}$$

Hence $n^{-1}\sum_{i=1}^n \mathbf{C}_i E\left(\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right)\mathbf{C}_i' = \frac{(q-1)}{6}\left(\frac{1}{q} + \frac{1}{T^2}\right)\boldsymbol{\Psi}_n$, where $\boldsymbol{\Psi}_n$ is defined by (3), and by Assumption 3 is assumed to have rank $m - r_0 > 0$. Furthermore, using $\sup_i \|\mathbf{C}_i\| < K$ and (12) then

$$\begin{aligned} n^{-1}\sum_{i=1}^n \mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i} &= O_p\left(T^{-1}\right), \; n^{-1}\sum_{i=1}^n \mathbf{Q}_{\bar{v}_i\bar{s}_i}'\mathbf{C}_i' = O_p\left(T^{-1}\right), \\ \text{and } \; n^{-1}\sum_{i=1}^n \mathbf{Q}_{\bar{v}_i\bar{v}_i} &= O_p\left(n^{-1/2}T^{-2}\right). \end{aligned} \tag{16}$$

Using the above results in (13) we have

$$\mathbf{Q}_{\bar{w}\bar{w}} - \frac{(q-1)}{6q}\boldsymbol{\Psi}_n = n^{-1}\sum_{i=1}^n \mathbf{C}_i\left[\mathbf{Q}_{\bar{s}_i\bar{s}_i} - E\left(\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right)\right]\mathbf{C}_i' + O_p\left(T^{-1}\right). \tag{17}$$

Further $n^{-1}\sum_{i=1}^n \mathbf{C}_i\left[\mathbf{Q}_{\bar{s}_i\bar{s}_i} - E\left(\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right)\right]\mathbf{C}_i' = q^{-1}\sum_{\ell=1}^q \mathbf{G}_\ell - \mathbf{G}_0$, where

$$\mathbf{G}_\ell = n^{-1}\sum_{i=1}^n \mathbf{C}_i\left\{\left(\frac{\bar{\mathbf{s}}_{i\ell}}{\sqrt{T}}\right)\left(\frac{\bar{\mathbf{s}}_{i\ell}}{\sqrt{T}}\right)' - E\left[\left(\frac{\bar{\mathbf{s}}_{i\ell}}{\sqrt{T}}\right)\left(\frac{\bar{\mathbf{s}}_{i\ell}}{\sqrt{T}}\right)'\right]\right\}\mathbf{C}_i', \text{ for } \ell = 1, 2, ..., q,$$

and

$$\mathbf{G}_0 = n^{-1}\sum_{i=1}^n \mathbf{C}_i\left\{\left(\frac{\bar{\mathbf{s}}_{i\circ}}{\sqrt{T}}\right)\left(\frac{\bar{\mathbf{s}}_{i\circ}}{\sqrt{T}}\right)' - E\left[\left(\frac{\bar{\mathbf{s}}_{i\circ}}{\sqrt{T}}\right)\left(\frac{\bar{\mathbf{s}}_{i\circ}}{\sqrt{T}}\right)'\right]\right\}\mathbf{C}_i'.$$

Under Assumptions 1 and 2, $\bar{\mathbf{s}}_{i\ell}$ and $\bar{\mathbf{s}}_{i\circ}$ are cross-sectionally independent random variables and $\sup_i \|\mathbf{C}_i\| < K$. Further, $T_q^{-1/2}\bar{\mathbf{s}}_{i\ell}$ and $T^{-1/2}\bar{\mathbf{s}}_{i\circ}$ are scaled partial sums of u_{it} and tend to bounded random variables. See, for example, result (d) of Proposition 17.1 in Hamilton (1994). Therefore, $\mathbf{G}_\ell$ and $\mathbf{G}_0$ both converge at the rate of $n^{-1/2}$ to their means that are zero, by construction. Namely $\mathbf{G}_\ell = O_p\left(n^{-1/2}\right)$ and $\mathbf{G}_0 = O_p\left(n^{-1/2}\right)$. Hence

$$n^{-1}\sum_{i=1}^{n} \mathbf{C}_i \left[\mathbf{Q}_{\bar{s}_i\bar{s}_i} - E\left(\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right)\right] \mathbf{C}_i' = O_p\left(n^{-1/2}\right), \tag{18}$$

and using this result in (17) yields

$$\mathbf{Q}_{\bar{w}\bar{w}} - \frac{(q-1)}{6q}\mathbf{\Psi}_n = O_p\left(n^{-1/2}\right) + O_p\left(T^{-1}\right).$$

Therefore, for a fixed $q\,(\geq 2)$, $\mathbf{Q}_{\bar{w}\bar{w}} \rightarrow_p \frac{(q-1)}{6q}\mathbf{\Psi}$, as $n, T \rightarrow \infty$ jointly such that $T_n \approx n^d$ and $d > 0$, where $\mathbf{\Psi} = lim_{n\rightarrow\infty}\mathbf{\Psi}_n$. This result is formally established in the following proposition.

Proposition 1 *Consider the panel data model for $\mathbf{w}_{it}$ given by (1) without the interactive time effects ($\mathbf{G}_i = 0$), and suppose Assumptions 1 to 3 hold. Consider the $m \times m$ pooled sample covariance matrix $\mathbf{Q}_{\bar{w}\bar{w}}$ defined by (9). Then for $q \geq 2$ and a fixed m we have*

$$\mathbf{Q}_{\bar{w}\bar{w}} = \frac{(q-1)}{6q}\mathbf{\Psi}_n + O_p\left(n^{-1/2}\right) + O_p\left(T^{-1}\right), \tag{19}$$

$$\mathbf{Q}_{\bar{w}\bar{w}}\boldsymbol{\beta}_{j0} = O_p\left(n^{-1/2}\right) + O_p\left(T^{-1}\right), \text{ for } j = 1, 2, ..., r_0, \tag{20}$$

and $\mathbf{Q}_{\bar{w}\bar{w}} \rightarrow_p \frac{(q-1)}{6q}\mathbf{\Psi}$, as $n, T \rightarrow \infty$ jointly such that $T_n \approx n^d$ and $d > 0$, where $\mathbf{\Psi}$ and $\mathbf{\Psi}_n$ are defined in Assumption 3.

For a known r_0 the PME estimator of $\mathbf{B}_0$, is given by $\hat{\mathbf{B}}_0 = \left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, ..., \hat{\boldsymbol{\beta}}_{r_0}\right)$, where $\hat{\boldsymbol{\beta}}_j$, $j = 1, 2, ..., m$ are the orthonormalized eigenvectors of $\mathbf{Q}_{\bar{w}\bar{w}}$, as set out below equation (13). Then

$$\hat{\mathbf{B}}_0'\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0 = \frac{(q-1)}{6q}\hat{\mathbf{B}}_0'\mathbf{\Psi}\hat{\mathbf{B}}_0 + O_p\left(n^{-1/2}\right) + O_p\left(T^{-1}\right). \tag{21}$$

and $\hat{\mathbf{B}}_0'\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0$ is asymptotically minimized when $\hat{\mathbf{B}}_0'\mathbf{\Psi}\hat{\mathbf{B}}_0 = \mathbf{0}$, and this occurs if $\hat{\mathbf{B}}_0$ lies in the space spanned by the r_0 eigenvectors of $\mathbf{\Psi}$ that are associated with its r_0 zero eigenvalues, namely if and only if $\hat{\mathbf{B}} = \mathring{\mathbf{B}}_0\mathbf{H}$, for some $r_0 \times r_0$ non-singular rotation matrix, $\mathbf{H}$. Hence, we have $\hat{\mathbf{B}}_r\mathbf{H} \rightarrow_p \mathring{\mathbf{B}}_0$ as $n, T \rightarrow \infty$ jointly such that $T \approx n^d$ and $d > 0$. A formal statement is provided in the following theorem with proofs in the Appendix.

Theorem 1 *Consider the panel data model for the $m \times 1$ vector $\mathbf{w}_{it}$ given by (1) without the interactive time effects ($\mathbf{G}_i = 0$), and suppose that Assumptions 1 to 3 hold and the number of long-run relations, r_0, is known. Let $\hat{\mathbf{B}}_0 = \left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, ..., \hat{\boldsymbol{\beta}}_{r_0}\right)$ be the $m \times r_0$ matrix formed from the orthonormalized eigenvectors of $\mathbf{Q}_{\bar{w}\bar{w}}$, defined by (9), associated with its r_0 smallest eigenvalues. Then for a fixed m and q (≥ 2), $\hat{\mathbf{B}}_0\mathbf{H} \rightarrow_p \mathring{\mathbf{B}}_0$ as $n, T \rightarrow \infty$ jointly such that $T_n \approx n^d$ and $d > 0$, for any $r_0 \times r_0$ non-singular matrix, $\mathbf{H}$.*

5 Identification and asymptotic distribution of PME estimator

We focus on the case of exact identification of the long-run relations, noting that the estimation of r_0 is invariant on how exact identification is achieved.

5.1 Exact identifying conditions

We assume there exist r_0^2 a *priori given and theoretically meaningful* exact identifying restrictions on $\mathbf{B}_0$ given by

$$\mathbf{R}\mathring{\mathbf{B}}_0 = (\mathbf{R}_1, \mathbf{R}_2)\begin{pmatrix} \mathring{\mathbf{B}}_{0,1} \\ \mathring{\mathbf{B}}_{0,2} \end{pmatrix} = \mathbf{A}, \tag{22}$$

where $\mathring{\mathbf{B}}_0$ is the $m \times r_0$ matrix of identified long-run relations, $\mathbf{R}_1$,$\mathbf{R}_2$ and $\mathbf{A}$ are $r_0 \times r_0$, $r_0 \times (m - r_0)$ and $r_0 \times r_0$ matrices of known fixed constants, with $rank\,(\mathbf{A}) = rank\,(\mathbf{R}_1) = r_0 < m$. Then it follows that $\mathbf{H} = (\mathbf{R}\mathbf{B}_0)^{-1}\mathbf{A}$, and the PME estimator of $\mathring{\mathbf{B}}_{\mathbf{0}}$ is given by

$$\hat{\mathring{\mathbf{B}}}_0 = \hat{\mathbf{B}}_0\left(\mathbf{R}\hat{\mathbf{B}}_0\right)^{-1}\mathbf{A}, \tag{23}$$

where $\hat{\mathbf{B}}_0 = \left(\hat{\boldsymbol{\beta}}_{10}, \hat{\boldsymbol{\beta}}_{20}, ..., \hat{\boldsymbol{\beta}}_{r_0,0}\right)$. The exact identifying restrictions, (22), will often take the form $\mathring{\mathbf{B}}_0 = \left(\mathbf{I}_{r_0}, \mathring{\mathbf{B}}'_{0,2}\right)'$, where $\mathring{\mathbf{B}}_{0,1} = \mathbf{I}_{r_0}$ is an identity matrix of order r_0. Without loss of generality, we consider this formulation and denote $\mathring{\mathbf{B}}_{0,2}$ by $\boldsymbol{\Theta}$. Once $\boldsymbol{\Theta}$ is estimated it is possible to estimate $\mathring{\mathbf{B}}_{0,1}$ under more general restrictions as $\mathring{\mathbf{B}}_{0,1} = \mathbf{R}_1^{-1}\left(\mathbf{A} - \mathbf{R}_2\boldsymbol{\Theta}\right)$.

Proposition 2 *Consider the r_0^2 exact identifying restrictions given by (22), and suppose $m \times r$ matrix of long-run relations $\mathring{\mathbf{B}}_0$ is normalized as $\mathring{\mathbf{B}}_0 = (\mathbf{I}_{r_0}, \boldsymbol{\Theta}')'$, and Assumption 3 holds. Partition $\boldsymbol{\Psi}$ conformably with $\mathring{\mathbf{B}}_0$ as $\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}'_{21} \\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22} \end{pmatrix}$. Then $\mathbf{I}_{r_0} + \boldsymbol{\Psi}_{11}$ and $\boldsymbol{\Psi}_{22}$ are respectively $r_0 \times r_0$ and $(m - r_0) \times (m - r_0)$ positive definite matrices and $\boldsymbol{\Theta}$ is uniquely determined by $\boldsymbol{\Theta} = -\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Psi}_{21}$, subject to the restrictions $\boldsymbol{\Psi}_{11} = \boldsymbol{\Psi}'_{21}\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Psi}_{21}$. A proof is provided in the Appendix.*

5.2 Asymptotic distribution

To derive the asymptotic distribution of $\hat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0$, note from (A.10) that

$$\mathbf{Q}_{\bar{w}\bar{w}}\sqrt{n}T\left(\widehat{\mathring{\mathbf{B}}_0} - \mathring{\mathbf{B}}_0\right) = -\left(n^{-1/2}\sum_{i=1}^{n} T\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0 - \frac{\sqrt{n}}{T}\left(n^{-1}\sum_{i=1}^{n} T^2\mathbf{Q}_{\bar{v}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0 + O_p\left(T^1\right). \tag{24}$$

By Lemma 5, $n^{-1}\sum_{i=1}^{n} T^2\mathbf{Q}_{\bar{v}_i\bar{v}_i} = O_p\left(n^{-1/2}\right)$ and since $\left\|\mathring{\mathbf{B}}_0\right\| < K$, then

$$\mathbf{Q}_{\bar{w}\bar{w}}\sqrt{n}T\left(\widehat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right) = -\left(n^{-1/2}\sum_{i=1}^{n} T\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0 + O_p\left(T^{-1}\right). \tag{25}$$

Further, write the first term as

$$\begin{aligned}\left(n^{-1/2}\sum_{i=1}^{n} T\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0 \quad &= \quad \left(n^{-1/2}\sum_{i=1}^{n}\left[T\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i} - \mathbf{C}_iE\left(T\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\right]\right)\mathring{\mathbf{B}}_0 \\ &\quad + \frac{\sqrt{n}}{T}\left(n^{-1}\sum_{i=1}^{n}\mathbf{C}_iE\left(T^2\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\right)\mathring{\mathbf{B}}_0,\end{aligned}$$

and using $\sup_i \|E(\mathbf{Q}_{\bar{s}_i\bar{v}_i})\| = O(T^{-2})$ (established in Lemma 3))

$$n^{-1}\sum_{i=1}^{n}\mathbf{C}_i E\left(T^2\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0 = O(1). \tag{26}$$

Using the above results in (25) we have

$$\mathbf{Q}_{\bar{w}\bar{w}}\sqrt{n}T\left(\hat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right) = -n^{-1/2}\sum_{i=1}^{n}\mathbf{Z}_i + O_p\left(\frac{\sqrt{n}}{T}\right), \tag{27}$$

where $\mathbf{Z}_i = \mathbf{C}_i\left[T\mathbf{Q}_{\bar{s}_i\bar{v}_i} - E\left(T\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\right]\mathring{\mathbf{B}}_0$. Recalling that $T\mathbf{Q}_{\bar{s}_i\bar{v}_i} = q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)'$, then $\mathbf{Z}_i$ can be written as

$$\mathbf{Z}_i = \mathbf{C}_i\left[q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)'\right]\mathring{\mathbf{B}}_0 - \mathbf{C}_i\left\{q^{-1}\sum_{\ell=1}^{q}E\left[\left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)'\right]\right\}\mathring{\mathbf{B}}_0. \tag{28}$$

Writing (27) in vec form

$$\left(\mathbf{I}_r\otimes\mathbf{Q}_{\bar{w}\bar{w}}\right)\sqrt{n}T\ vec\left(\hat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right) = n^{-1/2}\sum_{i=1}^{n}\left(\mathring{\mathbf{B}}_0'\otimes\mathbf{C}_i\right)\left[\bar{\boldsymbol{\xi}}_{iq} - E\left(\bar{\boldsymbol{\xi}}_{iq}\right)\right] + O_p\left(\frac{\sqrt{n}}{T}\right),$$

where $\bar{\boldsymbol{\xi}}_{iq} = q^{-1}\sum_{\ell=1}^{q}\boldsymbol{\xi}_{i\ell}$, and $\boldsymbol{\xi}_{i\ell} = vec\left[\left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)'\right] = \left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)\otimes\left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right)$. Under Assumption 1, $\bar{\mathbf{v}}_{i\ell}$, $\bar{\mathbf{s}}_{i\ell}$, $\bar{\mathbf{v}}_{i\circ}$ and $\bar{\mathbf{s}}_{i\circ}$ are distributed independently over i. Hence, $\bar{\boldsymbol{\xi}}_{iq}$ is distributed independently over i. In addition, $\sup_i\left\|\left(\mathring{\mathbf{B}}_0'\otimes\mathbf{C}_i\right)\right\| = \left\|\mathring{\mathbf{B}}_0\right\|\sup_i\|\mathbf{C}_i\| < K$ and using Lemma 8, it follows that for $(n,T)\to\infty$, jointly such that $T\approx n^d$ and $d>0$, we have

$$n^{-1/2}\sum_{i=1}^{n}\left(\mathring{\mathbf{B}}_0'\otimes\mathbf{C}_i\right)\left[\bar{\boldsymbol{\xi}}_{iq} - E\left(\bar{\boldsymbol{\xi}}_{iq}\right)\right]\to_d N\left(\mathbf{0},\boldsymbol{\Omega}_q\right), \tag{29}$$

where

$$\boldsymbol{\Omega}_q = lim_{n,T\to\infty}\left[n^{-1}\sum_{i=1}^{n}\left(\mathring{\mathbf{B}}_0'\otimes\mathbf{C}_i\right)\boldsymbol{\Omega}_{\bar{\xi}_{iq}}\left(\mathring{\mathbf{B}}_0\otimes\mathbf{C}_i'\right)\right], \tag{30}$$

and

$$\boldsymbol{\Omega}_{\bar{\xi}_{iq}} = Var\left(\bar{\boldsymbol{\xi}}_{iq}\right) = q^{-2}\sum_{\ell=1}^{q}\sum_{\ell'=1}^{q}Cov\left(\boldsymbol{\xi}_{i\ell},\boldsymbol{\xi}_{i\ell'}\right). \tag{31}$$

Using (29) in (27) yields asymptotic normality of $\hat{\mathring{\mathbf{B}}}$, which is formally established in the following theorem.

Theorem 2 *Consider the panel data model for the $m \times 1$ vector $\mathbf{w}_{it}$, given by (1) without interactive time effects ($\mathbf{G}_i = \mathbf{0}$). Suppose that Assumptions 1 to 3 hold, m and q (≥ 2) are fixed integers, and the number of long-run relations, r_0 $(m > r_0 > 0)$ is known. Suppose further that the long-run relations, $\mathring{\mathbf{B}}_0$, of interest are subject to the exact identifying restrictions, $\mathbf{R}\mathring{\mathbf{B}}_0 = \mathbf{A}$, given by (22), and consider the PME estimator of $\mathring{\mathbf{B}}_0$ given by*

$$\hat{\mathring{\mathbf{B}}}_0 = \hat{\mathbf{B}}_0 \left(\mathbf{R}\hat{\mathbf{B}}_0\right)^{-1} \mathbf{A},$$

where $\hat{\mathbf{B}}_0 = \left(\hat{\boldsymbol{\beta}}_{10}, \hat{\boldsymbol{\beta}}_{20}, ..., \hat{\boldsymbol{\beta}}_{r_0,0}\right)$ are the first r_0 orthonormalized eigenvectors of $\mathbf{Q}_{\bar{w}\bar{w}}$ defined by (9). Then

$$\sqrt{n}T\left(\mathbf{I}_{r_0} \otimes \mathbf{Q}_{\bar{w}\bar{w}}\right) vec\left(\hat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right) \rightarrow_d N\left(\mathbf{0}, \boldsymbol{\Omega}_q\right), \tag{32}$$

as $(n, T) \rightarrow \infty$, jointly such that $T \approx n^d$ for $d > 1/2$, where $\boldsymbol{\Omega}_q$ is defined by (30), and $\mathbf{Q}_{\bar{w}\bar{w}} \rightarrow_p \frac{(q-1)}{6q}\boldsymbol{\Psi}$. (see (19)). A proof is provided in the Appendix.

Theorem 2 can be readily used to obtain asymptotic distribution of any linear combination of $\hat{\mathring{\mathbf{B}}}_0$. One notable case of interest is to consider the exact identifying restrictions $\mathring{\mathbf{B}}_{0,1} = \mathbf{I}_{r_0}$ discussed above. Under these restrictions $\mathring{\mathbf{B}}_{0,1} = \hat{\mathring{\mathbf{B}}}_{0,1} = \mathbf{I}_{r_0}$, and $\hat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0 = \begin{pmatrix} \mathbf{0} & \hat{\boldsymbol{\Theta}}' - \boldsymbol{\Theta}_0' \end{pmatrix}$. Partitioning $\mathbf{Q}_{\bar{w}\bar{w}}$ and $\boldsymbol{\Omega}_z$ accordingly, we have

$$\sqrt{n}T\mathbf{Q}_{\bar{w}\bar{w}}\left(\hat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right) = \begin{pmatrix} \mathbf{Q}_{11,\bar{w}\bar{w}} & \mathbf{Q}_{12,\bar{w}\bar{w}} \\ \mathbf{Q}_{21,\bar{w}\bar{w}} & \mathbf{Q}_{22,\bar{w}\bar{w}} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \sqrt{n}T\left(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\right) \end{pmatrix},$$

where $\mathbf{Q}_{22,\bar{w}\bar{w}} \rightarrow_p \frac{(q-1)}{6q}\boldsymbol{\Psi}_{22}$, and $\boldsymbol{\Psi}_{22}$ is the $(m - r_0) \times (m - r_0)$ lower right block of $\boldsymbol{\Psi}$ which is positive definite (see Proposition 2). We obtain the following corollary.

Corollary 1 *Suppose assumptions of Theorem 2 hold and consider the exact identifying restrictions $\mathring{\mathbf{B}}_{0,1} = \hat{\mathring{\mathbf{B}}}_{0,1} = \mathbf{I}_{r_0}$. Suppose r_0 is known, $q \geq 2$, and let $\hat{\boldsymbol{\Theta}}$ and $\boldsymbol{\Theta}_0$ be the lower $(m - r_0) \times r_0$ block of $\hat{\mathring{\mathbf{B}}}_0$ and $\mathring{\mathbf{B}}_0$, respectively. Then,*

$$\sqrt{n}T\ vec\left(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\right) \rightarrow_d N\left(\mathbf{0}, \boldsymbol{\Omega}_{\theta q}\right), \tag{33}$$

as $n, T \to \infty$, jointly such that $T \approx n^d$ for $d > 1/2$,where

$$\mathbf{\Omega}_{\theta q} = \left(\frac{6q}{q-1}\right)^2 \left(\mathbf{I}_r \otimes \mathbf{\Psi}_{22}^{-1}\right) \mathbf{\Omega}_{q,22} \left(\mathbf{I}_r \otimes \mathbf{\Psi}_{22}^{-1}\right), \tag{34}$$

$\mathbf{\Omega}_{q,22}$ is the $(m-r_0) \times (m-r_0)$ lower right block of $\mathbf{\Omega}_q$ given by (30), and $\mathbf{\Psi}_{22}$ is the $(m-r_0) \times (m-r_0)$ lower right block of $\mathbf{\Psi}$.

5.3 Estimation of the asymptotic covariance of $\hat{\mathbf{\Theta}}$

To consistently estimate $\mathbf{\Omega}_{\theta q}$ we need a consistent estimator of $\mathbf{\Omega}_{q,22}$, since $[(q-1)/6q]\, \mathbf{\Psi}_{22}$ can be consistently estimated by $\mathbf{Q}_{\bar{w}\bar{w},22}$. Consider $\mathbf{\Omega}_q$ given by (30). Using (8) note that

$$\begin{aligned}
\left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}\right)\left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}\right)' \boldsymbol{\beta}_0 &= \left[\mathbf{C}_i\left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right) + \left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)\right]\left[\left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right)' \mathbf{C}_i' + \left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)'\right] \mathring{\mathbf{B}}_0, \\
&= \mathbf{C}_i\left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)' \mathring{\mathbf{B}}_0 + \left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)' \mathring{\mathbf{B}}_0.
\end{aligned}$$

Let $\boldsymbol{\zeta}_{i\ell} = vec\left[\left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}\right)\left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}\right)' \mathring{\mathbf{B}}_0\right]$, and $\boldsymbol{\eta}_{i\ell} = vec\left[\left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)'\right]$, where recall that $\boldsymbol{\eta}_{i\ell} = O_p(T^{-1})$ uniformly in i and ℓ. Then $\boldsymbol{\zeta}_{i\ell} = \left(\mathring{\mathbf{B}}_0' \otimes \mathbf{C}_i\right) \boldsymbol{\xi}_{i\ell} + \left(\mathring{\mathbf{B}}_0' \otimes \mathbf{I}_m\right) \boldsymbol{\eta}_{i\ell}$, and

$$\bar{\boldsymbol{\zeta}}_{iq} = q^{-1} \sum_{\ell=1}^{q} \boldsymbol{\zeta}_{i\ell} = \left(\mathring{\mathbf{B}}_0' \otimes \mathbf{C}_i\right)\left(q^{-1} \sum_{\ell=1}^{q} \boldsymbol{\xi}_{i\ell}\right) + \left(\mathring{\mathbf{B}}_0' \otimes \mathbf{I}_m\right)\left(q^{-1} \sum_{\ell=1}^{q} \boldsymbol{\eta}_{i\ell}\right) = \left(\mathring{\mathbf{B}}_0' \otimes \mathbf{C}_i\right) \bar{\boldsymbol{\xi}}_{iq} + O_p(T^{-1}),$$

where $\bar{\boldsymbol{\xi}}_{iq} = q^{-1} \sum_{\ell=1}^{q} \boldsymbol{\xi}_{i\ell}$, as before. It follows

$$n^{-1} \sum_{i=1}^{n} \bar{\boldsymbol{\zeta}}_{iq} \bar{\boldsymbol{\zeta}}_{iq}' = n^{-1} \sum_{i=1}^{n} \left(\mathring{\mathbf{B}}_0' \otimes \mathbf{C}_i\right) \bar{\boldsymbol{\xi}}_{iq} \bar{\boldsymbol{\xi}}_{iq}' \left(\mathring{\mathbf{B}}_0 \otimes \mathbf{C}_i'\right) + O_p(T^{-1}),$$

and $E\left(\bar{\boldsymbol{\xi}}_{iq}\right) = O(T^{-1})$. Hence, as $n, T \to \infty$ jointly such that $T \approx n^d$ for $d > 1/2$,

$$n^{-1} \sum_{i=1}^{n} \bar{\boldsymbol{\zeta}}_{iq} \bar{\boldsymbol{\zeta}}_{iq}' \to_p \; lim_{n,T} \left[n^{-1} \sum_{i=1}^{n} \left(\mathring{\mathbf{B}}_0' \otimes \mathbf{C}_i\right) E\left(\bar{\boldsymbol{\xi}}_{iq} \bar{\boldsymbol{\xi}}_{iq}'\right) \left(\mathring{\mathbf{B}}_0 \otimes \mathbf{C}_i'\right)\right] = \mathbf{\Omega}_q.$$

Therefore, $\boldsymbol{\Omega}_q$ can be consistently estimated by the $m^2 \times m^2$ matrix

$$\hat{\boldsymbol{\Omega}}_q = n^{-1} \sum_{i=1}^{n} \widehat{\bar{\boldsymbol{\zeta}}}_{iq} \widehat{\bar{\boldsymbol{\zeta}}}_{iq}' = \begin{pmatrix} \hat{\boldsymbol{\Omega}}_{q,11} & \hat{\boldsymbol{\Omega}}_{q,12} \\ \hat{\boldsymbol{\Omega}}_{q,21} & \hat{\boldsymbol{\Omega}}_{q,22} \end{pmatrix},$$

where

$$\widehat{\bar{\boldsymbol{\zeta}}}_{iq} = q^{-1} \sum_{\ell=1}^{q} vec \left[(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}) (\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ})' \widehat{\mathring{\mathbf{B}}_0} \right] = q^{-1} \sum_{\ell=1}^{q} \left[\widehat{\mathring{\mathbf{B}}}_0' (\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}) \otimes \mathbf{I}_m \right] (\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}).$$

Let $\widehat{\mathring{\mathbf{B}}_0}' (\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}) = \widehat{\bar{\mathbf{E}}}_{i\ell}$, the $r \times 1$ vector of error corrections for the sub-sample ℓ, then

$$\hat{\boldsymbol{\Omega}}_q = n^{-1} \sum_{i=1}^{n} q^{-2} \sum_{\ell=1}^{q} \sum_{\ell'=1}^{q} \left(\widehat{\bar{\mathbf{E}}}_{i\ell} \otimes \mathbf{I}_m \right) (\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}) (\bar{\mathbf{w}}_{i\ell'} - \bar{\mathbf{w}}_{i\circ})' \left(\widehat{\bar{\mathbf{E}}}_{i\ell'}' \otimes \mathbf{I}_m \right). \tag{35}$$

Using the above results we now have

$$\widehat{Var \left[vec \left(\hat{\boldsymbol{\Theta}} \right) \right]} = \frac{1}{nT^2} \mathbf{Q}_{\bar{w}\bar{w},22}^{-1} \hat{\boldsymbol{\Omega}}_{q,22} \mathbf{Q}_{\bar{w}\bar{w},22}^{-1}. \tag{36}$$

When $r_0 = 1$, estimate of the error correction term $\widehat{\mathring{\mathbf{B}}}_0' (\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}) = \widehat{\bar{e}}_{i\ell}$ is a scalar and we can write $\widehat{\bar{\boldsymbol{\zeta}}}_{iq} = q^{-1} \sum_{\ell=1}^{q} (\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}) \widehat{\bar{e}}_{i\ell}$, and

$$\begin{aligned} \hat{\boldsymbol{\Omega}}_q &= n^{-1} \sum_{i=1}^{n} \left[q^{-1} \sum_{\ell=1}^{q} (\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}) \widehat{\bar{e}}_{i\ell} \right] \left[q^{-1} \sum_{\ell=1}^{q} (\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ})' \widehat{\bar{e}}_{i\ell} \right], \\ &= n^{-1} \sum_{i=1}^{n} \left[q^{-2} \sum_{\ell=1}^{q} \sum_{\ell'=1}^{q} (\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}) (\bar{\mathbf{w}}_{i\ell'} - \bar{\mathbf{w}}_{i\circ})' \widehat{\bar{e}}_{i\ell} \widehat{\bar{e}}_{i\ell'} \right], \end{aligned}$$

which resembles the robust covariance matrix estimator that arises in estimation of panel data models with short T and large n. Here q plays the role of T.

6 Estimation of r_0 by eigenvalue thresholding

Under Assumption 3, the true number of common long-run relations, r_0, is defined by $rank(\boldsymbol{\Psi}_n) = m - r_0 > 0$, where $\boldsymbol{\Psi}_n = n^{-1} \sum_{i=1}^{n} \mathbf{C}_i \boldsymbol{\Sigma}_i \mathbf{C}_i'$. Subject to this condition, $\boldsymbol{\Psi}_n \boldsymbol{\beta}_{j,0} = 0$, for $j = 1, 2, ..., r_0$, where $\boldsymbol{\beta}_{j,0}$ is the j^{th} long-run relation ($j \leq r_0$). The

$m \times r_0$ matrix of long-run relations is denoted by $\mathbf{B}_0$. It is also worth bearing in mind that under Assumption 3, $\mathbf{\Psi}_n \boldsymbol{\beta}_j \neq \mathbf{0}$, for $j = r_0 + 1, r_0 + 2, ...m$, namely cannot be spanned by the r_0 columns of $\mathbf{B}_0$. See (4) and (5). There is a clear shift in the ordered eigenvalues of $\mathbf{\Psi}_n$ from $\lambda_{r_0} = 0$ to $\lambda_{r_0+1} > 0$, which allows us to propose a thresholding estimator of r_0 applied to the eigenvalues of $\mathbf{Q}_{\bar{w}\bar{w}}$, noting that under our assumptions $\mathbf{Q}_{\bar{w}\bar{w}}$ tends to $\left(\frac{q-1}{6q}\right) \mathbf{\Psi}$ as n and $T \rightarrow \infty$. See result (19) of Proposition 1. Such an estimator can be written conveniently as

$$\hat{r} = \sum_{j=1}^{m} \mathcal{I}\left(\hat{\lambda}_j < C_T\right), \tag{37}$$

where $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq .... \leq \hat{\lambda}_m$ are ordered eigenvalues of $\mathbf{Q}_{\bar{w}\bar{w}}$, and $\mathcal{I}(\mathcal{A}) = 1$ if $\mathcal{A}$ is true or zero otherwise, and $C_T = KT^{-\delta}$, for some $\delta > 0$. This estimator is invariant to the ordering of the eigenvalues, but using the ordering helps with the exposition and the rationale behind the proofs.

To establish the consistency of $\hat{r}$ as an estimator of r_0, we first note that (37) can be written equivalently as

$$\hat{r} - r_0 = -\sum_{j=1}^{r_0} \mathcal{I}\left(\hat{\lambda}_j \geq C_T\right) + \sum_{j=r_0+1}^{m} \mathcal{I}\left(\hat{\lambda}_j < C_T\right), \tag{38}$$

which in turn yields:

$$E\left|\hat{r} - r_0\right| \leq \sum_{j=1}^{r_0} \Pr\left(\hat{\lambda}_j \geq C_T\right) + \sum_{j=r_0+1}^{m} \Pr\left(\hat{\lambda}_j < C_T\right). \tag{39}$$

Again noting the ordering of the eigenvalues, $\Pr\left(\hat{\lambda}_j \geq C_T\right) \leq \Pr\left(\hat{\lambda}_1 \geq C_T\right)$, for $j = 2, 3, ..., r_0$, and $\Pr\left(\hat{\lambda}_{r_0+1} < C_T\right) \geq \Pr\left(\hat{\lambda}_j < C_T\right)$, for $j = r_0 + 2, r_0 + 3, ..., m$. Using these results in (39) we have

$$E\left|\hat{r} - r_0\right| \leq r_0 \Pr\left(\hat{\lambda}_1 \geq C_T\right) + (m - r_0) \Pr\left(\hat{\lambda}_{r_0+1} < C_T\right). \tag{40}$$

Similarly,

$$E\left(\hat{r}-r_0\right)^2 \le r_0^2 \Pr\left(\hat{\lambda}_1 \ge C_T\right)+(m-r_0)^2 \Pr\left(\hat{\lambda}_{r_0+1} < C_T\right) \quad (41)$$
$$+2r_0(m-r_0)\sqrt{\Pr\left(\hat{\lambda}_1 \ge C_T\right)\Pr\left(\hat{\lambda}_{r_0+1} < C_T\right)}.$$

Also, by Markov inequality there exists $\epsilon > 0$ such that

$$\Pr\left(|\hat{r}-r_0| > \epsilon\right) \le (r_0/\epsilon) \Pr\left(\hat{\lambda}_1 \ge C_T\right)+\left[(m-r_0)/\epsilon\right] \Pr\left(\hat{\lambda}_{r_0+1} < C_T\right), \quad (42)$$

Again by Markov inequality $\Pr\left(\hat{\lambda}_1 \ge C_T\right) \le C_T^{-1} E\left(\hat{\lambda}_1\right)$, and using result (S.49) established in Lemma 7, we have

$$\Pr\left(\hat{\lambda}_1 \ge C_T\right) = O\left(C_T^{-1} n^{-1/2} T^{-2}\right). \quad (43)$$

Consider now $\Pr\left(\hat{\lambda}_{r_0+1} < C_T\right)$, and recall that

$$\mathbf{Q}_{\bar{w}\bar{w}}\hat{\boldsymbol{\beta}}_j = \hat{\lambda}_j \hat{\boldsymbol{\beta}}_j, \text{ and } \hat{\lambda}_j = \hat{\boldsymbol{\beta}}_j' \mathbf{Q}_{\bar{w}\bar{w}} \hat{\boldsymbol{\beta}}_j, \text{ for } j = r_0+1, r_0+2, ..., m, \quad (44)$$

with associated population values given by (see Assumption 3).

$$\frac{(q-1)}{6q}\boldsymbol{\Psi}_n \boldsymbol{\beta}_j = \lambda_j \boldsymbol{\beta}_j, \text{ and } \frac{(q-1)}{6q}\boldsymbol{\beta}_j' \boldsymbol{\Psi}_n \boldsymbol{\beta}_j = \lambda_j, \text{ for } j = r_0+1, r_0+2, ..., m. \quad (45)$$

where $\boldsymbol{\Psi}_n = n^{-1}\sum_{i=1}^n \mathbf{C}_i \boldsymbol{\Sigma}_i \mathbf{C}_i' \succeq 0$, and $\boldsymbol{\beta}_j'\boldsymbol{\beta}_j = 1$. Furthermore, $\lambda_j > 0$ and $\boldsymbol{\Psi}_n \boldsymbol{\beta}_j \neq 0$, for $j > r_0$. But using (13) and results established in Lemma 3 and Proposition 1 we have $E\left(\mathbf{Q}_{\bar{w}\bar{w}}\right) = \frac{(q-1)}{6q}\boldsymbol{\Psi}_n + O\left(T^{-2}\right)$, and

$$\tilde{\mathbf{Q}}_{\bar{w}\bar{w}} = \mathbf{Q}_{\bar{w}\bar{w}} - E\left(\mathbf{Q}_{\bar{w}\bar{w}}\right) = n^{-1}\sum_{i=1}^n \mathbf{C}_i\left[\mathbf{Q}_{\bar{s}_i\bar{s}_i} - E\left(\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right)\right]\mathbf{C}_i' + O_p\left(T^{-1}\right) = O_p\left(n^{-1/2}\right) + O_p\left(T^{-1}\right).$$

Using the above results then (45) can be written equivalently as $E\left(\mathbf{Q}_{\bar{w}\bar{w}}\right)\boldsymbol{\beta}_j = \lambda_j\boldsymbol{\beta}_j + O\left(T^{-2}\right)$, and $\lambda_j = \boldsymbol{\beta}_j' E\left(\mathbf{Q}_{\bar{w}\bar{w}}\right)\boldsymbol{\beta}_j + O\left(T^{-2}\right)$. Therefore, together with (44), we have

$$\hat{\lambda}_j - \lambda_j = \hat{\boldsymbol{\beta}}_j' \mathbf{Q}_{\bar{w}\bar{w}} \hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j' E\left(\mathbf{Q}_{\bar{w}\bar{w}}\right)\boldsymbol{\beta}_j + O\left(T^{-2}\right)$$

and rearranged as

$$\begin{aligned}& \left(\left[E\left(\mathbf{Q}_{\bar{w}\bar{w}}\right)-\mathbf{I}_m\lambda_j\right]\right)\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right)-\left(\hat{\lambda}_j-\lambda_j\right)\boldsymbol{\beta}_j \\ = & -\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\boldsymbol{\beta}_j+\left(\hat{\lambda}_j-\lambda_j\right)\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right)-\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right)+O\left(T^{-2}\right).\end{aligned} \tag{46}$$

Also, since $\boldsymbol{\beta}_j'E\left(\mathbf{Q}_{\bar{w}\bar{w}}\right)=\lambda_j\boldsymbol{\beta}_j'+O\left(T^{-2}\right)$, then

$$\hat{\lambda}_j-\lambda_j=\boldsymbol{\beta}_j'\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\boldsymbol{\beta}_j+2\lambda_j\boldsymbol{\beta}_j'\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right)+2\boldsymbol{\beta}_j'\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right)+\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right)'\mathbf{Q}_{\bar{w}\bar{w}}\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right)+O\left(T^{-2}\right). \tag{47}$$

where $\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}=O_p\left(n^{-1/2}\right)+O_p\left(T^{-1}\right)$. Pre-multiplying both sides of the above equations by $\sqrt{n}$ and stacking them in matrix notation we will have

$$\begin{pmatrix} 1 & -2\lambda_j\boldsymbol{\beta}_j' \\ -\boldsymbol{\beta}_j & E\left(\mathbf{Q}_{\bar{w}\bar{w}}\right)-\lambda_j\mathbf{I}_m \end{pmatrix}\begin{pmatrix} \sqrt{n}\left(\hat{\lambda}_j-\lambda_j\right) \\ \sqrt{n}\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right) \end{pmatrix}=\begin{pmatrix} \sqrt{n}\boldsymbol{\beta}_j'\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\boldsymbol{\beta}_j \\ -\sqrt{n}\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\boldsymbol{\beta}_j \end{pmatrix}+\mathbf{p}_{nT},$$

where

$$\mathbf{p}_{nT}=n^{-1/2}\begin{pmatrix} 2\boldsymbol{\beta}_j'\sqrt{n}\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\sqrt{n}\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right)+\sqrt{n}\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right)'\mathbf{Q}_{\bar{w}\bar{w}}\sqrt{n}\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right) \\ \sqrt{n}\left(\hat{\lambda}_j-\lambda_j\right)\sqrt{n}\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right)-\sqrt{n}\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\sqrt{n}\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right) \end{pmatrix}+O(n^{1/2}T^{-2}).$$

It then follows that $\mathbf{p}_{nT}$ is of lower order as compared to $\sqrt{n}\left(\hat{\lambda}_j-\lambda_j\right)$ and $\sqrt{n}\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right)$, and as $n,T\to\infty$ jointly, such that $T\approx n^d$ for $d>1/2$, we have

$$\boldsymbol{\Omega}_j\begin{pmatrix} \sqrt{n}\left(\hat{\lambda}_j-\lambda_j\right) \\ \sqrt{n}\left(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\right) \end{pmatrix}=\begin{pmatrix} \sqrt{n}\boldsymbol{\beta}_j'\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\boldsymbol{\beta}_j \\ -\sqrt{n}\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\boldsymbol{\beta}_j \end{pmatrix}+o_p(1),$$

where $\boldsymbol{\Omega}_j=\begin{pmatrix} 1 & -2\lambda_j\boldsymbol{\beta}_j' \\ -\boldsymbol{\beta}_j & E\left(\mathbf{Q}_{\bar{w}\bar{w}}\right)-\lambda_j\mathbf{I}_m \end{pmatrix}$. Using partitioned inverse it is easily seen that $\boldsymbol{\Omega}_j$ has an inverse if $\boldsymbol{\Upsilon}_j=E\left(\mathbf{Q}_{\bar{w}\bar{w}}\right)-\lambda_j\mathbf{I}_m-2\lambda_j\boldsymbol{\beta}_j\boldsymbol{\beta}_j'$ is invertible. To check the invertibility of $\boldsymbol{\Upsilon}_j$ we note that $\boldsymbol{\Upsilon}_j\boldsymbol{\beta}_j=-2\lambda_j\boldsymbol{\beta}_j+O\left(T^{-2}\right)$, and since $\lambda_j>0$ for $j>r_0$ it then follows that $\boldsymbol{\Upsilon}_j$ must be invertible. Therefore, we can now solve for $\sqrt{n}\left(\hat{\lambda}_j-\lambda_j\right)$, in terms of a linear combination of $\sqrt{n}\boldsymbol{\beta}_j'\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\boldsymbol{\beta}_j$ and $\sqrt{n}\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\boldsymbol{\beta}_j$, and its asymptotic distribution can be derived accordingly. Consider the asymptotic distribution of these two terms, and note that since $\mathbf{Q}_{\bar{s}_i\bar{s}_i}=T^{-1}q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{s}}_{i\ell}-\bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{s}}_{i\ell}-\bar{\mathbf{s}}_{i\circ}\right)'$, then

$$\sqrt{n}\boldsymbol{\beta}_j'\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\boldsymbol{\beta}_j = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\beta}_j'\mathbf{C}_i\left[\mathbf{Q}_{\bar{s}_i\bar{s}_i} - E\left(\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right)\right]\mathbf{C}_i'\boldsymbol{\beta}_j = n^{-1/2}\sum_{i=1}^{n} q^{-1}\sum_{\ell=1}^{q}\left[\zeta_{ij,\ell}^2 - E\left(\zeta_{ij,\ell}^2\right)\right],$$

where $\zeta_{ij,\ell} = T^{-1/2}\boldsymbol{\beta}_j'\mathbf{C}_i\left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right)$. Also by Minkowski inequality

$$\left(E\left|\zeta_{ij,\ell}\right|^{4+\epsilon}\right)^{1/4+\epsilon} \leq \left\|\mathbf{C}_i\right\|\left(\left[E\left\|T^{-1/2}\bar{\mathbf{s}}_{i\ell}\right\|^{4+\epsilon}\right]^{1/4+\epsilon} + \left[E\left\|T^{-1/2}\bar{\mathbf{s}}_{i\ell}\right\|^{4+\epsilon}\right]^{1/4+\epsilon}\right).$$

Using Lemma 4 in the supplement we have $\sup_{i,\ell}\left\|T^{-1/2}\bar{\mathbf{s}}_{i\ell}\right\|^{4+\epsilon} < K$, and hence $\left(E\left|\zeta_{ij,\ell}\right|^{4+\epsilon}\right)^{1/4+\epsilon} < K$ and $\sqrt{n}\boldsymbol{\beta}_j'\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\boldsymbol{\beta}_j \rightarrow_d N(0,\varpi_j^2)$, for some $\varpi_j^2 > 0$. Similarly, it follows that all m elements of $\sqrt{n}\tilde{\mathbf{Q}}_{\bar{w}\bar{w}}\boldsymbol{\beta}_j = n^{-1/2}\sum_{i=1}^{n}\mathbf{C}_i\left[\mathbf{Q}_{\bar{s}_i\bar{s}_i} - E\left(\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right)\right]\mathbf{C}_i'\boldsymbol{\beta}_j$ are asymptotically normally distributed with zero means and finite variances. Therefore, it also follows that $\sqrt{n}\left(\hat{\lambda}_j - \lambda_j\right) \overset{a}{\sim} N(0,\varpi_{\lambda_j}^2)$, for some $\varpi_{\lambda_j}^2 > 0$. Using this result and setting $j = r_0 + 1$ we have

$$\Pr\left(\hat{\lambda}_{r_0+1} < C_T\right) \overset{a}{\sim} \Phi\left[\frac{\sqrt{n}}{\varpi_{\lambda_{r_0+1}}}\left(C_T - \lambda_{r_0+1}\right)\right] = \Phi\left[\frac{-\lambda_{r_0+1}\sqrt{n}}{\varpi_{\lambda_{r_0+1}}}\left(1 - \frac{C_T}{\lambda_{r_0+1}}\right)\right].$$

Since $\lambda_{r_0+1} > 0$, then $\Pr\left(\hat{\lambda}_{r_0+1} < C_T\right) \rightarrow 0$, as $n \rightarrow \infty$ if $C_T < \lambda_{r_0+1}$. Recall also from (43) that $\Pr\left(\hat{\lambda}_1 \geq C_T\right) = O\left(C_T^{-1}n^{-1/2}T^{-2}\right)$, and $\Pr\left(\hat{\lambda}_1 \geq C_T\right) \rightarrow 0$ if C_T^{-1} does not rise too fast with T. Setting $T = KT^{-\delta}$, and recalling that $n \approx T^{1/d}$, these two conditions on C_T are met if $0 < \delta < 2 + (1/2d)$. Using this result in (41) and (42), it follows that $\hat{r} \rightarrow_p r_0$ and $E\left(\hat{r} - r_0\right)^2 \rightarrow 0$, as n and $T \rightarrow \infty$, if δ is set close to zero such that $KT^{-\delta} < \lambda_{r_0+1}$. These results also suggest that the probability of selecting too many long-run relations is more affected by n than T, and the probability of selecting too few long-run relations tends to zero much faster with T than with n. We need large n for not selecting more than r_0 long-run relations. This latter probability, $\Pr\left(\hat{\lambda}_j < C_T\right)$ for $j > r_0$, is also affected by the size of λ_{r_0+1} which measures the degree to which there is a transition from a stationary linear combination under which $\lambda_j = 0$ for $j \leq r_0$, to $\lambda_j > 0$ for $j > r_0$.

