SFB 823

Accurate and (almost) tuning parameter free inference in cointegrating regressions

Karsten Reichold, Carsten Jentsch

Discussion Paper

Nr. 33/2020

SFB 823

Accurate and (Almost) Tuning Parameter Free Inference in Cointegrating Regressions*

Karsten Reichold
Department of Economics
University of Klagenfurt
Klagenfurt, Austria
&
Department of Statistics
TU Dortmund University
Dortmund, Germany

Carsten Jentsch
Department of Statistics
TU Dortmund University
Dortmund, Germany

December 17, 2020

Abstract

Tuning parameter choices complicate statistical inference in cointegrating regressions and affect finite sample distributions of test statistics. As commonly used asymptotic theory fails to capture these effects, tests often suffer from severe size distortions. We propose a novel self-normalized test statistic for general linear hypotheses, which avoids the choice of tuning parameters. Its limiting null distributions is nonstandard, but simulating asymptotically valid critical values is straightforward. To further improve the performance of the test in small to medium samples, we employ the vector autoregressive sieve bootstrap to construct critical values. To show its consistency, we establish a bootstrap invariance principle result under conditions that go beyond the assumptions commonly imposed in the literature. Simulation results demonstrate that our new test outperforms competing approaches, as it has good power properties and is considerably less prone to size distortions.

Keywords: Bootstrap invariance principle, Cointegration, IM-OLS, Inference, Self-normalization, Vector autoregressive sieve bootstrap

*Karsten Reichold gratefully acknowledges partial financial support by the German Research Foundation via the Collaborative Research Center SFB 823 *Statistical Modelling of Nonlinear Dynamic Processes*. We thank Katharina Hees and Fabian Knorre for valuable comments. `MATLAB` code for applications is available on www.github.com/kreichold/CointSelfNorm.

1 Introduction

Conducting inference in cointegrating regressions is cumbersome as it usually requires tuning parameter choices. Even in case these choices are "optimal", commonly used asymptotic theory usually fails to capture their effects on finite sample distributions of test statistics, often resulting in severe size distortions. In this paper we address these issues by proposing a novel test statistic for general linear hypotheses that itself is completely tuning parameter free. Test decisions can be based on either asymptotically valid critical values or bootstrap critical values, which might be preferred in small to medium samples.

Cointegration methods have been and are widely used to analyze long-run relationships between stochastically trending variables in many areas such as macroeconomics, environmental economics and finance, see, e. g., Benati *et al.* (2020), Wagner (2015) and Rad *et al.* (2016) for recent examples. In addition to these classical fields of application, cointegration methods have recently proven to be useful to describe phenomena also in other contexts. For instance, Dahlhaus *et al.* (2018) describe the close connection between cointegration and the theory of phase synchronization in physics and Phillips *et al.* (2020) apply cointegration-based methods to estimate Earth's climate sensitivity.

Although the OLS estimator is consistent in cointegrating regressions, its limiting distribution is usually contaminated by second order bias terms, reflecting the correlation structure between regressors and errors. This makes the OLS estimator infeasible for conducting inference based on (simulated) quantile tables of (non)standard distributions. The literature provides several estimators which overcome this difficulty at the cost of tuning parameter choices: the number of leads and lags for the dynamic OLS (D-OLS) estimator of Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993), kernel and bandwidth choices for the fully modified OLS (FM-OLS) estimator of Phillips and Hansen (1990) and the canonical cointegrating regression (CCR) estimator of Park (1992), or type and number of basis functions for the trend instrument variable (TIV) estimator of Phillips (2014). Such tuning parameters are often difficult to choose in practice and the finite sample performance of the estimators and tests based upon them often reacts sensitively to their choices. In particular, corresponding tests often suffer from severe size distortions, see, e. g., Vogelsang and Wagner (2014, Tables 2–3) for a detailed overview.

In contrast to the aforementioned approaches, the integrated modified OLS (IM-OLS) estimator of Vogelsang and Wagner (2014) avoids the choice of tuning parameters. However, standard asymptotic inference based on the IM-OLS estimator does require the estimation of a long-run variance parameter. This is typically achieved by non-parametric kernel estimators, which necessitate kernel and bandwidth choices. To capture their effects in finite samples, Vogelsang and Wagner (2014) propose fixed-b theory for obtaining critical values. However, their simulation results reveal that when endogeneity and/or error serial correlation is strong, a large sample size is needed for the procedure to yield reasonable sizes. Moreover, for small to medium samples, test performance still seems to be sensitive to the choice of b. Similarly, Hwang and Sun (2018) develop "partial" fixed-K theory for the TIV estimator, which captures the choice of the number of basis functions but ignores the impact of the basis functions itself.

Instead of trying to capture finite sample effects of tuning parameter choices, we propose a novel IM-OLS based test statistic for general linear hypotheses, which is completely tuning parameter free. The test statistic is based on a self-normalization approach that is similar in spirit to but different from the approach of Kiefer *et al.* (2000) for stationary data.[1] The limiting null distribution of the self-normalized test statistic is nonstandard but simulating asymptotically valid critical values is straightforward.

As asymptotically valid critical values might not be appropriate in small to medium samples, we use the bootstrap to construct critical values that might be more accurate in this case. In particular, we propose a residual-based vector autoregressive (VAR) sieve resampling procedure. In comparison to other resampling approaches broadly applicable to stationary processes – such as the residual-based block bootstrap proposed in Paparoditis and Politis (2003) for unit-root testing, or in Brüggemann *et al.* (2016) and Jentsch and Lunsford (2019) for heteroscedasticity-robust inference in (proxy) structural VARs – the VAR sieve bootstrap captures the second-order dependence structure of the original process, which is in our context both necessary *and* sufficient for the bootstrap to be consistent, in a simple manner. In particular, it only requires the determination of the order of the VAR, which is a relatively straightforward and well understood

[1]For a detailed review of recent developments on inference based on self-normalization in the stationary time series context, we refer to Shao (2015). For an application of self-normalization to high-dimensional stationary time series and monitoring cointegrating relationships, see Wang and Shao (2020) and Knorre *et al.* (2020), respectively.

task in practice.[2] The VAR sieve bootstrap is thus frequently used in related literature: Psaradakis (2001), inspired by the seminal work of Li and Maddala (1997), shows the usefulness of the sieve bootstrap in cointegrating regressions and Park (2002) provides its asymptotic justification by proving an underlying invariance principle result. Subsequently, Chang and Park (2003) and Chang *et al.* (2006) apply the sieve bootstrap to unit root testing and to conduct D-OLS based inference in cointegrating regressions, respectively, and Palm *et al.* (2010) use the sieve bootstrap in the context of testing for cointegration in conditional error correction models.[3] Moreover, residual-based bootstrap resampling in cointegrating VAR models of fixed order has been used in, e. g., Cavaliere *et al.* (2012) and Cavaliere *et al.* (2015) to determine the cointegration rank and to test hypotheses on the cointegrating vector, respectively.

To show consistency of the VAR sieve bootstrap in our context, we establish a bootstrap invariance principle result under relatively general conditions that go beyond the assumptions commonly imposed in the literature. Our framework allows for so-called weak white noises that are uncorrelated, but not necessarily independent and also for various concepts to quantify such weak forms of dependence of the innovation process. In particular, we do not impose the assumption of a causal linear process with i.i.d. innovations as in Park (2002). Given its wide applicability the bootstrap invariance principle result might be of interest in its own right.

The theoretical analysis is complemented by a simulation study to assess the performance of the proposed methods, benchmarked against competing approaches, including the commonly employed D-OLS and FM-OLS based tests. In contrast to the commonly applied tests, which show severe size distortions, our novel approach proves to hold the prescribed level approximately at the expense of only minor power losses. This might be seen as a huge advantage given that large size distortions under the null hypothesis are very common in the unit root and cointegrating literature.

The rest of the paper is organized as follows: Section 2 introduces the model and its underlying assumptions. Section 3 constructs the self-normalized test

[2]Choosing the block size in practice seems to be difficult. Although Politis and White (2004) and Patton *et al.* (2009) propose estimators of the optimal block size, these are tailor-made for the sample mean for univariate time series data.

[3]The approach of Chang *et al.* (2006) still requires the choice of leads and lags for estimation and additional kernel and bandwidth choices for inference.

statistic and derives its limiting null distribution. Section 4 presents the bootstrap procedure and derives its asymptotic validity. Section 5 presents finite sample results and Section 6 concludes. The proofs of the main results are provided in Appendix B, whereas those of auxiliary results are relegated to the Online Appendix.

We use the following notation. The integer part of a real number x is denoted by $\lfloor x \rfloor$. For a real matrix A we denote its transpose by A' and its Frobenius norm by $|A|_F = (\mathrm{tr}(A'A))^{1/2}$, where $\mathrm{tr}(\cdot)$ denotes the trace and for a vector the Frobenius norm becomes the Euclidean norm. The k-dimensional identity matrix is denoted by I_k and $0_{j \times k}$ (or simply 0) denotes a $(j \times k)$-dimensional matrix of zeros. With $\mathrm{diag}(\cdot)$ we denote a (block) diagonal matrix with diagonal elements specified throughout. Equality in distribution is signified by $\overset{d}{=}$. With $\xrightarrow{w}$ and $\xrightarrow{p}$ we denote weak convergence and convergence in probability, respectively. Adding the superscript "$*$" signifies convergence in the bootstrap probability space. The corresponding probability measure is denoted by $\mathbb{P}^*$ and $\mathbb{E}^*(\cdot)$ denotes the expectation with respect to $\mathbb{P}^*$. For notational simplicity, a Brownian motion $\{M(r),\ 0 \leq r \leq 1\}$ is denoted by $M(r)$.

2 The Model and Assumptions

We consider the cointegrating regression model

$$y_t = x_t'\beta + u_t, \tag{2.1}$$

$$x_t = x_{t-1} + v_t, \tag{2.2}$$

$t = 1, \ldots, T$, where $(y_t)_{t=1,\ldots,T}$ is a scalar time series and $(x_t)_{t=1,\ldots,T}$ is an $m \times 1$ vector of time series. For brevity we set $x_0 = 0$.[4] For $\{w_t\}_{t\in\mathbb{Z}} := \{[u_t, v_t']'\}_{t\in\mathbb{Z}}$ we assume the following:

Assumption 1. *Let $\{w_t\}_{t\in\mathbb{Z}}$ be an $\mathbb{R}^{1+m}$-valued, strictly stationary and purely nondeterministic stochastic process of full rank*[5] *with $\mathbb{E}(w_t) = 0$ and*

[4]Deterministic regressors are excluded from (2.1) to ease exposition of the main arguments. However, it is straightforward to incorporate, e. g., the leading case of an intercept and polynomial time trends, $d_t = [1, t, \ldots, t^p]'$, $p \geq 1$. Please note that the accompanying `MATLAB` code allows to handle this more general case.

[5]The process $\{w_t\}_{t\in\mathbb{Z}}$ is of full rank, if the components of its innovation process are linearly independent. For more details we refer to Meyer and Kreiss (2015, p. 379).

