REGULAR ARTICLE

A comparison of semiparametric tests for fractional cointegration

Christian Leschinski[1] · Michelle Voges[1] · Philipp Sibbertsen[1]

Abstract

There are various competing procedures to determine whether fractional cointegration is present in a multivariate time series, but no standard approach has emerged. We provide a synthesis of this literature and conduct a detailed comparative Monte Carlo study to guide empirical researchers in their choice of appropriate methodologies. Special attention is paid on empirically relevant issues such as assumptions about the form of the underlying process and the ability of the procedures to distinguish between short-run correlation and long-run equilibria. It is found that several approaches are severely oversized in presence of correlated short-run components and that the methods show different performance in terms of power when applied to common-component models instead of triangular systems.

Keywords Long memory · Fractional cointegration · Semiparametric estimation and testing

1 Introduction

The concept of cointegration derives its popularity from the fact that it allows to model equilibrium relationships between non-stationary time series. The most popular tests in the standard $I(1)/I(0)$ setting includes the two-step procedure by Engle and Granger (1987), the trace test by Johansen (1988) and the principal component test by Phillips and Ouliaris (1988) which are subject of several comparaitive studies like Reimers (1992) and Höglund and Östermark (2003). In practice, however, standard cointegration analysis can often not be applied, since the $I(1)/I(0)$ framework is too restrictive.

✉ Philipp Sibbertsen
sibbertsen@statistik.uni-hannover.de

[1] Leibniz Universität Hannover, Königsworther Platz 1, 30167 Hannover, Germany

For example, the series of interest may be persistent but not have a unit root, or the deviations from the equilibrium may be more persistent than the $I(0)$ model allows.

Fractional cointegration overcomes these shortcomings, by allowing for non-integer integration orders of the variables in the system and any (possibly non-zero) memory order in the cointegrating residuals as long as it is reduced compared to the original system. Consequently, fractional cointegration promises to facilitate the modeling of a larger number of equilibrium relationships compared to standard cointegration. This has led to the development of various testing and rank estimation procedures to determine whether fractional cointegration is present in a multivariate time series.

Parametric approaches include Johansen (2008), Łasak (2010), Johansen and Nielsen (2012), Łasak and Velasco (2015), and Johansen and Nielsen (2019), among others, who consider fractional extensions of the cointegrated VAR model of Johansen (1988). Furthermore, Breitung and Hassler (2002) introduce a trace test to determine the cointegrating rank, Avarucci and Velasco (2009) suggest rank estimation in a regression framework, and Hassler and Breitung (2006) develop a time domain residual-based test.

Semiparametric approaches, on the other hand, have the advantage that they allow the researcher to focus on the long-run relationship between the series and do not require the specification of short-run dynamics. This literature encompasses the spectral-based rank estimation procedure of Robinson and Yajima (2002) and its extension by Nielsen and Shimotsu (2007), a Hausmann-type test based on the multivariate local Whittle estimator introduced by Robinson (2008a), a number of residual-based tests for the null hypothesis of no fractional cointegration developed by Marmol and Velasco (2004), Chen and Hurvich (2006), Hualde and Velasco (2008), and Wang et al. (2015), a variance-ratio test proposed by Nielsen (2010), a test based on a GPH-type estimate of the cointegration strength introduced by Souza et al. (2018) and a rank estimation procedure based on an eigenanalysis of the autocovariance function from Zhang et al. (2019).

Unfortunately, the domain of applicability of most of these procedures is much more restrictive than the definition of fractional cointegration. Some are only applicable in stationary systems—some only in non-stationary systems. Some procedures require the reduction in memory to be more than 1/2—some require the memory of the cointegrating residuals to be less than 1/2.

Furthermore, there are different assumptions about the form of the fractionally cointegrated system. Some approaches assume that one of the observed series itself is an observation of the common underlying trend. Other approaches assume an unobserved common underlying trend. We refer to these models as the triangular system and the common-components model. Which of these assumptions is more suitable in practice depends on the specific application. On the one hand, it may be appropriate to think of the risk-free interest rate as an observed common component that is perturbed by risk premia in risky bonds so that a triangular model can be used. For cointegrated pairs of stocks, on the other hand, it is unclear why the price of one stock should be interpreted as a perturbed version of another stock price so that a common-components model is more appropriate. Finally, even though the development of each of these procedures to determine whether fractional cointegration is present is a major theoretical contri-

bution, relatively little effort has been devoted to analyze how they perform compared to each other.

Here, we try to address these issues by providing a survey of all the rank estimation and testing procedures discussed above. To study the relative performance of the competing approaches, we conduct an extensive Monte Carlo analysis of their size and power properties. It is found that several procedures - namely those of Nielsen and Shimotsu (2007) or Robinson and Yajima (2002), Marmol and Velasco (2004), and Hualde and Velasco (2008) show severe finite sample size distortions in systems with correlated short-run components. The relative performance in terms of power depends on the form of the system. For triangular systems and non-stationary common-components models the test of Souza et al. (2018) performs best overall, whereas the test of Chen and Hurvich (2006) is preferable for stationary common-components models.

The rest of the paper is structured as follows. The next section gives the definition and model of fractional cointegration we adopt and briefly reviews the basic estimation methods required by the tests. Section 3 is divided into two subsections describing two types of tests, 3.1 containing the tests based on a spectral matrix and 3.2 summarizing the tests based on cointegrating residuals, Sect. 4 presents finite sample results, and Sect. 5 concludes.

2 Fractional cointegration: models and definitions

A p-dimensional vector-valued time series X_t has long memory if its spectral density fulfills

$$f_X(\lambda) \sim \Lambda_j(d) \, G \, \overline{\Lambda_j(d)}, \quad \text{as } \lambda \to 0^+, \tag{1}$$

where G is a real, symmetric, and non-negative definite matrix, $\Lambda_j(d) = \text{diag}\left(\lambda^{-d_1} e^{i\pi d_1/2}, \ldots, \lambda^{-d_p} e^{i\pi d_p/2}\right)$ is a $p \times p$ diagonal matrix, $\overline{\Lambda_j(d)}$ is its complex conjugate transpose and '$\sim$' implies that for each element the ratio of real and imaginary parts on the left- and right-hand side tends to one. The element in the a-th row and b-th columns of the spectral matrix $f_X(\lambda)$ is denoted by $f_{ab}(\lambda) \sim g_{ab}\lambda^{-2d}$ for $a, b \in \{1, \ldots, p\}$ where g_{ab} denotes the respective element of G. The periodogram of X_t at the Fourier frequencies is given by

$$I_X(\lambda_j) = w_X(\lambda_j)\overline{w_X(\lambda_j)}, \tag{2}$$

with $w_X(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} X_t e^{i\lambda t}$, and $\lambda_j = 2\pi j/T$, for $j = 1, \ldots, \lfloor T/2 \rfloor$, where $\lfloor \cdot \rfloor$ denotes the greatest integer smaller than the argument.

There is a number of different definitions of fractional cointegration in the literature. The most common one goes back to Engle and Granger (1987). According to this definition the p-dimensional time series X_t is cointegrated of rank r, if all components of X_t are integrated of order d (denoted by $I(d)$), and there exists a non-singular matrix β so that the r linear combinations $v_t = \beta' X_t$ are $I(d-b_a) = I(d_{v_a})$ with $d > b_a > 0$

for all $a = 1, \ldots, r$. The matrix β is called the cointegrating matrix and each of its columns is a cointegrating vector. The elements of the vector v_t are the cointegrating residuals. Other definitions are given by Johansen (1995), Flôres Jr and Szafarz (1996), Marinucci and Robinson (2001), and Robinson and Yajima (2002) who also provide a discussion of the implications of the different definitions.

Standard cointegration is a special case of the definition above where $d = 1$ and $d_{v_a} = 0$ for all a. In this setup the system is non-stationary, whereas the cointegrating residuals are stationary. In contrast to that, fractional cointegration allows for a more flexible model so that several cases can be distinguished: weak cointegration ($b < 0.5$), strong cointegration ($b > 0.5$), stationary cointegration ($0 < d_v < d < 0.5$), or non-stationary cointegration ($0.5 < d_v < d$).

In general, (fractional) cointegration is an equilibrium concept where the persistence of the cointegrating residual d_v determines the speed of adjustment towards the cointegration equilibrium $\beta' X_t$, and shocks have no permanent influence on the equilibrium as long as $d_v < 1$ holds.

As an example, consider the fractionally (co-)integrated bivariate model with $X_t = (X_{1t}, X_{2t})'$, where

$$X_{1t} = c_1 + \xi_1 Y_t + \Delta^{-(d-b_1)} u_{1t} \mathbb{1}(t > 0) \tag{3}$$

$$X_{2t} = c_2 + \xi_2 Y_t + \Delta^{-(d-b_2)} u_{2t} \mathbb{1}(t > 0) \tag{4}$$

$$\text{and} \quad Y_t = \Delta^{-d} e_t \mathbb{1}(t > 0). \tag{5}$$

Here, $u_t = (u_{1t}, u_{2t})'$ is a weakly-dependent zero-mean process with constant covariance matrix Ω_u and spectral density matrix $f_u(\lambda)$, e_t (with variance σ_e^2 and spectral density $f_e(\lambda)$) is a univariate weakly-dependent zero-mean process that is allowed to be correlated with u_t, and L denotes the lag-operator so that $LY_t = Y_{t-1}$. The fractional difference operator $\Delta^d = (1-L)^d$ is defined in terms of the binomial expansion so that $(1 - L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k L^k$, with $\binom{d}{k} = \frac{d(d-1)(d-2)\ldots(d-(k-1))}{k!}$. Furthermore, $\mathbb{1}(\cdot)$ denotes the indicator function that takes the value one if its argument is true and is zero, otherwise. Finally, it is assumed that $d \geq b_1, b_2 \geq 0$.

The truncated processes $\Delta^{-(d-b_a)} u_{at} \mathbb{1}(t > 0)$ are fractionally-integrated processes of type-II which means they are only asymptotically stationary for $d < 1/2$, but in contrast to type-I processes they are still defined for $d > 1/2$. For a detailed discussion cf. Marinucci and Robinson (1999).

In this bivariate model there can be at most one cointegrating relationship. In this case $r = 1$ and β itself is a cointegrating vector. Obviously, if the linear combination $\beta' X_t = v_t$ has reduced memory, the same is true for every scalar multiple of it. To identify the cointegrating vector, it is therefore customary to apply some kind of normalization such as setting the first element of the vector to unity. In Eqs. (3) to (5), fractional cointegration arises if $\xi_1, \xi_2 \neq 0$, and $b_1, b_2 > 0$. In this case the normalized cointegrating vector is $\beta = \left(1, -\frac{\xi_1}{\xi_2}\right)' = \left(1, -\tilde{\beta}\right)'$ and the cointegrating residual v_t is $I(d - b) = I(d_v)$, where $b = \min(b_1, b_2)$. Note that this model is a common-components model, but it also nests a triangular system. This is obtained as a special case if $\Omega_{u,22} = 0$ so that X_{2t} is a direct (rescaled) observation of the underlying

common trend and only X_{1t} is perturbed with a cointegration error so that $b = b_1$. Standard cointegration in the $I(1)/I(0)$ framework is obtained as a special case if $d = 1$ and $b_1 = b_2 = 1$. It is also possible to have $\xi_1, \xi_2 \neq 0$, so that both X_{1t} and X_{2t} contain the common component Y_t, but they are not cointegrated if $b_1 = b_2 = 0$.

3 Tests for no fractional cointegration

In the following, we provide a comprehensive review of semiparametric tests and estimation procedures that can be used to determine the order of fractional cointe- gration in a p-dimensional vector-valued time series X_t. According to the definition discussed above, this requires that the components of X_t are integrated of the same order. In practice, this can either be assumed based on domain-specific knowledge, or it can be tested with tests for the equality of memory parameters that allow for cointegration introduced by, for example, Robinson and Yajima (2002), Nielsen and Shimotsu (2007), Hualde (2013), and Wang and Chan (2016). In particular Robin- son and Yajima (2002) discuss in detail how to partition a vector-valued time series into subvectors with equal memory parameters. These can then be used for further cointegration analysis.

In the following, it will be assumed that all components of X_t are $I(d)$, which means we abstract from these pre-testing issues to focus on the actual tests for the null of no fractional cointegration. For all tests the hypotheses are defined by

H_0: X_t is not fractionally cointegrated ($d = d_v$),
H_1: X_t is fractionally cointegrated ($d > d_v$).

In contrast to standard $I(1)/I(0)$ cointegration, the memory parameter d is unknown in fractionally cointegrated systems and has to be estimated. Since mul- tivariate memory estimation becomes inconsistent under cointegration, the memory parameters are estimated univariately and, if not stated otherwise, we employ the means of the univariate memory estimates in the tests.

The tests presented in this Section apply the most common estimators: the log- periodogram estimator $\widehat{d}_{GPH}$ of Geweke and Porter-Hudak (1983) and Robinson (1995b), the local Whittle estimator $\widehat{d}_{LW}$ of Künsch (1987) and Robinson (1995a), or the exact local Whittle estimator $\widehat{d}_{ELW}$ of Shimotsu and Phillips (2005) and Shimotsu (2010). All of these estimators are periodogram-based and employ the first m Fourier frequencies. The general requirement is that $m < \lfloor T/2 \rfloor$ tends to infinity more slowly than T so that $\frac{1}{m} + \frac{m}{T} \to 0$ as $T \to \infty$ and even the largest frequency $2\pi m/T$ is asymptotically local to the zero frequency.

