Journal of Econometrics 238 (2024) 105622

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

High-dimensional IV cointegration estimation and inference ☆

Peter C.B. Phillips [a,b,c,*], Igor L. Kheifets [d]

[a] Yale University, United States of America
[b] University of Auckland, New Zealand
[c] Singapore Management University, Singapore
[d] Belk College of Business, University of North Carolina at Charlotte, United States of America

A R T I C L E   I N F O

JEL classification:
C12
C13
C22

Keywords:
Cointegration
HAR inference
High-dimensional IV
Long run variance matrix
Multicointegration
Singularity
Trend IV estimation

A B S T R A C T

A semiparametric triangular systems approach shows how multicointegrating linkages occur naturally in an $I(1)$ cointegrated regression model when the long run error variance matrix in the system is singular. Under such singularity, cointegrated $I(1)$ systems embody a multicointegrated structure that makes them useful in many empirical settings. Earlier work shows that such systems may be analyzed and estimated without appealing to the associated $I(2)$ system but with suboptimal convergence rates and potential asymptotic bias. The present paper develops a robust approach to estimation and inference of such systems using high dimensional IV methods that have appealing asymptotic properties like those known to apply in the optimal estimation of cointegrated systems (Phillips, 1991). The approach uses an extended version of high-dimensional trend IV (Phillips, 2006, 2014) estimation with deterministic orthonormal instruments. The methods and derivations involve new results on high-dimensional IV techniques and matrix normalization in the limit theory that are of independent interest. Wald tests of general linear restrictions are constructed using a fixed-b long run variance estimator that leads to robust pivotal HAR inference in both cointegrated and multicointegrated cases. Simulations show good properties of the estimation and inferential procedures in finite samples. An empirical illustration to housing stocks, starts and completions is provided.

1. Introduction

Economic identities that link some variables to partial sums of constituent variables arise frequently in economic data. Examples include common relations between stock and flow versions of variables such as the capital stock and fixed investment, inventory investment and inventory stock, housing construction completions and housing units under construction. Many of these variables have nonstationary characteristics and cointegration models have proved a frequently used framework for empirical work investigating such time series.

The concept of multicointegration was introduced by Granger and Lee (1989, 1990) to allow explicitly for linkages among stock and flow forms of integrated order one ($I(1)$) variables. In particular, multicointegration was suggested to capture the notion that

☆ This paper is a revised and extended version of a paper entitled 'On Multicointegration' that was first presented in the Rimini Conference on Time Series Econometrics held in Larnaca, Cyprus, in June 2019 and later circulated in (Phillips and Kheifets, 2021, hereafter, PK). We thank the Editor, Serena Ng, and reviewers for helpful comments on earlier versions of the paper. Phillips acknowledges research support from the National Science Foundation under Grant No. SES 18-50860, a Kelly Fellowship at the University of Auckland, and an LKC Fellowship at Singapore Management University. Kheifets acknowledges financial support from the Spanish Ministerio de Ciencia, Innovación y Universidades (Grant No. ECO2017-86009-P), and the Jubiläumsfonds of the Oesterreichische Nationalbank (Grant No. 18692).

* Corresponding author at: Yale University, United States of America.
E-mail address: peter.phillips@yale.edu (P.C.B. Phillips).

https://doi.org/10.1016/j.jeconom.2023.105622
Received 13 May 2022; Received in revised form 9 November 2023; Accepted 10 November 2023
Available online 7 December 2023
0304-4076/

equilibrium errors in an $I(1)$ cointegrating relation may accumulate so that they cointegrate with the original variables. Engsted and Haldrup (1999) remark that this phenomenon is likely to occur in practice when characterizing the dynamic interactions of stock and flow variables. Granger and Lee (1990) and Lee (1996) showed how multicointegration can arise in the context of optimum control problems and infinite horizon quadratic adjustment cost models. One of the latest empirical applications of multicointegration has been to global climate change modeling, where the effects of accumulating cointegration disequilibria in temperature and surface radiation raise oceanic heat storage which leads to a multicointegrated linkage influencing global temperature (Bruns et al., 2020).

In these models the equilibrium errors (or residuals in an $I(1)$ cointegrating relation) are considered $I(0)$ or stationary, so that upon cumulation these errors become $I(1)$, and then subsequent cointegration may occur with the original variables or partial sums of them. Somewhat naturally it has therefore been posited in the multicointegration literature that multicointegration cannot occur in $I(1)$ systems (Engsted and Johansen, 1999; Engsted and Haldrup, 1999). This idea has been well accepted in the literature but was developed in a VAR framework and does not necessarily hold in general models such as semiparametric $I(1)$ cointegrating settings (Phillips, 1991). Instead, as demonstrated here and in recent related work on fully modified least squares (Kheifets and Phillips, 2023, hereafter, KP), multicointegration occurs in any cointegrated $I(1)$ model whenever there is a rank deficiency in the long run conditional covariance matrix of the cointegrating equation error. The phenomenon is a general one and rank deficiency turns out to be the determining factor of multicointegration in an $I(1)$ system. Multicointegration arises because singularity in the long run conditional covariance matrix induces a further long run cointegrating relation simply because the singularity implies a moving average $I(-1)$ (or higher level) component in that direction in the equation error, which leads directly to cointegration upon accumulation. This formulation of multicointegration in terms of rank deficiency in the long run conditional error covariance matrix is intuitive because it points directly to latent higher order relations in the $I(1)$ system and indicates their direction without further complications or the use of additional notation. The phenomenon has an analogue reduced rank structure in the parametric VAR model context and was noted but not analyzed by Engsted and Haldrup (1999, p.241).

Johansen (1992, 1995) provided explicit representations of reduced rank VAR structures which ensure the existence of cointegrated $I(1)$ and $I(2)$ systems. The implications of these conditions for characterizing models with multicointegration were employed in Engsted and Johansen (1999, hereafter, EJ), demonstrating the relevance of the $I(2)$ system for embodying multicointegrated structures in VAR systems. Outside the VAR setting, multicointegration can exist in an $I(1)$ reduced rank VARMA setting or in $I(1)$ cointegrated systems with infinite order bidirectional lags.[1] These models and approaches to multicointegration are reconciled in what follows.

The present paper makes three main contributions. First, a general analysis of multicointegration is provided within an $I(1)$ cointegrated system using the semiparametric triangular model framework. Multicointegration in such systems arises from singularity in the long run error covariance matrix, which in turn is shown to affect the asymptotic behavior of standard cointegrated system estimation procedures by introducing bias and non-pivotal inference. These findings are illustrated in PK (2021), Section 4.1.1, in the case of the integrated modified least squares (IM-OLS) approach (Vogelsang and Wagner, 2014). Similarly, KP (2023) recently developed asymptotics for the fully modified least squares (FM-OLS) cointegration coefficient estimator under multicointegration, showing degenerate limit theory in general but accelerated convergence over the usual $O(n)$ rate in the direction of multicointegration, accompanied by second order bias in the limit theory.[2] In a second contribution, it is shown that an extended version of high-dimensional trend IV (TIV) estimation with deterministic orthonormal instruments (Phillips, 2006, 2014) provides an approach to optimal estimation with mixed normal limit theory and pivotal inference in singular multicointegrated systems as well as standard cointegrated $I(1)$ systems. This TIV method therefore delivers a convenient IV approach to estimation and inference in $I(1)$ cointegrated systems that also applies under multicointegration. The methodology and asymptotic results are linked to maximum likelihood estimation in the specialized case of multicointegrated VAR systems studied in earlier work.

Third, high-dimensional TIV estimation has the further advantage of a faster convergence rate under multicointegration than the FM-OLS estimator studied in KP (2023) and this approach enables inference using standard Wald statistics formulated in the same way with a HAR variance matrix under both cointegration and multicointegration. This property opens the door to the analysis of cointegrated time series without full specification of the dynamics and without pretesting for multicointegration, thereby extending earlier work on VAR methods.

A final contribution of the paper is technical, with a group of new findings concerning the limit theory of functionals of trend transformed stationary and nonstationary variables in the case of asymptotically infinite instrument numbers. This contribution includes some new methods of developing limit theory for estimators and Wald statistics in highly complex cases involving singularities in signal matrices and partitioned regression asymptotics that require component-wise analysis or matrix normalization rather than diagonal matrix normalization to extract the correct asymptotics. These methods and results are of independent interest given recent research on large instrument numbers and deterministic transforms of variables that enable empirical investigations to focus on long run behavior.

The paper is organized as follows. The next section explains the source of multicointegration in the standard semiparametric triangular cointegrated system of $I(1)$ variables. Section 3 presents and analyzes TIV approaches to the estimation of cointegrated systems under conditions of multicointegration. Limit theory for these approaches is provided. Section 4 develops methods of inference using HAR methods that lead to pivotal asymptotics suited for inference in practical work. Section 5 shows optimality

[1] Estimation of $I(1)$ and $I(2)$ VAR processes where the lag order increases with the sample size is discussed in Li and Bauer (2020). Such a VAR can also be used to estimate invertible VARMA process.

[2] Other popular cointegration estimators such as dynamic OLS (Phillips and Loretan, 1991; Saikkonen, 1991; Stock and Watson, 1993) also generally suffer from second order asymptotic bias in the presence of multicointegration.

properties of the TIV estimation. Section 6 reports some simulation results and an empirical illustration is given in Section 7. Section 8 concludes with some general discussion and future research possibilities. Proofs of all the results in the paper are provided in a supplementary document for online publication (Phillips and Kheifets, 2023).[3] Some of the proofs in this Online Supplement involve new methods and technical calculations relevant to high-dimensional IV methods that are of wider interest. As an aid to readers, a glossary of notation is given in Section 9 at the end of the paper for the most common functionals that appear in the formulae of the paper.

2. Multicointegration in the $I(1)$ framework

The starting point in developing a framework for the source of multicointegration is the following $I(1)$ triangular matrix system of cointegration (Phillips, 1991)

$$y_t = Ax_t + u_{0t} \tag{1}$$

$$x_t = x_{t-1} + u_{xt}, \quad t = 1, \dots, T. \tag{2}$$

Here A is an $m_0 \times m_x$ cointegrating coefficient matrix, the $I(1)$ m_x-vector x_t is initialized at $t = 0$ by $x_0 = O_p(1)$, and the composite error vector $u_t = (u_{0t}', u_{xt}')'$ is assumed throughout the paper to follow the linear process

$$u_t = D(L)\eta_t = \sum_{j=0}^{\infty} D_j \eta_{t-j}, \quad \text{with } \sum_{j=0}^{\infty} j\|D_j\| < \infty, \quad \eta_t \sim iid(0, I_m), \tag{3}$$

where $m = m_0 + m_x$. Let $\Gamma_h = \mathbb{E}u_t u_{t+h}'$ and $\mathbb{V}^{\mathrm{LR}}(u_t) = \sum_{h=-\infty}^{\infty} \Gamma_h$ denote the long run variance matrix of u_t. The linear operator $D(L)$ and long run variance matrix $\mathbb{V}^{LR}(u_t) = \Omega = \sum_{h=-\infty}^{\infty} \Gamma_h = D(1)D(1)' = \sum_{j,k=0}^{\infty} D_j D_k'$ of u_t are partitioned conformably with u_t as

$$D(L) = \begin{bmatrix} D_{00}(L) & D_{0x}(L) \\ D_{x0}(L) & D_{xx}(L) \end{bmatrix}, \ \Omega = \begin{bmatrix} \Omega_{00} & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx} \end{bmatrix} \tag{4}$$

where $\Omega_{xx} > 0$ is positive definite, ensuring that x_t is a full rank unit root vector process which delivers m_x common stochastic trends to the $I(1)$ system (1)–(2). This full rank condition is maintained throughout the paper. The conditional long run variance matrix $\Omega_{00.x} = \Omega_{00} - \Omega_{0x}\Omega_{xx}^{-1}\Omega_{x0}$ is the Schur complement of the block Ω_{xx} in Ω and this matrix is positive (semi-) definite if and only if Ω is positive (semi-) definite by virtue of the Guttman rank additivity formula $\mathrm{rank}\,(\Omega) = \mathrm{rank}(\Omega_{xx}) + \mathrm{rank}\,(\Omega_{00.x})$.

The case of nonsingular Ω is well studied. The case where Ω_{xx} may be singular and the regressors x_t not full rank $I(1)$ processes was studied in Phillips (1995). But the situation where the conditional long run variance matrix $\Omega_{00.x}$ is singular seems largely to have been ignored[4] in the now vast literature on cointegration and, with the exception of KP (2023), none of the implications of singularity for estimation and inference have been explored in the (1)–(2) setting. This neglect is partly because, as we will show, singularity in the long run error covariance matrix leads to an $I(1)$ reduced rank VARMA representation rather than a reduced rank $I(1)$ VAR representation. So while such systems fall naturally within the semiparametric framework[5] above, they do not fall so neatly within the VAR framework, at least without raising the order of the system to $I(2)$. Nonetheless, the singular long run variance matrix case is especially interesting because it leads directly to a situation where partial sums of the observed variables y_t and x_t (which then become $I(2)$ variables) are cointegrated with x_t in some unknown direction — see (6) below. The importance of this situation is that it provides a primitive (that is, within the $I(1)$ system) link to the phenomenon of multicointegration, as envisaged in special cases by Granger and Lee (1989). But the source of the multicointegration is now firmly evident in the $I(1)$ framework (1)–(2). Moreover, the condition for multicointegration is straightforwardly expressed in terms of the existing parameters of the $I(1)$ system without further notation or complications.

An alternate representation of (1) which is useful in the development of efficient estimation methods of $I(1)$ cointegrated systems by FM-OLS or trend IV regression (Phillips, 2014) is the augmented regression

$$\begin{aligned} y_t &= Ax_t + \Omega_{0x}\Omega_{xx}^{-1}\Delta x_t + u_{0.xt}, \qquad u_{0.xt} := u_{0t} - \Omega_{0x}\Omega_{xx}^{-1}u_{xt} \\ &=: Ax_t + F\Delta x_t + u_{0.xt}, \quad \Delta x_t = u_{xt}, \end{aligned} \tag{5}$$

where both the cointegrating coefficient matrix A and the nonparametric long run regression coefficient $F = \Omega_{0x}\Omega_{xx}^{-1}$ are treated as unknown. The matrix F measures the long run endogeneity of the regressor x_t. Applying partial sum operations to (5) gives

$$Y_t = AX_t + F\left(x_t - x_0\right) + U_{0.xt}, \tag{6}$$

[3] Further material and discussions are available in the original working paper (Phillips and Kheifets, 2021).