$\sup_{t\in\mathbb{Z}} \mathbb{E}(|w_t|_F^a) < \infty$, for some $a > 2$. The autocovariance matrix function $\Gamma(\cdot)$ of $\{w_t\}_{t\in\mathbb{Z}}$ fulfills $\sum_{h=-\infty}^{\infty}(1+|h|)^k|\Gamma(h)|_F < \infty$ for some $k \geq 3/2$. For the spectral density matrix $f(\cdot)$ of $\{w_t\}_{t\in\mathbb{Z}}$ we assume that there exists a constant $c > 0$ such that $\min \sigma(f(\lambda)) \geq c$ for all frequencies $\lambda \in (-\pi, \pi]$, where $\sigma(f(\lambda))$ denotes the spectrum of $f(\cdot)$ at frequency λ, i. e., the set of the eigenvalues of $f(\cdot)$ at frequency λ.

The short memory condition $\sum_{h=-\infty}^{\infty}(1+|h|)^k|\Gamma(h)|_F < \infty$ for some $k \geq 3/2$ in Assumption 1 implies a continuously differentiable spectral density f, which is particularly bounded from below and from above, uniformly for all frequencies $\lambda \in (-\pi, \pi]$. As shown in Meyer and Kreiss (2015), a process fulfilling Assumption 1 possesses the one-sided representations

$$\Phi(L)w_t = \varepsilon_t, \tag{2.3}$$

$$w_t = \Psi(L)\varepsilon_t, \tag{2.4}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a strictly stationary uncorrelated – but not necessarily independent – white noise process with positive definite covariance matrix Σ, $\Phi(z) := I_{m+1} - \sum_{j=1}^{\infty} \Phi_j z^j$ and $\Psi(z) := I_{m+1} + \sum_{j=1}^{\infty} \Psi_j z^j$, with $\sum_{j=1}^{\infty}(1+j)^k|\Phi_j|_F < \infty$ and $\sum_{j=1}^{\infty}(1+j)^k|\Psi_j|_F < \infty$ for the $k \geq 3/2$ from Assumption 1. Moreover, it holds that $\det(\Phi(z)) \neq 0$ and $\det(\Psi(z)) \neq 0$ for all $|z| \leq 1$.

Assumption 2. *The process $\{w_t\}_{t\in\mathbb{Z}}$ has absolutely summable cumulants up to order four. More precisely, we have for all $j = 2, \ldots, 4$ and $\mathbf{a} = [a_1, \ldots, a_j]'$, with $a_1, \ldots, a_j \in \{1, \ldots, m+1\}$, that*

$$\sum_{h_2,\ldots,h_j=-\infty}^{\infty} |cum_{\mathbf{a}}(0, h_2, \ldots, h_j)| < \infty$$

where $cum_{\mathbf{a}}(0, h_2, \ldots, h_j)$ denotes the j-th joint cumulant of $w_{0,a_1}, w_{h_2,a_2}, \ldots, w_{h_j,a_j}$ and $w_{t,i}$ denotes the i-th element of w_t (see, e. g., Brillinger 1981).

Let Ω denote the long-run covariance matrix of $\{w_t\}_{t\in\mathbb{Z}}$, i. e.,

$$\Omega = \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix} = 2\pi f(0) = \sum_{h=-\infty}^{\infty} \Gamma(h) = \Psi(1)\Sigma\Psi(1)'.$$

From $\Sigma > 0$ and $\det(\Psi(1)) \neq 0$ it follows that $\Omega > 0$. In particular, positive definiteness of Ω_{vv} rules out cointegration among the elements of $\{x_t\}_{t\in\mathbb{Z}}$. As typical in the cointegration literature, we assume that $\{w_t\}_{t\in\mathbb{Z}}$ fulfills an invariance principle.

Assumption 3. *Let* $\{w_t\}_{t\in\mathbb{Z}}$ *fulfill*

$$B_T(r) := T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} w_t \xrightarrow{w} B(r) = \Omega^{1/2} W(r), \quad 0 \leq r \leq 1, \tag{2.5}$$

as $T \to \infty$*, where* $W(r) = [W_{u\cdot v}(r), W_v(r)']'$ *is an* $(1+m)$*-dimensional vector of independent standard Brownian motions and*

$$\Omega^{1/2} = \begin{bmatrix} \Omega_{u\cdot v}^{1/2} & \Omega_{uv}(\Omega_{vv}^{-1/2})' \\ 0 & \Omega_{vv}^{1/2} \end{bmatrix},$$

where $\Omega_{u\cdot v} := \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$*, such that* $\Omega^{1/2}(\Omega^{1/2})' = \Omega$*. For later usage we partition* $B(r) = [B_u(r), B_v(r)']'$*.*

We emphasize that Assumption 1 does explicitly not ask for invertibility or causality of the process $\{w_t\}_{t\in\mathbb{Z}}$ with respect to an i.i.d. white noise process, in contrast to the assumptions in related literature, compare, e. g., Park (2002), Chang *et al.* (2006) and Palm *et al.* (2010). In particular, in this paper, the innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ resulting from the representations in (2.3) and (2.4) will generally be uncorrelated, but not necessarily independent. Assumption 2 is of technical nature and satisfied if, e. g., $\{w_t\}_{t\in\mathbb{Z}}$ is α-mixing with strong-mixing coefficients $\alpha(j)$ such that $\sup_{t\in\mathbb{Z}} E(|w_t|_F^{4+\delta}) < \infty$ and $\sum_{j=1}^{\infty} j^2 \alpha(j)^{\delta/(4+\delta)} < \infty$ for some $\delta > 0$, see, e. g., Shao (2010, p. 221). In particular, Assumption 2 requires the existence of fourth moments of $\{w_t\}_{t\in\mathbb{Z}}$. To establish meaningful asymptotic theory, Assumptions 1 and 2 have to be complemented by an invariance principle in Assumption 3. This general formulation of an invariance principle allows for various concepts of choice to quantify weak forms of dependence of the innovation process $\{w_t\}_{t\in\mathbb{Z}}$. These include classical approaches that are sufficient to prove invariance principles – such as several variants of mixing properties, mixingale-type sequences, linear processes[6], or (Bernoulli) shift processes; see Merlevède *et al.* (2006) for an overview – but also other modern approaches that cover the

[6]Including all-pass filters, discussed for univariate times series in, e. g., Andrews *et al.* (2007), that lead to uncorrelated, but dependent white noise processes as well as their multivariate extensions based on non-causal and non-invertible vector-valued time series models as proposed by Lanne and Saikkonen (2013).

general notion of weakly dependent stationary time series discussed in Doukhan and Wintenberger (2007) or physical dependence proposed by Wu (2005) and employed in Wu (2007).

3 Testing General Linear Hypotheses

It is well known that in cointegrating regressions the OLS estimator is consistent despite the fact that the regressors are allowed to be endogenous and the errors are allowed to be serially correlated. However, its limiting distribution is contaminated by second order bias terms, reflecting the correlation structure between the regressors and the errors. This makes the OLS estimator unsuitable for conducting asymptotic inference using (simulated) quantile tables of (non)standard distributions. The literature provides several estimators that allow for such standard asymptotic inference. For our purposes, we choose the IM-OLS approach of Vogelsang and Wagner (2014) as the IM-OLS estimator avoids any tuning parameter choices. They propose to compute first the partial sum of both sides of (2.1), then to add x_t as a regressor to the partial sum regression and finally to estimate the regression coefficients by OLS.[7] That is, by computing the OLS estimator in the augmented partial sum regression

$$S_t^y = S_t^{x\prime}\beta + x_t'\gamma + S_t^u = Z_t'\theta + S_t^u, \tag{3.1}$$

with $S_t^y := \sum_{s=1}^t y_s$, $S_t^x := \sum_{s=1}^t x_s$, $S_t^u := \sum_{s=1}^t u_s$ and the $2m$-dimensional vector $Z_t := [S_t^{x\prime}, x_t']'$, the IM-OLS estimator $\hat{\theta}_{\mathrm{IM}} := [\hat{\beta}_{\mathrm{IM}}', \hat{\gamma}_{\mathrm{IM}}']'$ for $\theta := [\beta', \gamma']'$ in (3.1) is obtained. As shown in Vogelsang and Wagner (2014, Theorem 2) it holds under Assumption 3 that the limiting distribution of $\hat{\theta}_{\mathrm{IM}}$ is given by

$$\begin{bmatrix} T\left(\hat{\beta}_{\mathrm{IM}} - \beta\right) \\ \hat{\gamma}_{\mathrm{IM}} - \Omega_{vv}^{-1}\Omega_{vu} \end{bmatrix} \xrightarrow{w} \Omega_{u\cdot v}^{1/2} \left(\Pi'\right)^{-1} \mathcal{Z}, \tag{3.2}$$

[7] Adding x_t to the partial sum regression serves as an endogeneity correction, which is similar to the leads and lags augmentation in D-OLS estimation. Although similar in spirit, it is considerably simpler as it avoids choosing the numbers of leads and lags.

as $T \to \infty$, where $\Pi := \operatorname{diag}\left(\Omega_{vv}^{1/2}, \Omega_{vv}^{1/2}\right)$ and

$$\mathcal{Z} := \left(\int_0^1 g(r)g(r)'dr\right)^{-1} \int_0^1 [G(1) - G(r)]\, dW_{u\cdot v}(r), \tag{3.3}$$

with $g(r) := [\int_0^r W_v(s)'ds, W_v(r)']'$ and $G(r) := \int_0^r g(s)ds$.[8]

Conditional upon $W_v(r)$, the asymptotic distribution in (3.2) is normal with zero-mean and covariance matrix $\Omega_{u\cdot v}V$, where

$$\begin{aligned} V := (\Pi')^{-1} &\left(\int_0^1 g(r)g(r)'dr\right)^{-1} \left(\int_0^1 [G(1) - G(r)]\,[G(1) - G(r)]'\, dr\right) \\ &\times \left(\int_0^1 g(r)g(r)'dr\right)^{-1} \Pi^{-1}. \end{aligned}$$

Let $A_T := \operatorname{diag}(T^{-1}I_m, I_m)$ and define

$$\hat{V}_T := \left(\sum_{t=1}^T Z_t Z_t'\right)^{-1} \left(\sum_{t=1}^T c_t c_t'\right) \left(\sum_{t=1}^T Z_t Z_t'\right)^{-1},$$

where $c_1 := S_T^Z$ and $c_t := S_T^Z - S_{t-1}^Z$, with $S_t^Z := \sum_{j=1}^t Z_j$, for $t = 2, \ldots, T$. Then,

$$A_T^{-1}\hat{V}_T A_T^{-1} \xrightarrow{w} V, \tag{3.4}$$

as $T \to \infty$, compare Vogelsang and Wagner (2014, Proof of Theorem 3).