Our theoretical derivations also provide some insight on how to set K and δ when choosing C_T. It is clear that δ need not be too large, so long as $K \approx \lambda_{r_0+1}$. In practice, this can be achieved approximately, by appropriate scaling of the observations, $\mathbf{w}_{it}$, as discussed below.

Remark 5 *The above derivations also suggest that our proposed selection/estimation procedure would be valid even if there were near stationary relations, namely if* $\lambda_{r_0+1} \approx n^{-b}$ *for some* $b > 0$. *Then* $\Pr\left(\hat{\lambda}_{r_0+1} < C_T\right) \rightarrow 0$, *so long as* $b < 1/2$, *which could be viewed as local-to-zero eigenvalue. Such cases will not be pursued in this paper, where we require* $\lambda_{r_0+1} > 0$.

7 Allowing for interactive time effects

The model with interactive time effects is given by (1), which we reproduce here for convenience:

$$\mathbf{w}_{it} = \mathbf{a}_i + \mathbf{G}_i \mathbf{f}_t + \mathbf{C}_i \mathbf{s}_{it} + \mathbf{v}_{it}, \text{ for } t = 1, 2, ..., T; i = 1, 2, ..., n, \tag{48}$$

where $\mathbf{f}_t$ is an $m_f \times 1$ vector of latent factors with $\mathbf{G}_i$ the associated $m \times m_f$ matrix of factor loadings. We assume $\mathbf{f}_t$ is covariance stationary, and treat the factor loadings, $\mathbf{G}_i$, as nonstochastic, without placing any restrictions on them, besides being uniformly bounded. Hence, latent factors could be strong, semi-strong or weak. However, we do not allow for the possibility of unit roots in latent factors, and therefore do not consider the case of cointegration between $\mathbf{w}_{it}$ and latent factors. Specifically, we assume:

Assumption 4 *(i) The* $m_f \times 1$ *vector of latent factors,* $\mathbf{f}_t$, *is given by* $\mathbf{f}_t = \mathbf{\Phi}_f(L)\boldsymbol{\varepsilon}_{ft} = \sum_{j=0}^{\infty} \mathbf{\Phi}_{f\ell} L^{\ell} \boldsymbol{\varepsilon}_{f,t-\ell}$, *for* $t = ...0, 1, 2, ..., T$, *where* $\boldsymbol{\varepsilon}_{ft}$ *is an* $m_f \times 1$ *vector of errors distributed independently over* t *with* $E(\boldsymbol{\varepsilon}_{ft}) = \mathbf{0}$, *and* $\sup_{it} E \left\|\boldsymbol{\varepsilon}_{ft}\right\|^{4+\epsilon} < K$, *for some* $\epsilon > 0$. $\boldsymbol{\varepsilon}_{ft}$ *is independently distributed of* $\mathbf{u}_{it'}$, *for all* i, t, t'. *(ii) The* $m \times m_f$ *coefficient matrices* $\mathbf{\Phi}_{f\ell}$ *are non-stochastic constants such that* $\|\mathbf{\Phi}_{f\ell}\| < K\rho^{\ell}$, *where* ρ *lies in the range* $0 < \rho < 1$. *(iii)* $\mathbf{G}_i$ *are nonstochastic constants such that* $\sup_i \|\mathbf{G}_i\| < K$.

Under Assumption 4, $E\left(\mathbf{G}_i \mathbf{f}_t\right)$ is time invariant, and together with Assumption 1, $E\left(\mathbf{w}_{it}\right)$ continues to be time invariant. Subtracting sub-sample time averages from the full sample time average, $\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}$, will therefore continue to remove unit-specific means. More specifically, under (48) $\mathbf{Q}_{\bar{w}_i \bar{w}_i}$ given by (11) has the following expension

$$\mathbf{Q}_{\bar{w}_i \bar{w}_i} = \mathbf{C}_i \mathbf{Q}_{\bar{s}_i \bar{s}_i} \mathbf{C}_i' + \mathbf{C}_i \mathbf{Q}_{\bar{s}_i \bar{v}_i} + \mathbf{C}_i \mathbf{Q}_{\bar{s}_i \bar{f}_i} + \mathbf{Q}_{\bar{v}_i \bar{s}_i}' \mathbf{C}_i' + \mathbf{Q}_{\bar{f}_i \bar{s}_i}' \mathbf{C}_i + \mathbf{Q}_{\bar{f}_i \bar{f}_i} + \mathbf{Q}_{\bar{f}_i \bar{v}_i} + \mathbf{Q}_{\bar{v}_i \bar{f}_i} + \mathbf{Q}_{\bar{v}_i \bar{v}_i}, \tag{49}$$

where $\mathbf{Q}_{\bar{f}_i\bar{s}_i}=(Tq)^{-1}\sum_{\ell=1}^{q}\mathbf{G}_i\left(\bar{\mathbf{f}}_\ell-\bar{\mathbf{f}}_\circ\right)(\bar{\mathbf{s}}_{i\ell}-\bar{\mathbf{s}}_{i\circ})'=\mathbf{Q}'_{\bar{s}_i\bar{f}_i}$,

$$\mathbf{Q}_{\bar{f}_i\bar{v}_i}=(Tq)^{-1}\sum_{\ell=1}^{q}\mathbf{G}_i\left(\bar{\mathbf{f}}_\ell-\bar{\mathbf{f}}_\circ\right)(\bar{\mathbf{v}}_{i\ell}-\bar{\mathbf{v}}_{i\circ})'=\mathbf{Q}'_{\bar{v}_i\bar{f}_i}$$

$\mathbf{Q}_{\bar{f}_i\bar{f}_i}=(Tq)^{-1}\sum_{\ell=1}^{q}\mathbf{G}_i\left(\bar{\mathbf{f}}_\ell-\bar{\mathbf{f}}_\circ\right)\left(\bar{\mathbf{f}}_\ell-\bar{\mathbf{f}}_\circ\right)'\mathbf{G}'_i$, and the terms not involving the latent factor are as before. By Lemma 3 in the supplement we have

$$\sup_i E\left\|\mathbf{Q}_{\bar{f}_i\bar{f}_i}\right\|=O\left(T^{-2}\right),\ \sup_i E\left\|\mathbf{Q}_{\bar{f}_i\bar{v}_i}\right\|=O\left(T^{-2}\right),\text{ and }\sup_i E\left\|\mathbf{Q}_{\bar{f}_i\bar{s}_i}\right\|=O\left(T^{-1}\right). \tag{50}$$

Pooling over i, $\mathbf{Q}_{\bar{w}\bar{w}}=n^{-1}\sum_{i=1}^{n}\mathbf{Q}_{\bar{w}_i\bar{w}_i}$, and using the above results together with those already established in (16) and (18) now yields

$$\mathbf{Q}_{\bar{w}\bar{w}}=\frac{(q-1)}{6q}\boldsymbol{\Psi}_n+O_p\left(n^{-1/2}\right)+O_p\left(T^{-1}\right),\text{ and }\mathbf{Q}_{\bar{w}\bar{w}}\boldsymbol{\beta}_{j0}=O_p\left(n^{-1/2}\right)+O_p\left(T^{-1}\right), \tag{51}$$

which is identical to results (19)-(20) in Proposition 1 for models without interactive time effects. Similarly, inclusion of interactive time effects does not alter the convergence rate of the exactly identified PME estimator $\hat{\mathring{\mathbf{B}}}_0$, given by (23), and its asymptotic distribution will remain correctly centered at zero. To see this consider the following expression for $\mathbf{Q}_{\bar{w}\bar{w}}\sqrt{n}T\left(\hat{\mathring{\mathbf{B}}}_0-\mathring{\mathbf{B}}_0\right)$, which extends (24) to panel data models with interactive time effects,

$$\begin{aligned}\mathbf{Q}_{\bar{w}\bar{w}}\sqrt{n}T\left(\hat{\mathring{\mathbf{B}}}_0-\mathring{\mathbf{B}}_0\right) \;=\; & -\left(n^{-1/2}\sum_{i=1}^{n}T\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0-\left(n^{-1/2}\sum_{i=1}^{n}T\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{f}_i}\right)\mathring{\mathbf{B}}_0\\ & -\frac{\sqrt{n}}{T}\left(n^{-1}\sum_{i=1}^{n}T^2\mathbf{Q}_{\bar{f}_i\bar{f}_i}\right)\mathring{\mathbf{B}}_0-\frac{\sqrt{n}}{T}\left[n^{-1}\sum_{i=1}^{n}T^2\left(\mathbf{Q}_{\bar{f}_i\bar{v}_i}+\mathbf{Q}_{\bar{v}_i\bar{f}_i}\right)\right]\mathring{\mathbf{B}}_0\\ & -\frac{\sqrt{n}}{T}\left(n^{-1}\sum_{i=1}^{n}T^2\mathbf{Q}_{\bar{v}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0+O_p\left(T^{-1}\right).\end{aligned}$$

The new terms involve matrices $\mathbf{Q}_{\bar{s}_i\bar{f}_i}$, $\mathbf{Q}_{\bar{f}_i\bar{f}_i}$ and $\mathbf{Q}_{\bar{f}_i\bar{v}_i}=\mathbf{Q}'_{\bar{v}_i\bar{f}_i}$. Using the bounds in (50), we obtain $E\left\|n^{-1}\sum_{i=1}^{n}T^2\mathbf{Q}_{\bar{f}_i\bar{f}_i}\right\|\leq n^{-1}\sum_{i=1}^{n}T^2E\left\|\mathbf{Q}_{\bar{f}_i\bar{f}_i}\right\|=O\left(T^{-2}\right)$, and similarly
$E\left\|n^{-1}\sum_{i=1}^{n}T^2\left(\mathbf{Q}_{\bar{f}_i\bar{v}_i}+\mathbf{Q}_{\bar{v}_i\bar{f}_{ii}}\right)\right\|<K$. In addition, by Lemma 5 recall that $n^{-1}\sum_{i=1}^{n}T^2\mathbf{Q}_{\bar{v}_i\bar{v}_i}=$

$O_p\left(n^{-1/2}\right)$. Using these results and noting that $\left\|\mathring{\mathbf{B}}_0\right\| < K$, we have

$$\mathbf{Q}_{\bar{w}\bar{w}}\sqrt{n}T\left(\widehat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right) = -\left(n^{-1/2}\sum_{i=1}^{n} T\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0 - \left(n^{-1/2}\sum_{i=1}^{n} T\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{f}_i}\right)\mathring{\mathbf{B}}_0 + O_p\left(\frac{\sqrt{n}}{T}\right). \tag{52}$$

To simplify the notations we set $\boldsymbol{\omega}_{it} = \mathbf{f}_{it} + \mathbf{v}_{it}$, and note that

$$\mathbf{Q}_{\bar{s}_i\bar{\omega}_i} = \mathbf{Q}_{\bar{s}_i\bar{v}_i} + \mathbf{Q}_{\bar{s}_i\bar{f}_i} = T^{-1}q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\boldsymbol{\omega}}_{i\ell} - \bar{\boldsymbol{\omega}}_{i\circ}\right)'.$$

Then (52) can be written as

$$\mathbf{Q}_{\bar{w}\bar{w}}\sqrt{n}T\left(\widehat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right) = -\left(n^{-1/2}\sum_{i=1}^{n} T\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{\omega}_i}\right)\mathring{\mathbf{B}}_0 + O_p\left(\frac{\sqrt{n}}{T}\right).$$

Let $\mathbf{Z}_i^* = \mathbf{C}_i\left[T\mathbf{Q}_{\bar{s}_i\bar{\omega}_i} - E\left(T\mathbf{Q}_{\bar{s}_i\bar{\omega}_i}\right)\right]\mathring{\mathbf{B}}_0$. Since $\mathbf{s}_{it}$ is independent of $\mathbf{f}_t$ and $E\left(\mathbf{s}_{it}\right) = \mathbf{0}$, we have $E\left(\mathbf{Q}_{\bar{s}_i\bar{f}_i}\right) = \mathbf{0}$, and using (26) we obtain $\|n^{-1}\sum_{i=1}^{n}\mathbf{C}_iE\left(T^2\mathbf{Q}_{\bar{s}_i\bar{\omega}_i}\right)\boldsymbol{\beta}_0\| < K$. It now follows that

$$\mathbf{Q}_{\bar{w}\bar{w}}\sqrt{n}T\left(\widehat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right) = -n^{-1/2}\sum_{i=1}^{n}\mathbf{Z}_i^* + O_p\left(\frac{\sqrt{n}}{T}\right), \tag{53}$$

where $\mathbf{Z}_i^* = \mathbf{C}_i\left[T\mathbf{Q}_{\bar{s}_i\bar{\omega}_i} - E\left(T\mathbf{Q}_{\bar{s}_i\bar{\omega}_i}\right)\right]\boldsymbol{\beta}_0$. Writing (53) in vec form

$$\left(\mathbf{I}_r\otimes\mathbf{Q}_{\bar{w}\bar{w}}\right)\sqrt{n}T\ vec\left(\widehat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right) = n^{-1/2}\sum_{i=1}^{n}\left(\mathring{\mathbf{B}}_0'\otimes\mathbf{C}_i\right)\left[\boldsymbol{\xi}_{iq}^* - E\left(\boldsymbol{\xi}_{iq}^*\right)\right] + O_p\left(\frac{\sqrt{n}}{T}\right),$$

where $\boldsymbol{\xi}_{iq}^*$ is the $m^2\times 1$ vector given by $\boldsymbol{\xi}_{iq}^* = q^{-1}\sum_{s=1}^{q}\left(\bar{\boldsymbol{\omega}}_{is} - \bar{\boldsymbol{\omega}}_i\right)\otimes\left(\bar{\mathbf{s}}_{is} - \bar{\mathbf{s}}_i\right)$. Although $\boldsymbol{\xi}_{iq}^*$ is not independent over i, it is a martingale difference sequence, and $n^{-1/2}\sum_{i=1}^{n}\left(\mathring{\mathbf{B}}_0'\otimes\mathbf{C}_i\right)\left[\boldsymbol{\xi}_{iq}^* - E\left(\boldsymbol{\xi}_{iq}^*\right)\right]$ converges to a normal distribution as $n, T\to\infty$, jointly. Therefore, $\widehat{\mathring{\mathbf{B}}}$ will continue to be asymptotically normally distributed, with its asymptotic distribution correctly centered at zero. Even though the variance of the asymptotic distribution of PME estimator in general depends on the interactive time effects, the estimator of the asymptotic variance given by (36) will continue to be consistent under Assumption 4, and inference can be carried out in the same way as in panel data models without interactive time effects.

Overall, we find that the PME estimator is robust to interactive time effects so long as Assumption 4 holds. Theorems and 3 and 4 in the supplement provide formal statements regarding consistency and the asymptotic normality of the PME estimators in presence of interactive time effects.

8 How to choose q and C_T

To implement the estimation of r_0 and the associated long-run relations we need to decide on q, and $C_T = CT^{-\delta}$ that enter the thresholding estimation of r_0.

8.1 Choice of q

To ensure that $\mathbf{Q}_{\bar{w}\bar{w}} = n^{-1}T^{-1}q^{-1}\sum_{i=1}^{n}\sum_{\ell=1}^{q}\left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}\right)\left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}\right)'$ does not depend on the fixed effects, given by $E(\mathbf{w}_{it}) = \mathbf{a}_i$, we need at least two sub-samples, namely $q \geq 2$. To reduce the variance of time series dependence of $\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}$ over the sub-samples we need T/q to be reasonably large. An optimum choice of q most likely depends on T, and not so much on n. To ensure that the mathematical derivations are manageable and transparent, so far we have assumed that q is fixed as $T \to \infty$. But to select q we must allow q to depend on T, denoted as q_T, and consider values of q_T that rise with T, but at a slower rate such that $q_T/T \to 0$. Analogous to the problem of selecting the lag order in time series literature, we conjecture setting q_T to rise at the rate of $T^{1/3}$, and set q_T to the lower integer part of $\max(2, T^{1/3})$. This would suggest that values of q_T equal to 2, 3 and 4 for values of $T = 20, 50$ and 100, respectively, considered in our Monte Carlo simulations reported in Section 9 below, where we consider the values of 2 and 4, to save space. For estimation of r_0, the choice of $q = 2$ works perfectly well for all values of T considered. But the higher value of $q = 4$ does seem to perform slightly better than $q = 2$ for estimation of long-run coefficients when $T = 100$.

8.2 Choices of C and δ for estimation of r_0

It is clear that eigenvalues of $\mathbf{Q}_{\bar{w}\bar{w}}$ depend on the scale of the observations, $\mathbf{w}_{it}$, and some form of scaling of data is required to reduce the sensitivity of the eigenvalues to scale. One could resort to cross validation procedures to set C and δ, but based on extensive Monte Carlo experiments, we have found that setting $C = 1$ works well if we base our selection procedure on the eigenvalues of the following correlation matrix

$$\mathbf{R}_{\bar{w}\bar{w}} = \left[diag\left(\mathbf{Q}_{\bar{w}\bar{w}}\right)\right]^{-1/2} \mathbf{Q}_{\bar{w}\bar{w}} \left[diag\left(\mathbf{Q}_{\bar{w}\bar{w}}\right)\right]^{-1/2}. \quad (54)$$

Accordingly, our proposed estimator of r_0 is given by

$$\tilde{r} = \sum_{j=1}^{m} \mathcal{I}\left(\tilde{\lambda}_j < T^{-\delta}\right), \quad (55)$$

where $\tilde{\lambda}_j$, for $j = 1, 2, ..., m$ are the eigenvalues of $\mathbf{R}_{\bar{w}\bar{w}}$.[2]

Also, based on our theoretical derivations any value of δ close to *zero* should work. In the Monte Carlo experiments we consider the values of $\delta = 1/4$ and $1/2$ and conclude that $\delta = 1/4$ is a good overall choice and cross-validation is not necessary for the implementation of our estimation strategy.

9 Monte Carlo Evidence

We investigate small sample properties of the proposed PME estimator with Monte Carlo experiments using both VARMA(1,1) and VAR(1) designs, with and without interactive time effects. The designs we consider are all special cases of the general linear model (1). We set $m = 3$, and generate the 3×1 vector $\mathbf{w}_{it}$ as $I(1)$ variables under three scenarios: non-cointegration, $r_0 = 0$, and $r_0 = 1$ and $r_0 = 2$ cointegrating relations. We consider sample size combinations, $T = (20, 50, 100)$ and $n = (50, 500, 1000, 3000)$, and report results for $q = 2$ and $q = 4$ (sub-samples), which are in line with our conjecture of setting q in line with the $\max(2, T^{1/3})$ rule. See Section 8.1.

When $r_0 = 1$, there are a range of alternative estimators of the cointegrating relation in the literature that we can use for comparison, most of which assume the direction of long-run causality is known. To accommodate existing estimators,

[2] Using the correlation matrix, $\mathbf{R}_{\bar{w}\bar{w}}$, has the advantage that it is unaffected by scaling, so long as the same scaling is used across all cross section units. It is not invariant if the scaling varies across units as well as across the variables. This issues is addressed in Section S5 of the supplement where we investigate the small sample sensitivity of $\tilde{r}$ to differential scaling of the variables across the units. We find that the small sample performance of $\tilde{r}$ as an estimator of r_0, is hardly affected by such differential scaling.

we distinguish between experiments based on long-run causal ordering. The PME estimator does not require the direction of long-run causality to be known. When $r_0 > 1$, to the best of our knowledge, there are no obvious alternative estimators of r_0 and the associated cointegrating relations in the panel cointegration literature that we can use. Accordingly, for the purpose of comparison, we report results using a mean group version of Johansen's maximum likelihood procedure, whereby we estimate r_0 and associated cointegrating vectors (if any), for all individual units in the panel separately, and report the frequency with which r_0 is selected by Johansen procedure across the n units, and the mean group estimates of the cointegrating coefficients and their standard errors.

Subsection 9.1 outlines the data generating processes (DGPs). Subsection 9.2 gives the results for estimates of r_0. Our results show near perfect performance for our proposed estimator of r_0, even for samples as small as $T = 20$ and $n = 50$. This is in line with the theory developed in Section 6. Subsection 9.3 reports the results for the coefficients of the long-run relations assuming r_0 is known, which is justified considering the near perfect performance of our estimator of r_0. The MC results provide simulation evidence that the PME estimator performs well in panels with n as large as $1,000$ and T as small as 20.

9.1 Data generating processes

We consider experiments with and without long-run relations. In the experiments with long-run relations, $\mathbf{w}_{it}$ is generated as

$$\Delta \mathbf{w}_{it} = \mathbf{d}_i - \mathbf{\Pi}_i \mathbf{w}_{i,t-1} + \mathbf{u}_{it} - \mathbf{\Theta}_i \mathbf{u}_{i,t-1}, \tag{56}$$

for $i = 1, 2, ..., n$, $t = 1, 2, ..., T$, where $\mathbf{\Pi}_i = \mathbf{A}_i \mathbf{B}_0'$, $\mathbf{A}_i$ is $m \times r_0$, $\mathbf{B}_0$ is $m \times r_0$. We set $\mathbf{d}_i = \mathbf{\Pi}_i \boldsymbol{\mu}_{iw}$ to ensure no linear trends in data. See, for example, Section 5.7 in Johansen (1995). The elements of $\boldsymbol{\mu}_{iw}$ are generated as $IIDN(0, 1)$. We consider both VAR(1) and VARMA(1,1) designs. For VAR(1) we set $\mathbf{\Theta}_i = \mathbf{0}$, and for VARMA(1,1) with set $\mathbf{\Theta}_i = diag(\theta_{ij}, j = 1, 2, ..., m)$, and generate θ_{ij} as $IIDU\,[-0.5, 0.5]$. The errors $\mathbf{u}_{it}$, are generated following both Gaussian and chi-squared distributions.

We initially consider $m = 3$ variables in $\mathbf{w}_{it} = (w_{it,1}, w_{it,2}, w_{it,3})'$, with both one and two long-run relations. For $r_0 = 1$, we set $\mathbf{B}_0 = \boldsymbol{\beta}_{1,0} = (1, 0, -1)'$ and $\mathbf{A}_i =$

$(a_{i,11}, a_{i,21}, ..., a_{i,31})'$. In this case, the long-run relation is given by

$$\boldsymbol{\beta}'_{1,0}\mathbf{w}_{it} = w_{it,1} - w_{it,3} = \beta_{11,0}w_{it,1} + \beta_{12,0}w_{it,2} + \beta_{13,0}w_{it,3}, \tag{57}$$

with $\beta_{11,0} = 1$, $\beta_{12,0} = 0$ and $\beta_{13,0} = -1$. We identify the long-run relation by imposing $\beta_{11,0} = 1$ and estimate $\beta_{12,0}$ and $\beta_{13,0}$. When $r_0 = 2$, we set

$$\mathbf{B}_0 = \left(\boldsymbol{\beta}_{1,0}, \boldsymbol{\beta}_{2,0}\right) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}, \text{ and } \mathbf{A}_i = \begin{pmatrix} a_{i,11} & a_{i,12} \\ a_{i,21} & a_{i22} \\ a_{i,31} & a_{i,32} \end{pmatrix}.$$

In this case, the two long-run relations are given by

$$\boldsymbol{\beta}'_{1,0}\mathbf{w}_{it} = w_{it,1} - w_{it,3} = \beta_{11,0}w_{it,1} + \beta_{12,0}w_{it,2} + \beta_{13,0}w_{it,3}, \tag{58}$$
$$\boldsymbol{\beta}'_{2,0}\mathbf{w}_{it} = w_{it,2} - w_{it,3} = \beta_{21,0}w_{it,1} + \beta_{22,0}w_{it,2} + \beta_{23,0}w_{it,3}, \tag{59}$$

where $\beta_{11,0} = \beta_{22,0} = 1$, $\beta_{12,0} = \beta_{21,0} = 0$, and $\beta_{13,0} = \beta_{23,0} = -1$. We identify these long-run relations by imposing $\beta_{11,0} = \beta_{22,0} = 1$ and $\beta_{12,0} = \beta_{21,0} = 0$, and we estimate $\beta_{13,0}$ and $\beta_{23,0}$.

We set the values of $\mathbf{A}_i$ to control the average speed of convergence towards the long-run relations. For example, in the case where $r_0 = 1$ then $\mathbf{B}'_0\mathbf{A}_i = \rho_i = a_{i,11} - a_{i,31}$ and ρ_i, for $i = 1, 2, ..., n$ are generated as $IIDU[0.1, 0.2]$ representing slow convergence, and as $IIDU[0.1, 0.3]$ representing moderate convergence. We then set $a_{i,21} = 0$, which leaves us with one free parameter in $\mathbf{A}_i$ which we use to set the system measures of the fit, $PR^2_{nT} = 0.2$ and 0.3, defined as a pooled R^2, given by equation (S.15) in the supplement. Since $a_{i,11}$ and $a_{i,31}$ are both nonzero, the long-run causality runs from $(w_{it,2}, w_{it,3})$ to $w_{it,1}$ as well as from $w_{it,1}$ to $w_{it,3}$. For $r_0 = 2$ the rate of convergence will depend on the eigenvalues of $\mathbf{I}_2 - \mathbf{B}'_0\mathbf{A}_i$ and the details of how we generate the elements of $\mathbf{A}_i$ are given in the supplement.

We also consider experiments with interactive time effects, which are obtained by augmenting the solution of (56) with $\mathbf{G}_i\mathbf{f}_t$, where $\mathbf{f}_t$ is and $m_f \times 1$ vector of latent factors. Each of these factors are generated as AR(1) process with a break in the AR coefficient, and the individual elements of the $m \times m_f$ matrix of factor loadings $\mathbf{G}_i$ are generated as $IIDU\,[0.0.4]$. We set $m_f = 4$. Specifically, we augment the general linear process versions of the above VARMA(1,1) and VAR(1) specifications with $\mathbf{G}_i\mathbf{f}_t$, namely

$$\mathbf{w}_{it} = \mathbf{w}_{i0} + \mathbf{G}_i\mathbf{f}_t + \mathbf{C}_i\mathbf{s}_{it} + \mathbf{C}_i^*(L)\mathbf{u}_{it},$$

where $\mathbf{C}_i$ and $\mathbf{C}_i^*(L)$ are obtained from

$$(\mathbf{I}_3 - \mathbf{\Psi}_iL)\left[\mathbf{C}_i + \mathbf{C}_i^*(L)(1-L)\right] = (\mathbf{I}_3 - \mathbf{\Theta}_iL)(1-L),$$

and $\mathbf{\Psi}_i = \mathbf{I}_3 - \mathbf{A}_i\mathbf{B}_0'$. See the supplement for further details.

For the experiments with no long-run relations ($r_0 = 0$) we generate $\Delta\mathbf{w}_{it}$ using the VAR(1) model in first-differences:

$$\Delta\mathbf{w}_{it} = \mathbf{\Phi}_i\Delta\mathbf{w}_{i,t-1} + \mathbf{u}_{it}, \tag{60}$$

for $i = 1, 2, ..., n$, $t = 1, 2, ..., T$, where $\mathbf{u}_{it} \sim IIDN(\mathbf{0}, \mathbf{\Sigma}_{ui})$. The elements of the covariance matrix $\mathbf{\Sigma}_{ui} = (\sigma_{i,\ell\ell'})$ are generated as $\sigma_{i,\ell\ell} = 1$ for $i = 1, 2, ..., n$ and $\ell = 1, 2, 3$, and $\sigma_{i,\ell\ell'} \sim IIDU(0, 0.5)$, for $\ell \neq \ell'$, and $i = 1, 2, ..., n$. We use a diagonal matrix for $\mathbf{\Phi}_i = (\phi_{i,\ell\ell'})$, with $\phi_{i,\ell\ell}$ elements on its diagonal, for $r = 1, 2, ..., m$. We consider three options for $\phi_{i,\ell\ell}$: (i) low values $\phi_{i,\ell\ell} \sim U[0, 0.8]$, (ii) moderate values $\phi_{i,\ell\ell} \sim U[0.7, 0.9]$, and (iii) high values $\phi_{i,\ell\ell} \sim U[0.80, 0.95]$. $\mathbf{w}_{it}$ is then obtained by cumulating $\Delta\mathbf{w}_{it}$ from the initial value $\mathbf{w}_{i,0} = 0$. Similarly to the experiment with long-run relations given by (56), the model (60) is a special case of (1). Specifically, (60) leads to $\mathbf{w}_{it} = \mathbf{w}_{i0} + \mathbf{G}_i\mathbf{f}_t + \mathbf{C}_i\mathbf{s}_{it} + \mathbf{C}_i^*(L)\mathbf{u}_{it}$, where $\mathbf{C}_i = (\mathbf{I}_m - \mathbf{\Phi}_i)^{-1}$, and $\mathbf{C}_i^*(L) = -\mathbf{\Phi}_i(\mathbf{I}_m - \mathbf{\Phi}_i)^{-1}(\mathbf{I}_m - \mathbf{\Phi}_iL)^{-1}$. For experiments without interactive time effects we set $\mathbf{G}_i = \mathbf{0}$, for all i.

In addition to the designs described above, we also consider a data generating process taken from Section 3.1 of Chudik, Pesaran, and Smith (2023a). For this $m = 2$, and $\mathbf{w}_{it} = (w_{1,it}, w_{2,it})'$ is generated as

$$\begin{aligned}\Delta w_{1,it} &= c_i - a_{i,11}\left(w_{1,i,t-1} - w_{2,i,t-1}\right) + u_{1,it},\\ \Delta w_{2,it} &= u_{2,it},\end{aligned}$$

where $a_{i,11} \sim IIDU\left[0.2, 0.3\right]$, $\mathbf{u}_{it} = (u_{1,it}, u_{2,it})'$ is heteroskedastic and cross-sectionally independent, and $u_{1,it}$ is correlated with $u_{2,it}$.[3] We use this design to see how PME compares to single-equation estimators that correctly assume long-run causality runs from $w_{2,it}$ to $w_{1,it}$. We would expect such estimators to perform reasonably well, particularly when T is large relative to n, and provide a good baseline to evaluate the performance of PME in settings favorable to the single equation techniques advanced in the literature.

In short, we have 71 experiments. 6 experiments with $m = 3$ variables and no long-run relations ($r_0 = 0$), given by 3 choices for the distribution of ϕ_{ij}, and interactive time effects are included or not. $64 = 2^5$ experiments feature $m = 3$ variables with long-run relations, given by combinations of the choice of model (VAR(1) or VARMA(1,1)), $r_0 = 1$ or 2, Gaussian or chi-square error distributions, moderate or slow speed of convergence, system measure of fit $PR^2_{nT} = 0.2$ or 0.3, and interactive time effects are included or not. In addition, we have one experiment with $m = 2$ variables, $r_0 = 1$ long-run relation and one one-way long-run causality taken directly from Chudik, Pesaran, and Smith (2023a).

To save space, we report only summary results that are averages across a number of selected experiments, with the results for the individual experiments available from the authors upon request. Section S4 of the supplement also provides further details of the Monte Carlo designs, how the processes are initialized, and the rationale behind the parameterization adopted. Additionally, Section S6 of the supplement shows that the results for estimation of r_0 the associated long-run relations are robust to GARCH and threshold autoregressive effects.

9.2 Small sample evidence on estimation of r_0

We summarize the Monte Carlo findings for the estimation of r_0 by the eigenvalue thresholding estimator $\tilde{r}$ given by (55), using $\delta = 1/4$ and $1/2$ in Tables 1-2. Table 1 reports selection frequencies of $\tilde{r} = 0$, 1, 2, 3 using VAR(1) based DGPs without interactive time effects, for the three cases of $r_0 = 0, 1$ or 2.[4] The theory indicated very

[3] $u_{1,it} = \sigma_{1i} e_{1,it}$, $u_{2,it} = \sigma_{2i} e_{2,it}$, $\sigma^2_{1,i}, \sigma^2_{2,i} \sim IIDU\,[0.8, 1.2]$,

$$\begin{pmatrix} e_{1,it} \\ e_{2,it} \end{pmatrix} \sim IIDN\left(\mathbf{0}_2, \mathbf{\Sigma}_e\right), \mathbf{\Sigma}_e \sim \begin{pmatrix} 1 & \rho_{ei} \\ \rho_{ei} & 1 \end{pmatrix}, \text{ and } \rho_{ei} \sim IIDU\,[0.3, 0.7]\,.$$

See Section 3.1 of Chudik, Pesaran, and Smith (2023a) for details.

[4] These results are averaged over 3 VAR(1) experiments in first differences ($r_0 = 0$), differeing in terms of autoregressive coefficients (low, medium and high values), and over 8 VAR(1) experiments

fast convergence of $\tilde{r}$ to r_0, and this is confirmed by the Monte Carlo simulations. The eigenvalue thresholding estimator $\tilde{r}$ correctly identifies r_0 in 100 per cent of cases, except for the very small sample sizes considered. For $n = 50$ and $T = 20$, we see 5% probability of $\tilde{r}$ overestimating the true number of long-run relations when the smaller exponent $\delta = 1/4$ is used in experiments with $r_0 = 0$ in Table1. For comparison, we also report selection frequency for the Johansen procedure using the trace statistic and the conventional nominal level of 5 percent.[5] Whereas there is a single estimate for the whole panel per replication using $\tilde{r}$, there are n such estimates per replication using the Johansen procedure ($\hat{r}_i$, $i = 1, 2, ..., n$). We simply use them all in calculating the selection frequency.[6] Hence n will not influence these results, but increasing T does improve the frequency with which the correct value is selected using the trace statistic. For $T = 100$, frequency of correctly estimating the number of long-run relations using the Johansen trace statistics is 69 percent for $r_0 = 0$, 82 percent for $r_0 = 1$, and only 31 percent for $r_0 = 2$ in Table 1.

Simulation results for the performance of $\tilde{r}$ as an estimator of r_0 in the case of experiments with interactive time effects are presented in Section S5 of the supplement, to save space. These results continue to show near perfect performance of $\tilde{r}$ similarly to the experiments summarized above for the case of panels without interactive time effects.

Overall, our findings are in line with the theoretical insights of a very fast convergence of $\tilde{r}$ to r_0. For this reason, we report next on the small sample performance of the identified PME estimator assuming r_0 is known.

---

in levels featuring $r_0 = 1$ and 2 long-run relationships, differing in terms of the error distributions (Gaussian or chi-squared), fit (high or low), and speed of convergence towards long run (moderate or low).

[5] We assume the true lag order of the VAR design is known.

[6] There are a number of ways that one might choose r for the panel from these n tests, based on the modal selection or the average value of the test statistic for instance. We do not explore these avenues here.

TABLE 1: Selection frequencies, averaged across experiments, for the estimation of $r_0 = 0, 1, 2, 3$ by eigenvalue thresholding estimator, $\tilde{r}$, given by (55) with $\delta = 1/4$ and $1/2$ and by Johansen's trace statistics using $VAR(1)$ as the DGP with $r_0 = 0, 1, 2$, and without interactive time effects.

	Frequency $\tilde{r}=0$			Frequency $\tilde{r}=1$			Frequency $\tilde{r}=2$			Frequency $\tilde{r}=3$		
$n \setminus T$	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**
A. Experiments with $r_0 = 0$												
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/4$											
50	0.95	1.00	1.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
500	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/2$											
50	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
500	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Selection of r based on Johansen Trace statistics ($p = 0.05$)											
50	0.28	0.55	0.69	0.41	0.29	0.22	0.14	0.06	0.03	0.17	0.10	0.06
500	0.28	0.54	0.69	0.41	0.29	0.22	0.14	0.06	0.03	0.17	0.10	0.06
1,000	0.28	0.54	0.69	0.41	0.29	0.22	0.14	0.06	0.03	0.17	0.10	0.06
3,000	0.28	0.54	0.69	0.41	0.29	0.22	0.14	0.06	0.03	0.17	0.10	0.06
B. Experiments with $r_0 = 1$												
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/4$											
50	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
500	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/2$											
50	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
500	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Selection of r based on Johansen Trace statistics ($p = 0.05$)											
50	0.46	0.26	0.01	0.36	0.57	0.82	0.07	0.08	0.08	0.11	0.09	0.08
500	0.46	0.26	0.02	0.36	0.57	0.82	0.07	0.08	0.08	0.11	0.09	0.08
1,000	0.46	0.26	0.01	0.36	0.57	0.82	0.07	0.08	0.08	0.11	0.09	0.08
3,000	0.46	0.26	0.01	0.36	0.57	0.82	0.07	0.08	0.08	0.11	0.09	0.08
C. Experiments with $r_0 = 2$												
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/4$											
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/2$											
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
	Selection of r based on Johansen Trace statistics ($p = 0.05$)											
50	0.36	0.06	0.00	0.39	0.61	0.43	0.09	0.15	0.31	0.16	0.18	0.25
500	0.36	0.06	0.00	0.39	0.61	0.43	0.09	0.15	0.31	0.16	0.18	0.25
1,000	0.36	0.06	0.00	0.39	0.61	0.43	0.09	0.15	0.31	0.16	0.18	0.25
3,000	0.36	0.06	0.00	0.39	0.61	0.43	0.09	0.15	0.31	0.16	0.18	0.25

Notes: For $r_0 = 0$, (panel A) there are 3 experiments: with low, medium and high serial correlation in first-difference VAR(1) model. For $r_0 = 1$, (panel B) and $r_0 = 2$ (panel C) there are 8 VAR(1) experiments differing in terms of the error distributions (Gaussian or chi-square), fit (high or low), and speed of convergence towards long run (moderate or low). Lag order for the computation of Johansen's trace statistics is set equal to the true lag order for the VAR. All experiments based on $R = 2,000$ MC replications. Results for individual Monte Carlo experiments are available from the authors upon request. A summary of the different MC designs is given in Subsection 9.1, with a detailed account of the data generating processes provided in Section S4 of the supplement.

TABLE 2: Selection frequencies, averaged across experiments, for the estimation of $r_0 = 0, 1, 2, 3$ by eigenvalue thresholding estimator with $\delta = 1/4$ and $1/2$ and by Johansen's trace statistics using VARMA(1,1) as the DGP with $r_0 = 1, 2$, and without interactive time effects.

	Frequency $\tilde{r} = 0$			Frequency $\tilde{r} = 1$			Frequency $\tilde{r} = 2$			Frequency $\tilde{r} = 3$		
$n \setminus T$	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**
A. Experiments with $r_0 = 1$												
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/4$											
50	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
500	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/2$											
50	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
500	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Selection of r based on Johansen Trace statistics ($p = 0.05$)											
50	0.40	0.33	0.14	0.38	0.48	0.69	0.09	0.09	0.09	0.13	0.11	0.09
500	0.40	0.33	0.14	0.38	0.48	0.69	0.09	0.09	0.09	0.13	0.11	0.09
1,000	0.40	0.33	0.14	0.38	0.48	0.69	0.09	0.09	0.09	0.13	0.11	0.09
3,000	0.40	0.33	0.14	0.38	0.48	0.69	0.09	0.09	0.09	0.13	0.11	0.09
B. Experiments with $r_0 = 2$												
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/4$											
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/2$											
50	0.00	0.00	0.00	0.02	0.00	0.00	0.98	1.00	1.00	0.00	0.00	0.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
	Selection of r based on Johansen Trace statistics ($p = 0.05$)											
50	0.42	0.34	0.10	0.37	0.42	0.47	0.08	0.09	0.21	0.13	0.15	0.21
500	0.42	0.34	0.10	0.37	0.42	0.47	0.08	0.10	0.21	0.13	0.15	0.22
1,000	0.42	0.34	0.10	0.37	0.42	0.47	0.08	0.09	0.21	0.13	0.15	0.22
3,000	0.42	0.34	0.10	0.37	0.42	0.47	0.08	0.09	0.21	0.13	0.15	0.22

Notes: For $r_0 = 1$, (panel A) and $r_0 = 2$ (panel B) there are 8 VARMA(1,1) experiments differing in terms of the error distributions (Gaussian or chi-squared), fit (high or low), and speed of convergence towards long run (moderate or low). Lag order for the computation of Johansen's trace statistics is set equal to the integer part of $T^{-1/3}$. All experiments are based on $R = 2,000$ MC replications. Results for individual Monte Carlo experiments are available from the authors upon request. A summary of the different MC designs is given in Subsection 9.1, with a detailed account of the data generating processes provided in Section S4 of the supplement.

TABLE 3: Simulated bias, RMSE, size and power, averaged across 8 experiments, for PME and MG-Johansen estimators of long-run relations, using a three variable VAR(1) with $r_0 = 2$ and without interactive time effects

	Bias ($\times$100)			**RMSE** ($\times$100)			**Size**($\times$100)			**Power** ($\times$100)		
$n\backslash T$	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**
A. Results for $\beta_{13,0}$												
PME estimator with $q = 2$ sub-samples												
50	-0.15	-0.13	-0.07	4.06	1.96	1.10	7.66	7.10	7.39	15.21	42.18	81.30
500	-0.02	-0.15	-0.06	1.34	0.64	0.35	7.13	5.83	5.17	67.73	99.63	100.00
1,000	0.01	-0.15	-0.07	0.97	0.46	0.25	7.64	6.73	6.48	90.24	100.00	100.00
3,000	-0.05	-0.14	-0.06	0.66	0.29	0.15	12.84	8.16	7.76	99.96	100.00	100.00
PME estimator with $q = 4$ sub-samples												
50	-0.62	-0.39	-0.12	3.29	1.59	0.79	7.48	8.33	7.63	23.13	64.41	96.21
500	-0.49	-0.39	-0.11	1.22	0.62	0.27	10.58	13.04	7.58	91.19	100.00	100.00
1,000	-0.47	-0.38	-0.12	0.95	0.51	0.21	14.99	20.15	11.16	99.31	100.00	100.00
3,000	-0.53	-0.38	-0.12	0.76	0.43	0.15	35.96	47.89	20.91	100.00	100.00	100.00
MG-Johansen estimator												
50	15.78	-3.14	-0.39	>100	>100	65.53	2.66	2.96	3.26	2.67	10.45	56.48
500	-12.34	1.92	-0.86	>100	>100	85.82	1.98	1.89	2.66	2.27	10.77	65.62
1,000	-4.83	-1.07	0.18	>100	>100	36.39	2.16	2.24	2.31	2.38	11.44	65.38
3,000	-46.82	-40.69	-0.17	>100	>100	65.14	2.21	2.01	2.09	2.34	12.24	65.83
B. Results for $\beta_{23,0}$												
PME estimator with $q = 2$ sub-samples												
50	0.02	-0.15	-0.07	4.11	1.96	1.09	8.26	7.43	7.31	15.78	41.90	81.03
500	-0.21	-0.16	-0.07	1.34	0.65	0.35	6.99	5.88	5.63	67.41	99.65	100.00
1,000	-0.07	-0.14	-0.07	0.98	0.46	0.25	7.53	6.32	6.61	90.25	100.00	100.00
3,000	-0.08	-0.14	-0.07	0.66	0.29	0.16	13.13	8.56	7.98	99.97	100.00	100.00
PME estimator with $q = 4$ sub-samples												
50	-0.42	-0.38	-0.12	3.30	1.57	0.77	7.89	8.05	6.72	23.71	64.31	96.66
500	-0.71	-0.40	-0.12	1.31	0.63	0.28	13.23	13.69	8.26	91.11	100.00	100.00
1,000	-0.56	-0.38	-0.12	0.99	0.51	0.21	16.85	19.71	10.71	99.25	100.00	100.00
3,000	-0.57	-0.38	-0.12	0.78	0.43	0.16	37.17	48.16	21.74	100.00	100.00	100.00
MG-Johansen estimator												
50	-21.74	3.58	0.52	>100	>100	65.06	2.56	2.74	3.43	14.82	15.76	56.87
500	-6.62	-1.29	0.65	>100	>100	79.29	2.23	1.94	2.58	15.19	15.74	65.54
1,000	13.31	0.70	-0.18	>100	>100	40.82	2.34	2.14	2.56	15.74	15.99	65.71
3,000	13.99	55.84	0.04	>100	>100	56.04	2.09	2.17	2.23	15.49	17.06	65.86

Notes: The long-run relations are given by $\boldsymbol{\beta}'_{1,0}\mathbf{w}_{it} = w_{it,1} - w_{it,3}$ and $\boldsymbol{\beta}'_{2,0}\mathbf{w}_{it} = w_{it,2} - w_{it,3}$, and identified using $\beta_{11,0} = \beta_{22,0} = 1$, and $\beta_{12,0} = \beta_{21,0} = 0$. The 8 experiments differ with respect to distribution (Gaussian or chi-squared), fit (high or low), and speed of convergence toward long run (moderate or low). Size and power are computed at the five percent nominal level. Reported results are based on $R = 2,000$ Monte Carlo replications. Simulated power are computed under $H_1 : \beta_{13} = -0.97$, $\beta_{23} = -0.97$, as alternatives to -1 for both coefficients under the null. Results for individual Monte Carlo experiments are available from the authors upon request. A summary of the different MC designs is given in Subsection 9.1, with a detailed account of the data generating processes provided in Section S4 of the supplement.