[4] The possibility of full system singularity with rank failure in Ω is mentioned by Engsted and Haldrup (1999, p.241) but is not analyzed. Singularity in the long run conditional variance matrix and the implications of this singularity on estimation procedures such as fully modified least squares (FM-OLS) were the subject of a Yale Take Home Examination in 2011 (http://korora.econ.yale.edu/phillips/teach/ex/553a-ex11a.pdf). That approach was analyzed in KP (2023).

[5] The system is semiparametric because the nonparametric formulation of the system innovations in (3) generalizes the specific dynamics assumed to apply in a VAR system of cointegration or multicointegration.

with $Y_t = \sum_{s=1}^{t} y_s$, $X_t = \sum_{s=1}^{t} x_s$, and $U_{0.xt} = \sum_{s=1}^{t} u_{0.xs}$. Now suppose that the long run (conditional) variance matrix $\Omega_{00.x}$ of $u_{0.xt}$ is singular of rank $0 < p < m_0$ and H is an $m_0 \times (m_0 - p)$ matrix of full rank $m_0 - p$ spanning the null space of $\Omega_{00.x}$, so that

$$H' \Omega_{00.x} H = 0. \tag{7}$$

Then in this direction the transformed error $H' u_{0.xt}$ has zero long run variance matrix and zero spectral density matrix at the origin. There therefore exists some p dimensional $I(0)$ process ε_{Ht} for which $H' u_{0.xt} = \Delta \varepsilon_{Ht}$ *a.s.*, in the absence of fractional antipersistence[6] which is ruled out by the absolute 1-summability condition (3), leading to the representation

$$H' y_t = H' A x_t + H' F \Delta x_t + \Delta \varepsilon_{Ht},$$

and by partial summation to

$$H' Y_t = H' A X_t + H' F \left(x_t - x_0\right) + \left(\varepsilon_{Ht} - \varepsilon_{H0}\right).$$

It follows that

$$H' Y_t = H' A X_t + H' F x_t + \left(\varepsilon_{Ht} - \varepsilon_{H0} - H' F x_0\right) =: H' A X_t + H' F x_t + \eta_{Ht}, \tag{8}$$

where $\eta_{Ht} = \varepsilon_{Ht} - \varepsilon_{H0} - H' F x_0$ is $I(0)$ up to (and conditional on) the initial condition $x_0 = O_p(1)$, and provided no further level of long run degeneracy (or higher order multicointegration) is present for which $\mathbb{V}^{LR}(\varepsilon_{Ht}) = 0$. From (8) it follows that the variables $\left(Y_t, X_t, x_t\right)$ are cointegrated, involving both the $I(2)$ time series $\left(Y_t, X_t\right)$ and the $I(1)$ time series x_t. This accords with the conventional definition of multicointegration. Importantly, in this general framework the multicointegration coefficients, notably H and $H' F = H' \Omega_{0x} \Omega_{xx}^{-1}$, are nonparametric.

3. Estimation

With the exception of certain specialized models involving known relationships between variables such as stocks and flows, the existence of multicointegration will often not be anticipated in practical applied work on estimation and inference in $I(1)$ cointegrated systems. Tests for the presence of multicointegration have been developed for VAR systems (Engsted et al., 1997) but multicointegration may not be suspected, pre-test analyses may not be conducted or they may lead to pre-test bias and misleading outcomes; and empirical research may be conducted using triangular cointegrated systems rather than VAR specifications. In the absence of such tests it is obviously useful to have methods of estimating $I(1)$ cointegrated systems that are robust to the presence of multicointegration.[7]

Since semiparametric formulations of cointegrated $I(1)$ systems may be conducted in the presence of multicointegration, standard efficient methods of estimating these systems such as FM-OLS or dynamic OLS may continue to be employed in practical work. But the properties of such regressions are influenced by the singularity of the long run error covariance matrix. The typical impact of singularity is to raise the rate of convergence in the direction of singularity, thereby producing a degenerate limit theory for the estimate of the full cointegrating matrix. Moreover, common semiparametric methods of estimation such as FM-OLS involve the use of nonparametric kernel estimates of the long run variance and covariance matrices for bias correction and inference. In consequence, the accelerated rate of convergence in FM-OLS estimation is affected by the asymptotic behavior of these kernel estimates under rank degeneracy, as in the analysis of regression with cointegrated regressors and unrestricted VAR regression in the presence of cointegration (Phillips, 1995). Inference is correspondingly affected with further nonstandard limit distribution complications and non-pivotal limit theory in test statistics. These consequences may be analyzed[8] but are not pursued here. Instead, for reasons explained below the present paper develops new methods of estimation based on trend instrumental variable (TIV) methods that have clear advantages for efficient estimation and robust inference. To keep the analysis as brief as possible we confine attention to a scalar cointegrating relationship, which enables a convenient introduction of the basic ideas, highlights the main implications, and covers one of the most common cases arising in practice. Extension to the multivariate model follows usual lines but inferential analysis using Wald statistics is further complicated[9] by the need for matrix normalization to take account of differing rates of convergence in differing directions and arbitrary linear combinations of the matrix coefficients under test.

---

[6] Under fractional antipersistence in which ε_{Ht} is such that $H' u_{0.xt} = \Delta^d \varepsilon_{Ht} = (1-L)^d \varepsilon_{Ht}$ for some $d \in (0,1)$, the system (8) would be replaced by the equation $H'(1-L)^{-d} y_t = H' A (1-L)^{-d} x_t + H' F (1-L)^{-d} u_{xt} + \varepsilon_{Ht}$. Both the matrix transform H and the antipersistence parameter d would be unknown in this case, leading to further complications that are left for future research.

[7] Of course, the methods we introduce may result in misleading outcomes if the triangular system itself is misspecified.

[8] KP (2023) developed the limit theory of FM-OLS estimation under rank failure of the long run conditional variance matrix of the error in the augmented regression equation. Those results illustrate some of the effects that apply for other methods of cointegration system estimation.

[9] These complications are by no means trivial. Often in such cases, simplifying assumptions are made to assure no loss of rank or degrees of freedom in the limit, e.g., Andrews and Cheng (2012, 2014), and Vogelsang and Wagner (2014). Analyses of Wald statistic testing under matrix normalization without such prior requirements are developed in ongoing work in Magdalinos and Phillips (2019) for general nonstationary regression cases and for time polynomial trend regression in Phillips (2022).

3.1. Estimation approaches

To fix ideas, consider the following scalar version of the augmented $I(1)$ cointegrating Eq. (5)

$$y_t = a' x_t + f' \Delta x_t + u_{0.xt}, \quad \Delta x_t = u_{xt}, \; u_{0.xt} = u_{0t} - \Omega_{0x}\Omega_{xx}^{-1} u_{xt} \tag{9}$$

where $f' = \Omega_{0x}\Omega_{xx}^{-1}$ and the conditional long run variance $\Omega_{00.x} = \Omega_{00} - \Omega_{0x}\Omega_{xx}^{-1}\Omega_{x0} \geq 0$. We will consider both the standard form of the equation where $\Omega_{00.x} > 0$ and the singular form where $\Omega_{00.x} = \Omega_{00} - \Omega_{0x}\Omega_{xx}^{-1}\Omega_{x0} = 0$. In that event, we write $u_{0.xt} = \Delta e_t$ where e_t has variance σ_e^2 and long run variance $\omega_{ee} > 0$. The condition $\omega_{ee} > 0$ is maintained in the rest of the paper. Its relaxation leads to further complications involving higher order singularity that may be dealt with using similar methods to those developed here. In what follows, we consider two methods of estimation of the parameters in (9).

We start by requiring the following high-level conditions, which hold under well-known conditions (e.g., Phillips and Solo (1992)). Here and in what follows we use $\rightsquigarrow$ to signify weak convergence in the relevant probability space.

$$(a) \quad \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor n\cdot \rfloor} e_t \rightsquigarrow B_e(\cdot) \equiv BM\left(\omega_e^2\right), \quad \text{when } \Omega_{00.x} = 0, \tag{10}$$

$$(b) \quad \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor n\cdot \rfloor} u_{0.xt} \rightsquigarrow B_{0.x}(\cdot) = BM\left(\Omega_{00.x}\right), \quad \text{when } \Omega_{00.x} > 0. \tag{11}$$

In case (a) we further assume the joint functional law

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor n\cdot \rfloor} \left(e_t, u_{xt}'\right)' \rightsquigarrow \left(B_e(\cdot), B_x(\cdot)'\right)' \equiv BM\left(\begin{bmatrix} \omega_{ee} & \omega_{ex} \\ \omega_{xe} & \Omega_{xx} \end{bmatrix}\right) \text{ with } \begin{bmatrix} \omega_{ee} & \omega_{ex} \\ \omega_{xe} & \Omega_{xx} \end{bmatrix} > 0. \tag{12}$$

The functional law (11) already holds under (3), and (12) similarly holds under analogous linear process conditions, as in Phillips and Solo (1992). Although $u_{0.xt} = \Delta e_t$ has zero long run covariance with u_{xt} in case (a) the same is not necessarily so of e_t. For instance, if $e_t = \alpha' u_{xt} + \varepsilon_t$ where $\varepsilon_t \sim iid\left(0, \sigma_\varepsilon^2\right)$ and independent of u_{xt}, then $u_{0.xt} = \alpha' \Delta u_{xt} + \Delta \varepsilon_t$ has zero long run covariance with u_{xt} but the long run covariance of u_{xt} and e_t is $\mathbb{CV}^{\mathrm{LR}}\left(u_{xt}, e_t\right) = \omega_{xe} = \Omega_{xx}\alpha \neq 0$. Condition (12) allows for both $\omega_{xe} = 0$ and $\omega_{xe} \neq 0$ possibilities.

Using capitals as before to denote partial summation, write $Y_t = \sum_{s=1}^t y_s$, $X_t = \sum_{s=1}^t x_s$, and $U_{0.xt} = \sum_{s=1}^t u_{0.xs}$. The transformed system (9), up to initial conditions (in particular, taking $e_0 = 0$), is then

$$Y_t = a' X_t + f' x_t + e_t^+ \tag{13}$$

$$e_t^+ = e_t \mathbf{1}\left\{\Omega_{00.x} = 0\right\} + U_{0.xt} \mathbf{1}\left\{\Omega_{00.x} > 0\right\}, \tag{14}$$

a formulation that covers both singular and non-singular cases.

3.1.1. Trend instrumental variable estimation

The approach developed here is based on the trend IV (TIV) method of Phillips (2014)[10] which employs orthonormal (ON) deterministic trend functions as instrumental variables for the regressors in (9). These ON instruments are designed to transform the system so that its long run properties are brought into primary focus both for regression estimation (Phillips, 1998) and for long run variance matrix estimation (Phillips, 2005b; Müller, 2007). These methods have recently become popular in examining various properties of long run relations among time series variables (e.g., Phillips (2005a), Müller and Watson (2018) and Hwang and Sun (2018)) and have numerous empirical applications as revealed in these studies.

In the TIV method of estimating cointegrating equations such as (9), deterministic instrumental variables $\left\{\varphi_k\left(\frac{t}{n}\right)\right\}_{k=1}^K$ are employed, where $\left\{\varphi_k(r)\right\}_{k=1}^\infty$ are orthonormal basis functions of $L_2[0,1]$ and K is allowed to pass to infinity as $n \to \infty$. This approach is high-dimensional TIV estimation. An alternate version of this method is based on a fixed number K of orthonormal trend instruments and is used in recent work by Hwang and Sun (2018). We call this method the fixed-K approach of TIV regression. Various classes of orthonormal functions may be used in these regressions without materially affecting the limit theory or finite sample performance, as demonstrated in Phillips (2014) and Hwang and Sun (2018). The latter paper shows a particular advantage in terms of F and t distribution limit theory for conventional test statistics of coefficient restrictions, which can enhance inference in finite samples in standard cointegrated systems. This advantage has received wider attention recently (Lazarus et al., 2018). But as shown later in the current work, the fixed-K approach does not deal as effectively with multicointegrated systems.

In what follows, we let $\tilde{\varphi}_K(r) = \left(\varphi_1(r), \ldots, \varphi_K(r)\right)'$, and $\boldsymbol{\Phi}_K' = \left[\tilde{\varphi}_{K1}, \ldots, \tilde{\varphi}_{Kn}\right]$ where $\tilde{\varphi}_{Kt} = \tilde{\varphi}_K\left(\frac{t}{n}\right) = \left[\varphi_1\left(\frac{t}{n}\right), \ldots, \varphi_K\left(\frac{t}{n}\right)\right]'$. The projector matrix onto the space of the instruments is $P_{\Phi_K} = \boldsymbol{\Phi}_K\left(\boldsymbol{\Phi}_K'\boldsymbol{\Phi}_K\right)^{-1}\boldsymbol{\Phi}_K'$. For trigonometric orthonormal polynomials we

[10] The TIV approach was originally proposed in a York University Workshop conference presentation given in 2003. The same paper was presented in the Faro Time Series Econometrics Conference 2005 and distributed as a Cowles Foundation Discussion Paper (Phillips, 2006). That paper also introduced the concept of a trend likelihood associated with the low frequency components of a time series obtained by fitted regression on a number of deterministic orthonormal regressors. Phillips (2005b) introduced the related idea of trend coordinates based on these fitted regression components to study long run covariability among trending time series, a subject that has been extensively investigated recently by Müller and Watson (2018). The approach has earlier origins in band spectral regression (Hannan, 1963; Engle, 1974; Corbae et al., 2002) in the frequency domain.

have $n^{-1}\Phi_K'\Phi_K = I_K + O\left(\frac{1}{n}\right)$, as shown in Phillips (2005b, Lemma A) when $\varphi_k(r) = \sqrt{2}\sin\left\{\left(k-\frac{1}{2}\right)\pi r\right\}$, so that $P_{\Phi_K} \sim n^{-1}\Phi_K\Phi_K'$. TIV estimation of (9) is then asymptotically equivalent to simple least squares regression on the linearly transformed K-dimensional system

$$V_y = V_x a + V_{\Delta x} f + V_{u_{0.x}}, \tag{15}$$

where we use the general notation $V_c = \Phi_K' c = \sum_{t=1}^n \tilde{\varphi}_{Kt} c_t'$ for the trigonometric transform of a time series c_t. The resulting coefficient estimates of (15) have the following form in standard partitioned regression notation

$$\hat{a}_{TIV} - a = \left(V_x' Q_{V_{\Delta x}} V_x\right)^{-1} V_x' Q_{V_{\Delta x}} V_{u_{0.x}}, \tag{16}$$

$$\hat{f}_{TIV} - f = \left(V_{\Delta x}' Q_{V_x} V_{\Delta x}\right)^{-1} V_{\Delta x}' Q_{V_x} V_{u_{0.x}}, \tag{17}$$

where we use the usual general notation[11] $Q_x = I - P_x, P_x = x\left(x'x\right)^{-1}x'$. This least squares procedure is called transformed augmented least squares (TA-OLS) in Hwang and Sun (2018), who investigate its asymptotic properties when $\Omega_{00.x} > 0$ and K is fixed as $n \to \infty$. It is asymptotically equivalent to fixed-K TIV.