To estimate the cointegrating relationship $\beta' X_t = v_t$ when $r = 1$, the vector is partitioned such that $X_t = (y_t, x_t)$, where y_t is a scalar and x_t is $(p - 1) \times 1$. By doing so, the focus is on one possible cointegrating relation $y_t = \tilde{\beta} x_t + v_t$ where $\tilde{\beta}$ is $(p - 1)$-dimensional.

As in standard cointgration analysis the vector $\tilde{\beta}$ can be estimated with ordinary least squares (OLS) as long as $d > 1/2$ so that the series remains non-stationary. In stationary long-memory time series, OLS is inconsistent in presence of correlation between the stationary regressors and the innovation term v_t (cf. Robinson (1994)).

Robinson ([1994](#)) and Robinson and Marinucci ([2001](#)) introduce an alternative estimator of the cointegrating vector that is based on the periodogram local to the zero frequency. In contrast to OLS, this narrow-band frequency domain least squares (NBLS) estimator is consistent under cointegration for all values of d and has a non-normal limiting distribution in the non-stationary region. Christensen and Nielsen ([2006a](#)) extend the asymptotic results to the stationary region where the estimate follows an asymptotic normal distribution and Nielsen and Frederiksen ([2011](#)) provide a correction of the asymptotic bias under weak fractional cointegration.

Estimating the linear cointegrating relationship with NBLS requires calculating the averaged cross-periodogram of x_t with itself and y_t by $I_{xx}^{av}(\lambda_j) = \frac{2\pi}{T} \sum_{j=1}^{m} \omega_x(\lambda_j) \overline{\omega_x(\lambda_j)}$ and $I_{xy}^{av}(\lambda_j) = \frac{2\pi}{T} \sum_{j=1}^{m} \omega_x(\lambda_j) \overline{\omega_y(\lambda_j)}$. The NBLS estimate of $\tilde{\beta}$ is then defined by

$$\widehat{\beta}_m = I_{xx}^{av}(\lambda_j)^{-1} I_{xy}^{av}(\lambda_j). \tag{6}$$

The bandwidth m has to fulfill the usual local-to-zero condition as $T \to \infty$. If not specified otherwise, we employ NBLS to estimate the cointegrating vector. Other estimators suggested in the literature include estimation based on the eigenvectors of a version of $I_X^{av}(\lambda_j)$ (cf. Chen and Hurvich ([2006](#))) and joint estimation with the memory parameters in multivariate local Whittle approaches such as those of Nielsen ([2007](#)), Robinson ([2008b](#)) and Shimotsu ([2012](#)).

The following review is divided into tests based on the spectral density local to the origin (Sect. [3.1](#)) and tests based on estimates of the cointegrating residuals (Sect. [3.2](#)). Of course, this distinction is not clear cut, since some of the residual-based approaches also use the spectral properties of the potential cointegrating residuals and for example the test of Nielsen ([2010](#)) is presented as a variance-ratio test. Many different categorizations would be possible. Here, we refer to those approaches as "spectral-based" that rely on the properties of the spectrum of the observed series X_t itself, and those that rely on the spectrum of the cointegrating residual are called "residual-based".

3.1 Tests based on the spectral matrix

A number of procedures to determine the fractional cointegrating rank of the p-dimensional time series X_t are based on properties of the rescaled spectral matrix local to the zero frequency. This is denoted by G in Eq. ([1](#)) and has reduced rank if and only if X_t is fractionally cointegrated. If fractional cointegration is present, the number of eigenvalues that are equal to zero corresponds to the cointegrating rank r. More details on the connection between fractional cointegration, unit coherence and singularity of G are given in Velasco ([2003b](#)) and Nielsen ([2004](#)).

Based on this property Robinson and Yajima ([2002](#)) introduce an information criterion to determine the fractional cointegration rank that is extended to non-stationary processes by Nielsen and Shimotsu ([2007](#)). To obtain an estimate $\widehat{G}$ of G, the first step consists in applying the univariate exact local Whittle estimator of Shimotsu and Phillips ([2005](#)) and Shimotsu ([2010](#)) to each component of X_t separately, using band-

width m, and pooling them to the arithmetic mean $\widehat{d}_{ELW}$. The estimate of $\widehat{G}(\widehat{d}_{ELW})$ is then defined by $\widehat{G}(\widehat{d}_{ELW}) = \frac{1}{m_1} \sum_{j=1}^{m_1} \text{Re } I_{\Delta^d}(\lambda_j)$, where I_{Δ^d} is the periodogram of $\Delta^{\widehat{d}_{ELW}} X_t$. The bandwidths have to fulfill $\frac{m_1}{m} \to 0$ in order to ensure faster convergence of $\widehat{d}_{ELW}$ than of $\widehat{G}(\widehat{d}_{ELW})$.[1] Denote the empirical eigenvalues calculated from $\widehat{G}(\widehat{d}_{ELW})$ and sorted in descending order by $\widehat{\delta}_{a,G}$ for $a = 1, \ldots, p$. The cointegrating rank can then be estimated using a model selection criterion that is based on the partial sum of the sorted eigenvalues

$$\widehat{r}_{NS} = \arg \min_{k=0,\ldots,p-1} \left(n(T)(p-k) - \sum_{a=1}^{p-k} \widehat{\delta}_{a,G} \right), \qquad (7)$$

where $n(T)$ is a function which fulfills $n(T) + \frac{1}{\sqrt{m_1 n(T)}} \to 0$ as $T \to \infty$ so that $n(T)$ goes to zero more slowly than the estimation error in the eigenvalues that is of order $\mathcal{O}_P\left(m_1^{-1/2}\right)$. Asymptotically, the expression is therefore minimal if only estimates of non-zero eigenvalues are included in the sum.

To deal with situations in which the scales of the components in X_t are different, Nielsen and Shimotsu (2007) suggest to base the procedure on the correlation matrix $\widehat{P}(\widehat{d}_{ELW}) = \widehat{R}(\widehat{d}_{ELW})^{-1/2} \widehat{G}(\widehat{d}_{ELW}) \widehat{R}(\widehat{d}_{ELW})^{-1/2}$ instead of $\widehat{G}$, where $\widehat{R}(\widehat{d}_{ELW}) = \text{diag}(\widehat{g}_{11}, \ldots, \widehat{g}_{pp})$ contains the diagonal elements of $\widehat{G}(\widehat{d}_{ELW})$. This is admissible since the rank of $\widehat{P}$ is the same as that of $\widehat{G}$ in the limit. Nielsen and Shimotsu (2007) point out that this approach works better in simulations and also recommend to use the bandwidth $n(T) = m_1^{-0.3}$. The cointegrating rank estimate is consistent for $r \in \{0, \ldots, p-1\}$. It is applicable for systems of dimension $p \geq 2$, and it does not impose restrictions on d and b. A similar rank estimation procedure based on the average of finitely many tapered periodogram ordinates local to the origin was also proposed by Chen and Hurvich (2003).

The inconsistency of the multivariate local Whittle estimator under fractional cointegration is the basis for a test procedure originally proposed by Marinucci and Robinson (2001). They suggest a Hausman-type test that compares multivariate and univariate local Whittle estimates. Under the null hypothesis of no cointegration the multivariate estimator is efficient and both are consistent, whereas under the alternative of fractional cointegration the univariate estimator remains consistent, while the multivariate one does not.

This idea is formalized by Robinson (2008a). The test statistic is based on the objective function of the multivariate local Whittle estimator (cf. Lobato (1999), Shimotsu (2007)) $S(d) = \log \det \widehat{G}^*(d) - \frac{2pd}{m} \sum_{j=1}^{m} \log \lambda_j$ with $\widehat{G}^*(d) = \frac{1}{m} \sum_{j=1}^{m} I_X(\lambda_j) \lambda_j^{2d}$ and its derivative

$$s^*(d) = \text{tr}\left(\widehat{G}^*(d)^{-1} \widehat{H}^*(d) \right) \qquad (8)$$

---

[1] We follow the notation of Nielsen and Shimotsu (2007) and use m_1 for the bandwidth in the estimation of $G(d)$ and m for that of d. Note that Robinson and Yajima (2002) chose the opposite notation.

with $\widehat{H}^*(d) = \frac{1}{m} \sum_{j=1}^{m} \nu_j I_X(\lambda_j) \lambda_j^{2d}$ and $\nu_j = \log j - \frac{1}{m} \sum_{k=1}^{m} \log k$. Similar to the previous procedure, the memory parameter d is estimated by pooling the univariate estimates obtained by applying the local Whittle estimator to each of the component series. The equally weighted average is denoted by $\widehat{d}_{LW}$. To obtain a test statistic, the derivative $s^*(d)$ from (8) is evaluated at this averaged univariate estimate:

$$W_{Rob}^* = \frac{m s^*(\widehat{d}_{LW})^2}{N^2 \text{tr}(\widehat{F}^{*2}) - p} \tag{9}$$

with $\widehat{F}^* = \widehat{R}^{*-1/2} \widehat{G}^*(\widehat{d}_{LW}) \widehat{R}^{*-1/2}$ and $\widehat{R}^* = \text{diag}(\widehat{g}_{11}^*, \ldots, \widehat{g}_{pp}^*)$, where $\widehat{g}_{aa}^*$, $a = 1, \ldots, p$, are the diagonal elements of $\widehat{G}^*(\widehat{d}_{LW})$. The scaled derivative $m^{1/2} s^*(\widehat{d}_{LW})$ is asymptotically normal so that the test follows a χ_1^2-distribution if appropriately standardized by the term in the denominator.

The test generates power because $G(d)$ is singular under the alternative of fractional cointegration so that the inverse $\widehat{G}^*(\widehat{d}_{LW})^{-1}$ of the estimate and consequently the trace $s^*(\widehat{d}_{LW})$ become large. This is a score-type test that avoids the calculation of the multivariate local Whittle estimator that can be numerically expansive. Since the efficiency of the multivariate estimate is obtained with a single Newton step from the univariate estimate in direction of the multivariate one, $s^*(\widetilde{d}_{LW})$ is directly proportionate to the difference between the efficient and the inefficient estimate.

This test allows series of dimensions larger than two, but it is restricted to processes with $d \in (-1/2, 1/2)$ and focuses on the empirically relevant range $d \in (0, 1/2)$. A non-stationary extension based on a trimmed version of the local Whittle estimator is proposed, but the size and power properties of this test in simulations appear to depend heavily on the sample size.[2]

An alternative way to allow for non-stationary processes would be to base the test on the objective function of the multivariate exact local Whittle estimator (as in Shimotsu (2012), but without allowing for fractional cointegration) and univariate ELW estimates. Since the exact local Whittle estimates have the same asymptotic properties as the local Whittle estimate for $d \in (-1/2, 1/2)$, the test would have the same limiting distribution.

For a bivariate process with known $d \in (0, 1]$, Souza et al. (2018) propose a test based on an estimate of b obtained from the determinant of the trimmed and truncated spectral matrix of the fractionally differenced process via a log-periodogram regression. Denote the fractionally differenced process by $\Delta^d X_t = (\Delta^d X_{1t}, \Delta^d X_{2t})'$ with spectral density matrix $f_{\Delta^d}(\lambda)$, then the determinant $D_{\Delta^d}(\lambda)$ of $f_{\Delta^d}(\lambda)$ depends on the memory reduction parameter $b \in [0, d]$ and can be approximated by

$$D_{\Delta^d}(\lambda) \sim \tilde{g} |1 - e^{-i\lambda}|^{2b}, \quad \text{as} \quad \lambda \to 0^+, \tag{10}$$

where $\tilde{g}$ is a constant and finite scalar. Under cointegration, $f_{\Delta^d}(\lambda)$ does not have full rank near the origin (like G in (1)) so that its determinant $D_{\Delta^d}(\lambda)$ approaches zero as $\lambda \to 0^+$. The memory reduction b can be estimated from the logged version of Eq. (10) using a log-periodogram type regression,

---

[2] These results are available from the authors upon request.

$$\log D_{\Delta^d}(\lambda) \sim \log \tilde{g} + 2b \log |1 - e^{-i\lambda}| + \log \frac{\tilde{g}^*(\lambda)}{\tilde{g}}, \quad \text{as} \quad \lambda \to 0^+,$$

where $\lim_{\lambda \to 0^+} \tilde{g}^*(\lambda) = \tilde{g}$.

In order to make the estimation of b feasible, the empirical determinant $\widehat{D}_{\Delta^d}(\lambda)$ has to be calculated from an estimate $\widehat{f}_{\Delta^d}(\lambda)$ of the spectral density at the Fourier frequencies with $j = l, l + (2l - 1), l + 2(2l - 1), \ldots, m - (2l - 1), m$ with $l + 1 < m < T$. The latter is obtained from the locally averaged periodogram $\widehat{f}_{\Delta^d}(\lambda_j) = \frac{1}{2l-1} \sum_{k=j-(l-1)}^{j+(l-1)} I_{\Delta^d}(\lambda_k)$, where $I_{\Delta^d}(\lambda_k)$ is the periodogram of $\Delta^d X_t$. At each j the estimate $\widehat{f}_{\Delta^d}(\lambda_j)$ is thus a local average of the periodogram at frequency j and the $l - 1$ frequencies to its left and right and the λ_j are spaced so that the local averages are non-overlapping.

