TI 2022-093/III
Tinbergen Institute Discussion Paper

Fully Modified Estimation in Cointegrating Polynomial Regressions: Extensions and Monte Carlo Comparison

Yicong Lin[1]
Hanno Reuvers[2]

1 Vrije Universiteit Amsterdam and Tinbergen Institute

2 Erasmus University Rotterdam

Fully Modified Estimation in Cointegrating Polynomial Regressions: Extensions and Monte Carlo Comparison*

Yicong Lin[†1,2] and Hanno Reuvers[3]

[1]Department of Econometrics and Data Science, Vrije Universiteit Amsterdam

[2]Tinbergen Institute

[3]Department of Econometrics, Erasmus University Rotterdam

December, 2022

Abstract

We study a set of fully modified (FM) estimators in multivariate cointegrating polynomial regressions. Such regressions allow for deterministic trends, stochastic trends, and integer powers of stochastic trends to enter the cointegrating relations. A new feasible generalized least squares estimator is proposed. Our estimator incorporates: (1) the inverse autocovariance matrix of multidimensional errors and (2) second-order bias corrections. The resulting estimator has the intuitive interpretation of applying a weighted least squares objective function to filtered data series. Moreover, the required second-order bias corrections are convenient byproducts of our approach and lead to a conventional asymptotic inference. Based on different FM estimators, multiple multivariate KPSS-type of tests for the null of cointegration are constructed. We then undertake a comprehensive Monte Carlo study to compare the performance of the FM estimators and the related tests. We find good performance of the proposed estimator and the implied test statistics for linear hypotheses and cointegration.

JEL Classification: C12, C13, C32

Keywords: Cointegrating Polynomial Regression, Cointegration Testing, Fully Modified Estimation, Generalized Least Squares

*An earlier version of this paper was circulated under the title "Efficient Estimation by Fully Modified GLS with an Application to the Environmental Kuznets Curve."

†Corresponding author: Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, the Netherlands. E-mail address: yc.lin@vu.nl.

1 Introduction

In recent years, there has been an increasing interest in the theoretical properties and theoretical justifications of nonlinear cointegrating relations. For theoretical properties, we refer to the textbook treatise by Wang (2015), the recent review article by Tjøstheim (2020), and the extensive references found in either of them. Theoretical justifications are in some cases refinements of existing economic theory, e.g., nonlinear cointegration among bond yields with different times to maturity due to yield-dependent risk premia as discussed in Breitung (2001), or nonlinear purchasing power parity due to transaction/transportation costs and trade barriers (e.g., Hong and Phillips 2010). In other cases, economic theory postulates a nonlinear cointegrating relation from the outset. A popular example of the latter is the environmental Kuznets curve described in Grossman and Krueger (1995).[1]

There are three branches of literature on the estimation of such nonlinear cointegrating relations. First, the papers by Park and Phillips (1999) and Park and Phillips (2001) are concerned with nonlinear cointegration analysis of a parametric form. Second, there is literature on nonparametric kernel estimation of nonlinear cointegrating equations, see for example Wang and Phillips (2009) or Li et al. (2020). The third approach is reminiscent of a nonparametric sieve estimation with a power polynomial basis. That is, one estimates a cointegrating relation containing integer powers of integrated regressors. Wagner and Hong (2016) named this a cointegrating polynomial regression (CPR). The multivariate seemingly unrelated regressions (SURs) extension is available in Wagner et al. (2020). This paper builds on this SUR setup.

We make two theoretical contributions to the literature on CPRs. First, we propose a Fully Modified Generalized Least Squares (FM-GLS) estimator. This estimator requires two main steps: (1) It employs the inverse covariance matrix of the $2nT$-dimensional innovation vector, that is, the covariance matrix of the vector which stacks the n disturbances in the cointegrating equations and the n disturbances driving the $I(1)$ regressors over the time span T. The estimation of this inverse covariance matrix is based on the Modified Cholesky Decomposition originating from Pourahmadi (1999). The approach is computationally simple because the required quantities are obtained from the coefficients and prediction error variances of best linear least squares predictors. Sufficient conditions for consistency are provided. (2) We exploit the previous results to correct the second-order biases, resulting in a conventional chi-squared inference. Also, note that the approach differs from the linear cointegration results in Mark et al. (2005) and Moon and Perron (2005)

[1]There is no direct reference to Kuznets in the original paper by Grossman and Krueger (1995). But their nonlinear relations between environmental indicators and per capita GDP do remind strongly of the inverted U-shaped between income inequality and economic growth proposed by Kuznets (1901-1985).

since our bias corrections do not rely on leads and lags augmentation. Second, a multi-equation cointegration specification asks for a multivariate cointegration test. Building upon the work by Choi and Saikkonen (2010), we consider three such tests based on different FM estimators. The first test uses pre-filtered residuals to account for serial correlation, whereas the other two are direct multivariate generalizations of the KPSS-type of test in Wagner and Hong (2016). The estimators and cointegration tests are subsequently compared by an extensive Monte Carlo simulation. In our simulations, the FM-GLS estimator has a higher estimation accuracy and its implied Wald test has better size control and higher size-adjusted power compared to the existing methods. We find by simulation that pre-filtering improves the size control of the cointegration tests but has an adverse effect on power.

The plan of this paper is as follows. Section 2 introduces the model and the modified Cholesky block decomposition. This decomposition is the main ingredient for the FM-GLS estimator. The related asymptotic theory and stationarity tests are discussed in Section 3, whereas a finite sample simulation study is presented in Section 4. Section 5 concludes. All proofs are collected in the Appendices.

Some words on notation. Throughout this paper, C denotes a generic positive constant. The integer part of the number $a \in \mathbb{R}^+$ is denoted by $[a]$. For a vector $\boldsymbol{x} \in \mathbb{R}^n$, its p-norm is denoted by $\|\boldsymbol{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$. When applied to a matrix, $\|\boldsymbol{A}\|_p$ signifies the induced norm, i.e., $\|\boldsymbol{A}\|_p = \sup_{\boldsymbol{x}\neq\boldsymbol{0}} \|\boldsymbol{A}\boldsymbol{x}\|_p/\|\boldsymbol{x}\|_p$. The subscripts are omitted whenever $p = 2$, e.g., $\|\boldsymbol{x}\| = (\sum_{i=1}^n |x_i|^2)^{1/2}$ and $\|\boldsymbol{A}\| = (\lambda_{max}(\boldsymbol{A}'\boldsymbol{A}))^{1/2}$ where $\lambda_{max}(\cdot)$ is the largest eigenvalue. Similarly, $\lambda_{min}(\cdot)$ denotes the smallest eigenvalue. The Frobenius norm is denoted as $\|\cdot\|_{\mathcal{F}}$. The n-dimensional identity matrix is written as $\boldsymbol{I}_n$. The ith row or i^{th} column of an arbitrary matrix $\boldsymbol{A}$ are selected using $\text{col}_i(\boldsymbol{A})$ and $\text{row}_i(\boldsymbol{A})$, respectively. The Kronecker product is denoted "$\otimes$". We use the symbol "$\Rightarrow$" to signify weak convergence and the symbol "$\stackrel{d}{=}$" for equality in distribution. The stochastic order and strict stochastic order relations are indicated by $O_p(\cdot)$ and $o_p(\cdot)$.

2 The model and infeasible GLS

We first introduce the adopted model specification and propose an infeasible GLS estimator. As seen in Section 3 later, we adopt a fully modified approach to eliminate the dependence of the asymptotic distribution of the infeasible GLS estimator on nuisance parameters, in which we also compare the existing FM estimators. As mentioned, we study a system of seemingly unrelated

cointegrating polynomial regressions (SUCPR) as in Wagner et al. (2020), that is,

$$\boldsymbol{y}_t = \boldsymbol{Z}_t'\boldsymbol{\beta} + \boldsymbol{u}_t, \qquad t = 1, 2, \ldots, T, \tag{2.1}$$

where the dependent variable $\boldsymbol{y}_t := [y_{1t}, y_{2t}, \ldots, y_{nt}]'$ and innovations $\boldsymbol{u}_t := [u_{1t}, u_{2t}, \ldots, u_{nt}]'$ are $(n \times 1)$ random vectors. For the cross-sectional unit i, we use as explanatory variables: (1) deterministic components such as an intercept and polynomial time trends up to order d_i, and (2) integer powers of the $I(1)$ regressors (denoted x_{it}) up to degree s_i. Defining $\boldsymbol{d}_{it} = [1, t, \ldots, t^{d_i}]'$, $\boldsymbol{s}_{it} = [x_{it}, \ldots, x_{it}^{s_i}]'$, and $\boldsymbol{z}_{it} = [\boldsymbol{d}_{it}', \boldsymbol{s}_{it}']'$, we subsequently collect all explanatory variables in the block diagonal matrix $\boldsymbol{Z}_t = \mathrm{diag}[\boldsymbol{z}_{1t}, \ldots, \boldsymbol{z}_{nt}]$. We are interested in the d-dimensional parameter vector $\boldsymbol{\beta}$ where $d = \sum_{i=1}^{n}(d_i + s_i + 1)$. Overall, each cross-sectional unit in (2.1) specifies a single cointegrating relation containing polynomials in deterministic and stochastic trends. For each i, the highest orders of these polynomials, i.e., d_i and s_i, are assumed to be fixed and known. We do not allow for cointegration in the cross-sectional dimension.

The innovation series $\{\boldsymbol{u}_t\}$ is allowed to exhibit dependencies over time and across series. We assume that these dependencies can be modeled by a stationary VAR(∞) process, that is

$$\boldsymbol{\mathcal{A}}(L)\boldsymbol{u}_t = \left(\boldsymbol{I}_n - \sum_{j=1}^{\infty} \boldsymbol{A}_j L^j\right)\boldsymbol{u}_t = \boldsymbol{\eta}_t, \tag{2.2}$$

see Assumption 1 for further details. Efficient estimation of the parameter vector $\boldsymbol{\beta}$ now requires the use of generalized least squares (GLS). Our Zellner (1962)-type GLS estimator relies on the inverse of the $(nT \times nT)$ matrix $\boldsymbol{\Sigma}_{\boldsymbol{u}} = \mathbb{E}(\boldsymbol{u}\boldsymbol{u}')$ where $\boldsymbol{u} = [\boldsymbol{u}_1', \boldsymbol{u}_2', \ldots, \boldsymbol{u}_T']'$. In this paper, we directly estimate $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}$ using a multivariate extension of the modified Cholesky decomposition by Pourahmadi (1999). This extension was named the *Modified Cholesky Block Decomposition* (MCBD) by Kim and Zimmerman (2012) and Kohli et al. (2016). The latter papers used the MCBD to parametrize the covariance matrix of multivariate longitudinal data. As in Beutner et al. (2022), we use the MCBD for the time series application mentioned above, i.e., the computation of $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}$. The decomposition is closely related to linear minimum MSE predictors. We define

$$\begin{aligned} \boldsymbol{A}(\ell) &= \begin{bmatrix} \boldsymbol{A}_1(\ell) & \cdots & \boldsymbol{A}_\ell(\ell) \end{bmatrix} = \underset{(\boldsymbol{\Theta}_1,\ldots,\boldsymbol{\Theta}_\ell)\in\mathbb{R}^{n\times n\ell}}{\arg\min} \mathbb{E}\left\|\boldsymbol{u}_t - \boldsymbol{\Theta}_1\boldsymbol{u}_{t-1} - \cdots - \boldsymbol{\Theta}_\ell\boldsymbol{u}_{t-\ell}\right\|^2, \\ \boldsymbol{S}(\ell) &= \mathbb{E}\left[\boldsymbol{u}_t - \boldsymbol{A}_1(\ell)\boldsymbol{u}_{t-1} - \cdots - \boldsymbol{A}_\ell(\ell)\boldsymbol{u}_{t-\ell}\right]\left[\boldsymbol{u}_t - \boldsymbol{A}_1(\ell)\boldsymbol{u}_{t-1} - \cdots - \boldsymbol{A}_\ell(\ell)\boldsymbol{u}_{t-\ell}\right]', \end{aligned} \tag{2.3}$$

and $\boldsymbol{S}(0) = \mathbb{E}(\boldsymbol{u}_t\boldsymbol{u}_t')$. The inverse of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ is then given by $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1} = \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}'\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}$,

where $\boldsymbol{\mathcal{S}}_{\boldsymbol{u}} = \text{diag}\Big[\boldsymbol{S}(0), \boldsymbol{S}(1), \ldots, \boldsymbol{S}(T-1)\Big]$,

$$\boldsymbol{\mathcal{M}}_{\boldsymbol{u}} = \left[\boldsymbol{m}_{\boldsymbol{u}}^{ij}\right]_{1\le i,j\le T}, \text{ with } \boldsymbol{m}_{\boldsymbol{u}}^{ij} = \begin{cases} \mathbf{O}_{n\times n}, & \text{if} \quad i<j, \\ \boldsymbol{I}_n, & \text{if} \quad i=j, \\ -\boldsymbol{A}_{i-j}(i-1), & \text{if} \quad 2\le i\le T,\ 1\le j\le i-1, \end{cases} \tag{2.4}$$

and the $\boldsymbol{A}_j(i)$ follow from the partitioning $\boldsymbol{A}(\ell) = \big[\boldsymbol{A}_1(\ell), \ldots, \boldsymbol{A}_\ell(\ell)\big]$.

Weak stationarity of $\{\boldsymbol{u}_t\}$ implies that the block elements of $\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}$ being far below the main diagonal are small. This suggests a banding approach in which small elements are replaced by zeros. More specifically, we construct a *Banded Inverse Autocovariance Matrix* (BIAM) as

$$\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q) = \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}'(q)\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q)\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q), \tag{2.5}$$

where $1 \le q \ll T$ is called the banding parameter, $\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}(q) = \text{diag}\Big[\boldsymbol{S}(0), \boldsymbol{S}(1), \ldots, \boldsymbol{S}(q), \ldots, \boldsymbol{S}(q)\Big]$ and $\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q) = [\boldsymbol{m}_{\boldsymbol{u}}^{ij}(q)]_{1\le i,j\le T}$ with

$$\boldsymbol{m}_{\boldsymbol{u}}^{ij}(q) = \begin{cases} \mathbf{O}_{n\times n}, & \text{if} \quad i<j \text{ or } \{q+1<i\le T,\ 1\le j\le i-q-1\}, \\ \boldsymbol{I}_n, & \text{if} \quad i=j, \\ -\boldsymbol{A}_{i-j}(i-1), & \text{if} \quad 2\le i\le q,\ 1\le j\le i-1, \\ -\boldsymbol{A}_{i-j}(q), & \text{if} \quad q+1\le i\le T,\ i-q\le j\le i-1. \end{cases} \tag{2.6}$$

Example 1

Consider a weakly stationary n-dimensional VAR(3) process specified as $\boldsymbol{u}_t = \sum_{j=1}^{3}\boldsymbol{A}_j\boldsymbol{u}_{t-j} + \boldsymbol{\eta}_t$ with $\boldsymbol{\eta}_t \overset{i.i.d.}{\sim} (\mathbf{0}, \boldsymbol{\Sigma}_{\eta\eta})$. For $T=4$, the MCBD $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1} = \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}'\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}$ is based on

$$\boldsymbol{\mathcal{M}}_{\boldsymbol{u}} = \begin{bmatrix} \boldsymbol{I}_n & & & \\ -\boldsymbol{A}_1(1) & \boldsymbol{I}_n & & \\ -\boldsymbol{A}_2(2) & -\boldsymbol{A}_1(2) & \boldsymbol{I}_n & \\ -\boldsymbol{A}_3 & -\boldsymbol{A}_2 & -\boldsymbol{A}_1 & \boldsymbol{I}_n \end{bmatrix}, \qquad \boldsymbol{\mathcal{S}}_{\boldsymbol{u}} = \begin{bmatrix} \boldsymbol{S}(0) & & & \\ & \boldsymbol{S}(1) & & \\ & & \boldsymbol{S}(2) & \\ & & & \boldsymbol{\Sigma}_{\eta\eta} \end{bmatrix}.$$

With banding parameter $q = 2$, the resulting BIAM is $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(2) = \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}'(2)\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(2)\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(2)$, where

$$\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(2) = \begin{bmatrix} \boldsymbol{I}_n & & & \\ -\boldsymbol{A}_1(1) & \boldsymbol{I}_n & & \\ -\boldsymbol{A}_2(2) & -\boldsymbol{A}_1(2) & \boldsymbol{I}_n & \\ & -\boldsymbol{A}_2(2) & -\boldsymbol{A}_1(2) & \boldsymbol{I}_n \end{bmatrix}, \qquad \boldsymbol{\mathcal{S}}_{\boldsymbol{u}}(2) = \begin{bmatrix} \boldsymbol{S}(0) & & & \\ & \boldsymbol{S}(1) & & \\ & & \boldsymbol{S}(2) & \\ & & & \boldsymbol{S}(2) \end{bmatrix}.$$

The model of (2.1) can be stacked over time to yield the representation $\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{u}$ with $\boldsymbol{y} = [\boldsymbol{y}_1', \boldsymbol{y}_2', \ldots, \boldsymbol{y}_T']'$, $\boldsymbol{Z} = [\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_T]'$ and $\boldsymbol{u}$ as before. For the moment, we will assume $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q)$ to be *known* and focus on the following estimator:

$$\widehat{\boldsymbol{\beta}}_{GLS} = \left(\boldsymbol{Z}'\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q)\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q)\boldsymbol{y}. \tag{2.7}$$

A discussion on the theoretical properties of this infeasible estimator is informative because: (1) the incurred estimation error of an appropriately constructed estimator $\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q)$ will be asymptotically negligible, and (2) we can suppress the effect of banding by letting q increase with T.

Two remarks related to $\widehat{\boldsymbol{\beta}}_{GLS}$ are instructive. First, the GLS estimator differs from the usual least squares estimator $\widehat{\boldsymbol{\beta}}_{OLS} := (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}$ by a weighing with the inverse covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q)$. It is well documented in standard econometric textbooks (e.g., Davidson and MacKinnon 2004, chapter 7) that this weighing may lead to substantial efficiency gains. Second, it is illustrative to substitute the MCBD of $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q)$ into the definition of this infeasible GLS estimator. The result is $\widehat{\boldsymbol{\beta}}_{GLS} = (\boldsymbol{Z}_{filt}'\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q)\boldsymbol{Z}_{filt})^{-1}\boldsymbol{Z}_{filt}'\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q)\boldsymbol{y}_{filt}$ where $\boldsymbol{Z}_{filt} = \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)\boldsymbol{Z}$, and $\boldsymbol{y}_{filt} = \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)\boldsymbol{y}$. The premultiplications by $\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)$ have the effect of filtering and taking care of serial correlation. The block diagonal matrix $\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q)$ applies scaling and rotation to account for the correlations between the series. The following univariate autoregressive setting exemplifies this intuition.

Example 2

A regression model $y_t = \beta t + u_t$ has AR(1) innovations $u_t = \rho u_{t-1} + \eta_t$ where $\eta_t \overset{i.i.d.}{\sim} (0, \sigma^2)$ and $|\rho| < 1$. Taking $n = 1$, the expressions of Example 1 are easily adapted to yield:

$$\boldsymbol{\mathcal{S}}_{\boldsymbol{u}} = \operatorname{diag}\left[\frac{\sigma^2}{1-\rho^2}, \sigma^2, \ldots, \sigma^2\right], \qquad \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}\boldsymbol{y} = \begin{bmatrix} 1 & & & \\ -\rho & 1 & & \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & -\rho & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 - \rho y_1 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix}, \tag{2.8}$$

and a similar transformation for the linear trend. The implied GLS estimator coincides with the

estimator from Prais and Winsten (1954).

3 Asymptotic theory

In this section, we present the asymptotic results. More specifically, we derive: (1) the limiting distribution of the GLS estimator, (2) the fully modified GLS (FM-GLS) estimator that corrects for second-order bias terms, (3) Wald test statistics, and (4) multivariate KPSS-type of tests for the null of cointegration. As mentioned, the two fully modified estimators defined in Proposition 1 of Wagner et al. (2020) shall be adopted as a benchmark. The following assumption will facilitate the development of the asymptotic theory.

Assumption 1 (innovation processes)

The innovations processes in the model (2.1) satisfy the following assumptions:

(a) The process $\boldsymbol{\zeta}_t = [\boldsymbol{\eta}_t', \boldsymbol{\varepsilon}_t']'$ *is an independent and identically distributed (i.i.d.) sequence with* $\mathbb{E}(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_t') = \begin{bmatrix} \boldsymbol{\Sigma}_{\eta\eta} & \boldsymbol{\Sigma}_{\eta\epsilon} \\ \boldsymbol{\Sigma}_{\epsilon\eta} & \boldsymbol{\Sigma}_{\epsilon\epsilon} \end{bmatrix} \succ 0$ *and* $\mathbb{E}(\|\boldsymbol{\zeta}_t\|^{2r}) \leq C_r < \infty$ *for some constant* $C_r > 0$ *and some* $r > 2$.

(b) The lag polynomial $\boldsymbol{\mathcal{A}}(\cdot)$ *in Eq. (2.2) satisfies* $\det\big(\boldsymbol{\mathcal{A}}(z)\big) \neq 0$ *for all* $|z| \leq 1$ *and* $\sum_{j=0}^{\infty} j\|\boldsymbol{A}_j\|_{\mathcal{F}} < \infty$*, where* $\det(\boldsymbol{A})$ *denotes the determinant of a matrix* $\boldsymbol{A}$.

(c) $\Delta\boldsymbol{x}_t = \boldsymbol{v}_t$ *admits the VAR(*∞*) process* $\boldsymbol{\mathcal{D}}(L)\boldsymbol{v}_t = \boldsymbol{\varepsilon}_t$*, where* $\boldsymbol{\mathcal{D}}(L) = \boldsymbol{I}_n - \sum_{j=1}^{\infty} \boldsymbol{D}_j L^j$*. Moreover,* $\det\big(\boldsymbol{\mathcal{D}}(z)\big) \neq 0$ *for all* $|z| \leq 1$ *and* $\sum_{j=0}^{\infty} j\,\|\boldsymbol{D}_j\|_{\mathcal{F}} < \infty$.

The stationary VAR(∞) specifications for $\{\boldsymbol{u}_t\}$ and $\{\boldsymbol{v}_t\}$ are natural given the linear minimum MSE predictor formulae that underlie the definitions of the MCBD and BIAM. Moreover, the conditions in Assumption 1 ensure that the lag polynomials $\boldsymbol{\mathcal{A}}(L)$ and $\boldsymbol{\mathcal{D}}(L)$ are invertible (see for example Theorem 7.4.2 of Hannan and Deistler (2012)), thereby showing that our Assumption 1 is similar to the linear processes assumptions that are regularly adopted in the literature on nonlinear cointegration, cf. Choi and Saikkonen (2010), Wagner and Hong (2016), and Wagner et al. (2020). The assumption $\det\big(\boldsymbol{\mathcal{D}}(1)\big) \neq 0$ rules out cointegration among the components of $\{\boldsymbol{x}_t\}$.

Under Assumption 1(a), an invariance principle holds for $\boldsymbol{\zeta}_t$, i.e., $T^{-1/2}\sum_{t=1}^{[rT]} \boldsymbol{\zeta}_t \Rightarrow \boldsymbol{B}_{\boldsymbol{\zeta}}(r) \equiv \begin{bmatrix} \boldsymbol{B}_{\eta}(r) \\ \boldsymbol{B}_{\epsilon}(r) \end{bmatrix}$ where $\boldsymbol{B}_{\boldsymbol{\zeta}}$ denotes an $2n$-dimensional Brownian motion with covariance matrix $\begin{bmatrix} \boldsymbol{\Sigma}_{\eta\eta} & \boldsymbol{\Sigma}_{\eta\epsilon} \\ \boldsymbol{\Sigma}_{\epsilon\eta} & \boldsymbol{\Sigma}_{\epsilon\epsilon} \end{bmatrix}$. Moreover, Assumptions 1(b)-(c) justify the use of the Beveridge-Nelson decomposition (Phillips and Solo, 1992). A functional central limit theorem for linear processes is thus also applicable to

$\boldsymbol{\xi}_t = [\boldsymbol{u}_t', \boldsymbol{v}_t']'$, that is

$$\frac{1}{T^{1/2}} \sum_{t=1}^{[rT]} \boldsymbol{\xi}_t \Rightarrow \boldsymbol{B}_{\boldsymbol{\xi}}(r) \equiv \begin{bmatrix} \boldsymbol{B}_u(r) \\ \boldsymbol{B}_v(r) \end{bmatrix} \equiv \begin{bmatrix} \boldsymbol{\mathcal{A}}(1) & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\mathcal{D}}(1) \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{B}_\eta(r) \\ \boldsymbol{B}_\epsilon(r) \end{bmatrix}, \tag{3.1}$$

where the Brownian motion $\boldsymbol{B}_{\boldsymbol{\xi}}(r)$ of dimension $2n$ has covariance matrix

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{uu} & \boldsymbol{\Omega}_{uv} \\ \boldsymbol{\Omega}_{vu} & \boldsymbol{\Omega}_{vv} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mathcal{A}}(1) & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\mathcal{D}}(1) \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\Sigma}_{\eta\eta} & \boldsymbol{\Sigma}_{\eta\epsilon} \\ \boldsymbol{\Sigma}_{\epsilon\eta} & \boldsymbol{\Sigma}_{\epsilon\epsilon} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mathcal{A}}(1)' & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\mathcal{D}}(1)' \end{bmatrix}^{-1}. \tag{3.2}$$

Apart from this long-run covariance matrix $\boldsymbol{\Omega} = \sum_{h=-\infty}^{\infty} \mathbb{E}\big(\boldsymbol{\xi}_t \boldsymbol{\xi}_{t+h}'\big)$, we also introduce the one-sided long-run covariance matrix $\boldsymbol{\Delta} = \left[\begin{smallmatrix} \boldsymbol{\Delta}_{uu} & \boldsymbol{\Delta}_{uv} \\ \boldsymbol{\Delta}_{vu} & \boldsymbol{\Delta}_{vv} \end{smallmatrix}\right] = \sum_{h=0}^{\infty} \mathbb{E}\big(\boldsymbol{\xi}_t \boldsymbol{\xi}_{t+h}'\big)$. The Brownian motion defined by $\boldsymbol{B}_{u.v} = \boldsymbol{B}_u - \boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{B}_v$ is by construction independent of $\boldsymbol{B}_v$. Its $(n \times n)$ covariance matrix equals $\boldsymbol{\Omega}_{u.v} = \boldsymbol{\Omega}_{uu} - \boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{\Omega}_{vu}$.

3.1 Infeasible GLS

We start our analysis assuming that the $(nT \times nT)$ covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{u}}(q)$ is a known quantity for each q. The MCBD on p.4 can now be used to derive the limiting distribution of this infeasible GLS estimator. An insightful exposition of our results requires further notation.