The approach we suggest here is designed to robustify the TIV estimation procedure to the presence of multicointegration and singularity. The idea is to apply TIV regression to the following augmented regression form of (13)

$$Y_t = a'X_t + f'x_t + g'\Delta x_t + e_t^+ = a'X_t + f'x_t + g'u_{xt} + e_t^+, \tag{18}$$

where the additional (redundant) regressor Δx_t is included with coefficient $g = 0$ and the regression error is $e_t^+ = e_t \mathbf{1}\left\{\Omega_{00.x} = 0\right\} + U_{0.xt}\mathbf{1}\left\{\Omega_{00.x} > 0\right\}$ as before. Thus, this time aggregated version of the model is augmented by the inclusion of the additional regressor Δx_t, analogous to the original system (5). As before, in practical work it is useful to include a fitted intercept in (18), which is innocuous but takes care of initialization effects in the singular case where $e_t^+ = e_t\mathbf{1}\left\{\Omega_{00.x} = 0\right\}$ and $e_0 \neq 0$. Again, the limit theory is simply adjusted to employ deviations from means for the relevant stochastic processes, which for ease of notation is not done here.

In observation form, we write (18) as

$$Y = \left[X, C_x\right]\gamma + e^+, \text{ with } \gamma' = (a', f', g') =: \left(a', \ell'\right) \tag{19}$$

and

$$C_x' = \left[c_{x1}, \ldots, c_{xn}\right] = \begin{bmatrix} x_1 & \cdots & x_n \\ u_{x1} & \cdots & u_{xn} \end{bmatrix} =: \begin{bmatrix} x' \\ u_x' \end{bmatrix}.$$

Eq. (18) may, of course, also be estimated by direct application of least squares, leading to a form of augmented IM-OLS regression. This estimator, as well as IM-OLS, have nuisance parameter dependencies in the limit (PK (2021)). Therefore, we proceed with the analysis of TIV estimation of the augmented system. The TIV estimator of a in (18) has the form

$$\hat{a}_{TIV} = \arg\min_a (Y - Xa)' R_K (Y - Xa) = \left(X'R_K X\right)^{-1}\left(X'R_K Y\right), \tag{20}$$

where the projector matrix is $R_K = P_{\Phi_K} - P_{\Phi_K} C_x \left(C_x' P_{\Phi_K} C_x\right)^{-1} C_x' P_{\Phi_K}$ and $Y' = \left[Y_1, \ldots, Y_n\right]$. The TIV estimation procedure projects all the variables in the augmented system (18) onto the deterministic instruments using the projector P_{Φ_K}. For fixed K, this approach is, as above in (15), asymptotically equivalent to least squares regression on the transformed system

$$V_Y = V_X a + V_x f + V_{\Delta x} g + V_{e^+} =: V_X a + V_C \ell + V_{e^+}, \tag{21}$$

where we employ the notation $V_Z = \Phi_K' Z$ for an observation matrix Z. Standard partitioned least squares regression on (21) leads to the following estimator of a

$$\hat{a}_{fTIV} - a = \left(V_X' Q_{V_C} V_X\right)^{-1} V_X' Q_{V_C} V_{e^+}, \tag{22}$$

giving the fixed-K trend IV (fTIV) estimate.

The results that follow provide the asymptotic theory for TIV estimation with fixed-K and as $K \to \infty$ in both $\Omega_{00.x} > 0$ and $\Omega_{00.x} = 0$ cases. The proofs involve new complications due to the presence of the redundant regressor Δx_t in the fitted equation, the partially spurious nature of the regression equation when $\Omega_{00.x} > 0$, and the impact of singularity when $\Omega_{00.x} = 0$.

New asymptotic theory is provided to address these complications. The analysis is particularly difficult when $K \to \infty$ as $n \to \infty$ and the development of the asymptotic theory of inference in the following section involves new methods and results. But the final limit theory is satisfyingly simple for both the singular and nonsingular $\Omega_{00.x}$ cases. The result for fixed-K TIV estimation is given in Theorem 1. The main result is given in Theorem 2 for TIV estimation when $K \to \infty$. This approach leads to mixed normal limit distribution theory in both $\Omega_{00.x} > 0$ and $\Omega_{00.x} = 0$ cases, therefore providing a basis for robust estimation and inference in cointegrated/multicointegrated systems even when the presence of multicointegration is unknown a priori. For convenient reading of the following theorems and subsequent development, readers may be aided by consulting the notation glossary in Section 9.

[11] For convenience, a list of notation is provided in the glossary, Section 9.

Theorem 1 (*TIV Estimation with Fixed* K). *When* $\Omega_{00.x} > 0$, (11) *holds,* K *is fixed, and* $n \to \infty$

(i) $n\left(\hat{a}_{fTIV} - a\right) \rightsquigarrow S_K' \Psi_{0.xK} \equiv \mathcal{MN}\left(0, \Omega_{00.x} S_K' \left(\int_0^1 \int_0^1 (r \wedge s)\, \tilde{\varphi}_K(r)\, \tilde{\varphi}_K(s)'\, dr ds\right) S_K\right)$,

where $S_K = J_K \mu_K \left(\mu_K' J_K \mu_K\right)^{-1}$, $\mu_K = \int_0^1 \tilde{\varphi}_K B_X'$, $\Psi_{0.xK} = \int_0^1 \tilde{\varphi}_K B_{0.x}$, $J_K = Q_{\xi_K} - Q_{\xi_K} \eta_K \left(\eta_K' Q_{\xi_K} \eta_K\right)^{-1} \eta_K' Q_{\xi_K}$, $\eta_K = \int_0^1 \tilde{\varphi}_K(r) B_x(r)'\, dr$, $\xi_K = \int_0^1 \tilde{\varphi}_K(r)\, dB_x(r)'$, $B_X(r) = \int_0^r B_x$, *and* $\tilde{\varphi}_K$ *is defined before* (15).

When $\Omega_{00.x} = 0$, (12) *holds,* K *is fixed, and* $n \to \infty$

(ii) $n^2\left(\hat{a}_{fTIV} - a\right) \rightsquigarrow S_K' \psi_{eK}$,

where $\psi_{eK} = \int_0^1 \tilde{\varphi}_K\, dB_e$. *When* $\omega_{ex} = 0$, *the limit distribution is mixed normal and*

(ii)* $n^2\left(\hat{a}_{fTIV} - a\right) \rightsquigarrow \mathcal{MN}\left(0, \omega_{ee}\left(\mu_K' R_K \mu_K\right)^{-1}\right)$.

Theorem 2 (*TIV Estimation with* $K \to \infty$). *When* $\Omega_{00.x} > 0$, (11) *holds, and* $(K, n) \to \infty$ *with* $K = o\left(n^{4/5-\delta}\right)$ *for some* $\delta > 0$

(iii) $n\left(\hat{a}_{TIV} - a\right) \rightsquigarrow \mathcal{A}_{X.x}^{-1}\left(\int_0^1 B_{X.x} B_{0.x}\right) = \mathcal{A}_{X.x}^{-1} \int_0^1 \overrightarrow{B_{X.x}}\, dB_{0.x} \equiv \mathcal{MN}\left(0, \Omega_{00.x} \mathcal{A}_{X.x}^{-1} \int_0^1 \overrightarrow{B_{X.x}}\, \overrightarrow{B_{X.x}}' \mathcal{A}_{X.x}^{-1}\right)$, *where* $\mathcal{A}_{X.x} = \int_0^1 B_{X.x} B_{X.x}'$, $\overrightarrow{B_{X.x}}(r) = \int_r^1 B_{X.x}$, $B_{X.x}(r) = B_X(r) - \int_0^1 B_X B_x' \left(\int_0^1 B_x B_x'\right)^{-1} B_x(r)$

When $\Omega_{00.x} = 0$, (12) *holds, and* $(K, n) \to \infty$ *with* $K = o\left(n^{4/5-\delta}\right)$ *for some* $\delta > 0$

(iv) $n^2\left(\hat{a}_{TIV} - a\right) \rightsquigarrow \mathcal{A}_{X.x}^{-1}\left(\int_0^1 B_{X.x}\, dB_{e.x}\right) \equiv \mathcal{MN}\left(0, \omega_{ee.x} \mathcal{A}_{X.x}^{-1}\right)$,

where $B_{e.x}(r) = B_e(r) - \omega_{ex} \Omega_{xx}^{-1} B_x(r) \equiv BM\left(\omega_{ee.x}\right)$ *where* $B_{e.x}$ *is independent of the Brownian motion* B_x *and* $\omega_{ee.x} = \omega_{ee} - \omega_{ex} \Omega_{xx}^{-1} \omega_{xe}$.

As expected, in both Theorems 1 and 2 the limit distributions differ for the two cases $\Omega_{00.x} = 0$ and $\Omega_{00.x} > 0$. For Theorem 2 we employ the expansion rate condition on the instrument number $K = o\left(n^{4/5-\delta}\right)$ for some $\delta > 0$. The same condition was used in Phillips (2014) and facilitates the joint limit theory $(K, n) \to \infty$.

TIV regression has the usual $O(n)$ convergence rate for cointegrating regressions when $\Omega_{00.x} > 0$ in both fixed K and $K \to \infty$ cases. Just as in standard cointegrating regression theory with $\Omega_{00.x} > 0$ mixed normal limit theory applies, as it does for other methods of estimation such as FM-OLS regression in the nonsingular case; and when $K \to \infty$ as $n \to \infty$, Theorem 2(iii) shows that the mixed normal limit theory of TIV is the same as that of the IM-OLS estimator given in PK (2021), Theorem 1(i). OLS estimation may also be used in the augmented model (18) with the redundant regressor Δx_t and without the long run transforms but there are non-trivial effects on the estimated residuals from the use of OLS on the augmented system that make inference difficult with this approach.

The singular case with $\Omega_{00.x} = 0$ is more intriguing. First, rates of convergence rise to $O\left(n^2\right)$ as they do for IM-OLS. But the limit theory for TIV is much simpler because the long run transforms are effective in focusing attention on long run properties. Second, the TIV regression is successful in removing both endogeneity and serial correlation biases in both singular and nonsingular cases under joint convergence when $K \to \infty$ as $n \to \infty$. Third, the limit theory is mixed normal and conducive to pivotal inference in both cases, even though the rates of convergence are different for singular and nonsingular systems. Fourth, the mixed normal limit theory in (iv) may be written in standardized form as

$$\mathcal{MN}\left(0, \omega_{ee.x} \Omega_{xx}^{-1/2} \mathcal{A}_{W,X.x}^{-1} \Omega_{xx}^{-1/2}\right) \equiv \omega_{ee.x}^{-1/2} \Omega_{xx}^{-1/2} \times \mathcal{MN}\left(0, \mathcal{A}_{W,X.x}^{-1}\right), \tag{23}$$

with $\mathcal{A}_{W,X.x} = \int_0^1 W_{X.x} W_{X.x}'$, since by simple matrix scale manipulations we have the representation

$$\begin{aligned} B_{X.x}(r) &= B_X(r) - \int_0^1 B_X B_x' \left(\int_0^1 B_x B_x'\right)^{-1} B_x(r) \\ &= \Omega_{xx}^{1/2} \left\{ W_X(r) - \int_0^1 W_X W_x' \left(\int_0^1 W_x W_x'\right)^{-1} W_x(r) \right\} =: \Omega_{xx}^{1/2} W_{X.x}, \end{aligned} \tag{24}$$

where $B_x = \Omega_{xx}^{1/2} W_x$, $B_X(r) = \Omega_{xx}^{1/2} \int_0^r W_x$, and $W_x \equiv BM\left(I_{m_x}\right)$. The limit distribution (23) is then a matrix scaled form of a mixed normal distribution that depends only on functionals of standard Brownian motion. Importantly, the convergence rate of TIV regression is faster than that of FM-OLS in the multicointegrated case where the rate does not achieve $O(n^2)$ – see KP (2023).

Theorems 1 and 2 highlight differences in TIV estimation between the fixed K and high-dimensional $K \to \infty$ cases. For the fixed K case. TIV does not fully remove endogeneity bias as the limiting error transform $\psi_{eK} = \int_0^1 \tilde{\varphi}_K\, dB_e$ in the limit distribution (ii) remains correlated with the regressor variable limiting transforms $\left(\mu_K, \eta_K, \xi_K\right) = \left(\int_0^1 \tilde{\varphi}_K B_X', \int_{r=0}^1 \tilde{\varphi}_K B_x' dr, \int_0^1 \tilde{\varphi}_K\, dB_x\right)$ when the long run covariance $\omega_{ex} \neq 0$. But when $\omega_{ex} = 0$ and K is fixed the TIV estimator $\hat{a}_{TIV}$ does have mixed normal limit theory, given by

$$n^2\left(\hat{a}_{TIV} - a\right) \rightsquigarrow \mathcal{MN}\left(0, \omega_{ee}\left(\mu_K' R_K \mu_K\right)^{-1}\right), \tag{25}$$

which may be written in standardized Brownian motion form, analogous to (23) in this case. So under the long run orthogonality condition $\omega_{ex} = 0$ TIV estimation with fixed K instruments provides robust estimation and is effective in pivotal inference. But in the general case where the long run covariance $\mathbb{CV}^{\mathrm{LR}}\left(e_t, u_{xt}\right) = \omega_{ex} \neq 0$ and there is long run endogeneity in the singular multicointegrated model, the limit distribution in (iii) for the fixed K case is no longer mixed normal.

These results show the key advantage of high-dimensional trend IV regression on the augmented aggregated system (18). The limit theory of TIV regression is mixed normal in *both* non-singular and singular cases when $K \to \infty$ as $n \to \infty$. The method therefore provides a useful foundation for a robust approach to estimation and inference about cointegrating coefficients in both cointegrated and multicointegrated systems.