The zero mean Gaussian mixture limiting distribution of the IM-OLS estimator in conjunction with (3.4) forms the basis for standard asymptotic inference for the commonly considered Wald-type hypothesis test. To be more precise, for testing $s \leq m$ linearly independent restrictions on $\beta \in \mathbb{R}^m$ in (2.1), we consider the hypotheses

$$\mathrm{H}_0 : R_1\beta = r_0 \quad \text{versus} \quad \mathrm{H}_1 : R_1\beta \neq r_0, \tag{3.5}$$

where $R_1 \in \mathbb{R}^{s\times m}$ has full row rank s and $r_0 \in \mathbb{R}^s$. For deriving the limiting null distribution of the corresponding Wald-type test statistic, it is more convenient to rewrite the null hypothesis in terms of the correct centering parameter for $\hat{\theta}_{\mathrm{IM}}$,

[8] Since both x_t and S_t^u are I(1) processes, all correlation – between $B_v(r)$ and $B_u(r)$ – is soaked up in the long-run population regression vector $\Omega_{vv}^{-1}\Omega_{vu}$. Therefore, the correct centering parameter for $\hat{\gamma}_{\mathrm{IM}}$ in case of endogeneity is $\Omega_{vv}^{-1}\Omega_{vu}$ rather than the population value $\gamma = 0$. For more details see Vogelsang and Wagner (2014, p. 746).

given by $[\beta', (\Omega_{vv}^{-1}\Omega_{vu})']'$. To this end, we define $R_2 \coloneqq [R_1, 0_{s\times m}] \in \mathbb{R}^{s\times 2m}$ such that the null hypothesis in (3.5) reads as $R_1\beta = R_2[\beta', (\Omega_{vv}^{-1}\Omega_{vu})']'$.[9]

Under Assumption 3 it follows from Vogelsang and Wagner (2014, Theorem 3) that the limiting distribution of the Wald-type test statistic

$$\tau_{\mathrm{IM}} \coloneqq \left(R_2\hat{\theta}_{\mathrm{IM}} - r_0\right)' \left[R_2\hat{V}_T R_2'\right]^{-1} \left(R_2\hat{\theta}_{\mathrm{IM}} - r_0\right)$$

converges under the null hypothesis in distribution to

$$\mathcal{G}_\Omega \coloneqq \left(R_2\Omega_{u\cdot v}^{1/2} \left(\Pi'\right)^{-1} \mathcal{Z}\right)' \left(R_2 V R_2'\right)^{-1} \left(R_2\Omega_{u\cdot v}^{1/2} \left(\Pi'\right)^{-1} \mathcal{Z}\right) \overset{d}{=} \Omega_{u\cdot v}\chi_s^2, \tag{3.6}$$

as $T \to \infty$, where $\mathcal{Z}$ is defined in (3.3) and χ_s^2 denotes a chi-square distribution with s degrees of freedom.[10]

The limiting null distribution of τ_{IM} is contaminated by a nuisance parameter, $\Omega_{u\cdot v}$. The presence of the long-run variance parameter makes the limiting distribution highly case dependent and thus infeasible for inference based on tabulated critical values.[11] To remedy this problem, the literature suggests to plug-in a consistent estimator of $\Omega_{u\cdot v}$, $\hat{\Omega}_{u\cdot v}$ say, such that the nuisance parameter is scaled out in the limit. For this purpose, we denote

$$\tau_{\mathrm{IM}}(\kappa) \coloneqq \left(R_2\hat{\theta}_{\mathrm{IM}} - r_0\right)' \left[R_2\kappa\hat{V}_T R_2'\right]^{-1} \left(R_2\hat{\theta}_{\mathrm{IM}} - r_0\right) \tag{3.7}$$

such that $\tau_{\mathrm{IM}}(1) = \tau_{\mathrm{IM}}$. Hence, the literature typically considers Wald-type test statistics of the form

$$\tau_{\mathrm{IM}}(\hat{\Omega}_{u\cdot v}) = \left(R_2\hat{\theta}_{\mathrm{IM}} - r_0\right)' \left[R_2\hat{\Omega}_{u\cdot v}\hat{V}_T R_2'\right]^{-1} \left(R_2\hat{\theta}_{\mathrm{IM}} - r_0\right), \tag{3.8}$$

which converge under the null hypothesis in distribution to

$$\mathcal{G} \coloneqq \left(R_2 \left(\Pi'\right)^{-1} \mathcal{Z}\right)' \left(R_2 V R_2'\right)^{-1} \left(R_2 \left(\Pi'\right)^{-1} \mathcal{Z}\right) \overset{d}{=} \chi_s^2,$$

[9] Note that the auxiliary coefficient vector γ is not restricted under the null hypothesis and, in particular, $\Omega_{vv}^{-1}\Omega_{vu}$ does not have to be estimated.

[10] In practical applications it might be more convenient to express this – and the following – test statistic(s) in terms of $\hat{\beta}_{\mathrm{IM}}$ and R_1, only. This can be achieved by noting that $R_2\hat{\theta}_{\mathrm{IM}} = R_1\hat{\beta}_{\mathrm{IM}}$ and $R_2\hat{V}_T R_2' = R_1\hat{V}_T^{(1,1)}R_1'$, where $\hat{V}_T^{(1,1)}$ denotes the upper left $(m\times m)$-dimensional block of the $(2m\times 2m)$-dimensional matrix $\hat{V}_T$.

[11] Analogous results also hold for the Wald-type tests based on the D-OLS, FM-OLS and CCR estimators.

as $T \to \infty$. As $\mathcal{G}$ is nuisance parameter free, it allows for standard asymptotic inference based on tabulated critical values. However, estimation of $\Omega_{u\cdot v}$ is cumbersome and typically based on non-parametric kernel estimators that depend on kernel and bandwidth choices, see, e. g., Andrews (1991), Newey and West (1994) and Jansson (2002) for details. As these tuning parameter choices affect the finite sample distribution of $\tau_{\mathrm{IM}}(\hat{\Omega}_{u\cdot v})$ but are completely ignored in the conventional asymptotic framework, corresponding tests usually have large size distortions, especially when the level of endogeneity and/or error serial correlation is strong or sample size is small. Vogelsang and Wagner (2014) suggest to use fixed-b critical values to capture the effects of tuning parameter choices, but their simulation results reveal that in small to medium samples test performance, which still seems to be sensitive to the choice of b, worsens as endogeneity and/or error serial correlation increases.

To avoid any tuning parameter choices, we propose a novel test statistic based on a self-normalization approach. Instead of plugging-in a consistent estimator $\hat{\Omega}_{u\cdot v}$ of $\Omega_{u\cdot v}$ in (3.7) to get (3.8), we insert a quantity that is directly constructed from the data and asymptotically proportional to $\Omega_{u\cdot v}$. To this end, we define the OLS residuals in the augmented partial sum regression given in (3.1) as $\hat{S}_t^u := S_t^y - Z_t'\hat{\theta}_{\mathrm{IM}}$, $t = 1, \ldots, T$. For $t = 2, \ldots, T$ let $\Delta\hat{S}_t^u := \hat{S}_t^u - \hat{S}_{t-1}^u$ denote the first difference of the residuals and define the *self-normalizer* as

$$\hat{\eta}_T := T^{-2}\sum_{t=2}^{T}\left(\sum_{s=2}^{t}\Delta\hat{S}_s^u\right)^2.$$

The self-normalizer is similar to the test statistic of Shin (1994) to test the null hypothesis of cointegration in (2.1). In contrast to his test statistic, however, $\hat{\eta}_T$ is a) not divided by an estimator of the long-run variance $\Omega_{u\cdot v}$ and b) based on the first differences of the IM-OLS residuals rather than the IM-OLS residuals itself, which accounts for taking partial sums prior to estimation. We introduce our *self-normalized test statistic* as

$$\tau_{\mathrm{IM}}(\hat{\eta}_T) = \left(R_2\hat{\theta}_{\mathrm{IM}} - r_0\right)'\left[R_2\hat{\eta}_T\hat{V}_T R_2'\right]^{-1}\left(R_2\hat{\theta}_{\mathrm{IM}} - r_0\right).$$

Its limiting null distribution is given in the following proposition.

Proposition 1. *Let* $(y_t)_{t=1}^T$ *and* $(x_t)_{t=1}^T$ *be generated by* (2.1) *and* (2.2)*, respectively and let* $\{w_t\}_{t\in\mathbb{Z}}$ *satisfy Assumption 3. Then it holds under the null hypothesis* *given in* (3.5) *that*

$$\tau_{IM}(\hat{\eta}_T) \xrightarrow{w} \mathcal{G}_{SN} := \frac{\left(R_2\left(\Pi'\right)^{-1}\mathcal{Z}\right)'\left(R_2 V R_2'\right)^{-1}\left(R_2\left(\Pi'\right)^{-1}\mathcal{Z}\right)}{\int_0^1 \left(W_{u\cdot v}(r) - g(r)'\mathcal{Z}\right)^2 dr}$$
$$= \frac{\chi_s^2}{\int_0^1 \left(W_{u\cdot v}(r) - g(r)'\mathcal{Z}\right)^2 dr}, \tag{3.9}$$

as $T \to \infty$*, where the* χ_s^2 *distributed random variable in the numerator is correlated with the denominator* $\int_0^1 \left(W_{u\cdot v}(r) - g(r)'\mathcal{Z}\right)^2 dr$ *as both are driven by* $W_{u\cdot v}$ *and* $W_v(r)$*.*

The limiting null distribution of $\tau_{\mathrm{IM}}(\hat{\eta}_T)$ is nonstandard but free of nuisance parameters and only depends on the number of restrictions under the null hypothesis and the number of integrated regressors. Although the χ_s^2 distributed random variable in the numerator is correlated with the denominator, simulating critical values is straightforward. To this end, we approximate standard Brownian motions by normalized sums of 10,000 i.i.d. standard normal random variables and approximate the corresponding integrals accordingly. The fact that the numerator is invariant to Π and only depends on the number of linearly independent restrictions under the null hypothesis, justifies to set $\Pi = I_{2m}$ and $R_2 = [I_s, 0_{s\times(2m-s)}]$.[12] We tabulate critical values based on 10,000 replications for various values of m and s in Table 1.[13] As the test statistic $\tau_{\mathrm{IM}}(\hat{\eta}_T)$ does not depend on any tuning

Table 1: Asymptotic critical values for $\tau_{\mathrm{IM}}(\hat{\eta}_T)$

	$m=1$	$m=2$		$m=3$			$m=4$			
%	$s=1$	$s=1$	$s=2$	$s=1$	$s=2$	$s=3$	$s=1$	$s=2$	$s=3$	$s=4$
90.0	36.63	66.33	122.32	94.04	172.00	240.58	131.68	232.77	318.25	402.61
95.0	56.58	96.51	167.23	140.69	231.79	313.46	189.15	309.06	407.17	510.60
97.5	79.24	131.79	216.99	191.68	290.47	390.38	256.38	390.07	504.08	630.19
99.0	120.10	189.69	286.97	266.16	375.30	494.00	355.25	505.21	645.89	767.61

Notes: m is the number of integrated regressors and $s \leq m$ is the number of linearly independent restrictions under the null hypothesis.

parameter choices, inference based on simulated critical values is completely tun-

[12] We emphasize that the dependence between numerator and denominator in (3.9) implies that simply drawing from a chi-square distribution for the numerator is invalid.

[13] The presence of deterministic regressors in (2.1) affects the limiting null distribution of $\tau_{\mathrm{IM}}(\hat{\eta}_T)$. Asymptotic critical values in the presence of an intercept and polynomial time trends are tabulated in Table 4 in the Online Appendix.

ing parameter free. However, the asymptotically valid critical values might not serve as good approximations of the quantiles of the test statistic's finite sample distribution. Therefore, in the next section, we propose a VAR sieve based resampling procedure to obtain critical values that might be more accurate in finite samples.

4 Bootstrap Inference

4.1 Bootstrap Method

For asymptotic considerations, the representation given in (2.3) suggests to approximate $\{w_t\}_{t\in\mathbb{Z}} = \{[u_t, v_t']'\}_{t\in\mathbb{Z}}$ by a sequence of VAR processes with increasing order $q = q_T \to \infty$ as $T \to \infty$. These VAR approximations can be bootstrapped using the autoregressive sieve bootstrap method, first investigated by Kreiss (1988), Kreiss (1992) and Bühlmann (1997) for univariate time series and extended to the multivariate case by Paparoditis (1996). More recently, by extending the results of Kreiss *et al.* (2011) to the multivariate case, Meyer and Kreiss (2015) study the range of validity of the VAR sieve bootstrap in situations that go beyond the typically assumed setup of causal linear processes.