9.3 Estimation of coefficients of long-run relations

9.3.1 Comparison of PME with MG-Johansen in VAR(1) designs with multiple long-run relations

We first investigate how PME estimators of multiple long-run relations compare with the Mean Group estimator based on Johansen's estimator of individual cointegrating vectors (MG-Johansen). To this end, we focus on the VAR(1) design with multiple long-run relations ($r_0 = 2$ long-run relations among $m = 3$ variables) and no interactive time effects. We assume the true lag order is known in the case of MG-Johansen. This is a set up which we believe is most favorable to Johansen procedure when applied to the individual units. But we note that the PME estimator does not require the knowledge of the true lag order. These long-run relations are identified according to (58)-(59) for both PME, and MG-Johansen's estimators. While moments of Johansen's estimator do not exist, we expect the simulated size and power of MG-Johansen to be good for a sufficiently large T relative to n.

Table 3 provides a summary of the results for estimation of $\beta_{13,0}$ $(= -1)$ in part A, and for $\beta_{23,0}$ $(= -1)$ in part B. The table report bias, root mean square error (RMSE), size of the tests at 5% nominal level, and power of the tests ($H_1 : \beta_{13} = -0.97$, $\beta_{23,0} = -0.97$). All entries in this table are multiplied by 100, and averaged across 8 experiments, defined in terms of error distributions (Gaussian or chi-squared), fit (high or low), and speed of convergence towards the long run (moderate or low).

We first note that the PME estimator works quite well, with relatively small bias and RMSE that decline with n and T. The choice of $q = 2$ works well for most n and T combinations, and $q = 4$ performs better in terms of RMSE only when $T = 100$. In terms of size, PME with $q = 2$ does better than with $q = 4$ for all n and T combinations, and has size close to the nominal 5 per cent level, together with satisfactory power that rapidly rises to 100 per cent as n and T are increased. However, even when $q = 2$ there is some evidence of moderate size distortions, particularly when $T = 20$ and $n = 3,000$. The size distortion of PME when $q = 4$ seems to be largely due to its bias that does not decline with T, despite the fall we observe in its RMSE.

The results for the MG-Johansen estimator, also reported in Table 3, show size below 5 percent and a very low power in comparison to PME. The very large RMSE and bias entries could be due to lack of moments for the unit-specific estimators obtained using the Johansen procedure. Clearly, further research is needed for adapting the use of Johansen maximum likelihood approach for use with large panels. We are using MG-Johansen estimators here only for the purpose of comparisons when $r_0 > 1$. Below we do consider other panel cointegration procedures when $r_0 = 1$, and we will no longer report results for MG-Johansen as they are very similar to those summarized in Table 3.

The above summary results do not depend much on which of the two coefficients is considered.

TABLE 4: Simulated bias, RMSE, size and power, averaged across 8 experiments, for PME estimators of long-run relations in the case of a three variables VARMA(1,1) with $r_0 = 2$, and without interactive time effects

$n\backslash T$	**Bias** ($\times$100)			**RMSE** ($\times$100)			**Size**($\times$100)			**Power** ($\times$100)		
	20	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**
A. Results for $\beta_{13,0}$												
PME estimator with $q = 2$ sub-samples												
50	-0.20	-0.27	-0.14	6.92	4.05	2.44	8.49	7.43	7.99	10.44	16.84	33.66
500	0.16	-0.22	-0.13	2.24	1.31	0.77	7.25	5.80	5.81	29.62	72.41	96.88
1,000	0.23	-0.20	-0.13	1.66	0.93	0.55	8.40	5.98	5.79	48.41	92.58	99.92
3,000	0.16	-0.21	-0.11	1.08	0.56	0.32	12.96	6.78	6.73	86.03	99.99	100.00
PME estimator with $q = 4$ sub-samples												
50	-0.83	-0.73	-0.24	5.60	3.30	1.80	7.13	7.39	7.52	12.72	25.58	52.34
500	-0.51	-0.64	-0.20	1.94	1.20	0.60	8.33	9.94	7.17	53.05	92.56	99.86
1,000	-0.46	-0.63	-0.21	1.50	0.96	0.44	10.73	13.52	8.18	76.78	99.49	100.00
3,000	-0.53	-0.63	-0.20	1.11	0.75	0.30	24.57	32.18	14.53	98.21	100.00	100.00
B. Results for $\beta_{23,0}$												
PME estimator with $q = 2$ sub-samples												
50	0.06	-0.23	-0.12	6.84	4.07	2.40	8.50	7.71	7.47	10.87	17.45	33.51
500	-0.06	-0.25	-0.13	2.25	1.31	0.77	6.94	5.81	6.01	29.10	72.41	96.82
1,000	0.11	-0.21	-0.13	1.64	0.93	0.55	8.03	5.81	5.99	48.22	92.34	99.93
3,000	0.12	-0.20	-0.11	1.09	0.56	0.33	12.66	6.92	7.06	86.06	99.98	100.00
PME estimator with $q = 4$ sub-samples												
50	-0.51	-0.65	-0.22	5.48	3.28	1.74	7.36	7.20	6.68	13.36	26.04	51.75
500	-0.77	-0.67	-0.21	2.04	1.21	0.60	9.54	10.01	7.12	52.29	92.70	99.88
1,000	-0.59	-0.63	-0.21	1.53	0.96	0.44	11.66	14.24	8.08	76.76	99.49	100.00
3,000	-0.57	-0.63	-0.20	1.13	0.75	0.31	24.94	32.18	14.91	98.21	100.00	100.00

Notes: The long-run relations are given by $\boldsymbol{\beta}'_{1,0}\mathbf{w}_{it} = w_{it,1} - w_{it,3}$ and $\boldsymbol{\beta}'_{2,0}\mathbf{w}_{it} = w_{it,2} - w_{it,3}$, and identified using $\beta_{11,0} = \beta_{22,0} = 1$, and $\beta_{12,0} = \beta_{21,0} = 0$. The 8 experiments differ with respect to distribution (Gaussian or chi-squared), fit (high or low), and speed of convergence toward long run (moderate or low). Reported results are based on $R = 2,000$ Monte Carlo replications. Simulated power are computed under $H_1 : \beta_{13} = -0.97$ and $\beta_{23} = -0.97$, as alternatives to -1 for both coefficients under the null. Results for individual Monte Carlo experiments are available from the authors upon request. A summary of the different MC designs is given in Subsection 9.1, with a detailed account of the data generating processes provided in Section S4 of the supplement. Size and Power are computed at 5 percent nominal level.

TABLE 5: Simulated bias, RMSE, size and power, averaged across 8 experiments, for PME estimators of long-run relations in the case of three variables VARMA(1,1) with $r_0 = 2$, and with interactive time effects

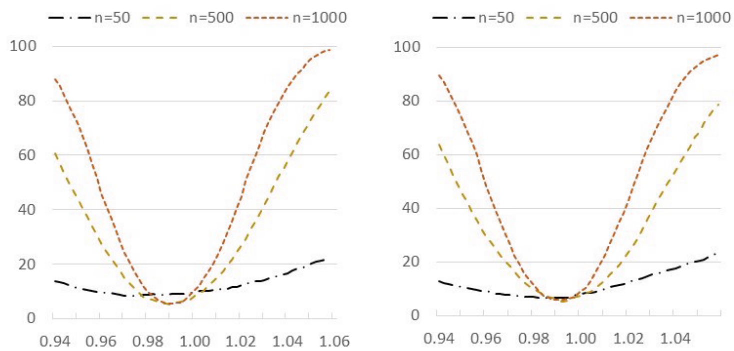
$n\backslash T$	**Bias** ($\times$100)			**RMSE** ($\times$100)			**Size**($\times$100)			**Power** ($\times$100)		
	20	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**
A. Results for $\beta_{13,0}$												
PME estimator with $q = 2$ sub-samples												
50	-0.20	-0.26	-0.14	6.89	4.05	2.44	8.56	7.46	8.00	10.59	16.85	33.76
500	0.16	-0.21	-0.13	2.23	1.31	0.77	7.15	5.83	5.79	29.86	72.38	96.91
1,000	0.24	-0.20	-0.13	1.65	0.93	0.55	8.38	5.94	5.76	48.74	92.63	99.92
3,000	0.16	-0.21	-0.11	1.08	0.56	0.32	12.91	6.78	6.71	86.23	99.99	100.00
PME estimator with $q = 4$ sub-samples												
50	-0.81	-0.72	-0.23	5.54	3.29	1.80	7.04	7.41	7.53	12.85	25.58	52.38
500	-0.50	-0.63	-0.20	1.92	1.20	0.60	8.21	9.96	7.13	53.69	92.61	99.85
1,000	-0.44	-0.63	-0.20	1.47	0.95	0.44	10.59	13.56	8.22	77.28	99.50	100.00
3,000	-0.52	-0.63	-0.20	1.09	0.75	0.30	24.09	32.06	14.57	98.36	100.00	100.00
B. Results for $\beta_{23,0}$												
PME estimator with $q = 2$ sub-samples												
50	0.05	-0.22	-0.12	6.81	4.06	2.40	8.41	7.76	7.54	11.00	17.61	33.53
500	-0.06	-0.24	-0.13	2.23	1.31	0.77	6.90	5.79	6.00	29.31	72.36	96.86
1,000	0.11	-0.21	-0.13	1.64	0.93	0.55	7.99	5.81	6.06	48.69	92.39	99.93
3,000	0.12	-0.20	-0.11	1.08	0.56	0.33	12.46	6.91	7.06	86.13	99.98	100.00
PME estimator with $q = 4$ sub-samples												
50	-0.49	-0.64	-0.22	5.43	3.27	1.74	7.45	7.23	6.63	13.33	25.83	51.71
500	-0.75	-0.67	-0.21	2.01	1.21	0.60	9.43	9.98	7.16	52.87	92.71	99.88
1,000	-0.57	-0.63	-0.21	1.51	0.96	0.44	11.54	14.21	8.06	77.26	99.50	100.00
3,000	-0.55	-0.62	-0.20	1.11	0.75	0.31	24.68	32.24	14.89	98.32	100.00	100.00

Notes: See the notes to Table 4.

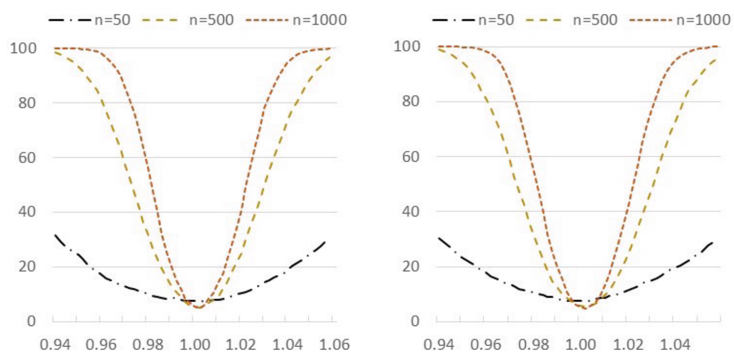
Long run parameters

-$\beta_{13,0}$ -$\beta_{23,0}$

A. $T = 20$



B. $T = 50$



C. $T = 100$

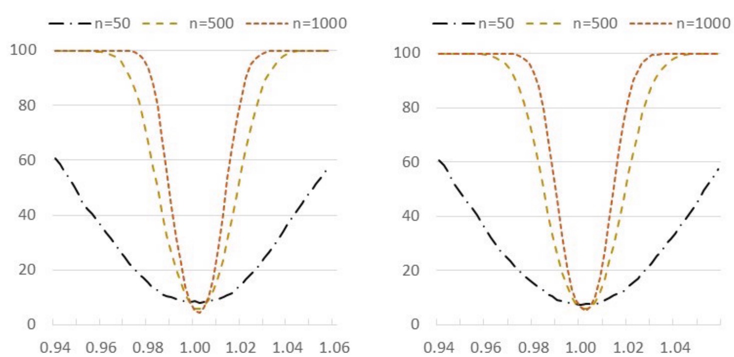
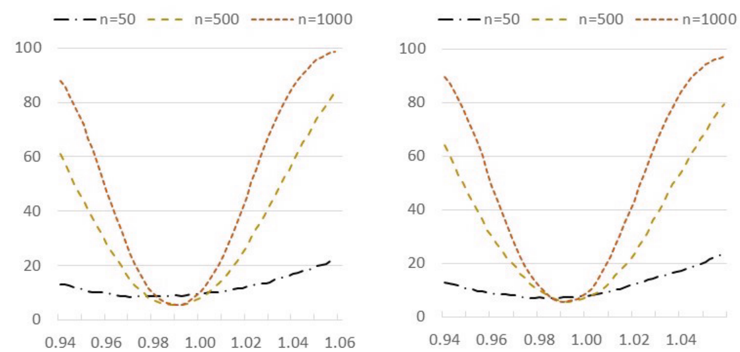


FIGURE 1: Empirical power curves for the tests based on PME estimators of $\beta_{13,0}$ and $\beta_{23,0}$ parameters of $r_0 = 2$ long-run relations, using VARMA(1,1) without interactive time effects, slow speed of convergence, $PR^2_{nT} = 0.2$, and Gaussian errors. PME estimators use $q = 2$ sub-samples.

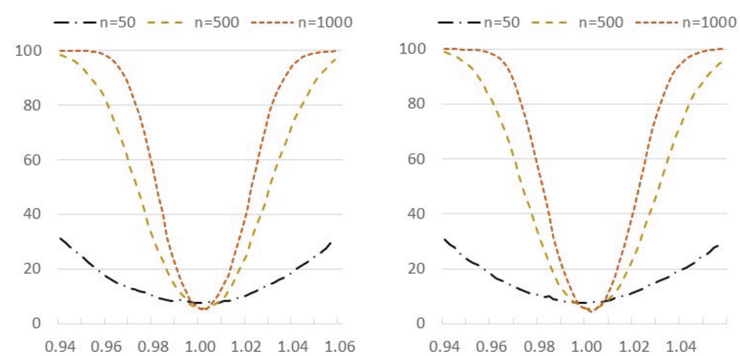
Long run parameters

$-\beta_{13,0}$	$-\beta_{23,0}$

A. $T = 20$



B. $T = 50$



C. $T = 100$

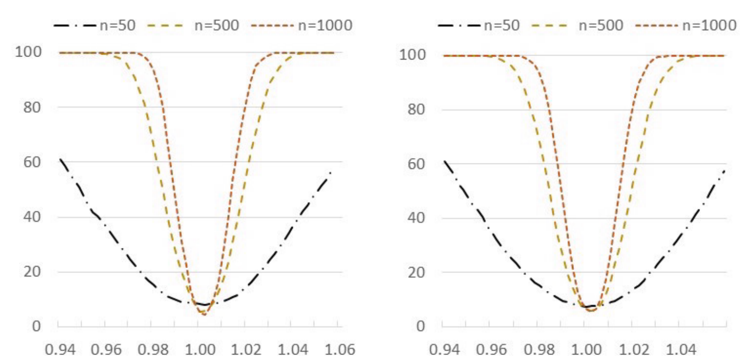


FIGURE 2: Empirical power curves for the tests based on PME estimators of $\beta_{13,0}$ and $\beta_{23,0}$ parameters of $r_0 = 2$ long-run relations, using VARMA(1,1) with interactive time effects, slow speed of convergence, $PR^2_{nT} = 0.2$, and Gaussian errors. PME estimators use $q = 2$ sub-samples.

TABLE 6: Simulated bias, RMSE, size and power for estimation of long-run relation using VAR(1) with $m = 2$ variables, $r_0 = 1$, and one-way long-run causality

	Bias ($\times$100)			**RMSE** ($\times$100)			**Size** ($\times$100)			**Power** ($\times$100)		
$n\backslash T$	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**
PME estimator with $q = 2$ sub-samples												
50	-0.98	-0.19	-0.03	4.92	2.11	1.05	7.95	6.90	6.30	15.90	37.65	82.40
500	-0.98	-0.16	-0.06	1.77	0.70	0.34	10.35	6.75	5.70	78.30	99.85	100.00
1,000	-0.93	-0.17	-0.06	1.42	0.51	0.24	15.90	7.40	4.70	96.30	100.00	100.00
3,000	-0.91	-0.18	-0.05	1.09	0.32	0.14	33.00	10.50	5.60	100.00	100.00	100.00
PME estimator with $q = 4$ sub-samples												
50	-1.46	-0.36	-0.09	4.09	1.79	0.85	9.05	7.35	6.30	23.30	52.00	95.40
500	-1.47	-0.33	-0.10	1.89	0.64	0.28	23.65	9.65	5.45	97.50	100.00	100.00
1,000	-1.41	-0.35	-0.10	1.65	0.53	0.21	39.95	15.50	6.20	100.00	100.00	100.00
3,000	-1.40	-0.35	-0.09	1.48	0.42	0.14	82.65	36.05	12.65	100.00	100.00	100.00
System Pooled Mean Group Estimator												
50	-0.33	0.05	-0.01	7.10	1.87	0.81	58.50	24.95	15.00	62.80	67.95	99.05
500	-0.05	0.03	0.00	2.21	0.58	0.25	59.45	25.10	13.20	83.15	100.00	100.00
1,000	-0.01	0.01	0.00	1.59	0.41	0.18	59.25	22.90	13.05	92.40	100.00	100.00
3,000	-0.01	0.00	0.00	0.89	0.24	0.10	57.45	24.30	13.65	99.75	100.00	100.00
Breitung's 2-Step Estimator												
50	5.25	1.28	0.36	6.45	2.03	0.84	53.95	23.30	12.30	28.55	33.05	96.15
500	5.25	1.26	0.36	5.39	1.35	0.43	99.70	84.00	41.90	70.55	97.95	100.00
1,000	5.27	1.25	0.35	5.34	1.29	0.39	99.90	97.75	63.75	91.00	100.00	100.00
3,000	5.29	1.24	0.36	5.32	1.26	0.37	100.00	100.00	98.20	99.90	100.00	100.00
Pooled Mean Group Estimator												
50	1.69	0.40	0.08	5.63	1.76	0.78	44.90	19.90	10.85	43.55	59.50	98.70
500	1.91	0.36	0.08	2.52	0.64	0.25	66.25	28.45	12.80	52.45	100.00	100.00
1,000	1.92	0.33	0.08	2.25	0.50	0.19	79.65	33.55	14.35	61.00	100.00	100.00
3,000	1.90	0.32	0.08	2.02	0.39	0.13	98.30	56.95	22.00	80.95	100.00	100.00
Pooled Bewley Estimator												
50	3.70	0.75	0.19	5.17	1.70	0.76	20.05	10.20	7.20	7.15	35.65	96.80
500	3.76	0.73	0.18	3.92	0.86	0.29	92.40	34.15	12.65	11.60	99.95	100.00
1,000	3.78	0.71	0.18	3.87	0.79	0.24	99.80	58.55	19.80	18.45	100.00	100.00
3,000	3.80	0.71	0.18	3.82	0.73	0.21	100.00	95.75	50.20	41.95	100.00	100.00
Panel FMOLS estimator												
50	10.27	4.27	2.01	11.08	4.69	2.24	96.10	90.85	85.70	85.30	47.05	52.85
500	10.30	4.27	2.03	10.38	4.32	2.05	100.00	100.00	100.00	100.00	88.95	98.10
1,000	10.33	4.25	2.03	10.37	4.28	2.04	100.00	100.00	100.00	100.00	97.90	99.95
3,000	10.36	4.25	2.03	10.37	4.26	2.04	100.00	100.00	100.00	100.00	100.00	100.00
Panel Dynamic OLS Estimator												
50	4.15	1.15	0.36	6.51	2.15	0.91	23.45	15.10	9.40	12.85	24.85	90.25
500	4.13	1.12	0.35	4.44	1.25	0.43	72.85	58.95	30.50	16.05	94.50	100.00
1,000	4.19	1.11	0.35	4.35	1.18	0.39	87.45	84.20	50.80	21.65	99.80	100.00
3,000	4.21	1.10	0.35	4.26	1.13	0.37	91.25	99.45	93.10	39.50	99.80	100.00

Notes: The long-run relation is given by $\boldsymbol{\beta}'_{1,0}\mathbf{w}_{it} = w_{it,1} - w_{it,2} = \beta_{11,0}w_{it,1} + \beta_{12,0}w_{it,2}$, and identified with $\beta_{11,0} = 1$. Coefficient $\beta_{12,0} = -1$ is estimated. Data generating process used for results reported in this table is taken from Chudik, Pesaran, and Smith (2023a). Section 3.1 of Chudik, Pesaran, and Smith (2023a) provides full account of this design. Reported results are based on $R = 2,000$ Monte Carlo replications. Size and Power are computed at 5 percent nominal level. Simulated powers are computed under $H_1 : \beta_{12} = -0.97$, compared to null value of -1.

9.3.2 Performance of PME in VARMA(1,1) designs

Results when using VARMA(1,1) designs with $r_0 = 2$ are presented in Tables 4 for models without time effects, and in Table 5 for models with interactive time effects. As before, these results are averages across eight VARMA(1,1) experiments featuring $r_0 = 2$ long-run relations that differ in terms of error distributions (Gaussian or chi-squared), fit (high or low), and speed of convergence toward long run (moderate or low). Figures 1-2 give empirical power curves for tests on $\beta_{13,0}$ and $\beta_{23,0}$ using PME estimators computed with $q = 2$ sub-samples, in VARMA(1,1) designs with $r_0 = 2$ long-run relations, and for $n = 50$, 500, and 1000. Separate panels show $T = 20$, $T = 50$, $T = 100$. This experiment has a slow speed of convergence, $PR_{nT}^2 = 0.2$, Gaussian errors and without interactive time effects (Figure 1) or with interactive time effects (Figure 2). In both cases, the power increases with n and T as expected.

Qualitatively, the VARMA results in Table 4 are similar to the VAR results summarized in Table 3, with one important exception. Allowing for an MA component in the DGP tends to reduce power of the tests. For example, in the case where $n = 500$, $T = 20$, and $q = 2$, the power of testing $\beta_{13,0} = -1$ against the alternative $\beta_{13} = -0.97$ is 67.73 per cent when using VAR design (Table 3) as compared to 29.62 per cent when using the VARMA design (Table 4). Otherwise, the results for bias, RMSE and size are comparable across the two designs. Similarly, allowing for interactive time effects (Table 5) does not alter these conclusions, and PME seems to be quite robust to the inclusion of interactive time effects, in line with the theoretical results of Section 7, so long as the time effects are not trended (deterministic or stochastic).

9.3.3 Comparisons with available estimators in the design with single long-run relation and one-way long-run causality

Last but not least, we investigate how PME compares with existing approaches for the estimation of a single long-run relation, in a design that is favorable to single equation estimators, some of which rely on the direction of long-run causality to be one-way and known, and the lag order of the VAR to be well specified. We report results for panel FMOLS estimator by Pedroni (1996, 2001a, 2001b), the Pooled Mean Group (PMG) estimator by Pesaran, Shin, and Smith (1999), panel Dynamic OLS (PDOLS) by Mark and Sul (2003), the two-step system estimator of Breitung (2005), the system PMG estimator of Chudik, Pesaran, and Smith (2023b), and the pooled Bewley estimator by Chudik, Pesaran, and Smith (2023a). For large T panels with moderate n, we would expect these single equation techniques to perform better than the PME that allows for MA components and does not assume long-run causality. This is confirmed by the results summarized in Table 6. When $T = 100$ and $n = 50$, PME ($q = 2$) has higher RMSE than all other estimators reported in Table 6 except for the FMOLS estimator. In contrast, PME estimator ($q = 2$) is much more balanced in terms of bias, RMSE and size of the tests, when T is small ($= 20$), and n quite large ($= 1000$). None of the alternative estimators to PME in the case of $r_0 = 1$ (and known one-way long-run causality) work well in samples where T is not very large and n much larger than T.

In terms of bias and size, the PME estimator with $q = 2$ performs much better than all the other single equation estimators under consideration, even if we consider $T = 100$ and $n = 50$. Overall, Monte Carlo findings show that the PME estimator with $q = 2$ can have satisfactory performance for panels where n is quite large relative to T, in particular for panels with n and T combinations similar to the ones we consider in our empirical application discussed below.

10 Empirical Applications

We provide two empirical applications to illustrate wide applicability of the PME approach to both micro and macro panels.

10.1 Estimation of long-run financial relations

The fist application considers micro panels of firms where n is quite large relative to the available time dimension. We use logarithms of six key financial variables from CRSP/Compustat, available from Wharton Research Data Services. The six variables (measured in logarithms) are book value (BV_{it}), market value (MV_{it}), short-term debt (SD_{it}), long-term debt (LD_{it}), total assets (TA_{it}) and total debt outstanding (DO_{it}). These are variables among which one would expect some key relations. We consider them in three sets of two or three variables, using two unbalanced panels both begin in 1950, the shorter one ends in 2010 and the longer one in 2021. The maximum T is 71. We set a minimum T of 20, and the average T is around 30.

The number of firms n varies from about $1,000$ to $2,500$ depending on the set of variables under consideration. As well as estimating the number of long-run relations and their parameters we test whether the coefficients in the linear combinations of logarithms take the value -1, to relate our results to the ratios used in corporate finance. In corporate finance accounting ratios, constructed from balance sheet data, are commonly used to measure the profitability, liquidity, and solvency of a firm. The rationales for the use of ratios include correcting for size in the cross section dimension and eliminating common trends in the time series dimension to render the ratios stationary. These two objectives are not always compatible. In an early contribution, that remains relevant, Lev and Sunder (1979) comment "It appears that the extensive use of financial ratios by both practitioners and researchers is often motivated by tradition and convenience rather than resulting from theoretical considerations or from a careful statistical analysis." Geelen et al. (2024) provide a more recent study of the use of financial ratios in empirical corporate finance literature.

Given that the theory is often not very specific, it is desirable to have a statistical criteria to judge which are the appropriate long-run relations among the set of variables considered and whether the logarithm of their ratios is stationary. The time series stationarity of finance ratios like the aggregate dividend price ratio have been studied, but the question has not, to our knowledge, been addressed in corporate finance, where the context is somewhat different. Corporate finance studies tend to use unbalanced panels with large n and relatively small T and and the vector of accounting variables, of the sort one gets from Compustat, may include multiple long-run relations. Thus the PME estimator which is appropriate for multiple long-run relations in a large n, moderate T panel seems well designed to determine whether there are long-run relations in accounting data.

Using a multiplicative specification, the relation between y_{it}, x_{it} and z_{it} can be readily cast in terms of $\mathbf{w}_{it} = (\ln y_{it}, \ln x_{it}, \ln z_{it})'$ and the PME procedure can be used to test (a) if $\boldsymbol{\beta}_0' \mathbf{w}_{it}$ is stationary; and (b) if $\beta_{11,0} = 1$, $\beta_{12,0} = -1$, $\beta_{13,0} = 0$; to validate the the use of $\ln(y_{it}/x_{it})$ or y_{it}/x_{it} in econometric analysis. If step (a) cannot be validated then $\boldsymbol{\beta}_0' \mathbf{w}_{it}$ will not be stationary and its use in econometric analysis could lead to spurious results. But if step (a) is validated but not (b), whilst it would not be advisable to use $\ln y_{it} - \ln x_{it}$, one can still consider using $\tilde{\boldsymbol{\beta}}' \mathbf{w}_{it}$, where $\tilde{\boldsymbol{\beta}}$ could be the exactly identified PME estimator of $\boldsymbol{\beta}_0$, in second stage panel regressions that allow for short term dynamics as well as other stationary variables.[7] Such a two-step procedure is justified, noting that $\tilde{\boldsymbol{\beta}}$ is super consistent, converging to $\boldsymbol{\beta}_0$ at the rate of $T\sqrt{n}$.

Data sources, full definitions of the variables, summary statistics and additional details on construction of the samples for each of the three variable sets are provided in Section S7 of the supplement. The following data filters are applied sequentially to each variable set and sample period, separately. Firms are omitted if, for a given variable set and sample: they do not have data for all variables in the set (without gaps and covering at least 20 time periods); have nonpositive entries on any of the variables, since we are using logarithms; and average value of key ratios fall below the 1st or above the 99th percentiles estimated after the application of the first two filters. This is similar to the filtering Geelen et al. (2024) use, except that we require a longer time series dimension.[8]

The variables are grouped into three sets, where we have prior expectations about possible cointegration amongst them. To illustrate the procedure we start with the simplest case where $m = 2$ and $\mathbf{w}_{it}$ include the logarithm of total debt outstanding and logarithm of total assets: $\{DO_{it}, TA_{it}\}$. The ratio of total debt to total assets is often used as a measure of leverage, so there is a single hypothesized long-run relation. To showcase the performance of the PME procedure with multiple long-run relations, we consider two other sets of variables with $m = 3$. They are: the logarithms of short and long term debt and total assets, $\{SD_{it}, LD_{it}, TA_{it}\}$; and the logarithms of total debt outstanding, book value and market value, $\{DO_{it}, BV_{it}, MV_{it}\}$. We expect two hypothesized long-run relations in both of these sets. Since these variables are often used to construct accounting ratios, whether the long-run relation has a unit coefficient is also of interest. Because of missing firm observations on some variables, the number of firms with minimum of 20 data points on all the variables under consideration falls as we include more variables in a set. For this reason we do not consider all six variables together.

Table 7 gives the estimates of the number of long-run relations for two and three

[7]To simplify the notations, we use the symbol tilde in this section to denote PME estimates of the exactly identified long-run relations.

[8]Geelen et al. (2024) state: "We winsorize all variables at the 1% and 99% levels to mitigate the impact of outliers. We drop all observations with missing values on one or more variables of interest. We remove observations with a market-to-book ratio larger than 20, negative book equity or negative EBITDA. Our final sample consists of 68,833 firm-year observations with 6,001 unique firms." This gives $\bar{T} = 11.5$, though some regressions use less.

variable models for the 1950-2021 and 1950-2010 unbalanced samples using the eigenvalue thresholding procedure, given by (55). It also reports the eigenvalues of the correlation matrix, $\mathbf{R}_{ww}$ defined by (54), together with the associated threshold values, $T_{ave}^{-\delta}$, where $T_{ave} = n^{-1}\sum_{i=1}^{n} T_i$. Since the panel is unbalanced we base the thresholds on T_{ave} and provide $\tilde{r}$ for $\delta = 1/2$ and $\delta = 1/4$, using $q = 2$ sub-sample time averages.

The preferred threshold based on $\delta = 1/4$ gives $\tilde{r} = 1$ long-run relation for the panel data models with $m = 2$, and $\tilde{r} = 2$ long-run relations for the two cases with $m = 3$. The threshold with $\delta = 1/2$ also yields $\tilde{r} = 1$ in the case with $m = 2$, but when used in the case of panels with $m = 3$ it selects $\tilde{r} = 1$ rather than $\tilde{r} = 2$. Given the theoretical discussion above and the Monte Carlo results, we proceed using the estimates of the number of long-run relations obtained using the preferred value of $\delta = 1/4$, namely *one* long-run relation when $m = 2$ and *two* long-run relations when $m = 3$.[9]

Tables 8-10 present PME estimates of the coefficients in the long-run relations, their standard errors and t-statistics for testing the null hypothesis that the long-run coefficient in question is equal to -1. The first set of estimates, in Table 8, are for panels with $\mathbf{w}_{it} = (DO_{it}, TA_{it})'$. Recall that DO_{it} and TA_{it} are *logarithms* of debt outstanding and total assets. There is a single hypothesized long-run relation: $\beta_{11,0}DO_{it} + \beta_{12,0}TA_{it}$. Panel A of Table 8 uses the exact identifying condition $\beta_{11,0} = 1$ and provides PME estimates of $\beta_{12,0}$ and t-statistics for the null value of $\beta_{12,0} = -1$. The PME estimates of $\beta_{12,0}$ at -1.142 and -1.113 for the two sample periods ending in 2021 and 2010, respectively, are similar and close to but significantly different from -1. To illustrate that, unlike regression based methods, the PME estimator is invariant to normalization, part B of Table 8 uses the exact identifying condition $\beta_{12,0} = 1$, and reports the PME estimate of $\beta_{11,0}$. It is confirmed that up to rounding error $\tilde{\beta}_{11} = 1/\tilde{\beta}_{12}$.

[9] Table S10 in the supplement reports IPS panel unit root tests by Im et al. (2003) (using a 40-year balanced panel), which do not reject the null of unit root in all cases except for short-term debt (SD). Given the 5 per cent chance that the IPS test could be in error, we proceed assuming that all the six variables are $I(1)$.

TABLE 7: Estimates of the number of long-run relations ($\tilde{r}$) by eigenvalue thresholding using firm-level data and $q = 2$ sub-sample time averages

Variables:	$\mathbf{w}_{it} = (DO_{it}, TA_{it})'$		$\mathbf{w}_{it} = (SD_{it}, LD_{it}, TA_{it})'$		$\mathbf{w}_{it} = (DO_{it}, BV_{it}, MV_{it})'$	
Sample end year:	1950-2021	1950-2010	1950-2021	1950-2010	1950-2021	1950-2010
Estimated number of long-run relations ($\tilde{r}$)						
$\tilde{r}$ ($\delta = 1/2$)	1	1	1	1	1	1
$\tilde{r}$ ($\delta = 1/4$)	1	1	2	2	2	2
Eigenvalues						
$\tilde{\lambda}_1$	0.090	0.079	0.106	0.100	0.058	0.046
$\tilde{\lambda}_2$	1.910	1.921	0.251	0.223	0.226	0.198
$\tilde{\lambda}_3$	-	-	2.644	2.678	2.717	2.756
Threshold $T_{ave}^{-\delta}$						
$\delta = 1/2$	0.176	0.178	0.177	0.179	0.178	0.182
$\delta = 1/4$	0.419	0.422	0.421	0.423	0.422	0.426
Sample dimensions						
n	2,555	1,901	1,373	1,101	1,415	1,164
T_{ave}	32.4	31.6	31.8	31.1	31.5	30.3
$\max_i T_i$	73	61	73	61	61	49
$\sum_{i=1}^{n} T_i$	82,837	60,118	43,621	34,262	44,592	35,293

Notes: This table reports eigenvalues $\tilde{\lambda}_j$ for $j = 1, 2..., m$ (in ascending order) of $\mathbf{R}_{\bar{w}\bar{w}}$ given by (54) using $q = 2$ sub-sample time averages and the corresponding eigenvalue thresholding estimates of the number of long-run relations given by $\tilde{r} = \sum_{j=1}^{m} \mathcal{I}\left(\tilde{\lambda}_j < T_{ave}^{-\delta}\right)$, for $\delta = 1/4, 1/2$, where $T_{ave} = n^{-1} \sum_{i=1}^{n} T_i$, and $T_i \geq 20$ for all i. The elements of $\mathbf{w}_{it}$ are logarithms of book value (BV_{it}), market value (MV_{it}), short-term debt (SD_{it}), long-term debt (LD_{it}), total assets (TA_{it}) and total debt outstanding (DO_{it}). See Section S7 of the supplement for variable definitions, data sources and availability, and filters applied.

For panels with $m = 3$ and $\tilde{r} = 2$ we need two exactly identifying conditions on each of the two long-run relations. In the case where $\mathbf{w}_{it} = (SD_{it}, LD_{it}, TA_{it})'$, we use the conditions $\beta_{11,0} = 1$ and $\beta_{12,0} = 0$ to exactly identify the first long-run relation, and conditions $\beta_{21,0} = 0$ and $\beta_{22,0} = 1$ to identify the second long-run relation. Hence the two identified long-run relations are $SD_{it} + \beta_{13,0} TA_{it}$ and $LD_{it} + \beta_{23,0} TA_{it}$. PME estimates for $\beta_{13,0}$ in Table 9 are -1.025 and -1.024 for the two unbalanced samples, both very close to -1 and statistically not different from -1. PME estimates for $\beta_{23,0}$ are -0.875 and -0.899, both statistically different from -1 at the one percent level. Rotation of the two identified long-run relations reported in Table 9 imply the long-run relations $SD_{it} - 1.156 LD_{it}$ and $SD_{it} - 1.130 LD_{it}$ for the samples ending in 2021 and 2010, respectively. Both of these PME estimates, -1.156 and -1.130, are statistically significantly different from -1 at the one per cent level. Hence, in the case of the variable set $\{SD_{it}, LD_{it}, TA_{it}\}$, the unit long-run elasticity hypothesis could be accepted for short term debt to total assets only..

TABLE 8: PME estimates for the set $\{DO_{it}, TA_{it}\}$, using $q = 2$ sub-sample time averages and one long-run relation

A. Exact identifying condition $\beta_{11,0} = 1$

	Exactly identified long-run relation $\boldsymbol{\beta}'_{1,0}\mathbf{w}_{it} = DO_{it} + \beta_{12,0}TA_{it}$	
Sample ends:	1950-2021	1950-2010
$\tilde{\beta}_{12}$	-1.142	-1.113
s.e.	(0.010)	(0.011)
t($\beta_{12,0} = -1$)	-14.6	-10.4
n	2,555	1,901
$\sum_{i=1}^{n} T_i$	82,837	60,118

B. Exact identifying condition $\beta_{12,0} = 1$

	Exactly identified long-run relation $\boldsymbol{\beta}'_{1,0}\mathbf{w}_{it} = \beta_{11,0}DO_{it} + TA_{it}$	
Sample ends:	1950-2021	1950-2010
$\tilde{\beta}_{11}$	-0.875	-0.899
s.e.	(0.010)	(0.012)
t($\beta_{11,0} = -1$)	-12.8	-8.7
n	2,555	1,901
$\sum_{i=1}^{n} T_i$	82,837	60,118

Notes: TA_{it} and DO_{it} are logarithms of total assets and total debt outstanding. Left panel reports PME estimate of $\beta_{12,0}$ using the exact identifying condition $\beta_{11,0} = 1$. Right panel reports PME estimate of $\beta_{11,0}$ using the exact identifying condition $\beta_{12,0} = 1$. PME estimator, given by (23), is computed using $q = 2$ sub-sample time averages, subject to $T_i \geq 20$, for all i. To simplify the notations we use the tilde symbols to denote the exactly identified PME estimates. The row labelled t($\beta_{12,0} = -1$) and t($\beta_{11,0} = -1$) gives the t statistic for testing $H_0 : \beta_{12,0} = -1$ and $H_0 : \beta_{11,0} = -1$, respectively. See Section S7 of the supplement for variable definitions, data sources and availability, and filters applied.

TABLE 9: PME estimates for the variable set $\{SD_{it}, LD_{it}, TA_{it}\}$, using $q = 2$ sub-sample time averages and two long-run relations

	First exactly identified long-run relation $\boldsymbol{\beta}'_{1,0}\mathbf{w}_{it} = SD_{it} + \beta_{13,0}TA_{it}$			Second exactly identified long-run relation $\boldsymbol{\beta}'_{2,0}\mathbf{w}_{it} = LD_{it} + \beta_{23,0}TA_{it}$	
Sample end year:	1950-2021	1950-2010	Sample end year:	1950-2021	1950-2010
$\tilde{\beta}_{13}$	-1.025	-1.024	$\tilde{\beta}_{23}$	-1.185	-1.158
s.e.	(0.019)	(0.019)	s.e.	(0.015)	(0.016)
t($\beta_{13,0} = -1$)	-1.4	-1.3	t($\beta_{23,0} = -1$)	-12.0	-9.7
n	1,373	1,101	n	1,373	1,101
$\sum_{i=1}^{n} T_i$	43,621	34,262	$\sum_{i=1}^{n} T_i$	43,621	34,262

Notes: Long run relations are $\boldsymbol{\beta}'_{j,0}\mathbf{w}_{it} = \beta_{j1,0}SD_{it} + \beta_{j2,0}LD_{it} + \beta_{j3,0}TA_{it}$, for $j = 1, 2$, where SD_{it}, LD_{it} and TA_{it} are logarithms of short-term and long-term debts, and total assets. The first long-run relation ($j = 1$) is identified using the exact identifying conditions $\beta_{11,0} = 1$ and $\beta_{12,0} = 0$. The second long-run relation ($j = 2$) is identified using the exact identifying conditions $\beta_{22,0} = 1$ and $\beta_{21,0} = 0$. This table reports estimates for $\beta_{13,0}$ and $\beta_{23,0}$ using PME estimator given by (23) with $q = 2$ sub-sample time averages, subject to $T_i \geq 20$, for all i. To simplify the notations we use tilde symbol to denote the exactly identified PME estimates. See Section S7 of the supplement for variable definitions, data sources and availability, and filters applied.

TABLE 10: PME estimation results for the variable set $\{DO_{it}, BV_{it}, MV_{it}\}$, using $q = 2$ sub-sample time averages and two long-run relations

	First exactly identified long-run relation $\boldsymbol{\beta}'_{1,0}\mathbf{w}_{it} = DO + \beta_{13,0}BV$			Second exactly identified long-run relation $\boldsymbol{\beta}'_{2,0}\mathbf{w}_{it} = BV + \beta_{23,0}MV$	
Sample end year:	2021	2010	Sample end year:	2021	2010
$\tilde{\beta}_{13}$	-1.055	-0.990	$\tilde{\beta}_{23}$	-0.937	-0.927
s.e.	(0.019)	(0.020)	s.e.	(0.008)	(0.009)
t($\beta_{13,0} = -1$)	2.8	-0.5	t($\beta_{23,0} = -1$)	-7.7	-8.6
n	1,415	1,164	n	1,415	1,164
$\sum_{i=1}^{n} T_i$	44,592	35,293	$\sum_{i=1}^{n} T_i$	44,592	35,293

Notes: Long run relations are $\boldsymbol{\beta}'_{j,0}\mathbf{w}_{it} = \beta_{j1,0}DO_{it} + \beta_{j2,0}BV_{it} + \beta_{j3,0}MV_{it}$, for $j = 1, 2$, where DO_{it}, BV_{it} and MV_{it} are logarithms of total debt outstanding, book, and market values. The first long-run relation ($j = 1$) is identified using the exact identifying conditions $\beta_{11,0} = 1$ and $\beta_{12,0} = 0$. The second long-run relation ($j = 2$) is identified using the exact identifying conditions $\beta_{22,0} = 1$ and $\beta_{21,0} = 0$. See also the notes to Table 9.

Table 10 reports similar results for $\mathbf{w}_{it} = (DO_{it},\ BV_{it},\ MV_{it})'$ and two exactly identified long-run relations associated with logarithms total debt, market value, and book value. PME estimates are close to -1, but the unit long-run elasticity between pairs of variables in this set is rejected for all cases except the logarithms of total debt outstanding and market value in the sample ending in 2010.[10]

Overall, the results support the existence of long-run relations between the logarithms of a number of key financial variables considered in the corporate finance literature. But with the notable exception of the logarithm of short term debt to total asset ratio, the use of other financial ratios as stationary variables in financial analysis is not supported by our empirical findings. Instead our study recommends using estimated long-run relations such as $DO_{i,t-1} - 1.14\, TA_{i,t-1}$, $LD_{i,t-1} - 1.19\, TA_{i,t-1}$, and $BV_{i,t-1} - 0.927 MV_{i,t-1}$, as error correction terms in dynamic panel regressors that allow for short term dynamics and other stationary variables. This allows the analysis of short term dynamics of financial variables to be coherently embedded in a long-run equilibrating framework.