4. Inference

Theorems 1–2 show that TIV methods provide consistent and asymptotically mixed normal estimation procedures which might be expected to form a basis for inference in the standard $I(1)$ cointegrating regression model with nonsingular $\Omega_{00.x} > 0$. But when $\Omega_{00.x} > 0$ the augmented system (18) is a partially spurious regression, just like the original aggregated system (13) with $I(1)$ regressors and an $I(1)$ error. The spurious nature of this regression complicates inference and requires special methods to estimate the long run variance $\Omega_{00.x}$ in constructing Wald tests. Moreover, when $\Omega_{00.x} = 0$, IM-OLS suffers from asymptotic second order bias and limit theory that is unsuited to pivotal inference, thereby failing to resolve endogeneity and serial correlation bias problems in the limit (PK (2021), Theorem 1). In what follows we therefore concentrate on the TIV approach to testing.

More specifically, consider a Wald test of the linear hypothesis $\mathcal{H}_0 : Ha = h$ about the cointegrating vector a where H is $q \times m_x$ of rank q and h is a q-vector. Just as in estimation, the problem of inference is complicated by the fact that it is unknown a priori whether the system is singular or not in the absence of prior information or pre-testing. Robust inference therefore requires that the same approach be employed in both cases since $\Omega_{00.x}$ is, of course, unknown. For this purpose it is convenient to employ a sandwich form in estimating the covariance matrix metric for the Wald statistic in order to deal in a comprehensive way with the different types of temporal dependencies that arise in the nonsingular $\Omega_{00.x} > 0$ and singular $\Omega_{00.x} = 0$ cases. This matrix can be constructed in a general way by using the form of the TIV estimate $\hat{a}_{TIV}$. In view of (19) and (20), $\hat{a}_{TIV}$ satisfies

$$\hat{a}_{TIV} - a = \left(X' R_K X\right)^{-1} \left(X' R_K e^+\right) = G_K \Phi_K' e^+ = G_K \sum_{t=1}^{n} \tilde{\varphi}_K\left(\frac{t}{n}\right) e_t^+, \tag{26}$$

where $R_K = P_{\Phi_K} - P_{\Phi_K} C_x \left(C_x' P_{\Phi_K} C_x\right)^{-1} C_x' P_{\Phi_K}$, so that

$$H\left(\hat{a}_{TIV} - a\right) = H G_K \Phi_K' e^+ = H G_K \left(\sum_{t=1}^{n} \tilde{\varphi}_K\left(\frac{t}{n}\right) e_t^+\right),$$

in which the coefficient matrix G_K has the form

$$G_K = \left(X' R_K X\right)^{-1} \left\{ X' \Phi_K \left(\Phi_K' \Phi_K\right)^{-1} - X' P_{\Phi_K} C_x \left(C_x' P_{\Phi_K} C_x\right)^{-1} C_x' \Phi_K \left(\Phi_K' \Phi_K\right)^{-1} \right\}, \tag{27}$$

and $\Phi_K' e^+ = \sum_{t=1}^{n} \tilde{\varphi}_K\left(\frac{t}{n}\right) e_t^+$ is the transformed error vector in the model after projection on the instruments Φ_K. We may estimate the residuals e_t^+ from the fitted TIV regression giving

$$\begin{aligned} \hat{e}_t^+ &= Y_t - \hat{a}_{TIV}' X_t - \hat{f}_{TIV}' x_t - \hat{g}_{TIV}' \Delta x_t \\ &= e_t^+ - \left(\hat{a}_{TIV} - a\right)' X_t - \left(\hat{f}_{TIV} - f\right)' x_t - \left(\hat{g}_{TIV} - g\right)' u_{xt}, \end{aligned}$$

and construct the kernel estimates[12]

$$\hat{V}_{Kn} = \sum_{j=-M+1}^{M} k\left(\frac{j}{M}\right) \frac{1}{n} \sum_{1 \le t, t+j \le n}^{n} \tilde{\varphi}_K\left(\frac{t}{n}\right) \tilde{\varphi}_K\left(\frac{t+j}{n}\right)' \hat{e}_t^+ \hat{e}_{t+j}^+, \tag{28}$$

$$\hat{\omega}_{e^+}^2 = \sum_{j=-M+1}^{M} k\left(\frac{j}{M}\right) \frac{1}{n} \sum_{1 \le t, t+j \le n}^{n} \hat{e}_t^+ \hat{e}_{t+j}^+, \tag{29}$$

as if we were estimating a long run variance matrix of $\tilde{\varphi}_K\left(\frac{t}{n}\right) e_t^+$ and long run variance of e_t^+, thereby ignoring the spurious nature of the regression when $\Omega_{00.x} > 0$.

The lag kernel function $k(\cdot) : \mathbb{R} \to [0,1]$ used in (28) and (29) is assumed to be a symmetric, piecewise smooth density with $k(x) = 0$ for $|x| > 1$, and $\int_{-1}^{1} k(x)\, dx = 1$. In the case of standard HAC estimation, the lag truncation parameter M is assumed to satisfy $\frac{1}{M} + \frac{M}{n} \to 0$ as $n \to \infty$. In the case of HAR inference with a fixed-b setting leading to $M = bn$, we use the notation $k_b(x) = k(\frac{x}{b})$ and correspondingly define the HAR kernel estimator as

$$\hat{V}_{bKn} = \sum_{j=-n+1}^{n-1} k_b\left(\frac{j}{n}\right) \frac{1}{n} \sum_{1 \le t, t+j \le n}^{n} \tilde{\varphi}_K\left(\frac{t}{n}\right) \tilde{\varphi}_K\left(\frac{t+j}{n}\right)' \hat{e}_t^+ \hat{e}_{t+j}^+. \tag{30}$$

[12] We warn the reader not to confuse the kernel estimators of the variance $\hat{V}_{Kn}$ and $\hat{V}_{bKn}$ in this section and transformed variables V_Y, V_X and such in (21).

With these components we can construct the following HAC and HAR Wald statistics in conventional form as follows

$$\mathrm{Wald}_{TIV} = \left(H\hat{a}_{TIV} - h\right)' \left[HG_K\left(n\hat{V}_{Kn}\right)G_K'H'\right]^{-1}\left(H\hat{a}_{TIV} - h\right), \tag{31}$$

$$\mathrm{Wald}_{TIV,b} = \left(H\hat{a}_{TIV} - h\right)' \left[HG_K\left(n\hat{V}_{bKn}\right)G_K'H'\right]^{-1}\left(H\hat{a}_{TIV} - h\right). \tag{32}$$

The regression error is $e_t^+ = e_t\mathbf{1}\left\{\Omega_{00.x} = 0\right\} + U_{0.xt}\mathbf{1}\left\{\Omega_{00.x} > 0\right\}$. So the asymptotic properties of (28), (29) and therefore both Wald test statistics Wald_{TIV} and $\mathrm{Wald}_{TIV,b}$ depend on the asymptotic behavior of the residuals $\hat{e}_t^+$, the long run error variance estimate $\hat{\omega}^2_{e^+}$, and the long run variance matrix estimates $\hat{V}_{Kn}$ and $\hat{V}_{bKn}$ associated with the transformed error components $\tilde{\varphi}_K\left(\frac{t}{n}\right)e_t^+$.

Two forms of TIV inference can be considered, corresponding to fixed-K and $K\to\infty$ cases, just as in estimation. A disadvantage of the fixed-K approach is that the partially spurious nature of the fitted regression carries the inconsistencies of the estimates $\left(\hat{f}_{TIV}, \hat{g}_{TIV}\right)$ into the regression residuals $\hat{e}_t^+$ and their $I(1)$ character in the usual $\Omega_{00.x} > 0$ case. This leads to divergence of statistical tests as $n\to\infty$ under the null hypothesis, just as in standard spurious regression limit theory (Phillips, 1986). Even with the use of sandwich formulae and HAC estimators such as (28) the divergence rate of the Wald test for fixed K is $O_p\left(\frac{n}{M}\right)$, as shown in the proof of Theorem 3 below.[13] This divergence rate for the Wald test with a HAC covariance matrix estimate is the same as that obtained in Phillips (1998) for standard spurious regression inference with HAC error variance matrix estimators. Hence, use of fixed K inference with HAC variance estimation is not readily compatible with both singular and nonsingular cases and encounters difficulties similar to those arising in the use of IM-OLS and FM-OLS. In view of these drawbacks, we do not pursue the fixed-K TIV approach further in this context of potential singularity and multicointegration in $I(1)$ systems.

The use of HAR inference leads to very different limit theory that is much more useful in practical work. Importantly, fixed-b settings for the bandwidth parameter as in (30) with $M = bn$ and $b\in(0,1]$ control divergence, just as in other work on spurious regressions with HAR inference methods (Sun, 2004; Phillips et al., 2019). As usual, the HAR approach introduces nonstandard limit theory. But, as we see below, the limit theory is pivotal even for quite general linear hypothesis tests such as $\mathcal{H}_0 : Ha = h$.

Under HAR inference, a substantial degree of robustness in terms of asymptotic size control in testing is achieved. Importantly, this robustness covers both cointegration and multicointegration cases. The following results give the limit theory of the two test statistics Wald_{TIV} and $\mathrm{Wald}_{TIV,b}$ when $(K,n)\to\infty$ when $\Omega_{00.x} > 0$ and $\Omega_{00.x} = 0$.

Theorem 3 (*TIV Inference with $K\to\infty$*). *Under the assumptions of Theorem 2 and under the null hypothesis $\mathcal{H}_0 : Ha = h$, the following hold as $\frac{1}{M} + \frac{M}{n} \to 0$, $(K,n)\to\infty$ with $K = o\left(n^{4/5-\delta}\right)$ for some $\delta > 0$:*

When $\Omega_{00.x} > 0$:

(v) $\mathrm{Wald}_{TIV} = O_p\left(\frac{n}{M}\right)$, $\quad \mathrm{Wald}_{TIV,b} \rightsquigarrow \eta'_{\mathcal{E}_W} L\left\{LL'\right\}^{-1} L'\eta_{\mathcal{E}_W}$,

where $L = \mathcal{E}_W^{1/2}\Omega_{xx}^{-1/2}H$, and setting $k_b(\cdot) = k\left(\frac{\cdot}{b}\right)$,

$$\mathcal{E}_W := \mathcal{A}^{-1}_{W,X.x}\left(\int_0^1\int_0^1 k_b(r-p)\left(W_{X.x}(r)W_{X.x}(p)'\right)\widetilde{W_{0.x}}(r)\widetilde{W_{0.x}}(p)\,dr\,dp\right)\mathcal{A}^{-1}_{W,X.x},$$

$$\eta_{\mathcal{E}_W} := \mathcal{E}_W^{-1/2}\left(\int_0^1 W_{X.x}W_{X.x}'\right)^{-1}\int_0^1 \overrightarrow{W_{X.x}}\,dW_{0.x}.$$

where

$$W_{0.x}(r) = W_0(r) - \Omega_{0x}\Omega_{xx}^{-1}W_x(r),$$

$$\widetilde{W_{0.x}}(r) = W_{0.x}(r) - \int W_{0.x}W_{X.x}'\left(\int W_{X.x}W_{X.x}'\right)^{-1}W_X(r) - \int W_{0.x}W_{x.X}'\left(\int W_{x.X}W_{x.X}'\right)^{-1}W_x(r),$$

$$W_{x.X}(r) = W_x(r) - \int W_xW_X'\left(\int W_XW_X'\right)^{-1}W_X(r),$$

$$W_{X.x}(r) = W_X(r) - \int W_XW_x'\left(\int W_xW_x'\right)^{-1}W_x(r)$$

When $\Omega_{00.x} = 0$:

(vi) $\mathrm{Wald}_{TIV} \rightsquigarrow \chi^2_q$, $\quad \mathrm{Wald}_{TIV,b} \rightsquigarrow \eta'_{e.x}\mathcal{J}_q'\left\{\mathcal{J}_q\mathcal{F}_W\mathcal{J}_q'\right\}^{-1}\mathcal{J}_q\eta_{e.x}$,

where

$$\mathcal{F}_W := \mathcal{A}^{-1}_{W,X.x}\left(\int_0^1\int_0^1 k_b(r-p)W_{X.x}(p)W_{X.x}(r)'\,d\mathcal{Q}_W(p)\,d\mathcal{Q}_W(r)\right)\mathcal{A}^{-1}_{W,X.x},$$

$$\eta_{e.x} := \mathcal{A}^{-1}_{W,X.x}\left(\int_0^1 W_{X.x}dW_{e.x}\right),\ W_{e.x}(r) = \omega_{ee.x}^{-1/2}B_{e.x} = \omega_{ee.x}^{-1/2}\left(B_e - \omega_{ex}\Omega_{xx}^{-1}B_x\right),$$

[13] In the Online Supplement, see equations (51) and equation (52) in the proof of Theorem 3 for the residual inconsistency and (60) for the divergence rate of the Wald test of $O_p\left(\frac{n}{M}\right)$ when K is fixed.

$$\mathcal{A}_{W,X.x} = \int_0^1 W_{X.x} W'_{X.x}, \text{ and } \mathcal{J}_q = \left[I_q, 0\right].$$

$\mathcal{F}_W$ and $\eta_{e.x}$ depend only on the vector standard Brownian motions (W_x, W_X) and the standard Brownian motion $W_{e.x}$ which is independent of (W_x, W_X). The stochastic process $\mathcal{Q}_W(\cdot)$ is also a functional of these standard Brownian motions and is defined in (76) in the Online Supplement.

Remarks

(a) The first result of **(v)** shows that the HAC-based Wald statistic Wald_{TIV} diverges at rate $O_p\left(\frac{n}{M}\right)$, just as the squared t-statistic in Phillips (1998). So HAC variance matrices in the construction of the Wald statistic fail to resolve the partially spurious nature of the regression (18) and are therefore not recommended in the present context where there is potential multicointegration.

