Testing for No Cointegration in Vector Autoregressions with Estimated Degree of Fractional Integration*

Matei Demetrescu† University of Kiel

Vladimir Kuzin Deutsche Bundesbank

Nazarii Salish Bonn Graduate School of Economics

This version: December 9, 2014

Abstract

The persistence of fractionally integrated VAR processes is characterized by the common integration parameter d and, under cointegration, by the persistence of the deviations from equilibrium. Finding long-run relations between fractionally integrated series is thus more difficult compared to classical (integer) cointegration since the two persistence parameters are typically unknown in practice. Moreover, even when d is known, the parameter characterizing the persistence of the departures from potential equilibrium strength of the potential return to equilibrium is not identified under no cointegration. To deal with the lack of identification under the null when testing, we resort to a linearization of the fractional difference filter, which leads to regression-based tests of no error-correction (if choosing a single-equation approach) and of no cointegration (if choosing a system approach). Short-run dynamics is accounted for via lag augmentation. When d is known, all test statistics possess standard (normal or χ^2) asymptotic limiting distributions under the respective null hypothesis and under local alternatives. For the case of not knowing d, we discuss conditions under which plugging in some consistent estimator does not affect the asymptotics. Estimation error of d induces size distortions even when no short-run dynamics is present; the key insight is that letting the number of lags increase as the sample size goes to infinity will account for the estimation error of d. Monte Carlo experiments illustrate the usefulness of the new tests and compare the single-equation no fractional cointegration tests with system fractional cointegration tests for both known and estimated fractional integration parameter in finite samples.

Key words: VAR model; Persistence; Strong and weak cointegration; Weak exogeneity; Simultaneity; Nonparametric estimator; Long autoregression

JEL classification: C12 (Hypothesis Testing), C22 and C32 (Time-Series Models)

---

*A previous version of this paper was circulated as "Single-Equation Error-Correction Testing and Endogeneity: The case of fractional cointegration". The authors would like to thank Jörg Breitung, Uwe Hassler, Søren Johansen, Michael Massmann and Morten Nielsen for very helpful comments and suggestions. The first and the third author gratefully acknowledge the support of the German Research Foundation (DFG) trough the grant DE 1617/2-1. The views expressed herein are those of the authors and do not necessarily reflect the views of institutions the authors belong to.

†**Corresponding Author:** Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40-60, D-24118 Kiel, Germany. E-mail address: `mdeme@stat-econ.uni-kiel.de`.

1 Motivation

The chances of detecting long-run relations between persistent series depend decisively on the strength of the reversion of the system to long-run equilibrium. While classical cointegration is characterized by integrated series exhibiting short-memory deviations from equilibrium, CI(1,1), the equilibrium deviations may be highly persistent in a number of applications. Nonstationary, yet mean-reverting equilibrium deviations (so-called weak cointegration) could explain for instance the occasional discrepancy between the expectation hypothesis of the terms structure of interest rates predicting $N-1$ cointegrating relations between N interest rates at different maturity, and applied work finding weak evidence for long-run relations between interest rates of higher maturity: the strength of reversion to equilibrium could well decrease with increasing maturity of the interest rates involved. Moreover, whether interest rates are precisely integrated of order one is questionable, and falsely imposing an integer order of integration has negative consequences. Thus, fractionally cointegrated models, CI(d,b), may provide a better framework for testing the null of no cointegration. They come at high cost, however: the integration order $d-b$ of the equilibrium deviations is not identified under the null of no cointegration, and estimation of the degree of fractional integration d of the series may also be problematic.

Along these lines, the estimation and especially testing of fractionally (co)integrated vector autoregressions have received increased attention in the literature. We shall focus on the time domain as it may be perceived as being more practitioner-friendly. Indeed, the typical time-domain procedure usually involves filters and regression techniques; see e.g. Tanaka (1999) or Breitung and Hassler (2002). Here, Avarucci and Velasco and Łasak (2010) propose for instance system no cointegration tests that set to deal with unknown b. But both assume known d (just like Breitung and Hassler, 2002 or Hassler and Breitung, 2006), and both impose restrictions on b: Avarucci and Velasco either use prior knowledge about $b \in [0, 0.5)$ or a consistent estimate thereof, whereas Łasak (2010) uses the method of Davies (1977) involving the maximum over a sequence of tests computed for a grid of values of $b \in [0.5+\epsilon, 1]$. The work of Johansen and Nielsen (2012a) does not impose limitations on d or b, but requires the analyzed series to be fractional white noise under the null of no cointegration.