(a) Introduce scaling matrices: $\boldsymbol{G}_{\boldsymbol{d}_i,T} := T^{-1/2}\,\mathrm{diag}[1, T^{-1}, \ldots, T^{-d_i}]$ for the time trends, and $\boldsymbol{G}_{\boldsymbol{s}_i,T} := T^{-1/2}\,\mathrm{diag}[T^{-1/2}, T^{-1}, \ldots, T^{-s_i/2}]$ for the stochastic trends. Moreover, we define $\boldsymbol{G}_T := \mathrm{diag}\,[\boldsymbol{G}_{1,T}, \ldots, \boldsymbol{G}_{n,T}]$, where $\boldsymbol{G}_{i,T} := \mathrm{diag}\,[\boldsymbol{G}_{\boldsymbol{d}_i,T}, \boldsymbol{G}_{\boldsymbol{s}_i,T}]$.

(b) Let $\boldsymbol{d}_i(r) := \big[1, r, \ldots, r^{d_i}\big]'$, $\boldsymbol{B}_{s_i}(r) := \big[B_{v_i}(r), B_{v_i}^2(r), \ldots, B_{v_i}^{s_i}(r)\big]'$ and $\boldsymbol{j}_i(r) := [\boldsymbol{d}_i(r)', \boldsymbol{B}_{s_i}(r)']'$. Define $d \times n$ block-diagonal random matrix $\boldsymbol{J}(r) := \mathrm{diag}\,[\boldsymbol{j}_1(r), \ldots, \boldsymbol{j}_n(r)]$.

(c) $\boldsymbol{b}_i := \left[\boldsymbol{0}_{d_i+1}', 1, 2\int_0^1 B_{v_i}(r)dr, \ldots, s_i \int_0^1 B_{v_i}^{s_i-1}(r)dr\right]'$.

Finally, we use $\boldsymbol{B}_{v_j}$ as shorthand notation for the jth component of $\boldsymbol{B}_v$.

Theorem 1 (limiting distribution of the infeasible GLS estimator)
If Assumption 1 holds, and if $q = q(T)$ satisfies $1/q + q/T \overset{T\to\infty}{\to} 0$, then

$$\begin{aligned} \boldsymbol{G}_T^{-1}\left(\widehat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta}\right) \Rightarrow & \left(\int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{uu}^{-1}\boldsymbol{J}(r)'dr\right)^{-1} \\ & \times \left(\int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{uu}^{-1}d\boldsymbol{B}_{u.v}(r) + \int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{uu}^{-1}\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}d\boldsymbol{B}_v(r) + \boldsymbol{\mathcal{B}}\right), \end{aligned} \tag{3.3}$$

as $T \to \infty$ (for a fixed n), where $\boldsymbol{\mathcal{B}} = [\boldsymbol{\mathcal{B}}_1', \ldots, \boldsymbol{\mathcal{B}}_n']'$ with $\boldsymbol{\mathcal{B}}_i = \mathrm{row}_i\left(\boldsymbol{\Sigma}_{\epsilon\eta}\right)\mathrm{col}_i\left(\boldsymbol{\Sigma}_{\eta\eta}^{-1}\right)\boldsymbol{b}_i$.

The limiting result in (3.3) coincides with the limiting distribution of the MSUR estimator, $\widetilde{\boldsymbol{\beta}}_{MSUR} := \left(\boldsymbol{Z}'(\boldsymbol{I}_T \otimes \widehat{\boldsymbol{\Omega}}_{uu}^{-1})\boldsymbol{Z}\right)^{-1}\left(\boldsymbol{Z}'(\boldsymbol{I}_T \otimes \widehat{\boldsymbol{\Omega}}_{uu}^{-1})\boldsymbol{y}\right)$, as reported in Wagner et al. (2020), see their Proof of Proposition 1. The equivalence of these limiting distributions is caused by the facts that: (1) applying a linear filter to an integrated series only affect its long-run variance (e.g., Phillips and Park 1988), and (2) the previous statement remains true when applying a linear filter to higher integer powers of integrated series. The terms $\int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{uu}^{-1}\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}d\boldsymbol{B}_v(r)$ and $\boldsymbol{\mathcal{B}}$ in (3.3) reflect the presence of second-order bias terms caused by serial correlation and endogeneity. In Section 3.3, we introduce the fully modified (FM) correction that adjusts these bias terms and leads to standard inference. We first introduce a feasible version of the GLS estimator.

3.2 Consistent estimation of BIAM and feasible GLS

Up to this point, we have discussed the infeasible estimator $\widehat{\boldsymbol{\beta}}_{GLS}$ given in Eq. (2.7). A feasible GLS approach requires a consistent estimator of the $(nT \times nT)$ matrix $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q)$. Several authors, e.g., Wu and Pourahmadi (2009) and McMurry and Politis (2010), have constructed consistent estimators of large covariance matrices using banding or tapering to reduce the number of unknown parameters. Direct usage of their results poses two difficulties. First, the numerical inversion of large matrices is computationally expensive for large nT. Second, matrix inversion might be impossible because the estimated covariance matrix cannot be guaranteed to be positive definite. In light of such considerations, we will estimate $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q)$ directly and ensure it to be positive definite. The approach is the sample counterpart of the BIAM described on p.5. That is, we replace true innovations by first-stage OLS residuals $\widehat{\boldsymbol{u}}_t = \boldsymbol{y}_t - \boldsymbol{Z}_t\widehat{\boldsymbol{\beta}}_{OLS}$, and subsequently minimize a sample moment in estimated residuals rather than the population mean squared forecasting error. This method was previously used by Cheng et al. (2015), Ing et al. (2016a), and Beutner et al. (2022), for univariate time series. For a multivariate time series, we define

$$\begin{aligned}
\widehat{\boldsymbol{A}}(\ell) &= \left[\widehat{\boldsymbol{A}}_1(\ell) \quad \cdots \quad \widehat{\boldsymbol{A}}_\ell(\ell)\right] = \underset{(\boldsymbol{\Theta}_1,\ldots,\boldsymbol{\Theta}_\ell)\in\mathbb{R}^{n\times n\ell}}{\arg\min} \sum_{t=\ell+1}^{T} \|\widehat{\boldsymbol{u}}_t - \boldsymbol{\Theta}_1\widehat{\boldsymbol{u}}_{t-1} - \cdots - \boldsymbol{\Theta}_\ell\widehat{\boldsymbol{u}}_{t-\ell}\|^2 \\
\widehat{\boldsymbol{S}}(\ell) &= \frac{1}{T-\ell}\sum_{t=\ell+1}^{T}\left[\widehat{\boldsymbol{u}}_t - \widehat{\boldsymbol{A}}_1(\ell)\widehat{\boldsymbol{u}}_{t-1} - \cdots - \widehat{\boldsymbol{A}}_\ell(\ell)\widehat{\boldsymbol{u}}_{t-\ell}\right]\left[\widehat{\boldsymbol{u}}_t - \widehat{\boldsymbol{A}}_1(\ell)\widehat{\boldsymbol{u}}_{t-1} - \cdots - \widehat{\boldsymbol{A}}_\ell(\ell)\widehat{\boldsymbol{u}}_{t-\ell}\right]',
\end{aligned} \tag{3.4}$$

$1 \le \ell \le q$, and $\widehat{\boldsymbol{S}}(0) = T^{-1}\sum_{t=1}^{T}\widehat{\boldsymbol{u}}_t\widehat{\boldsymbol{u}}_t'$. Similarly to (2.5)-(2.6), we subsequently construct the matrices $\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{u}}(q) = [\widehat{\boldsymbol{m}}_{\boldsymbol{u}}^{ij}(q)]_{1\le i,j\le T}$ and $\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{u}}(q) = \operatorname{diag}\left[\widehat{\boldsymbol{S}}(0), \widehat{\boldsymbol{S}}(1), \ldots, \widehat{\boldsymbol{S}}(q), \ldots, \widehat{\boldsymbol{S}}(q)\right]$, and obtain

the multivariate BIAM estimator as

$$\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q) = \widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{u}}'(q)\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{u}}^{-1}(q)\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{u}}(q). \tag{3.5}$$

Assumption 2

For $\widehat{\boldsymbol{u}} = [\widehat{\boldsymbol{u}}_1', \ldots, \widehat{\boldsymbol{u}}_T']'$ *and* $\boldsymbol{u} = [\boldsymbol{u}_1', \ldots, \boldsymbol{u}_T']'$, *assume* $\|\widehat{\boldsymbol{u}} - \boldsymbol{u}\|^2 = O_p(1)$.

Assumption 3

Assume $q = q_T$ *satisfies* $\dfrac{1}{q_T} + \dfrac{q_T^3}{T} \to 0$, *as* $T \to \infty$.

Assumption 2 requires the residuals to be sufficiently close to the true innovations. It is a rather mild assumption and it is satisfied if residuals are computed by least squares. Assumption 3 places constraints on the banding parameter q_T. First, it requires the banding parameter to diverge with the sample size. This ensures that no nonzero elements are (asymptotically) set to zero. Moreover, the assumption $q_T^3/T \overset{T\to\infty}{\to} 0$ establishes an upper bound for the growth rate of q_T. The following theorem shows the consistent estimation of $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}$, implying that the infeasible and feasible GLS estimators share the same limiting distribution.

Theorem 2 (consistency of $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}$)

If Assumptions 1-3 hold, then

$$\begin{aligned}\left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q_T) - \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\right\| &\leq \left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q_T) - \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q_T)\right\| + \left\|\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q_T) - \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\right\| \\ &= O_p\left(\sqrt{q_T^3/T}\right) + O\left(q_T^{-1/2}\sum_{s=q_T+1}^{\infty} s\,\|\boldsymbol{A}_s\|_{\mathcal{F}}\right) \longrightarrow_p 0, \quad \text{as } T \to \infty,\end{aligned} \tag{3.6}$$

where "$\longrightarrow_p$" denotes convergence in probability.

3.3 Fully modified inference

The asymptotic results of Theorem 1 are not immediately useful for statistical inference. There are two difficulties. First, the second-order bias dislocates the limiting distribution which can translate into substantial finite sample bias. Second, possible dependencies between the Brownian motions $\boldsymbol{B}_u$ and $\boldsymbol{B}_v$ cause the limiting distribution to depend on nuisance parameters. Critical values would therefore be nuisance parameter dependent as well.

These two issues have received extensive attention in the linear cointegration literature. A (non-exhaustive) list of solution methods is: joint modeling as in Phillips (1991), Saikkonen's (1992) dynamic least squares, and the integrated modified OLS and fixed-b approaches by Vogelsang and

Wagner (2014). We rely on the fully modified (FM) approach advocated by Phillips and Hansen (1990) and Phillips (1995). The idea is a twofold modification of the estimator: (1) second-order bias terms are subtracted, and (2) a transformation of the dependent variable is introduced to obtain a zero-mean Gaussian mixture limiting distribution. Recently, Wagner et al. (2020) have proposed two estimators within the framework of SUCPRs. These estimators, FM-SOLS and FM-SUR, rely on kernel estimators of the one- and two-sided long-run covariance matrices (see Theorem 3). As such, we impose the following assumption.

Assumption 4 (consistent estimation of long-run covariance matrices)
$\widehat{\boldsymbol{\Omega}}$ and $\widehat{\boldsymbol{\Delta}}$ are consistent kernel estimators of the long-run covariance matrix $\boldsymbol{\Omega}$ and the one-sided long-run covariance matrix $\boldsymbol{\Delta}$, respectively.

Andrews (1991) and Newey and West (1994) use kernel estimators for long-run covariance estimation. Their method involves the calculation of weighted sums of the autocovariance matrices of the residuals. These weights are determined by a kernel function and bandwidth parameter. Our Assumption 4 is easily satisfied by imposing suitable conditions on the kernel function and bandwidth parameter. We refer to Phillips (1995) and Jansson (2002) for an enumeration of such conditions.

Alternatively, we can obtain consistent one- and two-sided long-run covariance estimators within the BIAM framework of Section 3.2.[2] This approach resembles Berk (1974). The GLS estimator and its FM counterpart are thus constructed within a single framework. The estimators are as follows. For $t = 1, 2, \ldots, T$, we first stack $\widehat{\boldsymbol{u}}_t$ and $\Delta\boldsymbol{x}_t = \boldsymbol{v}_t$ in the $2n$-dimensional vector $\widehat{\boldsymbol{\xi}}_t = [\widehat{\boldsymbol{u}}_t', \Delta\boldsymbol{x}_t']'$. Since the BIAM estimator is essentially fitting VAR processes up to order q_T, we will use the estimated VAR(q_T) approximations to define the long-run covariance estimators. For $\boldsymbol{\Omega}$, the estimator is $\widehat{\boldsymbol{\Omega}}_{q_T} = \left(\boldsymbol{I}_{2n} - \sum_{j=1}^{q_T} \widehat{\boldsymbol{F}}_j(q_T)\right)^{-1} \widehat{\boldsymbol{\Sigma}}_{q_T} \left(\boldsymbol{I}_{2n} - \sum_{j=1}^{q_T} \widehat{\boldsymbol{F}}_j'(q_T)\right)^{-1}$, where $\widehat{\boldsymbol{\Sigma}}_{q_T} = \widehat{\boldsymbol{S}}(q_T)$ and $\widehat{\boldsymbol{F}}_j(q_T)$ denote respectively the estimated prediction error variance and the coefficient matrix of the jth lag when a VAR(q_T) is fitted to $\{\widehat{\boldsymbol{\xi}}_t\}_{t=1}^T$. Recall the population one-sided long-run covariance matrix is $\boldsymbol{\Delta} = \sum_{h=0}^{\infty} \mathbb{E}\big(\boldsymbol{\xi}_t\boldsymbol{\xi}_{t+h}'\big)$. It is thus intuitive to approximate this quantity by a finite sum of estimated covariance matrices of $\{\widehat{\boldsymbol{\xi}}_t\}_{t=1}^T$. These covariance matrices are nothing but subblocks of the matrix $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}}(q_T) = \widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{\xi}}^{-1}(q_T)\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\xi}}(q_T)\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{\xi}}^{-1\prime}(q_T)$.[3] We therefore use

$$\widehat{\boldsymbol{\Delta}}_{q_T,r_T} = \boldsymbol{Q}_{r_T}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}}(q_T)\boldsymbol{Q}_1 \tag{3.7}$$

[2]An overview of the procedure is given here. Section 2 in the Supplement provides further details.

[3]We use $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$ to denote the $(2nT \times 2nT)$ matrix $\mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}')$ where $\boldsymbol{\xi} = [\boldsymbol{\xi}_1', \boldsymbol{\xi}_2', \ldots, \boldsymbol{\xi}_T']'$. The matrices $\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{\xi}}(q)$ and $\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\xi}}(q)$ are defined similarly to respectively $\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{u}}(q)$ and $\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{u}}(q)$ (see p.10). The matrix $\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{\xi}}(q_T)$ is lower triangular with identity matrices on the main diagonal. Therefore, its matrix inverse exists and is fast to compute.

where $\boldsymbol{Q}_r = [\mathbf{O}_{2n\times 2n}, \cdots, \mathbf{O}_{2n\times 2n}, \boldsymbol{I}_{2n}, \cdots, \boldsymbol{I}_{2n}]'$ is a $\left(2nT \times 2n\right)$ block matrix of zeros of which the last r blocks have been replaced by identity matrices. To ensure consistency, we place the following rate restriction on the number of included autocovariance matrices.

Assumption 5
As $T \to \infty$, $\dfrac{1}{r_T} + \dfrac{r_T q_T^3}{T} \to 0$ *and* $r_T = O(q_T)$.

Definitions and limiting results for FM estimators are presented in Theorem 3. The FM-SOLS, FM-SUR, and FM-GLS estimators, all depend on estimators for $\boldsymbol{\Delta}$ and $\boldsymbol{\Omega}$. It is only the consistency of these estimators that is relevant for the asymptotic analysis, not whether the kernel or BIAM approach is employed. As such, we will not complicate notation by introducing additional notation to indicate whether the kernel or BIAM approach is used. In subsequent theorems and simulation results, we will use kernel estimators for FM-SOLS and FM-SUR, and the BIAM approach for FM-GLS. This seems to be the logical choice for these estimators.

Theorem 3
For $i = 1, \ldots, n$, *define* $\widehat{\boldsymbol{b}}_i = \left[\mathbf{0}'_{d_i+1}, T, 2\sum_{t=1}^T x_{it}, \ldots, s_i \sum_{t=1}^T x_{it}^{s_i-1}\right]'$. *Also, let the* $(n \times n)$ *matrix* $\widehat{\boldsymbol{\Delta}}^+_{vu}$ *be the (implied) consistent estimator of* $\boldsymbol{\Delta}^+_{vu} = \boldsymbol{\Delta}_{vu} - \boldsymbol{\Delta}_{vv}\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{\Omega}_{vu}$.

(a) Define the FM-SOLS estimator as

$$\widehat{\boldsymbol{\beta}}^+_{SOLS} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\left(\boldsymbol{Z}'\boldsymbol{y}^+ - \widehat{\boldsymbol{A}}\right), \tag{3.8}$$

where $\boldsymbol{y}^+ := [\boldsymbol{y}_1^{+\prime}, \boldsymbol{y}_2^{+\prime}, \ldots, \boldsymbol{y}_T^{+\prime}]'$ *with* $\boldsymbol{y}_t^+ = \boldsymbol{y}_t - \widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\Delta\boldsymbol{x}_t$, *and* $\widehat{\boldsymbol{A}} := [\widehat{\boldsymbol{A}}'_1, \ldots, \widehat{\boldsymbol{A}}'_n]'$ *with* $\widehat{\boldsymbol{A}}_i = \widehat{\boldsymbol{\Delta}}^+_{v_i u_i}\widehat{\boldsymbol{b}}_i$ *and* $\widehat{\boldsymbol{\Delta}}^+_{v_i u_i}$ *being the* i^{th} *element on the main diagonal of* $\widehat{\boldsymbol{\Delta}}^+_{vu}$. *If Assumptions 1 and 4 hold, then*

$$\boldsymbol{G}_T^{-1}\left(\widehat{\boldsymbol{\beta}}^+_{SOLS} - \boldsymbol{\beta}\right) \Rightarrow \left(\int_0^1 \boldsymbol{J}(r)\boldsymbol{J}(r)'dr\right)^{-1}\int_0^1 \boldsymbol{J}(r)d\boldsymbol{B}_{u.v}(r). \tag{3.9}$$

(b) Define the FM-SUR estimator as

$$\widehat{\boldsymbol{\beta}}^+_{SUR} = \left(\boldsymbol{Z}'(\boldsymbol{I}_T \otimes \widehat{\boldsymbol{\Omega}}_{u.v}^{-1})\boldsymbol{Z}\right)^{-1}\left(\boldsymbol{Z}'(\boldsymbol{I}_T \otimes \widehat{\boldsymbol{\Omega}}_{u.v}^{-1})\boldsymbol{y}^+ - \widetilde{\boldsymbol{A}}^*\right), \tag{3.10}$$

where $\widetilde{\boldsymbol{A}}^* := [\widetilde{\boldsymbol{A}}_1^*, \ldots, \widetilde{\boldsymbol{A}}_n^*]$ *with* $\widetilde{\boldsymbol{A}}_i^* = \mathrm{row}_i\left(\widehat{\boldsymbol{\Delta}}^+_{vu}\right)\mathrm{col}_i\left(\widehat{\boldsymbol{\Omega}}_{u.v}^{-1}\right)\widehat{\boldsymbol{b}}_i$. *If Assumptions 1 and 4 hold, then*

$$\boldsymbol{G}_T^{-1}\left(\widehat{\boldsymbol{\beta}}^+_{SUR} - \boldsymbol{\beta}\right) \Rightarrow \left(\int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{u.v}^{-1}\boldsymbol{J}(r)'dr\right)^{-1}\int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{u.v}^{-1}d\boldsymbol{B}_{u.v}(r). \tag{3.11}$$

(c) Define the FM-GLS estimator as

$$\widehat{\boldsymbol{\beta}}^{+}_{FGLS} = \left(\boldsymbol{Z}'\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q)\boldsymbol{Z}\right)^{-1}\left[\boldsymbol{Z}'\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q)\boldsymbol{y} - \boldsymbol{Z}'\left(\boldsymbol{I}_T \otimes \widehat{\boldsymbol{\Omega}}_{uu}^{-1}\widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\right)\boldsymbol{v} - \widehat{\boldsymbol{\mathcal{B}}}^{+}\right], \tag{3.12}$$

where $\boldsymbol{v} := [\Delta\boldsymbol{x}_1', \ldots, \Delta\boldsymbol{x}_T']' = [\boldsymbol{v}_1', \ldots, \boldsymbol{v}_T']'$, *and* $\widehat{\boldsymbol{\mathcal{B}}}^{+} = \left[\widehat{\boldsymbol{\mathcal{B}}}_1^{+\prime}, \ldots, \widehat{\boldsymbol{\mathcal{B}}}_n^{+\prime}\right]'$ *with*

$$\widehat{\boldsymbol{\mathcal{B}}}_i^{+} = \left[\mathrm{row}_i\left(\widehat{\boldsymbol{\Sigma}}_{\epsilon\eta}\right)\mathrm{col}_i\left(\widehat{\boldsymbol{\Sigma}}_{\eta\eta}^{-1}\right) - \mathrm{row}_i\left(\widehat{\boldsymbol{\Delta}}_{vv}\right)\mathrm{col}_i\left(\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\widehat{\boldsymbol{\Omega}}_{vu}\widehat{\boldsymbol{\Omega}}_{uu}^{-1}\right)\right]\ \widehat{\boldsymbol{b}}_i.$$

If Assumptions 1-3 and 5 hold, then

$$\boldsymbol{G}_T^{-1}\left(\widehat{\boldsymbol{\beta}}^{+}_{FGLS} - \boldsymbol{\beta}\right) \Rightarrow \left(\int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{uu}^{-1}\boldsymbol{J}(r)'dr\right)^{-1}\int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{uu}^{-1}d\boldsymbol{B}_{u.v}(r). \tag{3.13}$$

The FM-GLS estimator is new to the SUCPR literature, whereas the FM-SOLS and FM-SUR estimators have recently appeared in Wagner et al. (2020). Theorem 3 indicates that all three estimators have a zero-mean Gaussian mixture limiting distribution implying that standard inference is applicable for each. However, we also see from Theorem 3 that the asymptotic covariance matrices are generally different because different types of weighing are used in the construction of the estimators, see Corollary 1 below.

Corollary 1

Conditional on $\mathcal{F}_v = \sigma\big(\boldsymbol{B}_v(r), 0 \le r \le 1\big)$, *the asymptotic covariance matrices of FM-SOLS, FM-SUR, and FM-GLS are, respectively, given as follows.*

$$\begin{aligned}
\boldsymbol{\mathcal{V}}^{+}_{SOLS} &= \left(\int_0^1 \boldsymbol{J}(r)\boldsymbol{J}(r)'dr\right)^{-1}\left(\int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{u.v}\boldsymbol{J}(r)'dr\right)\left(\int_0^1 \boldsymbol{J}(r)\boldsymbol{J}(r)'dr\right)^{-1},\\
\boldsymbol{\mathcal{V}}^{+}_{SUR} &= \left(\int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{u.v}^{-1}\boldsymbol{J}(r)'dr\right)^{-1},\\
\boldsymbol{\mathcal{V}}^{+}_{FGLS} &= \left(\int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{uu}^{-1}\boldsymbol{J}(r)'dr\right)^{-1}\left(\int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{uu}^{-1}\boldsymbol{\Omega}_{u.v}\boldsymbol{\Omega}_{uu}^{-1}\boldsymbol{J}(r)'dr\right)\left(\int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{uu}^{-1}\boldsymbol{J}(r)'dr\right)^{-1}.
\end{aligned}$$

There are special cases in which some (pairs of) estimators become asymptotically equivalent. For example, if $n = 1$, *then all estimators are asymptotically equivalent because the weighting matrices* $\boldsymbol{\Omega}_{u.v}^{-1}$ *and* $\boldsymbol{\Omega}_{uu}^{-1}$ *are now scalars. Also, under exogeneity, we have* $\boldsymbol{\Omega}_{uu} = \boldsymbol{\Omega}_{u.v}$, *and the FM-SUR and FM-GLS estimators share the same limiting distribution.*

For completeness, we also detail how the FM-GLS estimator can be used to test linear hypotheses. A formal presentation of such a result is more involved because of the different convergence rates of the individual parameter estimators. That is, the parameters with the lowest

convergence rate will dominate the asymptotic distribution and one should take care to avoid a degenerate limiting distribution. We will rule out such complications by considering hypothesis tests on individual parameters.[4] Therefore, let $\boldsymbol{R}$ denote a $(k \times s)$ selection matrix in which every row contains a single 1 and zeros otherwise. The null hypothesis $\boldsymbol{R\beta} = \boldsymbol{r}$ can be tested using the standard chi-squared limiting distribution of the Wald statistic (Theorem 4). These tests are practically relevant. For example, exclusion restrictions of the type $\boldsymbol{R\beta} = \boldsymbol{0}$ allow us to test whether the cointegrating relation is linear.

Theorem 4 (linear hypothesis testings)
Consider the null hypothesis $H_0 : \boldsymbol{R\beta} = \boldsymbol{r}$, which imposes k linearly independent restrictions. Under the assumptions of Theorem 3(c), the Wald test statistic

$$\mathcal{W} = \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}^{+}_{FGLS} - \boldsymbol{r}\right)' \widehat{\boldsymbol{\Phi}}^{-1} \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}^{+}_{FGLS} - \boldsymbol{r}\right) \Rightarrow \chi^2_k, \qquad \text{as } T \to \infty, \tag{3.14}$$

where $\widehat{\boldsymbol{\Phi}} = \boldsymbol{R}\left[\boldsymbol{Z}'\left(\boldsymbol{I}_T \otimes \widehat{\boldsymbol{\Omega}}^{-1}_{uu}\right)\boldsymbol{Z}\right]^{-1}\left[\boldsymbol{Z}'\left(\boldsymbol{I}_T \otimes \widehat{\boldsymbol{\Omega}}^{-1}_{uu}\widehat{\boldsymbol{\Omega}}_{u.v}\widehat{\boldsymbol{\Omega}}^{-1}_{uu}\right)\boldsymbol{Z}\right]\left[\boldsymbol{Z}'\left(\boldsymbol{I}_T \otimes \widehat{\boldsymbol{\Omega}}^{-1}_{uu}\right)\boldsymbol{Z}\right]^{-1}\boldsymbol{R}'$.