10.2 International macro applications using Penn World Table

Our second empirical application considers cross country macroeconomic time series data from the Penn World Table[11] (PWT), where n (the number of countries) is smaller and the average time dimension is larger as compared to the corporate finance data. Given the good small sample performance of the PME approach even for smaller values of n in our Monte Carlo experiments, we have confidence in using the PME estimation also in the case of the macro application. We focus on four key macro variables, namely real merchandise exports per capita (ex_{it}), real merchandise imports per capita (im_{it}), real productivity per hour worked ($prod_{it}$) and real wages per hour worked ($wage_{it}$). The choice of these variables was motivated by two widely maintained hypotheses. Firstly, real wages and productivity should balance

[10] Rotation of the two identified long-run relationships reported in Table 10 imply the long-run relations $DO_{it} - 0.888 BV_{it}$ and $DO_{it} - 0.937 BV_{it}$ for the samples ending in 2021 and 2010, respectively, both of these PME estimates are statistically significanlty different from -1 at the 1 percent level.

[11] We use version 10.01 of PWT database, available at https://www.rug.nl/ggdc/productivity/pwt/, see Feenstra et al. (2015). See also the supplement for a detailed description of data constructions.

for steady state growth to be feasible. Secondly export and imports should balance for international solvency, though the constraint may not be binding for reserve-currency countries such as the US. We first consider the two pairs that we expect to cointegrate separately, then we consider all four variables together. As with the corporate finance application, we are interested both in whether they cointegrate and whether there is a unit coefficient. In both cases, labour market balance and trade balance, there is no clear causal ordering between the variables. Since the constraints that produce balance may be somewhat different for advanced and emerging economies, we report estimates for the country groupings separately as well as together.

The PME results for the analysis of a possible long-run relation between ex_{it} and im_{it} are summarized in Table 11. This table gives the two eigenvalues of $\mathbf{R}_{\bar{w}\bar{w}}$ for $\mathbf{w}_{it} = (ex_{it}, im_{it})'$ and $q = 2$. (See (54)) As can be seen there is a clear separation between the first and second eigenvalues supporting the existence of a long-run relation between ex_{it} and im_{it}. This result holds for both choices of the threshold parameter $\delta = 1/2$ and $1/4$, and for all three country groupings. Given this result we then estimated the long-run relation $\boldsymbol{\beta}'\mathbf{w}_{it} = \beta_{11}ex_{it} + \beta_{12}im_{it}$, normalizing on $\beta_{12} = 1$, for all three country groupings. There is a marked difference between the estimates of β_{11} across the advanced and emerging economies. For the advanced economies we have $\hat{\beta}_{11} = -0.914$ (0.029), which strongly rejects the null hypothesis of -1 and suggests systematic differences can persist between merchandise imports and exports for advanced economies. In contrast the estimate for the emerging economies, namely $\hat{\beta}_{11} = -0.992$ (0.045), is not significantly different from -1. For all countries pooled, $\hat{\beta}_{11} = -0.972$ (0.034) and does not reject the null of $\beta_{11} = -1$.[12] The difference in the estimates obtained for advanced and emerging economies could be due to greater ability of advanced economies in financing their goods trade imbalances by increasing their export of services and having easier access to international capital market.

Similar results are obtained when we consider the relation between wages and labour productivity. For all country groupings and both choices of the thresholds we find $\tilde{r} = 1$. See Table 12. The PME estimates of the long-run relation, $\boldsymbol{\beta}'\mathbf{w}_{it} = \beta_{11}wage_{it} + prod_{it}$ for advanced and emerging economies are $\hat{\beta}_{11} = -0.953$ (0.013) and $\hat{\beta}_{11} = -0.984$ (0.043), respectively. Once again estimates of β_{11} are quite close to -1, but as in the case of imports and exports the null of $\beta_{11} = -1$ is rejected for

[12] Normalizing on β_{11} yields the same results for β_{12} since PME estimate of β_{12} is exactly equal to $1/\hat{\beta}_{11}$ and by construction $Var(\hat{\beta}_{12}) = Var(1/\hat{\beta}_{11})$.

advanced economies but not for emerging economies.

TABLE 11: PME estimates for $\mathbf{w}_{it} = (ex_{it}, im_{it})'$, using $q = 2$ sub-sample time averages.

Sample:	Advanced	Emerging	All economies
Eigenvalues of $\mathbf{R}_{\bar{w}\bar{w}}$ given by (54) (in ascending order)			
$\tilde{\lambda}_1$	0.026	0.102	0.084
$\tilde{\lambda}_2$	1.974	1.898	1.916
Threshold $\bar{T}^{-\delta}$			
$\delta = 1/2$	0.128	0.134	0.132
$\delta = 1/4$	0.357	0.366	0.364
Estimated number of long-run relations ($\tilde{r}$)			
$\tilde{r}$ ($\delta = 1/2$)	1	1	1
$\tilde{r}$ ($\delta = 1/4$)	1	1	1
Exactly identified long-run relations			
$\boldsymbol{\beta}'\mathbf{w}_{it} = \begin{pmatrix} \beta_{11} & 1 \end{pmatrix} \begin{pmatrix} ex_{it} \\ im_{it} \end{pmatrix} = \beta_{11} ex_{it} + im_{it}$			
$\hat{\beta}_{11}$	-0.914	-0.992	-0.972
	(0.029)	(0.045)	(0.034)
Sample dimensions			
n	38	139	177
$\bar{T} = n^{-1}\sum_{i=1}^{n} T_i$	61.3	56.1	57.2

Notes: Standard errors are reported in parentheses. See Section S8 of the supplement for variable definitions, data sources and availability, and filters applied.

TABLE 12: PME estimates for $\mathbf{w}_{it} = (wage_{it}, prod_{it})'$, using $q = 2$ sub-sample time averages.

Sample:	Advanced	Emerging	All economies
Eigenvalues of $\mathbf{R}_{\bar{w}\bar{w}}$ given by (54) (in ascending order)			
$\tilde{\lambda}_1$	0.004	0.042	0.015
$\tilde{\lambda}_2$	1.996	1.958	1.985
Threshold $\bar{T}^{-\delta}$			
$\delta = 1/2$	0.135	0.146	0.139
$\delta = 1/4$	0.367	0.382	0.373
Estimated number of long-run relations ($\tilde{r}$)			
$\tilde{r}$ ($\delta = 1/2$)	1	1	1
$\tilde{r}$ ($\delta = 1/4$)	1	1	1
Exactly identified long-run relations			
$\boldsymbol{\beta}'\mathbf{w}_{it} = \begin{pmatrix} \beta_{11} & 1 \end{pmatrix} \begin{pmatrix} prod_{it} \\ wage_{it} \end{pmatrix} = \beta_{11} prod_{it} + wage_{it}$			
$\hat{\beta}_{11}$	-0.953	-0.984	-0.962
	(0.013)	(0.043)	(0.016)
Sample dimensions			
n	35	24	59
$\bar{T} = n^{-1}\sum_{i=1}^{n} T_i$	55.5	47.4	52.2

Notes: See notes to Table 11.

To illustrate how our proposed methods perform in the case of multiple long-run relations we now consider all the four variables together and set $\mathbf{w}_{it} = (ex_{it}, im_{it}, prod_{it}, wage_{it})'$. Given the above pair-wise results we would expect at least two long-run relations amongst these four variables. But the four eigenvalues of $\mathbf{R}_{\bar{w}\bar{w}}$ reported in Table 13 clearly suggest that there are three long-run relations among the four variables, irrespective whether we use $\delta = 1/4$ or $1/2$ as the threshold parameter. This result highlights the advantage of considering the possibility of multiple long-run relations and can help with discovery of hitherto unnoticed or overlooked long relations. For the third long-run relation we consider a possible long-run relation between exports and productivity. There is extensive microeconomic evidence suggesting exporting firms tend to have higher productivity. See, for example, Bernard and Jensen (2004). However, there is controversy about whether the relation arises because high productivity firms export more or whether the competitive pressure of exporting boosts productivity. See for example Aghion et al. (2018). Also, while the firm level micro relation has been examined for many countries, the country level macro relation has been less intensively explored. Our application provides cross country evidence on the relation between exports and productivity without making any assumption about the direction of causality between these variables. Accordingly, we consider the following exactly identified long-run relations assuming that $r_0 = 3$ amongst the four variables $\mathbf{w}_{it}$,

$$\mathbf{B}'\mathbf{w}_{it} = \begin{pmatrix} \beta_{11} & 1 & 0 & 0 \\ 0 & 0 & \beta_{23} & 1 \\ \beta_{31} & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} ex_{it} \\ im_{it} \\ prod_{it} \\ wage_{it} \end{pmatrix}. \tag{61}$$

PME estimates for β_{11} and β_{23} in Table 13 are in line with the estimates in Tables 11 and 12. For the third long-run relation, we obtain $\hat{\beta}_{31} = -0.509$ (0.023) for advanced economies and $\hat{\beta}_{31} = -0.426$ (0.032) for emerging economies. Hence, we discovered that exports and productivity are related with the expected sign. Unlike the other two long-run relations, we do not have any *a priori* reason to believe the estimates of β_{31} should be close to -1.

To corroborate the evidence in Table 13 regarding the third long-run relation, Table 14 reports PME findings for $\mathbf{w}_{it} = (prod_{it}, ex_{it})'$. There is a clear separation of eigenvalues, indicating existence of a long-run relation, in line with findings in Table 14 for the four-variable vector $\mathbf{w}_{it}$. In addition, the estimated coefficients in Table 14 are very similar to the corresponding coefficients reported in Table 13.

TABLE 13: PME estimates for the variable set $\{wage_{it}, prod_{it}, im_{it}, ex_{it}\}$, using $q = 2$ sub-sample time averages.

Sample:	Advanced	Emerging	All economies
Eigenvalues of $\mathbf{R}_{\bar{w}\bar{w}}$ given by (54) (in ascending order)			
$\tilde{\lambda}_1$	0.003	0.012	0.014
$\tilde{\lambda}_2$	0.010	0.037	0.015
$\tilde{\lambda}_3$	0.064	0.111	0.088
$\tilde{\lambda}_4$	3.923	3.840	3.883
Threshold $\bar{T}^{-\delta}$			
$\delta = 1/2$	0.135	0.146	0.139
$\delta = 1/4$	0.367	0.382	0.373
Estimated number of long-run relations ($\tilde{r}$)			
$\tilde{r}$ ($\delta = 1/2$)	3	3	3
$\tilde{r}$ ($\delta = 1/4$)	3	3	3
Exactly identified long-run relations			
$\mathbf{B}'\mathbf{w}_{it} = \begin{pmatrix} \beta_{11} & 1 & 0 & 0 \\ 0 & 0 & \beta_{23} & 1 \\ \beta_{31} & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} ex_{it} \\ im_{it} \\ prod_{it} \\ wage_{it} \end{pmatrix}$			
$\hat{\beta}_{11}$	-0.882	-1.005	-0.928
	(0.027)	(0.031)	(0.023)
$\hat{\beta}_{23}$	-0.952	-0.954	-0.953
	(0.013)	(0.041)	(0.015)
$\hat{\beta}_{31}$	-0.509	-0.426	-0.478
	(0.023)	(0.032)	(0.021)
Sample dimensions			
n	35	24	59
$\bar{T} = n^{-1}\sum_{i=1}^{n} T_i$	55.5	47.4	52.2

Notes: See Section S8 of the supplement for variable definitions, data sources and availability, and filters applied.

TABLE 14: PME estimates for $\mathbf{w}_{it} = (ex_{it}, prod_{it})'$, using $q = 2$ sub-sample time averages.

Sample:	Advanced	Emerging	All economies
Eigenvalues of $\mathbf{R}_{\bar{w}\bar{w}}$ given by (54) (in ascending order)			
$\tilde{\lambda}_1$	0.025	0.080	0.061
$\tilde{\lambda}2$	1.975	1.920	1.939
Threshold $\bar{T}^{-\delta}$			
$\delta = 1/2$	0.135	0.146	0.139
$\delta = 1/4$	0.367	0.382	0.373
Estimated number of long-run relations ($\tilde{r}$)			
$\tilde{r}$ ($\delta = 1/2$)	1	1	1
$\tilde{r}$ ($\delta = 1/4$)	1	1	1
Exactly identified long-run relations			
$\mathbf{B}'\mathbf{w}_{it} = \begin{pmatrix} \beta_{11} & 1 \end{pmatrix} \begin{pmatrix} ex_{it} \\ prod_{it} \end{pmatrix} = \beta_{11} ex_{it} + prod_{it}$			
$\hat{\beta}_{11}$	-0.510	-0.357	-0.432
	(0.023)	(0.044)	(0.036)
Sample dimensions			
n	35	29	64
$\bar{T} = n^{-1}\sum_{i=1}^{n} T_i$	55.5	47.0	51.7

Notes: See notes to Table 11.

TABLE 15: Comparison of PME estimates with alternative estimates of single long-run relation, $\mathbf{w}_{it} = (w_{1it}, w_{2,t})'$

w_{1it} :	ex_{it}						$prod_{it}$						ex_{it}					
$w_{2,t}$:	im_{it}						$wage_{it}$						$prod_{it}$					
Coint. relation:	$\beta_{11}w_{1it} + w_{2,t}$ $\hat{\beta}_{11}$			$w_{1it} + \beta_{12}w_{2,t}$ $\hat{\beta}_{12}$			$\beta_{11}w_{1it} + w_{2,t}$ $\hat{\beta}_{11}$			$w_{1it} + \beta_{12}w_{2,t}$ $\hat{\beta}_{12}$			$\beta_{11}w_{1it} + w_{2,t}$ $\hat{\beta}_{11}$			$w_{1it} + \beta_{12}w_{2,t}$ $\hat{\beta}_{12}$		
Sample:	Adv.	Eme.	All	Adv.	Eme.	All	Adv.	Eme.	All	Adv.	Eme.	All	Adv.	Eme.	All	Adv.	Eme.	All
PME	-0.914	-0.992	-0.972	-1.094	-1.008	-1.029	-0.953	-0.984	-0.962	-1.049	-1.016	-1.039	-0.510	-0.357	-0.432	-1.962	-2.803	-2.315
	(0.029)	(0.045)	(0.034)	(0.035)	(0.045)	(0.036)	(0.013)	(0.043)	(0.016)	(0.015)	(0.056)	(0.021)	(0.023)	(0.044)	(0.036)	(0.075)	(0.251)	(0.119)
SPMG	-0.962	-1.026	-0.976	-1.040	-0.974	-1.025	-1.072	-0.987	-1.043	-0.933	-1.013	-0.959	-0.367	-0.704	-0.371	-2.729	-1.421	-2.697
	(0.004)	(0.006)	(0.004)	(0.004)	(0.006)	(0.004)	(0.005)	(0.004)	(0.003)	(0.005)	(0.004)	(0.003)	(0.004)	(0.013)	(0.003)	(0.027)	(0.033)	(0.024)
Breitung's 2-step	-0.842	-0.924	-0.902	-1.048	-0.882	-0.922	-1.144	-1.010	-1.066	-0.832	-0.924	-0.900	-0.522	-0.371	-0.455	-1.701	-1.901	-1.756
	(0.010)	(0.021)	(0.016)	(0.019)	(0.016)	(0.013)	(0.008)	(0.020)	(0.005)	(0.005)	(0.018)	(0.004)	(0.013)	(0.013)	(0.010)	(0.042)	(0.095)	(0.040)
PMG	-0.985	-0.996	-0.989	-0.973	-0.949	-0.960	-1.140	-0.990	-1.100	-0.837	-0.962	-0.886	-0.291	-0.718	-0.306	-1.568	-1.350	-1.527
	(0.007)	(0.008)	(0.005)	(0.009)	(0.009)	(0.006)	(0.008)	(0.007)	(0.005)	(0.006)	(0.009)	(0.004)	(0.007)	(0.020)	(0.006)	(0.026)	(0.077)	(0.024)
PB	-0.868	-0.813	-0.828	-0.993	-0.869	-0.891	-1.136	-0.966	-1.097	-0.844	-0.961	-0.888	-0.486	-0.379	-0.437	-1.688	-1.967	-1.771
	(0.027)	(0.033)	(0.025)	(0.039)	(0.037)	(0.031)	(0.006)	(0.035)	(0.002)	(0.005)	(0.031)	(0.002)	(0.028)	(0.056)	(0.038)	(0.066)	(0.293)	(0.103)
PFMOLS	-0.857	-0.835	-0.841	-1.083	-0.886	-0.934	-1.241	-0.978	-1.070	-0.776	-0.944	-0.896	-0.528	-0.329	-0.445	-1.723	-2.051	-1.822
	(0.023)	(0.035)	(0.026)	(0.050)	(0.043)	(0.032)	(0.015)	(0.485)	(0.009)	(0.010)	(0.163)	(0.008)	(0.064)	(0.060)	(0.092)	(0.059)	(0.191)	(0.081)
PDOLS	-0.912	-0.841	-0.856	-1.045	-0.801	-0.853	-1.093	-1.029	-1.089	-0.868	-1.029	-0.891	-0.520	-0.409	-0.470	-1.853	-1.957	-1.901
	(0.004)	(0.006)	(0.004)	(0.005)	(0.006)	(0.004)	(0.006)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.008)	(0.004)	(0.015)	(0.038)	(0.015)
Sample size																		
n	38	139	177	38	139	177	35	24	59	35	24	59	35	29	64	35	29	64
$\bar{T} = n^{-1}\sum_{i=1}^{n} T_i$	61.3	56.1	57.2	61.3	56.1	57.2	55.5	47.4	52.2	55.5	47.4	52.2	55.5	47.0	51.7	55.5	47.0	51.7

Notes: PME and SPMG are the only estimators where $\hat{\beta}_{11} = \hat{\beta}_{12}^{-1}$. This is not the case for the remaining estimators. PME is the pooled minimum eigenvalue estimator using $q = 2$ subsamples. SPMG is the system PMG estimator of Chudik, Pesaran, and Smith (2023b), using $p = 2$ lags in levels (1 lag in first differences). Breitung's 2-step is the two-step system estimator of Breitung (2005), using $p = 2$ lags in levels. PMG is the Pooled Mean Group (PMG) estimator by Pesaran, Shin, and Smith (1999), using $p = 2$ lags in levels. PB is the pooled Bewley estimator by Chudik, Pesaran, and Smith (2023a), using $p = 2$ lags in levels. PFMOLS is the panel FMOLS estimator by Pedroni (1996, 2001a, 2001b). PDOLS is the panel Dynamic OLS (PDOLS) by Mark and Sul (2003), using one lead and one lag of first differences.

10.3 Comparison of PME and their estimates in the case of pair-wise relations

Here we provide comparative estimates for the pair-wise long-run relations for which a number of alternative estimators are proposed in the literature. Specifically, we compare PME estimates reported in Tables 11, 12, and 14 with the SPMG, Breitung's 2-step, PMG, PB, FMOLS, and PDOLS estimators discussed in Sub-Section 9.3.3.[13] The results are summarized in Table 15. PME and SPMG are the only estimators that are invariant to the normalization imposed on $\boldsymbol{\beta}'\mathbf{w}_{it} = \beta_{11}w_{1,it} + \beta_{12}w_{2,it}$. As a result, the equality $\hat{\beta}_{11}\hat{\beta}_{12} = 1$ holds only in the case of PME and SPMG estimators.

The estimates of β_{11} and β_{12} are generally similar but there are some large differences. For instance, the PME estimate of the long-run coefficient in $\beta_{11}ex_{it} + prod_{it}$ in the case of advanced economies is -0.510 (0.023) compared with SPMG estimate of -0.367 (0.004). Such difference could arise from the possibility that some of the SPMG modeling assumptions regarding short-run dynamics are not met, whereas the PME estimator is more general in that it does not require modeling of short-run dynamics. We also observe that standard errors of the estimates proposed in the literature are in some cases 2 to 10-fold smaller compared to PME. As we have seen in Monte Carlo experiments all of the estimators proposed in the literature severely over-reject the null, whilst this is not the case if we consider the MC results for size reported in Table 6.

11 Concluding remarks

This paper provides a new, pooled minimum eigenvalue (PME), methodology for the analysis of multiple long-run relations in panel data models where the cross section dimension, n, is large relative to the time series dimension, T. It uses non-overlapping sub-sample time averages as deviations from their full-sample counterpart and estimates the number of long-run relations and their coefficients using eigenvalues and eigenvectors of the pooled covariance matrix of these sub-sample deviations. It ap-

[13] The SPMG estimator is proposed by Chudik, Pesaran, and Smith (2023b), the Breitung's two-step system estimator is proposed by Breitung (2005), the PMG estimator is proposed by Pesaran, Shin, and Smith (1999), the PB (Pooled Bewley) estimator is proposed by Chudik, Pesaran, and Smith (2023a), the panel FMOLS estimator is proposed by Pedroni (1996, 2001a, 2001b), and the panel DOLS estimator is proposed by Mark and Sul (2003).

plies to unbalanced panels generated from general linear processes with interactive stationary time effects and does not require knowing long-run causal linkages. The PME estimator is consistent and asymptotically normally distributed as n and $T \to \infty$ jointly, such that $T \approx n^d$, with $d > 0$ for consistency and $d > 1/2$ for asymptotic normality. Extensive Monte Carlo studies show that the number of long-run relations can be estimated with high precision and the PME estimates of the long-run coefficients show small bias and $RMSE$ and have good size and power properties. The utility of our approach is illustrated with both micro and macro applications. The micro application uncovers long-run relations among key financial variables in an unbalanced panel of US firms from merged CRSP-Compustat data set covering $2,000$ plus firms over the period $1950 - 2021$. The macro application uses cross country macroeconomic time series data covering up to 177 countries over slightly shorter time period, $1950 - 2019$.

As well as application in other areas than corporate finance and macroeconomics the procedure opens up a range of theoretical developments. One question that we are considering investigating is whether it is possible to develop a unit root test based on similar principles. Our method is semi-parametric, in that it does not model the short run dynamics. Because the estimates of the long-run relations are super consistent, then having estimated them, the PME estimates of the long-run relations could be used as inputs into second stage models using stationary variables. For instance, they can provide estimates of the disequilibrium terms in error correction models of the short run dynamics, to measure speeds of adjustment. Such models could also be used for medium-term forecasting and counterfactual analysis.

A Appendix

This appendix provides proofs of propositions and theorems stated in the paper. This appendix follows the same notations as in Section 4. Sub-sample time averages for a generic vector $\mathbf{x}_{it}$ are defined as $\bar{\mathbf{x}}_{i\ell} = \frac{1}{T_q}\sum_{t=(\ell-1)T_q+1}^{\ell T_q} \mathbf{x}_{it}$, for $\ell = 1, 2, ..., q$, and the full-sample time average by $\bar{\mathbf{x}}_{i\circ} = q^{-1}\sum_{\ell=1}^{q} \bar{\mathbf{x}}_{i\ell}$, where q (≥ 2), m is fixed, and $T_q = T/q$ is an integer. Small and large finite positive constants that do not depend on sample sizes n and T are denoted by ϵ and K, respectively. They can take different values at different instances. All lemmas referenced in the proofs below are provided in the supplement.

Proof of Proposition 1. Consider the last three terms on the right side of (13). For its second term we have $E\left\|n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right\| \leq n^{-1}\sum_{i=1}^{n}\|\mathbf{C}_i\| E\left\|\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right\| \leq \sup_i\|\mathbf{C}_i\|\sup_i E\left\|\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right\|$. By Assumption 2 $\sup_i\|\mathbf{C}_i\| < K$, and using (S.11) in Lemma 3 of the supplement we have $\sup_i E\left\|\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right\| = O\left(T^{-1}\right)$. Hence, it follows that $E\left\|n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right\| = O\left(T^{-1}\right)$. Similarly, $E\left\|n^{-1}\sum_{i=1}^{n}\mathbf{Q}'_{\bar{v}_i\bar{s}_i}\mathbf{C}'_i\right\| = O\left(T^{-1}\right)$, and, using (S.9) of Lemma 3, $E\left\|n^{-1}\sum_{i=1}^{n}\mathbf{Q}_{\bar{v}_i\bar{v}_i}\right\| = O\left(T^{-2}\right)$. Using these results in (13) now yields

$$E\left\|\mathbf{Q}_{\bar{w}\bar{w}} - n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{s}_i}\mathbf{C}'_i\right\| = O\left(T^{-1}\right). \tag{A.1}$$

Consider $n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{s}_i}\mathbf{C}'_i$ and note that it can be written as

$$n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{s}_i}\mathbf{C}'_i = n^{-1}\sum_{i=1}^{n}\mathbf{C}_i E\left(\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right)\mathbf{C}'_i + n^{-1}\sum_{i=1}^{n}\mathbf{V}_i, \tag{A.2}$$

where $\mathbf{V}_i = \mathbf{C}_i\left[\mathbf{Q}_{\bar{s}_i\bar{s}_i} - E\left(\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right)\right]\mathbf{C}'_i$. Using result (S.3) of Lemma 2 and noting that $sup_i\|\mathbf{\Sigma}_i\| < K$ under Assumption 1, we obtain

$$n^{-1}\sum_{i=1}^{n}\mathbf{C}_i E\left(\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right)\mathbf{C}'_i = \frac{(q-1)}{6q}\mathbf{\Psi}_n + O\left(T^{-2}\right), \tag{A.3}$$

where $\mathbf{\Psi}_n = n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{\Sigma}_i\mathbf{C}'_i$. In addition, uniformly bounded fourth moments of individual elements of $\mathbf{u}_{it}$ ensure variances of individual elements $\mathbf{Q}_{\bar{s}_i\bar{s}_i}$ are bounded. Since $\sup_i\|\mathbf{C}_i\| < K$ by Assumption 2, it also follows that the variances of the individual elements of $\mathbf{V}_i$ are uniformly bounded. Noting that $E(\mathbf{V}_i) = \mathbf{0}$ by construction, and that $\mathbf{V}_i$ is independently distributed of $\mathbf{V}_j$ for all $i \neq j$ (by Assumption 1), we have

$$n^{-1}\sum_{i=1}^{n}\mathbf{V}_i = O_p\left(n^{-1/2}\right). \tag{A.4}$$

Using (A.3) and (A.4) in (A.2) yields

$$n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{s}_i}\mathbf{C}_i' = \frac{(q-1)}{6q}\mathbf{\Psi}_n + O_p\left(n^{-1/2}\right) + O\left(T^{-2}\right), \tag{A.5}$$

and using this result in (A.1), we obtain $\mathbf{Q}_{\bar{w}\bar{w}} \rightarrow_p \frac{(q-1)}{6q}\mathbf{\Psi}$, as $n, T \rightarrow \infty$ (in no particular order), where $\mathbf{\Psi} = \lim_{n\rightarrow\infty}\mathbf{\Psi}_n$. This completes the proof of (19). Result (20) follows from (19) by noting that $\mathbf{C}_i'\mathring{\mathbf{B}}_0 = \mathbf{0}$ for all i by Assumption 3, which in turn yields $\mathbf{\Psi}_n\mathring{\mathbf{B}}_0 = \mathbf{\Psi}\mathring{\mathbf{B}}_0 = \mathbf{0}$. ■

Proof of Proposition 2. Under Assumption 3 $\left(n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{\Sigma}_i\mathbf{C}_i'\right)\mathring{\mathbf{B}}_0 = \mathbf{\Psi}_n\mathring{\mathbf{B}}_0 = \mathbf{0}$, for any n and as $n \rightarrow \infty$. Then we have $\mathbf{\Psi}\mathring{\mathbf{B}}_\mathbf{0} = \mathbf{0} = (\mathbf{P}'\mathbf{P})\,\mathring{\mathbf{B}}_0 = \mathbf{0}$. But by condition $rank\,(\mathbf{\Psi}) = m - r_0$ of Assumption 3, $\mathbf{P}$ is an $(m-r_0)\times m$ full row rank matrix. Then $\mathbf{PP}'$ is non-singular, and we must also have $\mathbf{P}\mathring{\mathbf{B}}_\mathbf{0} = \mathbf{0}$. Combing this result with (22) now yields

$$\begin{pmatrix}\mathbf{R}\\ \mathbf{P}\end{pmatrix}\mathring{\mathbf{B}}_0 = \begin{pmatrix}\mathbf{A}\\ \mathbf{0}\end{pmatrix}. \tag{A.6}$$

Under Assumptions 3 and the exact r_0^2 restrictions given by (22) with $rank\,(\mathbf{A}) = rank\,(\mathbf{R}) = r_0 < m$, condition $rank\begin{pmatrix}\mathbf{P}_n\\ \mathbf{R}\end{pmatrix} = rank\begin{pmatrix}\mathbf{P}\\ \mathbf{R}\end{pmatrix} = m$ holds, $\mathbf{R}'\mathbf{R}+\mathbf{P}'\mathbf{P}$ is a positive definite matrix, and $\boldsymbol{\beta}$ is uniquely determined by

$$\mathring{\mathbf{B}}_0 = (\mathbf{R}'\mathbf{R}+\mathbf{P}'\mathbf{P})^{-1}\begin{pmatrix}\mathbf{A}\\ \mathbf{0}\end{pmatrix} = (\mathbf{R}'\mathbf{R}+\mathbf{\Psi})^{-1}\begin{pmatrix}\mathbf{A}\\ \mathbf{0}\end{pmatrix}.$$

Now using the normalizations $\mathbf{R} = (\mathbf{I}_{r_0}, \mathbf{0})'$ and $\mathbf{A} = \mathbf{I}_{r_0}$ and conformably partitioning $\mathbf{\Psi}$ we have

$$\mathbf{R}'\mathbf{R}+\mathbf{\Psi} = \begin{pmatrix}\mathbf{I}_{r_0}+\mathbf{\Psi}_{11} & \mathbf{\Psi}_{21}'\\ \mathbf{\Psi}_{21} & \mathbf{\Psi}_{22}\end{pmatrix}.$$

Since $\mathbf{R}'\mathbf{R}+\mathbf{\Psi}$ is a positive definite matrix then the $r_0\times r_0$ and $(m-r_0)\times(m-r_0)$ matrices $\mathbf{I}_{r_0}+\mathbf{\Psi}_{11}$ and $\mathbf{\Psi}_{22}$ are also positive definite. Solving the following system of equations now yields

$$\begin{pmatrix} \mathbf{I}_{r_0} + \mathbf{\Psi}_{11} & \mathbf{\Psi}_{21}' \\ \mathbf{\Psi}_{21} & \mathbf{\Psi}_{22} \end{pmatrix} \begin{pmatrix} \mathring{\mathbf{B}}_{0,1} \\ \mathbf{\Theta} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{r_0} \\ \mathbf{0} \end{pmatrix},$$

and we obtain $(\mathbf{I}_r + \mathbf{\Psi}_{11}) \mathring{\mathbf{B}}_{0,1} + \mathbf{\Psi}_{21}'\mathbf{\Theta} = \mathbf{I}_{r_0}$, and $\mathbf{\Psi}_{21}\boldsymbol{\beta}_1 + \mathbf{\Psi}_{22}\mathbf{\Theta} = \mathbf{0}$. Also, since $\mathring{\mathbf{B}}_{0,1} = \mathbf{I}_{r_0}$ it follows that $(\mathbf{I}_{r_0} + \mathbf{\Psi}_{11}) + \mathbf{\Psi}_{21}'\mathbf{\Theta} = \mathbf{I}_{r_0}$, and $\mathbf{\Theta} = -\mathbf{\Psi}_{22}^{-1}\mathbf{\Psi}_{21}$,which implies that $\mathbf{\Psi}_{11} = \mathbf{\Psi}_{21}'\mathbf{\Psi}_{22}^{-1}\mathbf{\Psi}_{21}$, and in turn ensures that $\mathbf{\Psi}$ is rank deficient, as required under Assumption 3). ■

Proof of Theorem 1. Multiplying (A.5) by $\hat{\mathbf{B}}_0'$ from the left and by $\hat{\mathbf{B}}_0$ from the right, and noting eigenvectors $\hat{\mathbf{B}}_0$ are normalized so that $\hat{\mathbf{B}}_0'\hat{\mathbf{B}}_0 = \mathbf{I}_r$ yields

$$\hat{\mathbf{B}}_0'\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0 = \frac{(q-1)}{6q}\hat{\mathbf{B}}_0'\mathbf{\Psi}_n\hat{\mathbf{B}}_0 + O_p\left(n^{-1/2}\right) + O_p\left(T^{-2}\right), \tag{A.7}$$

where $lim_{n\to\infty}\mathbf{\Psi}_n = \mathbf{\Psi}$ as $n \to \infty$. Under Assumption 3, we have $rank\left(\mathbf{\Psi}\right) = m - r_0$, $rank\left(\mathring{\mathbf{B}}_0\right) = r$ and $\mathring{\mathbf{B}}_0'\mathbf{\Psi} = \mathbf{0}$. Hence the space space spanned by the column vectors of $m \times r_0$ matrix $\mathring{\mathbf{B}}_0$ is the same as the space spanned by the first r_0 eigenvectors of $\mathbf{\Psi}$ associated with its r smallest eigenvalues. It now follows from (A.7) that the space spanned by the column vectors of $\hat{\mathbf{B}}_0$ converges to the space spanned by the column vectors of $\mathring{\mathbf{B}}_0$, as $n, T \to \infty$ (in no particular order), and $\hat{\mathbf{B}}_0\mathbf{H} \to_p \mathring{\mathbf{B}}_0$, for a suitable choice of $r_0 \times r_0$ nonsingular matrix $\mathbf{H}$.[14] ■

Proof of Theorem 2. Multiplying both sides of (13) by $\mathring{\mathbf{B}}_0$ from the right, and noting that $\mathbf{C}_i'\mathring{\mathbf{B}}_0 = \mathbf{0}$ under Assumption 3, we have

$$\mathbf{Q}_{\bar{w}\bar{w}}\mathring{\mathbf{B}}_0 = \left(n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0 + \mathbf{Q}_{\bar{v}\bar{v}}\mathring{\mathbf{B}}_0 \tag{A.8}$$

Lemma 7 established $\left\|\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0\right\| = O_p\left(n^{-1/2}T^{-2}\right)$, and given that $\hat{\mathring{\mathbf{B}}}_0$ is an $O_p(1)$ rotation of $\hat{\mathbf{B}}_0$, it follows

$$\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathring{\mathbf{B}}}_0 = O_p\left(n^{-1/2}T^{-2}\right). \tag{A.9}$$

[14]The rotation matrix is given by $\mathbf{H} = \left[\left(\mathring{\mathbf{B}}_0'\mathring{\mathbf{B}}_0\right)^{-1}\mathring{\mathbf{B}}_0'\hat{\mathbf{B}}\right]^{-1}$, see Lemma 13.1 of Johansen (1995).

Subtracting (A.8) from (A.9) yields

$$\mathbf{Q}_{\bar{w}\bar{w}}\left(\hat{\mathring{\mathbf{B}}}_0-\mathring{\mathbf{B}}_0\right)=-\left(n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0-\mathbf{Q}_{\bar{v}\bar{v}}\mathring{\mathbf{B}}_0+O_p\left(n^{-1/2}T^{-2}\right). \tag{A.10}$$

Result (S.47) of Lemma (5) implies $\mathbf{Q}_{\bar{v}\bar{v}}=O_p\left(n^{-1/2}T^{-2}\right)$, and we have

$$\mathbf{Q}_{\bar{w}\bar{w}}\sqrt{n}T\left(\hat{\mathring{\mathbf{B}}}_0-\mathring{\mathbf{B}}_0\right)=-\left(n^{-1/2}\sum_{i=1}^{n}T\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0+O_p\left(T^{-1}\right). \tag{A.11}$$

The first term on the right side of (A.11) can be written as

$$\left(n^{-1/2}\sum_{i=1}^{n}T\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0=\left(n^{-1/2}\sum_{i=1}^{n}\mathbf{Z}_i\right)\mathring{\mathbf{B}}_0+\frac{\sqrt{n}}{T}\left[n^{-1}\sum_{i=1}^{n}\mathbf{C}_iE\left(T^2\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\right]\mathring{\mathbf{B}}_0,$$

where $\mathbf{Z}_i=\mathbf{C}_i\left[T\mathbf{Q}_{\bar{s}_i\bar{v}_i}-E\left(T\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\right]\mathring{\mathbf{B}}_0$. Using result (S.12) of Lemma 3, we obtain

$$\left\|n^{-1}\sum_{i=1}^{n}\mathbf{C}_iE\left(T^2\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0\right\|\leq T^2\sup_i\left\|\mathbf{C}_i\right\|\sup_i\left\|E\left(\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\right\|\left\|\mathring{\mathbf{B}}_0\right\|<K, \tag{A.12}$$

where $\left\|\mathring{\mathbf{B}}_0\right\|<K$ and $\sup_i\left\|\mathbf{C}_i\right\|_1<K$ by Assumption 2. Using the above result in (A.11) it follows that

$$\mathbf{Q}_{\bar{w}\bar{w}}\sqrt{n}T\left(\hat{\mathring{\mathbf{B}}}_0-\mathring{\mathbf{B}}_0\right)=-n^{-1/2}\sum_{i=1}^{n}\mathbf{Z}_i+O_p\left(\frac{\sqrt{n}}{T}\right)+O_p\left(T^{-1}\right),$$

Vectorizing the above equation we have

$$\left(\mathbf{I}_r\otimes\mathbf{Q}_{\bar{w}\bar{w}}\right)\sqrt{n}T\ vec\left(\hat{\mathring{\mathbf{B}}}_0-\mathring{\mathbf{B}}_0\right)=n^{-1/2}\sum_{i=1}^{n}\left(\mathring{\mathbf{B}}_0'\otimes\mathbf{C}_i\right)\tilde{\boldsymbol{\xi}}_{iq}+O_p\left(\frac{\sqrt{n}}{T}\right)+O_p\left(T^{-1}\right), \tag{A.13}$$

where $\tilde{\boldsymbol{\xi}}_{iq}=\bar{\boldsymbol{\xi}}_{iq}-E\left(\bar{\boldsymbol{\xi}}_{iq}\right)$, and $\bar{\boldsymbol{\xi}}_{iq}$ is given by (recall $\mathbf{Z}_i=\mathbf{C}_i\left[T\mathbf{Q}_{\bar{s}_i\bar{v}_i}-E\left(T\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\right]\mathring{\mathbf{B}}_0$

and $\mathbf{Q}_{\bar{s}_i\bar{v}_i} = T^{-1}q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{s}}_{i\ell}-\bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell}-\bar{\mathbf{v}}_{i\circ}\right)')$

$$\begin{aligned}\bar{\boldsymbol{\xi}}_{iq} &= q^{-1}\sum_{\ell=1}^{q} vec\left[\left(\bar{\mathbf{s}}_{i\ell}-\bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell}-\bar{\mathbf{v}}_{i\circ}\right)'\right] = q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{v}}_{i\ell}-\bar{\mathbf{v}}_{i\circ}\right)\otimes\left(\bar{\mathbf{s}}_{i\ell}-\bar{\mathbf{s}}_{i\circ}\right), \\ &= q^{-1}\sum_{\ell=1}^{q}\bar{\mathbf{v}}_{i\ell}\otimes\bar{\mathbf{s}}_{i\ell} - \left(q^{-1}\sum_{\ell=1}^{q}\bar{\mathbf{v}}_{i\ell}\right)\otimes\bar{\mathbf{s}}_{i\circ} - \bar{\mathbf{v}}_{i\circ}\otimes\left(q^{-1}\sum_{\ell=1}^{q}\bar{\mathbf{s}}_{i\ell}\right) + \bar{\mathbf{v}}_{i\circ}\otimes\bar{\mathbf{s}}_{i\circ}.\end{aligned}$$

Using $q^{-1}\sum_{\ell=1}^{q}\bar{\mathbf{v}}_{i\ell} = \bar{\mathbf{v}}_{i\circ}$ and $q^{-1}\sum_{\ell=1}^{q}\bar{\mathbf{s}}_{i\ell} = \bar{\mathbf{s}}_{i\circ}$, the expression for $\bar{\boldsymbol{\xi}}_{iq}$ simplifies to $\bar{\boldsymbol{\xi}}_{iq} = q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{v}}_{i\ell}-\bar{\mathbf{v}}_{i\circ}\right)\otimes\left(\bar{\mathbf{s}}_{i\ell}-\bar{\mathbf{s}}_{i\circ}\right) = q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{v}}_{i\ell}\otimes\bar{\mathbf{s}}_{i\ell}\right) - \bar{\mathbf{v}}_{i\circ}\otimes\bar{\mathbf{s}}_{i\circ}$. Lemma 8 established convergence in distribution for $n^{-1/2}\sum_{i=1}^{n}\left(\mathring{\mathbf{B}}'\otimes\mathbf{C}_i\right)\tilde{\boldsymbol{\xi}}_{iq}$. Using this Lemma in (A.13), and noting that $d > 1/2$ implies$\sqrt{n}/T \to 0$ as $n, T \to \infty$, we obtain (32), as required. ■

References

Aghion, P., A. Bergeaud, M. Lequien, and M. J. Melitz (2018). The Impact of Exports on Innovation: Theory and Evidence. Banque de France Working Paper No. 678, April 2018.

Bai, J., C. Kao, and S. Ng (2009). Panel cointegration with global stochastic trends. *Journal of Econometrics 149*, 82–99.

Banerjee, A. and J. L. C. i Silvestre (2024). Panel Data Cointegration Testing with Structural Instabilities. *Journal of Business and Economic Statistics, 43*, 122–133.

Bernard, A. B. and J. B. Jensen (2004). Exporting and Productivity in the USA. *Oxford Review of Economic Policy 20*, 343–357.

Breitung, J. (2005). A parametric approach to the estimation of cointegration vectors in panel data. *Econometric Reviews 24*, 151–173.

Breitung, J. and M. H. Pesaran (2008). Unit roots and cointegration in panels. In L. Matyas and P. Sevestre (Eds.), *The Econometrics of Panel Data*, Volume 46, pp. 279–322. Berlin, Heidelberg: Elsevier BV.

Bykhovskaya, A. and V. Gorin (2022). Cointegration in large VARs. *Annals of Statistics 50*, 1593–1617.

Choi, I. (2015, 01). Panel cointegration. In *The Oxford Handbook of Panel Data*. Oxford University Press.

Chudik, A., M. H. Pesaran, and R. P. Smith (2023a). Pooled Bewley estimator of long-run relationships in dynamic heterogenous panels. *Econometrics and Statistics*. In Press, Corrected Proof. Available online 3 November 2023.

Chudik, A., M. H. Pesaran, and R. P. Smith (2023b). Revisiting the Great Ratios Hypothesis. *Oxford Bulletin of Economics and Statistics 85*, 1023–1047.

Coles, J. L. and Z. F. Li (2023). An Empirical Assessment of Empirical Corporate Finance. *Journal of financial and quantitative analysis 58*(4), 1391–1430.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press.

de Jong, R. M. and M. Wagner (2025). Panel cointegrating polynomial regression analysis and an illustration with the environmental kuznets curve. *Econometrics and Statistics 33*, 135–165.

Ditzen, J., Y. Karavias, and J. Westerlund (2025). Multiple Structural Breaks in Interactive Effects Panel Data Models. *Journal of Applied Econometrics 40*, 74–88.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). The Next Generation of the Penn World Table. *American Economic Review 105*, 3150–3182. available for download at www.ggdc.net/pwt.

Geelen, T., J. Hajda, E. Morellec, and A. Winegar (2024). Asset life, leverage, and debt maturity matching. *Journal of Financial Economics 154*, 103796.

Groen, J. J. J. and F. Kleibergen (2003). Likelihood-Based Cointegration Analysis in Panels of Vector Error-Correction Models. *Journal of Business and Economic Statistics 21*, 295–318.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Im, K. S., M. Pesaran, and Y. Shin (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics 115*, 53–74.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control 12*(2), 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica 59*, 1551.

Johansen, S. (1995). *Likelihood Based Inference on Cointegration in the Vector Autoregressive Model*. Oxford: Oxford University Press.

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

Larsson, R. and J. Lyhagen (2007). Inference in panel cointegration models with long panels. *Journal of Business and Economic Statistics 25*, 473–483.

Lev, B. and S. Sunder (1979). Methodological issues in the use of financial ratios. *Journal of Accounting and Economics 1*, 187–210.

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.

Müller, U. K. and M. W. Watson (2018). Long-run covariability. *Econometrica 86*, 775–804.

Onatski, A. and C. Wang (2018). Alternative asymptotics for cointegration tests in large VARs. *Econometrica 86*, 1465–1478.

Onatski, A. and C. Wang (2019). Extreme canonical correlations and high-dimensional cointegration analysis. *Journal of Econometrics 212*, 307–322.

Pedroni, P. (1996). Fully Modified OLS for Heterogeneous Cointegrated Panels and the Case of Purchasing Power Parity. Indiana University working papers in economics no. 96-020 (June 1996).

Pedroni, P. (2001a). Fully modified OLS for heterogeneous cointegrated panels. In B. Baltagi, T. Fomby, and R. C. Hill (Eds.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels, (Advances in Econometrics, Vol. 15)*, pp. 93–130. Emerald Group Publishing Limited, Bingley.

Pedroni, P. (2001b). Purchasing Power Parity Tests in Cointegrated Panels. *The Review of Economics and Statistics 83*, 727–731.

Pesaran, M. H. (2007). A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence. *Journal of Applied Econometrics 22*, 265–312.

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.

Petrov, V. V. (1992). Moments of Sums of Independent Random Variables. *Journal of Soviet Mathematics 61*, 1905–1906.