The resulting estimator for the cointegrating strength b is given by

$$\widehat{b}_{GPH} = \left(\sum_{j=l+1}^{m} \tilde{Z}_j^{*2} \right)^{-1} \sum_{j=l+1}^{m} \tilde{Z}_j^* \log \widehat{D}_{\Delta^d}(\lambda_j),$$

where $\tilde{Z}_j^* = Z_j^* - \bar{Z}^*$, $Z_j^* = \log |1 - e^{i\lambda}| = \log(2 - 2\cos(\lambda_j))$, and $\bar{Z}^*$ is the mean of the Z_j^*. Under the null hypothesis of no fractional cointegration we have $b = 0$. Under this condition, and assuming that l and m fulfill the condition $\frac{l+1}{m} + \frac{m}{T} + \frac{1}{m} + \frac{\log m}{m} \to 0$ as $T \to \infty$, the estimate $\widehat{b}_{GPH}$ is consistent and asymptotic normal with variance $\sigma_b^2 = \frac{1}{m}(\Psi^{(1)}(2l + 1) + \Psi^{(1)}(2l))$, where $\Psi^{(1)}(x) = \frac{\delta^2 \log \Gamma(x)}{\delta x^2}$ is the polygamma function of order 1 and $\Gamma(\cdot)$ denotes the gamma function.

The null hypothesis of no fractional cointegration can thus be tested using a simple t-test:

$$W_{SRFB} = \frac{\widehat{b}_{GPH}}{\sigma_b} \xrightarrow{d} N(0, 1). \tag{11}$$

The method has no restrictions regarding the range of d and b but is only applicable to bivariate processes. For practical purposes, d is usually unknown and has to be estimated, but as shown in our simulation study in Sect. 4 this has no severe implications for the quality of the test. However, a thorough theoretical examination of this aspect would be interesting for further research. Note that the work by Velasco (2003a) might help on this issue as he introduced a similar estimate focusing on knowing or not-knowing the true residuals.

3.2 Tests based on cointegrating residuals

By the definition of fractional cointegration the memory of the linear combination $v_t = \beta' X_t$ is lower than that of X_t itself. Under the null hypothesis of no fractional cointegration one can still write $v_t = \beta' X_t = y_t - \tilde{\beta} x_t$, since y_t can still depend on the values of the other components of X_t. The difference to the cointegrated case is

only that $d_v = d$. It is therefore natural to test for fractional cointegration by testing $d_v = d$ (or $b = 0$) versus $d_v < d$ (or $b > 0$) based on an estimate $\widehat{v}_t$ of the potential cointegrating residual.

Under weak non-stationary fractional cointegration, i.e., $d > d_v > 1/2$, Marmol and Velasco (2004) suggest a Hausman (1978)-type F-test that compares the OLS estimate $\widehat{\beta}^{OLS}$ of the cointegrating vector with an alternative estimate $\widehat{\beta}^{NB}$ with opposite consistency characteristics.

The OLS estimator $\widehat{\beta}^{OLS}$ is consistent for $\tilde{\beta}$ under the alternative (as long as $d > 1/2$) but inconsistent under the null hypothesis. Marmol and Velasco (2004) propose an alternative estimator $\widehat{\beta}^{NB}$ that is consistent for the vector $\tilde{\beta}$ under the null hypothesis but inconsistent under the alternative. The estimator is given by

$$\widehat{\beta}^{NB}(\widehat{d}_x, \widehat{d}_v) = \widehat{G}_{xx}^{MV}(\widehat{d}_x)^{-1}\widehat{g}_{xy}^{MV}(\widehat{d}_v),$$

where $\widehat{G}_{xx}^{MV}(d) = \frac{2\pi}{m_2}\sum_{j=1}^{m_2} \tilde{\Lambda}_j(d)^{-1}\text{Re}\left\{I_{xx}(\lambda_j)\right\}\tilde{\Lambda}_j^{-1}(d)$, and $\widehat{g}_{xy}^{MV}(d) = \sum_{j=1}^{m_2} \text{Re}\, I_{xy}(\lambda_j)\lambda_j^{2(d-1)}$, $\tilde{\Lambda}_j(d) = \text{diag}(\lambda_j^{1-d}, \ldots, \lambda_j^{1-d})$ and where $I_{xx}(\lambda_j)$ and $I_{xy}(\lambda_j)$ are the respective elements of the periodogram $I_{\Delta X \Delta X}(\lambda_j)$ of the differenced process ΔX_t and m_2 is subject to the usual bandwidth conditions. The estimator is closely related to the narrow band least squares estimator $\widehat{\beta}_m$ from (6) but uses a rescaled version of the periodogram. In fact, $\widehat{\beta}^{NB}(0, 0)$ would be equivalent to the NBLS estimate based only on the real part of the periodogram. Note that Nielsen (2005) introduced a very similar GLS-type estimate $\widehat{\beta}^{NB}(d, d)$.

Inconsistency under the alternative is only obtained through the choice $\widehat{\beta}^{NB}(\widehat{d}_x, \widehat{d}_v)$, where $\widehat{d}_v$ is estimated from the OLS residuals. Since under the alternative $\widehat{v}_t^{OLS}$ is a consistent estimate of the cointegrating residual, $\widehat{d}_v \to d_v < d$, whereas $\widehat{d}_x$ is estimated from the original series and is consistent for d. Under the null hypothesis, on the other hand, $\widehat{\beta}^{OLS}$ is inconsistent so that $\widehat{v}_t^{OLS}$ is just some linear combination of $I(d)$ series, $\widehat{d}_v \to d$, and $\widehat{\beta}^{NB}(\widehat{d}_x, \widehat{d}_v)$ is consistent for $\tilde{\beta}$.

Since the process is non-stationary, the memory is estimated by local Whittle from the differenced process. Alternatively, d could be estimated using a tapered local Whittle estimator, or by the exact or fully extended local Whittle estimator. The test statistic compares both estimates of $\tilde{\beta}$ where the normalizing variance $\widehat{V}^{MV}$ is estimated from the periodogram of the OLS residuals $\widehat{v}_t^{OLS}$ and that of x_t so that

$$\widehat{V}_{MV} = \left(\sum_{j=-m}^{m} I_{xx}(\lambda_j)\right)^{-1}\sum_{j=-m}^{m} I_{xx}(\lambda_j)I_{\widehat{v}\widehat{v}}(\lambda_j)\left(\sum_{j=-m}^{m} I_{xx}(\lambda_j)\right)^{-1}.$$

This leads to the test statistic

$$W_{MV} = \frac{1}{p-1}\left(\widehat{\beta}^{OLS} - \widehat{\beta}^{NB}\right)'\widehat{V}_{MV}^{-1}\left(\widehat{\beta}^{OLS} - \widehat{\beta}^{NB}\right). \tag{12}$$

The choices of m and m_2 are not linked, but both have to satisfy the condition $(m^{d-2} + m^{\gamma-1}\log T)\log^2 T + \frac{m}{T} \to 0$ as $T \to \infty$, with $\gamma > 0$ which is fulfilled if $m \sim T^\eta$,

$\eta \in (0, 1)$. The asymptotic distribution is non-standard and depends on the memory parameter d. It is given by

$$
W_{MV} \xrightarrow{d} \frac{1}{p-1} \int_0^1 W_y(d; r) W_x(d; r)' dr \, V^{-1} \int_0^1 W_x(d; r) W_y(d; r) dr,
$$

with $V = \int_0^1 \gamma_R(s) \left\{ \gamma_{xx}(s) + \gamma'_{xx}(s) + \gamma_{xx}(1-s) + \gamma'_{xx}(1-s) \right\} ds$, $\gamma_R(s) = \int_0^{1-s} W_y(d; r) W_y(d; r+s) dr$, and $\gamma_{xx}(s) = \int_0^{1-s} W_x(d; r) W_x(d; r+s)' dr$, where $W_y(d; r)$ is a fractional Brownian bridge, and $W_x(d; r)$ is a $p \times 1$ vector of independent fractional Brownian bridges.

Critical values are tabulated in Marmol and Velasco (2004) for dimensions up to $p = 5$ and different forms of detrending that affect the type of the fractional Brownian bridges. The test statistic W_{MV} diverges under the alternative since both $\widehat{\beta}^{NB}$ and $\widehat{V}_{MV}^{-1}$ diverge under fractional cointegration. Although the consistency of the test is derived assuming $d > d_v > 0.5$, Marmol and Velasco (2004) state that the test remains consistent if the stationarity border is crossed by the cointegrating residuals, i.e. $d > 0.5 > d_v$. Our simulations in Sect. 4 confirm this.

A direct residual-based test is proposed by Chen and Hurvich (2006) who estimate the possible cointegrating subspaces using eigenvectors of the averaged periodogram local to the zero frequency. The process X_t is assumed to be stationary after taking $(q - 1)$ integer differences which allows $d \in (q - 1.5, q - 0.5)$. In order to account for possible over-differentiation the complex-valued taper $h_t = 0.5(1 - e^{i2\pi t/T})$ of Hurvich and Chen (2000) is applied to the data. The tapered discrete Fourier transform (DFT) and periodogram of X_t are defined by

$$
w_X^{tap}(\lambda_j) = \frac{1}{\sqrt{2\pi \sum_t |h_t^{(q-1)}|^2}} \sum_{t=1}^T h_t^{(q-1)} X_t e^{i\lambda_j t},
$$

$$
I_X^{tap}(\lambda_j) = w_X(\lambda_j) \overline{w_X(\lambda_j)}.
$$

Based on the tapered periodogram, define the averaged periodogram matrix of X_t by $I_X^{av}(\lambda_j) = \sum_{j=1}^{m_3} \text{Re} \left(I_X^{tap}(\lambda_j) \right)$, where m_3 is a fixed positive integer fulfilling $m_3 > p + 3$. The eigenvalues of $I_X^{av}(\lambda_j)$ sorted in descending order are denoted by $\widehat{\delta}_{a, I_X^{av}}$ and the corresponding eigenvectors are given by $\widehat{\chi}_{a, I_X^{av}}$, for $a = 1, \ldots, p$. Under the alternative hypothesis, if there are $r > 0$ cointegrating relationships, the matrix consisting of the first r eigenvectors provides a consistent estimate of the cointegrating subspace.

To construct a test for the null hypothesis of no fractional cointegration the potential cointegrating residuals v_t are estimated by multiplying X_t with the eigenvectors $\widehat{\chi}_{a, I_X^{av}}$ so that $\widehat{v}_{at}^{av} = \widehat{\chi}_{a, I_X^{av}}' X_t$, for $a = 1, \ldots, p$.

The memory of the p residual processes is estimated with the local Whittle estimator using bandwidth m but calculated using shifted Fourier frequencies $\lambda_{\tilde{j}}$ with $\tilde{j} = j + (q - 1)/2$ to account for the tapering of order q. These estimates are denoted by $\widehat{d}_{v_a, \widetilde{LW}}$, and they remain consistent and asymptotic normal.

Since there can be at most $p - 1$ cointegrating relationships in a p-dimensional time series, the first residual corresponding to the largest eigenvalue cannot be a cointegrating residual. Its memory must therefore equal the common memory d of X_t. In contrast, the last residual $\widehat{v}_{pt}^{av}$ corresponding to the smallest eigenvalue is most likely to be a cointegrating residual if there is cointegration so that its memory is reduced by b under cointegration.

The test idea of Chen and Hurvich (2006) is therefore to compare the estimated memory orders from the residual series $\widehat{v}_{1t}^{av}$ and $\widehat{v}_{pt}^{av}$ that correspond to $\widehat{d}$ (first residual) and $\widehat{d}_v$ (last residual). Chen and Hurvich (2006) show that

$$\sqrt{m}\left(\widehat{d}_{v_a,\widetilde{LW}} - \widehat{d}_{v_b,\widetilde{LW}}\right) \xrightarrow{d} N\left(0, V_{CH,q}\left(1 - \frac{G_{ab}^2}{G_{aa}G_{bb}}\right)\right)$$

with $V_{CH,q} = \frac{1}{2}\frac{\Gamma(4q-3)\Gamma^4(q)}{\Gamma^4(2q-1)}$. A conservative test statistic is therefore given by

$$W_{CH} = \sqrt{m}\,\frac{\left(\widehat{d}_{v_1,\widetilde{LW}} - \widehat{d}_{v_p,\widetilde{LW}}\right)}{\sqrt{V_{CH,q}}}. \tag{13}$$

The test rejects if W_{CH} is larger than the standard normal quantile $z_{1-\alpha/2}$. It is very versatile, since it does not impose restrictions on the cointegration strengh b and can be applied to stationary as well as non-stationary long-memory processes, but it requires a priori knowledge about the location of d in the parameter space to determine the order of differencing.

Hualde and Velasco (2008) propose another testing strategy in a residual-based regression framework. As before, the series X_t is partitioned such that $X_t = (y_t, x_t')'$ and they consider the single-equation regression $y_t = \tilde{\beta}x_t + v_t$.

The test idea is based on the observation that the fractionally differenced residual $\Delta^{d_x}v_t$ is unrelated to the long-run level of x_t under the null hypothesis. This is because $\Delta^{d_x}v_t$ is $I(0)$ and x_t is $I(d)$. The cross-spectrum of x_t and $\Delta^{d_x}v_t$ should therefore be zero at frequencies local to zero. Possible dependence between the short-run components u_t and e_t in (3) would manifest itself in form of a non-zero cross-spectrum at higher frequencies.