3.4 Testing the null of cointegration

Stationarity tests are used to avoid spurious regressions and to verify the correct specification of the cointegrating relation. To test for stationarity of the SUCPR errors, we combine the test statistic from Nyblom and Harvey (2000) with the sub-sampling approach found in Choi and Saikkonen (2010). We consider three test statistics. To treat all test statistics in a unified framework, we define

$$\boldsymbol{\varphi}_{j,b}(\boldsymbol{x}) = \left[\boldsymbol{x}'_j, \sum_{s=j}^{j+1}\boldsymbol{x}'_s, \ldots, \sum_{s=j}^{j+b-1}\boldsymbol{x}'_s\right]', \tag{3.15}$$

that is, a vector of length nb stacking the cumulative sums of $\{\boldsymbol{x}_j, \ldots, \boldsymbol{x}_{j+b-1}\}$. If the true innovations $\{\boldsymbol{u}_t\}$ were observed, then we could use the full-sample KPSS-type of test statistic $T^{-2}\boldsymbol{\varphi}_{1,T}(\boldsymbol{u})'(\boldsymbol{I}_T \otimes \widehat{\boldsymbol{\Omega}}^{-1}_{uu})\boldsymbol{\varphi}_{1,T}(\boldsymbol{u}) = \operatorname{tr}\left[\widehat{\boldsymbol{\Omega}}^{-1}_{uu}T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\boldsymbol{u}_s\right)\left(\sum_{s=1}^{t}\boldsymbol{u}_s\right)'\right]$ to test for stationarity of the innovations. Under the null of stationarity, this test statistic would converge weakly to $\int_0^1 \|\boldsymbol{W}(r)\|^2 dr$ with $\boldsymbol{W}(r)$ denoting an n-dimensional standard Brownian motion. This limiting distribution is free of nuisance parameters and the cumulative distribution function is available as a series expansion (see Theorem S1 in the Supplement).

[4] For general linear hypothesis, we refer the reader to Sims et al. (1990) where a reordering based on convergence rates is used to establish the limiting distribution of the Wald F statistic for a general linear hypothesis. The same approach is applicable in our setting but we will not explore this in greater detail.

The innovations $\{\boldsymbol{u}_t\}$ are only available when cointegrating relations are pre-specified. If these coefficients are estimated, then this additional parameter uncertainty will contaminate the limiting distribution with nuisance parameters.[5] The idea behind the subsampling approach is to construct a test statistic incorporating $b = b_T$ residuals while computing parameter estimators from all T observations. If b_T increases slowly with sample size, then the parameter estimation error will be negligible relative to the randomness in the errors and the asymptotic distribution remains $\int_0^1 \|\boldsymbol{W}(r)\|^2 dr$.

The three KPSS-type of tests are based on the following residuals: $\hat{\boldsymbol{u}}^+_{t,SOLS} = \boldsymbol{y}^+_t - \boldsymbol{Z}_t\widehat{\boldsymbol{\beta}}^+_{SOLS}$, $\hat{\boldsymbol{u}}^+_{t,SUR} = \boldsymbol{y}^+_t - \boldsymbol{Z}_t\widehat{\boldsymbol{\beta}}^+_{SUR}$, and $\hat{\boldsymbol{u}}_{t,FGLS} = \boldsymbol{y}_t - \boldsymbol{Z}_t\widehat{\boldsymbol{\beta}}^+_{FGLS}$. The test statistics are given by:

$$K^i_{j,b_T} = \frac{1}{b_T^2}\boldsymbol{\varphi}_{j,b_T}(\hat{\boldsymbol{u}}^+_i)' \left(\boldsymbol{I}_{b_T} \otimes \widehat{\boldsymbol{\Omega}}^{-1}_{u.v}\right) \boldsymbol{\varphi}_{j,b_T}(\hat{\boldsymbol{u}}^+_i), \qquad i \in \{SOLS, SUR\}, \tag{3.16}$$

and

$$K^{FGLS}_{j,b_T} = \frac{1}{b_T^2}\boldsymbol{\varphi}_{j,b_T}(\hat{\boldsymbol{u}}_{FGLS})'\widehat{\boldsymbol{\Sigma}^{-1}_{\boldsymbol{u}}}(q_T, b_T)\boldsymbol{\varphi}_{j,b_T}(\hat{\boldsymbol{u}}_{FGLS}), \tag{3.17}$$

where $\widehat{\boldsymbol{\Sigma}^{-1}_{\boldsymbol{u}}}(q_T, b_T)$ is the $(nb_T \times nb_T)$ submatrix of $\widehat{\boldsymbol{\Sigma}^{-1}_{\boldsymbol{u}}}(q_T)$ obtained by selecting the rows and columns related to all time indices in the set $\{n(T - b_T) + 1, n(T - b_T) + 2, \ldots, nT\}$. The test statistic in (3.17) fits naturally into the FM-GLS estimation framework.

Theorem 5 (limiting distributions of KPSS tests)

Let the assumptions from Theorem 3 hold.

(a) If $\frac{1}{b_T} + \frac{b_T}{T} \overset{T\to\infty}{\to} 0$, *then as* $T \to \infty$,

$$K^i_{j,b_T} \Rightarrow \int_0^1 \|\boldsymbol{W}(r)\|^2 dr, \qquad 1 \le j \le T - b_T + 1, \qquad i \in \{SOLS, SUR\}.$$

(b) If $\frac{q_T}{b_T} + \frac{b_T}{T} \overset{T\to\infty}{\to} 0$, *then* $K^{FGLS}_{j,b_T} \Rightarrow \int_0^1 \|\boldsymbol{W}(r)\|^2 dr$ *for any* $1 \le j \le T - b_T + 1$.

A sample of size T allows for up to $M = \lfloor T/b_T \rfloor$ series of nonoverlapping blocks of residuals of length b_T. Similarly to Choi and Saikkonen (2010), we apply the Bonferroni procedure to use all these series and thereby increase power. The approach is applicable to any of the three test statistics in Theorem 5. As such, we keep the notation general and use a generic K_j to denote a test statistic based on the jth subseries, $j = 1, 2, \ldots, M$. In the Bonferroni procedure

[5] There are exceptions. Shin (1994) reports a nuisance parameter-free limiting distribution for a single-equation linear cointegrating relation. This remains true if only a single integrated variable enters the cointegrating regression with a higher power, see Proposition 5 in Wagner and Hong (2016).

we compute $K_{max} = \max\{K_1, K_2, \ldots, K_M\}$ and do not reject the null hypothesis whenever $K_{max} \leq c_{\alpha/M}$ with $c_{\alpha/M}$ defined by $\mathbb{P}\left(\int_0^1 \|\boldsymbol{W}(r)\|^2 dr \geq c_{\alpha/M}\right) = \alpha/M$. The Bonferroni inequality implies $\lim_{T\to\infty} \mathbb{P}\left(K_{max} \leq c_{\alpha/M}\right) \geq 1 - \lim_{T\to\infty} \sum_{j=1}^{M} \mathbb{P}\left(K_j > c_{\alpha/M}\right) = 1 - \alpha$ and we see that the probability of a type-I error does not exceed the significance level α.

Remark 1

We suggest following Choi and Saikkonen (2010) in terms of the implementation of the subsampling approach. That is, the block size b_T is selected using the minimum volatility rule by Romano and Wolf (2001). For this particular block size, we subsequently select subsamples by taking non-overlapping blocks from alternatively the start and the end of the sample.

Remark 2

The limiting results in Theorem 5 guarantee a correct asymptotic size. Our simulations show (1) that these tests have power against various alternative hypotheses and (2) that power increases with sample size. A theoretical investigation of the power properties is outside our scope.

4 Simulations

We now study the finite sample performance of the estimators and stationarity tests. First, we compare the FM-GLS estimator with the FM-SOLS and FM-SUR estimators from Wagner et al. (2020). All long-run covariance matrices are computed using a Bartlett kernel and the automatic bandwidth selection approach due to Andrews (1991). For FM-GLS, the banding parameter q_T is selected using the subsampling and risk-minimization approach explained in section 5 from Bickel and Levina (2008).[6] Infeasible counterparts of the estimator are constructed assuming the knowledge of the true serial correlation and/or cross-sectional dependence pattern. These estimators are denoted by infSOLS, infSUR, and infGLS. Second, we look at the cointegration tests. We consider three test statistics: K_{max}^{SOLS} and K_{max}^{SUR} use the residuals as in (3.16), whereas K_{max}^{FGLS} employs the pre-filtered residuals from (3.17). All tests are implemented with minimum volatility block size selection and Bonferroni correction. We consider $T \in \{100, 200, 500\}$ and $n \in \{3, 5\}$. All tests are performed at a nominal significance level of 5%. For stationary processes, a presample of 200 observations is used to remove the influence of the starting values. All results are based on 2.5×10^4 Monte Carlo replicates.

[6] More details concerning the implementation can be found in the Supplement.

4.1 Monte Carlo designs

We generate data according to a quadratic SUCPR. That is, we adopt the DGP in (2.1) with $\boldsymbol{z}_{it} = \left[1, t, x_{it}, x_{it}^2\right]'$. The integrated variables satisfy $\boldsymbol{x}_0 = \boldsymbol{0}$ and $\Delta\boldsymbol{x}_t = \boldsymbol{v}_t$. We explore two error processes.

Setting A (Errors as in Wagner et al. (2020)): As a benchmark, we revisit the simulation setting in Wagner et al. (2020) and generate innovations according to

$$\boldsymbol{u}_t = \rho_1 \boldsymbol{u}_{t-1} + \boldsymbol{\varepsilon}_t + \rho_2 \boldsymbol{e}_t, \qquad \boldsymbol{v}_t = \boldsymbol{e}_t + 0.5\boldsymbol{e}_{t-1}, \tag{4.1}$$

where $\boldsymbol{\varepsilon}_t \overset{i.i.d.}{\sim} \mathcal{N}\big(\boldsymbol{0}, \boldsymbol{\Sigma}(\rho_3)\big)$, $\boldsymbol{e}_t \overset{i.i.d.}{\sim} \mathcal{N}\big(\boldsymbol{0}, \boldsymbol{\Sigma}(\rho_4)\big)$ with

$$\boldsymbol{\Sigma}(\rho) = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} \tag{4.2}$$

is a symmetric Toeplitz matrix. The parameter ρ_1 controls the level of serial correlation and ρ_2 measures the degree of endogeneity. The parameters ρ_3 and ρ_4 indicate the extent of correlation across equations induced through $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{e}_t$, respectively. For simplicity, we assume identical values $\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho \in \{0, 0.3, 0.6, 0.8\}$. The true coefficient vector is $\boldsymbol{\beta} = \left[\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_n'\right]'$, where $\boldsymbol{\beta}_i = [1, 1, 5, \beta_{i,4}]'$ with $\beta_{i,4} = -0.3$, $i = 1, \ldots, n$.

Setting B (Vector ARMA errors): To further investigate the importance of serial correlation, we consider a second specification of the innovation process:

$$\boldsymbol{u}_t = \boldsymbol{\Lambda}_1 \boldsymbol{u}_{t-1} + \boldsymbol{\eta}_t + \boldsymbol{\Lambda}_2 \boldsymbol{\eta}_{t-1}, \qquad \boldsymbol{v}_t = \boldsymbol{\Lambda}_3 \boldsymbol{v}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{4.3}$$

where $\boldsymbol{\eta}_t$ and $\boldsymbol{\varepsilon}_t$ are generated as $\left[\begin{smallmatrix}\boldsymbol{\eta}_t \\ \boldsymbol{\varepsilon}_t\end{smallmatrix}\right] \overset{i.i.d.}{\sim} \mathcal{N}\big(\boldsymbol{0}, \boldsymbol{\Sigma}(\theta)\big)$ and $\boldsymbol{\Sigma}(\theta) \in \mathbb{R}^{2n \times 2n}$ as in (4.2) but with parameter θ. The matrices $\boldsymbol{\Lambda}_i$, $i = 1, 2, 3$, are generated independently and similarly to Chang et al. (2004). That is, we take the following three steps:

(a) Generate an $n \times n$ random matrix $\boldsymbol{U}_i$ from $\mathrm{U}[0, 1]$ and construct the orthogonal matrix $\boldsymbol{H}_i = \boldsymbol{U}_i \left(\boldsymbol{U}_i' \boldsymbol{U}_i\right)^{-1/2}$.

(b) Generate n eigenvalues $\lambda_{i1}, \ldots, \lambda_{in} \overset{i.i.d.}{\sim} \mathrm{U}\left[\underline{\lambda}, \bar{\lambda}\right]$.

(c) Let $\boldsymbol{L}_i = \mathrm{diag}\left(\lambda_{i1}, \ldots, \lambda_{in}\right)$ and compute $\boldsymbol{\Lambda}_i = \boldsymbol{H}_i \boldsymbol{L}_i \boldsymbol{H}_i'$.

The parameter $\theta \in \{0.3, 0.5\}$ governs regressor-error correlation and cross-equation correlation. The amount of serial correlation is specified through $\underline{\lambda}$ and $\bar{\lambda}$. The three scenarios $\left(\underline{\lambda}, \bar{\lambda}\right) \in \left\{ (0.1, 0.5), (0.5, 0.8), (0.8, 0.95) \right\}$ steadily increase the autocorrelation in the generated data.

Setting C (Cointegration tests): We continue to construct innovations according to Setting B. Moreover, we fix $\left[\begin{smallmatrix}\boldsymbol{\eta}_t\\ \boldsymbol{\varepsilon}_t\end{smallmatrix}\right] \overset{i.i.d.}{\sim} \mathcal{N}\big(\mathbf{0}, \boldsymbol{\Sigma}(\theta)\big)$ with $\theta = 0.3$, and we construct the matrices $\boldsymbol{\Lambda}_2$ and $\boldsymbol{\Lambda}_3$ using $\left(\underline{\lambda}, \bar{\lambda}\right) = (0.1, 0.5)$. The eigenvalues of $\boldsymbol{\Lambda}_1$ are varied to explore both size and power properties. We always estimate a *quadratic* seemingly unrelated CPR.

Size DGP. We generate the eigenvalues of $\boldsymbol{\Lambda}_1$ as before. That is, take $\lambda_{11}, \ldots, \lambda_{1n} \overset{i.i.d.}{\sim} \mathrm{U}\big[\underline{\lambda}, \bar{\lambda}\big]$, where $\left(\underline{\lambda}, \bar{\lambda}\right) \in \{(0.1, 0.5), (0.5, 0.8), (0.8, 0.95)\}$.

Power DGP1. We set $\lambda_{1j} = 1$ for $1 \leq j \leq J_1$ and generate $\lambda_{1j} \overset{i.i.d.}{\sim} \mathrm{U}\big[0.1, 0.5\big]$ for $J_1 + 1 \leq j \leq n$. The integer $J_1 \in \{1, 2, n\}$ represents the number of unit roots in $\{\boldsymbol{u}_t\}$.

Power DGP2. The eigenvalues of $\boldsymbol{\Lambda}_1$ are sampled as $\lambda_{11}, \ldots, \lambda_{1n} \overset{i.i.d.}{\sim} \mathrm{U}\big[0.1, 0.5\big]$, and the first $J_2 \in \{1, 2, n\}$ series follow a cubic SUCPR specification:

$$y_{it} = \begin{cases} 1 + t + 5x_{it} - 0.3x_{it}^2 + 0.01x_{it}^3 + u_{it}, & 1 \leq i \leq J_2, \\ 1 + t + 5x_{it} - 0.3x_{it}^2 + u_{it}, & J_2 + 1 \leq i \leq n. \end{cases}$$

Power DGP3. We again take $\lambda_{11}, \ldots, \lambda_{1n} \overset{i.i.d.}{\sim} \mathrm{U}\big[0.1, 0.5\big]$ and construct

$$y_{it} = \begin{cases} \sum_{s=1}^{t} u_{is}, & 1 \leq i \leq J_3, \\ 1 + t + 5x_{it} - 0.3x_{it}^2 + u_{it}, & J_3 + 1 \leq i \leq n, \end{cases}$$

where $J_3 \in \{1, 2, n\}$ represents for the number of equations that specify a spurious relation.

Overall, the Power DGPs 1-3 consider: missing $I(1)$ regressors, omitted higher order powers of the $I(1)$ regressor x_{it}, and spurious regressions, respectively.

4.2 The simulation results

Tables 1 and 2 report the empirical mean squared error (MSE) for both feasible and infeasible estimators. As results are qualitatively similar across equations, we only report on the estimators for $\beta_{1,4}$ (the coefficient in front of x_{1t}^2). The column with FGLS contains the numerical value of the MSE and the MSEs of all other estimators are expressed relative to this benchmark. Values above 1 indicate a better performance of FM-GLS. We make the following observations:

(a) The FM-GLS estimator generally has the lowest MSE among all feasible estimators. These finite sample efficiency gains are small at low levels of endogeneity and serial correlation, but become sizeable at higher levels. Moreover, the Monte Carlo outcomes for the infeasible estimators indicate that these gains remain when the estimators are informed about the true endogeneity and serial correlation properties. It is thus the GLS weighting of the data that improves estimation accuracy.

(b) There is one particular instance in Table 2 in which the performance of the FM-GLS estimator has a high MSE, namely the case of high persistency $\left(\underline{\lambda}, \bar{\lambda}\right) = (0.8, 0.95)$, high endogeneity $\theta = 0.5$, and small sample size $T = 100$. This is caused by an inaccurate BIAM estimator resulting from the combination of a small sample size, high endogeneity, and high persistency. The problem disappears when T increases.

The subsequent set of simulations evolves around hypothesis testing, see Table 3 and Figures 1-6. The errors are simulated using Setting A and we use the following Wald-type test statistics: the Wald-SOLS and Wald-SUR tests as developed in Proposition 2 in Wagner et al. (2020), and the Wald-FGLS test from Theorem 4. We consider: (i) the single equation test $H_0 : \beta_{1,4} = -0.3$ against the two-sided alternative $H_1 : \beta_{1,4} \neq -0.3$, and (ii) the joint test $H_0 : \beta_{1,4} = \beta_{2,4} = \ldots = \beta_{n,4} = -0.3$ against the alternative which rejects when at least one coefficient is unequal to -0.3. Some general remarks regarding size and size-corrected power are as follows.

(c) The Wald tests are typically oversized but the three tests react differently to changes in ρ. Increases in ρ result in an increasing size for the SOLS and SUR versions of the Wald test, whereas increases in ρ lead to size decreases for the GLS-type of Wald test. Overall, the GLS test provides better size control.

(d) In Figures 1 and 2, we vary the serial correlation parameter ρ_1 and the endogeneity parameter ρ_2 separately. Overall, variation in ρ_1 has a larger influence on size with the SUR test being the most sensitive and the GLS test being the least sensitive.

(e) For all three Wald-type of tests, the size of the tests improves with sample size T.

(f) The ordering in terms of size-corrected power is the same throughout Figures 3-6. That is, size-corrected power is lowest for Wald-SOLS, increases for Wald-SUR, and is highest for the Wald-FGLS test.

The simulation results for the KPSS-type of cointegration tests can be found in Table 4. The general conclusions are as follows.

(g) The empirical sizes of the K_{max}^{SOLS} and K_{max}^{SUR} tests are similar. We observe very conservative results for low serial correlation, decent size for medium serial correlation, and strongly oversized tests at high levels of serial correlation. These findings are completely in line with the simulation results that are reported in table 3 of Choi and Saikkonen (2010). The same behavior is observed for the K_{max}^{FGLS} test but the deviations from the 5% level are less extreme.

(h) The power of the K_{max}^{OLS}, K_{max}^{SUR}, and K_{max}^{FGLS} tests behave as expected: (1) power always increases with sample size, and (2) power increases when more unit roots, more misspecified equations, or more spurious relationships are incorporated in the DGP. The K_{max}^{FGLS} test has the lowest power among the three tests. This is caused by the fact that the filter can nearly difference the data and hence make it appear more stationary.

5 Conclusion

We investigated the performance of three fully modified (FM) estimators in seemingly unrelated cointegrating polynomial regressions. An FM-GLS estimator is proposed to eliminate nuisance parameters for conducting inference. Monte Carlo simulations revealed the advantages and disadvantages of these FM methods. Some interesting questions are left for future research. From a theoretical viewpoint, it is interesting to study the behavior of the modified Cholesky decomposition (and BIAM) when the series under consideration is nonstationary. This would give insights into the behavior of: (1) the FM-GLS estimator while estimating spurious regressions, and (2) the power properties of the cointegration tests.

Acknowledgements

This paper has been presented at the 2018 CFE meeting, the NESG 2019 conference, the 6th RCEA Time Series Econometrics Workshop, and the Department of Econometrics and Data Science at the Vrije Universiteit Amsterdam. We would like to thank the conference participants, especially Peter Pedroni and Peter Phillips, for their useful comments and suggestions. We extend our thanks to Eric Beutner, Dick van Dijk, Richard Paap, Franz Palm, and Stephan Smeekes, for their valuable feedback on earlier versions of this manuscript. All remaining errors are our own.

References

Abadir, K. M. and J. R. Magnus (2005). *Matrix Algebra*. Cambridge University Press.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics 2*, 489–502.

Beutner, E., Y. Lin, and S. Smeekes (2022). GLS estimation and confidence sets for the date of a single break in models with trends. *Econometric Reviews*. Forthcoming.

Bickel, P. J. and E. Levina (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics 36*, 199–227.

Breitung, J. (2001). Rank tests for nonlinear cointegration. *Journal of Business & Economic Statistics 19*, 331–340.

Chang, Y., J. Y. Park, and P. C. B. Phillips (2004). Bootstrap unit root tests in panels with cross-sectional dependency. *Journal of Econometrics 120*, 263–293.

Cheng, T.-C. F., C.-K. Ing, and S.-H. Yu (2015). Toward optimal model averaging in regression models with time series errors. *Journal of Econometrics 189*, 321–334.

Choi, I. and P. Saikkonen (2010). Tests for nonlinear cointegration. *Econometric Theory 26*, 682–709.

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

Davidson, R. and J. G. MacKinnon (2004). *Econometric Theory and Methods*. Oxford University Press.

de Jong, R. M. (2002). Nonlinear estimators with integrated regressors but without exogeneity. mimeo Michigan State University.

Grossman, G. M. and A. B. Krueger (1995). Economic growth and the environment. *The Quarterly Journal of Economics 110*, 353–377.

Hannan, E. and M. Deistler (2012). *The Statistical Theory of Linear Systems*. Society for Industrial and Applied Mathematics.

Hong, S. H. and P. C. B. Phillips (2010). Testing linearity in cointegrating relations with an application to purchasing power parity. *Journal of Business & Economic Statistics 28*, 96–114.

Ing, C.-K., H.-T. Chiou, and M. Guo (2016a). Estimation of inverse autocovariance matrices for long memory processes. *Bernoulli 22*, 1301–1330.

Ing, C.-K., H.-T. Chiou, and M. Guo (2016b). Estimation of inverse autocovariance matrices for long memory processes. *Bernoulli 22*, 1301–1330.

Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory 18*, 1449–1459.

Kim, C. and D. L. Zimmerman (2012). Unconstrained models for the covariance structure of multivariate longitudinal data. *Journal of Multivariate Analysis 107*, 104–118.

Kohli, P., T. P. Garcia, and M. Pourahmadi (2016). Modeling the cholesky factors of covariance matrices of multivariate longitudinal data. *Journal of Multivariate Analysis 145*, 87–100.

Li, D., P. C. B. Phillips, and J. Gao (2020). Kernel-based inference in time-varying coefficient cointegrating regression. *Journal of Econometrics 215*, 607–632.

Mark, N. C., M. Ogaki, and D. Sul (2005). Dynamic seemingly unrelated cointegrating regressions. *The Review of Economic Studies 72*, 797–820.

McMurry, T. L. and D. N. Politis (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis 31*, 471–482.

Moon, H. R. and B. Perron (2005). Efficient estimation of the seemingly unrelated regression cointegration model and testing for purchasing power parity. *Econometric Reviews 23*, 293–323.

Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies 61*, 631–653.

Nyblom, J. and A. Harvey (2000). Tests of common stochastic trends. *Econometric Theory 16*, 176–199.

Park, J. Y. and P. C. B. Phillips (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory 15*(3), 269–298.

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

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica 59*, 283–306.

Phillips, P. C. B. (1995). Fully modified least squares and vector autoregression. *Econometrica 63*, 1023–1078.

Phillips, P. C. B. and B. E. Hansen (1990). Statistical inference in instrumental variables regression with I(1) processes. *The Review of Economic Studies 57*, 99–125.

Phillips, P. C. B. and J. Y. Park (1988). Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors. *Journal of the American Statistical Association 83*, 111–115.

Phillips, P. C. B. and V. Solo (1992). Asymptotics for linear processes. *The Annals of Statistics 20*, 971–1001.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika 86*, 677–690.

Prais, S. J. and C. B. Winsten (1954). Trend estimators and serial correlation. Cowles Foundation, Discussion Paper 383.

Romano, J. P. and M. Wolf (2001). Subsampling intervals in autoregressive models with linear time trend. *Econometrica 69*, 1283–1314.

Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory 8*, 1–27.

Shin, Y. (1994). A residual-based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory 10*, 91–115.

Sims, C. A., J. H. Stock, and M. W. Watson (1990). Inference in linear time series mmodels with some unit roots. *Econometrica 58*, 113–144.

Tjøstheim, D. (2020). Some notes on nonlinear cointegration: A partial review with some novel perspectives. *Econometric Reviews 39*, 655–673.

Vogelsang, T. J. and M. Wagner (2014). Integrated modified OLS estimation and fixed-b inference for cointegrating regressions. *Journal of Econometrics 178*, 741–760.

Wagner, M., P. Grabarczyk, and S. H. Hong (2020). Fully modified OLS estimation and inference for seemingly unrelated cointegrating polynomial regressions and the environmental Kuznets curve for carbon dioxide emissions. *Journal of Econometrics 214*, 216–255.

Wagner, M. and S. H. Hong (2016). Cointegrating polynomial regressions: Fully modified OLS estimation and inference. *Econometric Theory 32*, 1289–1315.

Wang, Q. (2015). *Limit Theorems for Nonlinear Cointegrating Regression*. World Scientific.

Wang, Q. and P. C. B. Phillips (2009). Structural nonparametric cointegrating regression. *Econometrica 77*, 1901–1948.

Wu, W. B. and M. Pourahmadi (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica 19*, 1755–1768.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association 57*, 348–368.

Appendix A Proof of main results

We provide proof of the main results in this appendix. Some intermediate lemmas (indicated with S and a number) can be found in the online supplemental material.

Lemma 1

Let $\boldsymbol{\mathcal{A}}_q(L) = \boldsymbol{I}_n - \sum_{j=1}^{q} \boldsymbol{A}_j(q)L^j$ denote the lag polynomial implied by the coefficient matrices in (2.3). By the Beveridge-Nelson (BN) decomposition, we also have $\boldsymbol{\mathcal{A}}_q(L) = \boldsymbol{\mathcal{A}}_q(1) + (1-L)\widetilde{\boldsymbol{\mathcal{A}}}_q(L)$ where $\widetilde{\boldsymbol{\mathcal{A}}}_q(L) = \sum_{j=1}^{q} \widetilde{\boldsymbol{A}}_j(q)L^{j-1}$ with $\widetilde{\boldsymbol{A}}_j(q) = \sum_{i=j}^{q} \boldsymbol{A}_i(q)$. If Assumption 1 holds, then

(a) $\boldsymbol{\mathcal{A}}_q(1) = \boldsymbol{\mathcal{A}}(1) + O\left(\sum_{j=q+1}^{\infty} j^{1/2}\left\|\boldsymbol{A}_j\right\|_{\mathcal{F}}\right)$;

(b) there exists a $q^* > 0$ such that $\sum_{j=1}^{q}\left\|\widetilde{\boldsymbol{A}}_j(q)\right\|_{\mathcal{F}} < \infty$ for all $q > q^*$.