Phillips, P. C. B. and S. Ouliaris (1988). Testing for cointegration using principal component methods. *Journal of Economic Dynamics and Control 12*, 205–230.

Phillips, P. C. B. and S. Ouliaris (1990). Asymptotic Properties of Residual Based Tests for Cointegration. *Econometrica 58*, 165–193.

Westerlund, J. (2005). New Simple Tests for Panel Cointegration. *Econometric Reviews 24*, 297–316.

Supplement
"Analysis of Multiple Long-Run Relations in Panel Data Models"[15]

Alexander Chudik
Federal Reserve Bank of Dallas

M. Hashem Pesaran
Trinity College, Cambridge, UK and University of Southern California, USA

Ron P. Smith
Birkbeck, University of London, United Kingdom

September 11, 2025

This supplement is organized in seven sections. Section S1 presents lemmas and their proofs. Section S2 describes implementation of the PME estimator for unbalanced panels. Section S3 provides theorems and proofs for the consistency and asymptotic distribution of the PME estimator for the model with interactive time effects. Section S4 describes the Monte Carlo data generating processes and provides a list of individual Monte Carlo experiments. Section S5 investigates sensitivity of the PME estimator of r_0, $\tilde{r}$ given by (55), to different scaling of the observations on $\mathbf{w}_{it}$. Section S6 presents Monte Carlo findings for robustness of PME estimators to GARCH and threshold autoregressive effects. Section S7 describes data sources and variable constructions for the micro application, and Section S8 provides supplementary information for the macro application.

[15] The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas or the Federal Reserve System.

S1 Lemmas

Lemma 1 *Let* $\mathbf{v}_{it} = \mathbf{C}_i^*(L)\mathbf{u}_{it} = \sum_{j=0}^{\infty} \mathbf{C}_{ij}^* \mathbf{u}_{i,t-j}$ *and suppose Assumptions 1 and 2 hold. Then there exist finite positive constants* K_1*,* K_2*, and* $0 < \rho < 1$ *such that*

$$\sup_i \|\mathbf{\Gamma}_i(h)\| < K_1 \rho^h, \tag{S.1}$$

and

$$\sum_{h=0}^{\infty} \sup_i \|\mathbf{\Gamma}_i(h)\| < K_2, \tag{S.2}$$

where $\mathbf{\Gamma}_i(h) = E\left(\mathbf{v}_{it}\mathbf{v}'_{i,t-h}\right)$ *, for* $h = 0, 1, 2, ...$ *is the autocovariance function of* $\mathbf{v}_{it}$*.*

Proof. The autocovariance function of $\mathbf{v}_{it}$ is

$$\begin{aligned} \mathbf{\Gamma}_i(h) &= E\left(\mathbf{v}_{it}\mathbf{v}'_{i,t-h}\right) = E\left(\sum_{j=0}^{\infty} \mathbf{C}_{ij}^* \mathbf{u}_{i,t-j}\right)\left(\sum_{j'=0}^{\infty} \mathbf{u}'_{i,t-j'-h} \mathbf{C}_{ij'}^{*\prime}\right) \\ &= \sum_{j=0}^{\infty}\sum_{j'=0}^{\infty} \mathbf{C}_{ij}^* E\left(\mathbf{u}_{i,t-j}\mathbf{u}'_{i,t-j'-h}\right)\mathbf{C}_{ij'}^{*\prime} = \sum_{j=h}^{\infty} \mathbf{C}_{ij}^* \mathbf{\Sigma}_i \mathbf{C}_{i,j-h}^{*\prime} = \mathbf{\Gamma}'_i(-h). \end{aligned}$$

Taking spectral norm and supremum over i yields

$$\sup_i \|\mathbf{\Gamma}_i(h)\| \le \sum_{j=h}^{\infty} \sup_i \left\|\mathbf{C}_{ij}^*\right\| \sup_i \|\mathbf{\Sigma}_i\| \sup_i \left\|\mathbf{C}_{i,j-h}^*\right\| \le K_0 \sum_{j=h}^{\infty} \rho^j \rho^{j-h} = \frac{K_0 \rho^h}{1-\rho^2} < K_1 \rho^h,$$

where $\sup_i \|\mathbf{\Sigma}_i\| < K$ by Assumption 1 and $\sup_i \left\|\mathbf{C}_{ij}^*\right\| < K\rho^j$ with $0 < \rho < 1$ by Assumption 2. It follows $\sum_{h=0}^{\infty} \sup_i \|\mathbf{\Gamma}_i(h)\| < K_2$. ■

Lemma 2 *Consider the* $m \times m$ *matrix* $\mathbf{Q}_{\bar{s}_i\bar{s}_i} = T^{-1}q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right)'$*, where* $\bar{\mathbf{s}}_{i\ell}$ *and* $\bar{\mathbf{s}}_{i\circ}$ *are sub-sample and full sample time averages of the partial sum process* $\mathbf{s}_{it} = \sum_{\ell=1}^{t} \mathbf{u}_{i\ell}$*. Suppose* $E(\mathbf{u}_{it}\mathbf{u}'_{it}) = \mathbf{\Sigma}_i$ *for* $i = 1, 2, ..., n$ *and* $t = 1, 2, ..., T$*, and* $E(\mathbf{u}_{it}\mathbf{u}'_{it'}) = \mathbf{0}$ *for* $i = 1, 2, ..., n$*, and all* $t \neq t'$*,* $t, t' = 1, 2, ..., T$*. Then*

$$E\left(\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right) = \frac{(q-1)}{6}\left(\frac{1}{q} + \frac{1}{T^2}\right)\mathbf{\Sigma}_i, \tag{S.3}$$

for $q \ge 2$ *and* $i = 1, 2, ..., n$*.*

Proof. We note that

$$E\left(T\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right)=E\left[q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{s}}_{i\ell}-\bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{s}}_{i\ell}-\bar{\mathbf{s}}_{i\circ}\right)'\right]=q^{-1}\sum_{\ell=1}^{q}E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{s}}_{i\ell}'\right)-E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{s}}_{i\circ}'\right). \quad \text{(S.4)}$$

Also

$$\begin{aligned}\bar{\mathbf{s}}_{i1} &= \frac{1}{T_q}\left(\mathbf{s}_{i1}+\mathbf{s}_{i2}+....+\mathbf{s}_{i,T_q}\right)=\frac{1}{T_q}\left[\mathbf{u}_{i1}+(\mathbf{u}_{i1}+\mathbf{u}_{i2})+....\left(\mathbf{u}_{i1}+\mathbf{u}_{i2}+...+\mathbf{u}_{i,T_q}\right)\right],\\ &= \frac{1}{T_q}\left[T_q\mathbf{u}_{i1}+(T_q-1)\mathbf{u}_{i2}+(T_q-2)\mathbf{u}_{i3}+....+2\mathbf{u}_{i,T_q-1}+\mathbf{u}_{i,T_q}\right],\end{aligned}$$

which can be written more compactly as $\bar{\mathbf{s}}_{i1}=T_q^{-1}\sum_{t=1}^{T_q}\left(T_q-t+1\right)\mathbf{u}_{it}$. Similarly

$$\begin{aligned}\bar{\mathbf{s}}_{i2} &= \frac{1}{T_q}\left(\mathbf{s}_{i,T_q+1}+\mathbf{s}_{i,T_q+2}+....+\mathbf{s}_{i,2T_q}\right)\\ &= \frac{1}{T_q}\left[T_q\left(\mathbf{u}_{i1}+\mathbf{u}_{i2}+...+\mathbf{u}_{i,T_q}\right)+\sum_{t=T_q+1}^{2T_q}\left(2T_q-t+1\right)\mathbf{u}_{it}\right],\\ &= \sum_{t=1}^{T_q}\mathbf{u}_{it}+\frac{1}{T_q}\sum_{t=T_q+1}^{2T_q}\left(2T_q-t+1\right)\mathbf{u}_{it},\end{aligned}$$

and more generally

$$\bar{\mathbf{s}}_{i\ell}=\sum_{t=1}^{(\ell-1)T_q}\mathbf{u}_{it}+\frac{1}{T_q}\sum_{t=(\ell-1)T_q+1}^{\ell T_q}\left(\ell T_q+1-t\right)\mathbf{u}_{it},\text{ for }\ell=1,2,...,q. \quad \text{(S.5)}$$

It now readily follows that

$$E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{s}}_{i\ell}'\right)=\left[\left(\ell-1\right)T_q+\left(\frac{1}{T_q}\right)^2\sum_{t=(\ell-1)T_q+1}^{\ell T_q}\left(\ell T_q-t+1\right)^2\right]\boldsymbol{\Sigma}_i.$$

But

$$\sum_{t=(\ell-1)T_q+1}^{\ell T_q}\left(\ell T_q-t+1\right)^2=T_q^2+(T_q-1)^2+...+1=\frac{T_q(T_q+1)(2T_q+1)}{6},$$

and we obtain

$$E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{s}}_{i\ell}'\right)=\left[\left(\ell-1\right)T_q+\left(\frac{1}{T_q}\right)^2\frac{T_q(T_q+1)(2T_q+1)}{6}\right]\mathbf{\Sigma}_i=\left[\left(\ell-1\right)T_q+\frac{(T_q+1)(2T_q+1)}{6T_q}\right]\mathbf{\Sigma}_i,$$

and

$$q^{-1}\sum_{\ell=1}^{q}E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{s}}_{i\ell}'\right)=q^{-1}\sum_{\ell=1}^{q}\left(\left(\ell-1\right)T_q+\frac{(T_q+1)(2T_q+1)}{6T_q}\right)\mathbf{\Sigma}_i.$$

Noting that

$$q^{-1}\sum_{\ell=1}^{q}\left(\ell-1\right)T_q=q^{-1}\left[T_q+2T_q+...+(q-1)T_q\right]=q^{-1}T_q\frac{q(q-1)}{2}=\frac{(q-1)T_q}{2},$$

we have

$$q^{-1}\sum_{\ell=1}^{q}E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{s}}_{i\ell}'\right)=\left[\frac{(q-1)T_q}{2}+\frac{(T_q+1)(2T_q+1)}{6T_q}\right]\mathbf{\Sigma}_i.$$

Recalling that $T_q=T/q$, we can write the expression above as

$$q^{-1}\sum_{\ell=1}^{q}E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{s}}_{i\ell}'\right)=\left[\frac{(q-1)T}{2q}+\frac{(T+q)(2T+q)}{6qT}\right]\mathbf{\Sigma}_i. \tag{S.6}$$

Similarly, for the full sample average we have $\bar{\mathbf{s}}_{i\circ}=T^{-1}\sum_{t=1}^{T}\left(T-t+1\right)\mathbf{u}_{it}$, and

$$E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{s}}_{i\circ}'\right)=\frac{1}{T^2}\left[T^2+(T-1)^2+....+1\right]\mathbf{\Sigma}_i=\frac{T(T+1)(2T+1)}{6T^2}\mathbf{\Sigma}_i. \tag{S.7}$$

Using (S.6) and (S.7) in (S.4), yields

$$E\left(T\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right)=\left[\frac{T(q-1)}{2q}+\frac{(T+q)(2T+q)}{6Tq}\right]\mathbf{\Sigma}_i-\frac{(T+1)(2T+1)}{6T}\mathbf{\Sigma}_i=\frac{(q-1)}{6}\left(\frac{T}{q}+\frac{1}{T}\right)\mathbf{\Sigma}_i, \tag{S.8}$$

and result (S.3) follows. ■

Lemma 3 *Consider $m\times m$ matrices $\mathbf{Q}_{\bar{v}_i\bar{v}_i}=T^{-1}q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{v}}_{i\ell}-\bar{\mathbf{v}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell}-\bar{\mathbf{v}}_{i\circ}\right)'$ and $\mathbf{Q}_{\bar{s}_i\bar{v}_i}=T^{-1}q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{s}}_{i\ell}-\bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell}-\bar{\mathbf{v}}_{i\circ}\right)'$, where $\bar{\mathbf{s}}_{i\ell}$, $\bar{\mathbf{s}}_{i\circ}$ are the sub-sample and full sample time averages of the partial sum process $\mathbf{s}_{it}=\sum_{\ell=1}^{t}\mathbf{u}_{it}$, where m and q (≥ 2) are fixed. Also $\bar{\mathbf{v}}_{i\ell}$ and $\bar{\mathbf{v}}_{i\circ}$ are the sub-sample and full sample time averages of $\mathbf{v}_{it}=$*

$\mathbf{C}_i^*(L)\mathbf{u}_{it} = \sum_{j=0}^{\infty} \mathbf{C}_{ij}^* \mathbf{u}_{i,t-j}$, *and Assumptions 1 and 2 hold. Then*

$$\sup_i E \left\| \mathbf{Q}_{\bar{v}_i \bar{v}_i} \right\| = O\left(T^{-2}\right), \tag{S.9}$$

$$\sup_i \left\| T^2 \ E\left(\mathbf{Q}_{\bar{v}_i \bar{v}_i}\right) \right\| = O\left(T^{-1}\right) + O\left(\rho^{T/q}\right), \tag{S.10}$$

$$\sup_i E \left\| \mathbf{Q}_{\bar{s}_i \bar{v}_i} \right\| = O\left(T^{-1}\right), \tag{S.11}$$

and

$$\sup_i \left\| E\left(\mathbf{Q}_{\bar{s}_i \bar{v}_i}\right) \right\| = O\left(T^{-2}\right). \tag{S.12}$$

Proof. Consider

$$T\mathbf{Q}_{\bar{v}_i \bar{v}_i} = q^{-1} \sum_{\ell=1}^{q} \left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)' = q^{-1} \sum_{\ell=1}^{q} \bar{\mathbf{v}}_{i\ell}\bar{\mathbf{v}}_{i\ell}' - \bar{\mathbf{v}}_{i\circ}\bar{\mathbf{v}}_{i\circ}',$$

where $\bar{\mathbf{v}}_{i\circ} = q^{-1}\sum_{\ell=1}^{q} \bar{\mathbf{v}}_{i\ell}$. Taking spectral norm and expectations yields

$$E\left\| T\mathbf{Q}_{\bar{v}_i \bar{v}_i} \right\| \leq q^{-1} \sum_{\ell=1}^{q} E\left\| \bar{\mathbf{v}}_{i\ell} \right\|^2 + E\left\| \bar{\mathbf{v}}_{i\circ} \right\|^2,$$

and it follows

$$\sup_i E\left\| T\mathbf{Q}_{\bar{v}_i \bar{v}_i} \right\| \leq \sup_{i,\ell} E\left\| \bar{\mathbf{v}}_{i\ell} \right\|^2 + \sup_i E\left\| \bar{\mathbf{v}}_{i\circ} \right\|^2. \tag{S.13}$$

Term $E\left\| \bar{\mathbf{v}}_{i\ell} \right\|^2$ can be written as

$$E\left\| \bar{\mathbf{v}}_{i\ell} \right\|^2 = E\left(\bar{\mathbf{v}}_{i\ell}'\bar{\mathbf{v}}_{i\ell}\right) = E\left[tr\left(\bar{\mathbf{v}}_{i\ell}\bar{\mathbf{v}}_{i\ell}'\right)\right] = tr\left[E\left(\bar{\mathbf{v}}_{i\ell}\bar{\mathbf{v}}_{i\ell}'\right)\right],$$

where, under Assumptions 1-2 process $\mathbf{v}_{it} = \mathbf{C}_i^*(L)\mathbf{u}_{it}$ is covariance-stationary with uniformly (in i) absolutely summable autocovariances, and

$$E\left(\bar{\mathbf{v}}_{i\ell}\bar{\mathbf{v}}_{i\ell}'\right) = \frac{1}{T_q}\left\{\Gamma_i(0) + \sum_{h=1}^{T_q-1}\left(1 - \frac{h}{T_q}\right)\left[\Gamma_i(h) + \Gamma_i'(h)\right]\right\}, \tag{S.14}$$

in which $\mathbf{\Gamma}_i(h) = E\left(\mathbf{v}_{it}\mathbf{v}_{i,t-h}'\right)$. Using Lemma 1, it follows $\sup_{i,\ell} E\left\| \bar{\mathbf{v}}_{i\ell} \right\|^2 = O\left(T_q^{-1}\right)$, and given that q is fixed (does not change with the sample size), then $T_q \approx T$, and we in turn obtain

$$\sup_{i,\ell} E\left\|\bar{\mathbf{v}}_{i\ell}\right\|^2 = O\left(T^{-1}\right). \tag{S.15}$$

Similarly, $E\left\|\bar{\mathbf{v}}_{i\circ}\right\|^2 = tr\left[E\left(\bar{\mathbf{v}}_{i\circ}\bar{\mathbf{v}}_{i\circ}'\right)\right]$, where

$$E\left(\bar{\mathbf{v}}_{i\circ}\bar{\mathbf{v}}_{i\circ}'\right) = \frac{1}{T}\left\{\Gamma_i(0) + \sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\left[\Gamma_i(h) + \Gamma_i'(h)\right]\right\}, \tag{S.16}$$

and using Lemma 1, we obtain

$$\sup_{i} E\left\|\bar{\mathbf{v}}_{i\circ}\right\|^2 = O\left(T^{-1}\right). \tag{S.17}$$

Using (S.15) and (S.17) in (S.13) yields result (S.9). Consider now (S.10) and note that

$$T\ E\left(\mathbf{Q}_{\bar{v}_i\bar{v}_i}\right) = q^{-1}\sum_{\ell=1}^{q} E\left(\bar{\mathbf{v}}_{i\ell}\bar{\mathbf{v}}_{i\ell}'\right) - E\left(\bar{\mathbf{v}}_{i\circ}\bar{\mathbf{v}}_{i\circ}'\right),$$

and upon using (S.14) we have (recall that $T_q = T/q$)

$$q^{-1}\sum_{\ell=1}^{q} E\left(\bar{\mathbf{v}}_{i\ell}\bar{\mathbf{v}}_{i\ell}'\right) = \frac{1}{T}\left\{\Gamma_i(0) + \sum_{h=1}^{T_q-1}\left(1-\frac{h}{T_q}\right)\left[\Gamma_i(h) + \Gamma_i'(h)\right]\right\}.$$

Also using (S.16) we now have

$$\begin{aligned}
T\ E\left(\mathbf{Q}_{\bar{v}_i\bar{v}_i}\right) &= \frac{1}{T}\sum_{h=1}^{T_q-1}\left(1-\frac{h}{T_q}\right)\left[\Gamma_i(h) + \Gamma_i'(h)\right] - \frac{1}{T}\sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\left[\Gamma_i(h) + \Gamma_i'(h)\right], \\
&= \frac{1}{T}\sum_{h=1}^{T/q-1}\left[\left(1-\frac{qh}{T}\right) - \left(1-\frac{h}{T}\right)\right]\left[\Gamma_i(h) + \Gamma_i'(h)\right] \\
&\quad -\frac{1}{T}\sum_{h=T/q+1}^{T-1}\left(1-\frac{h}{T}\right)\left[\Gamma_i(h) + \Gamma_i'(h)\right].
\end{aligned}$$

Hence, as required we have

$$T^2\ E\left(\mathbf{Q}_{\bar{v}_i\bar{v}_i}\right) = \frac{(q-1)}{T}\sum_{h=1}^{T/q-1} h\left[\Gamma_i(h) + \Gamma_i'(h)\right] - \sum_{h=T/q+1}^{T-1}\left(1-\frac{h}{T}\right)\left[\Gamma_i(h) + \Gamma_i'(h)\right],$$

and

$$T^2 \ \left\| E\left(\mathbf{Q}_{\bar{v}_i\bar{v}_i}\right)\right\| \leq \frac{2\left(q-1\right)}{T}\sum_{h=1}^{T/q-1} h\left\|\Gamma_i(h)\right\| + \sum_{h=T/q+1}^{T-1}\left\|\Gamma_i(h)\right\|.$$

Since $\sup_i \left\|\mathbf{\Gamma}_i(h)\right\| < K_1\rho^h$, (see (S.1) in Lemma 1) it now follows that

$$T^2 \ \sup_i \left\| E\left(\mathbf{Q}_{\bar{v}_i\bar{v}_i}\right)\right\| \leq \frac{2\left(q-1\right)}{T} K_1 \sum_{h=1}^{T/q-1} h\rho^h + K_1 \sum_{h=T/q+1}^{T-1}\rho^h. \tag{S.18}$$

Also

$$\sum_{h=1}^{T/q-1} h\rho^h = \rho\left(\sum_{h=1}^{T/q-1} h\rho^{h-1}\right) = \rho\frac{d\left(\sum_{h=1}^{T/q-1}\rho^h\right)}{d\rho} = \rho\frac{d}{d\rho}\left(\frac{\rho-\rho^{T/q}}{1-\rho}\right) < K,$$

and $\sum_{h=T/q+1}^{T-1}\rho^h = \left(\rho^{T/q+1}-\rho^T\right)/\left(1-\rho\right)$. Using these results in (S.18) we have T^2 $\sup_i \left\| E\left(\mathbf{Q}_{\bar{v}_i\bar{v}_i}\right)\right\| = O\left(T^{-1}\right) + O\left(\rho^{T/q}\right)$, as required. We establish (S.11) next.

$$T\mathbf{Q}_{\bar{s}_i\bar{v}_i} = q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{s}}_{i\ell}-\bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell}-\bar{\mathbf{v}}_{i\circ}\right)' = q^{-1}\sum_{\ell=1}^{q}\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}_{i\ell}' - \bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\circ}',$$

where $\bar{\mathbf{v}}_{i\circ} = q^{-1}\sum_{\ell=1}^{q}\bar{\mathbf{v}}_{i\ell}$ and similarly $\bar{\mathbf{s}}_{i\circ} = q^{-1}\sum_{\ell=1}^{q}\bar{\mathbf{s}}_{i\ell}$. Taking spectral norm yields $\left\|T\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right\| \leq q^{-1}\sum_{\ell=1}^{q}\left\|\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}_{i\ell}'\right\| + \left\|\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\circ}'\right\|$. By Cauchy-Schwarz inequality,

$$E\left\|T\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right\| \leq q^{-1}\sum_{\ell=1}^{q}\left(E\left\|\bar{\mathbf{s}}_{i\ell}\right\|^2\right)^{1/2}\left(E\left\|\bar{\mathbf{v}}_{i\ell}\right\|^2\right)^{1/2} + \left(E\left\|\bar{\mathbf{s}}_{i\circ}\right\|^2\right)^{1/2}\left(E\left\|\bar{\mathbf{v}}_{i\circ}\right\|^2\right)^{1/2}. \tag{S.19}$$

Assumption 1 stipulates $\sup_i \left\|\mathbf{\Sigma}_i\right\| < K$. Using this bound in (S.6) and (S.7) yields

$$\sup_i E\left\|\bar{\mathbf{s}}_{i\circ}\right\|^2 = \sup_i tr\left[E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{s}}_{i\circ}'\right)\right] = O\left(T\right), \text{ and similarly } \sup_{i,\ell} E\left\|\bar{\mathbf{s}}_{i\ell}\right\|^2 = O\left(T\right). \tag{S.20}$$

In addition, we have already established $\sup_{i,\ell} E\left\|\bar{\mathbf{v}}_{i\ell}\right\|^2 = O\left(T^{-1}\right)$ (see (S.15)) and $\sup_i E\left\|\bar{\mathbf{v}}_{i\circ}\right\|^2 = O\left(T^{-1}\right)$ (see (S.17)). Using these results in (S.19) yields $\sup_i E\left\|T\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right\| = O\left(1\right)$, which in turn implies result (S.11).

We establish (S.12) next. Taking expectation of $\mathbf{Q}_{\bar{s}_i\bar{v}_i}$, we have

$$E\left(\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)=T^{-1}q^{-1}\sum_{\ell=1}^{q}\left[E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}_{i\ell}'\right)-E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}_{i\circ}'\right)-E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\ell}'\right)+E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\circ}'\right)\right], \tag{S.21}$$

where $\bar{\mathbf{s}}_{i\ell}=\sum_{t=1}^{(\ell-1)T_q}\mathbf{u}_{it}+T_q^{-1}\sum_{t=(\ell-1)T_q+1}^{\ell T_q}\left(\ell T_q+1-t\right)\mathbf{u}_{it}$, $\bar{\mathbf{s}}_{i\circ}=T^{-1}\sum_{t=1}^{T}\left(T-t+1\right)\mathbf{u}_{it}$, $\bar{\mathbf{v}}_{i\ell}=T_q^{-1}\sum_{t=(\ell-1)T_q+1}^{\ell T_q}\mathbf{v}_{it}$, $\bar{\mathbf{v}}_{i\circ}=T^{-1}\sum_{t=1}^{T}\mathbf{v}_{it}$, and $\mathbf{v}_{it}=\mathbf{C}_i^*(L)\mathbf{u}_{it}$ with $\mathbf{C}_i^*(L)=\sum_{j=0}^{\infty}\mathbf{C}_{ij}^*L^j$. We consider each of the four terms inside the summation operator on the right side of (S.21) in turn. For $E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}_{i\ell}'\right)$, we have

$$E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}_{i\ell}'\right)=\frac{1}{T_q}\sum_{t=1}^{(\ell-1)T_q}\sum_{t'=(\ell-1)T_q+1}^{\ell T_q}E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right)+\frac{1}{T_q^2}\sum_{t=(\ell-1)T_q+1}^{\ell T_q}\sum_{t'=(\ell-1)T_q+1}^{\ell T_q}\left(\ell T_q+1-t\right)E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right). \tag{S.22}$$

Noting that

$$E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right)=\begin{cases}\boldsymbol{\Sigma}_i\mathbf{C}_{i,t'-t}^{*\prime}, \text{ for } t\leq t',\\ \mathbf{0}, \text{ for } t>t',\end{cases} \tag{S.23}$$

we obtain

$$\frac{1}{T_q}\sum_{t=1}^{(\ell-1)T_q}\sum_{t'=(\ell-1)T_q+1}^{\ell T_q}E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right)=\frac{1}{T_q}\sum_{t=1}^{(\ell-1)T_q}\sum_{t'=(\ell-1)T_q+1}^{\ell T_q}\boldsymbol{\Sigma}_i\mathbf{C}_{i,t'-t}^{*\prime}.$$

Recalling that under Assumptions 1 and 2, $\sup_i\left\|\boldsymbol{\Sigma}_i\right\|<K$ and $\sup_i\left\|\mathbf{C}_{ij}^{*\prime}\right\|<K\rho^j$, for $\rho<1$, then it follows that

$$\begin{aligned}\sup_i\left\|\frac{1}{T_q}\sum_{t=1}^{(\ell-1)T_q}\sum_{t'=(\ell-1)T_q+1}^{\ell T_q}E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right)\right\| &\leq \frac{1}{T_q}\sum_{t=1}^{(\ell-1)T_q}\sum_{t'=(\ell-1)T_q+1}^{\ell T_q}\sup_i\left\|\boldsymbol{\Sigma}_i\right\|\sup_i\left\|\mathbf{C}_{i,t'-t}^{*\prime}\right\|,\\ &\leq \frac{1}{T_q}\sum_{t=1}^{(\ell-1)T_q}\sum_{t'=(\ell-1)T_q+1}^{\ell T_q}K\rho^{t'-t},\\ &\leq \frac{K}{T_q}\sum_{t=1}^{(\ell-1)T_q}\rho^{(\ell-1)T_q-t+1}\sum_{j=0}^{T_q}\rho^j.\end{aligned}$$

But both sums $\sum_{j=0}^{T_q} \rho^j$ and $\sum_{t=1}^{(\ell-1)T_q} \rho^{(\ell-1)T_q-t+1}$ are bounded in T (recall $T_q = T/q$ where q does not change with sample size and is fixed). Hence

$$\sup_i \left\| \frac{1}{T_q} \sum_{t=1}^{(\ell-1)T_q} \sum_{t'=(\ell-1)T_q+1}^{\ell T_q} E\left(\mathbf{u}_{it}\mathbf{v}'_{it'}\right) \right\| = O\left(T^{-1}\right). \tag{S.24}$$

For the second term on the right side of (S.22), we have

$$\begin{aligned} &\frac{1}{T_q^2} \sum_{t=(\ell-1)T_q+1}^{\ell T_q} \sum_{t'=(\ell-1)T_q+1}^{\ell T_q} \left(\ell T_q + 1 - t\right) E\left(\mathbf{u}_{it}\mathbf{v}'_{it'}\right) = \\ = \;& \frac{1}{T_q^2} \sum_{t=(\ell-1)T_q+1}^{\ell T_q} \left(\ell T_q + 1 - t\right) \left[\sum_{t'=(\ell-1)T_q+1}^{t-1} E\left(\mathbf{u}_{it}\mathbf{v}'_{it'}\right) + \sum_{t'=t}^{\ell T_q} E\left(\mathbf{u}_{it}\mathbf{v}'_{it'}\right) \right], \end{aligned}$$

and substituting (S.23), we obtain

$$\begin{aligned} &\frac{1}{T_q^2} \sum_{t=(\ell-1)T_q+1}^{\ell T_q} \sum_{t'=(\ell-1)T_q+1}^{\ell T_q} \left(\ell T_q + 1 - t\right) E\left(\mathbf{u}_{it}\mathbf{v}'_{it'}\right) \\ = \;& \frac{1}{T_q^2} \sum_{t=(\ell-1)T_q+1}^{\ell T_q} \left(\ell T_q + 1 - t\right) \sum_{t'=t}^{\ell T_q} \mathbf{\Sigma}_i \mathbf{C}^{*\prime}_{i,t'-t} = \frac{1}{T_q^2} \sum_{k=0}^{T_q-1} \left(\mathbf{\Sigma}_i \mathbf{C}^{*\prime}_{ik} \sum_{j=k+1}^{T_q} j \right), \\ = \;& \frac{1}{T_q^2} \sum_{k=0}^{T_q-1} \frac{\left(T_q + k + 1\right)\left(T_q - k\right)}{2} \mathbf{\Sigma}_i \mathbf{C}^{*\prime}_{ik}. \end{aligned} \tag{S.25}$$

Using (S.24) and (S.25) in (S.22) yields

$$\sup_i \left\| E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}'_{i\ell}\right) - \frac{1}{T_q^2} \sum_{k=0}^{T_q-1} \frac{\left(T_q + k + 1\right)\left(T_q - k\right)}{2} \mathbf{\Sigma}_i \mathbf{C}^{*\prime}_{ik} \right\| = O\left(T^{-1}\right). \tag{S.26}$$

Next consider $E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}'_{i\circ}\right)$, and note that

$$E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}'_{i\circ}\right) = \frac{1}{T} \sum_{t=1}^{(\ell-1)T_q} \sum_{t'=1}^{T} E\left(\mathbf{u}_{it}\mathbf{v}'_{it'}\right) + \frac{1}{T_q T} \sum_{t=(\ell-1)T_q+1}^{\ell T_q} \sum_{t'=1}^{T} \left(\ell T_q + 1 - t\right) E\left(\mathbf{u}_{it}\mathbf{v}'_{it'}\right). \tag{S.27}$$

For the first term on the right side of (S.27), we obtain

$$\frac{1}{T}\sum_{t=1}^{(\ell-1)T_q}\sum_{t'=1}^{T}E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right)=\frac{1}{T}\sum_{t=1}^{(\ell-1)T_q}\left[\sum_{t'=1}^{t-1}E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right)+\sum_{t'=t}^{T}E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right)\right]=\frac{1}{T}\sum_{t=1}^{(\ell-1)T_q}\sum_{t'=t}^{T}\mathbf{\Sigma}_i\mathbf{C}_{i,t'-t}^{*\prime},$$

and using $k=t'-t$, we have

$$\frac{1}{T}\sum_{t=1}^{(\ell-1)T_q}\sum_{t'=1}^{T}E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right)=\frac{1}{T}\sum_{t=1}^{(\ell-1)T_q}\sum_{k=0}^{T-t}\mathbf{\Sigma}_i\mathbf{C}_{ik}^{*\prime}=\frac{1}{T}\sum_{t=1}^{(\ell-1)T_q}\sum_{k=0}^{T_q-1}\mathbf{\Sigma}_i\mathbf{C}_{ik}^{*\prime}+\frac{1}{T}\sum_{t=1}^{(\ell-1)T_q}\sum_{k=T_q-1}^{T-t}\mathbf{\Sigma}_i\mathbf{C}_{ik}^{*\prime},$$

where (noting that $T=qT_q$) $T^{-1}\sum_{t=1}^{(\ell-1)T_q}\sum_{k=0}^{T_q-1}\mathbf{\Sigma}_i\mathbf{C}_{ik}^{*\prime}=q^{-1}\left(\ell-1\right)\sum_{k=0}^{T_q-1}\mathbf{\Sigma}_i\mathbf{C}_{ik}^{*\prime}$, and (noting that $\left\|\mathbf{C}_{ij}^{*\prime}\right\|<K\rho^j$ and $\left\|\mathbf{\Sigma}_i\right\|<K$)

$$\begin{aligned}\sup_i\left\|\frac{1}{T}\sum_{t=1}^{(\ell-1)T_q}\sum_{k=T_q-1}^{T-t}\mathbf{\Sigma}_i\mathbf{C}_{ik}^{*\prime}\right\| &\leq \frac{1}{T}\sum_{t=1}^{(\ell-1)T_q}\sum_{k=T_q}^{T-t}\sup_i\left\|\mathbf{\Sigma}_i\right\|\sup_i\left\|\mathbf{C}_{ik}^{*\prime}\right\|<\\ \frac{K}{T}\sum_{t=1}^{(\ell-1)T_q}\sum_{k=T_q}^{T-t}\rho^k &< K\frac{(\ell-1)T_q}{T}\rho^{T_q}\sum_{j=0}^{T-t-T_q}\rho^j<K_2\rho^{T_q},\end{aligned}$$

where $|\rho|<1$. Hence

$$\sup_i\left\|\frac{1}{T}\sum_{t=1}^{(\ell-1)T_q}\sum_{t'=1}^{T}E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right)-\frac{\ell-1}{q}\sum_{k=0}^{T_q-1}\mathbf{\Sigma}_i\mathbf{C}_{ik}^{*\prime}\right\|=O\left(\rho^{T_q}\right). \quad \text{(S.28)}$$

For the second term on the right side of (S.27), we have

$$\begin{aligned}&\frac{1}{T_qT}\sum_{t=(\ell-1)T_q+1}^{\ell T_q}\sum_{t'=1}^{T}\left(\ell T_q+1-t\right)E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right)=\\ &\frac{1}{T_qT}\sum_{t=(\ell-1)T_q+1}^{\ell T_q}\left[\sum_{t'=1}^{t-1}\left(\ell T_q+1-t\right)E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right)+\sum_{t'=t}^{T}\left(\ell T_q+1-t\right)E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right)\right],\\ =&\frac{1}{T_qT}\sum_{t=(\ell-1)T_q+1}^{\ell T_q}\sum_{t'=t}^{T}\left(\ell T_q+1-t\right)\mathbf{\Sigma}_i\mathbf{C}_{i,t'-t}^{*\prime}.\end{aligned}$$

Using $k = t' - t$ and $j = t - (\ell - 1)T_q$, we obtain

$$\frac{1}{T_q T} \sum_{t=(\ell-1)T_q+1}^{\ell T_q} \sum_{t'=t}^{T} (\ell T_q + 1 - t)\, \mathbf{\Sigma}_i \mathbf{C}^{*\prime}_{i,t'-t} = \frac{1}{T_q T} \sum_{j=1}^{T_q} \sum_{k=0}^{T-(\ell-1)T_q-j} (T_q + 1 - j)\, \mathbf{\Sigma}_i \mathbf{C}^{*\prime}_{ik}.$$

Consider the case $\ell < q$ and $\ell = q$ in turn. For $\ell < q$, we can write

$$\sum_{k=0}^{T-(\ell-1)T_q-j} \mathbf{C}^{*\prime}_{ik} = \sum_{k=0}^{T_q-1} \mathbf{C}^{*\prime}_{ik} + \sum_{k=T_q}^{T-(\ell-1)T_q-j} \mathbf{C}^{*\prime}_{ik},$$

where

$$\sup_i \left\| \sum_{k=T_q}^{T-(\ell-1)T_q-j} \mathbf{C}^{*\prime}_{ik} \right\| \leq \sum_{k=T_q}^{T-(\ell-1)T_q-j} \sup_i \|\mathbf{C}^{*\prime}_{ik}\| < K \sum_{k=T_q}^{T-(\ell-1)T_q-j} \rho^k < K_3 \rho^{T_q},$$

since $|\rho| < 1$. Hence, for $\ell < q$, we have

$$\begin{aligned}
\frac{1}{T_q T} \sum_{j=1}^{T_q} \sum_{k=0}^{T-(\ell-1)T_q-j} (T_q + 1 - j)\, \mathbf{\Sigma}_i \mathbf{C}^{*\prime}_{ik} &= \frac{1}{T_q T} \mathbf{\Sigma}_i \sum_{j=1}^{T_q} \left[(T_q + 1 - j) \left(\sum_{k=0}^{T_q-1} \mathbf{C}^{*\prime}_{ik} + O\left(\rho^{T_q}\right) \right) \right], \\
&= \frac{1}{T_q T} \mathbf{\Sigma}_i \sum_{j=1}^{T_q} \sum_{k=0}^{T_q-1} (T_q + 1 - j)\, \mathbf{C}^{*\prime}_{ik} + O\left(\rho^{T_q}\right), \\
&= \frac{1}{T_q T} \frac{(T_q + 1)\, T_q}{2} \mathbf{\Sigma}_i \sum_{k=0}^{T_q-1} \mathbf{C}^{*\prime}_{ik} + O\left(\rho^{T_q}\right).
\end{aligned}$$

where the $O\left(\rho^{T_q}\right)$ term is uniform in i. Noting that

$$\frac{1}{T_q T} \frac{(T_q + 1)\, T_q}{2} = \frac{1}{2q} + O\left(T^{-1}\right),$$

and using the fact that upper bound $O\left(\rho^{T_q}\right)$ is dominated by $O\left(T^{-1}\right)$, we obtain (for $\ell < q$)

$$\frac{1}{T_q T} \sum_{t=(\ell-1)T_q+1}^{\ell T_q} \sum_{t'=t}^{T} (\ell T_q + 1 - t)\, \mathbf{\Sigma}_i \mathbf{C}^{*\prime}_{i,t'-t} = \frac{1}{2q} \sum_{k=0}^{T_q-1} \mathbf{\Sigma}_i \mathbf{C}^{*\prime}_{ik} + O\left(T^{-1}\right),$$

where the $O\left(T^{-1}\right)$ term is uniform in i. For $\ell = q$, we have

$$\begin{aligned}\frac{1}{T_q T}\sum_{j=1}^{T_q}\sum_{k=0}^{T-H_{q-1}-j}\left(T_q+1-j\right)\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime} &= \frac{1}{T_q T}\sum_{j=1}^{T_q}\sum_{k=0}^{T_q-j}\left(T_q+1-j\right)\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime}, \\ &= \frac{1}{T_q T}\sum_{k=0}^{T_q-1}\frac{\left(T_q+k+1\right)\left(T_q-k\right)}{2}\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime}.\end{aligned}$$

But

$$\frac{1}{T_q T}\sum_{k=0}^{T_q-1}\frac{\left(T_q+k+1\right)\left(T_q-k\right)}{2}\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime} = \frac{1}{T_q T}\sum_{k=0}^{T_q-1}\frac{\left(T_q+1\right)T_q}{2}\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime} - \frac{1}{T_q T}\sum_{k=0}^{T_q-1}\frac{\left(k+1\right)k}{2}\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime},$$

where the second term can be bounded by

$$\sup_i\left\|\frac{1}{T_q T}\sum_{k=1}^{T_q-j}\frac{\left(k+1\right)k}{2}\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime}\right\| < \frac{K}{T_q T}\sum_{k=1}^{T_q-j}\frac{\left(k+1\right)k}{2}\rho^k = O\left(T^{-1}\right).$$

Hence regardless of $\ell = q$ or $\ell < q$, we have

$$\sup_i\left\|\frac{1}{T_q T}\sum_{t=(\ell-1)T_q+1}^{\ell T_q}\sum_{t'=t}^{T}\left(\ell T_q+1-t\right)\boldsymbol{\Sigma}_i\mathbf{C}_{i,t'-t}^{*\prime} - \frac{1}{2q}\sum_{k=0}^{T_q-1}\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime}\right\| = O\left(T^{-1}\right) \quad \text{(S.29)}$$

Using (S.29) and (S.28) in (S.27) and noting that $O\left(\rho^{T_q}\right)$ is dominated by $O\left(T^{-1}\right)$, we obtain

$$\sup_i\left\|E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}_{i\circ}\right) - \frac{\ell-1}{q}\sum_{k=0}^{T_q-1}\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime} - \frac{1}{2q}\sum_{k=0}^{T_q-1}\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime}\right\| = O\left(T^{-1}\right). \quad \text{(S.30)}$$

Consider the term $E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\ell}'\right)$ next.

$$
\begin{aligned}
E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\ell}'\right) &= \frac{1}{TT_q}\sum_{t=1}^{T}\sum_{t'=(\ell-1)T_q+1}^{\ell T_q}(T-t+1)\,E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right), \\
&= \frac{1}{TT_q}\sum_{t'=(\ell-1)T_q+1}^{\ell T_q}\left[\sum_{t=1}^{t'}(T-t+1)\,E\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right)+\sum_{t=t'+1}^{T}(T-t+1)\left(\mathbf{u}_{it}\mathbf{v}_{it'}'\right)\right], \\
&= \frac{1}{TT_q}\sum_{t'=(\ell-1)T_q+1}^{\ell T_q}\sum_{t=1}^{t'}(T-t+1)\,\boldsymbol{\Sigma}_i\mathbf{C}_{i,t'-t}^{*\prime}.
\end{aligned}
$$

Using $k=t'-t$ and $j=t'-H_{s-1}-1$, we obtain

$$
\begin{aligned}
E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\ell}'\right) &= \frac{1}{TT_q}\sum_{j=0}^{T_q-1}\sum_{k=0}^{j+(\ell-1)T_q}\left(T-(\ell-1)T_q-j+k\right)\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime} \qquad \text{(S.31)}\\
&= \frac{1}{TT_q}\sum_{j=0}^{T_q-1}\left(\begin{array}{c}\sum_{k=0}^{(\ell-1)T_q}\left(T-(\ell-1)T_q-j+k\right)\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime}\\ +\sum_{k=(\ell-1)T_q+1}^{j+(\ell-1)T_q}\left[T-(\ell-1)T_q-j+k\right]\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime}\end{array}\right).
\end{aligned}
$$

But

$$
\sup_i\left\|\sum_{k=(\ell-1)T_q+1}^{j+(\ell-1)T_q}\left(T-(\ell-1)T_q-j+k\right)\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime}\right\| \qquad \text{(S.32)}
$$

$$
\leq \sum_{k=(\ell-1)T_q+1}^{j+(\ell-1)T_q}\left(T-(\ell-1)T_q-j+k\right)\sup_i\left\|\boldsymbol{\Sigma}_i\right\|\sup_i\left\|\mathbf{C}_{ik}^{*\prime}\right\| < KT\sum_{k=(\ell-1)T_q+1}^{j+(\ell-1)T_q}\rho^k < K_1\rho_1^T,
$$

where $|\rho|<\rho_1<1$. In addition,

$$
\left\|\frac{1}{TT_q}\sum_{j=0}^{T_q-1}\sum_{k=0}^{(\ell-1)T_q}k\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime}\right\| \leq \frac{1}{TT_q}\sum_{j=0}^{T_q-1}\sum_{k=0}^{(\ell-1)T_q}k\left\|\boldsymbol{\Sigma}_i\right\|\left\|\mathbf{C}_{ik}^{*\prime}\right\| < \frac{K}{TT_q}\sum_{j=0}^{T_q-1}\sum_{k=0}^{(\ell-1)T_q}k\rho^k.
$$

Since $\sum_{k=0}^{(\ell-1)T_q}k\rho^k<K$, we obtain

$$
\sup_i\left\|\frac{1}{TT_q}\sum_{j=0}^{T_q-1}\sum_{k=0}^{(\ell-1)T_q}k\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime}\right\| = O\left(T^{-1}\right). \qquad \text{(S.33)}
$$

Using (S.32) and (S.33) in (S.31), it follows

$$\sup_i \left\| E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\ell}\right) - \frac{1}{TT_q}\sum_{j=0}^{T_q-1}\sum_{k=0}^{(\ell-1)T_q}\left(T-(\ell-1)T_q-j\right)\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime}\right\| = O\left(T^{-1}\right).$$

Given that

$$\frac{1}{TT_q}\sum_{j=0}^{T_q-1}\left(T-(\ell-1)T_q-j\right) = \frac{\left(2T-2(\ell-1)T_q-T_q+1\right)T_q}{2TT_q} = 1-\frac{\ell-1}{q}-\frac{1}{2q}+O\left(T^{-1}\right),$$

it further follows

$$\sup_i \left\| E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\ell}\right) - \left(1-\frac{\ell-1}{q}-\frac{1}{2q}\right)\sum_{k=0}^{(\ell-1)T_q}\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime}\right\| = O\left(T^{-1}\right). \tag{S.34}$$

For the purpose of this proof, it is convenient to decompose

$$\sum_{k=0}^{(\ell-1)T_q}\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime} = \sum_{k=0}^{T_q-1}\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime} + \sum_{k=T_q}^{(\ell-1)T_q}\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime},$$

where, as before,

$$\sup_i\left\|\sum_{k=T_q}^{(\ell-1)T_q}\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime}\right\| \leq \sum_{k=T_q}^{(\ell-1)T_q}\sup_i\left\|\boldsymbol{\Sigma}_i\right\|\sup_i\left\|\mathbf{C}_{ik}^{*\prime}\right\| < K\sum_{k=T_q}^{(\ell-1)T_q}\rho^k < K_1\rho^{T_q}.$$

Using this result in (S.34) yields

$$\sup_i \left\| E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\ell}^{\prime}\right) - \left(1-\frac{\ell-1}{q}-\frac{1}{2q}\right)\sum_{k=0}^{T_q-1}\boldsymbol{\Sigma}_i\mathbf{C}_{ik}^{*\prime}\right\| = O\left(T^{-1}\right). \tag{S.35}$$

Finally for the last term, $E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}'_{i\circ}\right)$, we have

$$
\begin{aligned}
E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}'_{i\circ}\right) &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}(T-t+1)\,E\left(\mathbf{u}_{it}\mathbf{v}'_{it'}\right), \\
&= \frac{1}{T^2}\sum_{t=1}^{T}(T-t+1)\left[\sum_{t'=1}^{t-1}E\left(\mathbf{u}_{it}\mathbf{v}'_{it'}\right)+\sum_{t'=t}^{T}E\left(\mathbf{u}_{it}\mathbf{v}'_{it'}\right)\right] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}(T-t+1)\sum_{t'=t}^{T}\mathbf{\Sigma}_i\mathbf{C}^{*\prime}_{i,t'-t}, \\
&= \frac{1}{T^2}\sum_{k=0}^{T-1}\left(\mathbf{\Sigma}_i\mathbf{C}^{*\prime}_{ik}\sum_{j=k+1}^{T}j\right), \\
&= \frac{1}{T^2}\sum_{k=0}^{T-1}\frac{(T+k+1)(T-k)}{2}\mathbf{\Sigma}_i\mathbf{C}^{*\prime}_{ik}. \qquad \text{(S.36)}
\end{aligned}
$$

It is further useful to split the sum above into first T_q terms and the remainder, which can be appropriately bounded, namely

$$
\begin{aligned}
&\frac{1}{T^2}\sum_{k=0}^{T-1}\frac{(T+k+1)(T-k)}{2}\mathbf{\Sigma}_i\mathbf{C}^{*\prime}_{ik} \\
=\ &\frac{1}{T^2}\sum_{k=0}^{T_q-1}\frac{(T+k+1)(T-k)}{2}\mathbf{\Sigma}_i\mathbf{C}^{*\prime}_{ik}+\frac{1}{T^2}\sum_{k=T_q}^{T-1}\frac{(T+k+1)(T-k)}{2}\mathbf{\Sigma}_i\mathbf{C}^{*\prime}_{ik},
\end{aligned}
$$

where (noting that $(T+k+1)(T-k)/2T^2 = O(1)$)

$$
\begin{aligned}
\sup_i\left\|\frac{1}{T^2}\sum_{k=T_q}^{T-1}\frac{(T+k+1)(T-k)}{2}\mathbf{\Sigma}_i\mathbf{C}^{*\prime}_{ik}\right\| &\leq \sum_{k=T_q}^{T-1}\frac{(T+k+1)(T-k)}{2T^2}\sup_i\left\|\mathbf{\Sigma}_i\right\|\sup_i\left\|\mathbf{C}^{*\prime}_{ik}\right\|, \\
&< K\sum_{k=T_q}^{T-1}\rho^k < K_1\rho^{T_q}.
\end{aligned}
$$

Hence, we obtain for the last term $E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}'_{i\circ}\right)$,

$$
\sup_i\left\|E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\circ}\right)-\frac{1}{T^2}\sum_{k=0}^{T_q-1}\frac{(T+k+1)(T-k)}{2}\mathbf{\Sigma}_i\mathbf{C}^{*\prime}_{ik}\right\| = O\left(\rho^{T_q}\right). \qquad \text{(S.37)}
$$

Consider now $E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}_{i\ell}'\right)-E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}_{i\circ}'\right)-E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\ell}'\right)+E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\circ}'\right)$. Using (S.26), (S.30), (S.35), and (S.37), we obtain

$$\sup_i \left\|E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}_{i\ell}'\right)-E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}_{i\circ}'\right)-E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\ell}'\right)+E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\circ}'\right)\right\|_1 < \sup_i \left\|\sum_{k=0}^{T_q-1} a_k \mathbf{\Sigma}_i \mathbf{C}_{ik}^{*\prime}\right\| + \frac{K}{T},$$

where a_k, for $k=0,1,...,T_q-1$, are given by

$$\begin{aligned} a_k &= \frac{(T_q+k+1)(T_q-k)}{2T_q^2} - \left(\frac{\ell-1}{q}+\frac{1}{2q}\right) - \left(1-\frac{\ell-1}{q}-\frac{1}{2q}\right) + \frac{(T+k+1)(T-k)}{2T^2}, \\ &= \frac{T_q-k-k^2}{2T_q^2} + \frac{T-k-k^2}{2T^2} = k\left(\frac{1}{2kT_q}-\frac{1}{2T_q^2}-\frac{k}{2T_q^2}+\frac{1}{2Tk}-\frac{1}{2T^2}-\frac{k}{2T^2}\right), \\ &= \frac{k}{T}\left(\frac{q}{2k}-\frac{q}{2T_q}-\frac{kq}{2T_q}+\frac{1}{2k}-\frac{1}{2T}-\frac{k}{2T}\right). \end{aligned}$$

It is easily seen that $|a_k| < \left(\frac{k}{T}\right)K$, and

$$\sup_i \left\|\sum_{k=0}^{T_q-1} a_k \mathbf{\Sigma}_i \mathbf{C}_{ik}^{*\prime}\right\| < \sum_{k=0}^{T_q-1} |a_k| \sup_i \|\mathbf{\Sigma}_i\| \sup_i \|\mathbf{C}_{ik}^{*\prime}\| < \frac{K}{T}\sum_{k=0}^{T_q-1} k\rho^k = O\left(T^{-1}\right).$$

Hence

$$\sup_i \left\|E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}_{i\ell}'\right)-E\left(\bar{\mathbf{s}}_{i\ell}\bar{\mathbf{v}}_{i\circ}'\right)-E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\ell}'\right)+E\left(\bar{\mathbf{s}}_{i\circ}\bar{\mathbf{v}}_{i\circ}'\right)\right\| = O\left(T^{-1}\right),$$

and using this bound in (S.21) yields $\sup_i \left\|E\left(\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\right\|_1 = O\left(T^{-2}\right)$, as required. This completes the proof of (S.12).