(b) On the other hand, the second result of **(v)** shows that the fixed-b HAR variance matrix estimator leads to the modified Wald statistic $\text{Wald}_{TIV,b}$ whose limit distribution can be represented by the pivotal quadratic form quantity $\eta'_{\mathcal{E}_W} L\left\{LL'\right\}^{-1} L'\eta_{\mathcal{E}_W}$. Importantly, the random projection matrix $P_L = L\left\{LL'\right\}^{-1} L'$ has rank $q = \text{rank}(L) = \text{rank}(H)$ $a.s.$ and is diagonalizable by an orthogonal matrix. Since the distribution of the random vector $\eta_{\mathcal{E}_W} = \mathcal{E}_W^{1/2}\eta_W$ is invariant to orthogonal transformation in the same way as the vector standard Brownian motions (W_x, W_X), the random quadratic form $\eta'_{\mathcal{E}_W} L\left\{LL'\right\}^{-1} L'\eta_{\mathcal{E}_W}$, which is a nonlinear functional of these standard Brownian motions and $W_{0.x}$, depends only on the rank of the matrix L, viz. the number of restrictions q. The HAR statistic $\text{Wald}_{TIV,b}$ is constructed in the usual manner for trend IV inference and in the cointegration case with $\Omega_{00.x} > 0$ provides a simple alternative to the procedures suggested in Vogelsang and Wagner (2014).[14]

(c) Analysis under the local alternative hypothesis $\mathcal{H}_A : Ha = h + \frac{d(a)}{n}$ shows that the Wald test based on the $\text{Wald}_{TIV,b}$ statistic has non-trivial asymptotic power under cointegration, with strength that depends on a random noncentrality parameter involving the quadratic form $\theta_d = d(a)'\mathcal{E}_W^{1/2} L\left\{LL'\right\}^{-1} L'\mathcal{E}_W^{1/2} d(a)$.

(d) In **(vi)** under multicointegration, the HAC-based Wald statistic $\text{Wald}_{TIV} \rightsquigarrow \chi^2_q$ and the HAR-based statistic

$$\text{Wald}_{TIV,b} \rightsquigarrow \eta'_{e.x}\mathcal{J}'_q\left\{\mathcal{J}_q\mathcal{F}_W\mathcal{J}'_q\right\}^{-1}\mathcal{J}_q\eta_{e.x}, \quad \text{with } \mathcal{J}_q = \left[I_q, 0\right].$$

Both test statistics have nontrivial asymptotic power under multicointegration and local alternative hypotheses of the form $\mathcal{H}_A : Ha = h + \frac{d(a)}{n^2}$. The statistic Wald_{TIV} has a noncentral χ^2_q limit distribution with noncentrality parameter $d(a)'d(a)$; and the $\text{Wald}_{TIV,b}$ statistic has a noncentral limit distribution involving the random noncentrality parameter $\vartheta_d = d(a)'\mathcal{J}'_q\left\{\mathcal{J}_q\mathcal{F}_W\mathcal{J}'_q\right\}^{-1}\mathcal{J}_q d(a)$.

(e) Theorem 3 **(v)** and **(vi)** show that the same HAR Wald statistic $\text{Wald}_{TIV,b}$ is asymptotically valid and pivotal for both cointegrated and multicointegrated systems, therefore providing a robust approach to inference concerning the cointegrating coefficients even under singularity. When no prior information on the presence of the multicointegration is available, appropriate critical values can be chosen after pretesting for multicointegration or more directly by using a bootstrap procedure for which the pivotal limit theory is an advantage. Exploration of these approaches will be considered in future research.

(f) These findings for the Wald test $\text{Wald}_{TIV,b}$ extend in a straightforward way to HAR-based t ratio statistics which produce asymptotically pivotal tests for both $\Omega_{00.x} = 0$ and $\Omega_{00.x} > 0$ cases.

In nonsingular systems with $\Omega_{00.x} > 0$ we can expect some loss of cointegration estimation efficiency and test power when using TIV estimation on the extended system (13) and the associated robust $\text{Wald}_{TIV,b}$ test rather than TIV estimation of (5) and associated Wald tests that rely on correct prior knowledge that the conditional error variance $\Omega_{00.x} > 0$. But when $\Omega_{00.x} = 0$, the faster $O(n^2)$ convergence rate of the estimator sharpens estimation efficiency and improves the discriminatory power of both the Wald_{TIV} test and the $\text{Wald}_{TIV,b}$ test.

5. Comparison with likelihood-based estimation

It is interesting to compare the performance of the TIV estimator with the maximum likelihood estimator in a correctly specified parametric model. Likelihood-based estimators of the cointegrating parameter in multicointegrated systems have been analyzed in the parametric $I(2)$ VAR framework (Johansen, 1997, 2006; Boswijk, 2000, 2010; Paruolo, 2000), and those asymptotics continue to hold in $I(2)$ VARs when the order of the autoregression is allowed to tend to infinity as a function of the sample size (Li and Bauer, 2020).

[14] The procedures suggested in Vogelsang and Wagner (2014) are designed only for the cointegration case with $\Omega_{00.x} > 0$ and do not apply under multicointegration.

For a comparison with VAR maximum likelihood estimation it is convenient to use the following notation from (Boswijk, 2010)

$$\Delta^2 Z_t = \alpha\beta' Z_{t-1} + \Gamma\Delta Z_{t-1} + \sum_{j=1}^{k-2}\Psi_j\Delta^2 Z_{t-j} + \varepsilon_t, \quad \bar{\alpha}_\perp\alpha'_\perp\Gamma\beta_\perp\bar{\beta}'_\perp = 0, \tag{33}$$

where the errors $\varepsilon_t \sim_{i.i.d.} \mathcal{N}(0,\Omega^z)$ with $m\times m$ variance matrix $\Omega^z > 0$, α and β are m-vectors, and Γ, $\{\Psi_j\}_{j=1}^{k-2}$ are $m\times m$ coefficient matrices.[15] In (33) the subscript $\perp$ signifies the orthogonal complement of a matrix or vector, and the overbar notation is defined by $\bar{\alpha} = \alpha(\alpha'\alpha)^{-1}$. Let $\delta = \bar{\alpha}'\Gamma\bar{\beta}_\perp$, then (33) becomes

$$\Delta^2 Z_t = \alpha(\beta' Z_{t-1} + \delta\beta'_\perp\Delta Z_{t-1}) + \Gamma\bar{\beta}\beta'\Delta Z_{t-1} + \sum_{j=1}^{k-2}\Psi_j\Delta^2 Z_{t-j} + \varepsilon_t. \tag{34}$$

This formulation yields the following cointegrating structure: the $I(2)$ time series Z_t cointegrates with coefficient β to the $I(1)$ times series $\beta' Z_t$; the $I(1)$ time series $\beta' Z_{t-1}$ and $\beta'_\perp\Delta Z_{t-1}$ cointegrate with coefficients 1 and δ to the $I(0)$ times series $\beta' Z_{t-1}+\delta\beta'_\perp\Delta Z_{t-1}$; and δ is called a multicointegrating coefficient for multicointegrated time series Z_t. As shown in the Online Supplement, the coordinates of Z_t can be rearranged so that $\beta' = [1,-a']$ and $\alpha' = [-1,0]$, with $\beta'_\perp = [a, I_{m_x}]$ and $\alpha'_\perp = [0, I_{m_x}]$ and Z_t partitioned with β as $Z'_t = [Y'_t, X'_t]$, translating (34) into the following equivalent system

$$\begin{bmatrix}\Delta y_t\\ \Delta x_t\end{bmatrix} = \frac{1}{1+a'a}\Gamma\begin{bmatrix}1\\ -a\end{bmatrix}(y_{t-1} - a'x_{t-1}) + \sum_{j=1}^{k-2}\Psi_j\begin{bmatrix}\Delta y_{t-j}\\ \Delta x_{t-j}\end{bmatrix} + \begin{bmatrix}-1\\ 0\end{bmatrix}(U_{0.x,t-1} - f'(aa'+I_{m_x})^{-1}au_{0,t-1}) + \varepsilon_t, \tag{35}$$

where $U_{0.x,t-1} = \sum_{s=1}^{t-1} u_{0.x,s} = \sum_{s=1}^{t-1}\Delta e_s = e_{t-1} - e_0$ which is $I(0)$ under multicointegration. Importantly, in (34) the cointegrating relationship is expressed in terms of the $I(2)$ variates Z_t and the equation has iid errors ε_t, whereas the rearranged system (35) is expressed in $I(1)$ form with the cointegrating relationship directly involving the $I(1)$ time series $\{y_{t-1}, x_{t-1}\}$ and this equation has moving average errors involving the pair of $I(0)$ time series $\{e_{t-1}, \varepsilon_t\}$. Calculations leading to (35) are given in the Online Supplement following the proof of Proposition 1, which is stated below.

Let $\hat{\beta}$ be the maximum likelihood estimator of β in (34) and $\hat{a}_{MLE}$ the corresponding estimator of a. The following proposition shows that time series generated by (33) correspond to a special case of (9) and (14) with multicointegration and that TIV estimation under this specialization is asymptotically equivalent to maximum likelihood estimation.

Proposition 1. *Time series $Z'_t = [Y'_t, X'_t]$ generated by the parametric VAR system (33) satisfy the semiparametric systems (9) and (14) with $\Omega_{00.x} = 0$. The TIV estimator $\hat{a}_{TIV}$ and maximum likelihood estimator $\hat{a}_{MLE}$ are then asymptotically equivalent with the same limit distribution.*

Under multicointegration the TIV estimator of the cointegrating coefficient is therefore asymptotically equivalent to the maximum likelihood estimator applied to the specialized and correctly specified finite dimensional VAR cointegrated system. Furthermore, the TIV estimator of the multicointegrating coefficient is asymptotically equivalent to the corresponding maximum likelihood estimator, again in the same specialized and correctly specified VAR multicointegrated systems. Indeed, for such multicointegrated systems $\delta = -f'(aa'+I_{m_x})^{-1}$ and the limiting distribution of $\hat{f}_{TIV}$, given in Phillips and Kheifets (2021), Theorem 4, is asymptotically equivalent to the limiting distribution of $-(\hat{a}_{MLE}\hat{a}'_{MLE}+I_{m_x})\hat{\delta}'$, and of $-(aa'+I_{m_x})\hat{\delta}'$, where $\hat{\delta}$ is the maximum likelihood estimator of δ in (34), which follows by a similar argument to that in proof of Proposition 1. Further analysis shows that the joint asymptotic distribution of the TIV and ML estimators of $[a', f']'$ is the same and has the following form

$$\begin{bmatrix}n^2(\hat{a}_{TIV}-a)\\ n(\hat{f}_{TIV}-f)\end{bmatrix} \rightsquigarrow \Omega_{xx}^{-1/2}\begin{bmatrix}\left(\int_0^1 W_{X.x}W'_{X.x}\right)^{-1}\int_0^1 W_{X.x}dW_{e.x}\\ \left(\int_0^1 W_{x.X}W'_{x.X}\right)^{-1}\int_0^1 W_{x.X}dW_{e.x}\end{bmatrix}\omega_{ee.x}^{1/2},$$

which will be explored in later work.

6. Simulations

This section reports the finite sample performance of TIV estimation of cointegrating relationships and compares TIV performance with IM-OLS estimation for various model specifications that include time series with and without multicointegration. Finite sample properties of the TIV Wald statistics are also studied in cases of cointegration and multicointegration. As a baseline for cointegrated series without multicointegration, simulations in past work (Phillips, 2014) showed good performance characteristics for TIV estimation in relation to other standard procedures such as FM-OLS and dynamic least squares in triangular systems as well as reduced rank regression (RRR) of Johansen (1988) in VAR system formulations with cointegration but not multicointegration. Those findings are now extended to include comparisons with IM-OLS in the present case.

[15] The present case is the one of $(r, s, p-r-s) = (1, 0, p-1)$ using notation in Johansen (1992).

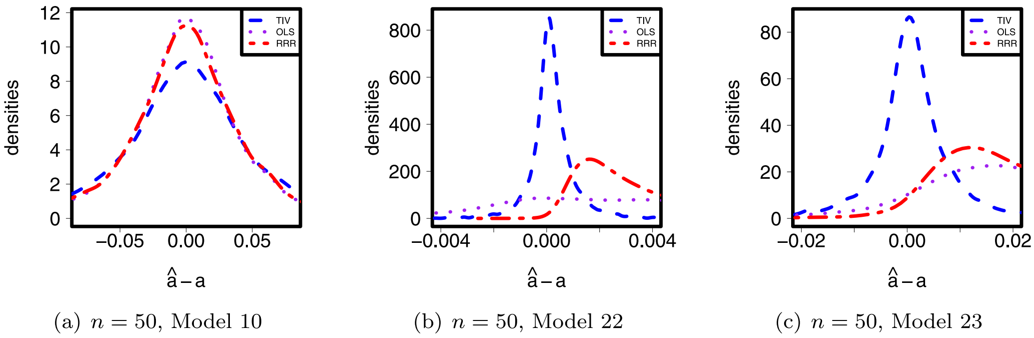


Fig. 1. Kernel estimates of the density functions of the estimation errors $\hat{a} - a$ for the TIV, RRR and LS estimators for sample size $n = 50$ in the pure cointegration model 10 and the multicointegration models 22 and 23.

Several experimental designs were employed based on the data generating process[16]

$$y_t = ax_t + u_{0t}$$
$$x_t = x_{t-1} + u_{xt}, \quad t = 1, \dots, n,$$

where $u_t = \eta_t^0 + D_1 \eta_{t-1}^0$, $\eta_t^0 \sim iid\,N(0, \Sigma)$, $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, the cointegrating coefficient $a = 2$, and the initialization of x_t is $x_0 = 0$. Both cointegrated and multicointegrated systems are considered and these are determined by the parameter settings of the (endogeneity) correlation coefficient ρ and the moving average coefficient matrix D_1. Various sample sizes are used and the number of replications in each experiment is 10,000. The following models were used.