Proof **(a)** By Cauchy-Schwartz inequality, we have $\left\|\boldsymbol{\mathcal{A}}_q(1) - \boldsymbol{\mathcal{A}}(1)\right\|_{\mathcal{F}} \leq \sum_{j=1}^{q}\left\|\boldsymbol{A}_j(q) - \boldsymbol{A}_j\right\|_{\mathcal{F}} + \sum_{j=q+1}^{\infty}\left\|\boldsymbol{A}_j\right\|_{\mathcal{F}} \leq \left(q\sum_{j=1}^{q}\left\|\boldsymbol{A}_j(q) - \boldsymbol{A}_j\right\|_{\mathcal{F}}^2\right)^{1/2} + \sum_{j=q+1}^{\infty}\left\|\boldsymbol{A}_j\right\|_{\mathcal{F}}$ and subsequently use Lemma S1 (see Supplement). **(b)** Using Baxter's inequality in Theorem 6.6.12 in Hannan and Deistler (2012) (also see their Remark 3) for the final inequality, we derive $\sum_{j=1}^{q}\left\|\widetilde{\boldsymbol{A}}_j(q)\right\|_{\mathcal{F}} \leq \sum_{j=1}^{q} j\left\|\boldsymbol{A}_j(q)\right\|_{\mathcal{F}} \leq \sum_{j=1}^{q} j\left\|\boldsymbol{A}_j(q) - \boldsymbol{A}_j\right\|_{\mathcal{F}} + \sum_{j=1}^{q} j\left\|\boldsymbol{A}_j\right\|_{\mathcal{F}} \leq C\sum_{j=1}^{q} j\left\|\boldsymbol{A}_j\right\|_{\mathcal{F}} \leq C$. ■

Proof of Theorem 1 The premultiplication by $\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)$ applies a linear filter whereas $\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q)$ implies weighting. Since the behaviour of the first $q \ll T$ elements does not affect the asymptotic results, we take $\boldsymbol{Z}_t = \boldsymbol{u}_t = \mathbf{O}$ for $t \leq 0$ and for all $t = 1, 2\ldots$, we apply the transformations implied by $\boldsymbol{\mathcal{A}}_q(L)$ and $\boldsymbol{S}^{-1}(q)$.[7] Consequently, we have

$$\begin{aligned}\boldsymbol{G}_T^{-1}\left(\widehat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta}\right) = &\left[\sum_{t=1}^{T}\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)'\boldsymbol{S}^{-1}(q)\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)\right]^{-1} \\ &\times\sum_{t=1}^{T}\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)'\boldsymbol{S}^{-1}(q)\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{u}_t\right) + o_p(1).\end{aligned} \tag{A.1}$$

Using the BN decomposition, Lemma 1, we first show

$$\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T = \boldsymbol{\mathcal{A}}_q(1)\boldsymbol{Z}_t'\boldsymbol{G}_T + \sum_{j=1}^{q}\widetilde{\boldsymbol{A}}_j(q)\Delta\boldsymbol{Z}_{t-j+1}'\boldsymbol{G}_T = \boldsymbol{\mathcal{A}}(1)\boldsymbol{Z}_t'\boldsymbol{G}_T + O_p\left(\frac{1}{\sqrt{qT}}\right). \tag{A.2}$$

Note that $\left\|\sum_{j=1}^{q}\widetilde{\boldsymbol{A}}_j(q)\Delta\boldsymbol{Z}_{t-j+1}'\boldsymbol{G}_T\right\| \leq \sum_{j=1}^{q}\left\|\widetilde{\boldsymbol{A}}_j(q)\right\|\left\|\Delta\boldsymbol{Z}_{t-j+1}'\boldsymbol{G}_T\right\|$ and that for any t,

$$\left\|\Delta\boldsymbol{Z}_t'\boldsymbol{G}_T\right\| = \max_{1\leq i\leq n}\left\|\boldsymbol{G}_{i,T}\Delta\boldsymbol{z}_{it}\right\| = \max_{1\leq i\leq n}\left(\left\|\boldsymbol{G}_{\boldsymbol{d}_i,T}\Delta\boldsymbol{d}_{it}\right\|^2 + \left\|\boldsymbol{G}_{\boldsymbol{s}_i,T}\Delta\boldsymbol{s}_{it}\right\|^2\right)^{1/2}. \tag{A.3}$$

[7] The same argumentation is used in Phillips and Park (1988). The modification to obtain rigorous proof is straightforward.

The vector $\boldsymbol{G}_{\boldsymbol{d}_i,T}\Delta\boldsymbol{d}_{it}$ typically contains elements $T^{-(k+1/2)}\left[t^k-(t-1)^k\right]$ where $k=0,1,\cdots,d_i$. By the inequality $(a+b)^n\le a^n+nb(a+b)^{n-1}$, for $a,b\ge 0$, $n\in\mathbb{N}$, we obtain $0\le t^k-(t-1)^k\le kt^{k-1}$, and thus $T^{-(k+1/2)}\left[t^k-(t-1)^k\right]\le d_iT^{-3/2}\le CT^{-3/2}$. As a result, $\|\boldsymbol{G}_{\boldsymbol{d}_i,T}\Delta\boldsymbol{d}_{it}\|^2\le CT^{-3}$. The vector $\boldsymbol{G}_{\boldsymbol{s}_i,T}\Delta\boldsymbol{s}_{it}$ typically contains elements $T^{-(k+1)/2}\left(x_{it}^k-x_{it-1}^k\right)$, where $k=1,\ldots,s_i$. The binomial expansion implies $x_{it}^k-x_{it-1}^k=\sum_{m=0}^{k-1}\binom{k}{m}x_{it-1}^m v_{it}^{k-m}=O_p\left(T^{(k-1)/2}\right)$, and thus $T^{-(k+1)/2}\left(x_{it}^k-x_{it-1}^k\right)=O_p\left(T^{-1}\right)$. It further implies that $\|\boldsymbol{G}_{\boldsymbol{s}_i,T}\Delta\boldsymbol{s}_{it}\|^2=O_p\left(T^{-2}\right)$. Combining $\left\|\Delta\boldsymbol{Z}_t'\boldsymbol{G}_T\right\|=O_p\left(T^{-1}\right)$ with Lemma 1(b), and using Lemma 1(a) and $\sum_{j=q+1}^{\infty}j^{1/2}\left\|\boldsymbol{A}_j\right\|_{\mathcal{F}}=o\left(q^{-1/2}\right)$, we have (A.2).

Using (A.2) and $\left\|\boldsymbol{S}(q)-\boldsymbol{\Sigma}_{\eta\eta}\right\|\to 0$ (see Eq. (S.17) in the supplementary material), we obtain

$$\sum_{t=1}^{T}\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)'\boldsymbol{S}^{-1}(q)\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)$$
$$=\sum_{t=1}^{T}\boldsymbol{G}_T\boldsymbol{Z}_t\boldsymbol{\Omega}_{uu}^{-1}\boldsymbol{Z}_t'\boldsymbol{G}_T+O_p\left(q^{-1/2}\right)\Rightarrow\int_0^1\boldsymbol{J}(r)\boldsymbol{\Omega}_{uu}^{-1}\boldsymbol{J}(r)'dr. \quad \text{(A.4)}$$

To continue, we rewrite the second part in (A.1) as

$$\begin{aligned}\sum_{t=1}^{T}\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)'\boldsymbol{S}^{-1}(q)\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{u}_t\right)&=\sum_{t=1}^{T}\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)'\boldsymbol{S}^{-1}(q)\boldsymbol{\eta}_t\\&\quad+\sum_{t=1}^{T}\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)'\boldsymbol{S}^{-1}(q)\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{u}_t-\boldsymbol{\eta}_t\right)=:I+II,\end{aligned} \quad \text{(A.5)}$$

and we will repeatedly use the identity

$$\sum_t\boldsymbol{J}_t\boldsymbol{D}\boldsymbol{e}_t=\sum_t\left(\boldsymbol{e}_t'\otimes\boldsymbol{J}_t\right)\mathrm{vec}(\boldsymbol{D})=\left[\sum_t\boldsymbol{J}_te_{1t},\sum_t\boldsymbol{J}_te_{2t},\cdots,\sum_t\boldsymbol{J}_te_{nt}\right]\mathrm{vec}(\boldsymbol{D}) \quad \text{(A.6)}$$

for any matrices $\boldsymbol{J}_t\in\mathbb{R}^{d\times n}$, $\boldsymbol{D}\in\mathbb{R}^{n\times n}$, and $\boldsymbol{e}_t=[e_{1t},\ldots,e_{nt}]'\in\mathbb{R}^{n\times 1}$. Then

$$I=\left[\sum_{t=1}^{T}\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)'\eta_{1t},\cdots,\sum_{t=1}^{T}\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)'\eta_{nt}\right]\mathrm{vec}\left(\boldsymbol{\Sigma}_{\eta\eta}^{-1}+o(1)\right), \quad \text{(A.7)}$$

where η_{it} is the ith entry of $\boldsymbol{\eta}_t$. For $1\le i\le n$, we use the BN decomposition to obtain

$$\begin{aligned}&\sum_{t=1}^{T}\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)'\eta_{it}\\&=\sum_{t=1}^{T}\boldsymbol{G}_T\boldsymbol{Z}_t\eta_{it}\boldsymbol{\mathcal{A}}_q(1)'+\sum_{t=1}^{T}\boldsymbol{G}_T\Delta\boldsymbol{Z}_t\eta_{it}\widetilde{\boldsymbol{A}}_1(q)'+\sum_{j=2}^{q}\sum_{t=1}^{T}\boldsymbol{G}_T\Delta\boldsymbol{Z}_{t-j+1}\eta_{it}\widetilde{\boldsymbol{A}}_j(q)'.\end{aligned} \quad \text{(A.8)}$$

By definition, we have $\sum_{t=1}^{T} \boldsymbol{G}_T \boldsymbol{Z}_t \eta_{it} = \text{diag}\left[\sum_{t=1}^{T} \boldsymbol{G}_{1,T} \boldsymbol{z}_{1t}, \ldots, \sum_{t=1}^{T} \boldsymbol{G}_{n,T} \boldsymbol{z}_{nt}\right] \eta_{it}$, where the limiting distribution of each block follows from Proposition 1 of Wagner and Hong (2016). More specifically, the kth block will converge to a stochastic integral and a second-order bias term which is proportional to $\boldsymbol{\Sigma}_{\epsilon_k \eta_i} := \mathbb{E}\left(\epsilon_{kt}\eta_{it}\right)$ (the $(k,i)^{th}$ element of $\boldsymbol{\Sigma}_{\epsilon\eta}$), $1 \leq k, i \leq n$, and thus

$$\sum_{t=1}^{T} \boldsymbol{G}_T \boldsymbol{Z}_t \eta_{it} \boldsymbol{\mathcal{A}}_q(1)' \Rightarrow \left(\int_0^1 \boldsymbol{J}(r) d\boldsymbol{B}_{\eta_i}(r) + \boldsymbol{B}_i\right) \boldsymbol{\mathcal{A}}(1)', \tag{A.9}$$

where $\boldsymbol{B}_i := \text{diag}\left[\boldsymbol{\Sigma}_{\epsilon_1 \eta_i}\ \boldsymbol{b}_1, \cdots, \boldsymbol{\Sigma}_{\epsilon_n \eta_i}\ \boldsymbol{b}_n\right]$.

As $\sum_{t=1}^{T} \boldsymbol{G}_T \Delta \boldsymbol{Z}_t \eta_{it} \widetilde{\boldsymbol{A}}_1(q)' = \text{diag}\left[\sum_{t=1}^{T} \boldsymbol{G}_{1,T} \Delta \boldsymbol{z}_{1t}, \ldots, \sum_{t=1}^{T} \boldsymbol{G}_{n,T} \Delta \boldsymbol{z}_{nt}\right] \eta_{it} \left(\sum_{j=1}^{\infty} \boldsymbol{A}_j + o(1)\right)'$, we again consider the limiting distributions block-wise. Every element in the blocks will rely on one of the following three results. (1) As derived below (A.3), we have $\left|T^{-(j+\frac{1}{2})} \sum_{t=1}^{T} \left[t^j - (t-1)^j\right] \eta_{it}\right| \leq CT^{-3/2} \sum_{t=1}^{T} |\eta_{it}| = o_p(1)$. (2) By Assumption 1, for any $1 \leq k, i \leq n$, v_{kt} and η_{it} are Near Epoch Dependent in L_4-norm on $\{[\boldsymbol{\eta}_t', \boldsymbol{\varepsilon}_t']'\}_{t \in \mathbb{Z}}$ of size -1 and arbitrary size, respectively. A small variation on Theorem 17.9 from Davidson (1994) shows that $\{v_{it}\eta_{it}\}$ are L_2-NED of size -1. The i.i.d. assumption on $\{[\boldsymbol{\eta}_t', \boldsymbol{\varepsilon}_t']'\}$ allows for a LLN for the sequence $\{v_{it}\eta_{it}\}$, see e.g., Theorem 20.21 of Davidson (1994), implying $T^{-1} \sum_{t=1}^{T} \Delta x_{kt} \eta_{it} = T^{-1} \sum_{t=1}^{T} v_{kt} \eta_{it} \longrightarrow_p \boldsymbol{\Sigma}_{\epsilon_k \eta_i}$. (3) $T^{-(j+1)/2} \sum_{t=1}^{T} \Delta x_{kt}^j \eta_{it} \Rightarrow j \boldsymbol{\Sigma}_{\epsilon_k \eta_i} \int_0^1 \boldsymbol{B}_{v_k}^{j-1}(r) dr$, where $j \geq 2$. The specific reason is as follows. By the binomial expansion (below (A.3)), we have

$$\begin{aligned}
T^{-(j+1)/2} \sum_{t=1}^{T} \Delta x_{kt}^j \eta_{it} &= T^{-(j+1)/2} \sum_{m=0}^{j-1} \binom{j}{m} \sum_{t=1}^{T} x_{kt-1}^m v_{kt}^{j-m} \eta_{it} = jT^{-(j+1)/2} \sum_{t=1}^{T} x_{kt-1}^{j-1} v_{kt} \eta_{it} + o_p(1) \\
&= j \boldsymbol{\Sigma}_{\epsilon_k \eta_i} \frac{1}{T} \sum_{t=1}^{T-1} \left(\frac{x_{kt}}{\sqrt{T}}\right)^{j-1} + \frac{j}{\sqrt{T}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} \left(\frac{x_{kt}}{\sqrt{T}}\right)^{j-1} \left(v_{kt}\eta_{it} - \boldsymbol{\Sigma}_{\epsilon_k \eta_i}\right) + o_p(1) \\
&\Rightarrow j \boldsymbol{\Sigma}_{\epsilon_k \eta_i} \int_0^1 \boldsymbol{B}_{v_k}^{j-1}(r) dr,
\end{aligned}$$

where $T^{-1/2} \sum_{t=1}^{T-1} \left(x_{kt}/\sqrt{T}\right)^{j-1} \left(v_{kt}\eta_{it} - \boldsymbol{\Sigma}_{\epsilon_k \eta_i}\right) = O_p(1)$. To see this, we refer to de Jong (2002). The moment and NED conditions in Assumption 1 in de Jong (2002) are satisfied. Moreover, since $F(x) = x^{j-1}$ is homogeneous of degree $j-1$, Assumption 2 in de Jong (2002) holds as well. The desired result now follows from Theorem 1 in de Jong (2002). Combining these results, we obtain

$$\sum_{t=1}^{T} \boldsymbol{G}_T \Delta \boldsymbol{Z}_t \eta_{it} \widetilde{\boldsymbol{A}}_1(q)' \Rightarrow \boldsymbol{B}_i \left(\sum_{j=1}^{\infty} \boldsymbol{A}_j\right)'. \tag{A.10}$$

Finally, the last term in (A.8) is bounded by $\sum_{j=2}^{q} \left\|\boldsymbol{A}_j(q)\right\| \left\|\sum_{t=1}^{T} \boldsymbol{G}_T \Delta \boldsymbol{Z}_{t-j+1} \eta_{it}\right\|$. Using similar arguments above, we conclude that $\sum_{t=1}^{T} \boldsymbol{G}_T \Delta \boldsymbol{Z}_{t-j+1} \eta_{it} = o_p(1)$ (lags of $\Delta \boldsymbol{Z}_t$ will lead

to $\mathbb{E}\big(v_{kt-j}\eta_{it}\big) = 0$ for any $j > 0$). Hence, $\sum_{j=2}^{q}\sum_{t=1}^{T} \boldsymbol{G}_T \Delta \boldsymbol{Z}_{t-j+1}\eta_{it}\widetilde{\boldsymbol{A}}_j(q)' = o_p(1)$. Combining (A.8), (A.9) and (A.10), we have

$$\sum_{t=1}^{T} \left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)'\eta_{it} \Rightarrow \int_0^1 \boldsymbol{J}(r)d\boldsymbol{B}_{\eta_i}(r)\boldsymbol{\mathcal{A}}(1)' + \boldsymbol{B}_i. \tag{A.11}$$

Note that $\operatorname{vec}\left(\boldsymbol{\Sigma}_{\eta\eta}^{-1}\right) = \begin{bmatrix} \operatorname{col}_1(\boldsymbol{\Sigma}_{\eta\eta}^{-1}) \\ \vdots \\ \operatorname{col}_n(\boldsymbol{\Sigma}_{\eta\eta}^{-1}) \end{bmatrix}$. Inserting (A.11) into (A.7), we eventually have

$$\begin{aligned} I &\Rightarrow \int_0^1 \left(d\boldsymbol{B}_\eta(r)' \otimes \boldsymbol{J}(r)\boldsymbol{\mathcal{A}}(1)'\right)\operatorname{vec}\left(\boldsymbol{\Sigma}_{\eta\eta}^{-1}\right) + [\boldsymbol{B}_1, \cdots, \boldsymbol{B}_n]\operatorname{vec}\left(\boldsymbol{\Sigma}_{\eta\eta}^{-1}\right) \\ &= \int_0^1 \boldsymbol{J}(r)\boldsymbol{\mathcal{A}}(1)'\boldsymbol{\Sigma}_{\eta\eta}^{-1}d\boldsymbol{B}_\eta(r) + \sum_{i=1}^{n}\boldsymbol{B}_i \operatorname{col}_i\left(\boldsymbol{\Sigma}_{\eta\eta}^{-1}\right) = \int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{uu}^{-1}d\boldsymbol{B}_u(r) + \boldsymbol{\mathcal{B}}_{\epsilon\eta}, \end{aligned} \tag{A.12}$$

where the symmetry property of $\boldsymbol{\Sigma}_{\eta\eta}^{-1}$ is used in the final step.

Now we consider the term II in (A.5). If we define $\boldsymbol{u}_t^* = \left[u_{1t}^*, \ldots, u_{nt}^*\right] := \boldsymbol{\mathcal{A}}_q(L)\boldsymbol{u}_t - \boldsymbol{\eta}_t = -\sum_{j=1}^{\infty}\left(\boldsymbol{A}_j(q) - \boldsymbol{A}_j\right)\boldsymbol{u}_{t-j}$, where $\boldsymbol{A}_j(q) = \mathbf{O}$ for $j > q$, and then apply (A.6), we have $II = \left[\sum_{t=1}^{T}\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)'u_{1t}^*, \ldots, \sum_{t=1}^{T}\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)'u_{nt}^*\right]\operatorname{vec}\left(\boldsymbol{\Sigma}_{\eta\eta}^{-1} + o(1)\right)$. For any block $i = 1, \ldots, n$, by the BN decomposition (A.2), we have

$$\begin{aligned} &\left\|\sum_{t=1}^{T}\left(\boldsymbol{\mathcal{A}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)'u_{it}^*\right\| \\ &= \left\|\sum_{t=1}^{T}\boldsymbol{G}_T\boldsymbol{Z}_t u_{it}^*\boldsymbol{\mathcal{A}}_q(1)' + O_p\big(T^{-1}\big)\sum_{t=1}^{T}u_{it}^*\right\| \\ &\le C\max_{1\le k\le n}\left\|\sum_{t=1}^{T}\boldsymbol{G}_{k,T}\boldsymbol{z}_{kt}u_{it}^*\right\| + O_p\big(T^{-1}\big)\sum_{j=1}^{\infty}\left\|\operatorname{row}_i\left(\boldsymbol{A}_j(q) - \boldsymbol{A}_j\right)\right\|\left\|\sum_{t=1}^{T}\boldsymbol{u}_{t-j}\right\| \\ &= O_p\left(\sum_{j=q+1}^{\infty} j^{1/2}\left\|\boldsymbol{A}_j\right\|_{\mathcal{F}}\right) = o_p(1). \end{aligned}$$

It implies $II = o_p(1)$. The theorem now follows from (A.1), (A.4), and (A.12). ■

Proof of Theorem 2 We start wih the estimation error $\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q) - \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q)$. Repeated addition and subtraction yields

$$\begin{aligned} \left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q) - \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q)\right\| &\le \left\|\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)\right\|\left\|\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{u}}^{-1}(q)\right\|\left\|\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{u}}(q)\right\| \\ &+ \left\|\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)\right\|\left\|\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{u}}^{-1}(q) - \boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q)\right\|\left\|\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{u}}(q)\right\| + \left\|\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)\right\|\left\|\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q)\right\|\left\|\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)\right\|. \end{aligned} \tag{A.13}$$

We will only consider the terms $\left\|\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)\right\|$ and $\left\|\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{u}}^{-1}(q) - \boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q)\right\|$. It is not hard to

derive that the remaining terms are bounded in probability. Define $\boldsymbol{\mathcal{G}} = \widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)$ and denote its $(n \times n)$ subblocks by $\boldsymbol{\mathcal{G}}_{ij}$, $1 \leq i, j \leq T$. This matrix $\boldsymbol{\mathcal{G}}$ is banded in such a way that there are at most $2q-1$ nonzero block in the block-columns of $\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{G}}'$. Using this observation and various norm properties, we find

$$\begin{aligned}
\|\boldsymbol{\mathcal{G}}\|^2 \leq \|\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{G}}'\|_1 &\leq \max_{1\leq j\leq T} \sum_{i=1}^{T} \left\| \left(\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{G}}'\right)_{ij} \right\|_1 \leq n(2q-1) \max_{1\leq j\leq T} \sum_{t=1}^{T} \|\boldsymbol{\mathcal{G}}_{jt}\|_1^2 \\
&= n(2q-1) \max_{1\leq \ell\leq q} \sum_{s=1}^{\ell} \left\|\widehat{\boldsymbol{A}}_s(\ell) - \boldsymbol{A}_s(\ell)\right\|_1^2 \leq n^3(2q-1) \max_{1\leq \ell\leq q} \left\|\widehat{\boldsymbol{A}}(\ell) - \boldsymbol{A}(\ell)\right\|^2 = O_p\left(\frac{q^3}{T}\right),
\end{aligned}$$

where the final step follows from Lemma S3.[8] We conclude that $\left\|\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)\right\| = O_p\left(\sqrt{q^3/T}\right)$. The difference $\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{u}}^{-1}(q) - \boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q)$ forms a symmetric and block diagonal matrix, hence $\left\|\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{S}}_{\boldsymbol{u}}(q)\right\| = \max\left\{\left\|\widehat{\boldsymbol{S}}(0) - \boldsymbol{S}(0)\right\|, \max_{1\leq\ell\leq q}\left\|\widehat{\boldsymbol{S}}(\ell) - \boldsymbol{S}(\ell)\right\|\right\}$. By Assumption 2,

$$\begin{aligned}
\left\|\widehat{\boldsymbol{S}}(0) - \boldsymbol{S}(0)\right\| &= \left\| T^{-1}\sum_{t=1}^{T} \widehat{\boldsymbol{u}}_t\widehat{\boldsymbol{u}}_t' - \mathbb{E}(\boldsymbol{u}_t\boldsymbol{u}_t') \right\| \\
&\leq T^{-1}\left\|\widehat{\boldsymbol{u}} - \boldsymbol{u}\right\|^2 + 2\left\| T^{-1}\sum_{t=1}^{T}(\widehat{\boldsymbol{u}}_t - \boldsymbol{u}_t)\boldsymbol{u}_t' \right\| + \left\| T^{-1}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{u}_t' - \mathbb{E}(\boldsymbol{u}_t\boldsymbol{u}_t') \right\| \\
&= O_p\big(T^{-1/2}\big).
\end{aligned}$$

Applying Lemma S3, we see $\left\|\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{u}}^{-1}(q) - \boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q)\right\| \leq \left\|\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{S}}_{\boldsymbol{u}}(q)\right\| \left\|\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{u}}^{-1}(q)\right\| \left\|\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q)\right\| = O_p\left(q/\sqrt{T}\right)$. Overall, recalling (A.13), a bound on the estimation eror is $\left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q) - \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q)\right\| = O_p\left(\sqrt{q^3/T}\right)$.

The bound on the truncation error, $\|\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q) - \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\| \leq Cq^{-1/2}\sum_{s=q+1}^{\infty} s\,\|\boldsymbol{A}_s\|_{\mathcal{F}}$, follows from a straightforward generalization of the results in Lemma 2 of Cheng et al. (2015) and Propositions 2.1-2.2 of Ing et al. (2016b). See Lemma S4 in the supplementary material for details. ■

Proof of Theorem 3 **(a)-(b)** See the proof of Proposition 1 in Wagner et al. (2020). **(c)** Since the residuals $\{\widehat{\boldsymbol{u}}_t\}$ are obtained by first stage OLS, we get $\|\widehat{\boldsymbol{u}} - \boldsymbol{u}\|^2 \leq \|\boldsymbol{G}_T^{-1}(\widehat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta})\|^2 \|\boldsymbol{G}_T\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{G}_T\| = O_p(1)$. Assumption 2 is thus satisfied and we can rely on the results in Theorem 2. From (3.12), the definition of the FM-GLS estimator, we have

$$\begin{aligned}
&\boldsymbol{G}_T^{-1}\left(\widehat{\boldsymbol{\beta}}_{FGLS}^{+} - \boldsymbol{\beta}\right) \\
&\quad = \left(\boldsymbol{G}_T\boldsymbol{Z}'\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q)\boldsymbol{Z}\boldsymbol{G}_T\right)^{-1}\left[\boldsymbol{G}_T\boldsymbol{Z}'\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q)\boldsymbol{u} - \boldsymbol{G}_T\boldsymbol{Z}'\left(\boldsymbol{I}_T \otimes \widehat{\boldsymbol{\Omega}}_{uu}^{-1}\widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\right)\boldsymbol{v} - \boldsymbol{G}_T\widehat{\boldsymbol{\mathcal{B}}}^{+}\right].
\end{aligned}$$

[8] More specifically, for any matrix $\boldsymbol{Q}$ we have $\|\boldsymbol{Q}\|^2 \leq \|\boldsymbol{Q}\boldsymbol{Q}'\|_1$. Moreover, if $\boldsymbol{Q}$ is an $(n \times n)$ matrix, then also $\|\boldsymbol{Q}\|_1 \leq \sqrt{n}\|\boldsymbol{Q}\|_{\mathcal{F}} \leq n\|\boldsymbol{Q}\|$.