Applying the VAR sieve bootstrap in our context requires to fit a finite order VAR to $w_t = [u_t, v_t']'$, $t = 1, \ldots, T$. However, while $v_t = x_t - x_{t-1}$ is simply given by the first difference of x_t, the regression error in (2.1), u_t, is unknown. We therefore fit a finite order VAR to $\hat{w}_t := [\hat{u}_t, v_t']'$, $t = 1, \ldots, T$, instead, where $\hat{u}_t := y_t - x_t'\hat{\beta}_{\text{IM}}$ denotes the IM-OLS residual in (2.1). In the following, let $\hat{\Phi}_1(q), \ldots, \hat{\Phi}_q(q)$ denote the solution of the sample Yule-Walker equations in the regression of $\hat{w}_t$ on $\hat{w}_{t-1}, \ldots, \hat{w}_{t-q}$, $t = q+1, \ldots, T$, and denote the corresponding residuals by $\hat{\varepsilon}_t(q) := \hat{w}_t - \sum_{j=1}^{q} \hat{\Phi}_j(q)\hat{w}_{t-j}$, $t = q+1, \ldots, T$.[14]

Bootstrap Scheme:

Step 1) Obtain the bootstrap sample $(\varepsilon_t^*)_{t=1}^T$ by randomly drawing T times with replacement from the centered residuals $\left(\hat{\varepsilon}_t(q) - \bar{\hat{\varepsilon}}_T(q)\right)_{t=q+1}^T$, where $\bar{\hat{\varepsilon}}_T(q) :=$

[14]It is well known that any finite order VAR estimated by the Yule-Walker estimator is causal and invertible in finite samples.

$(T-q)^{-1}\sum_{t=q+1}^{T}\hat{\varepsilon}_t(q)$, and construct $(w_t^*)_{t=1}^T$ recursively as

$$w_t^* = \hat{\Phi}_1(q)w_{t-1}^* + \ldots + \hat{\Phi}_q(q)w_{t-q}^* + \varepsilon_t^*,$$

given initial values $w_{1-q}^*, \ldots, w_0^*$.[15] Partition $w_t^* = [u_t^*, v_t^{*\prime}]'$ analogously to w_t and define $x_t^* := \sum_{s=1}^{t} v_s^*$.

Step 2) To generate data under the null hypothesis given in (3.5), define

$$y_t^* := x_t^{*\prime}\hat{\beta}_{\mathrm{IM}}^r + u_t^*,$$

with $\hat{\beta}_{\mathrm{IM}}^r$ such that $R_1\hat{\beta}_{\mathrm{IM}}^r = r_0$. That is, $\hat{\beta}_{\mathrm{IM}}^r$ is defined as the vector of the first m elements of

$$\hat{\theta}_{\mathrm{IM}} - \left(\sum_{t=1}^{T} Z_t Z_t'\right)^{-1} R_2' \left[R_2 \left(\sum_{t=1}^{T} Z_t Z_t'\right)^{-1} R_2'\right]^{-1} \left(R_2\hat{\theta}_{\mathrm{IM}} - r_0\right),$$

which in turn is the restricted OLS estimator of θ in (3.1) under the restriction $R_1\beta = R_2[\beta', (\Omega_{vv}^{-1}\Omega_{vu})']' = r_0$.

Step 3) Compute the OLS estimator in the bootstrap augmented partial sum regression

$$S_t^{y^*} = S_t^{x^*\prime}\beta + x_t^{*\prime}\gamma + S_t^{u^*} = Z_t^{*\prime}\theta + S_t^{u^*}, \tag{4.1}$$

where $S_t^{y^*} := \sum_{s=1}^{t} y_s^*$, $S_t^{x^*} := \sum_{s=1}^{t} x_s^*$, $S_t^{u^*} := \sum_{s=1}^{t} u_s^*$ and $Z_t^* := [S_t^{x^*\prime}, x_t^{*\prime}]'$, to obtain the bootstrap IM-OLS estimator $\hat{\theta}_{\mathrm{IM}}^* := [\hat{\beta}_{\mathrm{IM}}^{*\prime}, \hat{\gamma}_{\mathrm{IM}}^{*\prime}]'$ of θ in (4.1). Define the corresponding residuals as $\hat{S}_t^{u^*} := S_t^{y^*} - Z_t^{*\prime}\hat{\theta}_{\mathrm{IM}}^*$, $t = 1, \ldots, T$ and let $\Delta\hat{S}_t^{u^*} := \hat{S}_t^{u^*} - \hat{S}_{t-1}^{u^*}$, $t = 2, \ldots, T$, denote their first differences. Define

$$\hat{V}_T^* := \left(\sum_{t=1}^{T} Z_t^* Z_t^{*\prime}\right)^{-1} \left(\sum_{t=1}^{T} c_t^* c_t^{*\prime}\right) \left(\sum_{t=1}^{T} Z_t^* Z_t^{*\prime}\right)^{-1},$$

where $c_1^* := S_T^{Z^*}$ and $c_t^* := S_T^{Z^*} - S_{t-1}^{Z^*}$, with $S_t^{Z^*} := \sum_{j=1}^{t} Z_j^*$, for $t = 2, \ldots, T$.

[15] Though irrelevant for developing asymptotic theory, it is advantageous in practical applications to eliminate the dependencies of the results on the initial values of w_s^*, $1-q \le s \le 0$, to obtain a stationary sample. To this end, we suggest to generate a sufficiently large number of w_t^*'s and keep the last $T+q$ of them, only.

Step 4) Define the bootstrap version of the test statistic $\tau_{\mathrm{IM}}(\hat{\eta}_T)$ as

$$\tau^*_{\mathrm{IM}}(\hat{\eta}^*_T) := \left(R_2\hat{\theta}^*_{\mathrm{IM}} - r_0\right)' \left[R_2\hat{\eta}^*_T\hat{V}^*_T R_2'\right]^{-1} \left(R_2\hat{\theta}^*_{\mathrm{IM}} - r_0\right),$$

where

$$\hat{\eta}^*_T := T^{-2}\sum_{t=2}^{T}\left(\sum_{s=2}^{t}\Delta\hat{S}^{u^*}_s\right)^2.$$

Step 5) Let α denote the desired nominal size of the test. Repeat Steps 1) to 4) B times, where B is large and $(B+1)(1-\alpha)$ is an integer, to obtain B realizations of the bootstrap test statistic, $\tau^*_{\mathrm{IM}}(\hat{\eta}^*_T)$. Reject the null hypothesis given in (3.5), if the test statistic based on the original observations, $\tau_{\mathrm{IM}}(\hat{\eta}_T)$, is greater than the $(B+1)(1-\alpha)$-th largest realization of the bootstrap test statistic.

Note that we impose the null hypothesis when generating the bootstrap sample $(y^*_t)_{t=1}^T$ by using the restricted IM-OLS estimator, $\hat{\beta}^r_{\mathrm{IM}}$, instead of the unrestricted IM-OLS estimator, $\hat{\beta}_{\mathrm{IM}}$. However, we use the unrestricted residuals, $\hat{u}_t = y_t - x_t'\hat{\beta}_{\mathrm{IM}}$, instead of the restricted residuals, $y_t - x_t'\hat{\beta}^r_{\mathrm{IM}}$, in the definition of $\hat{w}_t$. It is advantageous to use the restricted residuals in the definition of $\hat{w}_t$ when the null hypothesis is true. However, the empirical distribution function of the restricted residuals will generally fail to mimic the population distribution under the alternative, which leads to a loss of power of the test compared to the case where $\hat{w}_t$ is based on the unrestricted residuals, compare, e. g., van Giersbergen and Kiviet (2002) and Paparoditis and Politis (2005) for a detailed discussion.

4.2 Bootstrap Consistency

In this subsection we show the asymptotic validity of the testing approach proposed in the previous subsection. To this end, we first prove an invariance principle result to hold for the bootstrap innovations $\{\varepsilon^*_t\}_{t\in\mathbb{Z}}$, which then enables us to show that an invariance principle result also holds for $\{w^*_t\}_{t\in\mathbb{Z}}$. For the sieve bootstrap based on $(\hat{w}_t)_{t=1}^T$ to work properly, it is necessary *and* sufficient that it mimics the second-order dependence structure of $\{w_t\}_{t\in\mathbb{Z}}$. This stems from the fact that the dependence structure in the limiting null distribution of $\tau_{\mathrm{IM}}(\hat{\eta}_T)$ depends only on the second moments of $\{w_t\}_{t\in\mathbb{Z}}$ and, with respect to second moments, independence and uncorrelatedness are indistinguishable. Hence, we approximate $(\hat{w}_t)_{t=1}^T$

by a sequence of VAR processes with increasing order $q_T \to \infty$ as $T \to \infty$. We impose the following condition on the rate at which q_T goes to infinity.

Assumption 4. *Let $q_T \to \infty$ and $q_T = O((T/\ln(T))^{1/3})$ as $T \to \infty$.*

For notational brevity, we suppress the dependence of q_T on T in the following, i. e., we simply write q instead of q_T. We are now in the position to prove the following invariance principle for the bootstrap innovations.

Proposition 2. *It holds under Assumptions 1, 2 and 4 that*

$$W_T^*(r) := T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} \varepsilon_t^* \xrightarrow{w^*} \Sigma^{1/2} W(r), \quad 0 \leq r \leq 1, \quad \textit{in } \mathbb{P},$$

as $T \to \infty$, with $\Sigma^{1/2}(\Sigma^{1/2})' = \Sigma$.

The preceding result together with the Beveridge-Nelson decomposition (cf. Phillips and Solo 1992) allows us to prove an invariance principle for $\{w_t^*\}_{t \in \mathbb{Z}}$.

Theorem 1. *It holds under Assumptions 1, 2 and 4 that*

$$B_T^*(r) := T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} w_t^* \xrightarrow{w^*} \Psi(1)\Sigma^{1/2} W(r), \quad 0 \leq r \leq 1, \quad \textit{in } \mathbb{P},$$

as $T \to \infty$. If, in addition, Assumption 3 is fulfilled, it holds that $\Psi(1)\Sigma^{1/2}W(r) \stackrel{d}{=} B(r)$, with $B(r)$ introduced in (2.5)*. In particular, $\Psi(1)\Sigma^{1/2}W(r)$ has covariance matrix Ω.*

The invariance principle for $\{w_t^*\}_{t \in \mathbb{Z}}$ is the key ingredient in showing that the bootstrap IM-OLS estimator $\hat{\theta}_{\mathrm{IM}}^*$ in (4.1) has, conditional on the original sample, the same limiting distribution as the IM-OLS estimator $\hat{\theta}_{\mathrm{IM}}$ in (3.1).