Cointegrated models

Model 10: $D_1 = 0_{2\times 2}, \ \rho = 0$
Model 11: $D_1 = 0_{2\times 2}, \ \rho = 0.5$
Model 12: $D_1 = \begin{bmatrix} 0.3 & 0.4 \\ 0.8 & 0.6 \end{bmatrix}, \ \rho = 0.5$

Multicointegrated models

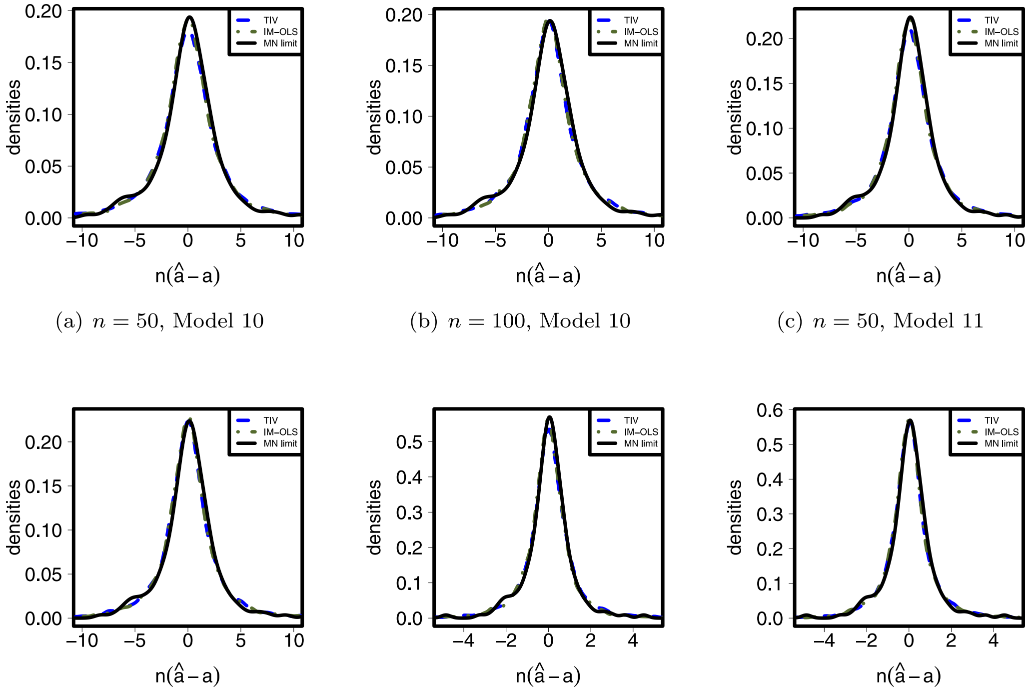
Model 20: $D_1 = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}, \ \rho = 0$
Model 21: $D_1 = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}, \ \rho = 0.5$
Model 22: $D_1 = \begin{bmatrix} 0.3 & 0.4 \\ 5.2 & 0.6 \end{bmatrix}, \ \rho = 0.5$
Model 23: $D_1 = \begin{bmatrix} -0.3 & 0.4 \\ 0.7 & -0.6 \end{bmatrix}, \ \rho = 0.5$

The models with $\rho = 0$ and zero diagonal elements in D_1 do not generate endogeneity or serial cross-correlation. So those models are pure cointegrated systems with exogenous regressors and iid innovations. Model 12 has been used in the cointegration literature in earlier work (Phillips and Loretan, 1991), and Model 22 modifies model 12 by introducing multicointegration into the system. Model 23 also generates a multicointegrated system, but with less variability in u_x compared[17] to Model 22.

For TIV estimation the orthonormal trigonometric polynomials $\varphi_k(r) = \sqrt{2}\sin\{(k - 1/2)\pi r\}$ were used as instrumental variables and the number of instruments was based on the setting $K = n^{3.8/5}$ in accord with the requirement in Theorem 2 that $K = o\left(n^{4/5-\delta}\right)$ for some $\delta > 0$. Following the recommendation in the paper the model was estimated by TIV with a fitted intercept. The asymptotic distributions in Theorem 3 were obtained by numerical computation from simulations with time series of length $n = 1000$ using 1000 replications.

16 To make explicit the role of endogeneity of stochastic regressors in the simulation analysis, we allow nonzero correlation between the first and second coordinates of η_t^0 as opposed to the equivalent linear process assumptions in Section 2, where $\eta_t \sim iid\,N(0, I_m)$.

17 The eigenvalues of D_1 are $(0,0)$, $(0,0)$, $(-0.1352, 1.0352)$ for cointegrated models 10, 11, 12 and $(-1,0)$, $(-1,0)$, $(-1,1.9)$, and $(-1,0.1)$ for cointegrated models 20, 21, 22, and 23, respectively.



(a) $n = 50$, Model 10 (b) $n = 100$, Model 10 (c) $n = 50$, Model 11

(d) $n = 100$, Model 11 (e) $n = 50$, Model 12 (f) $n = 100$, Model 12

Fig. 2. Kernel density estimates of the density functions of the estimation error $n(\hat{a}-a)$ for the TIV and the IM-OLS estimators and the density of the mixed-normal limit of the TIV estimator for sample sizes $n = 50$ and $n = 100$ and cointegration models 10, 11, and 12.

6.1. Finite sample distributions of the estimators

This subsection compares finite sample performance and convergence rates of TIV, RRR and OLS estimators of the cointegrating parameter a. Empirical densities of the centered and scaled TIV and IM-OLS estimators are compared with the asymptotic distributions given in Theorem 2.

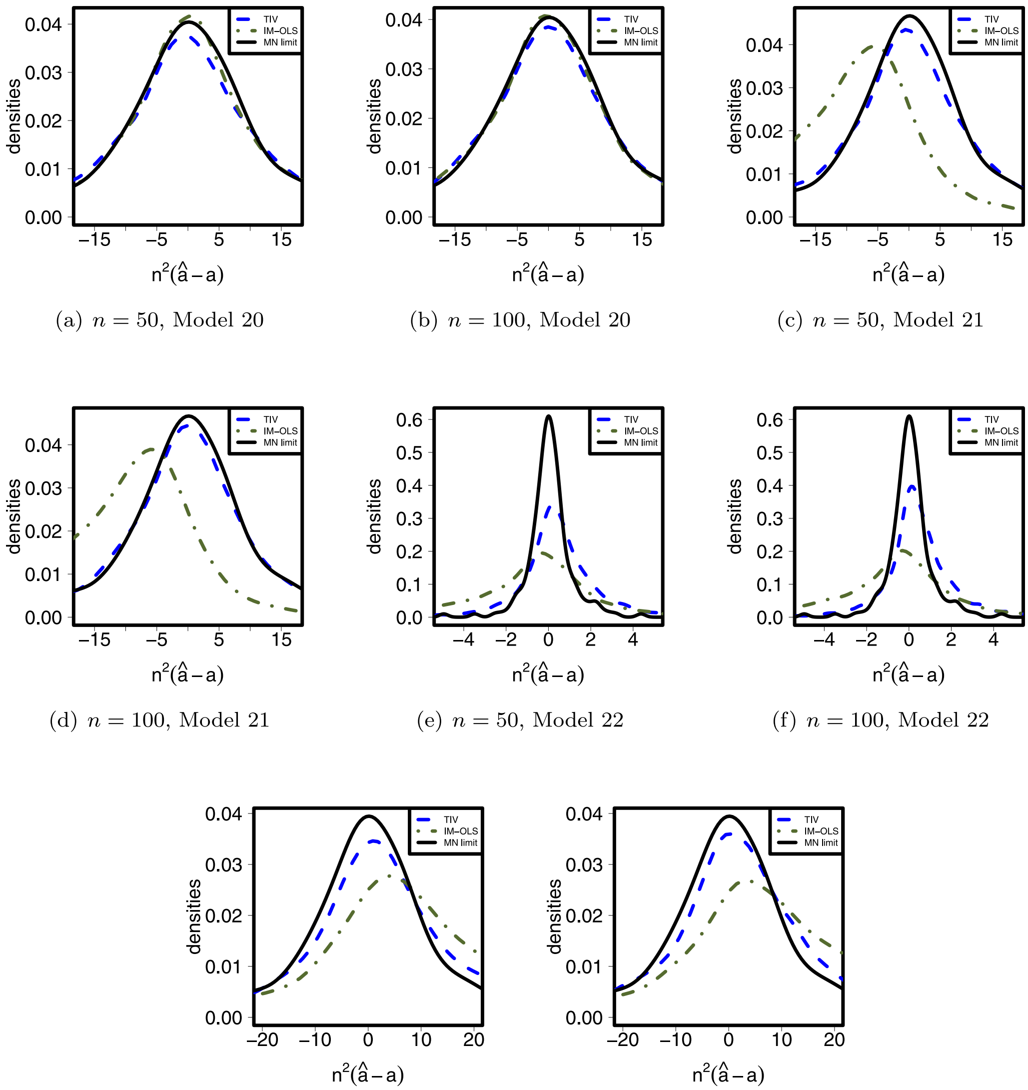
The centered densities of the TIV, RRR and OLS estimators are shown in Fig. 1 for $n = 50$ for three models. In the pure cointegration model 10, the three estimators show similar behavior although TIV, which is not needed in this pure cointegration case, shows somewhat greater dispersion than OLS and RRR. In models 22 and 23 under multicointegration the TIV estimator shows much greater concentration and little bias compared with OLS and RRR which are biased and skewed with greater dispersion. These results corroborate the limit theory in which TIV has an $O(n^2)$ convergence rate in multicointegrated models instead of the usual $O(n)$ rate for cointegrated systems.

We now compare the performance characteristics of TIV and IM-OLS in finite samples. Fig. 2 plots the densities of the centered TIV and IM-OLS estimators scaled by the appropriate convergence rate for each model against the mixed-normal asymptotic distribution. For the cointegration models 10–12, Fig. 2 plots the densities of the standardized TIV estimator $n(\hat{a}_{TIV} - a)$ based on the sample sizes $n = 50$ and $n = 100$ together with the asymptotic mixed normal density given in Theorem 2(iii). For the three models, the mixed-normal approximation works well as an approximation to the finite sample distributions of TIV, even for $n = 50$. The same is true for the densities of the standardized IM-OLS estimators, confirming the result in PK (2021), Theorem 1(i) and earlier results in Vogelsang and Wagner (2014).

For the multicointegrated models 20–23, the densities of the standardized TIV estimator $n^2(\hat{a}_{TIV} - a)$, based on sample sizes $n = 50$ and $n = 100$ and the simulated asymptotic mixed normal density, based on Theorem 2(iv), are plotted in Fig. 3. For all these models and cases the mixed-normal approximation to the distribution of the TIV estimator works well, again even for $n = 50$, whereas the IM-OLS estimator shows clear evidence of bias, skewness and greater dispersion for models 21–23. For model 20, where no endogeneity or serial correlation is present, which is the perfect set of conditions for the IM-OLS estimator, the densities of both estimators are approximated well by the mixed normal density, as predicted by PK (2021), Theorems 1(ii) and 2(iv) with some finite sample advantage in terms of reduced dispersion to the IM-OLS estimator in this case.

6.2. Size and power properties of the Wald test

Finite sample performance of Wald test statistics for testing the null hypothesis $\mathcal{H}_0 : a = 2$ were explored next. The empirical rejection rates under the null for the Wald statistics using the HAR variance estimate and the fixed-b asymptotic distribution given



(a) $n = 50$, Model 20 (b) $n = 100$, Model 20 (c) $n = 50$, Model 21

(d) $n = 100$, Model 21 (e) $n = 50$, Model 22 (f) $n = 100$, Model 22

(g) $n = 50$, Model 23 (h) $n = 100$, Model 23

Fig. 3. Kernel density estimates of the density functions of the estimation error $n^2(\hat{a}-a)$ for the TIV and the IM-OLS estimators and the density of the mixed-normal limit of the TIV estimator for sample sizes $n = 50$ and $n = 100$ and multicointegration models 20, 21, 22 and 23.

in Theorem 3 were calculated with the setting $b = 1$ and are reported in Table 1 for levels 10%, 5% and 1%. The results show excellent size control in all cases even for $n = 50$ in both the cointegrated and multicointegrated models.

For the Wald statistics using the HAC variance estimate calculated with the setting $M = 3n^{1/5}$ and using a χ^2 asymptotic distribution are presented in Table 2. For the cointegration models size is not controlled and the statistics diverge with the sample size. For the multicointegration models the rejection rates are 2–3 times larger than the nominal ones. Both cases show the importance of the HAR specification and appropriate limit theory for controlling size in Wald statistic testing.

Two control parameters – the number of instruments K and the bandwidth M (or b, the sample fraction) – are used in variance estimation. These parameters need to be set by the user. We analyzed the sensitivity of the Wald test to these parameter settings for models 12 and 22. Empirical rejection rates of the Wald test at the 5% nominal level were studied, taking $K \in \{10, 20, 30, 40\}$ for $n = 50$ and $K \in \{20, 40, 60, 80\}$ for $n = 100$ and varying M as fractions $\{0.2, 0.4, 0.6, 0.8, 1\}$ of the sample size n.[18] The rates under the null in Table 3 show: (i) that size is stable across a wide range of values of K and b in the cointegrated case; and (ii) that the size is stable across a wide range of values of K, when $b > 0.5$ in the multicointegrated case.

[18] Our asymptotic theory requires that $K = o(n^{4/5-\delta})$ for some $\delta > 0$ and $M = bn$ for some $b \in (0, 1]$.

Table 1
Test size using HAR variance estimates. Empirical rejection rates are shown at nominal 10%, 5% and 1% levels Wald test using the fixed-b asymptotic approximation, calculated for different models and sample sizes.

Model	n	10%	5%	1%
10	50	0.1178	0.0612	0.0163
10	100	0.1100	0.0591	0.0139
11	50	0.1178	0.0612	0.0163
11	100	0.1100	0.0591	0.0139
12	50	0.1139	0.0585	0.0154
12	100	0.1141	0.0581	0.0163
20	50	0.1070	0.0552	0.0130
20	100	0.0958	0.0479	0.0095
21	50	0.1070	0.0552	0.0130
21	100	0.0958	0.0479	0.0095
22	50	0.1242	0.0653	0.0139
22	100	0.1161	0.0623	0.0138
23	50	0.1201	0.0613	0.0135
23	100	0.0950	0.0513	0.0121

Table 2
Test size using HAC variance estimates. Empirical rejection rates are shown at nominal 10%, 5% and 1% levels for the Wald test using χ^2 critical values as approximations, calculated for different models and sample sizes.

Model	n	10%	5%	1%
10	50	0.6981	0.6442	0.5494
10	100	0.7302	0.6852	0.5949
11	50	0.6981	0.6442	0.5494
11	100	0.7302	0.6852	0.5949
12	50	0.6893	0.6348	0.5391
12	100	0.7359	0.6833	0.5907
20	50	0.2212	0.1488	0.0661
20	100	0.1722	0.1056	0.0396
21	50	0.2212	0.1488	0.0661
21	100	0.1722	0.1056	0.0396
22	50	0.2492	0.1730	0.0788
22	100	0.2318	0.1560	0.0638
23	50	0.2601	0.1871	0.0932
23	100	0.1983	0.1312	0.0536

Size-adjusted power calculations under the alternative $H_1 : a = 2.1$ are reported in Table 4. The results show that power is stable across all K values with a minor drop for larger bandwidths in the cointegration case. The size-adjusted power results in the multicointegration case under the alternative $H_1 : a = 2.001$ in Table 4 show that the power is high and increases with the sample size but with a minor drop for larger K and bandwidth size. In view of the faster convergence rate in the multicointegration case, local power in this case is evident for the much smaller departure $H_1 : a = 2.001$ from the null compared with the cointegration case where results for $H_1 : a = 2.1$ are reported.