Given Theorem 2, we have $\boldsymbol{G}_T\boldsymbol{Z}'\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q)\boldsymbol{Z}\boldsymbol{G}_T = \boldsymbol{G}_T\boldsymbol{Z}'\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q)\boldsymbol{Z}\boldsymbol{G}_T + o_p(1)$ and it converges weakly to the expression in (A.4).

To continue, we define $\widehat{\boldsymbol{\mathcal{A}}}_q(L) = \boldsymbol{I}_n - \sum_{j=1}^q \widehat{\boldsymbol{A}}_j(q)L^j$ and its BN decomposition $\widehat{\boldsymbol{\mathcal{A}}}_q(L) = \widehat{\boldsymbol{\mathcal{A}}}_q(1) + (1-L)\boldsymbol{\mathcal{A}}_q^*(L)$ through $\boldsymbol{\mathcal{A}}_q^*(L) = \sum_{j=1}^q \boldsymbol{A}_j^*(q)L^{j-1}$ with $\boldsymbol{A}_j^*(q) = \sum_{i=j}^q \widehat{\boldsymbol{A}}_i(q)$. $\widehat{\boldsymbol{\mathcal{A}}}_q(1) = \boldsymbol{\mathcal{A}}_q(1) + o_p(1)$ and $\sum_{j=1}^q \left\|\boldsymbol{A}_j^*(q)\right\|_{\mathcal{F}} \leq \sum_{j=1}^q \left\|\widetilde{\boldsymbol{A}}_j(q)\right\|_{\mathcal{F}} + o_p(q)$ are obtained from the following two results: (1) $\left\|\widehat{\boldsymbol{\mathcal{A}}}_q(1) - \boldsymbol{\mathcal{A}}_q(1)\right\|_{\mathcal{F}} \leq \sum_{j=1}^q \left\|\widehat{\boldsymbol{A}}_j(q) - \boldsymbol{A}_j(q)\right\|_{\mathcal{F}} \leq C\sqrt{q}\left\|\widehat{\boldsymbol{A}}(q) - \boldsymbol{A}(q)\right\| = O_p\left(\frac{q^{3/2}}{T^{1/2}}\right) = o_p(1)$, where the last step follows from Lemma S3, and (2) $\sum_{j=1}^q \left\|\boldsymbol{A}_j^*(q) - \widetilde{\boldsymbol{A}}_j(q)\right\|_{\mathcal{F}} \leq q\sum_{j=1}^q \left\|\widehat{\boldsymbol{A}}_j(q) - \boldsymbol{A}_j(q)\right\|_{\mathcal{F}} = o_p(q)$. Using the BN decomposition of $\widehat{\boldsymbol{\mathcal{A}}}_q(L)$ and similar steps as those below (A.5), we have

$$\boldsymbol{G}_T\boldsymbol{Z}'\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q)\boldsymbol{u} = \sum_{t=1}^T \left(\widehat{\boldsymbol{\mathcal{A}}}_q(L)\boldsymbol{Z}_t'\boldsymbol{G}_T\right)' \widehat{\boldsymbol{S}}^{-1}(q)\left(\widehat{\boldsymbol{\mathcal{A}}}_q(L)\boldsymbol{u}_t\right) + o_p(1) \Rightarrow \int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{uu}^{-1}d\boldsymbol{B}_u(r) + \boldsymbol{\mathcal{B}}_{\epsilon\eta},$$

where $\widehat{\boldsymbol{S}}(q) = \boldsymbol{S}(q) + o_p(1)$ given in Lemma S3. Using the identity (A.6), it is not hard to deduce

$$\boldsymbol{G}_T\boldsymbol{Z}'\left(\boldsymbol{I}_T \otimes \widehat{\boldsymbol{\Omega}}_{uu}^{-1}\widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\right)\boldsymbol{v} \Rightarrow \int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{uu}^{-1}\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}d\boldsymbol{B}_v(r) + \underbrace{\begin{bmatrix} \mathrm{row}_1\left(\boldsymbol{\Delta}_{vv}\right)\mathrm{col}_1\left(\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{\Omega}_{vu}\boldsymbol{\Omega}_{uu}^{-1}\right)\boldsymbol{b}_1 \\ \vdots \\ \mathrm{row}_n\left(\boldsymbol{\Delta}_{vv}\right)\mathrm{col}_n\left(\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{\Omega}_{vu}\boldsymbol{\Omega}_{uu}^{-1}\right)\boldsymbol{b}_n \end{bmatrix}}_{\boldsymbol{\mathcal{B}}_{vu}}.$$

Combining the results above leads to:

$$\boldsymbol{G}_T\boldsymbol{Z}'\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q)\boldsymbol{u} - \boldsymbol{G}_T\boldsymbol{Z}'\left(\boldsymbol{I}_T \otimes \widehat{\boldsymbol{\Omega}}_{uu}^{-1}\widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\right)\boldsymbol{v} \Rightarrow \int_0^1 \boldsymbol{J}(r)\boldsymbol{\Omega}_{uu}^{-1}d\boldsymbol{B}_{u.v}(r) + \boldsymbol{\mathcal{B}}^+,$$

where $\boldsymbol{\mathcal{B}}^+ := \boldsymbol{\mathcal{B}}_{\epsilon\eta} - \boldsymbol{\mathcal{B}}_{vu}$. By construction, we have $\boldsymbol{G}_T\widehat{\boldsymbol{\mathcal{B}}}^+ \Rightarrow \boldsymbol{\mathcal{B}}^+$. Altogether this implies the limiting distribution in the theorem. ■

Proof of Theorem 4 We first introduce the appropriate scaling into the test statistic, that is

$$\mathcal{W} \equiv \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}_{FGLS}^+ - \boldsymbol{r}\right)'\left(\boldsymbol{R}\boldsymbol{G}_T^{-1}\boldsymbol{R}'\right)\left[\left(\boldsymbol{R}\boldsymbol{G}_T^{-1}\boldsymbol{R}'\right)\widehat{\boldsymbol{\Phi}}\left(\boldsymbol{R}\boldsymbol{G}_T^{-1}\boldsymbol{R}'\right)\right]^{-1}\left(\boldsymbol{R}\boldsymbol{G}_T^{-1}\boldsymbol{R}'\right)\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}_{FGLS}^+ - \boldsymbol{r}\right).$$

Since the matrices $\boldsymbol{G}_T^{-1}$ and $\boldsymbol{R}'\boldsymbol{R}$ commute and $\boldsymbol{R}\boldsymbol{R}' = \boldsymbol{I}_k$, we have $(\boldsymbol{R}\boldsymbol{G}_T^{-1}\boldsymbol{R}')(\boldsymbol{R}\widehat{\boldsymbol{\beta}}_{FGLS}^+ - \boldsymbol{r}) = \boldsymbol{R}\boldsymbol{G}_T^{-1}(\widehat{\boldsymbol{\beta}}_{FGLS}^+ - \boldsymbol{\beta})$ under the null hypothesis. Conditional on $\mathcal{F}_v$ (defined in Corollary 1), this quantity is asymptotically normally distributed by Theorem 3(c) with asymptotic covariance matrix $\boldsymbol{R}\boldsymbol{\mathcal{V}}_{FGLS}^+\boldsymbol{R}'$, where $\boldsymbol{\mathcal{V}}_{FGLS}^+$ is given in Corollary 1. The consistent estimation of all the quantities involved ensures that $(\boldsymbol{R}\boldsymbol{G}_T^{-1}\boldsymbol{R}')\,\widehat{\boldsymbol{\Phi}}\,(\boldsymbol{R}\boldsymbol{G}_T^{-1}\boldsymbol{R}')$ has the same limit. Therefore, the Wald statistics are conditionally chi-square distributed with k degrees of freedom. Since this distribution does not depend on $\mathcal{F}_v$, we conclude that the unconditional distribution of $\mathcal{W}$ is also χ_k^2. ■

Proof of Theorem 5 The results for K_{j,b_T}^{SOLS} and K_{j,b_T}^{SUR} follow from a straightforward multivariate extension of the proof of Proposition 6 in Wagner and Hong (2016). For K_{j,b_T}^{FGLS}, we first let $\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})$ and $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q_T,b_T)$ be the population counterparts of $\boldsymbol{\varphi}_{j,b_T}(\hat{\boldsymbol{u}}_{FGLS})$ and $\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q_T,b_T)$, respectively, where $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q_T,b_T)$ denotes the subblock matrix of $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q_T)$ formed by taking the elements with row and column indices belonging to the set $\{n(T-b_T)+1, n(T-b_T)+2,\ldots,nT\}$. By rearrangement, we have

$$\begin{aligned} K_{j,b_T}^{FGLS} &= b_T^{-2}\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})'\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q_T,b_T)\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u}) \\ &\quad + b_T^{-2}\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})'\left(\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q_T,b_T)-\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q_T,b_T)\right)\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u}) + R(q_T,j,b_T), \quad \text{(A.14)} \end{aligned}$$

where the remainder term is bounded as

$$\begin{aligned} |R(q_T,j,b_T)| &\le \left\|b_T^{-1}\left(\boldsymbol{\varphi}_{j,b_T}(\hat{\boldsymbol{u}}_{FGLS})-\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})\right)\right\|^2 \left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q_T,b_T)\right\| \\ &\quad + 2\left\|b_T^{-1}\left(\boldsymbol{\varphi}_{j,b_T}(\hat{\boldsymbol{u}}_{FGLS})-\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})\right)\right\| \left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q_T,b_T)\right\| \left\|b_T^{-1}\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})\right\|. \end{aligned}$$

Poincaré's separation theorem (e.g., p.347-348 of Abadir and Magnus (2005)) implies $\lambda_{min}(\boldsymbol{A}) \le \lambda_{min}(\boldsymbol{B}) \le \lambda_{max}(\boldsymbol{B}) \le \lambda_{max}(\boldsymbol{A})$ when $\boldsymbol{B}$ is a principal submatrix of $\boldsymbol{A}$. By this inequality and Theorem 2, we conclude that $\left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q_T,b_T)-\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q_T,b_T)\right\| \le \left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q_T)-\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q_T)\right\| = o_p(1)$ and $\left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q_T,b_T)\right\| = O_p(1)$. Moreover,

$$\left\|b_T^{-1}\left(\boldsymbol{\varphi}_{j,b_T}(\hat{\boldsymbol{u}}_{FGLS})-\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})\right)\right\| = \left(b_T^{-1}\sum_{t=j}^{j+b_T-1}\left\|b_T^{-1/2}\sum_{s=j}^{t}(\widehat{\boldsymbol{u}}_{s,FGLS}-\boldsymbol{u}_s)\right\|^2\right)^{1/2} = o_p(1),$$

where $\left\|b_T^{-1/2}\sum_{s=j}^{t}(\widehat{\boldsymbol{u}}_{s,FGLS}-\boldsymbol{u}_s)\right\| \le \left\|b_T^{-1/2}\sum_{s=j}^{t}\boldsymbol{Z}_s'\boldsymbol{G}_T\right\| \left\|\boldsymbol{G}_T^{-1}(\widehat{\boldsymbol{\beta}}_{FGLS}^{+}-\boldsymbol{\beta})\right\| = o_p(1)$ by Theorem 3 and the assumption $b_T/T \to 0$ as $T\to\infty$. Standard weak convergence arguments imply $\left\|b_T^{-1}\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})\right\| = O_p(1)$. Combining these results, we have $b_T^{-2}\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})'\Big(\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}}(q_T,b_T) - \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q_T,b_T)\Big)\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u}) = o_p(1)$ and $R(q_T,j,b_T) = o_p(1)$. By (A.14), it remains to consider $b_T^{-2}\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})'\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q_T,b_T)\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})$. Construct the $nT\times nb_T$ selection matrix $\boldsymbol{R}_{j,b_T}$ such that

$$\boldsymbol{R}_{j,b_T}\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u}) = \left[\boldsymbol{0}',\cdots,\boldsymbol{0}',\boldsymbol{u}_j',\sum_{s=j}^{j+1}\boldsymbol{u}_s',\ldots,\sum_{s=j}^{j+b_T-1}\boldsymbol{u}_s',\boldsymbol{0}',\cdots,\boldsymbol{0}'\right]'.$$

Then, by the MCBD (2.5), we have

$$
\begin{aligned}
b_T^{-2}\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q_T,b_T)\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u}) &= b_T^{-2}\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})'\boldsymbol{R}_{j,b_T}'\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q_T)\boldsymbol{R}_{j,b_T}\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u}) \\
&= b_T^{-2}\left(\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q_T)\boldsymbol{R}_{j,b_T}\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})\right)'\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q_T)\left(\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q_T)\boldsymbol{R}_{j,b_T}\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})\right).
\end{aligned}
$$

As argued in the proof of Theorem 1, by the assumption $q_T/b_T \to 0$ as $T \to \infty$, we can treat the premultiplication of $\boldsymbol{R}_{j,b_T}\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})$ by $\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q_T)$ as applying the filter $\boldsymbol{\mathcal{A}}_{q_T}(L)$ block-wise. Under the same condition, $\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q_T)$ implies a scaling $\boldsymbol{S}^{-1}(q_T)$. By the BN decomposition in Lemma 1, and similarly $\boldsymbol{\mathcal{C}}(L) = \boldsymbol{\mathcal{C}}(1) + (1-L)\widetilde{\boldsymbol{\mathcal{C}}}(L)$ with $\boldsymbol{\mathcal{C}}(L) = [\boldsymbol{\mathcal{A}}(L)]^{-1}$,

$$
\begin{aligned}
\boldsymbol{\mathcal{A}}_q(L)\sum_{s=j}^{t}\boldsymbol{u}_s &= \boldsymbol{\mathcal{A}}_{q_T}(1)\sum_{s=j}^{t}\boldsymbol{u}_s + \widetilde{\boldsymbol{\mathcal{A}}}_{q_T}(L)\boldsymbol{u}_t \\
&= \boldsymbol{\mathcal{A}}_{q_T}(1)\boldsymbol{\mathcal{C}}(1)\sum_{s=j}^{t}\boldsymbol{\eta}_s + \widetilde{\boldsymbol{\mathcal{A}}}_{q_T}(1)\widetilde{\boldsymbol{\mathcal{C}}}(L)\big(\boldsymbol{\eta}_t - \boldsymbol{\eta}_{j-1}\big) + \widetilde{\boldsymbol{\mathcal{A}}}_{q_T}(L)\boldsymbol{u}_t.
\end{aligned}
$$

For $t = j + [rb_T] - 1$, a FCLT for i.i.d. sequences gives

$$
b_T^{-1/2}\boldsymbol{S}^{-1/2}(q_T)\boldsymbol{\mathcal{A}}_{q_T}(L)\sum_{s=j}^{t}\boldsymbol{u}_s = \boldsymbol{S}^{-1/2}(q_T)\boldsymbol{\mathcal{A}}_{q_T}(1)\boldsymbol{\mathcal{C}}(1)\ b_T^{-1/2}\sum_{s=j}^{t}\boldsymbol{\eta}_s + O_p(b_T^{-1/2}) \Rightarrow \boldsymbol{W}(r).
$$

The partial sum process $\sum_{s=j}^{t}\boldsymbol{\eta}_s$ thus dominates the asymptotic distribution:

$$
\begin{aligned}
&b_T^{-2}\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u})'\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q_T,b_T)\boldsymbol{\varphi}_{j,b_T}(\boldsymbol{u}) \\
&\qquad = b_T^{-1}\sum_{t=j}^{j+b_T-1}\left\|b_T^{-1/2}\boldsymbol{S}^{-1/2}(q_T)\boldsymbol{\mathcal{A}}_{q_T}(L)\sum_{s=j}^{t}\boldsymbol{u}_s\right\|^2 + o_p(1) \Rightarrow \int_0^1 \|\boldsymbol{W}(r)\|^2\,dr.
\end{aligned}
$$

using the continuous mapping theorem. ■

Table 1: Empirical MSE for the coefficient $\beta_{i,4}$ of x_{it}^2 with $i = 1$ under error Setting A. The column labeled FGLS contains the numerical value of the MSE of feasible FM-GLS. Other MSEs are expressed relative to this benchmark. Values above 1 indicate a better performance of feasible FM-GLS.

	$n = 3$						$n = 5$					
ρ	SOLS	SUR	FGLS	infSOLS	infSUR	infGLS	SOLS	SUR	FGLS	infSOLS	infSUR	infGLS
$T = 100$												
0	0.999	1.048	3.56E-05	0.937	0.937	0.937	0.988	1.047	3.81E-05	0.894	0.894	0.894
0.3	1.170	1.077	5.50E-05	1.098	0.959	0.907	1.186	1.064	5.43E-05	1.090	0.899	0.851
0.6	2.166	1.474	7.09E-05	2.017	1.180	0.864	2.260	1.426	6.85E-05	2.025	1.046	0.785
0.8	5.247	2.607	7.51E-05	6.547	2.474	0.878	5.720	2.589	7.40E-05	6.591	1.984	0.676
$T = 200$												
0	1.012	1.042	4.09E-06	0.961	0.961	0.961	1.019	1.069	4.25E-06	0.939	0.939	0.939
0.3	1.146	1.043	6.64E-06	1.101	0.960	0.942	1.216	1.069	6.62E-06	1.124	0.937	0.907
0.6	2.045	1.295	1.01E-05	1.920	1.101	0.906	2.222	1.345	9.37E-06	2.032	1.056	0.851
0.8	5.206	2.434	1.29E-05	5.502	1.971	0.916	6.197	2.701	1.11E-05	6.106	1.778	0.813
$T = 500$												
0	1.013	1.028	2.41E-07	0.988	0.988	0.988	1.016	1.039	2.55E-07	0.973	0.973	0.973
0.3	1.195	1.054	4.03E-07	1.162	0.998	0.975	1.224	1.052	4.02E-07	1.176	0.978	0.961
0.6	1.877	1.154	7.45E-07	1.789	1.054	0.952	2.142	1.217	6.49E-07	1.993	1.023	0.914
0.8	4.158	1.771	1.26E-06	4.011	1.456	0.935	5.341	2.031	1.03E-06	5.038	1.394	0.880

Table 2: Empirical MSE for the coefficient $\beta_{i,4}$ of x_{it}^2 with $i = 1$. under error Setting B. The column labeled FGLS contains the numerical value of the MSE of feasible FM-GLS. Other MSEs are expressed relative to this benchmark. Values above 1 indicate a better performance of feasible FM-GLS.

		$n = 3$						$n = 5$					
$(\underline{\lambda}, \bar{\lambda})$	T	SOLS	SUR	FGLS	infSOLS	infSUR	infGLS	SOLS	SUR	FGLS	infSOLS	infSUR	infGLS
Panel A: Low endogeneity $\theta = 0.3$													
$(0.1, 0.5)$	100	1.290	1.270	9.46E-05	1.270	1.131	0.865	1.256	1.275	9.51E-05	1.215	1.052	0.822
	200	1.216	1.163	1.27E-05	1.191	1.077	0.916	1.220	1.164	1.22E-05	1.201	1.038	0.896
	500	1.137	1.088	8.58E-07	1.133	1.033	0.961	1.159	1.106	7.65E-07	1.151	0.997	0.936
$(0.5, 0.8)$	100	1.790	1.825	3.99E-05	1.792	1.522	0.638	1.812	1.896	3.81E-05	1.802	1.431	0.594
	200	1.677	1.681	4.91E-06	1.597	1.408	0.763	1.682	1.645	4.73E-06	1.581	1.271	0.733
	500	1.467	1.404	3.34E-07	1.392	1.248	0.897	1.522	1.415	3.01E-07	1.432	1.185	0.864
$(0.8, 0.95)$	100	0.123	0.136	1.48E-04	0.212	0.145	0.022	0.001	0.001	2.67E-02	0.002	0.001	0.000
	200	2.257	2.322	8.59E-07	2.537	1.986	0.517	1.622	1.577	1.04E-06	2.218	1.343	0.366
	500	2.098	2.050	5.47E-08	2.021	1.686	0.703	2.031	2.014	4.90E-08	1.964	1.394	0.632
Panel B: High endogeneity $\theta = 0.5$													
$(0.1, 0.5)$	100	1.327	1.269	7.48E-05	1.579	1.230	0.899	1.335	1.299	7.26E-05	1.825	1.211	0.817
	200	1.232	1.161	9.91E-06	1.394	1.131	0.949	1.269	1.225	9.26E-06	1.653	1.166	0.886
	500	1.179	1.091	6.66E-07	1.297	1.079	0.980	1.162	1.093	5.99E-07	1.464	1.071	0.942
$(0.5, 0.8)$	100	1.880	1.892	3.08E-05	2.732	1.861	0.661	1.850	1.833	3.05E-05	3.170	1.718	0.575
	200	1.793	1.727	3.72E-06	2.132	1.610	0.798	1.713	1.639	3.76E-06	2.143	1.381	0.697
	500	1.480	1.370	2.62E-07	1.678	1.299	0.941	1.577	1.461	2.44E-07	1.739	1.256	0.863
$(0.8, 0.95)$	100	0.030	0.031	5.14E-04	0.172	0.077	0.005	0.000	0.000	3.27E+03	0.000	0.000	0.000
	200	2.381	2.514	6.85E-07	5.002	2.825	0.541	1.880	1.771	7.83E-07	5.026	2.030	0.389
	500	2.186	2.090	4.23E-08	2.756	1.850	0.712	2.184	2.044	4.03E-08	2.786	1.560	0.619

Table 3: Empirical size (%) of the single-equation Wald tests for $H_0 : \beta_{1,4} = -0.3$ and the joint Wald tests for $H_0 : \beta_{1,4} = \beta_{2,4} = \cdots = \beta_{n,4} = -0.3$, where $\beta_{i,4}$ denote the coefficients in front of x_{it}^2. The columns labeled by Wald-SOLS and Wald-SUR display the results of the Wald-type test statistics given in Proposition 2 by Wagner et al. (2020). The Wald-FGLS test is constructed as in Theorem 4.

	$n = 3$			$n = 5$		
ρ	Wald-SOLS	Wald-SUR	Wald-FGLS	Wald-SOLS	Wald-SUR	Wald-FGLS
Panel A: Single-equation test						
$T = 100$						
0	11.75	13.44	9.89	14.30	17.84	13.44
0.3	13.25	15.09	9.92	15.32	19.59	13.55
0.6	16.13	19.03	7.54	18.80	27.65	11.86
0.8	20.56	26.60	4.70	26.72	43.11	10.13
$T = 200$						
0	9.02	10.00	7.48	9.90	12.00	8.47
0.3	9.96	10.93	7.24	11.32	13.62	8.81
0.6	12.68	14.10	5.93	14.62	19.40	7.71
0.8	15.72	19.50	2.95	19.58	30.95	4.90
$T = 500$						
0	7.02	7.44	5.89	7.32	8.47	6.19
0.3	8.07	8.41	5.76	8.54	9.50	6.04
0.6	9.41	9.68	4.78	10.34	12.54	5.94
0.8	10.92	12.48	3.09	13.47	18.96	3.73
Panel B: Joint test						
$T = 100$						
0	17.85	21.68	14.60	29.13	39.40	26.65
0.3	20.67	24.62	14.44	32.60	44.91	28.14
0.6	27.13	33.59	11.40	45.65	65.44	27.72
0.8	36.99	48.49	8.30	63.29	86.38	26.57
$T = 200$						
0	11.66	13.74	9.00	17.49	22.90	13.26
0.3	14.11	16.25	8.77	21.35	28.23	14.55
0.6	19.08	22.71	7.10	30.58	44.07	12.88
0.8	26.21	33.92	4.19	44.82	69.44	10.60
$T = 500$						
0	8.70	9.64	6.60	10.83	13.38	7.91
0.3	9.84	10.74	6.23	13.29	16.19	8.19
0.6	12.79	14.09	5.13	18.69	25.15	7.67
0.8	16.02	19.56	3.35	26.86	41.72	5.28

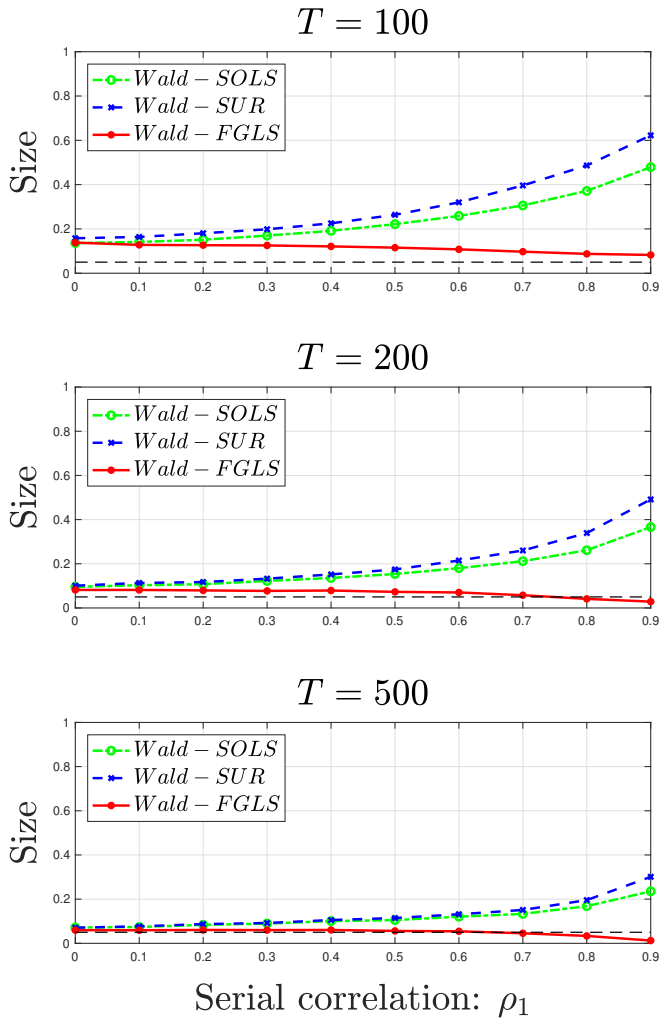

Figure 1: Empirical size of the joint Wald tests $H_0 : \beta_{1,4} = \beta_{2,4} = \cdots = \beta_{n,4} = -0.3$, where $\beta_{i,4}$ denote the coefficients in front of x_{it}^2. The Wald-SOLS test (green) and Wald-SUR test (blue) are based on Proposition 2 by Wagner et al. (2020). The Wald-FGLS test (red) is found in Theorem 4. We vary the serial correlation parameter ρ_1 from 0 to 0.9 while keeping $\rho_2 = \rho_3 = \rho_4 = 0.8$ fixed. The cross-sectional dimension is $n = 3$.