Theorem 2. *Let the bootstrap quantities be generated as described in Section 4.1. Then it holds under Assumptions 1, 2, 3 and 4 that*

$$\begin{bmatrix} T\left(\hat{\beta}_{IM}^* - \hat{\beta}_{IM}^r\right) \\ \hat{\gamma}_{IM}^* - \Omega_{vv}^{-1}\Omega_{vu} \end{bmatrix} \xrightarrow{w^*} \Omega_{u \cdot v}^{1/2} \left(\Pi'\right)^{-1} \mathcal{Z} \quad \textit{in } \mathbb{P}, \tag{4.2}$$

with $\mathcal{Z}$ defined in (3.3)*. Moreover, it holds under the null hypothesis given in* (3.5) *that*

$$\tau_{IM}^*(\hat{\eta}_T^*) \xrightarrow{w^*} \mathcal{G}_{SN} \quad \textit{in } \mathbb{P},$$

with $\mathcal{G}_{SN}$ defined in (3.9).

Theorem 2 shows that the VAR sieve bootstrap, described in Section 4.1, is consistently estimating the limiting distributions of the IM-OLS estimator and the self-normalized test statistic based upon it.[16]

5 Finite Sample Performance

We generate data according to (2.1) and (2.2) with $m = 2$ regressors, i. e.,

$$\begin{aligned} y_t &= x_{1t}\beta_1 + x_{2t}\beta_2 + u_t, \\ x_{it} &= x_{i,t-1} + v_{it}, \quad x_{i0} = 0, \quad i = 1, 2, \end{aligned}$$

for $t = 1, \ldots, T$. The regression errors and the innovations for the stochastic regressors are generated similarly as in Vogelsang and Wagner (2014), i. e.,

$$\begin{aligned} u_t &= \rho_1 u_{t-1} + \varepsilon_t + \rho_2(\nu_{1t} + \nu_{2t}), \quad u_{-100} = 0, \\ v_{it} &= \nu_{it} + 0.5\nu_{i,t-1}, \quad \nu_{i,-100} = 0, \quad i = 1, 2, \end{aligned}$$

for $t = -99, \ldots, 0, 1, \ldots, T$, where ε_t, ν_{1t} and ν_{2t} are i.i.d. standard normal random variables. The period $t = -99, \ldots, 0$ serves as a burn-in period to ensure stationarity. The parameters ρ_1 and ρ_2 control the level of serial correlation in the regression errors and the extent of endogeneity, respectively. The regression parameters are chosen as $\beta_1 = \beta_2 = 1$ and the order $1 \leq q \leq \lfloor T^{1/3} \rfloor =: q_{\max}$ of the VAR sieve is chosen as the one that minimizes the Akaike information criterion (AIC) computed on the evaluation period $t = q_{\max} + 1, \ldots, T$, as suggested by Kilian and Lütkepohl (2017, p. 56).[17] We consider results for $T \in \{50, 100, 200\}$ and $\rho_1 = \rho_2 \in \{0, 0.3, 0.6, 0.9\}$. In all cases, the number of Monte Carlo and bootstrap replications is 5,000 and 1,999, respectively.

We first briefly compare the IM-OLS estimator with the D-OLS and FM-OLS estimators in terms of bias and root mean squared error (RMSE). Implementing

[16]Following the bootstrap literature, it would be more common to use $\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$ rather than $\Omega_{vv}^{-1}\Omega_{vu}$ as the centering coefficient vector in (4.2). However, both versions lead to the same limiting distribution. As (an estimate of) $\Omega_{vv}^{-1}\Omega_{vu}$ is not needed to construct the bootstrap samples, we use $\Omega_{vv}^{-1}\Omega_{vu}$ as the centering coefficient vector in (4.2) to stress that estimating Ω is not necessary for our procedure.

[17]Results based on the Bayesian information criterion (BIC) are similar and therefore not reported.

the D-OLS estimator requires choosing the numbers of leads and lags of the first differences of the integrated regressors. To this end, we use the Bayesian information criterion (BIC) analyzed in Choi and Kurozumi (2012) as it appears to be the most successful criterion – among those considered by the authors – in reducing the mean squared error of the D-OLS estimator. The FM-OLS estimator is based on an estimate of the long-run covariance matrix of $\{[u_t, v_t']'\}_{t\in\mathbb{Z}}$, which is typically obtained by kernel estimators, which require kernel and bandwidth choices. We analyze the results for the Bartlett and the Quadratic Spectral (QS) kernel together with the corresponding data-dependent bandwidth selection rules of Andrews (1991). The results are displayed in Table 3 in Appendix A. The table shows that, although asymptotically unbiased, the D-, FM- and IM-OLS estimators are biased in finite samples. Bias and RMSE increase in $\rho_1 = \rho_2$ and decrease in T. Often, the IM-OLS estimator has the smallest bias, followed by the D-OLS estimator, but the IM-OLS estimator has in genral a slightly larger RMSE than the FM- and D-OLS estimators, which coincides with the findings in Vogelsang and Wagner (2014). For $T = 50$, however, the D-OLS estimator performs poorly, which is mainly attributed to the tendency of the BIC (and the AIC) to suggest a relatively large number of leads and lags in this case. In total, the IM-OLS estimator seems to be an appropriate initial estimator for the VAR sieve procedure.

We now turn to the performance of the self-normalized (bootstrap) test under the null hypothesis $\mathrm{H}_0 : \beta_1 = 1,\ \beta_2 = 1$. The results are benchmarked against the commonly used Wald-type tests based on the D-, FM- and IM-OLS estimators, in the following denoted by $\tau_{\mathrm{D}}(\hat{\Omega}_{u\cdot v}), \tau_{\mathrm{FM}}(\hat{\Omega}_{u\cdot v})$ and $\tau_{\mathrm{IM}}(\hat{\Omega}_{u\cdot v})$, respectively, which rely on a kernel estimator of a long-run variance parameter, $\Omega_{u\cdot v}$. We again analyze results for the Bartlett and the QS kernel together with the corresponding data-dependent bandwidth selection rules of Andrews (1991). In addition, we compare the self-normalized test, $\tau_{\mathrm{IM}}(\hat{\eta}_T)$, and the self-normalized bootstrap test, $\tau^*_{\mathrm{IM}}(\hat{\eta}_T)$, with the IM-OLS based Wald-type test that relies on bootstrap rather than chi-square critical values, $\tau^*_{\mathrm{IM}}(\hat{\Omega}_{u\cdot v})$, in the following referred to as the IM-OLS based Wald-type bootstrap test. We also analyze the results of the IM-OLS based test statistic that is neither divided by $\hat{\Omega}_{u\cdot v}$ nor by $\hat{\eta}_T$, in conjunction with corresponding bootstrap critical values, $\tau^*_{\mathrm{IM}}(1)$.[18]

[18]Bootstrap critical values for the IM-OLS based Wald-type statistic are obtained using the scheme described in Section 4.1, with obvious modifications. In particular, note that in each

Figure 1 displays the empirical null rejection probabilities – in the following referred to as (empirical) sizes – of the tests for $T = 100$, $\rho_1 = \rho_2 \in \{0.6, 0.9\}$ and nominal sizes α in the range $[0.01, 0.20]$, using 20 values on a grid with mesh size 0.01. Whenever necessary, long-run covariance matrix estimation is based on the Bartlett kernel. The following four key observations emerge: First, the Wald-type test based on the IM-OLS estimator clearly outperforms the Wald-type tests based on the D- and FM-OLS estimators, but is still severely size distorted. Second, empirical sizes of the commonly employed tests are generally much larger than those of the tests based on the self-normalized test statistic. Third, using bootstrap critical values instead of asymptotically valid critical values reduces size distortions of the self-normalized test and the IM-OLS based Wald-type test considerably when the level of error serial correlation and/or endogeneity is large.[19] Finally, the self-normalized bootstrap test outperforms the IM-OLS based Wald-type bootstrap test, with the performance advantages being more pronounced the larger the level of error serial correlation and/or endogeneity. The results are similar for the QS kernel and the other choices of T and $\rho_1 = \rho_2$.

To analyze the effects of different tuning parameter choices, the sample size and the extent of error serial correlation and endogeneity in more detail, we now focus on the results for $\alpha = 0.05$. Corresponding results for $T \in \{50, 100, 200\}$ and $\rho_1 = \rho_2 \in \{0, 0.3, 0.6, 0.9\}$ are displayed in Table 2. The table shows that the test based on the self-normalized test statistic that relies on asymptotically valid critical values clearly outperforms the commonly used tests in terms of size. In particular, it yields sizes that are close to the nominal level unless $\rho_1 = \rho_2$ is large relative to T. Using bootstrap critical values improves the performance of the test based on the self-normalized test statistic, especially for large values of $\rho_1 = \rho_2$. The self-normalized bootstrap test also outperforms the IM-OLS based Wald-type bootstrap test and the bootstrap test based on the test statistic that is neither divided by $\hat{\Omega}_{u \cdot v}$ nor by $\hat{\eta}_T$, especially when $\rho_1 = \rho_2$ is large. Table 2 further shows that the results are robust to different kernel choices. In addition, Figure 2 displays the effect of different bandwidth choices $(1, 2, 4, \ldots, T)$ on the sizes of the

bootstrap iteration, the estimate of the long-run variance $\Omega_{u \cdot v}$ is based on the corresponding bootstrap sample rather than on the original sample, which has turned out to be beneficial. Bootstrap critical values for the test statistic $\tau_{\mathrm{IM}}(1)$ are obtained similarly.

[19] For a given nominal size α, the size distortion of a test is defined as the difference between its empirical size and α, which corresponds to the vertical distance of the empirical size curve to the 45-degree reference line in Figure 1.

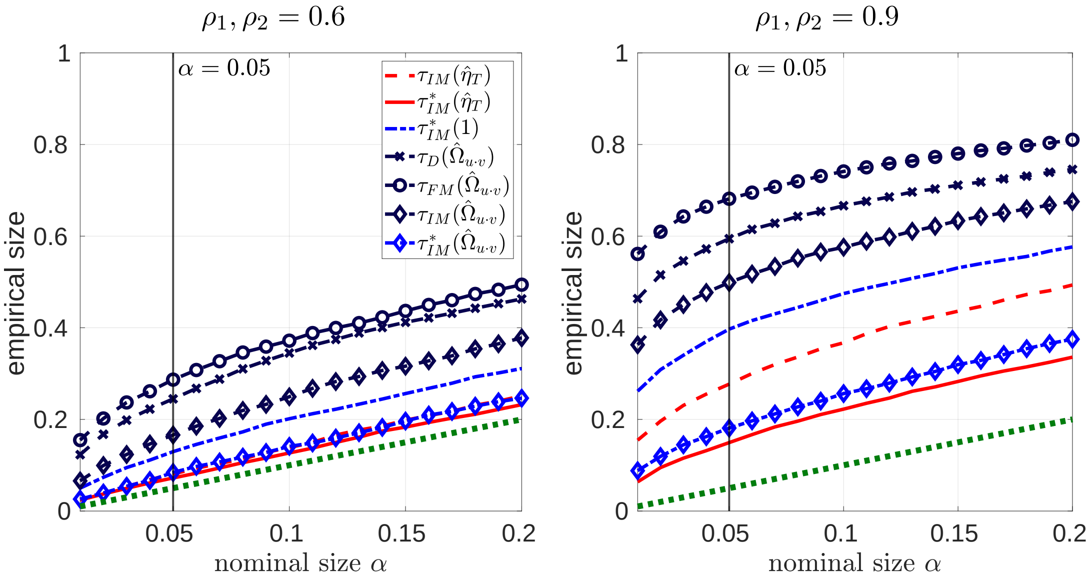

Figure 1: Empirical versus nominal size of the tests for $H_0 : \beta_1 = 1, \ \beta_2 = 1$ for $T = 100$. Notes: Superscript "$*$" signifies the use of bootstrap critical values. The asymptotically valid critical values for the test based on the self-normalized test statistic are simulated as described in Section 3. The dotted green 45-degree line is the target.

three commonly used Wald-type tests and the IM-OLS based Wald-type bootstrap test. It shows that the self-normalized tests outperform the three commonly used tests *for all bandwidth choices*. Figure 2 further shows that sizes of the IM-OLS based Wald-type bootstrap test decrease as the bandwidth increases, especially for the QS kernel. However, as additional simulation results (not reported) reveal, this is accompanied by huge power losses.[20] This indicates that simply replacing asymptotically valid critical values by bootstrap critical values does not capture the finite sample effects of kernel and bandwidth choices completely and thus shows the need for self-normalized test statistics.