The test statistic of Hualde and Velasco (2008) is therefore based on the quantity $\widehat{\tau}_m$ defined as

$$\widehat{\tau}_m = \sum_{j=1}^{m} w_x(-\lambda_j)\zeta(\lambda_j)w_{\Delta^{d_v,d}X}(\lambda_j)$$

where $\Delta^{d_v,d}X_t = \left(\Delta^{\widehat{d}_v}y_t, \Delta^{\widehat{d}}x_t'\right)'$ and $\zeta(\lambda_j) = (1, 0'_{p-1})\,\widehat{f}_X(\lambda_j)^{-1}$. The projection vector $\zeta(\lambda_j)$ estimates the DFT of the residual process v_t from $w_{\Delta^{d_v,d}X}(\lambda_j)$—the DFT of the fractionally differenced process $\Delta^{d_v,d}X_t$. As usual for these semiparametric approaches, it is assumed that $m \leq T/2$ and $m/T \to 0$, as $T \to \infty$. This leads to the test statistic

$$W_{HV} = \widehat{\tau}_m' \widehat{V}_{HV}^{-1} \widehat{\tau}_m \tag{14}$$

with $\widehat{V}_{HV} = \sum_{j=0}^m a_j \operatorname{Re} \kappa(\lambda_j) I_{XX}(\lambda_j)$, and $\kappa(\lambda_j) = \zeta(\lambda_j) (1, 0_{p-1}')'$, where the weights are defined by $a_j = 1$ if $j \in \{0, T/2\}$ and $a_j = 2$ otherwise. Under the null hypothesis this test statistic follows an asymptotic χ_{p-1}^2-distribution. Under the alternative the test develops power, since d_v is estimated from the NBLS estimate of the cointegrating residuals. Since these have reduced memory under the alternative, the first component of $\Delta^{d_v, d} X_t$ (y_t) is $I(b)$ instead of $I(0)$ and the cross spectrum of the underdifferenced estimate of v_t and x_t in $\widehat{\tau}_m$ becomes non-zero. As before, the memory orders are estimated using consistent estimators that account for the (possible) non-stationarity of the data—for example the exact local Whittle estimator of Shimotsu and Phillips (2005).

A modified test with more power in bivariate systems $X_t = (X_{1t}, X_{2t})'$ is calculated with $\widetilde{\tau}_m$ instead of $\widehat{\tau}_m$:

$$\widetilde{\tau}_m = \sum_{j=0}^m a_j \frac{\operatorname{Re}\left(I_{\Delta^{\widehat{d}_v} X_1, X_2}(\lambda_j) - \frac{\widetilde{f}_{12}(\lambda_j)}{\widehat{f}_{22}(\lambda_j)} I_{\Delta^{\widehat{d}_v} X_2, X_2}(\lambda_j)\right)}{\widehat{f}_{11}(\lambda_j) - \frac{\widehat{f}_{12}(\lambda_j)\widehat{f}_{21}(\lambda_j)}{\widehat{f}_{22}(\lambda_j)}}.$$

Here, the respective elements of the spectral matrices of differenced processes $\widehat{f}_{\Delta^{\widehat{d}}}(\lambda_j) = \frac{1}{2m+1} \sum_{k=j-m}^{j+m} I_{\Delta^{\widehat{d}} X}(\lambda_k)$, $\widetilde{f}_{\Delta^{\widehat{d}_v}}(\lambda_j) = \frac{1}{2m+1} \sum_{k=j-m}^{j+m} I_{\Delta^{\widehat{d}_v} X}(\lambda_k)$ are denoted by $\widehat{f}_{ab}(\lambda_j)$ and $\widetilde{f}_{ab}(\lambda_j)$ with $a, b \in \{1, 2\}$. This is the same as $\widehat{\tau}_m$ but with $\widehat{f}_{12}(\lambda_j)$ replaced by $\widetilde{f}_{12}(\lambda_j)$ that is constructed using $\widehat{d}_v$ so that it also diverges under the alternative and constitutes an additional source of power. The asymptotic χ_{p-1}^2-distribution is unaffected by this modification.

It is not necessary to impose any restrictions on the range of d and d_v except for those implied by fractional cointegration, and processes of dimensions higher than two are allowed. The asymptotic χ_{p-1}^2 distribution depends only on the dimension of the process. Furthermore, the memory parameters are allowed to differ as long as two components of X_t share the same memory parameter and the vector is sorted so that the component with the highest memory comes first.

Nielsen (2010) introduces a sequential testing approach to test fractional cointegration and to determine the cointegrating rank. The method is based on a variance-ratio statistic and imposes the assumption that the process X_t is non-stationary and the potential cointegrating residual process is stationary with $d_v < 0.5 < d$.

Denote the demeaned process by $Z_t = X_t - \overline{X}_t$, where $\overline{X}_t$ is the vector of arithmetic means of the component series. The fractionally integrated version of Z_t is denoted by $\widetilde{Z}_t = \Delta^{-\epsilon} Z_t$. Then the variance ratio is given by $K_T(\epsilon) = A_T C_T^{-1}$, with $A_T = \sum_{t=1}^T Z_t Z_t'$, and $C_T = \sum_{t=1}^T \widetilde{Z}_t \widetilde{Z}_t'$. Taking the ratio has the advantage of eliminating the processes' variance from the asymptotic distribution. The eigenvalues of $K_T(\epsilon)$ sorted in ascending order are denoted by $\delta_{a,K}$ with $a = 1, \ldots, p$.

Similar to the spectral matrix G, the rank of $K_T(\epsilon)$ is reduced to $p - r$ under fractional cointegration. This leads to a non-parametric trace statistic whose structure is similar to the trace statistic of Johansen (1991) in the parametric context

$$W_{Niel}(\epsilon) = T^{2\epsilon} \sum_{k=1}^{p-r} \widehat{\delta}_{k,K}, \quad r = 1, \ldots, p-1, \tag{15}$$

where r is the number of cointegrating relations under the null hypothesis. Using (15) the cointegrating rank can be determined by a sequence of tests of the null hypothesis $H_0\colon r = r_0$ vs. $H_1\colon r > r_0$.

The limiting distribution is given by

$$W_{Niel}(\epsilon) \xrightarrow{d} \operatorname{tr}\left\{ \left[\int_0^1 W_{n-r}(d;s) W_{n-r}(d;s)' ds \right. \right.$$
$$\left. \left. \left(\int_0^1 \widetilde{W}_{n-r}(d+\epsilon;s) \widetilde{W}_{n-r}(d+\epsilon;s)' ds \right)^{-1} \right] \right\},$$

where $W_{n-r}(d,u) = B_d^{n-r}(u) - \int_0^1 B_d^{n-r}(v)dv$, $\widetilde{W}_{n-r}(d+\epsilon;u) = B_{d+\epsilon}^{n-r}(u) - \int_0^u \frac{(u-v)^{\epsilon-1}}{\Gamma(\epsilon)} dv \int_0^1 B_{d+\epsilon}^{n-r}(v)dv$, B_d^{n-r} is a $n-r$ dimensional vector of mutually independent standard fractional Brownian motions of type II, and the Brownian motions driving the fractional Brownian motions B_d^{n-r} and $B_{d+\epsilon}^{n-r}$ are identical.

This asymptotic distribution is non-standard and depends on the dimension p, the cointegrating rank r, the order of fractional integration ϵ and d. In practice d can be estimated consistently, and the other parameters are known. Critical values for $d = 1$, $\epsilon = 0.1$, and $p - r = 1, 2, \ldots, 8$ are given by Nielsen (2010), who recommends to use $\epsilon = 0.1$ to integrate the process because it leads to higher power than larger values whereas smaller values improve power slightly but lead to size distortions at the same time. For more details confer Nielsen (2009). Note that choosing a different order of fractional summation changes the limiting distribution which implies that the test performance is free from user-chosen tuning parameters.

To see why this test can be considered to be residual-based, note that

$$\widehat{\delta}_{a,K} = \frac{\widehat{\eta_a}' A_T \widehat{\eta_a}}{\widehat{\eta_a}' C_T \widehat{\eta_a}} = \frac{\sum_{t=1}^{T} \widehat{v}_t^2}{\sum_{t=1}^{T} \widetilde{v}_t^2},$$

where $\widehat{\eta}_a$ denotes the eigenvector corresponding to $\widehat{\delta}_{a,K}$. Since the first r eigenvectors are consistent estimates of the cointegrating space, the first r eigenvalues are thus given by the ratio of the sum of the squared cointegrating residuals and the sum of squares of their ϵ times integrated version $\widetilde{v}_t$. Here the squares are estimators of the respective process variances and it is assumed that $d > 1/2 > d_v$. Therefore, under the null hypothesis of no fractional cointegration the enumerator grows with rate $\mathcal{O}_P(T^{2d})$ and the more persistent denominator grows with rate $\mathcal{O}_P(T^{2(d+\epsilon)})$, so that the eigenvalue has rate $\mathcal{O}_P(T^{-2\epsilon})$. Under the alternative of fractional cointegration with $d_v < 1/2$, the process v_t is stationary so that the process variance is finite and the enumerator grows with rate $\mathcal{O}_P(T)$. The denominator that may or may not be stationary due to the integration with ϵ is $\mathcal{O}_P(T^{\max\{1/2, d-b+\epsilon\}})$. Consequently, the

eigenvalue is $\mathcal{O}_P(T^{\min\{0,1-2(d-b+\epsilon)\}})$, so that it goes to zero more slowly than under the null hypothesis.

The test is restrictive in that it requires non-stationary processes and, preferably, stationary residual processes, but as shown by his Monte Carlo simulation the test still exhibits power if $d_v > 0.5$ and $b > 0$. Furthermore, it is applicable to multivariate systems and is able to estimate the number of cointegrating relations.

Wang et al. (2015) propose a simple residual-based test in a bivariate setting where $X_t = (X_{1t}, X_{2t})'$. The test statistic is based on the partial sum of $\Delta^{d_v} X_{2t}$, which is the demeaned second component series fractionally differenced with the memory order of the potential cointegrating residual v_t. It is given by

$$W_{WWC} = T^{-1/2} \frac{\sum_{t=1}^{T} \Delta^{\widehat{d}_v} Z_{2t}}{\sqrt{2\pi \widehat{f}_{22}(0)}}, \tag{16}$$

where f_{22} is the spectral density of either u_{2t} or e_t in (3), depending on whether a triangular model or a common-components model is assumed.

Under the null hypothesis $d_v = d$ so that $\Delta^{d_v} Z_{2t}$ is $I(0)$ and the appropriately rescaled sum is asymptotically standard normal. Under the alternative $\Delta^{d_v} Z_{2t}$ is $I(b)$, so that the test statistic diverges with rate $\mathcal{O}_P(T^b)$.

To make this test statistic feasible the spectral density f_{22} can be estimated from the periodogram of the fractionally differenced process $\Delta^{\widehat{d}} Z_{2t}$ following the approach of Hualde (2013): $\widehat{f}_{22}(0) = \frac{1}{(2m+1)} \sum_{j=-m}^{m} I_{\Delta^{\widehat{d}} Z_2}(\lambda_j)$, where $I_{\Delta^{\widehat{d}} Z_2}(\lambda_j)$ is the periodogram of $\Delta^{\widehat{d}} Z_{2t}$.

While Wang et al. (2015) are agnostic about the method that is used for the estimation of the memory parameters d and d_v, they assume that $d > 1/2$ so that the cointegrating vector can be estimated using ordinary least squares. The memory orders can be estimated from $\widehat{v}_t^{OLS}$ and Z_{2t} using any of the common semiparametric estimates such as ELW with bandwidth m as in $\widehat{f}_{22}$ that fulfills the usual bandwidth conditions. The method does not impose any restrictions on the fractional cointegrating strength b. As the Monte Carlo simulations below show, the non-stationarity requirement ($d > 1/2$) can be circumvented if the cointegrating residual v_t is based on the NBLS estimate of the cointegrating vector instead of the OLS estimate.

Zhang et al. (2019) propose an alternative estimator of the cointegrating space that is based on the eigenvectors of the non-negative matrix $\widehat{M} = \sum_{j=0}^{j_0} \widehat{\Omega}_Z(j)\widehat{\Omega}_Z(j)'$, where $\widehat{\Omega}_Z(j) = \frac{1}{T} \sum_{t=1}^{T-j} Z_{t+j} Z_t'$ is the autocovariance matrix at lag j and j_0 is a fixed integer. The matrix $\widehat{M}$ is thus the sum of the outer products of the first j_0 autocovariance matrices with themselves. The outer product is used instead of the covariance matrices $\widehat{\Omega}_Z(j)$ to ensure that there is no information cancellation over different lags in $\widehat{M}$. It is assumed that $d > 0.5$ and $d_v < 0.5$.

The eigenvalues of $\widehat{M}$ in descending order are denoted by $\widehat{\delta}_{a,M}$ for $a = 1, \ldots, p$ and the corresponding eigenvectors are denoted by $\widehat{\chi}_{a,M}$. Similar to the matrix G in (1), the first $p - r$ eigenvalues of M are non-zero, whereas the remaining r are zero. For known r the eigenvectors corresponding to the r smallest eigenvalues provide a consistent estimate of the cointegrating space.

If r is unknown, the p potential cointegrating residuals are estimated using the eigenvectors so that $\widehat{v}_{at}^M = \widehat{\chi}_{a,M}' X_t$. By the same argument as in the procedure of Chen and Hurvich (2006), the residual corresponding to the smallest eigenvalue is most likely a cointegrating residual with reduced memory of $d_v = d - b$ and the residual corresponding to the largest eigenvalue is $I(d)$.