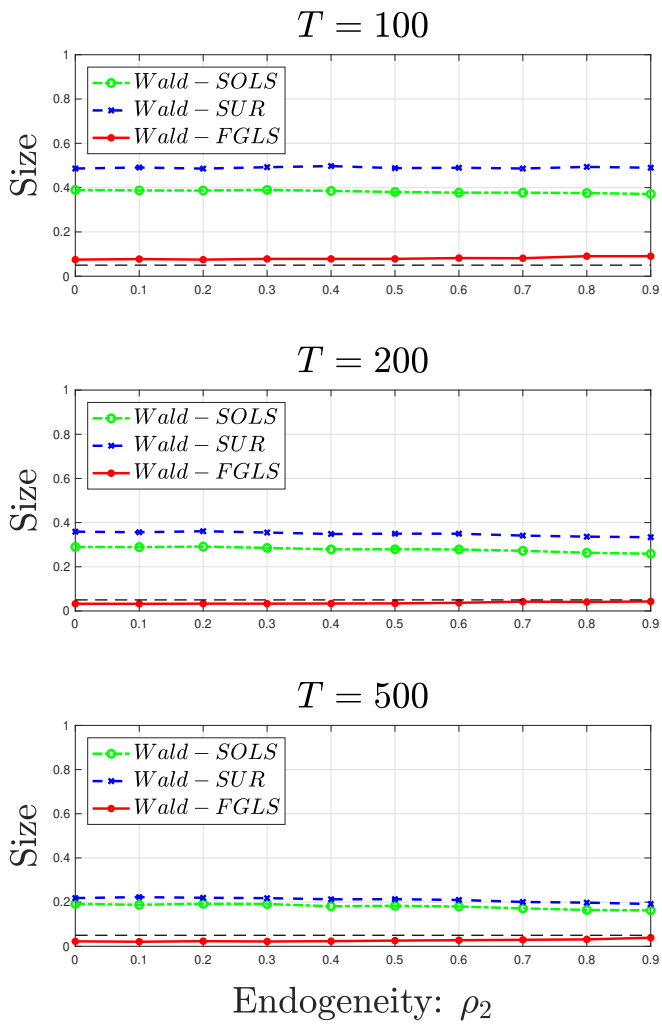

Figure 2: Empirical size of the joint Wald tests $H_0 : \beta_{1,4} = \beta_{2,4} = \cdots = \beta_{n,4} = -0.3$, where $\beta_{i,4}$ denote the coefficients in front of x_{it}^2. We vary the endogeneity parameter ρ_2 from 0 to 0.9 while keeping $\rho_1 = \rho_3 = \rho_4 = 0.8$ fixed. The cross-sectional dimension is $n = 3$. See Figure 1 for further information.

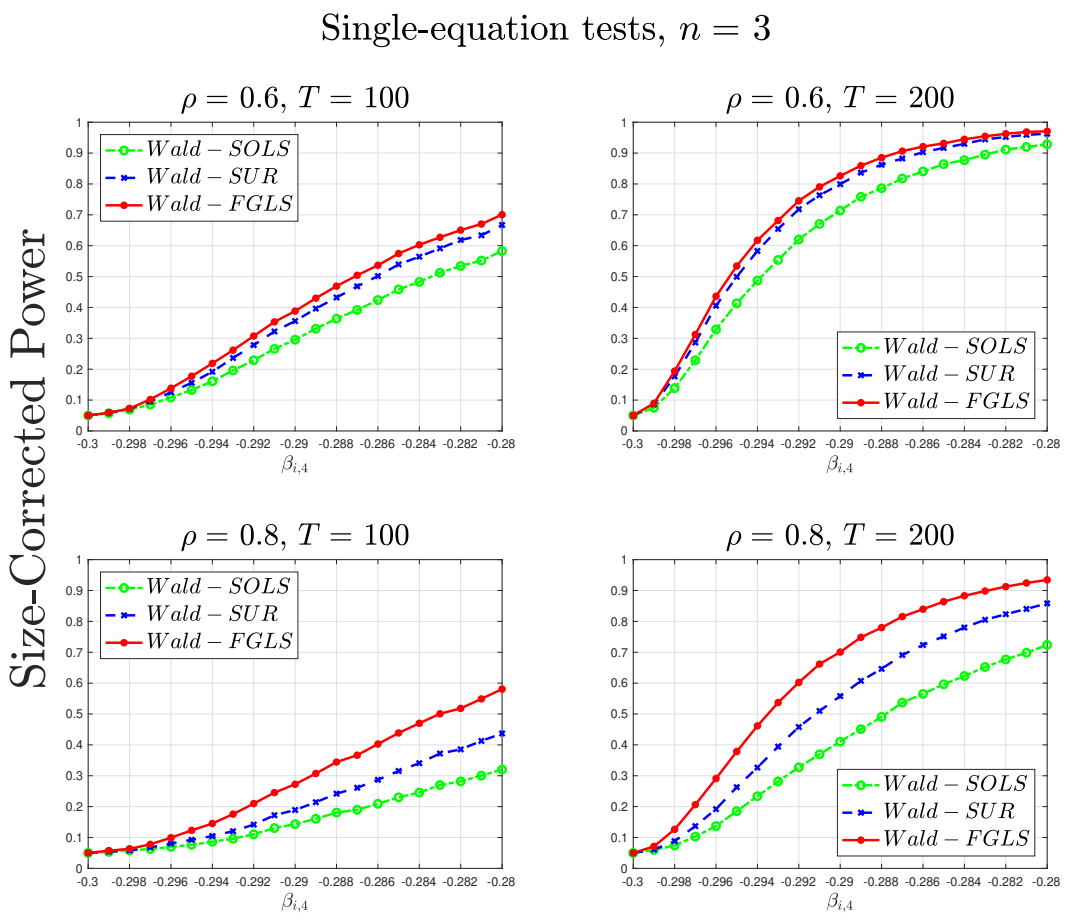

Figure 3: Empirical size-corrected power of the single-equation Wald tests $H_0 : \beta_{1,4} = -0.3$ where $\beta_{1,4}$ is the coefficient in front of x_{1t}^2. We consider $n = 3$, $T \in \{100, 200\}$, and $\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho$ with $\rho \in \{0.6, 0.8\}$. See Figure 1 for further information.

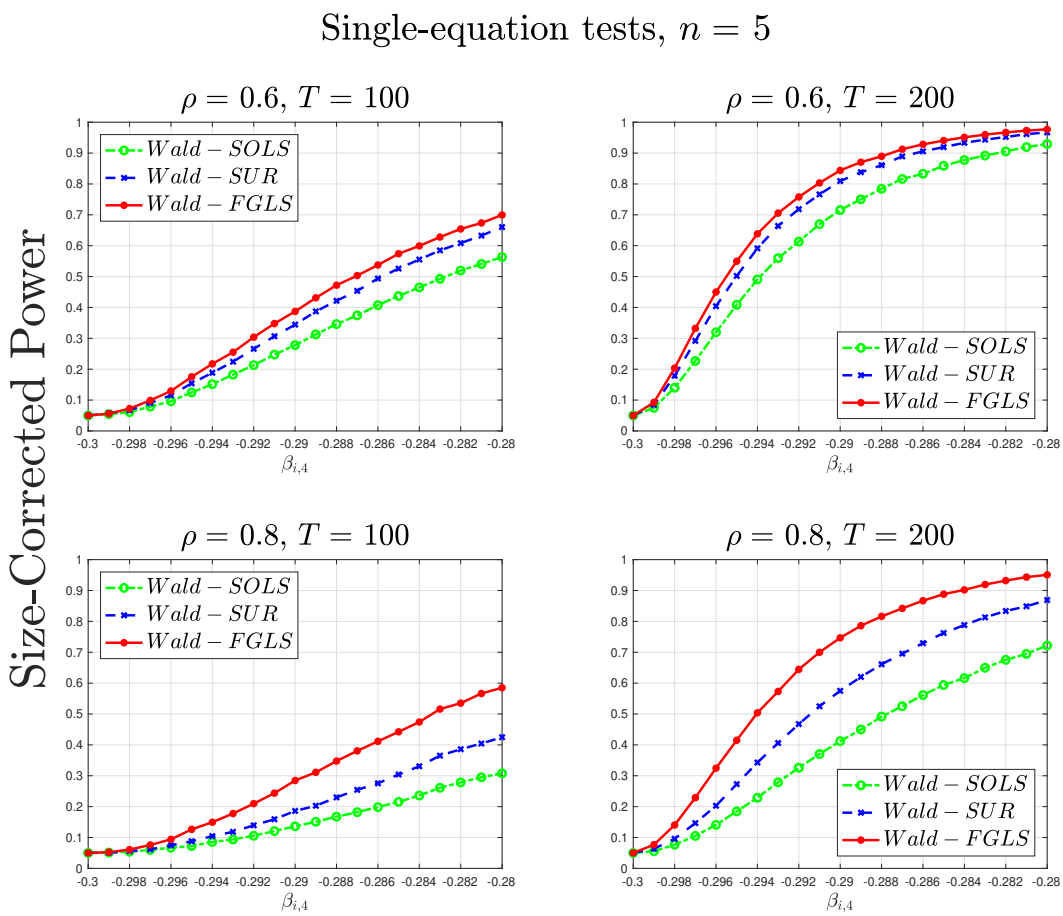

Figure 4: Empirical size-corrected power of the single-equation Wald tests $H_0 : \beta_{1,4} = -0.3$ where $\beta_{1,4}$ is the coefficient in front of x_{1t}^2. We consider $n = 5$, $T \in \{100, 200\}$, and $\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho$ with $\rho \in \{0.6, 0.8\}$. See Figure 1 for further information.

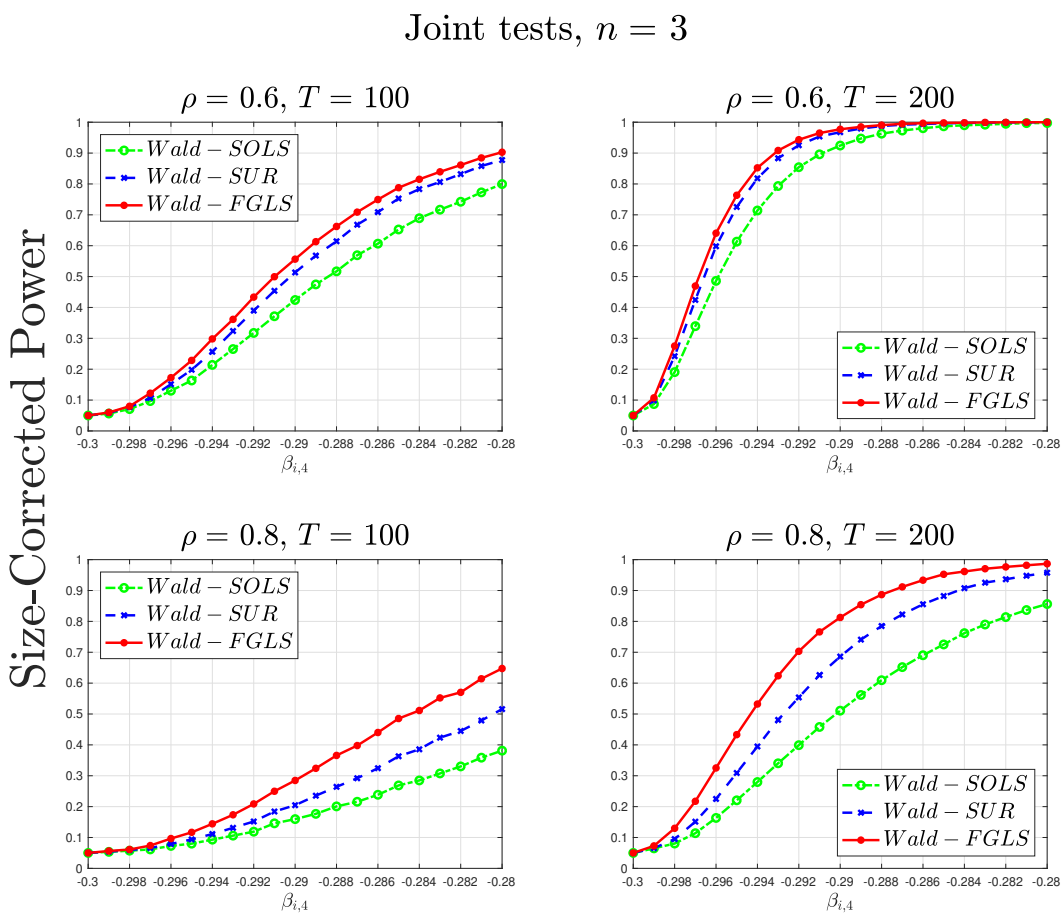

Figure 5: Empirical size-corrected power of the joint Wald tests $H_0 : \beta_{1,4} = \beta_{2,4} = \cdots = \beta_{n,4} = -0.3$ where $\beta_{i,4}$ are the coefficients in front of x_{it}^2. We consider $n = 3$, $T \in \{100, 200\}$, and $\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho$ with $\rho \in \{0.6, 0.8\}$. See Figure 1 for further information.

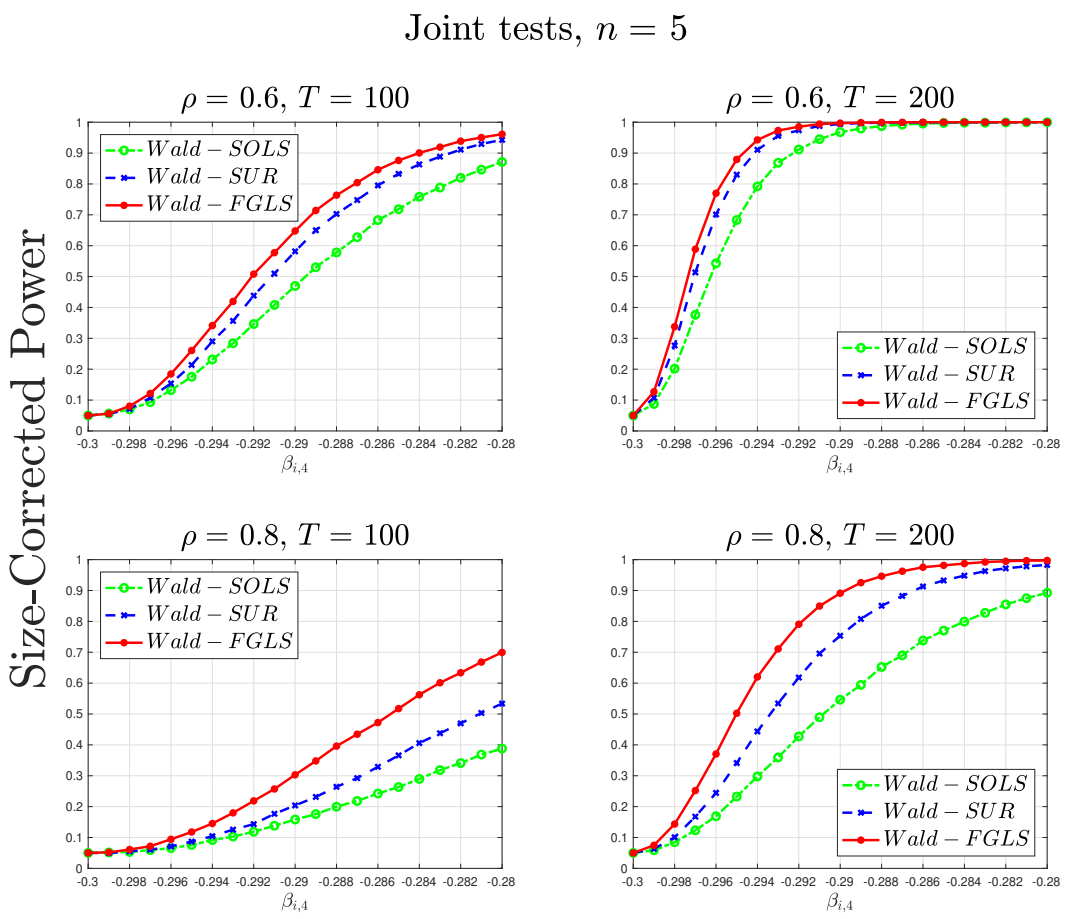

Figure 6: Empirical size-corrected power of the joint Wald tests $H_0 : \beta_{1,4} = \beta_{2,4} = \cdots = \beta_{n,4} = -0.3$ where $\beta_{i,4}$ are the coefficients in front of x_{it}^2. We consider $n = 5$, $T \in \{100, 200\}$, and $\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho$ with $\rho \in \{0.6, 0.8\}$. See Figure 1 for further information.

Table 4: Empirical size (%) and power (%) of Bonferroni-type (multivariate) subsampling KPSS tests. The integer J_1 in Power DGP1 indicates the number of unit roots contained in errors $\{\boldsymbol{u}_t\}$, J_2 is related to the number of equations that exclude cubic power terms x_{it}^3, J_3 specifies the number of spurious relations.

			$n=3$			$n=5$		
		T	K_{max}^{SOLS}	K_{max}^{SUR}	K_{max}^{FGLS}	K_{max}^{SOLS}	K_{max}^{SUR}	K_{max}^{FGLS}
Panel A: Size								
$(\underline{\lambda}, \bar{\lambda})$, serial correlation	$(0.1, 0.5)$	100	0.37	0.34	0.47	0.33	0.25	0.26
		200	0.42	0.37	0.53	0.42	0.28	0.18
		500	0.71	0.60	1.24	0.64	0.58	0.80
	$(0.5, 0.8)$	100	8.10	7.28	1.86	22.35	19.65	1.29
		200	3.54	3.22	2.91	8.70	7.57	1.09
		500	1.92	1.69	4.54	4.02	3.40	4.78
	$(0.8, 0.95)$	100	50.63	48.67	16.17	90.99	89.69	18.41
		200	40.48	38.56	14.26	83.71	81.78	12.16
		500	21.87	20.36	16.25	60.09	56.87	15.49
Panel B: Power DGP1								
J_1, # {unit roots}	1	100	44.03	39.92	21.14	57.65	52.05	14.00
		200	57.97	50.90	33.85	74.09	66.64	27.24
		500	67.23	57.37	56.63	88.41	80.23	49.93
	2	100	66.64	63.83	38.07	84.77	80.36	29.12
		200	77.77	73.00	52.88	94.38	91.16	49.87
		500	84.19	78.83	78.40	98.57	96.89	72.02
	n	100	78.89	77.41	54.26	99.39	99.24	64.89
		200	88.40	86.53	70.07	99.97	99.95	88.75
		500	91.28	90.19	89.55	100.00	100.00	96.53
Panel C: Power DGP2								
J_2, # {misspecified equations}	1	100	10.92	10.51	3.47	18.67	18.40	1.91
		200	26.41	25.58	7.72	41.90	41.41	4.95
		500	55.72	55.28	24.10	77.76	77.22	18.36
	2	100	17.83	16.97	6.15	30.77	29.77	3.70
		200	40.29	39.04	15.57	62.01	61.15	10.51
		500	69.64	68.50	38.48	91.66	91.27	30.87
	n	100	23.16	22.53	9.88	53.66	51.21	10.45
		200	47.25	45.60	22.83	83.78	82.73	29.05
		500	75.60	73.98	56.79	98.23	98.15	69.35
Panel D: Power DGP3								
J_3, # {spurious relations}	1	100	77.03	77.06	28.34	94.26	94.12	21.04
		200	89.55	89.14	35.64	98.97	98.84	31.79
		500	97.26	97.05	55.61	99.91	99.91	48.48
	2	100	87.40	86.75	52.85	98.95	98.81	43.05
		200	94.95	94.33	57.95	99.89	99.89	58.15
		500	98.69	98.49	77.59	100.00	100.00	69.59
	n	100	91.55	90.60	71.95	99.97	99.92	87.87
		200	96.71	96.33	74.85	99.99	100.00	93.86
		500	98.04	97.62	91.66	100.00	100.00	98.02

Note: To decrease the computational burden we reduced the number of Monte Carlo replications to 10^4.

Supplemental Appendix to:

Fully Modified Estimation in Cointegrating Polynomial Regressions: Extensions and Monte Carlo Comparison

Yicong Lin[1,2] and Hanno Reuvers[3]

[1]Department of Econometrics and Data Science, Vrije Universiteit Amsterdam

[2]Tinbergen Institute

[3]Department of Econometrics, Erasmus University Rotterdam

December, 2022

1 Additional Proofs

Lemma S1

If $\{\boldsymbol{u}_t\}$ satisfies Assumption 1, then for any $m \geq 1$, there exists a constant $C > 0$ such that

$$\sum_{j=1}^{m} \|\boldsymbol{A}_j(m) - \boldsymbol{A}_j\|_{\mathcal{F}}^2 \leq C \sum_{j=m+1}^{\infty} \|\boldsymbol{A}_j\|_{\mathcal{F}}^2 . \tag{S.1}$$

Proof In view of page 257 of Hannan and Deistler (2012), the summability condition of Assumption 1 implies that the spectral density matrix is bounded and bounded away from zero. The boundedness condition in Cheng and Pourahmadi (1993) is thus satisfied and (S.1) follows from their Theorem 2.2. ■

Lemma S2 (implications of the first moment bound theorem)

Let Assumption 1 hold, and define

$$\boldsymbol{\eta}_{t+1,\ell} = \boldsymbol{u}_{t+1} - \boldsymbol{A}(\ell)\boldsymbol{u}_t(\ell), \qquad \text{where} \quad \boldsymbol{u}_t(\ell) = \left[\boldsymbol{u}_t', \boldsymbol{u}_{t-1}', \cdots, \boldsymbol{u}_{t-\ell+1}'\right]' . \tag{S.2}$$

The following three inequalities are true:

(a) $\mathbb{E}\left\| \frac{1}{T-q} \sum_{t=q}^{T-1} \boldsymbol{u}_t(q)\boldsymbol{u}_t(q)' - \mathbb{E}\left(\boldsymbol{u}_t(q)\boldsymbol{u}_t(q)'\right) \right\|^2 \leq C \frac{q^2}{T-q}$;

(b) $\mathbb{E}\left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \left(\boldsymbol{\eta}_{t+1,\ell} - \boldsymbol{\eta}_{t+1}\right)\boldsymbol{u}_t(\ell)' \right\|^r \leq C \left(\frac{\ell}{T-\ell}\right)^{r/2} \left(\sum_{j=\ell+1}^{\infty} \|\boldsymbol{A}_j\|_{\mathcal{F}}^2\right)^{r/2}$, *for some $r \geq 2$ and any $1 \leq \ell \leq q$;*

(c) $\mathbb{E}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\left(\boldsymbol{\eta}_{t+1,\ell}\boldsymbol{\eta}'_{t+1,\ell}\right)-\mathbb{E}\left(\boldsymbol{\eta}_{t+1,\ell}\boldsymbol{\eta}'_{t+1,\ell}\right)\right\|^r \leq C(T-\ell)^{-r/2}$, *for some* $r \geq 2$ *and any* $1 \leq \ell \leq q$.

Proof **(a)** Since $\|\cdot\|^2 \leq \|\cdot\|^2_{\mathcal{F}}$, we obtain

$$\mathbb{E}\left\|\frac{1}{T-q}\sum_{t=q}^{T-1}\boldsymbol{u}_t(q)\boldsymbol{u}_t(q)' - \mathbb{E}\left(\boldsymbol{u}_t(q)\boldsymbol{u}_t(q)'\right)\right\|^2 \leq \frac{1}{(T-q)^2}\sum_{k,\ell=1}^{q}\sum_{i,j=1}^{n}\mathbb{E}\left[\sum_{t=q}^{T-1}\left[u_{i,t-k+1}u_{j,t-\ell+1} - \mathbb{E}(u_{i,t-k+1}u_{j,t-\ell+1})\right]\right]^2, \quad \text{(S.3)}$$

where $u_{i,t}$ denotes the ith element $\boldsymbol{u}_t$. As remarked in the main text, the lag polynomial $\boldsymbol{\mathcal{A}}(L)$ is invertible by Assumption 1. Recall $\boldsymbol{\mathcal{C}}(L) = [\boldsymbol{\mathcal{A}}(L)]^{-1} = \sum_{j=0}^{\infty}\boldsymbol{C}_jL^j$ with $\boldsymbol{C}_0 = \boldsymbol{I}_n$, and $\sum_{j=0}^{\infty} j\,\|\boldsymbol{C}_j\|_{\mathcal{F}} < \infty$. We observe that $u_{i,t} = \sum_{j=0}^{\infty}\mathrm{row}_i\left(\boldsymbol{C}_j\right)\boldsymbol{\eta}_{t-j}$. By Proposition 10.2(b) of Hamilton (1994), absolute summability of the coefficient matrices $\{\boldsymbol{C}_j\}_{j=0}^{\infty}$ implies $\sum_{s=0}^{\infty}|\gamma_{u,k}(s)| < \infty$ where $\gamma_{u,k}(s) = \mathbb{E}(u_{k,t}u_{k,t-s})$. The conditions for the First Moment Bound Theorem (FMBT) in Findley and Wei (1993) are thus satisfied. Choosing $q(t,s) = 1$ if $t = s \geq q$ (the banding parameter) and $q(t,s) = 0$ otherwise,

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=q}^{T-1}\left[u_{i,t-k+1}u_{j,t-\ell+1} - \mathbb{E}(u_{i,t-k+1}u_{j,t-\ell+1})\right]\right]^2 &= \mathbb{E}\left[\sum_{t,s=1}^{T-1}q(t,s)\left[u_{i,t-k+1}u_{j,s-\ell+1} - \mathbb{E}(u_{i,t-k+1}u_{j,s-\ell+1})\right]\right]^2 \\
&\leq C\sum_{t,s,l,w=1}^{T-1}q(t,s)q(l,w)\gamma_{u,i}(t-l)\gamma_{u,j}(s-w) = C\sum_{t,l=q}^{T-1}\gamma_{u,i}(t-l)\gamma_{u,j}(t-l) \\
&\leq C\gamma_{u,j}(0)\sum_{t,l=q}^{T-1}|\gamma_{u,i}(t-l)| \leq C(T-q)\left[\gamma_{u,j}(0)\sum_{t=-\infty}^{\infty}|\gamma_{u,i}(t)|\right] \leq C(T-q),
\end{aligned}$$

by the FMBT. This bound holds for general k, ℓ, i and j, and **(a)** thereby follows from (S.3).