Finally, we analyze the properties of the self-normalized (bootstrap) test under deviations from the null hypothesis. To this end, we generate data for $\beta_1 = \beta_2 \in (1, 1.4]$ using 20 values on a grid with mesh size 0.02. The power of the self-normalized tests is again benchmarked against the commonly used Wald-type tests based on the D-, FM- and IM-OLS estimators and against the IM-OLS based Wald-type bootstrap test. Results are displayed in Figure 3. The large differences in sizes (i. e., under the null hypothesis), however, make a meaningful comparison

[20] In this setting, the procedure of Andrews (1991) yields relatively small bandwidths.

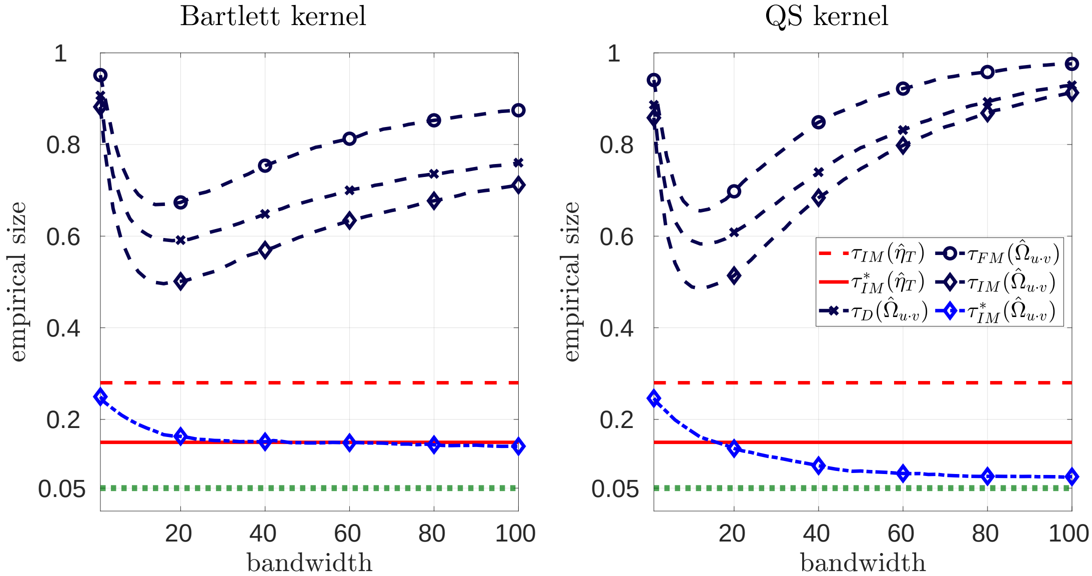

Figure 2: Empirical sizes of the tests for $\mathrm{H}_0 : \beta_1 = 1, \ \beta_2 = 1$ at 5% level for $\rho_1, \rho_2 = 0.9$ and $T = 100$. Notes: Superscript "$*$" signifies the use of bootstrap critical values. As the self-normalized test statistic does not depend on kernel and bandwidth choices, the corresponding curves are constant within and equal across subfigures. The dotted green target line displays the nominal size of the tests.

of the performances under the alternative difficult. To enable a "fair" comparison, we – similar as, e. g., Cavaliere *et al.* (2015, p. 826) – first simulate under the null hypothesis and record for each test the nominal size $\tilde{\alpha}$ that yields an empirical size equal to the desired $\alpha = 0.05$. We then use critical values corresponding to $\tilde{\alpha}$ in the simulations under the alternative hypotheses. Therefore, by construction, all curves in Figure 3 start at $\alpha = 0.05$ for $\beta_1, \beta_2 = 1$. The following three main observations emerge: First, the tests based on the IM-OLS estimator have slightly smaller power than the tests based on the D- and FM-OLS estimators, which coincides with the findings in Vogelsang and Wagner (2014). Second, replacing the long-run variance estimator in the IM-OLS based test statistic with the self-normalizer decreases the power of the test slightly. Third, the self-normalized test based on bootstrap critical values has similar but marginally smaller power than the self-normalized test based on the asymptotically valid critical values but this difference vanishes as the sample size increases.

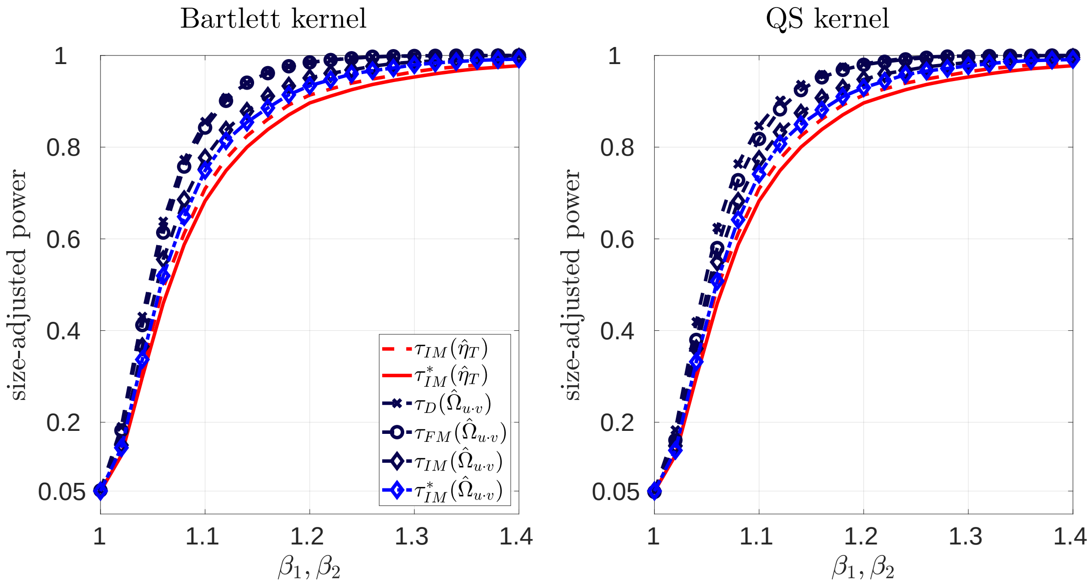

Figure 3: Size-adjusted power of the tests for $\mathrm{H}_0 : \beta_1 = 1,\ \beta_2 = 1$ at 5% level for $\rho_1, \rho_2 = 0.6$ and $T = 100$. Notes: Superscript "$*$" signifies the use of bootstrap critical values. As the self-normalized test statistic does not depend on kernel (and bandwidth) choices, the corresponding curves are equal across subfigures.

Table 2: Empirical sizes of the tests for $\mathrm{H}_0 : \beta_1 = 1,\ \beta_2 = 1$ at 5% level

		Self-normalization			Bartlett kernel				QS kernel			
	ρ_1, ρ_2	$\tau_{\mathrm{IM}}(\hat{\eta}_T)$	$\tau^*_{\mathrm{IM}}(\hat{\eta}_T)$	$\tau^*_{\mathrm{IM}}(1)$	$\tau_{\mathrm{D}}(\hat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{FM}}(\hat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{IM}}(\hat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{IM}}(\hat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{D}}(\hat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{FM}}(\hat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{IM}}(\hat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{IM}}(\hat{\Omega}_{u\cdot v})$
$T = 50$	0	0.03	0.08	0.12	0.74	0.18	0.13	0.09	0.77	0.25	0.19	0.07
	0.3	0.05	0.08	0.16	0.75	0.23	0.17	0.10	0.76	0.28	0.19	0.08
	0.6	0.11	0.10	0.23	0.73	0.40	0.25	0.12	0.72	0.43	0.25	0.11
	0.9	0.40	0.23	0.58	0.76	0.78	0.67	0.29	0.79	0.82	0.71	0.32
$T = 100$	0	0.03	0.07	0.08	0.12	0.13	0.10	0.06	0.15	0.16	0.11	0.05
	0.3	0.05	0.07	0.09	0.15	0.17	0.13	0.07	0.15	0.18	0.12	0.06
	0.6	0.08	0.07	0.13	0.24	0.29	0.17	0.08	0.23	0.28	0.15	0.08
	0.9	0.28	0.15	0.40	0.59	0.68	0.50	0.18	0.62	0.72	0.53	0.19
$T = 200$	0	0.03	0.05	0.06	0.09	0.10	0.07	0.05	0.09	0.10	0.08	0.05
	0.3	0.05	0.05	0.06	0.11	0.12	0.09	0.06	0.10	0.12	0.08	0.05
	0.6	0.06	0.06	0.09	0.17	0.19	0.12	0.07	0.16	0.18	0.10	0.06
	0.9	0.15	0.09	0.22	0.38	0.52	0.28	0.11	0.39	0.54	0.29	0.11

Notes: Superscript "$*$" signifies the use of bootstrap critical values. The asymptotically valid critical value for the test based on the self-normalized test statistic is taken from Table 1 (167.23; $m = 2$, $s = 2$).

6 Conclusion

We propose a novel self-normalized test statistic for general linear restrictions in cointegrating regressions, which avoids the estimation of nuisance parameters. In combination with the IM-OLS estimator, the test statistic is completely tuning parameter free. Its limiting null distribution is nonstandard, but we provide asymptotically valid critical values. To further improve the performance of the test in small to medium samples, we employ the VAR sieve bootstrap to construct critical values. To show its consistency, we establish a bootstrap invariance principle result under conditions that go beyond the assumptions commonly imposed in the literature, which might be of interest in its own right. Obtaining bootstrap critical values requires the choice of a single tuning parameter, the order of a VAR, which is a relatively straightforward and well understood task in practice.

Simulation results suggest that our new self-normalized bootstrap test works extremely well in finite samples and outperforms competing approaches, including the commonly employed Wald-type test, which depends on the choice of several tuning parameters to estimate a long-run variance parameter. In particular, the self-normalized bootstrap test has good power properties and is considerably less prone to size distortions, which are often observed to be tremendous for the commonly used approaches, especially when sample size is small or the level of error serial correlation and/or endogeneity is large. Moreover, the results are robust to information criteria choices to determine the order of the VAR. In larger samples the bootstrap critical values can be replaced by asymptotically valid critical values without major performance losses. Given that the self-normalized bootstrap test is easy to implement, it should therefore become a serious competitor to commonly used tests in practice.

Self-normalized bootstrap inference might also be a promising approach to address the enormous size distortions of hypothesis tests often observed in cointegrated panels. This interesting direction of future research is currently under investigation by the authors.

References

Andrews, D.W.K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* **59**, 817–858.

Andrews, B., Davis, R.A., Breidt, F.J. (2007). Rank-Based Estimation for All-Pass Time Series Models. *Annals of Statistics* **35**, 844–869.