Finally, in Table 5 we calculate the empirical rejection rates of Wald test statistics at the 5% level taking K as in Table 3 and (small) bandwidth as fractions $\{0.02, 0.04, 0.06, 0.08, 0.1\}$ of the sample size n using a χ^2 approximation instead of the correct limit theory. The test statistic diverges for all values of K and bandwidths in the cointegration case, while the size in the multicointergation case is sensitive to both number of instruments and bandwidth size.

7. Empirical illustration

Lee (1996) considered a model of the housing market that implies a long run equilibrium relationship between time series of housing starts and housing completions. If these series are multicointegrated then a parametric VAR $I(1)$ model will be misspecified. Engsted and Haldrup (1999) therefore analyzed the time series within an $I(2)$ framework allowing for multicointegration. In this section, we analyze the long run relationship between housing starts and completions over the five decade period 1968–2020 in an $I(1)$ semiparametric triangular model using the new TIV estimator and associated Wald tests. The number of instruments is again set to $K = n^{3.8/5}$, and the HAC and HAR variance estimates are calculated with the settings $M = 3n^{1/5}$ and $b = 1$.

The data are provided by the U.S. Census Bureau and the U.S. Department of Housing and Urban Development. They were obtained from FRED, the Federal Reserve Bank of St. Louis on March 16, 2021. We consider two series: $starts$ = Housing Starts, which comprise Total New Privately-Owned Housing Units Started [HOUST]; and $completions$ = Total New Privately-Owned Housing Units Completed [COMPUTSA]. Both series are reported in thousands of units and are seasonally adjusted. Our empirical analysis considers the following five decadal periods: (1) 1968-01-01–1979-12-31, (2) 1980-01-01–1989-12-31, (3) 1990-01-01–1999-12-31, (4) 2000-01-01–2009-12-31, (5) 2010-01-01–2019-12-31.

The cointegration relationship between $completions$ and $starts$ is estimated in each of these decades. In estimation no *a priori* assumption is made about the existence or non-existence of multicointegration. The results are given in Table 6. Over decades (1)

Table 3
Test size across K and b. Empirical rejection rates at nominal 5% level Wald test using the fixed-b asymptotic approximation, calculated for different models and sample sizes, for a range of instrument numbers K (in rows) and a range of bandwidths used in the kernel estimation of the variance determined by b (in columns).

Model	n	K\b	0.2	0.4	0.6	0.8	1
12	50	10	0.0533	0.0550	0.0576	0.0622	0.0633
12	50	20	0.0582	0.0541	0.0553	0.0593	0.0577
12	50	30	0.0580	0.0549	0.0540	0.0582	0.0581
12	50	40	0.0577	0.0543	0.0543	0.0587	0.0585
12	100	20	0.0612	0.0594	0.0606	0.0646	0.0625
12	100	40	0.0607	0.0557	0.0555	0.0588	0.0583
12	100	60	0.0599	0.0546	0.0544	0.0579	0.0567
12	100	80	0.0594	0.0548	0.0542	0.0576	0.0554
22	50	10	0.0345	0.0604	0.0660	0.0682	0.0662
22	50	20	0.0744	0.0766	0.0753	0.0716	0.0670
22	50	30	0.0982	0.0856	0.0803	0.0781	0.0713
22	50	40	0.1070	0.0935	0.0889	0.0839	0.0766
22	100	20	0.0622	0.0634	0.0627	0.0631	0.0599
22	100	40	0.0886	0.0723	0.0715	0.0672	0.0612
22	100	60	0.1020	0.0798	0.0716	0.0710	0.0629
22	100	80	0.1011	0.0855	0.0823	0.0754	0.0672

Table 4
Size-adjusted power across K and b. Empirical rejection rates at nominal 5% level Wald test using fixed-b approximation, calculated for different models and sample sizes, for a range of number of instruments, K (shown in rows), and a range of bandwidths used in the kernel estimation of the variance determined by b (shown in columns).

Model	n	K\b	0.2	0.4	0.6	0.8	1
12	50	10	0.8011	0.7657	0.7312	0.7036	0.6773
12	50	20	0.8143	0.7883	0.7593	0.7278	0.7057
12	50	30	0.8177	0.7883	0.7644	0.7320	0.7058
12	50	40	0.8208	0.7884	0.7624	0.7311	0.7019
12	100	20	0.9440	0.9233	0.9044	0.8819	0.8650
12	100	40	0.9484	0.9276	0.9095	0.8905	0.8747
12	100	60	0.9493	0.9289	0.9116	0.8921	0.8778
12	100	80	0.9494	0.9281	0.9126	0.8926	0.8786
22	50	10	0.5742	0.5313	0.4955	0.4627	0.4291
22	50	20	0.5222	0.4793	0.4369	0.3966	0.3621
22	50	30	0.4229	0.3749	0.3280	0.2976	0.2782
22	50	40	0.3129	0.2693	0.2423	0.2148	0.2049
22	100	20	0.9671	0.9577	0.9453	0.9282	0.9108
22	100	40	0.9396	0.9246	0.9084	0.8891	0.8652
22	100	60	0.8941	0.8722	0.8533	0.8258	0.8006
22	100	80	0.8274	0.8013	0.7787	0.7488	0.7238

and (2) to 1990, the estimate 0.97 is basically the same as that found in Lee (1996) and Engsted and Haldrup (1999). The estimate then declines to 0.96 in 1990–2000, then it is again 0.97 and goes to 0.95 in recent years. A possible interpretation is that 5% of houses under construction were never completed in those decades. A practical question is whether this fraction of uncompleted houses is significant, which can be formalized as a test of the null hypothesis $\mathcal{H}_0 : a = 1$ against the alternative $\mathcal{H}_1 : a < 1$.

The equilibrium errors from the cointegrated relationship between *completions* and *starts* accumulate into a stock variable of incomplete constructions. In each period, the inventory stock variable is measured as

$$Stock_t = \sum_{j=1}^{t} \left(\hat{a}_{TIV} \cdot start_j - completed_j \right), \tag{36}$$

and is plotted together with the flow variables *starts* and *completions* in Fig. 4. The figure reveals that these variables are again cointegrated, revealing a multicointegrated relationship between *completions* and *starts*. To conduct a test of the null $\mathcal{H}_0$, the asymptotic distributions of the Wald test statistic given in Theorem 4 are approximated by Monte Carlo simulations with 1000 replications for a sample size of 1000, and p-values for the two distributions (under cointegration and multicointegration) are calculated for each period.

The empirical findings for these tests are shown in Table 6. Allowing for multicointegration in the relationship we conclude that the null hypothesis $\mathcal{H}_0 : a = 1$ is rejected at the 5% level for all periods, except for period (5), as indicated by the p-values shown in the column 'pvalue-M'. If the multicointegrated relationship is ignored, the null hypothesis would not be rejected for any period, except for period (4), as indicated by the p-values given in the column 'pvalue-C'. Allowing for the presence of a multicointegrated relationship among starts, completions, and the housing stock therefore has a material impact on the empirical (cointegrating) relationship between starts and completions that suggests a significant shift in the relationship that raises the fraction of uncompleted houses.

Table 5
Test size across K and small b. Empirical rejection rates at nominal 5% level for the Wald test using χ^2 critical values, calculated for different models and sample sizes, for a range instrument numbers K (shown in rows), and a range of bandwidths used in the kernel estimation of the variance determined by b (shown in columns).

Model	n	K\b	0.02	0.04	0.06	0.08	0.1
12	50	10	0.7086	0.7086	0.6790	0.6450	0.6281
12	50	20	0.7460	0.7460	0.7174	0.6838	0.6614
12	50	30	0.7497	0.7497	0.7222	0.6891	0.6660
12	50	40	0.7518	0.7518	0.7222	0.6905	0.6658
12	100	20	0.8140	0.7525	0.7081	0.6773	0.6565
12	100	40	0.8238	0.7617	0.7182	0.6853	0.6615
12	100	60	0.8250	0.7635	0.7189	0.6865	0.6612
12	100	80	0.8256	0.7631	0.7183	0.6863	0.6607
22	50	10	0.0009	0.0009	0.0025	0.0129	0.0340
22	50	20	0.0172	0.0172	0.0301	0.0694	0.1172
22	50	30	0.0686	0.0686	0.0953	0.1524	0.2047
22	50	40	0.1481	0.1481	0.1757	0.2214	0.2549
22	100	20	0.0008	0.0062	0.0401	0.0863	0.1238
22	100	40	0.0104	0.0544	0.1367	0.1946	0.2285
22	100	60	0.0483	0.1284	0.2100	0.2511	0.2702
22	100	80	0.1172	0.1890	0.2378	0.2617	0.2761

Table 6
US housing construction data. Wald test statistics and p-values for the null hypothesis $\mathcal{H}_0 : a = 1$ under cointegration and multicointegration for successive decades over 1968–2020.

Period begins	Period ends	TIV	pvalue-M	pvalue-C
1968-01-01	1979-12-31	0.9722	0.0398	0.1418
1980-01-01	1989-12-31	0.9735	0.0149	0.0746
1990-01-01	1999-12-31	0.9607	0.0274	0.1169
2000-01-01	2009-12-31	0.9709	0.0050	0.0448
2010-01-01	2019-12-31	0.9455	0.0522	0.1667

The parameter estimates are similar to those in Lee (1996) and Engsted and Haldrup (1999) for the corresponding periods. What our analysis and findings reveal in addition to the results in those papers is that the long run relationship between housing starts and completions is significantly different from a one-to-one relation when a multicointegrating linkage among the variables is present.

To keep this illustration of the methodology brief we focus on parameter estimation and inference, omitting detailed analysis of model assumptions and fit. Choosing time periods by decade, while not data-driven and somewhat arbitrary, displays some instabilities in the coefficients over these periods. The results provide motivation to develop formal tests for structural breaks within the present setup of potential multicointegration, a subject that is left for future research.

A primary contribution of the present work is the construction of a Wald test statistic that has a pivotal asymptotic distribution with or without multicointegration. This appealing property of the statistic opens the door to designing a test procedure and bootstrap algorithm that do not require a practitioner to choose a particular cointegration regime and associated critical values. Such a procedure is currently under development.

8. Conclusion

This paper has studied the effects of singularities in long run conditional covariance matrices on estimation and inference in cointegrating regression models. Such singularities are shown to be present whenever a cointegrated $I(1)$ system happens to involve multicointegrated time series. Singularities complicate estimation and inference by leading to non-pivotal, nuisance parameter dependencies in all existing methods of estimating nonstationary time series regressions. But in view of their natural focus on the analysis of long run properties, instrumental variable regression with deterministic trend regressors or similar trend transforms have appealing properties even under singularities. The results of the present analysis show that, in spite of the complications introduced by long run variance matrix singularities, certain key advantages of the trend IV regression approach continue to apply. Notably, the limit theory of trend IV regression is mixed normal and Wald tests based on traditional sandwich formulae may be conducted under pivotal asymptotics without knowledge of potential singularities or the presence of multicointegration in the time series. Use of fixed-b methods in conjunction with trend IV regression are particularly helpful in achieving pivotal limit theory when the regression equation is partially spurious with nonstationary errors and usual HAC-based test statistics are divergent under the null.

The analysis in this paper deals with estimation and inference in a scalar cointegrating relationship. The main ideas and methods of estimation and inference extend to systems estimation. In such cases, the higher convergence rate $O\left(n^2\right)$ applies in the (possibly matrix) direction L_1 of singularity of $\Omega_{00.x}$ for which $L_1'\Omega_{00.x}L_1 = 0$ and the slower $O(n)$ rate applies in the orthogonal direction L_2. The full matrix of cointegrating coefficients then converges to a mixed normal limit distribution which is a matrix transform of the slower rate limit distribution, just as in usual cointegration limit theory (Park and Phillips, 1988, 1989; Phillips, 1988, 1989). The analysis and algebra in this general case follows the same approach as that in cointegrated regression systems with cointegrated regressors and unrestricted VAR estimation with cointegrated variates, as detailed in Phillips (1995). But inferential limit theory is more subtle in this case of singularity in the matrix $\Omega_{00.x}$ because of interaction between the restriction matrix H, the rotation matrix $L = [L_1, L_2]$ isolating the two directions of convergence, and the matrix normalization involved in standardizing the TIV estimation errors. A full analysis of this case requires the use of methods and limit theory for Wald tests under general conditions of matrix normalization as recently developed in Magdalinos and Phillips (2019). The application of those methods in the present context is left for subsequent work.