(b) For $1 \leq \ell \leq q$ and $r \geq 2$, we have

$$\mathbb{E}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\left(\boldsymbol{\eta}_{t+1,\ell} - \boldsymbol{\eta}_{t+1}\right)\boldsymbol{u}_t(\ell)'\right\|^r \leq \ell^{r/2-1}(T-\ell)^{-r}\sum_{s=0}^{\ell-1}\mathbb{E}\left\|\sum_{t=\ell}^{T-1}\left(\boldsymbol{\eta}_{t+1,\ell} - \boldsymbol{\eta}_{t+1}\right)\boldsymbol{u}'_{t-s}\right\|^r_{\mathcal{F}}, \quad \text{(S.4)}$$

by $\|\cdot\|^r \leq (\|\cdot\|^2_{\mathcal{F}})^{r/2}$ and the c_r-inequality. By assumption, $\boldsymbol{\eta}_{t+1}$ is uncorrelated with $\left[\boldsymbol{u}'_{t-\ell+1},\ldots,\boldsymbol{u}'_t\right]'$ implying that $\mathbb{E}\left[\sum_{t=\ell}^{T-1}(\boldsymbol{\eta}_{t+1,\ell} - \boldsymbol{\eta}_{t+1})\boldsymbol{u}'_{t-s}\right] = \mathbf{O}$. The FMBT can thus be applied directly without having to express the quadratic form in deviations from the mean. However, some rewriting is needed to obtain expressions in scalar random sequences. To this end, use $A_{j,kl}$ and $A_{j,kl}(\ell)$ to denote the $(k,l)^{th}$ element of $\boldsymbol{A}_j$ and $\boldsymbol{A}_j(\ell)$, respectively. Setting $\boldsymbol{A}_j(\ell) = \mathbf{O}$ for $j > \ell$, we have $\boldsymbol{\eta}_{t+1,\ell} - \boldsymbol{\eta}_{t+1} = \sum_{j=1}^{\infty}\left[\boldsymbol{A}_j - \boldsymbol{A}_j(\ell)\right]\boldsymbol{u}_{t+1-j}$,

and hence

$$\mathbb{E}\left\|\sum_{t=\ell}^{T-1}\left(\boldsymbol{\eta}_{t+1,\ell}-\boldsymbol{\eta}_{t+1}\right)\boldsymbol{u}_{t-s}'\right\|_{\mathcal{F}}^{r}=\mathbb{E}\left\{\sum_{k,m=1}^{n}\left[\sum_{t=\ell}^{T-1}u_{m,t-s}\sum_{j=1}^{\infty}\sum_{l=1}^{n}\left(A_{j,kl}-A_{j,kl}(\ell)\right)u_{l,t+1-j}\right]^{2}\right\}^{r/2}$$
$$\leq n^{r-2}\sum_{k,m=1}^{n}\mathbb{E}\left|\sum_{t=\ell}^{T-1}u_{m,t-s}\sum_{j=1}^{\infty}\sum_{l=1}^{n}\left(A_{j,kl}-A_{j,kl}(\ell)\right)u_{l,t+1-j}\right|^{r}=n^{r-2}\sum_{k,m=1}^{n}\mathbb{E}\left|\sum_{t=\ell}^{T-1}u_{m,t-s}u_{t}^{*}\right|^{r} \tag{S.5}$$

with $u_t^*=\sum_{j=1}^{\infty}\sum_{l=1}^{n}\left(A_{j,kl}-A_{j,kl}(\ell)\right)u_{l,t+1-j}$, where we suppress the dependence on the index k (also below) without confusion. To apply the FMBT, we define the autocovariances $\gamma_{u^*}(t-h)=\mathbb{E}\big(u_t^*u_h^*\big)$, the difference in lag polynomial coefficients $\boldsymbol{a}_l(\ell)=[A_{1,kl}-A_{1,kl}(\ell),A_{2,kl}-A_{2,kl}(\ell),\ldots]'$ and $\boldsymbol{\Sigma}_{u_l,\infty}=\big[\gamma_{u,l}(i-j),1\leq i,j<\infty\big]$. By the Cauchy-Schwartz inequality, the c_r-inequality, and boundedness of the maximum eigenvalue of $\boldsymbol{\Sigma}_{u_l,\infty}$, we obtain

$$\gamma_{u^*}(t-h)\leq\gamma_{u^*}(0)=\mathbb{E}\left[\sum_{l=1}^{n}\sum_{j=1}^{\infty}\left(A_{j,kl}-A_{j,kl}(\ell)\right)u_{l,t+1-j}\right]^{2}\leq n\sum_{l=1}^{n}\left[\boldsymbol{a}_{kl}(\ell)'\boldsymbol{\Sigma}_{u_l,\infty}\boldsymbol{a}_{kl}(\ell)\right]$$
$$\leq Cn\sum_{l=1}^{n}\|\boldsymbol{a}_{kl}(\ell)\|^{2}\leq Cn\sum_{j=1}^{\infty}\|\boldsymbol{A}_j-\boldsymbol{A}_j(\ell)\|_{\mathcal{F}}^{2}\leq Cn\sum_{j=\ell+1}^{\infty}\|\boldsymbol{A}_j\|_{\mathcal{F}}^{2}. \tag{S.6}$$

Applying the FMBT, we have

$$\left(\mathbb{E}\left|\sum_{t=\ell}^{T-1}u_{m,t-s}u_t^*\right|^{r}\right)^{1/r}\leq C\left[\sum_{t,h=\ell}^{T-1}\gamma_{u,m}(t-h)\gamma_{u^*}(t-h)\right]^{1/2}$$
$$\leq C\left[\gamma_{u^*}(0)(T-\ell)\sum_{t=-\infty}^{\infty}|\gamma_{u,m}(t)|\right]^{1/2}\leq C(T-\ell)^{1/2}\left(\sum_{j=\ell+1}^{\infty}\|\boldsymbol{A}_j\|_{\mathcal{F}}^{2}\right)^{1/2}, \tag{S.7}$$

using (S.6) and the absolute summability of $\{\gamma_{u,m}(t)\}$. Combining (S.4), (S.5) and (S.7) leads to the desired inequality.

(c) The equality $\boldsymbol{\eta}_{t+1,\ell}=\Big(\boldsymbol{I}_n-\sum_{j=1}^{\ell}\boldsymbol{A}_j(\ell)L^j\Big)\boldsymbol{\mathcal{C}}(L)\boldsymbol{\eta}_{t+1}$ shows that $\boldsymbol{\eta}_{t+1,\ell}$ has a linear process representation in terms of $\boldsymbol{\eta}_t$. Theorem 6.6.12 of Hannan and Deistler (2012) implies that $\sup_{1\leq\ell<\infty}\sum_{j=0}^{\ell}\|\boldsymbol{A}_j(\ell)\|_{\mathcal{F}}<\infty$. By Propositions 10.2(b) and 10.3 of Hamilton (1994), both the coefficient matrices associated with $\Big(\boldsymbol{I}_n-\sum_{j=1}^{\ell}\boldsymbol{A}_j(\ell)L^j\Big)\boldsymbol{\mathcal{C}}(L)$ and the autocovariances $\{\mathbb{E}\left(\eta_{k,t+1,\ell}\eta_{k,t+1-s,\ell}\right)\}_{s=0}^{\infty}$ are absolutely summable, where $\eta_{k,t+1,\ell}$ is the kth entry of $\boldsymbol{\eta}_{t+1,\ell}$. The proof is completed using the

c_r-inequality and the FMBT, that is, for $r \geq 2$,

$$\begin{aligned}
&\mathbb{E}\left\| \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \left(\boldsymbol{\eta}_{t+1,\ell} \boldsymbol{\eta}_{t+1,\ell}' \right) - \mathbb{E}\left(\boldsymbol{\eta}_{t+1,\ell} \boldsymbol{\eta}_{t+1,\ell}' \right) \right\|^r \\
&\leq n^{r-2} \frac{1}{(T-\ell)^r} \sum_{k,m=1}^{n} \mathbb{E} \left| \sum_{t=\ell}^{T-1} \left[\eta_{k,t+1,\ell} \eta_{m,t+1,\ell} - \mathbb{E}\left(\eta_{k,t+1,\ell} \eta_{m,t+1,\ell} \right) \right] \right|^r \\
&\leq C n^{r-2} \frac{1}{(T-\ell)^r} \sum_{k,m=1}^{n} \left[\sum_{t,l=\ell}^{T-1} \mathbb{E}\left(\eta_{k,t+1,\ell} \eta_{k,l+1,\ell} \right) \mathbb{E}\left(\eta_{m,t+1,\ell} \eta_{m,l+1,\ell} \right) \right]^{r/2} \leq C \frac{1}{(T-\ell)^{r/2}}.
\end{aligned}$$

■

Lemma S3

If Assumptions 1-3 hold, then

$$\max_{1 \leq \ell \leq q} \left\| \widehat{\boldsymbol{A}}(\ell) - \boldsymbol{A}(\ell) \right\| = O_p\left(q/\sqrt{T} \right), \qquad \max_{1 \leq \ell \leq q} \left\| \widehat{\boldsymbol{S}}(\ell) - \boldsymbol{S}(\ell) \right\| = O_p\left(q/\sqrt{T} \right). \tag{S.8}$$

Proof Recall the definition of $\boldsymbol{\eta}_{t+1,\ell}$ and $\boldsymbol{u}_t(\ell)$ in (S.2). Similarly, define

$$\widehat{\boldsymbol{\eta}}_{t+1,\ell} = \widehat{\boldsymbol{u}}_{t+1} - \boldsymbol{A}(\ell) \widehat{\boldsymbol{u}}_t(\ell), \quad \widetilde{\boldsymbol{\eta}}_{t+1,\ell} = \widehat{\boldsymbol{u}}_{t+1} - \widehat{\boldsymbol{A}}(\ell) \widehat{\boldsymbol{u}}_t(\ell), \quad \widehat{\boldsymbol{u}}_t(\ell) = \left[\widehat{\boldsymbol{u}}_t', \widehat{\boldsymbol{u}}_{t-1}', \cdots, \widehat{\boldsymbol{u}}_{t-\ell+1}' \right]'. \tag{S.9}$$

We first prove $\max_{1 \leq \ell \leq q} \left\| \widehat{\boldsymbol{A}}(\ell) - \boldsymbol{A}(\ell) \right\| = O_p\left(q/\sqrt{T} \right)$. Since $\left(\widehat{\boldsymbol{A}}(\ell) - \boldsymbol{A}(\ell) \right) \widehat{\boldsymbol{u}}_t(\ell) = \widehat{\boldsymbol{\eta}}_{t+1,\ell} - \widetilde{\boldsymbol{\eta}}_{t+1,\ell}$ and $\frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widetilde{\boldsymbol{\eta}}_{t+1,\ell} \widehat{\boldsymbol{u}}_t(\ell)' = \mathbf{O}$ (the first-order condition from (3.4)), we have

$$\left(\widehat{\boldsymbol{A}}(\ell) - \boldsymbol{A}(\ell) \right) \left(\frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\boldsymbol{u}}_t(\ell) \widehat{\boldsymbol{u}}_t(\ell)' \right) = \frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\boldsymbol{\eta}}_{t+1,\ell} \widehat{\boldsymbol{u}}_t(\ell)'. \tag{S.10}$$

If we can show that $\frac{1}{T-q} \sum_{t=q}^{T-1} \widehat{\boldsymbol{u}}_t(q) \widehat{\boldsymbol{u}}_t(q)'$ is asymptotically invertible, then $\frac{1}{T-\ell} \sum_{t=\ell}^{T-1} \widehat{\boldsymbol{u}}_t(\ell) \widehat{\boldsymbol{u}}_t(\ell)'$ must also be asymptotically invertible with probability 1, for any $1 \leq \ell \leq q$.[1] By the triangular inequality, $\left\| \frac{1}{T-q} \sum_{t=q}^{T-1} \widehat{\boldsymbol{u}}_t(q) \widehat{\boldsymbol{u}}_t(q)' - \mathbb{E}\left(\boldsymbol{u}_t(q) \boldsymbol{u}_t(q)' \right) \right\| \leq I_a + I_b$, where

$$I_a = \left\| \frac{1}{T-q} \sum_{t=q}^{T-1} \boldsymbol{u}_t(q) \boldsymbol{u}_t(q)' - \mathbb{E}\left(\boldsymbol{u}_t(q) \boldsymbol{u}_t(q)' \right) \right\| = O_p\left(q/\sqrt{T} \right)$$

by Chebyshev's inequality and Lemma S2 (i), and

$$\begin{aligned}
I_b &= \left\| \frac{1}{T-q} \sum_{t=q}^{T-1} \widehat{\boldsymbol{u}}_t(q) \widehat{\boldsymbol{u}}_t(q)' - \frac{1}{T-q} \sum_{t=q}^{T-1} \boldsymbol{u}_t(q) \boldsymbol{u}_t(q)' \right\| \\
&\leq \frac{1}{T-q} \sum_{t=q}^{T-1} \left\| \widehat{\boldsymbol{u}}_t(q) - \boldsymbol{u}_t(q) \right\|^2 + 2 \sqrt{\frac{1}{T-q} \sum_{t=q}^{T-1} \left\| \widehat{\boldsymbol{u}}_t(q) - \boldsymbol{u}_t(q) \right\|^2} \sqrt{\frac{1}{T-q} \sum_{t=q}^{T-1} \left\| \boldsymbol{u}_t(q) \right\|^2},
\end{aligned}$$

[1] If the matrix $\boldsymbol{Q}$ is invertible, then each leading principle submatrix of $\boldsymbol{Q}$ is invertible as well.

since $\|\sum_t \boldsymbol{a}_t\boldsymbol{a}_t' - \sum_t \boldsymbol{b}_t\boldsymbol{b}_t'\| \le \sum_t \|\boldsymbol{a}_t - \boldsymbol{b}_t\|^2 + 2\sqrt{\sum_t \|\boldsymbol{a}_t - \boldsymbol{b}_t\|^2}\sqrt{\sum_t \|\boldsymbol{b}_t\|^2}$. We have $\frac{1}{T-q}\sum_{t=q}^{T-1} \|\widehat{\boldsymbol{u}}_t(q) - \boldsymbol{u}_t(q)\|^2 = \frac{1}{T-q}\sum_{t=q}^{T-1}\sum_{s=t-q+1}^{t} \|\widehat{\boldsymbol{u}}_s - \boldsymbol{u}_s\|^2 \le \frac{q}{T-q}\|\widehat{\boldsymbol{u}} - \boldsymbol{u}\|^2 = \frac{q}{T}O_p(1)$ by Assumption 2. Because $\frac{1}{T-q}\sum_{t=q}^{T-1}\|\boldsymbol{u}_t(q)\| = O_p(q)$ by Markov's inequality, we conclude $I_b = O_p\left(q/\sqrt{T}\right)$. Overall, this gives

$$\left\|\frac{1}{T-q}\sum_{t=q}^{T-1}\widehat{\boldsymbol{u}}_t(q)\widehat{\boldsymbol{u}}_t(q)' - \mathbb{E}\big(\boldsymbol{u}_t(q)\boldsymbol{u}_t(q)'\big)\right\| = O_p\left(\frac{q}{\sqrt{T}}\right) + O_p\left(\frac{q}{\sqrt{T}}\right) = o_p(1). \tag{S.11}$$

Now observe that $\mathbb{E}\big(\boldsymbol{u}_t(q)\boldsymbol{u}_t(q)'\big)$ is a leading principal submatrix of $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ (thus invertible, see footnote 1). As a result, $\frac{1}{T-q}\sum_{t=q}^{T-1}\widehat{\boldsymbol{u}}_t(q)\widehat{\boldsymbol{u}}_t(q)'$ is asymptotically invertible.

We subsequently bound the RHS of (S.10) as follows: $\max_{1\le\ell\le q}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\widehat{\boldsymbol{\eta}}_{t+1,\ell}\widehat{\boldsymbol{u}}_t(\ell)'\right\| \le II_a + \ldots + II_e$, where $II_a = \max_{1\le\ell\le q}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\boldsymbol{\eta}_{t+1}\boldsymbol{u}_t(\ell)'\right\|$, $II_b = \max_{1\le\ell\le q}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\left(\boldsymbol{\eta}_{t+1,\ell} - \boldsymbol{\eta}_{t+1}\right)\boldsymbol{u}_t(\ell)'\right\|$, and

$$\begin{aligned}
II_c &= \max_{1\le\ell\le q}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\left(\widehat{\boldsymbol{\eta}}_{t+1,\ell} - \boldsymbol{\eta}_{t+1,\ell}\right)\boldsymbol{u}_t(\ell)'\right\|,\\
II_d &= \max_{1\le\ell\le q}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\boldsymbol{\eta}_{t+1,\ell}\left(\widehat{\boldsymbol{u}}_t(\ell) - \boldsymbol{u}_t(\ell)\right)'\right\|,\\
II_e &= \max_{1\le\ell\le q}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\left(\widehat{\boldsymbol{\eta}}_{t+1,\ell} - \boldsymbol{\eta}_{t+1,\ell}\right)\left(\widehat{\boldsymbol{u}}_t(\ell) - \boldsymbol{u}_t(\ell)\right)'\right\|.
\end{aligned}$$

We consider these terms separately starting from II_a. Using the properties of Frobenius norm,

$$\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\boldsymbol{\eta}_{t+1}\boldsymbol{u}_t(\ell)'\right\|^2 \le \sum_{s=0}^{\ell-1}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\boldsymbol{\eta}_{t+1}\boldsymbol{u}_{t-s}'\right\|_{\mathcal{F}}^2 = \frac{1}{(T-\ell)^2}\sum_{s=0}^{\ell-1}\sum_{i,j=1}^{n}\left|\sum_{t=\ell}^{T-1}\eta_{i,t+1}u_{j,t-s}\right|^2.$$

Assumption 1 justifies the use of Lemma 2 in Wei (1987) which gives $\mathbb{E}\left|\sum_{t=\ell}^{T-1}\eta_{i,t+1}u_{j,t-s}\right|^2 \le C\sum_{t=\ell}^{T-1}\mathbb{E}\left(u_{j,t-s}^2\right) \le C(T-\ell)$. By Chebyshev's inequality, $\forall\varepsilon > 0$, there exists $\alpha_\varepsilon > 0$ such that

$$\mathbb{P}\left(II_a \ge \alpha_\varepsilon\frac{q}{\sqrt{T}}\right) \le \frac{1}{\alpha_\varepsilon^2}\frac{T}{q^2}\sum_{\ell=1}^{q}\mathbb{E}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\boldsymbol{\eta}_{t+1}\boldsymbol{u}_t(\ell)'\right\|^2 \le \frac{C}{\alpha_\varepsilon^2} \le \varepsilon,$$

and thus $II_a = O_p\left(q/\sqrt{T}\right)$. Furthermore, we deduce that $II_b = O_p\left(q/\sqrt{T}\right)$ by Lemma S2(ii) and Chebyshev's inequality. For II_c, if we write $\widehat{\boldsymbol{\eta}}_{t+1,\ell} - \boldsymbol{\eta}_{t+1,\ell} = \left[\boldsymbol{I}_n, -\boldsymbol{A}(\ell)\right]\left[\widehat{\boldsymbol{u}}_{t+1}(\ell+1) - \boldsymbol{u}_{t+1}(\ell+1)\right]$, then by Cauchy-Schwarz inequality and Baxter's inequality (leads to $\max_{1\le\ell\le q}\|\boldsymbol{A}(\ell)\|^2 \le C$),

$$II_c \le C\sqrt{\max_{1\le\ell\le q}\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\|\widehat{\boldsymbol{u}}_{t+1}(\ell+1) - \boldsymbol{u}_{t+1}(\ell+1)\|^2}\sqrt{\max_{1\le\ell\le q}\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\|\boldsymbol{u}_t(\ell)\|^2} = O_p\left(\frac{q}{\sqrt{T}}\right),$$

where the last step follows from arguments similar to those preceding (S.11). Similarly, $II_d = O_p\left(q/\sqrt{T}\right)$

and $II_e = O_p\left(q/T\right)$. Combining all results, we finally have

$$\max_{1\leq\ell\leq q}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\widehat{\boldsymbol{\eta}}_{t+1,\ell}\widehat{\boldsymbol{u}}_t(\ell)'\right\| = O_p\left(\frac{q}{\sqrt{T}}\right). \tag{S.12}$$

By invertibility of $\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\widehat{\boldsymbol{u}}_t(\ell)\widehat{\boldsymbol{u}}_t(\ell)'$, (S.10) and (S.12), $\max_{1\leq\ell\leq q}\left\|\widehat{\boldsymbol{A}}(\ell)-\boldsymbol{A}(\ell)\right\| = O_p\left(q/\sqrt{T}\right)$ follows.

We continue with $\max_{1\leq\ell\leq q}\left\|\widehat{\boldsymbol{S}}(\ell)-\boldsymbol{S}(\ell)\right\| = O_p\left(q/\sqrt{T}\right)$. Since $\widetilde{\boldsymbol{\eta}}_{t+1,\ell} = \widehat{\boldsymbol{\eta}}_{t+1,\ell} - \left(\widehat{\boldsymbol{A}}(\ell)-\boldsymbol{A}(\ell)\right)\widehat{\boldsymbol{u}}_t(\ell)$ (see (S.9)), we can use (S.10) and the invertibility of $\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\widehat{\boldsymbol{u}}_t(\ell)\widehat{\boldsymbol{u}}_t(\ell)'$ to write

$$\begin{aligned}
&\max_{1\leq\ell\leq q}\left\|\widehat{\boldsymbol{S}}(\ell)-\boldsymbol{S}(\ell)\right\| = \max_{1\leq\ell\leq q}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\left(\widetilde{\boldsymbol{\eta}}_{t+1,\ell}\widetilde{\boldsymbol{\eta}}'_{t+1,\ell}\right)-\mathbb{E}\left(\boldsymbol{\eta}_{t+1,\ell}\boldsymbol{\eta}'_{t+1,\ell}\right)\right\| \\
&= \max_{1\leq\ell\leq q}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\left(\widehat{\boldsymbol{\eta}}_{t+1,\ell}\widehat{\boldsymbol{\eta}}'_{t+1,\ell}\right)-\mathbb{E}\left(\boldsymbol{\eta}_{t+1,\ell}\boldsymbol{\eta}'_{t+1,\ell}\right)-\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\widehat{\boldsymbol{\eta}}_{t+1,\ell}\widehat{\boldsymbol{u}}_t(\ell)'\left(\widehat{\boldsymbol{A}}(\ell)-\boldsymbol{A}(\ell)\right)'\right\| \\
&\leq \max_{1\leq\ell\leq q}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\left(\widehat{\boldsymbol{\eta}}_{t+1,\ell}\widehat{\boldsymbol{\eta}}'_{t+1,\ell}\right)-\mathbb{E}\left(\boldsymbol{\eta}_{t+1,\ell}\boldsymbol{\eta}'_{t+1,\ell}\right)\right\| + C\max_{1\leq\ell\leq q}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\widehat{\boldsymbol{\eta}}_{t+1,\ell}\widehat{\boldsymbol{u}}_t(\ell)'\right\|^2 \\
&\leq \max_{1\leq\ell\leq q}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\left(\widehat{\boldsymbol{\eta}}_{t+1,\ell}\widehat{\boldsymbol{\eta}}'_{t+1,\ell}-\boldsymbol{\eta}_{t+1,\ell}\boldsymbol{\eta}'_{t+1,\ell}\right)\right\| + O_p\left(\sqrt{\frac{q}{T}}\right)+O_p\left(\frac{q^2}{T}\right),
\end{aligned} \tag{S.13}$$

where the last step in (S.13) follows from $\max_{1\leq\ell\leq q}\left\|\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\left(\boldsymbol{\eta}_{t+1,\ell}\boldsymbol{\eta}'_{t+1,\ell}\right)-\mathbb{E}\left(\boldsymbol{\eta}_{t+1,\ell}\boldsymbol{\eta}'_{t+1,\ell}\right)\right\| = O_p\left(\sqrt{q/T}\right)$ using Lemma S2(iii) and (S.12). By the inequality $\left\|\sum_t \boldsymbol{a}_t\boldsymbol{a}'_t-\sum_t\boldsymbol{b}_t\boldsymbol{b}'_t\right\| \leq \sum_t\left\|\boldsymbol{a}_t-\boldsymbol{b}_t\right\|^2 + 2\sqrt{\sum_t\left\|\boldsymbol{a}_t-\boldsymbol{b}_t\right\|^2}\sqrt{\sum_t\left\|\boldsymbol{b}_t\right\|^2}$ and similar arguments as for II_c and II_d above, the first term in (S.13) is bounded by

$$\begin{aligned}
&\max_{1\leq\ell\leq q}\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\left\|\widehat{\boldsymbol{\eta}}_{t+1,\ell}-\boldsymbol{\eta}_{t+1,\ell}\right\|^2 \\
&\qquad + 2\sqrt{\max_{1\leq\ell\leq q}\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\left\|\widehat{\boldsymbol{\eta}}_{t+1,\ell}-\boldsymbol{\eta}_{t+1,\ell}\right\|^2}\sqrt{\max_{1\leq\ell\leq q}\frac{1}{T-\ell}\sum_{t=\ell}^{T-1}\left\|\boldsymbol{\eta}_{t+1,\ell}\right\|^2} = O_p\left(\frac{q}{\sqrt{T}}\right).
\end{aligned}$$

Overall, we obtain $\max_{1\leq\ell\leq q}\left\|\widehat{\boldsymbol{S}}(\ell)-\boldsymbol{S}(\ell)\right\| = O_p\left(q/\sqrt{T}\right)$ as well. ■

Lemma S4

Under Assumptions 1 and 3, we have

$$\left\|\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q)-\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\right\| \leq C\frac{1}{\sqrt{q}}\sum_{s=q+1}^{\infty}s\left\|\boldsymbol{A}_s\right\|_{\mathcal{F}}. \tag{S.14}$$

Proof Consider $\left\|\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q)-\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\right\|$. A rewriting as in (A.13) shows that $\left\|\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)-\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}\right\|$ and $\left\|\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q)-\right.$

$\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}\|$ are the two important terms to bound. Hölder's inequality implies

$$\left\|\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)-\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}\right\| \leq \sqrt{\left\|\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)-\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}\right\|_1 \left\|\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)-\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}\right\|_\infty}. \tag{S.15}$$

For the matrix 1-norm we are concerned with the maximum absolute column sum. For an arbitrary $(nT \times nT)$ matrix $\boldsymbol{Q}$ partitioned (block) column-wise, i.e. $\boldsymbol{Q}=[\boldsymbol{Q}_1,\boldsymbol{Q}_2,\ldots,\boldsymbol{Q}_T]$, we have the bound $\|\boldsymbol{Q}\|_1 = \max_{1\leq t\leq T}\|\boldsymbol{Q}_t\|_1 \leq \sqrt{n}\max_{1\leq t\leq T}\|\boldsymbol{Q}_t\|_{\mathcal{F}}$. This implies $\left\|\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q)-\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}\right\|_1 \leq \sqrt{n}\max\{I_a, I_b\}$ where

$$I_a = \max_{0\leq j\leq T-q-2}\left(\sum_{i=0}^{T-q-2-j}\|\boldsymbol{A}_{q+1+i}(q+1+i+j)\|_{\mathcal{F}}^2 + \sum_{i=\max(1,q+1-j)}^{q}\|\boldsymbol{A}_i(i+j)-\boldsymbol{A}_i(q)\|_{\mathcal{F}}^2\right)^{1/2},$$

$$I_b = \max_{T-q-1\leq j\leq T-2}\left(\sum_{i=1}^{T-1-j}\|\boldsymbol{A}_i(i+j)-\boldsymbol{A}_i(q)\|_{\mathcal{F}}^2\right)^{1/2}.$$