■

Lemma 4 *Suppose Assumptions 1 and 2 hold, m and q (≥ 2) are fixed, and consider $\tilde{\boldsymbol{\xi}}_{iq} = \bar{\boldsymbol{\xi}}_{iq} - E\left(\bar{\boldsymbol{\xi}}_{iq}\right)$, where $\bar{\boldsymbol{\xi}}_{iq} = q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{v}}_{i\ell}\otimes\bar{\mathbf{s}}_{i\ell}\right) - \bar{\mathbf{v}}_{i\circ}\otimes\bar{\mathbf{s}}_{i\circ}$, $\bar{\mathbf{s}}_{i\ell}$ and $\bar{\mathbf{s}}_{i\circ}$ are, respectively, the sub-sample and full sample time averages of the partial sum process $\mathbf{s}_{it} = \sum_{\ell=1}^{t}\mathbf{u}_{it}$, and $\bar{\mathbf{v}}_{i\ell}$ and $\bar{\mathbf{v}}_{i\circ}$ are, respectively, the sub-sample and full sample time averages of $\mathbf{v}_{it} = \mathbf{C}_i^*(L)\mathbf{u}_{it} = \sum_{j=0}^{\infty}\mathbf{C}_{ij}^*\mathbf{u}_{i,t-j}$. Let $\kappa = (4+\epsilon)/2 > 2$, where $\epsilon > 0$ is the constant from Assumption 1. Then*

$$\sup_{i,k,T} E\left|\tilde{\xi}_{iqk}\right|^{\kappa} < K, \tag{S.38}$$

$$\sup_{i,\ell,T} E \left\| \sqrt{T} \left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ} \right) \right\|^{2\kappa} < K, \tag{S.39}$$

and

$$\sup_{i,\ell,T} \left\| T^{-1/2} \bar{\mathbf{s}}_{i\ell} \right\|^{2\kappa} < K, \tag{S.40}$$

where $\tilde{\xi}_{iqk}$ *is* k*-th element of* $\tilde{\boldsymbol{\xi}}_{iq}$*.*

Proof. Denote the individual elements of $\bar{\mathbf{v}}_{i\ell}$ and $\bar{\mathbf{s}}_{i\ell}$ by $\bar{v}_{i\ell k}$ and $\bar{s}_{i\ell k}$, for $k = 1, 2, ..., m$, respectively. These vectors depend on T_q but the subscript T_q is suppressed to simplify notations. Given that m and q are fixed, sufficient condition for (S.38) is

$$\sup_{i,\ell,\ell',k,k',T} E \left| \bar{v}_{i\ell k} \bar{s}_{i\ell' k'} \right|^{\kappa} < K, \tag{S.41}$$

where $\kappa = (4 + \epsilon)/2 > 2$ since $\epsilon > 0$ by Assumption 1. It is convenient to write $\bar{v}_{i\ell k} \bar{s}_{i\ell' k'}$ as $\left(T^{1/2} \bar{v}_{i\ell k} \right) \left(T^{-1/2} \bar{s}_{i\ell' k'} \right)$. Using Cauchy-Schwarz inequality, we obtain

$$E \left| \left(\sqrt{T} \bar{v}_{i\ell k} \right) \frac{\bar{s}_{i\ell' k'}}{\sqrt{T}} \right|^{\kappa} \leq \left(E \left| \sqrt{T} \bar{v}_{i\ell k} \right|^{2\kappa} \right)^{1/2} \left(E \left| \frac{\bar{s}_{i\ell' k'}}{\sqrt{T}} \right|^{2\kappa} \right)^{1/2}. \tag{S.42}$$

Consider $\sqrt{T} \bar{v}_{i\ell k}$ first, and note that it can be written as

$$\begin{aligned} \sqrt{T} \bar{v}_{i\ell k} &= \sqrt{T} T_q^{-1} \sum_{t=(\ell-1)T_q+1}^{\ell T_q} \mathbf{e}_k' \mathbf{v}_{it} = q T^{-1/2} \sum_{t=(\ell-1)T_q+1}^{\ell T_q} \sum_{h=0}^{\infty} \mathbf{e}_k' \mathbf{C}_{ih}^* \mathbf{u}_{i,t-h}, \\ &= q \sum_{h=0}^{\infty} \mathbf{e}_k' \mathbf{C}_{ih}^* \left(T^{-1/2} \sum_{t=(\ell-1)T_q+1}^{\ell T_q} \mathbf{u}_{i,t-h} \right) = q \sum_{h=0}^{\infty} \mathbf{e}_k' \mathbf{C}_{ih}^* \left(T^{-1/2} \boldsymbol{\vartheta}_{i\ell h,T} \right), \end{aligned} \tag{S.43}$$

where $\mathbf{e}_k$ is $m \times 1$ selection vector for k-th element, and $\boldsymbol{\vartheta}_{i\ell h,T} = \sum_{t=(\ell-1)T/q+1}^{\ell T/q} \mathbf{u}_{i,t-h}$ is sum of $T_q = T/q$ independent random vectors distributed with zero means. Under Assumption 1, $\sup_{it} E \left\| \mathbf{u}_{it} \right\|^{2\kappa} = \sup_{it} E \left\| \mathbf{u}_{it} \right\|^{4+\epsilon} < K$. Hence, using the result on moments of the sums of independent random variables given by equation (2) of Petrov

(1992),[16] we have (for a fixed q) $\sup_{i\ell hT} E\left\|\boldsymbol{\vartheta}_{i\ell h,T}\right\|^{2\kappa} = O\left(T^{\kappa}\right)$, and it follows

$$\sup_{i\ell hT} E\left\|T^{-1/2}\boldsymbol{\vartheta}_{i\ell h,T}\right\|^{2\kappa} < K, \tag{S.44}$$

Consider $E\left\|\sqrt{T}\bar{v}_{i\ell k}\right\|^{2\kappa}$ next. Using (S.43) and the Minkowski inequality, we obtain the following upper bound

$$\begin{aligned}\left(E\left\|\sqrt{T}\bar{v}_{i\ell k}\right\|^{2\kappa}\right)^{\frac{1}{2\kappa}} &\leq q\sum_{h=0}^{\infty}\left[E\left\|\mathbf{e}_k'\mathbf{C}_{ih}^{*}\left(T^{-1/2}\boldsymbol{\vartheta}_{i\ell h,T}\right)\right\|^{2\kappa}\right]^{\frac{1}{2\kappa}}, \\ &\leq q\sum_{h=0}^{\infty}\left\|\mathbf{C}_{ih}^{*}\right\|\left(E\left\|T^{-1/2}\boldsymbol{\vartheta}_{i\ell h,T}\right\|^{2\kappa}\right)^{\frac{1}{2\kappa}}.\end{aligned}$$

Using (S.44) and noting that under Assumption 2 $\sup_{ih}\sum_{h=0}^{\infty}\left\|\mathbf{C}_{ih}^{*}\right\| < K$, we obtain

$$\sup_{i\ell kT} E\left\|\sqrt{T}\bar{v}_{i\ell k}\right\|^{2\kappa} < K. \tag{S.45}$$

Existence of a uniform bound for $E\left|T^{-1/2}\bar{s}_{i\ell k}\right|^{2\kappa}$ is established next in a similar way. Using (S.5), $s_{i\ell k}$ is given by

$$s_{i\ell k} = \sum_{t=1}^{(\ell-1)T_q} u_{ikt} + \frac{1}{T_q}\sum_{t=(\ell-1)T_q+1}^{\ell T_q}\left(\ell T_q + 1 - t\right)u_{ikt} = \sum_{t=1}^{\ell T_q}\zeta_{i\ell kt,T},$$

where $\zeta_{i\ell kt,T} = a_{t\ell,T}u_{ikt}$

$$a_{t\ell,T} = \begin{cases} 1 & \text{for } t = 1, 2, ..., (\ell-1)T_q \\ T_q^{-1}\left(\ell T_q + 1 - t\right) & \text{for } t = (\ell-1)T_q + 1, ..., \ell T_q \end{cases},$$

and we have $\left|a_{t\ell,T}\right| \leq 1$. Hence $s_{i\ell k}$ is sum of independent random variables distributed with zero means. In addition, Assumption 1 and $\left|a_{t\ell,T}\right| \leq 1$ imply $\sup_{i\ell ktT} E\left|\zeta_{i\ell kt,T}\right|^{2\kappa}$. Using equation (2) of Petrov (1992) we obtain $\sup_{i\ell kT} E\left|\bar{s}_{i\ell k}\right|^{2\kappa} = O\left(T^{\kappa}\right)$, and it follows

$$\sup_{i\ell kT} E\left|\frac{\bar{s}_{i\ell k}}{\sqrt{T}}\right|^{2\kappa} < K. \tag{S.46}$$

[16] Let $X_1, X_2, ..., X_n$ be independent random variables and let $S_n = \sum_{\ell=1}^{n} X_\ell$. If $E\left(X_\ell\right) = 0$ for $\ell = 1, 2, ..., n$, and $p \geq 2$ then $E\left|S_n\right|^p \leq c\left(p\right)n^{p/2-1}\sum_{\ell=1}^{n} E\left|X_\ell\right|^p$, where $c\left(p\right) > 0$ depends only on p. See Petrov (1992).

Using (S.45) and (S.46) in (S.42) establishes (S.41). This completes the proof of (S.38). Consider (S.39) next. Given that m and q are fixed, sufficient condition for (S.39) to hold is (S.45), which was already established. Similarly, the last result of this lemma, (S.40), follows from (S.46). This completes the proof. ■

Lemma 5 *Consider* $\mathbf{Q}_{\bar{v}\bar{v}} = T^{-1}n^{-1}q^{-1}\sum_{i=1}^{n}\sum_{\ell=1}^{q}\left(\bar{\mathbf{v}}_{i\ell}-\bar{\mathbf{v}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell}-\bar{\mathbf{v}}_{i\circ}\right)'$ *and suppose Assumptions 1 and 2 hold,* $q\,(\geq 2)$ *is fixed and* m *is fixed. Let* $\kappa = (4+\epsilon)/2 > 2$*, where* $\epsilon > 0$ *is the constant from Assumption 1. Then for any* $m \times 1$ *vector* $\mathbf{b}$ *of unit length,* $\|\mathbf{b}\| = 1$*, we have*

$$E\left(\left|\mathbf{b}'\mathbf{Q}_{\bar{v}\bar{v}}\mathbf{b}\right|^{\kappa}\right) = O\left(n^{-\kappa/2}T^{-2\kappa}\right). \tag{S.47}$$

Proof. Denote (k,ℓ)-th element of the scaled $m \times m$ matrix $\sqrt{n}T^2\mathbf{Q}_{\bar{v}\bar{v}}$ as $\tilde{q}_{vv,k\ell}$, which can be written as

$$\tilde{q}_{vv,k\ell} = n^{-1/2}\sum_{i=1}^{n}\nu_{ik}\nu_{i\ell},$$

where ν_{ik} is k-th element of $q^{-1}\sum_{\ell=1}^{q}\sqrt{T}\left(\bar{\mathbf{v}}_{i\ell}-\bar{\mathbf{v}}_{i\circ}\right)$. Consider $\nu_{ik}\nu_{i\ell}$. Using Cauchy-Schwarz inequality,

$$E\left|\nu_{ik}\nu_{i\ell}\right|^{\kappa} \leq \left(E\left|\nu_{ik}\right|^{2\kappa}\right)^{1/2}\left(E\left|\nu_{i\ell}\right|^{2\kappa}\right)^{1/2},$$

and by result (S.39) of Lemma 4, we obtain $E\left|\nu_{ik}\nu_{i\ell}\right|^{\kappa}$ is uniformly bounded, namely

$$\sup_{k,\ell} E\left|\nu_{ik}\nu_{i\ell}\right|^{\kappa} < K.$$

Since $\nu_{ik}\nu_{i\ell}$ is independently distributed over i, we can use the result on moments of the sums of independent random variables given by equation (2) of Petrov (1992), and obtain $E\left|\tilde{q}_{vv,k\ell}\right|^{\kappa} = O(1)$. This implies $E\left(\left|\mathbf{b}'\left(\sqrt{n}T^2\mathbf{Q}_{\bar{v}\bar{v}}\right)\mathbf{b}\right|^{\kappa}\right) = O(1)$ for any $m \times 1$ vector $\mathbf{b}$ of unit length. Result (S.47) now readily follows. ■

Lemma 6 *Consider* $\mathbf{Q}_{\bar{w}\bar{w}} = T^{-1}n^{-1}q^{-1}\sum_{i=1}^{n}\sum_{\ell=1}^{q}\left(\bar{\mathbf{w}}_{i\ell}-\bar{\mathbf{w}}_{i\circ}\right)\left(\bar{\mathbf{w}}_{i\ell}-\bar{\mathbf{w}}_{i\circ}\right)'$ *and suppose the* $m \times 1$ *vector* $\mathbf{w}_{it}$ *is given by (1) without interactive time effects (*$\mathbf{G}_i = \mathbf{0}$*), Assumptions 1 to 3 hold,* $q\,(\geq 2)$ *is fixed and* m *is fixed. Then*

$$E\left|\boldsymbol{\beta}_{j0}'\mathbf{Q}_{\bar{w}\bar{w}}\boldsymbol{\beta}_{j0}\right| = O\left(n^{-1/2}T^{-2}\right), \text{ for } j = 1,2,...,r_0, \tag{S.48}$$

where $\boldsymbol{\beta}_{j0}$ *are eigenvectors of* $(6q)^{-1}(q-1)\boldsymbol{\Psi}_n$ *defined in Assumption 3.*

Proof. Premultiplying (13) by $\mathbf{B}_0'$, postmultiplying by $\mathbf{B}_0$, and noting $\mathbf{B}_0'\mathbf{C}_i = \mathbf{0}$ for $i = 1, 2, ..., n$ under Assumption 3, we obtain

$$\mathbf{B}_0'\mathbf{Q}_{\bar{w}\bar{w}}\mathbf{B}_0 = \mathbf{B}_0'\mathbf{Q}_{\bar{v}\bar{v}}\mathbf{B}_0,$$

where by the orthonormality requirement $\mathbf{B}_0'\mathbf{B}_0 = \mathbf{I}_{r_0}$. Using Lemma 5, there exists a positive constant K such that for any $j = 1, 2, ..., r$,

$$E\left(\boldsymbol{\beta}_{j0}\mathbf{Q}_{\bar{v}\bar{v}}\boldsymbol{\beta}_{j0}\right)^2 < Kn^{-1}T^{-4}.$$

Result (S.48) directly follows. ■

Lemma 7 *Consider* $\mathbf{Q}_{\bar{w}\bar{w}} = T^{-1}n^{-1}q^{-1}\sum_{i=1}^{n}\sum_{\ell=1}^{q}\left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}\right)\left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}\right)'$ *and the associated* $m \times r$ *matrix* $\hat{\mathbf{B}}_0$ *given by orthonormal eigenvectors of* $\mathbf{Q}_{\bar{w}\bar{w}}$ *corresponding to its* r_0 *smallest eigenvalues collected in the* $r_0 \times r_0$ *diagonal matrix* $\hat{\boldsymbol{\Lambda}}_0 = diag\left(\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_{r_0}\right)$*. Suppose the* $m \times 1$ *vector* $\mathbf{w}_{it}$ *is given by (1) without interactive time effects (*$\mathbf{G}_i = \mathbf{0}$*), Assumptions 1 to 3 hold,* m *and* $q\ (\geq 2)$ *are fixed integers. Then*

$$E\left\|\hat{\boldsymbol{\Lambda}}_0\right\| = O\left(n^{-1/2}T^{-2}\right). \tag{S.49}$$

$$\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0 = O_p\left(n^{-1/2}T^{-2}\right). \tag{S.50}$$

Proof. We have $\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0 = \hat{\mathbf{B}}_0\hat{\boldsymbol{\Lambda}}_0$, where $\hat{\boldsymbol{\Lambda}}_0 = diag\left(\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_{r_0}\right)$ is consisting of r_0 smallest eigenvalues of $\mathbf{Q}_{\bar{w}\bar{w}}$, denoted as $0 \leq \hat{\lambda}_1 \leq \hat{\lambda}_2 \leq ... \leq \hat{\lambda}_{r_0}$, and $\hat{\mathbf{B}}_0'\hat{\mathbf{B}}_0 = \mathbf{I}_r$. Noting $\mathbf{B}_0'\mathbf{B}_0 = \mathbf{I}_r$ then

$$0 \leq tr\left(\hat{\boldsymbol{\Lambda}}_0\right) = tr\left(\hat{\mathbf{B}}_0'\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0\right) \leq tr\left(\mathbf{B}_0'\mathbf{Q}_{\bar{w}\bar{w}}\mathbf{B}_0\right).$$

But result (S.48) of Lemma 6 implies $E\left[tr\left(\mathbf{B}_0'\mathbf{Q}_{\bar{w}\bar{w}}\mathbf{B}_0\right)\right] = O\left(n^{-1/2}T^{-2}\right)$ and it follows that $E\left\|\hat{\boldsymbol{\Lambda}}_0\right\| = O\left(n^{-1/2}T^{-2}\right)$, as required. Taking norm of $\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0 = \hat{\mathbf{B}}_0\hat{\boldsymbol{\Lambda}}_0$, and noting that $\left\|\hat{\mathbf{B}}_0\right\| = 1$, we obtain $\left\|\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0\right\| \leq \left\|\hat{\boldsymbol{\Lambda}}_0\right\|$. Taking expectations and using (S.49) yields $E\left\|\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0\right\| = O\left(n^{-1/2}T^{-2}\right)$, and result (S.50) follows. ■

Lemma 8 *Suppose Assumptions 1 to 3 hold,* $q\,(\geq 2)$ *is fixed and* m *is fixed. Consider* $\bar{\boldsymbol{\xi}}_{iq} = q^{-1}\sum_{\ell=1}^{q}\left(\bar{\mathbf{v}}_{i\ell}\otimes\bar{\mathbf{s}}_{i\ell}\right) - \bar{\mathbf{v}}_{i\circ}\otimes\bar{\mathbf{s}}_{i\circ}$*, where* $\bar{\mathbf{s}}_{i\ell}$*,* $\bar{\mathbf{s}}_{i\circ}$ *are the sub-sample and full sample* *time averages of the partial sum process* $\mathbf{s}_{it} = \sum_{\ell=1}^{t} \mathbf{u}_{it}$*, and* $\bar{\mathbf{v}}_{i\ell}, \bar{\mathbf{v}}_{i\circ}$ *are the sub-sample and full sample time averages of* $\mathbf{v}_{it} = \mathbf{C}_i^*(L)\mathbf{u}_{it} = \sum_{j=0}^{\infty} \mathbf{C}_{ij}^* \mathbf{u}_{i,t-j}$*. Let* $\boldsymbol{\Omega}_{\xi_{qi}} = Var\left(\bar{\boldsymbol{\xi}}_{iq}\right)$ *and suppose* $\boldsymbol{\Omega}_q = \lim_{n,T\to\infty}\left(\mathring{\mathbf{B}}_0' \otimes \mathbf{C}_i\right) \boldsymbol{\Omega}_{\xi_{qi}} \left(\mathring{\mathbf{B}}_0 \otimes \mathbf{C}_i'\right)$ *is positive definite. Then*

$$n^{-1/2} \sum_{i=1}^{n} \left(\mathring{\mathbf{B}}_0' \otimes \mathbf{C}_i\right) \tilde{\boldsymbol{\xi}}_{iq} \to_d N\left(\mathbf{0}, \boldsymbol{\Omega}_q\right), \tag{S.51}$$

for $n, T \to \infty$ *jointly (in no particular order), where* $\tilde{\boldsymbol{\xi}}_{iq} = \bar{\boldsymbol{\xi}}_{iq} - E\left(\bar{\boldsymbol{\xi}}_{iq}\right)$*.*

Proof. Let $n_T = n(T)$ be any non-decreasing integer-valued function of T such that $\lim_{T\to\infty} n_T = \infty$, and let $\boldsymbol{\omega}$ be any $m^2 \times 1$ vector of unit length, namely $\|\boldsymbol{\omega}\| = 1$. Define the following triangular array

$$a_{n_T,i} = \frac{1}{\sqrt{n_T}} \boldsymbol{\omega}' \boldsymbol{\Omega}_q^{-1/2} \left(\mathring{\mathbf{B}}_0' \otimes \mathbf{C}_i\right) \left[\bar{\boldsymbol{\xi}}_{iq} - E\left(\bar{\boldsymbol{\xi}}_{iq}\right)\right],$$

where $m^2 \times 1$ vector $\bar{\boldsymbol{\xi}}_{iq}$ depends on T, but this subscript is suppressed for notational simplicity. By construction array $\{a_{n_T,i}\}$ is zero mean. By assumption 1, $\mathbf{u}_{it}$ is independently distributed over i and t, and therefore $a_{n_T,i}$ is independent of $a_{n_T,j}$ for $i \neq j$. In addition,

$$\sum_{i=1}^{n_T} Var\left(a_{n_T,i}\right) = 1.$$

We establish next the Liapunov's condition, given by

$$\lim_{n\to\infty} \sum_{i=1}^{n_T} E\left|a_{n_T,i}\right|^{2+\epsilon} = 0 \text{ for some } \epsilon > 0, \tag{S.52}$$

is met. Since $\boldsymbol{\Omega}_{\xi_q}$ is positive definite then $\left\|\boldsymbol{\omega}' \boldsymbol{\Omega}_{\xi_q}^{-1/2} \left(\mathring{\mathbf{B}}_0' \otimes \mathbf{C}_i\right)\right\| < K$, and (S.52) will hold if

$$\sup_{i,k,T} E\left|\tilde{\xi}_{iqk}\right|^{2+\epsilon} < K \text{ for some } \epsilon > 0, \tag{S.53}$$

where $\tilde{\xi}_{iqk}$ is k-th element of $\tilde{\boldsymbol{\xi}}_{iq}$. Lemma 4 establish (S.53) holds, and therefore (S.52) is met, as required. Then, by Liapunov's theorem (Theorem 23.11 in Davidson (1994)), the Lindeberg condition is also met, and (S.51) follows by Theorem 23.6 in Davidson (1994). ■

Lemma 9 *Consider* $m \times m$ *matrices* $\mathbf{Q}_{\bar{s}_i \bar{v}_i} = T^{-1} q^{-1} \sum_{\ell=1}^{q} \left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)'$*, where* $\bar{\mathbf{s}}_{i\ell}$*,* $\bar{\mathbf{s}}_{i\circ}$ *are the sub-sample and full sample time averages of the partial sum* *process* $\mathbf{s}_{it} = \sum_{\ell=1}^{t} \mathbf{u}_{it}$*, and* $\bar{\mathbf{v}}_{i\ell}$*,*$\bar{\mathbf{v}}_{i\circ}$ *are the sub-sample and full sample time averages of* $\mathbf{v}_{it} = \mathbf{C}_i^*(L)\mathbf{u}_{it} = \sum_{j=0}^{\infty} \mathbf{C}_{ij}^* \mathbf{u}_{i,t-j}$*. Suppose Assumptions 1 and 2 hold,* $q(\geq 2)$ *is fixed and* m *is fixed. Then*

$$n^{-1} \sum_{i=1}^{n} \mathbf{C}_i \mathbf{Q}_{\bar{s}_i \bar{v}_i} = O_p\left(n^{-1/2}T^{-1}\right) + O\left(T^{-2}\right) \tag{S.54}$$

Proof. For the purpose of this proof, define $m^2 \times 1$ vector $\mathbf{b}_{nT}$ obtained by vectorization of $n^{-1}\sum_{i=1}^{n} \mathbf{C}_i \mathbf{Q}_{\bar{s}_i \bar{v}_i}$,

$$\mathbf{b}_{nT} = vec\left(n^{-1}\sum_{i=1}^{n} \mathbf{C}_i \mathbf{Q}_{\bar{s}_i \bar{v}_i}\right) = T^{-1}n^{-1}\sum_{i=1}^{n} \left(\mathbf{I}_m \otimes \mathbf{C}_i\right) \bar{\boldsymbol{\xi}}_{iq},$$

where $\bar{\boldsymbol{\xi}}_{iq} = q^{-1}\sum_{\ell=1}^{q} \left(\bar{\mathbf{v}}_{i\ell} \otimes \bar{\mathbf{s}}_{i\ell}\right) - \bar{\mathbf{v}}_{i\circ} \otimes \bar{\mathbf{s}}_{i\circ}$. Denote the k-th element of $\mathbf{b}_{nT}$ as $b_{k,nT} = \mathbf{e}_k' \mathbf{b}_{nT}$, where $\mathbf{e}_k$ is $m^2 \times 1$ selection vector for the k-th element, $k = 1, 2, ..., m^2$. Define $\tilde{b}_{k,nT} = b_{k,nT} - E\left(b_{k,nT}\right)$. Using result (S.12) of Lemma 3 and noting $\sup_i \left\|\mathbf{C}_i\right\| < K$ by Assumption 2, we obtain

$$\sup_k E\left(b_{k,nT}\right) = O\left(T^{-2}\right). \tag{S.55}$$

Consider $E\left(\tilde{b}_{k,nT}\right)^2$ next. We have

$$E\left(\tilde{b}_{k,nT}\right)^2 = E\left\{\left[T^{-1}n^{-1}\sum_{i=1}^{n} \mathbf{e}_k' \left(I_m \otimes \mathbf{C}_i\right) \tilde{\boldsymbol{\xi}}_{iq}\right]^2\right\},$$

where $\tilde{\boldsymbol{\xi}}_{iq} = \bar{\boldsymbol{\xi}}_{iq} - E\left(\bar{\boldsymbol{\xi}}_{iq}\right)$. By construction $E\left(\tilde{\boldsymbol{\xi}}_{iq}\right) = 0$. In addition, $\tilde{\boldsymbol{\xi}}_{iq}$ is independent of $\tilde{\boldsymbol{\xi}}_{jq}$ for any $i \neq j$. It follows

$$\begin{aligned} E\left(\tilde{b}_{k,nT}\right)^2 &= n^{-2}\sum_{i=1}^{n} E\left(\mathbf{e}_k' \left(\mathbf{I}_m \otimes \mathbf{C}_i\right) \tilde{\boldsymbol{\xi}}_{iq}\right)^2, \\ &\leq T^{-2}n^{-2}\sum_{i=1}^{n} \left\|\mathbf{e}_k' \left(\mathbf{I}_m \otimes \mathbf{C}_i\right)\right\|^2 E\left(\left\|\tilde{\boldsymbol{\xi}}_{iq}\right\|^2\right). \end{aligned}$$

Using result (S.38) of Lemma 4 and $\sup_i \|\mathbf{C}_i\| < K$ (by Assumption 2), then

$$\sup_k E\left(\tilde{b}_{k,nT}\right)^2 = O\left(n^{-1}T^{-2}\right). \tag{S.56}$$

Using (S.55) and (S.56) in $b_{k,nT} = \tilde{b}_{k,nT} + E\left(b_{k,nT}\right)$, we obtain (S.54). ■

Lemma 10 *Consider $m \times m$ matrices $\mathbf{Q}_{\bar{f}_i\bar{f}_i} = T^{-1}q^{-1}\sum_{\ell=1}^{q}\mathbf{G}_i\left(\overline{\mathbf{f}}_\ell - \overline{\mathbf{f}}_\circ\right)\left(\overline{\mathbf{f}}_\ell - \overline{\mathbf{f}}_\circ\right)'\mathbf{G}_i'$, $\mathbf{Q}_{\bar{f}_i\bar{v}_i} = T^{-1}q^{-1}\sum_{\ell=1}^{q}\mathbf{G}_i\left(\overline{\mathbf{f}}_\ell - \overline{\mathbf{f}}_\circ\right)\left(\overline{\mathbf{v}}_{i\ell} - \overline{\mathbf{v}}_{i\circ}\right)'$ and $\mathbf{Q}_{\bar{f}_i\bar{s}_i} = T^{-1}q^{-1}\sum_{\ell=1}^{q}\mathbf{G}_i\left(\overline{\mathbf{f}}_\ell - \overline{\mathbf{f}}_\circ\right)\left(\overline{\mathbf{s}}_{i\ell} - \overline{\mathbf{s}}_{i\circ}\right)'$, where $\overline{\mathbf{f}}_\ell$ and $\overline{\mathbf{f}}_\circ$ are the sub-sample and full sample time averages of $\mathbf{f}_t = \sum_{j=0}^{\infty}\boldsymbol{\Phi}_{f\ell}L^{\ell}\boldsymbol{\varepsilon}_{f,t-\ell}$, $\overline{\mathbf{s}}_{i\ell}$ and $\overline{\mathbf{s}}_{i\circ}$ are the sub-sample and full sample time averages of the partial sum process $\mathbf{s}_{it} = \sum_{\ell=1}^{t}\mathbf{u}_{it}$, and $\overline{\mathbf{v}}_{i\ell}$ and $\overline{\mathbf{v}}_{i\circ}$ are the sub-sample and full sample time averages of $\mathbf{v}_{it} = \mathbf{C}_i^*(L)\mathbf{u}_{it} = \sum_{j=0}^{\infty}\mathbf{C}_{ij}^*\mathbf{u}_{i,t-j}$. Suppose Assumptions 1, 2 and 4 hold, and $q(\geq 2)$, m and m_f are fixed. Then,*

$$\sup_i E\left\|\mathbf{Q}_{\bar{f}_i\bar{f}_i}\right\| = O\left(T^{-2}\right), \tag{S.57}$$

$$\sup_i E\left\|\mathbf{Q}_{\bar{f}_i\bar{v}_i}\right\| = O\left(T^{-2}\right), \tag{S.58}$$

and

$$\sup_i E\left\|\mathbf{Q}_{\bar{f}_i\bar{s}_i}\right\| = O\left(T^{-1}\right). \tag{S.59}$$

Proof. Consider (S.57) first. We can write $\mathbf{Q}_{\bar{f}_i\bar{f}_i}$ as

$$\mathbf{Q}_{\bar{f}_i\bar{f}_i} = \mathbf{G}_i\mathbf{Q}_{\bar{f}\bar{f}}\mathbf{G}_i',$$

where

$$\mathbf{Q}_{\bar{f}\bar{f}} = T^{-1}q^{-1}\sum_{\ell=1}^{q}\left(\overline{\mathbf{f}}_\ell - \overline{\mathbf{f}}_\circ\right)\left(\overline{\mathbf{f}}_\ell - \overline{\mathbf{f}}_\circ\right)'.$$

Taking norm, expectation and sup over i, we have

$$\sup_i E\left\|T\mathbf{Q}_{\bar{f}_i\bar{f}_i}\right\| \leq \left(\sup_i\|\mathbf{G}_i\|\right)^2 E\left\|\mathbf{Q}_{\bar{f}\bar{f}}\right\| \leq K\cdot E\left\|T\mathbf{Q}_{\bar{f}\bar{f}}\right\|, \tag{S.60}$$

where $\sup_i\|\mathbf{G}_i\| < K$ by Assumption 4. In addition, $\mathbf{f}_t$ is a covariance stationary process that satisfies essentially the same assumptions as the covariance stationary process $\mathbf{v}_{it}$. Hence, using the same arguments as in the proof of result (S.9) in Lemma

3 in Appendix, we obtain

$$E\left\|T\mathbf{Q}_{\bar{f}\bar{f}}\right\| \leq \sup_{\ell} E\left\|\bar{\mathbf{f}}_{\ell}\right\|^{2} + \sup_{i} E\left\|\bar{\mathbf{f}}_{\circ}\right\|^{2} = O\left(T^{-1}\right), \tag{S.61}$$

where

$$\sup_{\ell} E\left\|\bar{\mathbf{f}}_{\ell}\right\|^{2} = O\left(T^{-1}\right) \text{ and } \sup_{i} E\left\|\bar{\mathbf{f}}_{\circ}\right\|^{2} = O\left(T^{-1}\right). \tag{S.62}$$

Using (S.61) in (S.60), we obtain (S.57).

Consider (S.58) next. We can write $T\mathbf{Q}_{\bar{f}_i\bar{v}_i}$ as

$$T\mathbf{Q}_{\bar{f}_i\bar{v}_i} = q^{-1}\sum_{\ell=1}^{q}\mathbf{G}_i\left(\bar{\mathbf{f}}_{\ell} - \bar{\mathbf{f}}_{\circ}\right)\left(\bar{\mathbf{v}}_{i\ell} - \bar{\mathbf{v}}_{i\circ}\right)' = q^{-1}\sum_{\ell=1}^{q}\mathbf{G}_i\bar{\mathbf{f}}_{\ell}\bar{\mathbf{v}}_{i\ell}' - \mathbf{G}_i\bar{\mathbf{f}}_{\circ}\bar{\mathbf{v}}_{i\circ}'.$$

Hence

$$E\left\|T\mathbf{Q}_{\bar{f}_i\bar{v}_i}\right\| \leq q^{-1}\sum_{\ell=1}^{q}\left\|\mathbf{G}_i\right\|\left(E\left\|\bar{\mathbf{f}}_{\ell}\right\|^{2}\right)^{1/2}\left(E\left\|\bar{\mathbf{v}}_{i\ell}\right\|^{2}\right)^{1/2} + \left\|\mathbf{G}_i\right\|\left(E\left\|\bar{\mathbf{f}}_{\circ}\right\|^{2}\right)^{1/2}\left(E\left\|\bar{\mathbf{v}}_{i\circ}\right\|^{2}\right)^{1/2}. \tag{S.63}$$

Using (S.15), (S.17), (S.62) and $\sup_i \left\|\mathbf{G}_i\right\| < K$ (by Assumption 4) in (S.63), result (S.58) follows.

Consider the last result, (S.59). Similarly to (S.63), we have

$$E\left\|T\mathbf{Q}_{\bar{f}_i\bar{s}_i}\right\| \leq q^{-1}\sum_{\ell=1}^{q}\left\|\mathbf{G}_i\right\|\left(E\left\|\bar{\mathbf{f}}_{\ell}\right\|^{2}\right)^{1/2}\left(E\left\|\bar{\mathbf{s}}_{i\ell}\right\|^{2}\right)^{1/2} + \left\|\mathbf{G}_i\right\|\left(E\left\|\bar{\mathbf{f}}_{\circ}\right\|^{2}\right)^{1/2}\left(E\left\|\bar{\mathbf{s}}_{i\circ}\right\|^{2}\right)^{1/2}. \tag{S.64}$$

Using (S.20), (S.62) and $\sup_i \left\|\mathbf{G}_i\right\| < K$ (by Assumption 4) in (S.64), result (S.59) follows. ■

Lemma 11 *Consider* $\mathbf{Q}_{\bar{w}\bar{w}} = T^{-1}n^{-1}q^{-1}\sum_{i=1}^{n}\sum_{\ell=1}^{q}\left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}\right)\left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}\right)'$ *and suppose the* $m \times 1$ *vector* $\mathbf{w}_{it}$ *is given by (48) featuring interactive time effects, Assumptions 1 to 4 hold, and* $q\,(\geq 2)$, m *and* m_f *are fixed. Then*

$$E\left|\boldsymbol{\beta}_{j0}'\mathbf{Q}_{\bar{w}\bar{w}}\boldsymbol{\beta}_{j0}'\right| = O\left(T^{-2}\right), \text{ for } j = 1, 2, ..., r_0, \tag{S.65}$$

where $\boldsymbol{\beta}_{j0}$ *(for* $j = 1, 2, ..., r_0$*) are defined in Assumption 3.*

Proof. Premultiplying (13) by $\mathbf{B}_0'$, postmultiplying by $\mathbf{B}_0$, and noting $\mathbf{B}_0'\mathbf{C}_i = \mathbf{0}$ for $i = 1, 2, ..., n$ under Assumption 3, we obtain

$$\mathbf{B}_0'\mathbf{Q}_{\bar{w}\bar{w}}\mathbf{B}_0 = \mathbf{B}_0'\mathbf{Q}_{\bar{f}\bar{f}}\mathbf{B}_0 + \mathbf{B}_0'\left(\mathbf{Q}_{\bar{f}\bar{v}} + \mathbf{Q}_{\bar{v}\bar{f}}\right)\mathbf{B}_0 + \mathbf{B}_0'\mathbf{Q}_{\bar{v}\bar{v}}\mathbf{B}_0,$$

where by the orthonormality requirement $\mathbf{B}_0'\mathbf{B}_0 = \mathbf{I}_{r_0}$. By Lemma 5, $E\left\|\mathbf{B}_0'\mathbf{Q}_{\bar{v}\bar{v}}\mathbf{B}_0\right\| = O\left(n^{-1/2}T^{-2}\right)$. Using results (S.57) and (S.58) of Lemma 10, we obtain $E\left\|\mathbf{B}_0'\mathbf{Q}_{\bar{f}\bar{f}}\mathbf{B}_0\right\| = O\left(T^{-2}\right)$ and
$E\left\|\mathbf{B}_0'\left(\mathbf{Q}_{\bar{f}\bar{v}} + \mathbf{Q}_{\bar{v}\bar{f}}\right)\mathbf{B}_0\right\| = O\left(T^{-2}\right)$. Hence, the dominant term is $O\left(T^{-2}\right)$ and result (S.65) follows. ■

Lemma 12 *Consider* $\mathbf{Q}_{\bar{w}\bar{w}} = T^{-1}n^{-1}q^{-1}\sum_{i=1}^{n}\sum_{\ell=1}^{q}\left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}\right)\left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}\right)'$ *and the associated* $m \times r$ *matrix* $\hat{\mathbf{B}}_0$ *given by orthonormal eigenvectors of* $\mathbf{Q}_{\bar{w}\bar{w}}$ *corresponding to its* r_0 *smallest eigenvalues. Suppose the* $m \times 1$ *vector* $\mathbf{w}_{it}$ *is given by (48) featuring interactive time effects, Assumptions 1 to 4 hold, and* $q\,(\geq 2)$, m *and* m_f *are fixed. Then*

$$\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0 = O_p\left(T^{-2}\right). \tag{S.66}$$

Proof. Similarly to proof of result (S.50) in Lemma 7 in Appendix, when $\mathbf{w}_{it}$ is given by (48), we continue to have $0 \leq tr\left(\hat{\mathbf{\Lambda}}\right) = tr\left(\hat{\mathbf{B}}_0'\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0\right) \leq tr\left(\mathbf{B}_0'\mathbf{Q}_{\bar{w}\bar{w}}\mathbf{B}_0\right)$. But result (S.65) of Lemma 11 implies $E\left[tr\left(\mathbf{B}_0'\mathbf{Q}_{\bar{w}\bar{w}}\mathbf{B}_0\right)\right] = O\left(T^{-2}\right)$ and it follows $E\left\|\hat{\mathbf{\Lambda}}\right\| = O\left(T^{-2}\right)$. Taking norm of $\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0 = \hat{\mathbf{B}}_0\mathbf{\Lambda}$ and noting $\left\|\hat{\mathbf{B}}_0\right\| = 1$, we obtain $\left\|\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0\right\| \leq \left\|\hat{\mathbf{\Lambda}}\right\|$. Taking expectations and using $E\left\|\hat{\mathbf{\Lambda}}\right\| = O\left(T^{-2}\right)$ yields $E\left\|\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0\right\| = O\left(T^{-2}\right)$, which is sufficient for (S.66). This completes the proof. ■

Lemma 13 *Suppose Assumptions 1 to 4 hold, and* $q\,(\geq 2)$, m *and* m_f *are fixed. Consider* $\boldsymbol{\xi}_{iq}^{*} = q^{-1}\sum_{\ell=1}^{q}\left(\bar{\boldsymbol{\omega}}_{i\ell} - \bar{\boldsymbol{\omega}}_{i}\right)\otimes\left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i}\right)$, *where* $\bar{\mathbf{s}}_{i\ell}$, $\bar{\mathbf{s}}_{i\circ}$ *are the sub-sample and full sample time averages of the partial sum process* $\mathbf{s}_{it} = \sum_{\ell=1}^{t}\mathbf{u}_{it}$, *and* $\bar{\boldsymbol{\omega}}_{i\ell}$,$\bar{\boldsymbol{\omega}}_{i}$ *are the sub-sample and full sample time averages of* $\boldsymbol{\omega}_{it} = \mathbf{f}_{it} + \mathbf{v}_{it}$, $\mathbf{f}_{it} = \sum_{j=0}^{\infty}\mathbf{\Phi}_{fi\ell}L^{\ell}\boldsymbol{\varepsilon}_{f,t-\ell}$, $\mathbf{v}_{it} = \sum_{j=0}^{\infty}\mathbf{C}_{ij}^{*}\mathbf{u}_{i,t-j}$. *Let* $\mathbf{\Omega}_{\xi_{qi}^{*}} = Var\left(\bar{\boldsymbol{\xi}}_{iq}^{*}\right)$ *and suppose* $\mathbf{\Omega}_{q}^{*} = \lim_{n,T\to\infty}\left(\mathring{\mathbf{B}}_0'\otimes\mathbf{C}_i\right)\mathbf{\Omega}_{\xi_{qi}^{*}}\left(\mathring{\mathbf{B}}_0\otimes\mathbf{C}_i'\right)$ *is positive definite. Then*

$$n^{-1/2}\sum_{i=1}^{n}\left(\mathring{\mathbf{B}}_0'\otimes\mathbf{C}_i\right)\tilde{\boldsymbol{\xi}}_{iq}^{*} \rightarrow_d N\left(\mathbf{0}, \mathbf{\Omega}_{q}^{*}\right), \tag{S.67}$$

for $n, T \to \infty$ *jointly (in no particular order), where* $\tilde{\boldsymbol{\xi}}_{iq}^{*} = \bar{\boldsymbol{\xi}}_{iq}^{*} - E\left(\bar{\boldsymbol{\xi}}_{iq}^{*}\right)$.