9. Glossary of notation

We use the following notation for data matrices, various functionals of the Brownian motions $(B_X, B_x, B_e, B_{0.x})$, and associated stochastic processes including their standard Brownian motion analogues $(W_X, W_x, W_e, W_{0.x})$. The functionals are defined in the paper and repeated here for convenience. In the following formulae $\int$ represents $\int_0^1$ when the limits are not provided.

$$P_x = x\left(x'x\right)^{-1}x', \quad Q_x = I - x\left(x'x\right)^{-1}x'$$

$$X = [X_1, \dots, X_n],\ x = [x_1, \dots, x_n],\ Y = [Y_1, \dots, Y_n],\ u_x = [u_{x1}, \dots, u_{xn}]$$

$$B_X(r) = \int_0^r B_x = \Omega_{xx}^{1/2} W_X = \Omega_{xx}^{1/2} \int_0^r W_x$$

$$B_{X.x}(r) = B_X(r) - \int B_X B_x \left(\int B_x B_x'\right)^{-1} B_x(r) = \Omega_{xx}^{1/2} W_{X.x}(r)$$

$$W_{X.x}(r) = W_X(r) - \int W_X W_x' \left(\int W_x W_x'\right)^{-1} W_x(r)$$

$$\overrightarrow{B_{X.x}}(r) = \int_r^1 B_{X.x} = \Omega_{xx}^{1/2}\, \overrightarrow{W_{X.x}}(r) = \Omega_{xx}^{1/2} \int_r^1 W_{X.x}$$

$$\mathcal{A}_{X.x} = \int B_{X.x} B_{X.x}' = \Omega_{xx}^{1/2} \mathcal{A}_{W,X.x} \Omega_{xx}^{1/2} = \Omega_{xx}^{1/2} \int W_{X.x} W_{X.x}'\, \Omega_{xx}^{1/2}$$

$$B_{e.x}(r) = B_e(r) - \omega_{ex}\Omega_{xx}^{-1} B_x(r) =_d \omega_{ee.x}^{1/2} W_{e.x},\quad \omega_{ee.x} = \omega_{ee} - \omega_{ex}\Omega_{xx}^{-1}\omega_{xe},$$

$$B_{0.x}(r) = B_0(r) - \Omega_{0x}\Omega_{xx}^{-1} B_x(r) =_d \Omega_{00.x}^{1/2} W_{0.x}$$

$$B_{x.X}(r) = B_x(r) - \int B_x B_X' \left(\int B_X B_X'\right)^{-1} B_X(r) =_d \Omega_{xx}^{1/2} W_{x.X}$$

$$W_{x.X}(r) = W_x(r) - \int W_x W_X' \left(\int W_X W_X'\right)^{-1} W_X(r)$$

$$\begin{aligned}\widetilde{B_{0.x}}(r) &= B_{0.x}(r) - \int B_{0.x} B_{X.x}' \left(\int B_{X.x} B_{X.x}'\right)^{-1} B_X(r) \\ &\quad - \int B_{0.x} B_{x.X}' \left(\int B_{x.X} B_{x.X}'\right)^{-1} B_x(r) \\ &= \Omega_{xx}^{1/2} \widetilde{W_{0.x}}(r)\end{aligned}$$

$$\eta_K = \int_{r=0}^1 \tilde{\varphi}_K(r) B_x(r)'\, dr,\ \xi_K = \int_{r=0}^1 \tilde{\varphi}_K(r)\, dB_x(r)',\ \mu_K = \int_{r=0}^1 \tilde{\varphi}_K(r) B_X(r)'\, dr$$

$$\eta_W = \left(\int W_{X.x} W_{X.x}'\right)^{-1} \int \overrightarrow{W_{X.x}}\, dW_{0.x},$$

$$\eta_{e.x} = \left(\int W_{X.x} W_{X.x}'\right)^{-1} \left(\int W_{X.x}\, dW_{e.x}\right)$$

$$\Psi_{0.xK} = \int \tilde{\varphi}_K B_{0.x} \equiv \mathcal{N}\left(0, \Omega_{00.x}\left(\int\int (r \wedge s)\, \tilde{\varphi}_K(r)\, \tilde{\varphi}_K(s)'\, dr ds\right)\right)$$

$$\psi_{eK} = \int \tilde{\varphi}_K\, dB_e$$

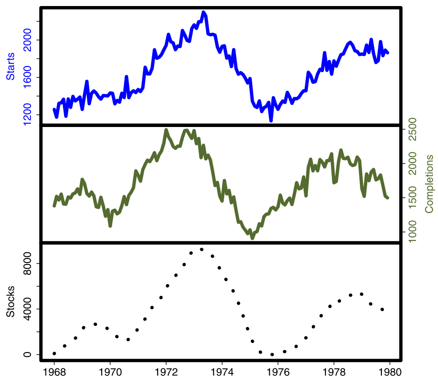
$$J_K = Q_{\xi_K} - Q_{\xi_K}\eta_K \left(\eta_K' Q_{\xi_K} \eta_K\right)^{-1} \eta_K' Q_{\xi_K}$$

$$\mathcal{J}_q = \left[I_q, 0\right]$$

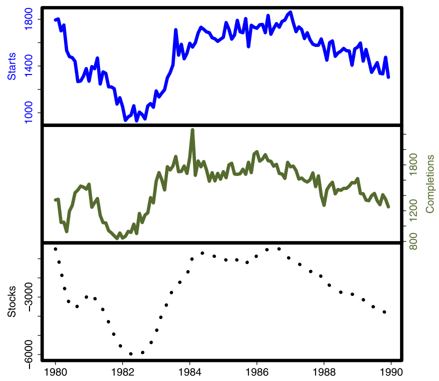
$$\vartheta_K = \int B_{X.x}(r)\, \tilde{\varphi}_K(r)'\, dr = \left(\zeta_1, \dots, \zeta_K\right)$$

$$\ell_X' = \int dB_{e.x} B_{X.x}' \left(\int B_{X.x} B_{X.x}'\right)^{-1} =_d \omega_{ee.x}\, \ell_{W,X}'\, \Omega_{xx}^{-1/2}$$

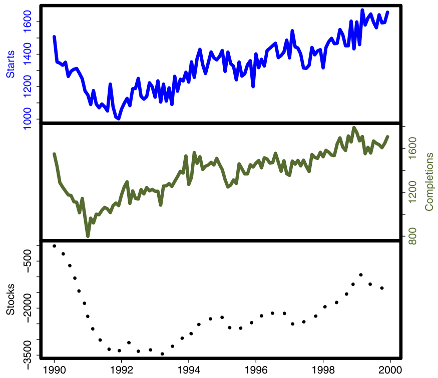
$$\ell_{W,X}' = \int dW_{e.x} W_{X.x}' \left(\int W_{X.x} W_{X.x}'\right)^{-1}$$



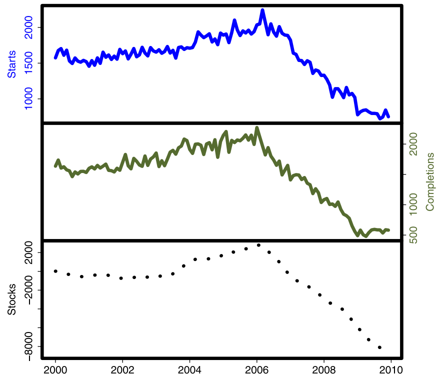
(a) 1970s



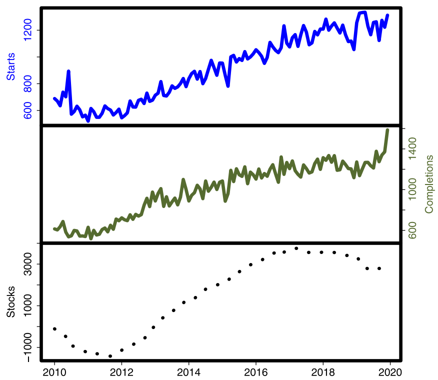
(b) 1980s



(c) 1990s



(d) 2000s



(e) 2010s

Fig. 4. Housing starts (Starts), completions (Completions) and accumulated difference (Stock) data for successive decades (a) 1970s, (b) 1980s, (c) 1990s (d) 2000s and (e) 2010s.

$$\ell'_x = \int dB_{e.x} B_{x.X} \left(\int B_{x.X} B'_{x.X} \right)^{-1} =_d \omega_{ee.x} \ell'_{W,x} \Omega_{xx}^{-1/2}$$

$$\ell'_{W,x} = \int dW_{e.x} W'_{x.X} \left(\int W_{x.X} W'_{x.X} \right)^{-1}$$

$$\begin{aligned}
\ell_+ &= (\ell_X', \ell_x')' =_d \omega_{ee.x} \ell_{W,+}' \Omega_{xx}^{-1/2} \\
\ell_{W,+}' &= (\ell_{W,X}', \ell_{W,x}')' \\
B_+(r) &= \left(B_X(r)', B_x(r)'\right)' = \Omega_{xx}^{1/2} W_+(r) = \Omega_{xx}^{1/2} \left(W_X(r)', W_x(r)'\right)' \\
\mathcal{Q}_B &= B_{e.x}(r) + \ell_+' \int_0^r B_+ = \Omega_{ee.x}^{1/2} \mathcal{Q}_W(r) \\
\mathcal{Q}_W(r) &= W_{e.x}(r) + \ell_{W,+}' \int_0^r W_+,
\end{aligned}$$

where $\tilde{\varphi}_K(r) = \left(\varphi_1(r), \ldots, \varphi_K(r)\right)'$.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2023.105622.

References

Andrews, D.W., Cheng, X., 2012. Estimation and inference with weak, semi-strong, and strong identification. Econometrica 80 (5), 2153–2211.
Andrews, D.W., Cheng, X., 2014. GMM estimation and uniform subvector inference with possible identification failure. Econometric Theory 30 (2), 287–333.
Boswijk, H.P., 2000. Mixed normality and ancillarity in i(2) systems. Econometric Theory 16 (6), 878–904.
Boswijk, H.P., 2010. Mixed normal inference on multicointegration. Econometric Theory 26 (5), 1565–1576.
Bruns, S.B., Csereklyei, Z., Stern, D.I., 2020. A multicointegration model of global climate change. J. Econometrics 214 (1), 175–197.
Corbae, D., Ouliaris, S., Phillips, P.C.B., 2002. Band spectral regression with trending data. Econometrica 70 (3), 1067–1109.
Engle, R.F., 1974. Band spectrum regression. Internat. Econom. Rev. 15, 1–11.
Engsted, T., Gonzalo, J., Haldrup, N., 1997. Testing for multicointegration. Econom. Lett. 56 (3), 259–266.
Engsted, T., Haldrup, N., 1999. Multicointegration in stock-flow models. Oxf. Bull. Econ. Stat. 61 (2), 237–254.
Engsted, T., Johansen, S., 1999. Granger's representation theorem and multicointegration. In: Cointegration, Causality and Forecasting: Festschrift in Honour of Clive Granger. pp. 200–212.
Granger, C.W.J., Lee, T.-H., 1989. Investigation of production, sales and inventory relationships using multicointegration and non-symmetric error correction models. J. Appl. Econom. 4 (S1), S145–S159.
Granger, C.W.J., Lee, T.-H., 1990. Investigation of production, sales and inventory relationships using multicointegration and non-symmetric error correction models. Adv. Econom. 8, 71–84.
Hannan, E.J., 1963. Regression for time series. In: Rosenblatt, M. (Ed.), Time Series Analysis. Wiley, pp. 17–37.
Hwang, J., Sun, Y., 2018. Simple, robust, and accurate F and t tests in cointegrated systems. Econometric Theory 34, 949–984.
Johansen, S., 1988. Statistical analysis of cointegration vectors. J. Econom. Dynam. Control 12 (2), 231–254.
Johansen, S., 1992. A representation of vector autoregressive processes integrated of order 2. Econom. Theory 8, 188–202.
Johansen, S., 1995. A statistical analysis of cointegration for I(2) variables. Econometric Theory 11, 25–59.
Johansen, S., 1997. Likelihood analysis of the I(2) model. Scand. J. Stat. 24 (4), 433–462.
Johansen, S., 2006. Statistical analysis of hypotheses on the cointegrating relations in the I(2) model. J. Econometrics 132 (1), 81–115.
Kheifets, I.L., Phillips, P.C.B., 2023. Fully modified least squares cointegrating parameter estimation in multicointegrated system. J. Econometrics 232, 300–319.
Lazarus, E., Lewis, D.J., Stock, J.H., Watson, M.W., 2018. HAR inference: Recommendations for practice. J. Bus. Econom. Statist. 36 (4), 541–559.
Lee, T.-H., 1996. Stock adjustment for multicointegrated series. Empir. Econ. 21 (4), 633–639.
Li, Y., Bauer, D., 2020. Modeling I(2) processes using vector autoregressions where the lag length increases with the sample size. Econometrics 8 (3).
Magdalinos, T., Phillips, P.C.B., 2019. Wald Tests under Matrix Normalization. Working Paper, University of Southampton.
Müller, U.K., 2007. A theory of robust long-run variance estimation. J. Econometrics 141 (2), 1331–1352.
Müller, U.K., Watson, M.W., 2018. Long-run covariability. Econometrica 86 (3), 775–804.
Park, J.Y., Phillips, P.C.B., 1988. Statistical inference in regressions with integrated processes: Part 1. Econometric Theory 4 (3), 468–497.
Park, J.Y., Phillips, P.C.B., 1989. Statistical inference in regressions with integrated processes: Part 2. Econometric Theory 5 (1), 95–131.
Paruolo, P., 2000. Asymptotic efficiency of the two stage estimator in I(2) systems. Econometric Theory 16 (4), 524–550.
Phillips, P.C.B., 1986. Understanding spurious regressions in econometrics. J. Econometrics 33 (3), 311–340.
Phillips, P.C.B., 1988. Multiple regression with integrated time series. In: Prabhu, N.U. (Ed.), Statistical Inference from Stochastic Processes. In: Contemporary Mathematics, vol. 80, pp. 79–106.
Phillips, P.C.B., 1989. Partially identified econometric models. Econometric Theory 5 (2), 181–240.
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., 1998. New tools for understanding spurious regressions. Econometrica 66, 1299–1325.
Phillips, P.C.B., 2005a. Challenges of trending time series econometrics. Math. Comput. Simulation 68 (5–6), 401–416.
Phillips, P.C.B., 2005b. HAC estimation by automated regression. Econometric Theory 21 (1), 116–142.
Phillips, P.C.B., 2006. Optimal Estimation of Cointegrated Systems with Irrelevant Instruments. Cowles Foundation Discussion Paper No. 1547, Yale University.
Phillips, P.C.B., 2014. Optimal estimation of cointegrated systems with irrelevant instruments. J. Econometrics 178, 210–224.
Phillips, P.C.B., 2022. Asymptotics of Polynomial Time Trend Estimation and Hypothesis Testing Under Rank Deficiency. Working Paper, Yale University, p. 24.
Phillips, P.C.B., Kheifets, I.L., 2021. On Multicointegration. Cowles Foundation Discussion Paper, Yale University.
Phillips, P.C.B., Kheifets, I.L., 2023. Online Supplement to 'Robust High-Dimensional IV Cointegration Estimation and Inference'. Cowles Foundation Discussion Paper, Yale University.
Phillips, P.C.B., Loretan, M., 1991. Estimating long-run economic equilibria. Rev. Econom. Stud. 58 (3), 407–436.
Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. Ann. Statist. 20, 971–1001.
Phillips, P.C.B., Wang, X., Zhang, Y., 2019. HAR testing for spurious regression in trend. Econometrics 7 (4), 50.
Saikkonen, P., 1991. Asymptotically efficient estimation of cointegration regressions. Econometric Theory 7 (1), 1–21.
Stock, J.H., Watson, M.W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. Econometrica 61, 783–820.
Sun, Y., 2004. A convergent t-statistic in spurious regressions. Econometric Theory 20 (5), 943–962.
Vogelsang, T.J., Wagner, M., 2014. Integrated modified OLS estimation and fixed-b inference for cointegrating regressions. J. Econometrics 178 (2), 741–760.