The cointegrating rank can be estimated using a simple criterion based on the summed autocorrelations of the potential cointegrating residuals. Define

$$Q_a(k_0) = \sum_{k=1}^{k_0} \widehat{\rho}_a(k),$$

$$\text{with } \widehat{\rho}_a(k) = \frac{\frac{1}{T-k}\sum_{t=1}^{T-k}(\widehat{v}_{a,t+k}^M - \overline{\widehat{v}_{at}^M})(\widehat{v}_{at} - \overline{\widehat{v}_{at}^M})'}{\frac{1}{T}\sum_{t=1}^{T}(\widehat{v}_{at}^M - \overline{\widehat{v}_{at}^M})^2},$$

where $\overline{\widehat{v}_{at}^M}$ is the mean of $\widehat{v}_{at}^M$. The cointegrating rank estimator counts the instances when the averaged autocorrelation is smaller than a threshold $c_0 \in (0, 1)$:

$$\widehat{r}_{ZRY} = \sum_{a=1}^{p} \mathbb{1}\left\{\frac{Q_a(k_0)}{k_0} < c_0\right\}. \tag{17}$$

If the residual $\widehat{v}_{at}^M$ is stationary ($d_v < 1/2$), the rescaled sum of autocorrelations $Q_a(k_0)/k_0$ converges to zero asymptotically for $k_0 \to \infty$, since the autocorrelations are asymptotically proportionate to k^{2d_v-1}. Under certain regularity conditions this estimate is consistent. Even though the consistency is only proven for $r \geq 1$ in Theorem 4.2 of Zhang et al. (2019), our simulations below show that it also works well in discriminating between $r = 0$ and $r = 1$.

It should be noted that the authors define $r = p$ if all components of X_t are $I(0)$. This leads to some abuse of notation and r cannot be interpreted as the cointegrating rank in a narrow sense. Based on their simulations Zhang et al. (2019) recommend to use $j_0 = 5$, $k_0 = 20$ and $c_0 = 0.3$. The estimator is easy to implement and applicable to higher dimensional processes. However, the requirement of $d > 0.5$ and $d_v < 0.5$ is restrictive.

4 Monte Carlo Study

The asymptotic properties of all tests and rank estimates presented in Sect. 3 are derived by the respective authors, and some of them also present simulations to explore the finite sample results of the test statistics. This however is not the case for all tests and a comprehensive comparative study suited to guide the choice of appropriate methods in practical applications is entirely missing. To close this gap, we conduct an extensive Monte Carlo study. In addition to general results, we are particularly interested in answering two empirically motivated questions.

(i) How does correlation between the underlying short-run components influence the size of the tests? This question is important, since applied researchers will generally want to test for fractional cointegration if two related series seem to be co-moving. Similar trajectories, however, can also be generated by persistent processes with highly correlated innovations. Tests for the null hypothesis of no fractional cointegration should therefore be robust to a relatively high degree of correlation between the short-run components of the series.

(ii) Is there a notable difference in the power of the tests depending on whether the data is generated from a triangular model or from a common-components model? Both models are used in the literature to motivate and construct testing procedures, but to our knowledge simulation results are typically based on the triangular representation. In practice, either model could be justified—depending on the application. For example, if one is considered with potential fractional cointegrating relationships between stock prices, it is not clear why one of the stock prices should be seen as a perturbed version of the other one (as it is the case in the triangular model that treats the series in an asymmetric way) so that the common-components model is more suitable. In contrast to that, in the case of the potential parity between implied volatility and the expected average realized volatility over the next month (the so-called implied-realized parity analyzed by Christensen and Prabhala (1998), Christensen and Nielsen (2006b), and Nielsen (2007), among others), there is theoretical reason to assume that the implied volatility is a perturbed version of the expected average future realized volatility, since it contains a variance-risk premium (cf. Chernov (2007)). Therefore, a triangular model is more suitable.

We focus on three data generating processes (DGPs) based on the general model from Eqs. (3) to (5). For simplicity we set $c_1 = c_2 = 0$ and $b = b_1 = b_2$ so that the processes are mean zero and have a common memory reduction parameter. A simple bivariate model without fractional cointegration is constructed by setting $\xi_1 = \xi_2 = 0$. This model—referred to as (size) DGP1—is given by

$$X_{1t} = \Delta^{-d} u_{1t} \mathbb{1}\{t > 0\}, \tag{18}$$
$$X_{2t} = \Delta^{-d} u_{2t} \mathbb{1}\{t > 0\}, \tag{19}$$

where correlation between u_{1t} and u_{2t} is allowed. This is our size-DGP. For the power simulations, we consider a triangular model and a common-components model. In both cases we set $\xi_1 = \xi_2 = 1$ which implies a cointegrating vector of $\beta = (1, -1)'$. The triangular model DGP2 is given by

$$X_{1t} = Y_t + \Delta^{-(d-b)} u_{1t} \mathbb{1}\{t > 0\}, \tag{20}$$
$$X_{2t} = Y_t, \tag{21}$$

and the common-components model DGP3 is defined by

$$X_{1t} = Y_t + \Delta^{-(d-b)} u_{1t} \mathbb{1}\{t > 0\}, \tag{22}$$
$$X_{2t} = Y_t + \Delta^{-(d-b)} u_{2t} \mathbb{1}\{t > 0\}. \tag{23}$$

In both DGP2 and DGP3 we have $Y_t = \Delta^{-d} e_t \mathbb{1}\{t > 0\}$. The underlying short-run components u_{1t} and u_{2t}, or u_{1t} and e_t—depending on the DGP—have unit variance and correlation ρ.

We consider sample sizes of $T \in \{100, 500, 1000, 2500\}$ and values of $d \in \{0.4, 0.7, 1\}$ in the stationary and non-stationary region. Under fractional cointegration, the memory reduction b is linked to the value of d so that $b \in \{d/3, d\}$. Consequently, there is either a memory reduction to 0 if $b = d$ or a weaker form of cointegration if $b = d/3$. In order to examine the impact of correlation between the short-run components, we consider $\rho \in \{0, 0.45, 0.9, 0.99\}$. Note that the results for $b = d/3$, $\delta_m = 0.55$ and further robustness tests (other size DGPs and $p = 3$-dimensional processes) are available online as supplementary material.

The semiparametric nature of the tests and rank estimates requires several bandwidth choices. The memory estimation with (E)LW estimators involved in all methods is based on the bandwidth m that determines the number of frequencies included in the estimation. We use $m = \lfloor T^{\delta_m} \rfloor$ with $\delta_m = \{0.55, 0.75\}$ to account for sensitivities regarding bandwidth choice. With regard to the other bandwidth choices, we follow the recommendations by the authors: $m_1 = \lfloor T^{\delta_m - 0.1} \rfloor$ and $p(T) = m_1^{-0.3}$ for Nielsen and Shimotsu ([2007](#)) or Robinson and Yajima ([2002](#)), $l = 1$ for Souza et al. ([2018](#)), $m_3 = 25$ for Chen and Hurvich ([2006](#)), $c_0 = 0.3$, $j_0 = 5$ and $k_0 = 20$ for Zhang et al. ([2019](#)), and for Marmol and Velasco ([2004](#)) we set $m = \lfloor T^{2/3} \rfloor$ and $m_2 = \lfloor T^{\delta_m} \rfloor$. All tests are carried out allowing for a non-zero mean.

The results presented are based on 5000 replications and a nominal significance level of $\alpha = 0.05$. Since the tests impose different conditions on d and d_v, we mark the cells in the tables in bold where the methods have well-defined asymptotic properties and are supposed to deliver good results. In some cases the methods give satisfactory results beyond these limitations. For example, we implement the method of Wang et al. ([2015](#)) using a NBLS estimate of the cointegrating vector instead of the OLS estimate. This makes the test applicable in stationary time series as well as in non-stationary ones.

Since the limiting distributions of the non-pivotal test statistics of Marmol and Velasco ([2004](#)) and Nielsen ([2010](#)) depend on d and it is assumed that $d > 1/2$, it is unclear which critical values would be used in the stationary region. The respective fields are therefore left blank.

Further, it should be noted that the methods of Nielsen and Shimotsu ([2007](#)) (or Robinson and Yajima ([2002](#))) and Zhang et al. ([2019](#)) are not tests but rank estimates. Instead of the rejection frequency, we therefore report the ratio of correctly estimated cointegrating ranks. Therefore, the results cannot be interpreted as size or power, and in the size table and graphs the estimates should yield 0 instead of 0.05, since the estimates do not involve any significance level.

Table [1](#) displays size results based on DGP 1 with $\delta_m = 0.75$. The methods that have well defined asymptotic properties across all parameter constellations covered in the table are those of Nielsen and Shimotsu ([2007](#)), Chen and Hurvich ([2006](#)), Hualde and Velasco ([2008](#)), and Souza et al. ([2018](#)). It can be observed that all of these methods achieve good size properties for $\rho = 0$, except for the test of Chen and Hurvich ([2006](#)), when $d = 0.4$. Among those four procedures, only the tests of Souza et al. ([2018](#)) (disregarding the smallest sample) and Chen and Hurvich ([2006](#)) do not

Table 1 Size (*rank estimation) based on DGP1 with $\delta_m = 0.75$

Method	ρ	0			0.45			0.9			0.99		
	T/d	0.4	0.7	1	0.4	0.7	1	0.4	0.7	1	0.4	0.7	1
NS07*	100	0.000	0.000	0.014	0.130	0.138	0.241	1.000	1.000	0.998	1.000	1.000	1.000
	500	0.000	0.000	0.001	0.000	0.000	0.049	1.000	1.000	0.994	1.000	1.000	1.000
	1000	0.000	0.000	0.000	0.000	0.000	0.021	1.000	1.000	0.989	1.000	1.000	1.000
	2500	0.000	0.000	0.000	0.000	0.000	0.006	1.000	1.000	0.976	1.000	1.000	1.000
CH06	100	0.219	0.119	0.021	0.105	0.073	0.029	0.077	0.040	0.032	0.075	0.033	0.033
	500	0.177	0.058	0.032	0.076	0.031	0.021	0.060	0.026	0.017	0.064	0.027	0.020
	1000	0.136	0.051	0.031	0.064	0.025	0.018	0.045	0.018	0.018	0.046	0.020	0.017
	2500	0.129	0.044	0.023	0.059	0.023	0.015	0.039	0.018	0.012	0.040	0.017	0.012
HV08	100	0.001	0.008	0.015	0.017	0.022	0.024	0.128	0.098	0.058	0.312	0.263	0.130
	500	0.002	0.018	0.020	0.027	0.029	0.022	0.309	0.141	0.028	0.764	0.440	0.098
	1000	0.003	0.023	0.022	0.039	0.029	0.021	0.399	0.128	0.028	0.805	0.501	0.076
	2500	0.003	0.027	0.022	0.060	0.035	0.021	0.494	0.140	0.024	0.835	0.535	0.057
SRFB18	100	0.114	0.117	0.101	0.107	0.111	0.110	0.112	0.121	0.114	0.109	0.115	0.105
	500	0.054	0.054	0.049	0.049	0.052	0.052	0.046	0.055	0.046	0.056	0.052	0.045
	1000	0.041	0.047	0.042	0.043	0.043	0.040	0.044	0.047	0.048	0.045	0.043	0.044
	2500	0.037	0.039	0.029	0.035	0.037	0.037	0.036	0.036	0.034	0.033	0.035	0.035

Table 1 continued

Method	T/d	ρ 0			0.45			0.9			0.99		
		0.4	0.7	1	0.4	0.7	1	0.4	0.7	1	0.4	0.7	1
R08	100	**0.174**	0.183	0.092	**0.050**	0.059	0.066	**0.036**	0.048	0.039	**0.042**	0.045	0.039
	500	**0.233**	0.254	0.104	**0.049**	0.080	0.063	**0.052**	0.066	0.041	**0.049**	0.062	0.043
	1000	**0.239**	0.275	0.090	**0.053**	0.080	0.064	**0.051**	0.076	0.046	**0.057**	0.074	0.042
	2500	**0.265**	0.304	0.084	**0.055**	0.094	0.056	**0.054**	0.084	0.039	**0.050**	0.084	0.040
WWC15	100	0.080	**0.095**	**0.090**	0.079	**0.087**	**0.094**	0.081	**0.096**	**0.098**	0.078	**0.091**	**0.094**
	500	0.069	**0.068**	**0.075**	0.068	**0.068**	**0.074**	0.065	**0.074**	**0.074**	0.072	**0.072**	**0.068**
	1000	0.066	**0.064**	**0.067**	0.069	**0.062**	**0.066**	0.065	**0.056**	**0.067**	0.055	**0.060**	**0.068**
	2500	0.057	**0.059**	**0.056**	0.059	**0.052**	**0.061**	0.060	**0.057**	**0.059**	0.055	**0.049**	**0.061**
ZRY19*	100	0.101	**0.644**	**0.652**	0.071	**0.597**	**0.627**	0.057	**0.560**	**0.573**	0.062	**0.565**	**0.576**
	500	0.401	**0.058**	**0.000**	0.288	**0.043**	**0.000**	0.270	**0.036**	**0.000**	0.281	**0.030**	**0.000**
	1000	0.548	**0.000**	**0.000**	0.410	**0.000**	**0.000**	0.408	**0.000**	**0.000**	0.401	**0.000**	**0.000**
	2500	0.677	**0.000**	**0.000**	0.517	**0.000**	**0.000**	0.505	**0.000**	**0.000**	0.510	**0.000**	**0.000**
N10	100		0.035	0.059		0.041	0.061		0.053	0.053		0.058	0.063
	500		0.048	0.055		0.048	0.052		0.064	0.057		0.060	0.055
	1000		0.057	0.056		0.056	0.055		0.070	0.060		0.064	0.055
	2500		0.067	0.054		0.078	0.056		0.078	0.056		0.076	0.053
MV04	100		0.034	0.057		0.036	0.066		0.104	0.101		0.332	0.219
	500		0.041	0.052		0.047	0.056		0.158	0.092		0.496	0.234
	1000		0.039	0.052		0.047	0.055		0.141	0.081		0.490	0.224
	2500		0.036	0.052		0.045	0.056		0.127	0.069		0.442	0.197

We abbreviate the methods with the initial letters of the authors' names and the year of publication

over-reject if ρ increases.[3] For low values of d the test of Hualde and Velasco (2008) already becomes oversized for $\rho = 0.45$ and as ρ increases it becomes oversized for higher values of d, too. The rank estimation procedure of Robinson and Yajima (2002) and Nielsen and Shimotsu (2007) is even more affected and estimates a cointegrating rank of one in nearly all cases if $\rho \geq 0.9$.