Baxter, G. (1962). An Asymptotic Result for the Finite Predictor. *Mathematica Scandinavica* **10**, 137–144.

Benati, L., Lucas Jr., R.E., Nicolini, J.P., Weber, W. (2020). International Evidence on Long-Run Money Demand. *Journal of Monetary Economics*. Forthcoming.

Brillinger, D.R. (1981). Time Series: Data Analysis and Theory. Holden-Day, Inc., San Francisco.

Brüggemann, R., Jentsch, C., and Trenkler, C. (2016). Inference in VARs with Conditional Heteroskedasticity of Unknown Form. *Journal of Econometrics* **191**, 69–85.

Bühlmann, P. (1997). Sieve Bootstrap for Time Series. *Bernoulli* **3**, 123–148.

Cavaliere, G., Nielsen, H.B., Rahbek, A. (2015). Bootstrap Testing of Hypotheses on Co-Integration Relations in Vector Autoregressive Models. *Econometrica* **83**, 813–831.

Cavaliere, G., Rahbek, A., Taylor, A.M.R. (2012). Bootstrap Determination of the Co-Integration Rank in Vector Autoregressive Models. *Econometrica* **80**, 1721–1740.

Chang, Y., Park, J.Y. (2003). A Sieve Bootstrap for the Test of a Unit Root. *Journal of Time Series Analysis* **24**, 379–400.

Chang, Y., Park, J.Y., Song, K. (2006). Bootstrapping Cointegrating Regressions. *Journal of Econometrics* **133**, 703–739.

Choi, I., Kurozumi, E. (2012). Model Selection Criteria for the Leads-and-Lags Cointegrating Regression. *Journal of Econometrics* **169**, 224–238.

Dahlhaus, R., Kiss, I.Z., Neddermeyer, J.C. (2018). On the Relationship Between the Theory of Cointegration and the Theory of Phase Synchronization. *Statistical Science* **33**, 334–357.

Doukhan, P., Wintenberger, O. (2007). An Invariance Principle for Weakly Dependent Stationary General Models. *Probability and Mathematical Statistics* **27**, 45–73.

Einmahl, U. (1987). A Useful Estimate in the Multidimensional Invariance Principle. *Probability Theory and Related Fields* **76**, 81–101.

Hannan, E.J., Deistler, M. (1988). The Statistical Theory of Linear Systems. Wiley, New York.

Hwang, J., Sun, Y. (2018). Simple, Robust, and Accurate F and t Tests in Cointegrated Systems. *Econometric Theory* **34**, 949–984.

Jansson, M. (2002). Consistent Covariance Matrix Estimation for Linear Processes. *Econometric Theory* **18**, 1449–1459.

Jentsch, C., Lunsford, K. (2019). The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Comment. *American Economic Review* **109**, 2655–2678.

Kiefer, N.M., Vogelsang, T.J., Bunzel, H. (2000). Simple Robust Testing of Regression Hypotheses. *Econometrica* **68**, 695–714.

Kilian, L., Lütkepohl, H. (2017). Structural Vector Autoregressive Analysis. Cambridge University Press, Cambridge.

Knorre, F., Wagner, M., Grupe, M. (2020). Monitoring Cointegrating Polynomial Regressions: Theory and Application to the Environmental Kuznets Curves for Carbon and Sulfur Dioxide Emissions. Mimeo.

Kreiss, J.-P. (1988). Asymptotical Inference for a Class of Stochastic Processes. Universität Hamburg: Habilitationsschrift.

Kreiss, J.-P. (1992). Bootstrap Procedures for AR(∞) Processes. In Bootstrapping and Related Techniques, Jöckel, K.H., Rothe, G., Sender, W. (eds.), Lecture Notes in Economics and Mathematical Systems, vol. 376 Springer: Heidelberg, 107–113.

Kreiss, J.-P., Paparoditis, E., Politis, D.N. (2011). On the Range of Validity of the Autoregressive Sieve Bootstrap. *Annals of Statistics* **39**, 2103–2130.

Lanne, M., Saikkonen, P. (2013). Noncausal Vector Autoregression. *Econometric Theory* **29**, 447–481.

Li, H., Maddala, G.S. (1997). Bootstrapping Cointegrating Regressions. *Journal of Econometrics* **80**, 297–318.

Merlevède, F., Peligrad, M., Utev, S. (2006). Recent Advances in Invariance Principles for Stationary Sequences. *Probability Surveys* **3**, 1–36.

Meyer, M., Kreiss, J.-P. (2015). On the Vector Autoregressive Sieve Bootstrap. *Journal of Time Series Analysis* **36**, 377 – 397.

Newey, W.K., West, K.D. (1994). Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies* **61**, 631–653.

Palm, F.C., Smeekes, S., Urbain, J.P. (2010). A Sieve Bootstrap Test for Cointegration in a Conditional Error Correction Model. *Econometric Theory* **26**, 647 – 681.

Paparoditis, E. (1996). Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes. *Journal of Multivariate Analysis* **57**, 277 – 296.

Paparoditis, E., Politis, D.N. (2003). Residual-Based Block Bootstrap for Unit Root Testing. *Econometrica* **71**, 813 – 855.

Paparoditis, E., Politis, D.N. (2005). Bootstrap Hypothesis Testing in Regression Models. *Statistics & Probability Letters* **74**, 356 – 365.

Park, J.Y. (1992). Canonical Cointegrating Regressions. *Econometrica* **60**, 119–143.

Park, J.Y. (2002). An Invariance Principle for Sieve Bootstrap in Time Series. *Econometric Theory* **18**, 469–490.

Patton, A., Politis, D.N., White, H. (2009). Correction to "Automatic Block-Length Selection for the Dependent Bootstrap" by D. Politis and H. White. *Econometric Reviews* **28**, 372–375.

Phillips, P.C.B. (2014). Optimal Estimation of Cointegrated Systems With Irrelevant Instruments. *Journal of Econometrics* **178**, 210–224.

Phillips, P.C.B., Hansen, B.E. (1990). Statistical Inference in Instrumental Variables Regression with I(1) Processes. *Review of Economic Studies* **57**, 99–125.

Phillips, P.C.B., Leirvik, T., Storelvmo, T. (2020). Econometric Estimates of Earth's Transient Climate Sensitivity. *Journal of Econometrics* **214**, 6 – 32.

Phillips, P.C.B., Loretan, M. (1991). Estimating Long Run Economic Equilibria. *Review of Economic Studies* **58**, 407–436.

Phillips, P.C.B., Solo, V. (1992). Asymptotics for Linear Processes. *Annals of Statistics* **20**, 971–1001.

Politis, D.N., White, H. (2004). Automatic Block-Length Selection for the Dependent Bootstrap. *Econometric Reviews* **23**, 53-70.

Psaradakis, Z. (2001). On Bootstrap Inference in Cointegrating Regressions. *Economics Letters* **72**, 1–10.

Rad, H., Low, R.K.Y., Faff, R. (2016). The Profitability of Pairs Trading Strategies: Distance, Cointegration and Copula Methods. *Quantitative Finance* **16**, 1541–1558.

Saikkonen, P. (1991). Asymptotically Efficient Estimation of Cointegrating Regressions. *Econometric Theory* **7**, 1–21.

Shao, X. (2010). The Dependent Wild Bootstrap. *Journal of the American Statistical Association* **105**, 218–235.

Shao, X. (2015). Self-Normalization for Time Series: A Review of Recent Developments. *Journal of the American Statistical Association* **110**, 1797–1817.

Shin, Y. (1994). A Residual-Based Test of the Null of Cointegration Against the Alternative of No Cointegration. *Econometric Theory* **10**, 91–115.

Stock, J.H., Watson, M.W. (1993). A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica* **61**, 783–820.

van Giersbergen, N.P.A., Kiviet, J.F. (2002). How to Implement the Bootstrap in Static or Stable Dynamic Regression Models: Test Statistic Versus Confidence Region Approach. *Journal of Econometrics* **108**, 133–156.

Vogelsang, T.J., Wagner, M. (2014). Integrated Modified OLS Estimation and Fixed-b Inference for Cointegrating Regressions. *Journal of Econometrics* **178**, 741–760.

Wagner, M. (2015). The Environmental Kuznets Curve, Cointegration and Nonlinearity. *Journal of Applied Econometrics* **30**, 948–967.

Wang, R., Shao, X. (2020). Hypothesis Testing for High-Dimensional Time Series via Self-Normalization. *Annals of Statistics* **48**, 2728–2758.

Wu, W. (2005). Nonlinear System Theory: Another Look at Dependence. *Proceedings of the National Academy of Sciences of the USA* **102**, 14150–14154.

Wu, W. (2007). Strong Invariance Principles for Dependent Random Variables. *Annals of Probability* **35**, 2294–2320.

Appendices

A Additional Finite Sample Results

Table 3: Bias and RMSE of the IM-, D- and FM-OLS estimator of β_1 and β_2, respectively

						Bartlett kernel		QS kernel	
	ρ_1, ρ_2	$\hat{\beta}_{\mathrm{IM},1}$	$\hat{\beta}_{\mathrm{IM},2}$	$\hat{\beta}_{\mathrm{D},1}$	$\hat{\beta}_{\mathrm{D},2}$	$\hat{\beta}_{\mathrm{FM},1}$	$\hat{\beta}_{\mathrm{FM},2}$	$\hat{\beta}_{\mathrm{FM},1}$	$\hat{\beta}_{\mathrm{FM},2}$
					Bias				
$T = 50$	0	0.00	-0.00	-0.01	0.00	0.00	-0.00	0.00	-0.00
	0.3	0.00	-0.00	-0.01	-0.00	0.01	0.01	0.01	0.01
	0.6	0.02	0.02	-0.01	-0.00	0.05	0.05	0.05	0.05
	0.9	0.24	0.22	0.14	0.13	0.29	0.27	0.30	0.28
$T = 100$	0	0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00
	0.3	0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00
	0.6	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02
	0.9	0.13	0.14	0.11	0.12	0.17	0.18	0.17	0.18
$T = 200$	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.6	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01
	0.9	0.05	0.05	0.06	0.06	0.09	0.09	0.09	0.09
					RMSE				
$T = 50$	0	0.07	0.07	0.21	0.21	0.05	0.05	0.05	0.05
	0.3	0.10	0.10	0.26	0.26	0.07	0.07	0.07	0.07
	0.6	0.18	0.18	0.36	0.36	0.14	0.14	0.15	0.15
	0.9	0.70	0.67	0.77	0.76	0.51	0.50	0.59	0.57
$T = 100$	0	0.04	0.04	0.02	0.02	0.02	0.02	0.02	0.02
	0.3	0.05	0.05	0.03	0.03	0.03	0.03	0.03	0.03
	0.6	0.09	0.09	0.06	0.06	0.06	0.07	0.07	0.07
	0.9	0.38	0.39	0.29	0.31	0.31	0.32	0.34	0.34
$T = 200$	0	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01
	0.3	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02
	0.6	0.05	0.05	0.03	0.03	0.03	0.03	0.03	0.03
	0.9	0.20	0.20	0.15	0.15	0.18	0.18	0.19	0.19

Notes: $\hat{\beta}_{\mathrm{IM},i}$, $\hat{\beta}_{\mathrm{D},i}$ and $\hat{\beta}_{\mathrm{FM},i}$ denote the IM-, D- and FM-OLS estimator of β_i, respectively, $i = 1, 2$.