We will bound the three summations that are encountered in the expressions for I_a and I_b. First, changing the summation index and using c_r-inequality,

$$\begin{aligned}\sum_{i=0}^{T-q-2-j}\left\|\boldsymbol{A}_{q+1+i}(q+1+i+j)\right\|_{\mathcal{F}}^2 &= \sum_{s=q+1}^{T-1-j}\|\boldsymbol{A}_s(s+j)\|_{\mathcal{F}}^2 \\ &\leq 2\sum_{s=q+1}^{T-1-j}\|\boldsymbol{A}_s(s+j)-\boldsymbol{A}_s\|_{\mathcal{F}}^2 + 2\sum_{s=q+1}^{T-1-j}\|\boldsymbol{A}_s\|_{\mathcal{F}}^2 \\ &\leq 2\sum_{s=q+1}^{T-1-j}s^{-2}\left(s^2\sum_{k=1}^{s+j}\|\boldsymbol{A}_k(s+j)-\boldsymbol{A}_k\|_{\mathcal{F}}^2\right) + 2\sum_{s=q+1}^{\infty}\|\boldsymbol{A}_s\|_{\mathcal{F}}^2.\end{aligned}$$

For convenience, we define

$$\mathcal{K}_q = \left(\sum_{s=q+1}^{\infty} s\,\|\boldsymbol{A}_s\|_{\mathcal{F}}\right)^2\left(\sum_{s=q+1}^{\infty}\frac{1}{s^2}\right). \tag{S.16}$$

For any $j \geq 0$ and $s \geq q+1$, we have $s^2\sum_{k=1}^{s+j}\|\boldsymbol{A}_k(s+j)-\boldsymbol{A}_k\|_{\mathcal{F}}^2 \leq Cs^2\sum_{k=s+j+1}^{\infty}\|\boldsymbol{A}_k\|_{\mathcal{F}}^2 \leq C\left(\sum_{k=s+j+1}^{\infty}k\,\|\boldsymbol{A}_k\|_{\mathcal{F}}\right)^2$ by the L^2-Baxter's inequality. The first term in the RHS above is thus bounded by $C\mathcal{K}_q$. Moreover, by Cauchy-Schwartz inequality, the second term can be bounded as $\sum_{s=q+1}^{\infty}\|\boldsymbol{A}_s\|_{\mathcal{F}}^2 = \sum_{s=q+1}^{\infty}\left(s^2\|\boldsymbol{A}_s\|_{\mathcal{F}}^2\right)s^{-2} \leq \left(\sum_{s=q+1}^{\infty}s^2\|\boldsymbol{A}_s\|_{\mathcal{F}}^2\right)\left(\sum_{s=q+1}^{\infty}s^{-2}\right) \leq \mathcal{K}_q$. For the second

summation in I_a, we first consider the case $0 \le j \le q$, or $\max(1, q+1-j) = q+1-j$, such that

$$\begin{aligned}\sum_{i=\max(1,q+1-j)}^{q} \|\boldsymbol{A}_i(i+j) - \boldsymbol{A}_i(q)\|_{\mathcal{F}}^2 &= \sum_{i=q+1-j}^{q} \|\boldsymbol{A}_i(i+j) - \boldsymbol{A}_i(q)\|_{\mathcal{F}}^2 \\ &\le 2 \sum_{i=q+1-j}^{q} \|\boldsymbol{A}_i(i+j) - \boldsymbol{A}_i\|_{\mathcal{F}}^2 + 2 \sum_{i=q+1-j}^{q} \|\boldsymbol{A}_i(q) - \boldsymbol{A}_i\|_{\mathcal{F}}^2 \\ &\le 2 \sum_{s=q+1}^{q+j} \|\boldsymbol{A}_{s-j}(s) - \boldsymbol{A}_{s-j}\|_{\mathcal{F}}^2 + 2 \sum_{i=1}^{q} \|\boldsymbol{A}_i(q) - \boldsymbol{A}_i\|_{\mathcal{F}}^2, \le C\mathcal{K}_q\end{aligned}$$

using arguments detailed before. This upper bound remains valid for $q+1 \le j \le T-q-2$. It is likewise straightforward to derive $\sum_{i=1}^{T-1-j} \|\boldsymbol{A}_i(i+j) - \boldsymbol{A}_i(q)\|_{\mathcal{F}}^2 \le C\mathcal{K}_q$. Collecting all the results, we have $I_a \le C\sqrt{\mathcal{K}_q}$, $I_b \le C\sqrt{\mathcal{K}_q}$, and thus $\left\|\boldsymbol{F}_{\boldsymbol{u}}(q) - \boldsymbol{F}_{\boldsymbol{u}}\right\|_1 \le C\sqrt{\mathcal{K}_q}$.

For $\|\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}\|_\infty$, we are bounding the maximum absolute row sums. For an arbitrary $(nT \times nT)$ matrix $\boldsymbol{Q}$ partitioned as $\boldsymbol{Q} = [\boldsymbol{Q}_1', \boldsymbol{Q}_2', \ldots, \boldsymbol{Q}_T']'$, we have $\|\boldsymbol{Q}\|_\infty = \max_{1\le t\le T} \|\boldsymbol{Q}_t\|_\infty \le \sqrt{n} \max_{1\le t\le T} \|\boldsymbol{Q}_t\|_{\mathcal{F}}$, such that

$$\|\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}\|_\infty \le \sqrt{n} \max_{q+1\le m\le T-1} \left\{ \sum_{j=q+1}^{m} \|\boldsymbol{A}_j(m)\|_{\mathcal{F}}^2 + \sum_{j=1}^{q} \|\boldsymbol{A}_j(q) - \boldsymbol{A}_j(m)\|_{\mathcal{F}}^2 \right\}^{1/2},$$

where $\sum_{j=q+1}^{m} \|\boldsymbol{A}_j(m)\|_{\mathcal{F}}^2 \le C\mathcal{K}_q$ and $\sum_{j=1}^{q} \|\boldsymbol{A}_j(q) - \boldsymbol{A}_j(m)\|_{\mathcal{F}}^2 \le C\mathcal{K}_q$, for any $q+1 \le m \le T-1$, using the L^2-Baxter's inequality and the previous upper bound on $\sum_{s=q+1}^{\infty} \|\boldsymbol{A}_s\|_{\mathcal{F}}^2$. We conclude that $\|\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}\|_\infty \le C\sqrt{\mathcal{K}_q}$. Together with our previous we result, we obtain $\left\|\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}\right\| \le C\sqrt{\mathcal{K}_q}$ from (S.15).

From $\left\|\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q) - \boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}\right\| \le \left\|\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{S}}_{\boldsymbol{u}}\right\| \left\|\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}\right\| \left\|\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q)\right\|$ we see that it suffices to inspect $\left\|\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{S}}_{\boldsymbol{u}}\right\|$ (the other norms are bounded). Exploiting the fact that both $\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}(q)$ and $\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}$ are block-diagonal, we have $\left\|\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{S}}_{\boldsymbol{u}}\right\| = \max_{q+1\le k\le T-1} \|\boldsymbol{S}(q) - \boldsymbol{S}(k)\| \le 2\max_{q\le k\le T-1} \|\boldsymbol{S}(k) - \boldsymbol{\Sigma}_{\eta\eta}\|$. Let $\boldsymbol{A}_j(\ell) = \mathbf{O}$ for $j > \ell$, and recall the definition of $\boldsymbol{\eta}_{t+1,\ell}$ in (S.2). We find, for any $k \ge q$,

$$\begin{aligned}\|\boldsymbol{S}(k) - \boldsymbol{\Sigma}_{\eta\eta}\| &= \left\|\mathbb{E}\left(\boldsymbol{\eta}_{t+1,k} - \boldsymbol{\eta}_{t+1}\right)\left(\boldsymbol{\eta}_{t+1,k} - \boldsymbol{\eta}_{t+1}\right)'\right\| \\ &\le C \sum_{s=1}^{\infty} \|\boldsymbol{A}_s(k) - \boldsymbol{A}_s\|^2 \le C \sum_{s=q+1}^{\infty} \|\boldsymbol{A}_s\|_{\mathcal{F}}^2 \le C\mathcal{K}_q. \end{aligned} \tag{S.17}$$

We thereby obtain $\left\|\boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}(q) - \boldsymbol{\mathcal{S}}_{\boldsymbol{u}}^{-1}\right\| \le C\mathcal{K}_q$. Together with the bound on $\left\|\boldsymbol{\mathcal{M}}_{\boldsymbol{u}}(q) - \boldsymbol{\mathcal{M}}_{\boldsymbol{u}}\right\|$, we deduce

$$\left\|\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(q) - \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\right\| \le C\sqrt{\mathcal{K}_q} = C\left(\sum_{s=q+1}^{\infty} s\,\|\boldsymbol{A}_s\|_{\mathcal{F}}\right)\sqrt{\sum_{s=q+1}^{\infty} \frac{1}{s^2}} \le C\frac{1}{\sqrt{q}} \sum_{s=q+1}^{\infty} s\,\|\boldsymbol{A}_s\|_{\mathcal{F}}.$$

The proof is complete. ■

Theorem S1

Let $\boldsymbol{W}(r) = [W_1(r), W_2(r), \ldots, W_n(r)]'$ *denote an* n*-dimensional standard Brownian motion. The cumulative distribution function (CDF) of* $\int_0^1 \|\boldsymbol{W}(r)\|^2 dr$ *is given by*

$$F_n(x) = 2^{n/2} \sum_{j=0}^{\infty} k_{j,n} \operatorname{Erfc}\left(\frac{l_{j,n}}{2\sqrt{x}}\right), \qquad x > 0,$$

where $k_{j,n} = (-1)^j \frac{\Gamma(j+n/2)}{j!\Gamma(n/2)}$, $l_{j,n} = 2\sqrt{2}j + \frac{n}{\sqrt{2}}$ *and* $\operatorname{Erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt$.

Proof We follow the approach from Example 1 of Anderson and Darling (1952) or equivalently appendix B of Choi and Saikkonen (2010). Let f_n denote the probability density function of $\int_0^1 \|\boldsymbol{W}(r)\|^2 dr$ and write $\mathcal{L}\{\cdot\}$ and $\mathcal{L}^{-1}\{\cdot\}$ for the Laplace and inverse Laplace operator, respectively. From the equality $\int_0^1 \|\boldsymbol{W}(r)\|^2 dr = \sum_{i=1}^n \int_0^1 W_i(r)^2 dr$, independence of the components of $\boldsymbol{W}(r)$, and the known univariate result in Choi and Saikkonen (2010), we have

$$\mathcal{L}\{f_n(x)\}(t) = \int_0^{\infty} e^{-tx} f_n(x) dx = \left[\cosh\left(\sqrt{2t}\right)\right]^{-n/2}. \tag{S.18}$$

According to equation (4.28) in Anderson and Darling (1952), the CDF is

$$\begin{aligned} F_n(x) &= \mathcal{L}^{-1}\left\{\frac{1}{t}\left[\cosh\left(\sqrt{2t}\right)\right]^{-n/2}\right\}(x) = \mathcal{L}^{-1}\left\{\frac{1}{t}\left(\frac{e^{\sqrt{2t}}}{2}\right)^{-n/2}\left[1+e^{-2\sqrt{2t}}\right]^{-n/2}\right\}(x) \\ &= \mathcal{L}^{-1}\left\{\frac{1}{t}\left(\frac{e^{\sqrt{2t}}}{2}\right)^{-n/2}\sum_{j=0}^{\infty} k_{j,n} e^{-2j\sqrt{2t}}\right\}(x) = 2^{n/2}\sum_{j=0}^{\infty} k_{j,n}\mathcal{L}^{-1}\left\{\frac{1}{t}e^{-l_{j,n}\sqrt{t}}\right\}(x) \end{aligned} \tag{S.19}$$

where we use (1) a t with a positive real part, (2) linearity of the inverse Laplace operator, and (3) the binomial expansion of $[1+e^{-2\sqrt{2t}}]^{-n/2}$. The identity from Choi and Saikkonen (2010), $\mathcal{L}^{-1}\left\{\frac{1}{t}e^{-u\sqrt{t}}\right\}(x) = 1 - \operatorname{Erf}\left(\frac{u}{2\sqrt{x}}\right) = \operatorname{Erfc}\left(\frac{u}{2\sqrt{x}}\right)$, completes the proof. ■

2 Estimation of quantities for fully modified inference

The FM-GLS estimator relies on $\boldsymbol{\Omega}$, $\boldsymbol{\Delta}$, and $\mathbb{E}(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_t')$ (see Assumption 1). For convenience, we denote this $(2n \times 2n)$ matrix $\begin{bmatrix} \boldsymbol{\Sigma}_{\eta\eta} & \boldsymbol{\Sigma}_{\eta\epsilon} \\ \boldsymbol{\Sigma}_{\epsilon\eta} & \boldsymbol{\Sigma}_{\epsilon\epsilon} \end{bmatrix}$ by $\boldsymbol{\Sigma}$. Please note the difference between $\boldsymbol{\Sigma}$ and the large-dimensional matrix $\boldsymbol{\Sigma}_{\boldsymbol{u}}$. In this section, we consider the estimation of these three quantities within the BIAM framework. For convenience, we recall $\boldsymbol{\xi}_t = [\boldsymbol{u}_t', \boldsymbol{v}_t']'$ and define $\boldsymbol{\xi} = [\boldsymbol{\xi}_1', \boldsymbol{\xi}_2', \ldots, \boldsymbol{\xi}_T']'$. Similarly to the definition of $\boldsymbol{\Sigma}_u$, we used $\boldsymbol{\Sigma}_{\boldsymbol{\xi}} := \mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}')$ to denote the $(2nT \times 2nT)$ autocovariance matrix of $\{\boldsymbol{\xi}_t\}$. As a sample counterpart, we stack $\widehat{\boldsymbol{u}}_t$ and $\Delta\boldsymbol{x}_t = \boldsymbol{v}_t$ in the $2n$-dimensional vector $\widehat{\boldsymbol{\xi}}_t = \left[\widehat{\boldsymbol{u}}_t', \boldsymbol{v}_t'\right]'$. Using $\{\widehat{\boldsymbol{\xi}}_t\}_{t=1}^T$, the BIAM estimator for $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$ is now constructed as $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}}(q) = \widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{\xi}}^{-1}(q)\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\xi}}(q)\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{\xi}}^{-1\prime}(q)$, where the matrices $\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{\xi}}(q)$ and $\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\xi}}(q)$ are defined similarly to $\widehat{\boldsymbol{\mathcal{M}}}_{\boldsymbol{u}}(q)$ and $\widehat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{u}}(q)$, respectively. Since the BIAM estimator is fitting VAR processes up to order q_T (see (3.4)), the coefficient estimates $\widehat{\boldsymbol{F}}_j(q_T)$ of the jth lag when a VAR(q_T) is fitted, $j = 1, 2, \ldots, q_T$, are immediate byproducts of the BIAM procedure and can thus be used to construct our

estimators. Finally, if $\boldsymbol{\mathcal{F}}(L) = \text{diag}\left[\boldsymbol{\mathcal{A}}(L), \boldsymbol{\mathcal{D}}(L)\right] := \boldsymbol{I}_{2n} - \sum_{j=1}^{\infty} \boldsymbol{F}_j L^j$, where $\boldsymbol{F}_j = \text{diag}\left[\boldsymbol{A}_j, \boldsymbol{D}_j\right]$, then $\boldsymbol{\mathcal{F}}(L)\boldsymbol{\xi}_t = \boldsymbol{\zeta}_t$ holds.

Theorem S2

Recall $\widehat{\boldsymbol{\Omega}}_{q_T} = \left(\boldsymbol{I}_{2n} - \sum_{j=1}^{q_T} \widehat{\boldsymbol{F}}_j(q_T)\right)^{-1} \widehat{\boldsymbol{\Sigma}}_{q_T} \left(\boldsymbol{I}_{2n} - \sum_{j=1}^{q_T} \widehat{\boldsymbol{F}}_j(q_T)\right)^{-1\prime}$, $\widehat{\boldsymbol{\Delta}}_{q_T,r_T} = \boldsymbol{Q}_{r_T}' \widehat{\boldsymbol{\Sigma}_{\boldsymbol{\xi}}}(q_T)\boldsymbol{Q}_1$, *and*

$$\widehat{\boldsymbol{\Sigma}}_{q_T} = \frac{1}{T - q_T} \sum_{t=q_T+1}^{T} \left[\widehat{\boldsymbol{\xi}}_t - \widehat{\boldsymbol{F}}_1(q_T)\widehat{\boldsymbol{\xi}}_{t-1} - \cdots - \widehat{\boldsymbol{F}}_{q_T}(q_T)\widehat{\boldsymbol{\xi}}_{t-q_T}\right]\left[\widehat{\boldsymbol{\xi}}_t - \widehat{\boldsymbol{F}}_1(q_T)\widehat{\boldsymbol{\xi}}_{t-1} - \cdots - \widehat{\boldsymbol{F}}_{q_T}(q_T)\widehat{\boldsymbol{\xi}}_{t-q_T}\right]',$$

where $\boldsymbol{Q}_r = [\mathbf{O}_{2n\times 2n}, \cdots, \mathbf{O}_{2n\times 2n}, \boldsymbol{I}_{2n}, \cdots, \boldsymbol{I}_{2n}]'$ *is an* $(2nT \times 2n)$ *block matrices of zeros of which the last* r *blocks have been replaced by identity matrices. If Assumptions 1-3 and 5 hold, then*

$$\left\|\widehat{\boldsymbol{\Sigma}}_{q_T} - \boldsymbol{\Sigma}\right\| = O_p\left(\frac{q_T}{\sqrt{T}}\right) + O\left(\sum_{s=q_T+1}^{\infty} \|\boldsymbol{F}_s\|_{\mathcal{F}}^2\right) = o_p(1), \tag{S.1}$$

$$\left\|\widehat{\boldsymbol{\Omega}}_{q_T} - \boldsymbol{\Omega}\right\| = O_p\left(\sqrt{\frac{q_T^3}{T}} + \frac{1}{q_T}\sum_{s=q_T+1}^{\infty} s\|\boldsymbol{F}_s\|_{\mathcal{F}}\right) = o_p(1), \tag{S.2}$$

$$\left\|\widehat{\boldsymbol{\Delta}}_{q_T,r_T} - \boldsymbol{\Delta}\right\| = o(r_T^{-1}) + O_p\left(\sqrt{\frac{q_T^3}{T}} r_T + \sqrt{\frac{r_T}{q_T}}\sum_{s=q_T+1}^{\infty} s\|\boldsymbol{F}_s\|_{\mathcal{F}}\right) = o_p(1). \tag{S.3}$$

Proof Note that $\widehat{\boldsymbol{\xi}}_t - \boldsymbol{\xi}_t = [(\widehat{\boldsymbol{u}}_t - \boldsymbol{u}_t)', \mathbf{0}']'$ and hence $\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2 = \|\widehat{\boldsymbol{u}} - \boldsymbol{u}\|^2 = O_p(1)$ by Assumption 2. The conditions for Lemmas S3 – S4 and Theorem 2 are thus satisfied and we can use these results in subsequent proofs. **(a)** The result (S.1) follows from the triangle inequality, Lemma S3 and (S.17). **(b)** The second result (S.2) is obtained by the definition $\boldsymbol{\Omega} = (\boldsymbol{I}_{2n} - \sum_{j=1}^{\infty} \boldsymbol{F}_j)^{-1}\boldsymbol{\Sigma}(\boldsymbol{I}_{2n} - \sum_{j=1}^{\infty} \boldsymbol{F}_j)^{-1\prime}$, Lemma S3 and a straightforward modification of (A.13). **(c)** By $\boldsymbol{\Delta} = \sum_{h=r_T}^{\infty} \mathbb{E}\left(\boldsymbol{\xi}_t\boldsymbol{\xi}_{t+h}'\right) + \boldsymbol{Q}_{r_T}'\boldsymbol{\Sigma}_{\boldsymbol{\xi}}\boldsymbol{Q}_1$, the LHS of (S.3) can be bounded

$$\left\|\widehat{\boldsymbol{\Delta}}_{q_T,r_T} - \boldsymbol{\Delta}\right\| \leq \sum_{h=r_T}^{\infty} \left\|\mathbb{E}\left(\boldsymbol{\xi}_t\boldsymbol{\xi}_{t+h}'\right)\right\|_{\mathcal{F}} + \left\|\boldsymbol{Q}_{r_T}'\boldsymbol{\Sigma}_{\boldsymbol{\xi}}\right\| \left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}}(q_T) - \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}\right\| \left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{\xi}}}(q_T)\boldsymbol{Q}_1\right\|.$$

Since summability conditions on the coefficient matrices carry over to the autocovariances, we have $\sum_{h=r_T}^{\infty} \left\|\mathbb{E}\left(\boldsymbol{\xi}_t\boldsymbol{\xi}_{t+h}'\right)\right\|_{\mathcal{F}} \leq r_T^{-1}\sum_{h=r_T}^{\infty} h \left\|\mathbb{E}\left(\boldsymbol{\xi}_t\boldsymbol{\xi}_{t+h}'\right)\right\|_{\mathcal{F}} = o(r_T^{-1})$ by Assumption 1. Moreover, $\left\|\boldsymbol{Q}_{r_T}'\boldsymbol{\Sigma}_{\boldsymbol{\xi}}\right\| \leq C\sqrt{r_T}$ and $\left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}}(q_T) - \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}\right\|$ is discussed in Theorem 2. Finally, showing $\left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{\xi}}}(q_T)\boldsymbol{Q}_1\right\| = O_p(1)$ will complete the proof after a straightforward comparison of the established stochastic orders. It suffices to prove $\left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{\xi}}}(q_T)\right\| = O_p(1)$. Weyl's inequality (e.g. pages 40 and 46 in Tao (2012)) and Theorem 2 imply

$$\left|\lambda_{min}\left(\widehat{\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}}(q_T)\right) - \lambda_{min}\left(\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}\right)\right| \leq \left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}}(q_T) - \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}\right\| = o_p(1).$$

By the uniform boundedness of $\left\|\boldsymbol{\Sigma}_{\boldsymbol{\xi}}\right\|$, for a sufficiently large T, there exists a constant $C > 0$ such that $\left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{\xi}}}(q_T)\right\|^{-1} = \lambda_{min}\left(\widehat{\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}}(q_T)\right) \leq C$ and thus $\left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{\xi}}}(q_T)\right\| \leq C^{-1}$ with arbitrarily high probability. ■

3 Details on implementation

The implementation of the BIAM estimator and the subsampling KPSS tests requires selecting the banding parameter q and the block length b. In our simulations and empirical application, we follow the subsampling and risk-minimization approach previously used by Bickel and Levina (2008), Wu and Pourahmadi (2009) and Ing et al. (2016) to select q. The steps are as follows:

Step 1 Split the series of (first-step OLS) residuals, $\{\widehat{\boldsymbol{u}}_t\}_{t=1}^{T}$, into J_0 non-overlapping subsequences of length l_0. These subsequences are $\{\widehat{\boldsymbol{u}}_t\}_{t=(j-1)l_0+1}^{jl_0}$ for $j = 1, \ldots, J_0$ with $J_0 = [T/l_0]$.

Step 2 Select an integer H, $1 \leq H < l_0$, and construct the $(nH \times nH)$ sample autocovariance matrix $\widehat{\boldsymbol{\Pi}}_{\boldsymbol{u},nH} = \frac{1}{T-H}\sum_{t=H}^{T-1}\widehat{\boldsymbol{u}}_t(H)\widehat{\boldsymbol{u}}_t(H)'$ which is an estimator of $\boldsymbol{\Sigma}_{\boldsymbol{u},nH} := \mathbb{E}\left[\boldsymbol{u}(H)\boldsymbol{u}(H)'\right]$ with $\boldsymbol{u}(H) := \begin{bmatrix} \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_H \end{bmatrix}$, where $\widehat{\boldsymbol{u}}_t(\ell) = \left(\widehat{\boldsymbol{u}}_t', \cdots, \widehat{\boldsymbol{u}}_{t-\ell+1}'\right)'$. Compute $\widehat{\boldsymbol{\Pi}}_{\boldsymbol{u},nH}^{-1}$.

Step 3 For every subsequence of residuals $1 \leq j \leq J_0$, compute the BIAM estimate of $\boldsymbol{\Sigma}_{\boldsymbol{u},nH}$ repeatedly for all possible banding parameters $1 \leq \bar{q} < H$, denoted as $\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u},nH}^{-1}}(\bar{q}; j)$.

Step 4 Select the banding parameter that minimizes the feasible average risk, i.e.

$$q := \underset{\bar{q}\in[1,H)}{\arg\min} \frac{1}{J_0}\sum_{j=1}^{J_0}\left\|\widehat{\boldsymbol{\Sigma}_{\boldsymbol{u},nH}^{-1}}(\bar{q}; j) - \widehat{\boldsymbol{\Pi}}_{\boldsymbol{u},nH}^{-1}\right\|_p .$$

We take $p = 1$, $H = [2 \times T^{1/4}]$ and $l_0 = [T/5]$ and obtain satisfactory results for all the settings we have explored. As mentioned in Bickel and Levina (2008), the use of another vector norm (e.g. $p = 2$) does not lead to qualitatively different results.

When we implement the minimum volatility rule as mentioned in Section 3 to select b, the values of tuning parameters are adapted from Wagner and Hong (2016). More specifically, we set $b_{max} = \lceil 2T^{0.5}\rceil$ instead of $b_{max} = \lceil 2.5T^{0.5}\rceil$.

References

Anderson, T. W. and D. A. Darling (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics 23*, 193–212.

Bickel, P. J. and E. Levina (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics 36*, 199–227.

Cheng, R. and M. Pourahmadi (1993). Baxter's inequality and convergence of finite predictors of multivariate stochastic processes. *Probability Theory and Related Fields 95*, 115–124.

Choi, I. and P. Saikkonen (2010). Tests for nonlinear cointegration. *Econometric Theory 26*, 682–709.

Findley, D. F. and C.-Z. Wei (1993). Moment bounds for deriving time series CLT's and model selection procedures. *Statistica Sinica 3*, 453–480.

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

Hannan, E. and M. Deistler (2012). *The Statistical Theory of Linear Systems*. Society for Industrial and Applied Mathematics.

Ing, C.-K., H.-T. Chiou, and M. Guo (2016). Estimation of inverse autocovariance matrices for long memory processes. *Bernoulli 22*, 1301–1330.

Tao, T. (2012). *Topics in Random Matrix Theory*. American Mathematical Society.

Wagner, M. and S. H. Hong (2016). Cointegrating polynomial regressions: Fully modified OLS estimation and inference. *Econometric Theory 32*, 1289–1315.

Wei, C.-Z. (1987). Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *The Annals of Statistics 15*, 1667–1682.

Wu, W. B. and M. Pourahmadi (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica 19*, 1755–1768.