In addition to the tests of Souza et al. (2018) and Chen and Hurvich (2006), the modified version of the test by Wang et al. (2015) that is based on the NBLS estimator instead of OLS also maintains satisfactory size properties across all values of ρ and d.

The group of procedures that is only applicable to non-stationary systems consists of Marmol and Velasco (2004), Nielsen (2010), and Zhang et al. (2019). It can be observed that the procedure of Marmol and Velasco (2004) behaves similar to that of Hualde and Velasco (2008) in the sense that it is very liberal for higher values of ρ and lower values of d. For non-stationary series and larger sample sizes the procedure by Zhang et al. (2019) estimates correctly the cointegrating rank to be zero—independently of the degree of correlation. The variance-ratio statistic of Nielsen (2010) turns out to be slightly liberal for $d = 0.7$ in larger samples, but holds the nominal size for $d = 1$ even in small samples. In particular, the performance is independent of the degree of correlation.

Finally, the test of Robinson (2008a) is only applicable for stationary systems. Here, it can be observed that the test does not hold its size for $\rho = 0$. This is because the Hausman-testing principle requires one of the estimates of the memory parameter to be more efficient than the other one, but the multivariate estimate is not more efficient in absence of correlation. For other values of ρ, however, the test has good size properties. Interestingly, the test also has good size properties if $d = 1$, even though it assumes stationarity. The intermediate value of $d = 0.7$, on the other hand, leads to a moderately oversized test.

Figure 1 analyzes the interaction between the degree of correlation ρ and the choice of the bandwidth δ_m. It shows the size of the tests in scatterplots where the results with no correlation ($\rho = 0$) are plotted against results with high correlation ($\rho = 0.99$). In the upper panel, tests that allow for stationary processes ($d = 0.4$) and in the lower panel ($d = 1$) the non-stationarity-robust tests, i.e. all except that of Robinson (2008a), are displayed. The dashed lines mark the nominal size level of 0.05 so that ideally all points would lie on the intersection between these two lines. The dotted line is the bisector implying that methods above the bisector do better with correlation and methods below the bisector do better without. Black symbols give results with a bandwidth parameter of $\delta_m = 0.75$ and gray symbols with $\delta_m = 0.55$.

It can be observed that the procedures by Marmol and Velasco (2004), Nielsen and Shimotsu (2007) and Hualde and Velasco (2008) lie below the bisector and are thus negatively affected by high correlation, whereas the tests by Chen and Hurvich (2006) and Robinson (2008a) lie above the bisector. The remaining tests lie on the bisector indicating robustness to correlation. Regarding bandwidth choice, the tests by Marmol and Velasco (2004), Chen and Hurvich (2006), Hualde and Velasco (2008) and Wang et al. (2015) are more liberal with a small bandwidth, whereas Nielsen and Shimotsu

---

[3] The test of Chen and Hurvich (2006) is conservative by construction as discussed in the previous section.
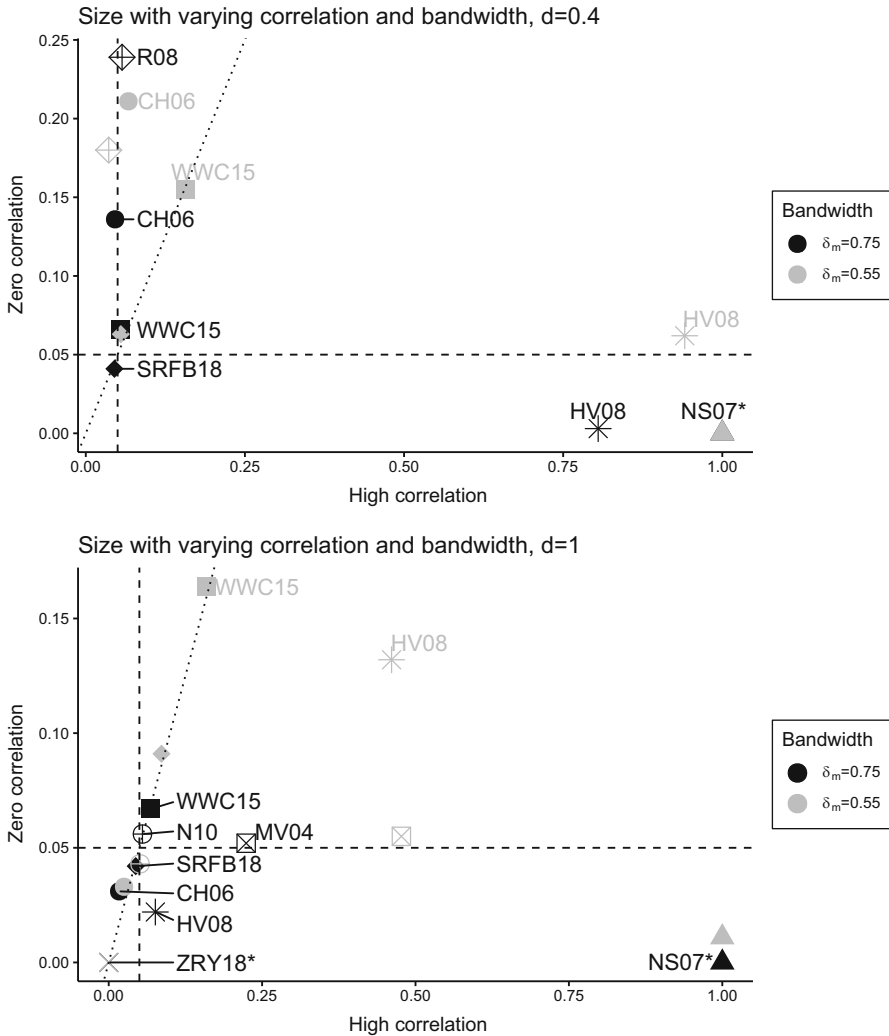
Fig. 1 Size (*rank estimation) based on DGP1 depending on correlation $\rho \in \{0, 0.99\}$ and bandwidth $\delta_m \in \{0.55, 0.75\}$ with $T = 1000$

(2007), Robinson (2008a), Nielsen (2010), Souza et al. (2018) and Zhang et al. (2019) are relatively robust in terms of size. In general, correlation in the underlying short-run component is mistaken for cointegration more often in stationary systems than in non-stationary ones.

Overall, in terms of size for bivariate systems and taking the range of admissible parameter values into account, we find that the test of Souza et al. (2018) has the best performance, followed by those of Chen and Hurvich (2006) and Wang et al. (2015). Considering procedures only applicable to non-stationary systems, Nielsen (2010) and Zhang et al. (2019) are very reliable options as well.

Table 2 Power (*rank estimation), $b = d$ and $\delta_m = 0.75$ for the triangular model (DGP2)

Method	ρ	0			0.45			0.9			0.99		
	T/d	0.4	0.7	1	0.4	0.7	1	0.4	0.7	1	0.4	0.7	1
	d_v	0	0	0	0	0	0	0	0	0	0	0	0
NS07*	100	0.989	0.998	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
CH06	100	0.556	0.349	0.337	0.317	0.291	0.110	0.046	0.205	0.020	0.007	0.179	0.010
	500	0.998	0.443	1.000	0.991	0.179	1.000	0.999	0.081	1.000	1.000	0.071	1.000
	1000	1.000	0.811	1.000	1.000	0.518	1.000	1.000	0.277	1.000	1.000	0.241	1.000
	2500	1.000	0.998	1.000	1.000	0.972	1.000	1.000	0.869	1.000	1.000	0.833	1.000
HV08	100	0.788	0.989	1.000	0.882	0.998	1.000	0.958	1.000	1.000	0.959	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SRFB18	100	0.720	0.976	0.999	0.709	0.982	0.999	0.642	0.928	0.993	0.259	0.586	0.914
	500	0.993	1.000	1.000	0.994	1.000	1.000	0.983	1.000	1.000	0.794	0.961	0.993
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.949	0.994	0.999
	2500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000

Table 2 continued

Method	T/d	$\rho=0$			$\rho=0.45$			$\rho=0.9$			$\rho=0.99$		
	d_v	0.4 (0)	0.7 (0)	1 (0)	0.4 (0)	0.7 (0)	1 (0)	0.4 (0)	0.7 (0)	1 (0)	0.4 (0)	0.7 (0)	1 (0)
R08	100	**0.044**	0.245	0.572	**0.009**	0.023	0.200	**0.049**	0.060	0.065	**0.671**	0.463	0.147
	500	**0.870**	1.000	1.000	**0.368**	0.869	0.998	**0.202**	0.239	0.454	**1.000**	0.957	0.599
	1000	**0.997**	1.000	1.000	**0.799**	0.999	1.000	**0.347**	0.343	0.573	**1.000**	0.991	0.709
	2500	**1.000**	1.000	1.000	**0.999**	1.000	1.000	**0.576**	0.486	0.707	**1.000**	1.000	0.804
WWC15	100	0.634	**0.896**	0.966	0.618	**0.887**	0.970	0.426	**0.855**	0.965	0.307	**0.832**	0.959
	500	0.829	**0.967**	0.993	0.807	**0.968**	0.994	0.694	**0.961**	0.993	0.630	**0.956**	0.995
	1000	0.867	**0.982**	0.998	0.850	**0.979**	0.998	0.763	**0.977**	0.997	0.695	**0.978**	0.997
	2500	0.917	**0.992**	0.998	0.896	**0.988**	0.999	0.817	**0.990**	0.999	0.787	**0.990**	0.999
ZRY19*	100	0.020	**0.403**	0.797	0.007	**0.339**	0.780	0.006	**0.289**	0.770	0.008	**0.289**	0.764
	500	0.082	**0.983**	1.000	0.032	**0.974**	1.000	0.016	**0.953**	1.000	0.010	**0.954**	1.000
	1000	0.119	**1.000**	1.000	0.042	**1.000**	1.000	0.011	**0.998**	1.000	0.011	**0.998**	1.000
	2500	0.200	**1.000**	1.000	0.049	**1.000**	1.000	0.013	**1.000**	1.000	0.007	**1.000**	1.000
N10	100		**0.431**	0.978		**0.356**	0.965		**0.270**	0.950		**0.262**	0.954
	500		**0.996**	1.000		**0.988**	1.000		**0.981**	1.000		**0.974**	1.000
	1000		**1.000**	1.000		**1.000**	1.000		**1.000**	1.000		**1.000**	1.000
	2500		**1.000**	1.000		**1.000**	1.000		**1.000**	1.000		**1.000**	1.000
MV04	100		0.481	0.974		0.748	0.994		0.854	0.998		0.866	0.999
	500		0.985	1.000		0.993	1.000		0.997	1.000		0.997	1.000
	1000		0.997	1.000		0.999	1.000		1.000	1.000		1.000	1.000
	2500		1.000	1.000		1.000	1.000		1.000	1.000		1.000	1.000

Table 3 Power (*rank estimation) with $b = d$ and $\delta_m = 0.75$ for the common-component model (DGP3)