Proof. (S.67) can be established in a similar way as the proof of (S.51), but we rely on a central limit theorems for martingales (Theorem 24.3 in Davidson (1994)), since $\tilde{\boldsymbol{\xi}}_{iq}^{*}$ is no longer independently distributed over i in the presence of unobserved common factors. ■

S2 PME estimator for unbalanced panels

We assume there are n cross section units, $i = 1, 2, ..., n$, with each cross section unit having T_i consecutive observations. We do not allow for gaps. Let $T_m = \max T_i$. We also assume all T_i's expand at the same rate with n. Specifically, there exists a positive constant $\varkappa_u > 0$, which does not depend on the sample size, such that $T_i > \varkappa_u \cdot T_m$, for $i = 1, 2...., n$.

In practice, these assumption may require excluding some observations to avoid gaps, and excluding individual cross section units where T_i is very small compared with the rest of the panel (see Subsection S2.1). In the exposition below, we assume that the unbalanced panel satisfies the assumptions above.

To accommodate unbalanced panels, we consider the following generalization of sub-sample time averages defined in Section 4 (see (6)),

$$\bar{\mathbf{w}}_{i\ell} = \frac{1}{T_{i\ell}} \sum_{t=H_{i,\ell-1}+1}^{H_{i\ell}} \mathbf{w}_{it}, \text{ for } \ell = 1, 2, ..., q_i, \tag{S.1}$$

where $T_{i\ell} = H_{i\ell} - H_{i,\ell-1}$ is the sample size for computing sub-sample time average ℓ for unit i. We continue to assume $H_{i\ell} > H_{i,\ell-1}$, $T_i = \sum_{\ell=1}^{q_i} T_{i\ell}$, with $T_{i\ell}$ being of the same order as T_i, namely there exists positive constants $K > 0$ that do not depend on the sample size such that $T_{i\ell}/\max_{\ell=1,2,...,q_i}\{T_{i\ell}\} > K_0$ for all i and all ℓ. The number of sub-sample time averages, q_i, in (S.1) are allowed to vary across units with $2 \leq q_i \leq q_m$, where q_m does not depend on the sample size. The time average $\bar{\mathbf{w}}_{i\circ}$ is now defined by

$$\bar{\mathbf{w}}_{i\circ} = q_i^{-1} \sum_{\ell=1}^{q_i} \bar{\mathbf{w}}_{i\ell},$$

and

$$\mathbf{Q}_{\bar{w}_i \bar{w}_i} = T_i^{-1} q_i^{-1} \sum_{\ell=1}^{q_i} \left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}\right) \left(\bar{\mathbf{w}}_{i\ell} - \bar{\mathbf{w}}_{i\circ}\right)'.$$

with the associated pooled version given by

$$\mathbf{Q}_{\bar{w}\bar{w}} = n^{-1}\sum_{i=1}^{n}\mathbf{Q}_{\bar{w}_i\bar{w}_i} = T_{ave}^{-1}n^{-1}\sum_{i=1}^{n}\left(\phi_i q_i^{-1}\sum_{\ell=1}^{q_i}\left(\bar{\mathbf{w}}_{i\ell}-\bar{\mathbf{w}}_{i\circ}\right)\left(\bar{\mathbf{w}}_{i\ell}-\bar{\mathbf{w}}_{i\circ}\right)'\right),$$

where

$$\phi_i = \left(\frac{T_i}{T_{ave}}\right)^{-1} \text{ and } T_{ave} = \left(n^{-1}\sum_{i=1}^{n}T_i^{-1}\right)^{-1}.$$

Following the same arguments as in the case of balanced panels, it can be established that $E\left\|\mathbf{Q}_{\bar{v}_i\bar{v}_i}\right\|_1 < K/T_i^2$, $E\left\|\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right\|_1 < K/T_i$, and $E\left\|\mathbf{Q}_{\bar{s}_i\bar{s}_i}\right\| < K$. These results imply $\mathbf{Q}_{\bar{w}\bar{w}}\boldsymbol{\beta} = O_p\left(T_{ave}^{-1}\right)$ and we can use the eigenvectors corresponding to the r_0 smallest eigenvalues of $\mathbf{Q}_{\bar{w}\bar{w}}$ to estimate long-run relations subject to the exact identifying assumptions.

To derive asymptotic distribution, let

$$\mathbf{Z}_i = \mathbf{C}_i\left[\phi_i q_i^{-1}\sum_{\ell=1}^{q_i}\left(\bar{\mathbf{s}}_{i\ell}-\bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell}-\bar{\mathbf{v}}_{i\circ}\right)'\right]\mathring{\mathbf{B}}_0 - \mathbf{C}_i\left[\phi_i q_i^{-1}\sum_{\ell=1}^{q_i}E\left(\bar{\mathbf{s}}_{i\ell}-\bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\mathbf{v}}_{i\ell}-\bar{\mathbf{v}}_{i\circ}\right)'\right]\mathring{\mathbf{B}}_0, \tag{S.2}$$

and following the same arguments as in Section 4, we obtain

$$\mathbf{Q}_{\bar{w}\bar{w}}\sqrt{n}T_{ave}\left(\widehat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right) = -n^{-1/2}\sum_{i=1}^{n}\mathbf{Z}_i + O_p\left(\frac{\sqrt{n}}{T_{ave}}\right), \tag{S.3}$$

Consider the exact identifying restrictions $\mathring{\mathbf{B}}_0 = \left(\mathbf{I}_{r_0}, \mathring{\mathbf{B}}_{0,2}'\right)'$, where $\mathbf{I}_{r_0}$ is an identity matrix of order r_0. Let $\hat{\boldsymbol{\Theta}}$ and $\boldsymbol{\Theta}_0$ be the lower $(m-r_0)\times r_0$ block of $\widehat{\mathring{\mathbf{B}}}_0$ and $\mathring{\mathbf{B}}_0$, respectively. Then, $\sqrt{n}T_{ave}vec\left(\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0\right)$ converges to a normal distribution and the variance of $\hat{\boldsymbol{\Theta}}$ can be estimated as

$$\widehat{Var}\left(\hat{\boldsymbol{\Theta}}\right) = \frac{1}{nT_{ave}^2}\mathbf{Q}_{22,\bar{w}\bar{w}}^{-1}\hat{\boldsymbol{\Omega}}_{z,22}\mathbf{Q}_{22,\bar{w}\bar{w}}^{-1}, \tag{S.4}$$

where

$$\hat{\boldsymbol{\Omega}}_z = n^{-1}\sum_{i=1}^{n}\phi_i^2 q_i^{-2}\sum_{\ell=1}^{q_i}\sum_{\ell'=1}^{q_i}\left(\bar{\mathbf{E}}_{i\ell}\otimes\mathbf{I}_m\right)\left(\bar{\mathbf{w}}_{i\ell}-\bar{\mathbf{w}}_{i\circ}\right)\left(\bar{\mathbf{w}}_{i\ell'}-\bar{\mathbf{w}}_{i\circ}\right)'\left(\bar{\mathbf{E}}_{i\ell'}'\otimes\mathbf{I}_m\right). \tag{S.5}$$

and, as before, $\mathbf{Q}_{22,\bar{w}\bar{w}}$ and $\hat{\boldsymbol{\Omega}}_{z,22}$ are the $(m-r_0)\times(m-r_0)$ lower blocks of matrices

$\mathbf{Q}_{\bar{w}\bar{w}}$ and $\hat{\mathbf{\Omega}}_z$, respectively.

S2.1 Implementation of unbalanced PME estimator in empirical application

Let $T_m = \max T_i$. In empirical application in Section 10 we exclude cross section units with gaps in data and we exclude units with $T_i < 20$ years. The remaining units are used for estimation, and we denote the number of these units as n.

Definition of sub-sample sample time averages in (S.1) is quite general, and allows for a variety of asymptotically justified options to implement the PME estimator for unbalanced panels. In view of inference in small samples being adversely affected by a larger choices of q, we set $q_i = 2$ for all i, and we divide the T_i consecutive observations into two halves when T_i is even. If T_i is odd, we compute the first sub-sample time average based on $(T_i + 1)/2$ observations, and the second sub-sample time average based on the remaining $(T_i - 1)/2$ observations. This version of unbalanced PME estimator is considered in the empirical application.

S3 Consistency and asymptotic distribution of PME estimator in the model with interactive time effects

Theorems 3-4 extends Theorems 1-2 to model with interactive time effects.

Theorem 3 *Consider the panel data model for the $m \times 1$ vector $\mathbf{w}_{it}$ given by (48) and suppose that Assumptions 1 to 4 hold, and the number of long-run relations, r_0, is known. Let $\hat{\mathbf{B}}_0$ be formed from the first r_0 orthonormalized eigenvectors of $\mathbf{Q}_{\bar{w}\bar{w}}$ given by (9), associated with its r_0 smallest eigenvalues. Then for a fixed m, m_f and $q\,(\geq 2)$, $\hat{\mathbf{B}}_0\mathbf{H} \to_p \mathring{\mathbf{B}}_0$ as $n, T \to \infty$ jointly such that $T_n \approx n^d$ and $d > 0$, for some $r_0 \times r_0$ non-singular matrix $\mathbf{H}$.*

Proof of Theorem 3. Using (18) in (51), we have

$$\mathbf{Q}_{\bar{w}\bar{w}} = \mathbf{\Psi}_n + O_p\left(n^{-1/2}\right) + O_p\left(T^{-1}\right).$$

Multiplying this expression by $\hat{\mathbf{B}}_0'$ from the left and by $\hat{\mathbf{B}}_0$ from the right, and noting eigenvectors $\hat{\mathbf{B}}_0$ are normalized so that $\hat{\mathbf{B}}_0'\hat{\mathbf{B}}_0 = \mathbf{I}_r$ yields

$$\hat{\mathbf{B}}_0'\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0 = \frac{(q-1)}{6q}\hat{\mathbf{B}}_0'\mathbf{\Psi}_n\hat{\mathbf{B}}_0 + O_p\left(n^{-1/2}\right) + O_p\left(T^{-1}\right), \tag{S.1}$$

where $lim_{n\to\infty}\mathbf{\Psi}_n = \mathbf{\Psi}$ as $n \to \infty$. (S.1) is the same as (A.7) with the exception of the $O_p\left(T^{-1}\right)$ term (as opposed to $O_p\left(T^{-2}\right)$). Consistency of $\hat{\mathbf{B}}_0$ now follows using the same arguments as in the proof of Theorem 1. ■

Theorem 4 *Consider the panel data model for the $m\times 1$ vector $\mathbf{w}_{it}$ given by (48), and suppose that Assumptions 1 to 4 hold, m, m_f and $q(\geq 2)$ are fixed, and the number of long-run relations, r_0 $(m > r_0 > 0)$ is known. Suppose further that the long-run relations, $\mathring{\mathbf{B}}_0$, of interest are subject to the exact identifying restrictions, $\mathbf{R}\mathring{\mathbf{B}}_0 = \mathbf{A}$, given by (22), and consider the PME estimator of $\mathring{\mathbf{B}}_0$ given by $\hat{\mathring{\mathbf{B}}}_0 = \hat{\mathbf{B}}_0\left(\mathbf{R}\hat{\mathbf{B}}_0\right)^{-1}\mathbf{A}$, where $\hat{\mathbf{B}}_0 = \left(\hat{\boldsymbol{\beta}}_{10}, \hat{\boldsymbol{\beta}}_{20}, ..., \hat{\boldsymbol{\beta}}_{r_0,0}\right)$ are the first r_0 orthonormalized eigenvectors of $\mathbf{Q}_{\bar{w}\bar{w}}$ defined by (9). Then*

$$\sqrt{n}T\left(\mathbf{I}_r\otimes\mathbf{Q}_{\bar{w}\bar{w}}\right) vec\left(\hat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right) \to_d N\left(\mathbf{0}, \mathbf{\Omega}_q^*\right), \tag{S.2}$$

as $n, T \to \infty$, jointly such that $T \approx n^d$ for $d > 1/2$, where

$$\mathbf{\Omega}_q^* = lim_{n,T\to\infty}\left[n^{-1}\sum_{i=1}^{n}\left(\mathring{\mathbf{B}}_0'\otimes\mathbf{C}_i\right)\mathbf{\Omega}_{\xi_{qi}^*}\left(\mathring{\mathbf{B}}_0\otimes\mathbf{C}_i'\right)\right], \tag{S.3}$$

and $\mathbf{\Omega}_{\xi_{qi}^*} = Var\left(\boldsymbol{\xi}_{iq}^*\right)$, and $\mathbf{Q}_{\bar{w}\bar{w}} \to_p \frac{(q-1)}{6q}\mathbf{\Psi}$.

Proof of Theorem 4. Averaging (49) over i, we have

$$\begin{aligned}\mathbf{Q}_{\bar{w}\bar{w}} &= n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{s}_i}\mathbf{C}_i' + n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i} + n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{f}_i} + \\ &\quad + n^{-1}\sum_{i=1}^{n}\mathbf{Q}_{\bar{v}_i\bar{s}_i}'\mathbf{C}_i' + n^{-1}\sum_{i=1}^{n}\mathbf{Q}_{\bar{f}_i\bar{s}_i}'\mathbf{C}_i + \mathbf{Q}_{\bar{f}\bar{f}} + \mathbf{Q}_{\bar{f}\bar{v}} + \mathbf{Q}_{\bar{v}\bar{f}} + \mathbf{Q}_{\bar{v}\bar{v}}.\end{aligned} \tag{S.4}$$

Multiplying both sides of (S.4) by $\mathring{\mathbf{B}}_0$ from the right, and noting that $\mathbf{C}_i'\mathring{\mathbf{B}}_0 = \mathbf{0}$ under Assumption 3, we have

$$\mathbf{Q}_{\bar{w}\bar{w}}\mathring{\mathbf{B}}_0 = \left(n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0 + \left(n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{f}_i}\right)\mathring{\mathbf{B}}_0 + \mathbf{Q}_{\bar{f}\bar{f}}\mathring{\mathbf{B}}_0 + \mathbf{Q}_{\bar{f}\bar{v}}\mathring{\mathbf{B}}_0 + \mathbf{Q}_{\bar{v}\bar{f}}\mathring{\mathbf{B}}_0 + \mathbf{Q}_{\bar{v}\bar{v}}\mathring{\mathbf{B}}_0, \tag{S.5}$$

where by results (S.57)-(S.58) of Lemma 3, $\mathbf{Q}_{\bar{f}\bar{f}} = O_p\left(T^{-2}\right)$, $\mathbf{Q}_{\bar{f}\bar{v}} = \mathbf{Q}_{\bar{v}\bar{f}}' = O_p\left(T^{-2}\right)$, and by result (S.47) of Lemma (5) $\mathbf{Q}_{\bar{v}\bar{v}} = O_p\left(n^{-1/2}T^{-2}\right)$. Using result (S.66) of Lemma 12, we have $\left\|\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathbf{B}}_0\right\| = O_p\left(T^{-2}\right)$, and given that $\hat{\mathring{\mathbf{B}}}_0$ is an $O_p(1)$ rotation of $\hat{\mathbf{B}}_0$, it follows

$$\mathbf{Q}_{\bar{w}\bar{w}}\hat{\mathring{\mathbf{B}}}_0 = O_p\left(T^{-2}\right). \tag{S.6}$$

Subtracting (S.5) from (S.6) yields (noting $O_p\left(T^{-2}\right)$ dominates $O_p\left(n^{-1/2}T^{-2}\right)$)

$$\mathbf{Q}_{\bar{w}\bar{w}}\left(\hat{\mathring{\mathbf{B}}} - \mathring{\mathbf{B}}_0\right) = -\left(n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\mathring{\mathbf{B}}_0 + \left(n^{-1}\sum_{i=1}^{n}\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{f}_i}\right)\mathring{\mathbf{B}}_0 + O_p\left(T^{-2}\right).$$

Defining $\mathbf{Q}_{\bar{s}_i\bar{\omega}_i} = \mathbf{Q}_{\bar{s}_i\bar{v}_i} + \mathbf{Q}_{\bar{s}_i\bar{f}_i}$ and multiplying above equation by $\sqrt{n}T$ we obtain,

$$\mathbf{Q}_{\bar{w}\bar{w}}\sqrt{n}T\left(\widehat{\mathring{\mathbf{B}}_0} - \mathring{\mathbf{B}}_0\right) = -\left(n^{-1/2}\sum_{i=1}^{n}T\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{\omega}_i}\right)\mathring{\mathbf{B}}_0 + O_p\left(\frac{\sqrt{n}}{T}\right). \tag{S.7}$$

Noting $\mathbf{s}_{it}$ is independently distributed of $\mathbf{f}_{it'}$ for all t, t', the first term on the right side of (S.7) can be written as

$$\left(n^{-1/2}\sum_{i=1}^{n}T\mathbf{C}_i\mathbf{Q}_{\bar{s}_i\bar{\omega}_i}\right)\mathring{\mathbf{B}}_0 = \left(n^{-1/2}\sum_{i=1}^{n}\mathbf{Z}_i^*\right)\mathring{\mathbf{B}}_0 + \frac{\sqrt{n}}{T}\left[n^{-1}\sum_{i=1}^{n}\mathbf{C}_iE\left(T^2\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)\right]\mathring{\mathbf{B}}_0,$$

where $\mathbf{Z}_i^* = \mathbf{C}_i\left[T\mathbf{Q}_{\bar{s}_i\bar{\omega}_i} - E\left(T\mathbf{Q}_{\bar{s}_i\bar{\omega}_i}\right)\right]\mathring{\mathbf{B}}_0$, and $E\left(T^2\mathbf{Q}_{\bar{s}_i\bar{\omega}_i}\right) = E\left(T^2\mathbf{Q}_{\bar{s}_i\bar{v}_i}\right)$. Using (A.12), it follows that

$$\mathbf{Q}_{\bar{w}\bar{w}}\sqrt{n}T\left(\hat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right) = -n^{-1/2}\sum_{i=1}^{n}\mathbf{Z}_i^* + O_p\left(\frac{\sqrt{n}}{T}\right),$$

Vectorizing the above equation we have

$$\left(\mathbf{I}_r \otimes \mathbf{Q}_{\bar{w}\bar{w}}\right) \sqrt{n}T \ vec\left(\hat{\mathring{\mathbf{B}}}_0 - \mathring{\mathbf{B}}_0\right) = n^{-1/2} \sum_{i=1}^{n} \left(\mathring{\mathbf{B}}_0' \otimes \mathbf{C}_i\right) \tilde{\boldsymbol{\xi}}_{iq}^* + O_p\left(\frac{\sqrt{n}}{T}\right), \tag{S.8}$$

where $\tilde{\boldsymbol{\xi}}_{iq}^* = \bar{\boldsymbol{\xi}}_{iq}^* - E\left(\bar{\boldsymbol{\xi}}_{iq}^*\right)$, and $\bar{\boldsymbol{\xi}}_{iq}^*$ is given by (recall $\mathbf{Z}_i^* = \mathbf{C}_i \left[T\mathbf{Q}_{\bar{s}_i\bar{\omega}_i} - E\left(T\mathbf{Q}_{\bar{s}_i\bar{\omega}_i}\right)\right] \mathring{\mathbf{B}}_0$ and $\mathbf{Q}_{\bar{s}_i\bar{\omega}_i} = T^{-1}q^{-1}\sum_{\ell=1}^{q} \left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\boldsymbol{\omega}}_{i\ell} - \bar{\boldsymbol{\omega}}_{i\circ}\right)'$)

$$\bar{\boldsymbol{\xi}}_{iq}^* = q^{-1} \sum_{\ell=1}^{q} vec\left[\left(\bar{\mathbf{s}}_{i\ell} - \bar{\mathbf{s}}_{i\circ}\right)\left(\bar{\boldsymbol{\omega}}_{i\ell} - \bar{\boldsymbol{\omega}}_{i\circ}\right)'\right] = q^{-1} \sum_{\ell=1}^{q} \left(\bar{\boldsymbol{\omega}}_{i\ell} \otimes \bar{\mathbf{s}}_{i\ell}\right) - \bar{\boldsymbol{\omega}}_{i\circ} \otimes \bar{\mathbf{s}}_{i\circ}.$$

Lemma 13 established convergence in distribution for $n^{-1/2}\sum_{i=1}^{n}\left(\mathring{\mathbf{B}}_0' \otimes \mathbf{C}_i\right)\tilde{\boldsymbol{\xi}}_{iq}^*$. Using this Lemma in (S.8), and noting that $d > 1/2$ implies$\sqrt{n}/T \to 0$ as $n, T \to \infty$, we obtain (S.2), as required. ■

S4 Description of Data Generating Processes

We have experiments with $r_0 = 0, 1, 2$ long-run relations, and with and without interactive time effects. Overview of all experiments is provided in the paper. Subsection S4.1 below provides full details of the DGPs with long-run relations and without interactive time effects. Subsection S4.2 provides details of the DGPs without long-run relations and without interactive time effects. Subsection S4.3 provides details on augmentation of each of these DGPs with interactive time effects.

S4.1 Experiments with $r_0 = 1$ and 2 long-run relations and $m = 3$ variables

We consider the following data generating process for $\mathbf{w}_{it}$,

$$\Delta \mathbf{w}_{it} = \mathbf{d}_i - \boldsymbol{\Pi}_i \mathbf{w}_{i,t-1} + \mathbf{u}_{it} - \boldsymbol{\Theta}_i \mathbf{u}_{i,t-1}, \tag{S.1}$$

where

$$\boldsymbol{\Pi}_i = \mathbf{A}_i \mathbf{B}_0',$$

$\Delta \mathbf{w}_{it}$ is $m \times 1$, $\mathbf{A}_i$ is $m \times r_0$, $\mathbf{B}_0$ is $m \times r_0$. To ensure that $\mathbf{w}_{it}$ is not trended we impose the restriction

$$\mathbf{d}_i = \mathbf{\Pi}_i \boldsymbol{\mu}_{iw} = \mathbf{A}_i \mathbf{B}_0' \boldsymbol{\mu}_{iw}. \tag{S.2}$$

We set $m = 3$ and consider two cases, one and two long-run relations, $r_0 = 1$ and $r_0 = 2$.

For $r_0 = 1$, $\mathbf{B}_0 = \boldsymbol{\beta}_0 = (1, 0, -1)'$, and for $r_0 = 2$

$$\mathbf{B}_0 = \begin{pmatrix} \boldsymbol{\beta}_{10} & \boldsymbol{\beta}_{20} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}.$$

The VARMA representation of (S.1) is given by

$$\mathbf{w}_{it} = \mathbf{d}_i + (\mathbf{I}_m - \mathbf{A}_i \mathbf{B}_0') \, \mathbf{w}_{i,t-1} + \mathbf{u}_{it} - \mathbf{\Theta}_i \mathbf{u}_{i,t-1} \tag{S.3}$$

which can be written more compactly as

$$\mathbf{\Psi}_i(L) \mathbf{w}_{it} = \mathbf{d}_i + (\mathbf{I}_m - \mathbf{\Theta}_i L) \mathbf{u}_{it}, \tag{S.4}$$

where $\mathbf{\Psi}_i(L) = \mathbf{I}_m - \mathbf{\Psi}_i L$, and $\mathbf{\Psi}_i = \mathbf{I}_m - \mathbf{A}_i \mathbf{B}_0'$.

S4.1.1 Relation to Granger Representation

$\Delta \mathbf{w}_{it}$ can be written as

$$\Delta \mathbf{w}_{it} = \left[\mathbf{C}_i + (1 - L) \mathbf{C}_i^*(L) \right] \mathbf{u}_{it} \tag{S.5}$$

where $\mathbf{C}_i^*(L) = \sum_{\ell=0}^{\infty} \mathbf{C}_{i\ell}^* L^{\ell}$. First differencing (S.4),

$$\mathbf{\Psi}_i(L) \Delta \mathbf{w}_{it} = (\mathbf{I}_m - \mathbf{\Theta}_i L) \Delta \mathbf{u}_{it},$$

and using (S.5), we obtain

$$\mathbf{\Psi}_i(L) \left[\mathbf{C}_i + (1 - L) \mathbf{C}_i^*(L) \right] = (\mathbf{I}_m - \mathbf{\Theta}_i L)(1 - L). \tag{S.6}$$

Using the above we obtain $\mathbf{C}_i$ and $\mathbf{C}_i^*$, $\ell = 1, 2, ...$ in terms of $\mathbf{\Psi}_i$ and $\mathbf{\Theta}_i$. Hence we can relate the parameters of (S.1) to the general linear representation

$$\mathbf{w}_{it} = \mathbf{a}_i + \mathbf{C}_i \mathbf{s}_{it} + \mathbf{C}_i^*(L)\mathbf{u}_{it}. \tag{S.7}$$

where $\mathbf{s}_{it} = \mathbf{u}_{i1} + \mathbf{u}_{i2} + ... + \mathbf{u}_{it}$. Pre-multiplying by $\mathbf{B}_0'$, we obtain

$$\mathbf{B}_0'\mathbf{w}_{it} = \mathbf{B}_0'\mathbf{a}_i + \mathbf{B}_\mathbf{0}'\mathbf{C}_i^*(L)\mathbf{u}_{it}.$$

S4.1.2 Parameterization

We would expect $\mathbf{B}_0$ to be more difficult to estimate the more persistent is $\mathbf{B}_0'\mathbf{w}_{it}$. In the extreme case where $\mathbf{A}_i = \mathbf{0}$, $\mathbf{B}_0$ is not identified. Pre-multiplying both sides of (S.1) by $\mathbf{B}_0'$ yields

$$\mathbf{B}_0'\Delta\mathbf{w}_{it} = \mathbf{B}_0'\mathbf{d}_i - \mathbf{B}_0'\mathbf{A}_i\mathbf{B}_0'\mathbf{w}_{i,t-1} + \mathbf{B}_0'\mathbf{u}_{it} - \mathbf{B}_0'\mathbf{\Theta}_i\mathbf{u}_{i,t-1}. \tag{S.8}$$

Let $\boldsymbol{\xi}_{it} = \mathbf{B}_0'\mathbf{w}_{it}$ then

$$\boldsymbol{\xi}_{it} = \mathbf{B}_0'\mathbf{d}_i + (\mathbf{I}_r - \mathbf{B}_0'\mathbf{A}_i)\boldsymbol{\xi}_{i,t-1} + \mathbf{B}_0'\mathbf{u}_{it} - \mathbf{B}_0'\mathbf{\Theta}_i\mathbf{u}_{i,t-1}. \tag{S.9}$$

In the VAR(1) case where $\mathbf{\Theta}_i = \mathbf{0}$, the persistence of $\boldsymbol{\xi}_{it} = \mathbf{B}_0'\mathbf{w}_{it}$ is fully determined by the largest eigenvalue of $(\mathbf{I}_r - \mathbf{B}_0'\mathbf{A}_i)$. Allowing for the MA component complicates the analysis of persistence of $\boldsymbol{\xi}_{it}$, but it will still depend on the size of the eigenvalues of $(\mathbf{I}_r - \mathbf{B}_0'\mathbf{A}_i)$. As an approximation it is reasonable to control the eigenvalues of $(\mathbf{I}_r - \mathbf{B}_0'\mathbf{A}_i)$.

S4.1.3 Long run relations

When $r_0 = 1$, we set $\mathbf{B}_0' = \boldsymbol{\beta}_0' = (1, 0, -1)$ and $\mathbf{A}_i = (a_{i,11}, a_{i,21}, a_{i,31})'$, and hence

$$\mathbf{B}_0'\mathbf{A}_i = a_{i,11} - a_{i,31} = \rho_i. \tag{S.10}$$

In the case of VAR(1) the persistence of $\boldsymbol{\xi}_{it}$ does not depend on $a_{i,21}$, which we set to 0, for all i. We consider two sets of values for ρ_i and generate them as (for $i = 1, 2, ..., n$)

$$\begin{aligned} \text{slow } \rho_i &\sim IIDU\left[0.1, 0.2\right], \text{ and} \\ \text{moderate } \rho_i &\sim IIDU\left[0.1, 0.3\right]. \end{aligned}$$

Slow speed of convergence corresponds to a median half-life of 4.3 periods (years) and moderate speed of convergence corresponds to a median half-life of 3.1 periods (years). Since both $a_{i,11}$ and $a_{i,31}$ are nonzero, long-run causality runs from $w_{it,1}$ to $(w_{it,2}, w_{it,3})$ as well as from $w_{it,3}$ to $w_{it,1}$. (S.10) leaves us with one free parameter to determine $\mathbf{A}_i$ which we choose to control a system measure of fit, as described below.

When $r_0 = 2$, we set the long-run relations as

$$\mathbf{B}_0 = (\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}.$$

Let

$$\mathbf{A}_i = \begin{pmatrix} a_{i,11} & a_{i,12} \\ a_{i,21} & a_{i,22} \\ a_{i,31} & a_{i,32} \end{pmatrix},$$

then

$$\begin{aligned} \mathbf{I}_2 - \mathbf{B}_0'\mathbf{A}_i &= \mathbf{I}_2 - \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} a_{i,11} & a_{i,12} \\ a_{i,21} & a_{i,22} \\ a_{i,31} & a_{i,32} \end{pmatrix}, \\ &= \begin{pmatrix} 1 - (a_{i,11} - a_{i,31}) & -(a_{i,12} - a_{i,32}) \\ -(a_{i,21} - a_{i,31}) & 1 - (a_{i,22} - a_{i,32}) \end{pmatrix}. \end{aligned}$$

Hence

$$\mathbf{I}_2 - \mathbf{B}_0'\mathbf{A}_i = \begin{pmatrix} 1 - \rho_{i,11} & -\rho_{i,12} \\ -\rho_{i,21} & 1 - \rho_{i,22} \end{pmatrix}. \tag{S.11}$$

The eigenvalues of this matrix are given by

$$\begin{aligned}
\lambda_1 &= 1-\frac{1}{2}\rho_{i,11}-\frac{1}{2}\rho_{i,22}-\frac{1}{2}\sqrt{\rho_{i,11}^2-2\rho_{i,11}\rho_{i,22}+\rho_{i,22}^2+4\rho_{i,12}\rho_{i,21}}, \\
\lambda_2 &= 1-\frac{1}{2}\rho_{i,11}-\frac{1}{2}\rho_{i,22}+\frac{1}{2}\sqrt{\rho_{i,11}^2-2\rho_{i,11}\rho_{i,22}+\rho_{i,22}^2+4\rho_{i,12}\rho_{i,21}}.
\end{aligned}$$

To simplify the design we set $\rho_{i,12} = (a_{i,12} - a_{i,32}) = 0 = \rho_{i,21} = (a_{i,21} - a_{i,31}) = 0$. then

$$\rho_{i,11} = (a_{i,11} - a_{i,31}) \text{ and } \rho_{i,22} = (a_{i,22} - a_{i,32}),$$

and the eigenvalues are $\lambda_1 = 1 - \rho_{i,11}$, and $\lambda_2 = 1 - \rho_{i,22}$. A stable solution arises so long as $sup_{i,j}\left|1 - \rho_{i,jj}\right| < 1$. We consider two sets of values for $\rho_{i,11}$ and $\rho_{i,22}$ and generate them as (for $i = 1, 2, ..., n$)

$$\text{Slow } \rho_{i,11} \sim IIDU\left[0.1, 0.2\right] \text{ ; Moderate } \rho_{i,11} \sim IIDU\left[0.1, 0.3\right],$$

$$\text{Slow } \rho_{i,22} \sim IIDU\left[0.1, 0.2\right] \text{ ; Moderate } \rho_{i,22} \sim IIDU\left[0.1, 0.3\right].$$

We set $a_{i,j\ell}$ values below, to achieve a balance between the speed of convergence to equilibrium, ρ_{ij}, and the fit of the error correction equations.

S4.1.4 Deterministic terms, errors and initial values

We generate $\boldsymbol{\mu}_{iw} = (\mu_{ij,w})$ as $\mu_{ij,w} \sim IIDN(0,1)$, for $j = 1, 2, 3$; and $i = 1, 2, ..., n$. These parameters do not enter the distribution of the estimators of $\mathbf{B}_0$.

Errors: $\mathbf{u}_{it}$ are generated as

$$\mathbf{u}_{it} = \mathbf{P}_{ui}\boldsymbol{\varepsilon}_{it}, \tag{S.12}$$

for $i = 1, 2, ..., n$, and $t = 1, 2, ..., T$, where two cases for considered for the distribution of $\boldsymbol{\varepsilon}_{it}$: Gaussian case with elements of $\boldsymbol{\varepsilon}_{it}$ generated as $IIDN\left(0, 1\right)$, and non-Gaussian case with individual elements of $\boldsymbol{\varepsilon}_{it}$ generated as IID from a chi-squared distribution with 4 degrees of freedom normalized to mean zero and unit variance, namely $\left[\chi^2\left(4\right) - 4\right]/\sqrt{8}$. Matrix $\mathbf{P}_{ui}$ is a lower-triangular matrix computed as Choleski factorization of $\boldsymbol{\Sigma}_i = \mathbf{P}_{ui}\mathbf{P}_{ui}'$, where individual elements of $\boldsymbol{\Sigma}_i$, denoted as $\sigma_{i,pq}$, for $p, q = 1, 2, 3$ are generated as follows. We set the diagonal elements to one, namely $\sigma_{i,pp} = 1$ for $p = 1, 2, 3$, and $i = 1, 2, ..., n$. We generate the off-diagonal elements as

$\sigma_{i,pq} = \sigma_{i,qp} \sim IIDU\left(0, 0.5\right)$ for $\left(p, q\right) \in \left\{\left(1, 2\right), \left(1, 3\right), \left(2, 3\right)\right\}$, and $i = 1, 2, ..., n$. This ensures $\mathbf{\Sigma}_i$ is symmetric and positive definite with a probability one.

For the MA part we consider two cases $\mathbf{\Theta}_i = 0$ (VAR design), and $\mathbf{\Theta}_i = diag(\theta_{ij}, j = 1, 2, 3)$, with θ_{ij}, $j = 1, 2, 3 \sim U(0, 0.50)$.

Initial values: We first generate $\Delta\mathbf{w}_{it}$, for $t = 1, 2, ..., T$, using the representation, (S.1), and then cumulate these differences to obtain $\mathbf{w}_{it}$. To this end we require the initial values $\mathbf{B}'\mathbf{w}_{i0}$ and $\Delta\mathbf{w}_{i0}$. Using the Granger representation let $\Upsilon_i\left(L\right) = \left[\mathbf{C}_i + (1 - L)\mathbf{C}_i^*(L)\right] = \sum_{\ell=0}^{\infty} \Upsilon_{i\ell} L^\ell$, and

$$\Delta\mathbf{w}_{i0} = \sum_{\ell=0}^{\infty} \Upsilon_{i,-\ell}\mathbf{u}_{i,-\ell}, \tag{S.13}$$

and $\mathbf{B}_0'\mathbf{w}_{i0} = \mathbf{B}_0'\mathbf{a}_i + \sum_{\ell=0}^{\infty} \mathbf{B}_0'\mathbf{C}_{i,-\ell}^*\mathbf{u}_{i,-\ell}$. Hence, $E\left(\mathbf{B}_0'\mathbf{w}_{i0}\left|\mathbf{a}_i\right.\right) = E\left(\boldsymbol{\xi}_{i0}\left|\mathbf{a}_i\right.\right) = \mathbf{B}_0'\mathbf{a}_i$. Also using (S.9), and noting that $E\left(\Delta\mathbf{w}_{i,t-1}\right) = \mathbf{0}$, we have

$$E\left(\boldsymbol{\xi}_{it}\left|\mathbf{d}_i, \boldsymbol{\alpha}_i\right.\right) = \mathbf{B}'\mathbf{d}_i + (\mathbf{I}_r - \mathbf{B}_0'\mathbf{A}_i)E\left(\boldsymbol{\xi}_{i,t-1}\left|\boldsymbol{\alpha}_i\right.\right),$$

and assuming $|\lambda_{\min}(\mathbf{I}_r - \mathbf{B}_0'\mathbf{A}_i)| < 1$ we obtain $E\left(\boldsymbol{\xi}_{i0}\right) = (\mathbf{B}_0'\mathbf{A}_i)^{-1}\mathbf{B}_0'\mathbf{d}_i$. Therefore, recalling that $\mathbf{d}_i = \mathbf{\Pi}_i\boldsymbol{\mu}_{iw}$ (see (S.2))

$$\mathbf{B}_0'\mathbf{w}_{i0} = \mathbf{B}_0'\boldsymbol{\mu}_{iw} + \sum_{\ell=0}^{\infty} \mathbf{B}_0'\mathbf{C}_{i,-\ell}^*\mathbf{u}_{i,-\ell}.$$

Using (S.7) we also have

$$Var\left(\mathbf{B}_0'\mathbf{w}_{it}\right) = Var\left(\mathbf{B}_0'\mathbf{w}_{i0}\right) = \mathbf{B}_\mathbf{0}'\left(\sum_{s=0}^{\infty} \mathbf{C}_{is}^{*\prime}\mathbf{V}_i\mathbf{C}_{is}^*\right)\mathbf{B}_0. \tag{S.14}$$

We generate $\Delta\mathbf{w}_{i0}$ and $\mathbf{B}_0'\mathbf{w}_{i0}$ according to

$$\Delta\mathbf{w}_{i0} = \sum_{\ell=0}^{M} \Upsilon_{i,-\ell}\mathbf{u}_{i,-\ell}, \text{ and } \mathbf{B}_0'\mathbf{w}_{i0} = \mathbf{B}_0'\boldsymbol{\mu}_{iw} + \sum_{\ell=0}^{M} \mathbf{B}_0'\mathbf{C}_{i,-\ell}^*\mathbf{u}_{i,-\ell},$$

where we set $M = 50$.

S4.1.5 Fit of error correction equations and setting the error correction coefficients

An average measure of fit for the $m \times 1$ system of equations for $\Delta \mathbf{w}_{it} = (\Delta \mathring{w}_{it,1}, \Delta \mathring{w}_{it,2}, ..., \Delta \mathring{w}_{it,m})'$ is given by

$$PR_{nT}^2 = 1 - \frac{\sum_{j=1}^{m} \sum_{t=1}^{T} \sum_{i=1}^{n} u_{it,j}^2}{\sum_{j=1}^{m} \sum_{t=1}^{T} \sum_{i=1}^{n} (\Delta w_{it,j} - \Delta \bar{w}_{iT,j})^2}, \tag{S.15}$$

where $\Delta \bar{w}_{iT,j} = T^{-1} \sum_{t=1}^{T} \Delta w_{it,j}$. This system measure places equal weights on the fit of the m different error-correcting (EC) equations. We can control for PR_{nT}^2 by generating remaining free parameters in $\mathbf{A}_i$.

S4.1.6 Case of $r_0 = 1$ long-run relation in VAR(1) design

Using (S.15), and for $\mathbf{\Theta}_i = \mathbf{0}$, under stationarity and for large n, (T does not need to be large given the cross-sectional independence of $\mathbf{u}_{it}$) we have

$$\begin{aligned} PR_n^2 &= 1 - \frac{\sum_{j=1}^{m} \sum_{i=1}^{n} E\left(u_{it,j}^2\right)}{\sum_{j=1}^{m} \sum_{i=1}^{n} E(\Delta w_{it,j} - \Delta \bar{w}_{iT,j})^2} \\ &= \frac{\sum_{i=1}^{n} tr\left(\mathbf{A}_i \mathbf{\Omega}_i \mathbf{A}_i'\right)}{\sum_{i=1}^{n} tr\left(\mathbf{A}_i \mathbf{\Omega}_i \mathbf{A}_i'\right) + \sum_{i=1}^{n} tr(\mathbf{V}_i)}, \end{aligned} \tag{S.16}$$

where $\mathbf{\Omega}_i = Var\left(\mathbf{B}_0' \mathbf{w}_{i,t-1}\right)$. Using (S.9) (when $\mathbf{\Theta}_i = \mathbf{0}$)we note that

$$\boldsymbol{\xi}_{it} = \mathbf{B}_0' \mathbf{w}_{i,t-1} = \mathbf{B}_0' \mathbf{d}_i + (\mathbf{I}_r - \mathbf{B}_0' \mathbf{A}_i) \boldsymbol{\xi}_{i,t-1} + \mathbf{B}_0' \mathbf{u}_{it},$$

and $\mathbf{\Omega}_i$ is given by

$$\mathbf{\Omega}_i = (\mathbf{I}_r - \mathbf{B}_0' \mathbf{A}_i) \mathbf{\Omega}_i (\mathbf{I}_r - \mathbf{B}_0' \mathbf{A}_i)' + \mathbf{B}_0' \mathbf{V}_i \mathbf{B}_0. \tag{S.17}$$

In the case where $r_0 = 1$, $\mathbf{\Omega}_i$ is a scalar, $\mathbf{B}_0 = \boldsymbol{\beta}_{1,0}$, and we have

$$\mathbf{\Omega}_i = \frac{\boldsymbol{\beta}_{1,0}' \mathbf{V}_i \boldsymbol{\beta}_{1,0}.}{1 - (1 - \rho_i)^2},$$

Hence

$$PR_n^2 = \frac{\sum_{i=1}^{n} \frac{\left(\mathbf{A}_i' \mathbf{A}_i\right) \boldsymbol{\beta}_{1,0}' \mathbf{V}_i \boldsymbol{\beta}_{1,0}}{1 - (1 - \rho_i)^2}}{\sum_{i=1}^{n} \frac{\left(\mathbf{A}_i' \mathbf{A}_i\right) \boldsymbol{\beta}_{1,0}' \mathbf{V}_i \boldsymbol{\beta}_{1,0}}{1 - (1 - \rho_i)^2} + \sum_{i=1}^{n} tr(\mathbf{V}_i)}.$$

For a given value of ρ_i, it is now possible to use $\mathbf{A}_i'\mathbf{A}_i = a_{i,11}^2 + \alpha_{i,21}^2 + \alpha_{i,31}^2$ as a scaling factor to achieve a desired value of PR_n^2. But we need to take account of the fact that $\rho_i = a_{i,11} - a_{i,31}$ and both $a_{i,11}$ and $a_{i,31}$ can not be scaled up. Given ρ_i we set

$$\mathbf{A}_i'\mathbf{A}_i = (\rho_i + a_{i,31})^2 + a_{i,21}^2 + a_{i,31}^2 = \varkappa^2,$$

and then derive $a_{i,21}$ and $a_{i,31}$ that satisfy the above equation. Then $\varkappa^2$ can be set in terms of PR_n^2 :

$$\varkappa^2 = \left(\frac{1 - PR_n^2}{PR_n^2}\right)\left(\frac{\sum_{i=1}^n tr(\mathbf{V}_i)}{\sum_{i=1}^n \frac{\boldsymbol{\beta}_{1,0}'\mathbf{V}_i\boldsymbol{\beta}_{1,0}}{1-(1-\rho_i)^2}}\right). \tag{S.18}$$

We can allow variations across $a_{i,31}$ and $a_{i,21}$, so long as $(\rho_i + a_{i,31})^2 + a_{i,21}^2 + a_{i,31}^2 = \varkappa^2$. To check the feasibility of the above procedure (results in real-valued $a_{i,j\ell}$) we set $a_{i,21} = 0$, and note that $a_{i,31}$ must now satisfy the quadratic equation $a_{i,31}^2 + \rho_i a_{i,31} + \frac{1}{2}(\rho_i^2 - \varkappa^2) = 0$. For this equation to have real solutions we must have $\rho_i^2 - 2(\rho_i^2 - \varkappa^2) > 0$, or $\varkappa^2 > \sup_i \rho_i^2/2$. We consider two values of $PR_n^2 = 0.20$ and 0.30.