B Proofs of the Main Results

Proof of Proposition 1. Vogelsang and Wagner (2014, Lemma 2) show that

$$T^{-1/2}\sum_{t=2}^{\lfloor rT\rfloor}\Delta\hat{S}_t^u \xrightarrow{w} \Omega_{u\cdot v}^{1/2}\left(W_{u\cdot v}(r)-g(r)'\mathcal{Z}\right), \quad 0\leq r\leq 1,$$

as $T\to\infty$. The continuous mapping theorem thus yields

$$\hat{\eta}_T = T^{-1}\sum_{t=2}^{T}\left(T^{-1/2}\sum_{s=2}^{t}\Delta\hat{S}_s^u\right)^2 \xrightarrow{w} \Omega_{u\cdot v}\int_0^1\left(W_{u\cdot v}(r)-g(r)'\mathcal{Z}\right)^2 dr,$$

as $T\to\infty$. The final result now follows with standard arguments from (3.6). □

The proof of the remaining results relies on the following lemma.

Lemma 1. *Under Assumptions 1, 2 and 4, it holds that*

$$\max_{q+1\leq t\leq T}|\hat{w}_t - w_t|_F = O_{\mathbb{P}}(T^{-1/2}) \tag{B.1}$$

and

$$q^{1/2}\sum_{j=1}^{q}|\hat{\Phi}_j(q)-\tilde{\Phi}_j(q)|_F = O_{\mathbb{P}}(q^3/T) = o_{\mathbb{P}}(1), \tag{B.2}$$

where $\tilde{\Phi}_1(q),\ldots,\tilde{\Phi}_q(q)$ *denote the solution of the sample Yule-Walker equations in the regression of* w_t *on* $w_{t-1},\ldots,w_{t-q}$, $t=q+1,\ldots,T$.

Proof. The proof is given in the Online Appendix. □

Two key ingredients in the proof of Lemma 1, which are also useful hereafter, are the following. First, it holds under Assumptions 1 and 4 that

$$q^{3/2}\sup_{1\leq j\leq q}|\tilde{\Phi}_j(q)-\Phi_j(q)|_F = q^{3/2}O_{\mathbb{P}}((\ln(T)/T)^{1/2}) = O_{\mathbb{P}}(1), \tag{B.3}$$

compare Meyer and Kreiss (2015, Remark 3.3.), where $\Phi_1(q),\ldots,\Phi_q(q)$ denote the finite predictor coefficients, i.e., the solution of the population Yule-Walker equations based on the true moments. Second, under Assumption 1 with $k\geq 3/2$, there exist constants $q_0 \in \mathbb{N}$ and $c < \infty$ such that

$$\sum_{j=1}^{q}(1+j)^k|\Phi_j(q) - \Phi_j|_F \leq c \sum_{j=q+1}^{\infty}(1+j)^k|\Phi_j|_F, \tag{B.4}$$

for all $q \geq q_0$ and the right-hand side converges to zero as $q \to \infty$, see Meyer and Kreiss (2015, Lemma 3.1).[21]

Lemma 1 is used to prove the following two lemmas, which in turn are used to prove Proposition 2.

Lemma 2. *It holds under Assumptions 1 and 4 that*

$$\mathbb{E}^*\left(|\varepsilon_t^*|_F^a\right) = (T-q)^{-1}\sum_{t=q+1}^{T}|\hat{\varepsilon}_t(q) - \bar{\hat{\varepsilon}}_T(q)|_F^a = O_{\mathbb{P}}(1),$$

in $\mathbb{P}$*, for the* $a > 2$ *from Assumption 1.*

Proof. The proof is given in the Online Appendix. □

Lemma 3. *It holds under Assumptions 1, 2 and 4 that*

$$\mathbb{E}^*\left(\varepsilon_t^*\varepsilon_t^{*\prime}\right) = (T-q)^{-1}\sum_{t=q+1}^{T}\left(\hat{\varepsilon}_t(q) - \bar{\hat{\varepsilon}}_T(q)\right)\left(\hat{\varepsilon}_t(q) - \bar{\hat{\varepsilon}}_T(q)\right)' = \Sigma + o_{\mathbb{P}}(1),$$

in $\mathbb{P}$*.*

Proof. The proof is given in the Online Appendix. □

Proof of Proposition 2. Given Lemma 2 and Lemma 3, the result follows immediately from Einmahl (1987), as in Chang *et al.* (2006, p. 714).[22] □

Proof of Theorem 1. Using similar arguments as Palm *et al.* (2010, p. 670), it follows that

$$B_T^*(r) = T^{-1/2}\sum_{t=1}^{\lfloor rT \rfloor} w_t^* = \left(I - \sum_{j=1}^{q}\hat{\Phi}_j(q)\right)^{-1} W_T^*(r) + T^{-1/2}(\bar{w}_0^* - \bar{w}_{\lfloor rT \rfloor}^*),$$

[21]This result is known as the generalized Baxter's inequality, see Baxter (1962) and Hannan and Deistler (1988, p. 269).

[22]For more details we refer to the pre-print of Palm *et al.* (2010), which is available on the website http://researchers-sbe.unimaas.nl/stephansmeekes/research/publications/ (Accessed: 2020-12-17).

where $\bar{w}^*_{t-1} := \left(I - \sum_{j=1}^{q} \hat{\Phi}_j(q)\right)^{-1} \sum_{i=1}^{q} \left(\sum_{j=i}^{q} \hat{\Phi}_j(q)\right) w^*_{t-i}$. It thus remains to show that

$$I - \sum_{j=1}^{q} \hat{\Phi}_j(q) \xrightarrow{p} \Phi(1) \tag{B.5}$$

and

$$\mathbb{P}^*\left(\max_{0\leq t\leq T} |T^{-1/2}\bar{w}^*_t|_F > \delta\right) = o_{\mathbb{P}}(1). \tag{B.6}$$

We first show (B.5). From Lemma 1, (B.3) and (B.4), we obtain

$$\begin{aligned}
|I - \sum_{j=1}^{q} \hat{\Phi}_j(q) - \Phi(1)|_F &\leq \sum_{j=1}^{q} |\hat{\Phi}_j(q) - \tilde{\Phi}_j(q)|_F + \sum_{j=1}^{q} |\tilde{\Phi}_j(q) - \Phi_j(q)|_F \\
&\quad + \sum_{j=1}^{q} |\Phi_j(q) - \Phi_j|_F + \sum_{j=q+1}^{\infty} |\Phi_j|_F \\
&\leq \sum_{j=1}^{q} |\hat{\Phi}_j(q) - \tilde{\Phi}_j(q)|_F + q \sup_{1\leq j\leq q} |\tilde{\Phi}_j(q) - \Phi_j(q)|_F \\
&\quad + c \sum_{j=q+1}^{\infty} |\Phi_j|_F + \sum_{j=q+1}^{\infty} |\Phi_j|_F \\
&= o_{\mathbb{P}}(1) + o_{\mathbb{P}}(1) + o(1) + o(1) = o_{\mathbb{P}}(1).
\end{aligned}$$

To prove (B.6), we note that it follows from strict stationarity of $\{\bar{w}^*_t\}_{t\in\mathbb{Z}}$ and Markov's inequality, that

$$\begin{aligned}
\mathbb{P}^*\left(\max_{0\leq t\leq T} |T^{-1/2}\bar{w}^*_t|_F > \delta\right) &\leq \sum_{t=0}^{T} \mathbb{P}^*\left(|T^{-1/2}\bar{w}^*_t|_F > \delta\right) \\
&\leq (T+1)\mathbb{P}^*\left(|T^{-1/2}\bar{w}^*_t|_F > \delta\right) \\
&\leq \delta^{-a}(T^{1-a/2} + T^{-a/2})\mathbb{E}^*\left(|\bar{w}^*_t|_F^a\right),
\end{aligned}$$

with the $a > 2$ from Assumption 1, compare Park (2002, p. 486). Similarly as in Palm *et al.* (2010, p. 671), we obtain

$$\mathbb{E}^*\left(|\bar{w}^*_t|_F^a\right) \leq c\,(m+1)^{a/2-1} \left(\sum_{j=0}^{\infty} |\bar{\hat{\Psi}}_j(q)|_F^2\right)^{a/2} \mathbb{E}^*\left(|\varepsilon^*_t|_F^a\right),$$

for some constant c and $\bar{\hat{\Psi}}_j(q) := \sum_{i=j+1}^{\infty} \hat{\Psi}_i(q)$, where the matrices $\hat{\Psi}_j$ are determined by the power series expansion of the inverse of $I - \sum_{j=1}^{q} \hat{\Phi}_j(q) z^j$. As discussed in Palm *et al.* (2010, p. 671) it follows that $\sum_{j=0}^{\infty} |\bar{\hat{\Psi}}_j(q)|_F^2 = O_{\mathbb{P}}(1)$ if we can show that $\sum_{j=1}^{q} j^{1/2} |\hat{\Psi}_j(q)|_F = O_{\mathbb{P}}(1)$, which in turn holds if $\sum_{j=1}^{q} j^{1/2} |\hat{\Phi}_j(q)|_F = O_{\mathbb{P}}(1)$. Using again Lemma 1, (B.3) and (B.4), we obtain

$$\begin{aligned}
\sum_{j=1}^{q} j^{1/2} |\hat{\Phi}_j(q)|_F &\leq \sum_{j=1}^{q} j^{1/2} |\hat{\Phi}_j(q) - \tilde{\Phi}_j(q)|_F + \sum_{j=1}^{q} j^{1/2} |\tilde{\Phi}_j(q) - \Phi_j(q)|_F \\
&\quad + \sum_{j=1}^{q} j^{1/2} |\Phi_j(q) - \Phi_j|_F + \sum_{j=1}^{q} j^{1/2} |\Phi_j|_F \\
&\leq q^{1/2} \sum_{j=1}^{q} |\hat{\Phi}_j(q) - \tilde{\Phi}_j(q)|_F + q^{1/2} \sum_{j=1}^{q} |\tilde{\Phi}_j(q) - \Phi_j(q)|_F \\
&\quad + \sum_{j=1}^{q} (1+j) |\Phi_j(q) - \Phi_j|_F + \sum_{j=1}^{q} j^{1/2} |\Phi_j|_F \\
&\leq q^{1/2} \sum_{j=1}^{q} |\hat{\Phi}_j(q) - \tilde{\Phi}_j(q)|_F + q^{3/2} \sup_{1 \leq j \leq q} |\tilde{\Phi}_j(q) - \Phi_j(q)|_F \\
&\quad + \sum_{j=1}^{q} (1+j) |\Phi_j(q) - \Phi_j|_F + \sum_{j=1}^{q} j^{1/2} |\Phi_j|_F \\
&= o_{\mathbb{P}}(1) + O_{\mathbb{P}}(1) + o(1) + O(1) = O_{\mathbb{P}}(1).
\end{aligned}$$

This completes the proof, since $\mathbb{E}^* \left(|\varepsilon_t^*|_F^a \right) = O_{\mathbb{P}}(1)$ by Lemma 2 for the $a > 2$ from Assumption 1. □

Proof of Theorem 2. The first result in Theorem 2 follows from the bootstrap invariance principle result in Theorem 1 and similar arguments as used in Vogelsang and Wagner (2014, Proof of Theorem 2). The second result then follows from the bootstrap invariance principle and similar arguments as used in the proof of Proposition 1. □