Method	ρ	0			0.45			0.9			0.99		
	T/d	0.4	0.7	1	0.4	0.7	1	0.4	0.7	1	0.4	0.7	1
	d_v	0	0	0	0	0	0	0	0	0	0	0	0
NS07*	100	**0.624**	**0.866**	**0.968**	**0.993**	**0.999**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	500	**0.658**	**0.999**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	1000	**0.688**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	2500	**0.803**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
CH06	100	**0.646**	**0.912**	**0.696**	**0.599**	**0.896**	**0.628**	**0.530**	**0.872**	**0.547**	**0.532**	**0.873**	**0.524**
	500	**0.995**	**0.756**	**1.000**	**0.991**	**0.661**	**1.000**	**0.985**	**0.611**	**0.999**	**0.981**	**0.592**	**1.000**
	1000	**1.000**	**0.783**	**1.000**	**1.000**	**0.592**	**1.000**	**1.000**	**0.437**	**1.000**	**1.000**	**0.423**	**1.000**
	2500	**1.000**	**0.988**	**1.000**	**1.000**	**0.950**	**1.000**	**1.000**	**0.871**	**1.000**	**1.000**	**0.847**	**1.000**
HV08	100	**0.172**	**0.711**	**0.986**	**0.387**	**0.875**	**0.994**	**0.725**	**0.977**	**1.000**	**0.887**	**0.995**	**1.000**
	500	**0.883**	**1.000**	**1.000**	**0.986**	**1.000**	**1.000**	**0.999**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	1000	**0.994**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	2500	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
SRFB18	100	**0.366**	**0.775**	**0.979**	**0.417**	**0.837**	**0.986**	**0.467**	**0.868**	**0.990**	**0.474**	**0.874**	**0.990**
	500	**0.800**	**1.000**	**1.000**	**0.856**	**1.000**	**1.000**	**0.895**	**1.000**	**1.000**	**0.904**	**1.000**	**1.000**
	1000	**0.964**	**1.000**	**1.000**	**0.983**	**1.000**	**1.000**	**0.992**	**1.000**	**1.000**	**0.991**	**1.000**	**1.000**
	2500	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
R08	100	0.109	0.533	0.860	0.054	0.368	0.770	0.018	0.183	0.601	0.013	0.158	0.573
	500	0.950	1.000	1.000	0.904	1.000	1.000	0.837	1.000	1.000	0.816	1.000	1.000
	1000	0.999	1.000	1.000	0.999	1.000	1.000	0.996	1.000	1.000	0.996	1.000	1.000
	2500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3 continued

Method	ρ	0			0.45			0.9			0.99		
	T/d												
	d_v	0.4	0.7	1	0.4	0.7	1	0.4	0.7	1	0.4	0.7	1
		0	0	0	0	0	0	0	0	0	0	0	0
WWC15	100	0.391	0.769	0.933	0.477	0.798	0.939	0.500	0.835	0.939	0.509	0.833	0.942
	500	0.634	0.928	0.989	0.688	0.942	0.988	0.733	0.943	0.989	0.742	0.945	0.990
	1000	0.721	0.959	0.992	0.761	0.968	0.994	0.806	0.966	0.990	0.802	0.970	0.993
	2500	0.807	0.979	0.997	0.826	0.982	0.997	0.856	0.983	0.999	0.873	0.982	0.998
ZRY19*	100	0.015	0.363	0.785	0.010	0.347	0.770	0.005	0.308	0.754	0.006	0.300	0.756
	500	0.045	0.978	1.000	0.029	0.966	1.000	0.016	0.958	1.000	0.014	0.955	1.000
	1000	0.072	1.000	1.000	0.035	0.999	1.000	0.017	0.999	1.000	0.014	0.998	1.000
	2500	0.091	1.000	1.000	0.042	1.000	1.000	0.018	1.000	1.000	0.015	1.000	1.000
N10	100		0.151	0.879		0.175	0.887		0.202	0.881		0.201	0.874
	500		0.962	1.000		0.965	1.000		0.966	1.000		0.967	1.000
	1000		1.000	1.000		1.000	1.000		1.000	1.000		1.000	1.000
	2500		1.000	1.000		1.000	1.000		1.000	1.000		1.000	1.000
MV04	100		0.162	0.439		0.141	0.611		0.297	0.919		0.548	0.975
	500		0.217	0.999		0.553	1.000		0.911	1.000		0.953	1.000
	1000		0.491	1.000		0.872	1.000		0.992	1.000		0.993	1.000
	2500		0.952	1.000		0.996	1.000		1.000	1.000		1.000	1.000

To analyze the power of the procedures, we focus on the triangular representation in DGP2 with $b = d$ so that the memory reduces to zero in the cointegrating relation. Again, δ_m is set to 0.75. The results are shown in Table 2. In the following, we focus on the results for parameter constellations for which the tests have reasonable size properties.

It can be seen that the rank estimate of Nielsen and Shimotsu (2007) correctly identifies the presence of fractional cointegration even in relatively small samples. Since the estimate works well under the null hypothesis if ρ is low, it clearly outperforms its competitors in this situation. The power of the test of Hualde and Velasco (2008) is also high, but it suffers from similar size issues in case of strongly correlated short-run components.

Among the tests that are more widely applicable the approach of Souza et al. (2018) generates higher power than that of Wang et al. (2015) (except for $\rho = 0.99$), which in turn outperforms the approach of Chen and Hurvich (2006). Furthermore, it can be seen that the test of Souza et al. (2018) outperforms more restrictive approaches in small samples such as those of Robinson (2008a) and Nielsen (2010). In large sample this observation vanishes. For the test of Chen and Hurvich (2006) we can observe that the power is lower for $d = 0.7$ than for other values of d. Furthermore, the power becomes non-monotonic in T in some cases. This effect is likely to be caused by the fact that the order of differentiation required may be estimated incorrectly for intermediate values of d. The approach of Zhang et al. (2019) behaves similarly as that of Nielsen (2010).

With regard to the test of Robinson (2008a), it is noteworthy that the power is considerably lower for $\rho = 0.9$ than it is for $\rho = 0.45$ or $\rho = 0.99$. Further simulation results on this V-shaped dependence pattern between the power of the test and ρ (not reported here) show that the test has no power if $\rho = 0.8$ and its power is very low in a neighborhood of this point.

The test of Marmol and Velasco (2004) develops good power for stationary values of d_v, even though its theoretical properties are derived under the assumption that $d_v > 0.5$.

Overall, we find that the rank estimation of Nielsen and Shimotsu (2007) performs best in identifying the correct order of fractional cointegration if the correlation between the series is low. For non-stationary data Nielsen (2010) is a good choice, and among the more broadly applicable methods the test of Souza et al. (2018) clearly performs best in terms of size and power.

The previous table is generated based on the triangular model (DGP2), but we are also interested in the performance based on the common-components model (DGP3). Those results are displayed in Table 3. It can be seen that there is a number of striking differences in the relative performance of the tests. For low values of d, the rank estimation procedure of Nielsen and Shimotsu (2007)/Robinson and Yajima (2002) loses precision. At the same time, the test of Chen and Hurvich (2006) becomes more powerful so that overall the two procedures become comparable in terms of their ability to identify the correct rank. Unfortunately, the non-monotonicity of the test of Chen and Hurvich (2006) for intermediate values of d becomes even more apparent.
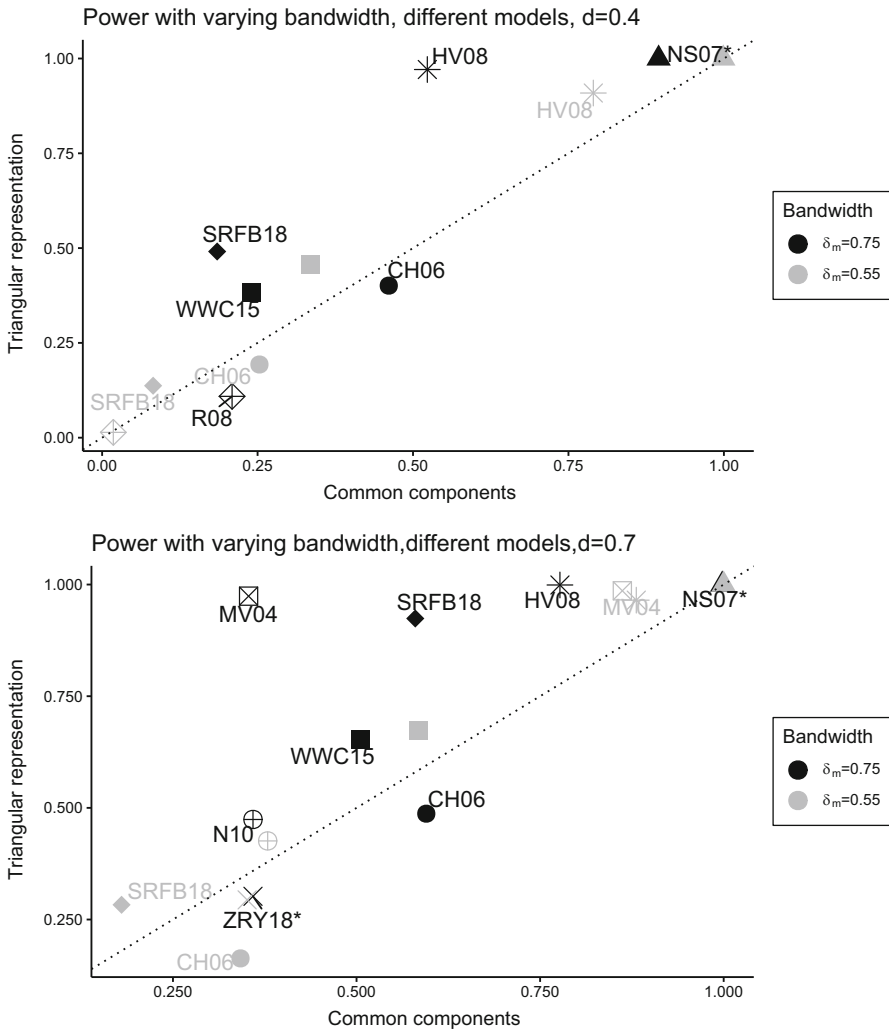
Fig. 2 Power (*rank estimation) depending on model specification (DGP2 or DGP3) and bandwidth $\delta_m \in \{0.55, 0.75\}$ with $T = 1000$, $\rho = 0.45$, and $b = d/3$

The test of Souza et al. (2018) still performs relatively well—especially for larger values of d. The same holds true for that of Wang et al. (2015) which reaches a relatively high power in smaller samples but approaches 1 only slowly.

With respect to the other tests, it can be seen that the test of Hualde and Velasco (2008) has very good power properties—also for low values of ρ where it maintains its size. The procedures of Nielsen (2010) and Zhang et al. (2019) have lower power for $d = 0.7$ in small samples but in all other constellations their power results are very good.

We further conduct a similar analysis to that for the size in Figure 2 in order to analyze the effect of the model construction and that of the bandwidth choice on the

power of the procedures. As before, black symbols represent results with $\delta_m = 0.75$ and gray symbols represent $\delta_m = 0.55$. The values of d and b are selected such that the power of the procedures tends to be low and changes in their behavior are easier to identify. First of all, most procedures perform better in the triangular model except for the tests by Chen and Hurvich (2006) and Robinson (2008a). Furthermore, while an increase of the bandwidth leads to a considerable power gain for the tests of Chen and Hurvich (2006), Robinson (2008a), and Souza et al. (2018), the approaches of Marmol and Velasco (2004), Hualde and Velasco (2008) and Nielsen and Shimotsu (2007) have higher power with a smaller bandwidth—at least in the common-components model. This, however, might be due to the larger size distortions visible in Figure 1. The performance of the approaches of Nielsen (2010) and Wang et al. (2015) is relatively independent of the bandwidth choice. For the test of Nielsen (2010) this is explained by the fact that the bandwidth only influences the estimate of d that determines the correct set of critical values. The test statistic itself does not depend on the bandwidth.

A strong advantage of semiparametric methods is the theoretical robustness to short-run dynamics. We therefore consider the previous experiments with u_{1t}, u_{2t}, e_t as AR(1) processes with $\phi \in \{-0.5, 0.5\}$. Table 4 contains both size and power results. Most tests and rank estimates exhibit performance differences compared to the white noise case, in particular with respect to size. For Souza et al. (2018), Wang et al. (2015) and Zhang et al. (2019) we observe sensitivity to the sign of the AR parameter as they tend to become liberal (or have a higher error rate for Zhang et al. (2019)) with $\phi = -0.5$ and become conservative (have a lower error rate) with $\phi = 0.5$. Additionally, the test by Souza et al. (2018) does not hold the nominal size if $d = 1$ and $\phi = 0.5$. Similar as the estimate of Zhang et al. (2019) does not yield 0 anymore, the one by Nielsen and Shimotsu (2007) has an error rate of 2-10% instead of 0%. The test of Chen and Hurvich (2006) becomes more conservative with stationary data and gets closer to the nominal significance level for non-stationary data, and the procedures by Hualde and Velasco (2008) and Marmol and Velasco (2004) both turn out very conservative. The test by Nielsen (2010) holds the nominal size only with $d = 1$ and a negative AR parameter and becomes very liberal in the other cases. In comparison to analogous results without short-run dynamics we find weaker performances in small samples for the tests by Robinson (2008a) and Marmol and Velasco (2004) in terms of size, and for the methods by Nielsen and Shimotsu (2007), Souza et al. (2018) and Marmol and Velasco (2004) in terms of power. In general, we observe a tendency of slightly lower power in some occasions like stationary data with the tests of Chen and Hurvich (2006) and Souza et al. (2018), but overall the power results differ not that much compared to the white noise case.

5 Conclusion

This review is written with the objective to provide guidance for the selection of methods in practical applications. We judge the methods based on (i) the range of values of d and b that are allowed, (ii) the ability to distinguish correctly between common trends and correlated innovations, and (iii) the performance across different DGPs—namely triangular systems as well as common-components models.