S4.1.7 Case of $r_0 = 2$ long-run relations in VAR(1) design

Setting $\mathbf{Q}_i = (\mathbf{I}_{r_0} - \mathbf{B}_0'\mathbf{A}_i)$, and using (S.17), we have[17]

$$vec(\boldsymbol{\Omega}_i) = (\mathbf{Q}_i \otimes \mathbf{Q}_i)\, vec(\boldsymbol{\Omega}_i) + vec(\mathbf{B}_0'\mathbf{V}_i\mathbf{B}_0)$$

Then

$$vec(\boldsymbol{\Omega}_i) = \left[\mathbf{I}_{r_0^2} - (\mathbf{Q}_i \otimes \mathbf{Q}_i)\right]^{-1} vec(\mathbf{B}_0'\mathbf{V}_i\mathbf{B}_0).$$

Also since

$$tr(\mathbf{A}_i\boldsymbol{\Omega}_i\mathbf{A}_i') = tr(\boldsymbol{\Omega}_i\mathbf{A}_i'\mathbf{A}_i) = vec(\mathbf{A}_i'\mathbf{A}_i)'\, vec(\boldsymbol{\Omega}_i),$$

then

$$tr(\mathbf{A}_i\boldsymbol{\Omega}_i\mathbf{A}_i') = vec(\mathbf{A}_i'\mathbf{A}_i)'\left[\mathbf{I}_{r_0^2} - (\mathbf{Q}_i \otimes \mathbf{Q}_i)\right]^{-1} vec(\mathbf{B}_0'\mathbf{V}_i\mathbf{B}_0),$$

$$\mathbf{I}_{r_0^2} - (\mathbf{Q}_i \otimes \mathbf{Q}_i) = diag(1 - \rho_{i,11}^2, 1 - \rho_{i,11}\rho_{i,22}, 1 - \rho_{i,11}\rho_{i,22}, 1 - \rho_{i,22}^2).$$

[17] Note that $vec(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\, vec(\mathbf{B})$.

As before $\mathbf{A}_i'\mathbf{A}_i$ can be scaled up/down fixing the value of $\mathbf{Q}_i$. Recall from (S.11) that (since $\rho_{i,12} = (a_{i,12} - a_{i,32}) = 0$, and $\rho_{i,21} = (a_{i,21} - a_{i,31}) = 0$),

$$\boldsymbol{\Upsilon}_i = \begin{pmatrix} \rho_{i,11} & 0 \\ 0 & \rho_{i,22} \end{pmatrix}$$

where $\rho_{i,11} = (a_{i,11} - a_{i,31})$ and $\rho_{i,22} = (a_{i,32} - a_{i,22})$. Noting that

$$PR_n^2 = \frac{\sum_{i=1}^n tr\left(\mathbf{A}_i \boldsymbol{\Omega}_i \mathbf{A}_i'\right)}{\sum_{i=1}^n tr\left(\mathbf{A}_i \boldsymbol{\Omega}_i \mathbf{A}_i'\right) + \sum_{i=1}^n tr(\mathbf{V}_i)},$$

then

$$\frac{1 - PR_n^2}{PR_n^2} = \frac{\sum_{i=1}^n tr(\mathbf{V}_i)}{\sum_{i=1}^n vec\left(\mathbf{A}_i'\mathbf{A}_i\right)' \left[\mathbf{I}_{r^2} - (\mathbf{Q}_i \otimes \mathbf{Q}_i)\right]^{-1} vec\left(\mathbf{B}_0'\mathbf{V}_i\mathbf{B}_0\right)},$$

where

$$\begin{aligned} \mathbf{A}_i'\mathbf{A}_i &= \begin{pmatrix} a_{i,11} & a_{i,21} & a_{i,31} \\ a_{i,12} & a_{i,22} & a_{i,32} \end{pmatrix} \begin{pmatrix} a_{i,11} & a_{i,12} \\ a_{i,21} & a_{i,22} \\ a_{i,31} & a_{i,32} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{s=1}^3 a_{i,s1}^2 & \sum_{s=1}^3 a_{i,s1}a_{i,s2} \\ \sum_{s=1}^3 a_{i,s1}a_{i,s2} & \sum_{s=1}^3 a_{i,s2}^2 \end{pmatrix}. \end{aligned}$$

As before $\rho_{i,11}$ and $\rho_{i,22}$ are pre-set:

$$\text{Slow } \rho_{i,11} \sim IIDU\left[0.1, 0.2\right] \text{ ; Moderate } \rho_{i,11} \sim IIDU\left[0.1, 0.3\right],$$

$$\text{Slow } \rho_{i,22} \sim IIDU\left[0.1, 0.2\right] \text{ ; Moderate } \rho_{i,22} \sim IIDU\left[0.1, 0.3\right].$$

and $\rho_{i,12} = (a_{i,12} - a_{i,32}) = 0 = \rho_{i,21} = (a_{i,21} - a_{i,31}) = 0$. Overall, we have the following restrictions

$$\begin{aligned} a_{i,11} &= a_{i,31} + \rho_{i,11}, \text{ and } a_{i,22} = a_{i,32} + \rho_{i,22} \\ a_{i,12} &= a_{i,32} \text{ and } a_{i,21} = a_{i,31} \end{aligned}$$

We are left with two free parameters to control the fit of the model. To this end note that

$$\mathbf{A}_i'\mathbf{A}_i = \begin{pmatrix} (a_{i,31}+\rho_{i,11})^2+2a_{i,31}^2 & (a_{i,31}+\rho_{i,11})\, a_{i,32}+a_{i,31}\,(a_{i,32}+\rho_{i,22})+a_{i,31}a_{i,32} \\ (a_{i,31}+\rho_{i,11})\, a_{i,32}+a_{i,31}\,(a_{i,32}+\rho_{i,22})+a_{i,31}a_{i,32} & (a_{i,32}+\rho_{i,22})^2+2a_{i,32}^2 \end{pmatrix},$$

and $\mathbf{A}_i'\mathbf{A}_i$ can scaled up by scaling on $a_{i,31}$ and $a_{i,32}$. We set $a_{i,31} = a_{i,32} = \varkappa$ which yields

$$\mathbf{A}_i'\mathbf{A}_i = \begin{pmatrix} \rho_{i,11}^2+2\varkappa\rho_{i,11}+3\varkappa^2 & (\rho_{i,11}+\rho_{i,22})\,\varkappa+3\varkappa^2 \\ (\rho_{i,11}+\rho_{i,22})\,\varkappa+3\varkappa^2 & \rho_{i,22}^2+2\varkappa\rho_{i,22}+3\varkappa^2 \end{pmatrix}.$$

Then we solve the following equation for $\varkappa$ that yields a desired value of PR_n^2

$$\frac{1-PR_n^2}{PR_n^2} = \frac{\sum_{i=1}^n tr(\mathbf{V}_i)}{\sum_{i=1}^n vec\,(\mathbf{A}_i'\mathbf{A}_i)\,[\mathbf{I}_{r^2}-(\mathbf{Q}_i\otimes\mathbf{Q}_i)]^{-1}\, vec\,(\mathbf{B}_0'\mathbf{V}_i\mathbf{B}_0)},$$

where

$$[\mathbf{I}_{r^2}-(\mathbf{Q}_i\otimes\mathbf{Q}_i)]^{-1} = diag(\frac{1}{1-\rho_{i,11}^2}, \frac{1}{1-\rho_{i,11}\rho_{i,22}}, \frac{1}{1-\rho_{i,11}\rho_{i,22}}, \frac{1}{1-\rho_{i,22}^2}).$$

We consider the same two values of $PR_n^2 = 0.20$ and 0.30.

S4.1.8 Case of VARMA(1,1) design

When $\mathbf{\Theta}_i \neq \mathbf{0}$, it is cumbersome to solve for PR_n^2 analytically. We proceed by computing $\varkappa$ using stochastic simulations to ensure desired value of $PR_{RnT}^2(\varkappa)$, where

$$PR_{RnT}^2(\varkappa) = \frac{1}{R}\sum_{rep=1}^{R} PR_{nT}^2(rep,\varkappa),$$

and

$$PR_{nT}^2(rep,\varkappa) = 1 - \frac{\sum_{j=1}^m\sum_{t=1}^T\sum_{i=1}^n \left(u_{it,j}^{(rep)}\right)^2}{\sum_{j=1}^m\sum_{t=1}^T\sum_{i=1}^n (\Delta w_{it,j}^{(rep)} - \Delta\bar{w}_{iT,j}^{(rep)})^2},$$

in which we use $rep = 1, 2, ..., R$ to denote individual MC replications ($R = 2000$). We solve $PR_{RnT}^2(\varkappa) = 0.2$ or 0.3 using grid search method.

S4.2 Experiments with no long-run relations

Experiment with $I(1)$ variables and $r_0 = 0$ are based on the DGP given by

$$\Delta \mathbf{w}_{it} = \mathbf{\Phi}_i \Delta \mathbf{w}_{i,t-1} + \mathbf{u}_{it},$$

for $i = 1, 2, ..., n$, $t = 1, 2, ..., T$, where $\mathbf{u}_{it} = \mathbf{P}_i \boldsymbol{\varepsilon}_{it}$, $\boldsymbol{\varepsilon}_{it} \sim IIDN(\mathbf{0}, \mathbf{I}_m)$, the error covariance matrix $\mathbf{P}_i \mathbf{P}_i' \equiv \mathbf{\Sigma}_{ui} = [\sigma_{i,\ell q}]$, is generated using $\sigma_{i,\ell\ell} = 1$ for $i = 1, 2, ..., n$ and $\ell = 1, 2, 3$, and $\sigma_{i,\ell q} \sim IIDU(0, 0.5)$, for $\ell \neq q$, and $i = 1, 2, ..., n$. Matrix $\mathbf{\Phi}_i$ is diagonal with ϕ_{ij} elements on the diagonal, for $j = 1, 2, ..., m$. We consider three options for ϕ_{ij}: (i) low values $\phi_{ij} \sim U[0, 0.8]$, (ii) moderate values $\phi_{ij} \sim U[0.7, 0.9]$, and (iii) high values $\phi_{ij} \sim U[0.80, 0.95]$. Initial values are generated as $\Delta w_{i0,j} \sim IIDN\left[0, \left(1 - \phi_{ij}^2\right)^{-1}\right]$ for $j = 1, 2, ..., m$, and $i = 1, 2, ..., n$, and $\mathbf{w}_{i,-1}$ is set to $\mathbf{0}$. This DGP is also a special case of (1). Specifically, it leads to $\mathbf{w}_{it} = \mathbf{w}_{i0} + \mathbf{G}_i \mathbf{f}_t + \mathbf{C}_i \mathbf{s}_{it} + \mathbf{C}_i^*(L) \mathbf{u}_{it}$, where $\mathbf{C}_i = (\mathbf{I}_m - \mathbf{\Phi}_i)^{-1}$, and $\mathbf{C}_i^*(L) = -\mathbf{\Phi}_i (\mathbf{I}_m - \mathbf{\Phi}_i)^{-1} (\mathbf{I}_m - \mathbf{\Phi}_i L)^{-1}$.

S4.3 Experiments with interactive time effects

We augment the general linear process versions of the above VARMA and VAR specifications with $\mathbf{G}_i \mathbf{f}_t$, namely

$$\mathbf{w}_{it} = \mathbf{w}_{i0} + \mathbf{G}_i \mathbf{f}_t + \mathbf{C}_i \mathbf{s}_{it} + \mathbf{C}_i^*(L) \mathbf{u}_{it}, \tag{S.19}$$

where $\mathbf{C}_i$ and $\mathbf{C}_i^*(L)$ are obtained from the VARMA and VAR models above. $\mathbf{f}_t$ is an $m_f \times 1$ vector of unobserved common factors, $\mathbf{G}_i$ is an $m \times m_f$ matrix of factor loadings. We set $m_f = 4$ and generate latent factors, $\mathbf{f}_t$, to be serially correlated with a break,

$$\begin{aligned} \mathbf{f}_t &= \rho_{f_1} \mathbf{f}_{t-1} + \sqrt{1 - \rho_{f1}^2} \mathbf{v}_{ft}, \text{ for } t = 1, 2, ..., [T/2] - 1, \\ \mathbf{f}_t &= \rho_{f2} \mathbf{f}_{t-1} + \sqrt{1 - \rho_{f2}^2} \mathbf{v}_t, \text{ for } t = [T/2], [T/2] + 1, ..., T - 1, \end{aligned}$$

where $\rho_{f1} = 0.6$, $\rho_{f2} = 0.4$, and $\mathbf{v}_t \sim IIDN\left(\mathbf{0}, \mathbf{I}_{m_f}\right)$. Individual elements of $\mathbf{G}_i$ are generated as $IIDU\,[0.0.4]$.

S4.4 Design from Chudik, Pesaran and Smith (2021)

This design features $m = 2$ variables in $\mathbf{w}_{it} = (w_{1,it}, w_{2,it})'$, generated according to the cross-sectionally independent DGP described in Section 3.1 of Chudik, Pesaran, and Smith (2023a). In this design, $\mathbf{w}_{it}$ is generated as

$$\Delta w_{1,it} = c_i - a_i (w_{1,i,t-1} - w_{2,i,t-1}) + u_{1,it}, \tag{S.20}$$

$$\Delta w_{2,it} = u_{2,it}, \tag{S.21}$$

where $a_i \sim IIDU[0.2, 0.3]$, $u_{1,it} = \sigma_{1i} e_{1,it}$, $u_{2,it} = \sigma_{2i} e_{2,it}$, $\sigma^2_{1,i}, \sigma^2_{2,i} \sim IIDU[0.8, 1.2]$,

$$\begin{pmatrix} e_{1,it} \\ e_{2,it} \end{pmatrix} \sim IIDN(\mathbf{0}_2, \mathbf{\Sigma}_e), \; \mathbf{\Sigma}_e \sim \begin{pmatrix} 1 & \rho_{ei} \\ \rho_{ei} & 1 \end{pmatrix}, \text{ and } \rho_{ei} \sim IIDU[0.3, 0.7].$$

Full details of this design are provided in Section 3.1 of Chudik, Pesaran, and Smith (2023a).

S4.5 List of individual experiments

Overall, we conducted 71 experiments for the estimation of r_0 and 65 experiments for the estimation of $\mathbf{B}_0$. Table provide a summary of these experiments. Monte Carlo findings for individual experiments are available from authors upon request.

TABLE S1: List of Monte Carlo experiments

A. Experiments with no long-run relations used for estimation of r_0 only

Number of experiments	Interactive effects	$\phi_{i,\ell\ell}$
6	yes, no	low, moderate, high

B. Experiments with long-run relations and two-way long-run causality

Number of experiments	r_0	Interactive effects	Model	Error distribution	PR^2	Speed of convergence
64	1,2	yes, no	VAR(1), VARMA(1,1)	Gaussian, chi-squared	0.2, 0.3	moderate, slow

C. *Experiments with single long-run relation and one-way long-run causality*

Number of experiments	
1	Design from Chudik, Pesaran and Smith (2021)

Notes: This table lists 71 Monte Carlo experiments.

S5 Sensitivity of $\tilde{r}$ to scaling

It is clear that eigenvalues of $\mathbf{Q}_{\bar{w}\bar{w}}$ depend on the scale of the observations, $\mathbf{w}_{it} = (w_{it,1}, w_{it,2}, ..., w_{it,m})'$, and some form of scaling of data is required to reduce or eliminate the sensitivity of the estimator of r_0 to scaling. In Section 8 in the body of the paper, we proposed using eigenvalues of the correlation matrix $\mathbf{R}_{\bar{w}\bar{w}}$ given by (54), which, for convenience, we reproduce here:

$$\mathbf{R}_{\bar{w}\bar{w}} = \left[diag\left(\mathbf{Q}_{\bar{w}\bar{w}}\right)\right]^{-1/2} \mathbf{Q}_{\bar{w}\bar{w}} \left[diag\left(\mathbf{Q}_{\bar{w}\bar{w}}\right)\right]^{-1/2}.$$

Accordingly, we defined in equation (55) the following estimator of r_0,

$$\tilde{r} = \sum_{j=1}^{m} \mathcal{I}\left(\tilde{\lambda}_j < T^{-\delta}\right),$$

where $\tilde{\lambda}_j$, for $j = 1, 2, ..., m$ are the eigenvalues of $\mathbf{R}_{\bar{w}\bar{w}}$.

Consider the scaled vector $\dot{\mathbf{w}}_{it} = \mathbf{D}\mathbf{w}_{it}$, where $\mathbf{D} = diag\left(d_{11}, d_{22}, ..., d_{mm}\right)$ is a diagonal $m \times m$ scaling matrix with $d_{kk} > 0$ for all $k = 1, 2, ..., m$. Then matrix $\dot{\mathbf{R}}_{\bar{w}\bar{w}}$ computed based on scaled variables $\dot{\mathbf{w}}_{it}$ remains unaffected by the scaling matrix $\mathbf{D}$, namely $\mathbf{R}_{\bar{w}\bar{w}} = \dot{\mathbf{R}}_{\bar{w}\bar{w}}$. However, this no longer true when differential scaling is considered, given by $\ddot{\mathbf{w}}_{it} = \mathbf{D}_i\mathbf{w}_{it}$, where $\mathbf{D}_i = diag\left(d_{i,11}, d_{i,22}, ..., d_{i,mm}\right)$, for $i = 1, 2, ..., n$, are unit-specific diagonal scaling matrices with $d_{i,kk} > 0$ for all $k = 1, 2, ..., m$ and $i = 1, 2, ..., n$. In this case, matrix $\ddot{\mathbf{R}}_{\bar{w}\bar{w}}$, computed based on $\ddot{\mathbf{w}}_{it}$, is affected by scaling, namely $\ddot{\mathbf{R}}_{\bar{w}\bar{w}} \neq \mathbf{R}_{\bar{w}\bar{w}}$ in the general case where $\mathbf{D}_i$ differs over i.

To investigate impact of differential scaling given by $\ddot{\mathbf{w}}_{it} = \mathbf{D}_i \mathbf{w}_{it}$ on the small sample performance of $\tilde{r}$, we consider additional Monte Carlo experiments below. We generate $\mathbf{D}_i = \varkappa_i \mathbf{I}_3$, for $i = 1, 2, ..., n$, where $\varkappa_i \sim IIDU[1, 2]$. We compare the small sample performance of $\tilde{r}$ based on the original data $\mathbf{w}_{it}$ and the scaled data $\ddot{\mathbf{w}}_{it} = \mathbf{D}_i \mathbf{w}_{it} = \varkappa_i \mathbf{w}_{it}$, for three experiments. Table S2 reports selection frequencies for the estimation of r_0 using $\tilde{r}$ based on original and scaled data as well as two choices of $q = 2$ and 4 and two choices of exponent $\delta = 1/4$ and $1/2$, in experiments with $r_0 = 0$ (no long-run relations), given by the first-differenced VAR(1) design (60) with high persistence $\phi_{i,\ell\ell} \sim U[0.80, 0.95]$. Table S3 reports the same set of results in the case of VAR(1) experiments with $r_0 = 1$ long-run relation, interactive effects, low speed of convergence, low value of $PR^2 = 0.2$ and chi-squared distributed errors. Table S4 below reports findings for VAR(1) experiments with $r_0 = 2$ long-run relations, interactive effects, low speed of convergence, low value of $PR^2 = 0.2$ and chi-squared distributed errors. Overall, these Monte Carlo findings show a very small dependence of $\tilde{r}$ on the adopted differential scaling. In addition, these findings also show that performance of $\tilde{r}$ is almost identical for the two choices of $q = 2$ and 4.

TABLE S2: Selection frequencies for the estimation of r_0 by eigenvalue thresholding estimator $\tilde{r}$ with $\delta = 1/4$ and $1/2$, $q = 2$ and 4, and using original and scaled data, in experiments with no long-run relation ($r_0 = 0$), $\phi_{ij} \sim U[0.8, 0.95]$, and with interactive time effects

	Frequency $\tilde{r}=0$			Frequency $\tilde{r}=1$			Frequency $\tilde{r}=2$			Frequency $\tilde{r}=3$		
$n \setminus T$	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**
A. $\tilde{r}$ is computed using original data w_{it}												
	Estimator $\tilde{r}$ with $\delta = 1/4$ and $q = 2$											
50	0.95	1.00	1.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
500	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/4$ and $q = 4$											
50	0.97	1.00	1.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
500	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/2$ and $q = 2$											
50	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
500	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/2$ and $q = 4$											
50	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
500	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B. $\tilde{r}$ is computed using scaled data $\varkappa_i \mathbf{w}_{it}$, $\varkappa_i \sim IIDU[1,2]$												
	Estimator $\tilde{r}$ with $\delta = 1/4$ and $q = 2$											
50	0.92	1.00	1.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
500	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/4$ and $q = 4$											
50	0.95	1.00	1.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
500	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/2$ and $q = 2$											
50	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
500	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/2$ and $q = 4$											
50	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
500	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes: This table reports selection frequencies for the number of estimated long-run relations. r_0 denotes the true number of long-run relations. $\tilde{r}$ is given by (55), namely $\tilde{r} = \sum_{j=1}^{m} \mathcal{I}\left(\tilde{\lambda}_j < T^{-\delta}\right)$, with $\tilde{\lambda}_j$, $j = 1, 2, ..., m$, being the eigenvalues of $\mathbf{R}_{\bar{w}\bar{w}}$ defined by (54). See Subsection 9.1 in the paper for a summary of the design and Section S4 in the supplement for the full description. Reported results are based on $R = 2000$ MC replications.

TABLE S3: Selection frequencies for the estimation of r_0 by eigenvalue thresholding estimator $\tilde{r}$ with $\delta = 1/4$ and $1/2$, $q = 2$ and 4, and using original and scaled data, in VAR(1) experiments with $r_0 = 1$ long-run relation, slow speed of convergence toward long run, chi-squared distributed errors, $PR^2_{nT} = 0.2$, and with interactive time effects

	Frequency $\tilde{r}=0$			Frequency $\tilde{r}=1$			Frequency $\tilde{r}=2$			Frequency $\tilde{r}=3$		
$n \setminus T$	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**
A. $\tilde{r}$ is computed using original data w_{it}												
	Estimator $\tilde{r}$ with $\delta = 1/4$ and $q = 2$											
50	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
500	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/4$ and $q = 4$											
50	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
500	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/2$ and $q = 2$											
50	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
500	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/2$ and $q = 4$											
50	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
500	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
B. $\tilde{r}$ is computed using scaled data $\varkappa_i \mathbf{w}_{it}$, $\varkappa_i \sim IIDU[1,2]$												
	Estimator $\tilde{r}$ with $\delta = 1/4$ and $q = 2$											
50	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
500	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/4$ and $q = 4$											
50	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
500	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/2$ and $q = 2$											
50	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
500	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/2$ and $q = 4$											
50	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
500	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes: See notes to Table S2.

TABLE S4: Selection frequencies for the estimation of r_0 by eigenvalue thresholding estimator $\tilde{r}$ with $\delta = 1/4$ and $1/2$, $q = 2$ and 4, and using original and scaled data, in VAR(1) experiments with $r_0 = 2$ long-run relations, slow speed of convergence toward long run, chi-squared distributed errors, $PR^2_{nT} = 0.2$, and with interactive time effects

	Frequency $\tilde{r} = 0$			Frequency $\tilde{r} = 1$			Frequency $\tilde{r} = 2$			Frequency $\tilde{r} = 3$		
$n \setminus T$	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**
A. $\tilde{r}$ is computed using original data w_{it}												
	Estimator $\tilde{r}$ with $\delta = 1/4$ and $q = 2$											
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/4$ and $q = 4$											
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/2$ and $q = 2$											
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/2$ and $q = 4$											
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
B. $\tilde{r}$ is computed using scaled data $\varkappa_i \mathbf{w}_{it}$, $\varkappa_i \sim IIDU[1,2]$												
	Estimator $\tilde{r}$ with $\delta = 1/4$ and $q = 2$											
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/4$ and $q = 4$											
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/2$ and $q = 2$											
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
	Estimator $\tilde{r}$ with $\delta = 1/2$ and $q = 4$											
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00

Notes: See notes to Table S2.

S6 Robustness of PME estimators to GARCH and threshold autoregressive effects

We also conduct Monte Carlo experiments to investigate robustness of PME estimators to GARCH and threshold autoregressive (TAR) effects. To this end, we replace $\mathbf{u}_{it}$ with

$$\boldsymbol{\xi}_{it} = \Phi^{+}\mathbf{I}\left(\boldsymbol{\xi}_{i,t-1} > 0\right)\boldsymbol{\xi}_{i,t-1} + \Phi^{-}\mathbf{I}\left(\boldsymbol{\xi}_{i,t-1} \leq 0\right)\boldsymbol{\xi}_{i,t-1} + \mathbf{v}_{it},$$

for $t = -50, -49, ..., 0, 1, 2, ..., T$, with initial values $\boldsymbol{\xi}_{i,-51} = 0$,

$$\mathbf{I}\left(\boldsymbol{\xi}_{i,t-1} > 0\right) = \begin{pmatrix} I\left(\xi_{it,1} > 0\right) & 0 & 0 \\ 0 & I\left(\xi_{it,2} > 0\right) & 0 \\ 0 & 0 & I\left(\xi_{it,3} > 0\right) \end{pmatrix},$$

$I\left(.\right)$ is an indicator function,

$$\Phi^{-} = \phi^{-}\mathbf{I}_3, \text{ and } \Phi^{+} = \phi^{+}\mathbf{I}_3.$$

$\mathbf{v}_{it} = \left(v_{it,1}, v_{it,2}, v_{it,3}\right)'$ is generated according to $\mathbf{v}_{it} = \mathbf{D}_{it}\mathbf{P}_i\boldsymbol{\varepsilon}_{it}$, where $\boldsymbol{\varepsilon}_{it}$ is generated as before, namely individual elements are either $IIDN\left(0,1\right)$, or generated as IID from a chi-squared distribution with 4 degrees of freedom normalized to mean zero and unit variance, namely $\left[\chi^2\left(4\right) - 4\right]/\sqrt{8}$. $\mathbf{P}_i$ is Choleski factor of $\mathbf{R}_i = \mathbf{P}_i\mathbf{P}_i'$, and

$$\mathbf{R}_i \sim \begin{pmatrix} 1 & \rho_{\varepsilon i,12} & \rho_{\varepsilon i,13} \\ \rho_{\varepsilon i,12} & 1 & \rho_{\varepsilon i,23} \\ \rho_{\varepsilon i,13} & \rho_{\varepsilon i,23} & 1 \end{pmatrix}, \; \rho_{\varepsilon i,pq} \sim IIDU\left[0, 0.5\right],$$

for $\left(p,q\right) \in \left\{\left(1,2\right), \left(1,3\right), \left(2,3\right)\right\}$, and $i = 1, 2, ..., n$. We generate $\mathbf{D}_{it}$, for $i = 1, 2, ..., n$, and $t = -50, -49, ..., 0, 1, 2, ..., T$, as

$$\mathbf{D}_{it} = \begin{pmatrix} \sigma_{it,1} & 0 & 0 \\ 0 & \sigma_{it,2} & 0 \\ 0 & 0 & \sigma_{it,3} \end{pmatrix},$$

where

$$\sigma^2_{it,q} = \left(1 - \psi_{i1,q} - \psi_{i2,q}\right) + \psi_{i1,q}\varepsilon^2_{i,t-1,q} + \psi_{i1,q}\sigma^2_{i,t-1,q},$$

for $q = 1, 2, 3,$ and $t = -50, -49, ..., 0, 1, 2, ..., T$, with initial values $\sigma^2_{i,-51,q} = 1$.

Experiments with GARCH effects are given by

$$\psi_{i1,q} \sim IIDU\left(0.2, 0.3\right), \, \psi_{i2,q} \sim IIDU\left(0.3, 0.6\right),$$

for $q = 1, 2, 3$ and $i = 1, 2, ..., n$. Experiments with TAR effects are given by $\phi^- = 0.2$ and $\phi^+ = 0.6$.

Overall, we consider baseline experiments without GARCH and TAR effects ($\psi_{i1,q} = \psi_{i2,q} = \phi^- = \phi^+ = 0$), experiments with GARCH effects, experiments with TAR effects, and experiments featuring both GARCH and TAR effects. Tables S5 and S6 report Monte Carlo findings for three-variable VARMA(1,1) as the DGP with $r_0 = 2$ long-run rlations, chi-squared distributed errors, $PR^2_{nT} = 0.2$, moderate speed of convergence toward long run, and with interactive time effects.

TABLE S5: Selection frequencies for the estimation of $r_0 = 2$ by eigenvalue thresholding estimator, $\tilde{r}$ with $\delta = 1/4$ and $1/2$ using three-varaible $VARMA(1,1)$ as the DGP with $r_0 = 2$ long-run relations, chi-squared distributed errors, $PR_{nT}^2 = 0.2$, moderate speed of convergence toward long run, and with interactive time effects.

	Frequency $\tilde{r}=0$			Frequency $\tilde{r}=1$			Frequency $\tilde{r}=2$		
$n \setminus T$	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**
A. Baseline experiments without GARCH and without TAR effects									
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/4$								
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/2$								
50	0.00	0.00	0.00	0.05	0.00	0.00	0.95	1.00	1.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
B. Experiments with GARCH effects									
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/4$								
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/2$								
50	0.00	0.00	0.00	0.07	0.00	0.00	0.93	1.00	1.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
C. Experiments with TAR effects									
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/4$								
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/2$								
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
D. Experiments with GARCH and TAR effects									
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/4$								
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
	Correlation matrix eigenvalue thresholding estimator $\tilde{r}$, with $\delta = 1/2$								
50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
500	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
1,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
3,000	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00

Notes: This table reports selection frequencies for the number of estimated long-run relations. r_0 denotes the true number of long-run relations. $\tilde{r}$ is given by (55), namely $\tilde{r} = \sum_{j=1}^{m} \mathcal{I}\left(\tilde{\lambda}_j < T^{-\delta}\right)$, with $\tilde{\lambda}_j$, $j = 1, 2, ..., m$, being the eigenvalues of $\mathbf{R}_{\tilde{w}\tilde{w}}$ defined by (54). See Sections S4 and S6 in the supplement for the full description of the design. Reported results are based on $R = 2000$ MC replications.

TABLE S6: Simulated bias, RMSE, size and power for the PME estimation of $\beta_{13,0}$ in three variable VARMA(1,1) experiments with $r_0 = 2$ long run relations, chi-squared distributed errors, $PR^2_{nT} = 0.2$, moderate speed of convergence toward long run, and with interactive time effects

	Bias ($\times 100$)			**RMSE** ($\times 100$)			**Size** ($\times 100$)			**Power** ($\times 100$)		
$n\backslash T$	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**	**20**	**50**	**100**
A. Baseline experiments without GARCH and without TAR effects												
PME estimator with $q = 2$ sub-samples												
50	-1.12	-0.38	-0.20	8.60	4.85	2.91	7.95	7.20	7.65	10.50	13.40	25.95
500	-0.59	-0.34	-0.15	2.76	1.58	0.90	6.85	5.65	5.70	28.10	59.50	94.20
1,000	-0.44	-0.28	-0.14	1.95	1.13	0.66	6.35	6.05	6.00	46.60	85.55	99.70
3,000	-0.59	-0.29	-0.11	1.22	0.70	0.38	9.05	7.15	5.35	92.10	99.95	100.00
PME estimator with $q = 4$ sub-samples												
50	-1.86	-0.90	-0.30	7.00	3.97	2.18	6.50	6.50	7.00	12.55	19.30	38.25
500	-1.48	-0.81	-0.23	2.62	1.47	0.70	10.50	9.55	5.90	52.95	86.95	99.60
1,000	-1.39	-0.75	-0.22	2.07	1.14	0.53	14.70	12.20	7.55	81.00	98.90	100.00
3,000	-1.51	-0.76	-0.21	1.74	0.91	0.34	38.25	29.70	11.60	99.95	100.00	100.00
B. Experiments with GARCH effects												
PME estimator with $q = 2$ sub-samples												
50	-1.55	-0.45	-0.25	9.95	5.29	3.07	9.05	8.80	8.15	11.70	14.45	26.90
500	-0.54	-0.32	-0.15	3.28	1.76	0.94	6.70	6.35	4.65	22.85	51.05	90.55
1,000	-0.43	-0.32	-0.13	2.33	1.27	0.71	6.90	6.45	6.65	34.50	77.15	99.50
3,000	-0.58	-0.31	-0.11	1.46	0.79	0.41	7.60	7.50	5.90	76.10	99.00	100.00
PME estimator with $q = 4$ sub-samples												
50	-2.23	-0.97	-0.34	8.22	4.37	2.30	7.80	7.65	7.85	14.10	19.30	36.95
500	-1.45	-0.82	-0.25	2.97	1.61	0.75	8.95	9.70	6.90	41.60	77.35	98.90
1,000	-1.35	-0.78	-0.22	2.30	1.25	0.56	11.20	12.20	6.85	63.25	96.00	100.00
3,000	-1.50	-0.77	-0.21	1.83	0.96	0.36	27.50	26.40	10.20	97.50	100.00	100.00
C. Experiments with TAR effects												
PME estimator with $q = 2$ sub-samples												
50	1.56	-0.12	-0.08	6.54	3.50	1.71	9.60	7.90	8.55	7.90	17.70	50.25
500	2.78	0.30	0.05	3.44	1.11	0.53	30.40	5.90	4.80	6.30	71.10	100.00
1,000	2.84	0.33	0.08	3.18	0.82	0.39	52.15	6.90	5.70	5.95	94.25	100.00
3,000	2.74	0.31	0.07	2.86	0.55	0.23	92.00	12.10	6.15	6.65	100.00	100.00
PME estimator with $q = 4$ sub-samples												
50	1.48	-0.36	-0.12	5.38	3.02	1.43	9.50	7.40	8.45	7.10	22.95	64.40
500	2.64	-0.03	0.06	3.12	0.91	0.43	36.70	3.95	4.80	6.00	90.80	100.00
1,000	2.68	0.01	0.07	2.93	0.65	0.32	63.40	4.95	5.60	6.15	99.35	100.00
3,000	2.58	0.01	0.07	2.67	0.38	0.19	96.75	5.25	6.10	9.50	100.00	100.00
D. Experiments with GARCH and TAR effects												
PME estimator with $q = 2$ sub-samples												
50	1.32	-0.09	-0.06	6.80	3.45	1.66	8.35	8.20	6.85	7.80	18.55	51.60
500	2.64	0.34	0.07	3.41	1.12	0.52	23.95	6.25	4.70	6.55	70.65	100.00
1,000	2.66	0.34	0.09	3.04	0.82	0.39	42.45	7.30	6.70	5.55	94.05	100.00
3,000	2.54	0.31	0.07	2.69	0.55	0.22	82.10	10.85	5.90	8.35	100.00	100.00
PME estimator with $q = 4$ sub-samples												
50	1.23	-0.32	-0.12	5.52	2.97	1.37	7.25	7.00	7.20	7.00	22.90	66.10
500	2.46	0.01	0.07	3.02	0.92	0.43	28.00	4.90	5.40	5.50	89.20	100.00
1,000	2.48	0.04	0.08	2.76	0.65	0.32	50.15	4.90	7.05	5.70	99.40	100.00
3,000	2.36	0.02	0.07	2.47	0.38	0.19	91.05	5.05	6.40	13.00	100.00	100.00

Notes: The long-run relations are given by $\boldsymbol{\beta}'_{1,0}\mathbf{w}_{it} = w_{it,1} - w_{it,3}$ and $\boldsymbol{\beta}'_{2,0}\mathbf{w}_{it} = w_{it,2} - w_{it,3}$, and identified using $\beta_{11,0} = \beta_{22,0} = 1$, and $\beta_{12,0} = \beta_{21,0} = 0$. Reported results are based on $R = 2,000$ Monte Carlo replications. Simulated power are computed under $H_1 : \beta_{13} = -0.97$, as alternatives to -1 under the null. A detailed account of the data generating processes provided in Section S4 and S6 of this supplement. Size and Power are computed at 5 percent nominal level.

S7 Supplementary information for micro application

S7.1 Data sources

All data is from Wharton Research Data Services, WRDS, available at wrds.wharton.upenn.edu, accessed on 2023-04-20. Specifically, our variables come from annual "CRSP/Compustat Merged" database available through WRDS. We follow literature (Geelen et al. (2024)) in constructing the variables listed in Table S5.

TABLE S5: *Variable description**

Variable	Definition	Compustat item
MV	Market value	CSHO × PRCC_F
BV	Book value	See Note 1
DO	Total debt outstanding, DO=SD+LT	DLC+DLTT
SD	Short-term debt	DLC
LD	Long-term debt	DLTT
TA	Total assets	AT

Notes: (*) The source for all variables is Wharton Research Data Services, wrds.wharton.upenn.edu, accessed on 2023-04-20. Variables come from annual "CRSP/Compustat Merged" database available through WRDS.
(1) Book value = the book equity of shareholders + compustat item $TXDITC$ - book value of preferred stocks. The book equity of shareholders is given by (in sequential order, depending on availability): (*i*) compustat item SEQ, or (*ii*), common equity (compustat item CEQ) + par value of preferred stock (compustat item PSTK), or (*iii*) total assets (compustat item AT) - total liabilities (compustat item LT). The book value of preferred stocks is computed as (in sequential order, based on availability): (*i*) the redemption value (compustat item PSTKRV), or (*ii*) the liquidation value (compustat item PSTKL), or (*iii*) the par value (compustat item PSTK).

S7.2 Data availability, variable construction and summary statistics

Data availability is summarized in Table S6, for the three variable sets we considered: {DO, TA}, {SD, LD, TA},{DO, BV, MV}. We apply following data filters (in sequential order) to each of the variable set and data sample separately.

Filter 1. We omit firms with gaps in the data for the given variable set and sample, and firms with $T_i < 20$ (unbalanced samples).

Filter 2. We omit firms with nonpositive entries on any of the variable in the given set and sample.

Filter 3. We omit firms where, for a given sample and variable set, average value of key ratios fall below 1 or above 99 percentiles estimated after the application of the first two filters.

Filter 1 ensures that sufficient data exists, Filter 2 excludes firms with negative or zero values, and Filter 3 is the outlier filter based on percentiles of key ratios. Tables S3-S5 report summary statistics for each variable set after all filters were applied.

TABLE S6: Number of firms with available data

	1950-2021	1950-2010
Variable set DO, TA		
Filter 1	4193	3010
Filter 1, 2	2546	1909
Filter 1, 2, 3	2531	1901
Variable set SD, LD, TA		
Filter 1	4193	3010
Filter 1, 2	1379	1110
Filter 1, 2, 3	1365	1101
Variable set DO, BV, MV		
Filter 1	2907	2196
Filter 1, 2	1419	1172
Filter 1, 2, 3	1403	1164

Notes: Variable definitions are provided in Table S5. Filter 1 omits firms with gaps in the data for the given variable set and sample, and firms with fewer than 20 years. Filter 2 omits firms with nonpositive entries. Filter 3, omits firms where average value of key ratios fall below 1 or above 99 percentiles after the application of the first two filters

TABLE S7: Summary Statistics (min, mean, median, max) for individual variables and ratios in the variable set {DO, TA} after the application of all filters.

	1950-2021	1950-2010
Variable DO		
min	0.001	0.001
median	160.4	94.77
mean	3,405	1,837
max	889,300	889,300
Variable TA		
min	0.283	0.126
median	689.8	371.7
mean	13,809	6,621
max	3,743,567	3,001,251
Ratio DO/TA		
min	0.000	0.000
median	0.270	0.276
mean	0.290	0.294
max	4.434	6.789

Notes: Unit for the top part of this table reporting the summary statistics for individual variables is million U.S. dollars.

TABLE S8: Summary Statistics (min, mean, median, max) for individual variables and ratios in the variable set {SD, LD, TA} after the application of all filters.

	1950-2021	1950-2010
Variable SD		
min	0.001	0.001
median	21.77	12.98
mean	2,081	1,071
max	614,237	562,857
Variable LD		
min	0.001	0.001
median	133.4	78.77
mean	2,882	1,453
max	486,876	486,876
Variable TA		
min	0.652	0.652
median	795.7	419.2
mean	19,510	8,482
max	3,743,567	2,264,909
Ratio SD/TA		
min	0.000	0.000
median	0.039	0.041
mean	0.066	0.067
max	1.585	1.585
Ratio LD/TA		
min	0.000	0.000
median	0.215	0.217
mean	0.233	0.233
max	1.752	1.752

Notes: Unit for the top part of this table reporting the summary statistics for individual variables is million U.S. dollars.

TABLE S9: Summary Statistics (min, mean, median, max) for individual variables and ratios in the variable set {DO, BV, MV} after the application of all filters.

	1950-2021	1950-2010
Variable DO		
min	0.002	0.002
median	138.9	91.61
mean	2,178	1,212
max	889,300	889,300
Variable BV		
min	0.079	0.079
median	269.2	178.7
mean	2,752	1,594
max	595,878	212,294
Variable MV		
min	0.177	0.177
median	377.0	227.4
mean	5,007	2,908
max	662,627	504,240
Ratio DO/MV		
min	0.000	0.000
median	0.455	0.486
mean	0.832	0.869
max	63.26	63.26
Ratio BV/MV		
min	0.002	0.003
median	0.773	0.833
mean	0.997	1.066
max	27.85	31.89

Notes: Unit for the top part of this table reporting the summary statistics for individual variables is million U.S. dollars.

TABLE S10: *IPS unit root test results for 40-year balanced sample ending 2021*

Panel unit root test results						
Lag order:	$p=1$	$p=2$	$p=1$	$p=2$	$p=1$	$p=2$
A. Panel unit root test results for the variable set DO, TA ($n=336$)						
	DO		TA			
IPS stat.:	-1.534	-1.504	-1.45	-1.457		
B. Panel unit root test results for the variable set SD, LD, TA ($n=176$)						
	SD		LD		TA	
IPS stat.:	-2.344**	-2.067**	-1.492	-1.419	-1.561	-1.585
C. Panel unit root test results for the variable set DO, BV, MV ($n=175$)						
	DO		BV		MV	
IPS stat.:	-1.328	-1.215	-1.153	-1.165	-1.328	-1.351

Notes: This table reports IPS panel unit root test statistics by Im, Pesaran, and Shin (2003) for balanced sample using 40 years ending 2021. Rejections at 5 and 1 percent nominal level are highlighted by * and **, respectively.

S8 Supplementary information for macro applications

We use Penn World Table database, version 10.01, available at https://www.rug.nl/ggdc/productivity/pwt/. This dataset contains unbalanced annual macroeconomic data covering the period 1950−2019. The following variables are constructed (variable names correspond to the variable identifiers in PWT database)

1. (Exports per-capita and imports per-capita), *exppc=csh_x×rgdpna/pop and imppc=-csh_m×rgdpna/pop*

2. (Real wages and productivity per hour worked) *ewageph=labsh×rgdpna/(emp×avh)* and *prodph=rgdpna/(emp×avh)*

3. (Output and Capital in efficiency units) *connapc=(csh_c×rgdpna)/pop*, *invnapc=(csh_i×rgdpna)/pop*, *rgdpnapc=rgdpna/pop,*

where: *rgdpna* is Real GDP at constant 2017 national prices (in mil. 2017US$), *pop* is population (in millions), *csh_x* is share of merchandise exports at current PPPs, *csh_m* is share of merchandise imports at current PPPs*,* *labsh* is share of labour compensation in GDP at current national prices, *emp* is number of persons engaged (in millions), *avh* is average annual hours worked by persons engaged, *csh_ c* is share of household consumption at current PPPs, and *csh_ i* is Share of gross capital formation at current PPPs.

We use the following variables in our analysis: $ex_{it} = \ln(exppc_{it})$, $im_{it} = \ln(imppc_{it})$, $wage_{it} = ln(ewageph_{it})$, $prod_{it} = \ln(prodph_{it})$, $inv_{it} = \ln(invnapc_{it})$, $cons_{it} = \ln(connapc)$, and $gdp_{it} = \ln(rgdpnapc_{it})$.

For for each variable-pair, we apply the following filters : (A) Incldue countries with $T_i \geq 20$. (B) Drop countries with annual per capita observations that are below $\epsilon = 0.01$.