Table 4 Size (*rank estimation) based on DGP1 and power (*rank estimation) based on DGP2 with $\delta_m = 0.75$, $\rho = 0.45$, and $b = 0$/$b = d$ including short-run components

Method		size						power					
	φ	−0.5			0.5			−0.5			0.5		
	T/d	0.4	0.7	1	0.4	0.7	1	0.4	0.7	1	0.4	0.7	1
	d_v	0.4	0.7	1	0.4	0.7	1	0	0	0	0	0	0
NS07*	100	0.098	0.152	0.061	0.053	0.034	0.015	0.269	0.716	0.107	0.052	0.066	0.315
	500	0.080	0.053	0.026	0.048	0.033	0.041	0.990	0.175	1.000	0.268	0.996	1.000
	1000	0.070	0.033	0.021	0.033	0.029	0.036	1.000	0.247	1.000	0.406	1.000	1.000
	2500	0.053	0.024	0.018	0.034	0.020	0.033	1.000	0.890	1.000	0.757	1.000	1.000
CH06	100	0.011	0.020	0.070	0.011	0.020	0.021	0.085	0.311	0.938	0.518	0.969	0.999
	500	0.003	0.030	0.060	0.018	0.029	0.037	0.190	0.899	1.000	0.659	1.000	1.000
	1000	0.003	0.031	0.058	0.009	0.035	0.040	0.327	0.975	1.000	0.697	1.000	1.000
	2500	0.003	0.031	0.057	0.007	0.041	0.042	0.580	0.995	1.000	0.846	1.000	1.000
HV08	100	0.123	0.141	0.227	0.138	0.163	0.311	1.000	1.000	1.000	1.000	1.000	1.000
	500	0.000	0.000	0.049	0.000	0.000	0.082	1.000	1.000	1.000	1.000	1.000	1.000
	1000	0.000	0.000	0.022	0.000	0.000	0.036	1.000	1.000	1.000	1.000	1.000	1.000
	2500	0.000	0.000	0.009	0.000	0.000	0.016	1.000	1.000	1.000	1.000	1.000	1.000
SRFB18	100	0.093	0.110	0.101	0.013	0.023	0.166	0.022	0.039	0.226	0.002	0.015	0.103
	500	0.077	0.109	0.082	0.013	0.014	0.245	0.409	0.882	0.997	0.192	0.825	0.998
	1000	0.067	0.113	0.068	0.013	0.020	0.251	0.834	0.999	1.000	0.644	0.998	1.000
	2500	0.064	0.099	0.071	0.017	0.024	0.264	0.998	1.000	1.000	0.995	1.000	1.000

Table 4 continued

In the header blocks, the top level is *size* / *power*; within each, φ = −0.5 and φ = 0.5; the three columns per φ-block correspond to the parameter (T/d, d_v). For *size* the columns are (0.4, 0.7, 1) with d_v = (0.4, 0.7, 1); for *power* the columns are (0.4, 0.7, 1) with d_v = (0, 0, 0).

Method	T	size φ=−0.5 (0.4)	(0.7)	(1)	size φ=0.5 (0.4)	(0.7)	(1)	power φ=−0.5 (0.4)	(0.7)	(1)	power φ=0.5 (0.4)	(0.7)	(1)
R08	100	0.133	0.147	0.131	0.146	0.179	0.410	0.757	0.955	0.994	0.684	0.961	1.000
	500	0.053	0.067	0.068	0.088	0.093	0.351	0.991	1.000	1.000	0.993	1.000	1.000
	1000	0.051	0.048	0.051	0.069	0.065	0.332	1.000	1.000	1.000	1.000	1.000	1.000
	2500	0.029	0.038	0.048	0.053	0.049	0.294	1.000	1.000	1.000	1.000	1.000	1.000
WWC15	100	0.188	0.181	0.190	0.004	0.004	0.004	0.736	0.920	0.973	0.291	0.729	0.914
	500	0.121	0.113	0.120	0.001	0.001	0.001	0.846	0.973	0.996	0.619	0.930	0.988
	1000	0.109	0.098	0.100	0.001	0.001	0.002	0.873	0.985	0.997	0.746	0.963	0.996
	2500	0.077	0.081	0.087	0.001	0.002	0.002	0.904	0.990	0.999	0.829	0.985	0.998
ZRY19*	100	0.041	0.107	0.134	0.013	0.008	0.004	0.930	1.000	1.000	0.924	0.996	1.000
	500	0.039	0.082	0.096	0.032	0.010	0.001	1.000	1.000	1.000	1.000	1.000	1.000
	1000	0.056	0.078	0.088	0.038	0.009	0.002	1.000	1.000	1.000	1.000	1.000	1.000
	2500	0.063	0.073	0.077	0.046	0.008	0.002	1.000	1.000	1.000	1.000	1.000	1.000
N10	100	0.017	0.036		0.138	0.267		0.141	0.781		0.932	0.998	
	500	0.084	0.048		0.153	0.292		0.999	1.000		1.000	1.000	
	1000	0.110	0.046		0.134	0.282		1.000	1.000		1.000	1.000	
	2500	0.130	0.053		0.124	0.234		1.000	1.000		1.000	1.000	
MV04	100	0.446	0.667		0.576	0.634		0.136	0.700		0.526	0.834	
	500	0.156	0.000		0.008	0.008		0.796	1.000		0.997	1.000	
	1000	0.007	0.000		0.000	0.000		0.966	1.000		1.000	1.000	
	2500	0.000	0.000		0.000	0.000		1.000	1.000		1.000	1.000	

Based on our Monte Carlo studies, we find that some of the proposed approaches have weaknesses in their finite sample behavior in some empirically relevant scenarios—especially in presence of correlated short-run components. This concerns mostly the methods of Nielsen and Shimotsu (2007) (or Robinson and Yajima (2002)), Marmol and Velasco (2004), and Hualde and Velasco (2008) that have the highest power but have size issues in case of strongly correlated short-run components. With regard to iii.), we find that the size properties of the tests in the triangular case and the common-components model is generally comparable (see online material). For the power of the tests, however, there are important differences between the two cases. In particular, the test of Chen and Hurvich (2006) has much better power for stationary systems under the common components specification, whereas the methods of Robinson and Yajima (2002) and Hualde and Velasco (2008) become worse in their ability to detect fractional cointegration.

Although the methods of Robinson (2008a), Nielsen (2010), and Zhang et al. (2019) turn out to be robust to short-run correlation and are appealing due to their simplicity, they impose practically relevant restrictions on the permissible range of d and b. However, if there is prior knowledge about the (non-) stationarity of the data, those procedures are very good options.

Overall, we conclude that the test of Souza et al. (2018) for bivariate systems has the best properties, both theoretically and empirically, and is a good choice for the applied econometrician. It allows for the whole empirically relevant range of d and b, it is robust to correlation and short-run dynamics with positive coefficients, and it provides comparable performance in both—triangular systems and common-components models.

In higher dimensional systems, however, the test of Souza et al. (2018) is no longer applicable and that of Chen and Hurvich (2006) turns out to be liberal in finite samples from stationary processes. Here, the procedure of Robinson (2008a) can be recommended for stationary processes and the rank estimation by Nielsen (2010) and Zhang et al. (2019) should be preferred for non-stationary systems if the cointegrating residuals can be expected to be stationary.

References

Avarucci M, Velasco C (2009) A Wald test for the cointegration rank in nonstationary fractional systems. J Econ 151(2):178–189

Breitung J, Hassler U (2002) Inference on the cointegration rank in fractionally integrated processes. J Econ 110(2):167–185

Chen WW, Hurvich CM (2003) Semiparametric estimation of multivariate fractional cointegration. J Am Stat Assoc 98(463):629–642

Chen WW, Hurvich CM (2006) Semiparametric estimation of fractional cointegrating subspaces. Ann Stat 34(6):2939–2979

Chernov M (2007) On the role of risk premia in volatility forecasting. J Bus Econ Stat 25(4):411–426

Christensen BJ, Nielsen MØ (2006a) Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. J Econ 133(1):343–371

Christensen BJ, Nielsen MØ (2006b) Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. J Econ 133(1):343–371

Christensen BJ, Prabhala NR (1998) The relation between implied and realized volatility. J Financ Econ 50(2):125–150

Engle RF, Granger CW (1987) Co-integration and error correction: representation, estimation, and testing. Econometrica 55(2):251–276

Flôres RG Jr, Szafarz A (1996) An enlarged definition of cointegration. Econ Lett 50(2):193–195

Geweke J, Porter-Hudak S (1983) The estimation and application of long memory time series models. J Time Ser Anal 4(4):221–238

Hassler U, Breitung J (2006) A residual-based LM-type test against fractional cointegration. Econ Theory 22(6):1091–1111

Hausman JA (1978) Specification tests in econometrics. Econometrica 46(6):1251–1271

Höglund R, Östermark R (2003) Size and power of some cointegration tests under structural breaks and heteroskedastic noise. Stat Pap 44(1):1–22

Hualde J (2013) A simple test for the equality of integration orders. Econ Lett 119(3):233–237

Hualde J, Velasco C (2008) Distribution-free tests of fractional cointegration. Econ Theory 24(01):216–255

Hurvich CM, Chen WW (2000) An efficient taper for potentially overdifferenced long-memory time series. J Time Ser Anal 21(2):155–180

Johansen S (1988) Statistical analysis of cointegration vectors. J Econ Dyn Control 12(2–3):231–254

Johansen S (1991) Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. Econometrica 59(6):1551–1580

Johansen S (1995) Likelihood-based inference in cointegrated vector autoregressive models. Oxford University Press, Oxford

Johansen S (2008) Representation of cointegrated autoregressive processes with application to fractional processes. Econ Rev 28(1–3):121–145

Johansen S, Nielsen MØ (2012) Likelihood inference for a fractionally cointegrated vector autoregressive model. Econometrica 80(6):2667–2732

Johansen S, Nielsen MØ (2019) Nonstationary cointegration in the fractionally cointegrated VAR model. J Time Ser Anal 40(4):519–543

Künsch H (1987) Statistical aspects of self-similar processes. Proc First World Congr Bernoulli Soc 1:67–74

Łasak K (2010) Likelihood based testing for no fractional cointegration. J Econ 158(1):67–77

Łasak K, Velasco C (2015) Fractional cointegration rank estimation. J Bus Econ Stat 33(2):241–254

Lobato IN (1999) A semiparametric two-step estimator in a multivariate long memory model. J Econ 90(1):129–153

Marinucci D, Robinson PM (1999) Alternative forms of fractional brownian motion. J Stat Plan Inference 80(1–2):111–122

Marinucci D, Robinson PM (2001) Semiparametric fractional cointegration analysis. J Econ 105(1):225–247

Marmol F, Velasco C (2004) Consistent testing of cointegrating relationships. Econometrica 72(6):1809–1844

Nielsen MØ (2004) Spectral analysis of fractionally cointegrated systems. Econ Lett 83(2):225–231

Nielsen MØ (2005) Semiparametric estimation in time-series regression with long-range dependence. J Time Ser Anal 26(2):279–304

Nielsen MØ (2007) Local Whittle analysis of stationary fractional cointegration and the implied-realized volatility relation. J Bus Econ Stat 25(4):427–446

Nielsen MØ (2009) A powerful test of the autoregressive unit root hypothesis based on a tuning parameter free statistic. Econ Theory 25(6):1515–1544

Nielsen MØ (2010) Nonparametric cointegration analysis of fractional systems with unknown integration orders. J Econ 155(2):170–187

Nielsen MØ, Frederiksen P (2011) Fully modified narrow-band least squares estimation of weak fractional cointegration. Econ J 14(1):77–120

Nielsen MØ, Shimotsu K (2007) Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach. J Econ 141(2):574–596

Phillips PC, Ouliaris S (1988) Testing for cointegration using principal components methods. J Econ Dyn Control 12(2–3):205–230

Reimers HE (1992) Comparisons of tests for multivariate cointegration. Stat Pap 33(1):335–359

Robinson PM (1994) Semiparametric analysis of long-memory time series. Ann Stat 22(1):515–539

Robinson PM (1995a) Gaussian semiparametric estimation of long range dependence. Ann Stat 23(5):1630–1661

Robinson PM (1995b) Log-periodogram regression of time series with long range dependence. Ann Stat 23(3):1048–1072

Robinson P (2008a) Diagnostic testing for cointegration. J Econ 143(1):206–225

Robinson PM (2008b) Multiple local Whittle estimation in stationary systems. Ann Stat 36(5):2508–2530

Robinson PM, Marinucci D (2001) Narrow-band analysis of nonstationary processes. Ann Stat 29(4):947–986

Robinson PM, Yajima Y (2002) Determination of cointegrating rank in fractional systems. J Econ 106(2):217–241

Shimotsu K (2007) Gaussian semiparametric estimation of multivariate fractionally integrated processes. J Econ 137(2):277–310

Shimotsu K (2010) Exact local Whittle estimation of fractional integration with unknown mean and time trend. Econ Theory 26(2):501–540

Shimotsu K (2012) Exact local Whittle estimation of fractionally cointegrated systems. J Econ 169(2):266–278

Shimotsu K, Phillips PC (2005) Exact local Whittle estimation of fractional integration. Ann Stat 33(4):1890–1933

Souza IVM, Reisen VA, Franco GdC, Bondon P (2018) The estimation and testing of the cointegration order based on the frequency domain. J Bus Econ Stat 36(4):695–704

Velasco C (2003a) Gaussian semi-parametric estimation of fractional cointegration. J Time Ser Anal 24(3):345–378

Velasco C (2003b) Nonparametric frequency domain analysis of nonstationary multivariate time series. J Stat Plan Inference 116(1):209–247

Wang M, Chan NH (2016) Testing for the equality of integration orders of multiple series. Econometrics 4(4):49

Wang B, Wang M, Chan NH (2015) Residual-based test for fractional cointegration. Econ Lett 126:43–46

Zhang R, Robinson P, Yao Q (2019) Identifying cointegration by eigenanalysis. J Am Stat Assoc 114(526):916–927