We therefore discuss time-domain based tests for no fractional cointegration with particular attention paid to dealing with the two unknown parameters d and b. The test procedures we propose rely on a linearization of the fractional integration filter, and their analysis is carried out in two steps. We first work under the assumption of known d, and show in the second step that plugging in an estimate $\widehat{d}$ does not have any adverse effects asymptotically provided that additional mild conditions are fulfilled.

Section 2 thus starts with known fractional integration parameter d. To deal with the lack of identification of b under the null of no cointegration, we resort to Taylor approximations as used e.g. by Luukkonen et al. (1988). This leads to a regression-based system test belonging to the family of tests proposed by Breitung and Hassler (2002). As a complement, we provide tests for no error correction in a given equation of the system. In the integer case, the behavior of single-equation tests is decisively influenced by (lack of) weak exogeneity, as discussed by Engle et al. (1983); see e.g. Urbain (1993) for a more detailed examination. Moreover, the endogeneity issue affects the alternative hypothesis of cointegration as well: Johansen (1992) shows that, under (integer) cointegration, weak exogeneity implies only the studied equation to exhibit error-correction, which leads to possible inconsistency of the single-equation test if error correction is only present in the other equations of the system.

We study in Section 3 the asymptotic behavior of the proposed test statistics. Subsection 3.1 builds on a known fractional integration parameter d. For the system test, we show the limiting distribution of the test statistic to be chi-squared, or noncentral chi-squared under local alternatives. For the error-correction tests, the limiting null distribution is normal and, under the null, does not depend on either the number of regressors or, more importantly, on their exogeneity. We also discuss a variant of the single-equation test based on the conditional version of the single-equation model and show that it can have nontrivial power even if error-correction is not present in the test equation.

Subsection 3.2 then deals with the thorny issue of plugging in an estimate of d. We find the asymptotics for known d to remain valid even when plugging in an estimate $\widehat{d}$ whenever two additional conditions are met. They are (a) a long autoregressive augmentation with suitable rates for the lag order and (b) a consistent estimator $\widehat{d}$ with suitable convergence rates. The results of the second step apply in the univariate case as well, such that one can e.g. check whether the prerequisite of a common d for the series of the system plausibly holds before testing for no cointegration or no error-correction.

Finally, we compare our single-equation tests to system fractional cointegration tests in Section 4. We find in Monte Carlo experiments that fractional single-equation tests can be much more powerful than their system counterparts when error correction is present in the test equation, and that they have better size properties, also when the degree of fractional integration is estimated.

Before proceeding to the main part of the paper, let us set some notation. Boldface symbols denote real vectors; e.g. the multivariate process of interest, $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}} = \boldsymbol{y}_t$. The lag operator is denoted by L, $L\{\boldsymbol{y}_t\} = \{\boldsymbol{y}_{t-1}\}$. The fractional difference operator is given by the usual series expansion, $\Delta^d = (1-L)^d = 1 - dL - \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 - \ldots$, and its truncated version by $\Delta_+^d \equiv \Delta^d \mathbb{I}(t > 0)$ with $\mathbb{I}$ the indicator function. Finally, the symbols "$\xrightarrow{d}$" and "$\xrightarrow{p}$" stand for convergence in distribution and convergence in probability.

2 Model and test procedures

2.1 Basic idea

The starting point is the data generating process [DGP] proposed by Granger (1986), i.e. the fractional error-correction model given by the following

Assumption 1 *Let* $\boldsymbol{y}_t = \Delta_+^{-d}\boldsymbol{x}_t$ *and*

$$\boldsymbol{x}_t = \boldsymbol{\alpha}\left(\Delta_+^{-b} - 1\right)\boldsymbol{\beta}'\boldsymbol{x}_t + \boldsymbol{u}_t, \tag{1}$$

where $\boldsymbol{u}_t$ *is a* K*-variate* $I(0)$ *process subject to further conditions and* $\boldsymbol{\alpha}, \boldsymbol{\beta}$ *are* $K \times r$ *matrices,* $0 \leq r < K$*, such that* $\boldsymbol{\alpha}'_\perp\boldsymbol{\beta}_\perp$ *is nonsingular when* $b \neq 0$*, with* $\perp$ *denoting the orthogonal complement w.r.t.* $\mathbb{R}^K$*.*

Considering zero initial values ensures $\boldsymbol{y}_t$ to be well defined in mean square sense for $d \geq 0.5$; we shall equivalently work with truncated difference operators. The assumption about the initial condition is not innocuous, but conditioning on observations at the beginning of the observed series helps deal with the problem; see Johansen and Nielsen (2012b). To illustrate the derivation of the new tests we start with $\boldsymbol{u}_t$ being stationary white noise with covariance matrix Σ. This will be relaxed in the following subsection, where $\boldsymbol{u}_t$ is allowed to exhibit serial correlation.

It is the parameter b that controls the persistence of equilibrium deviations. For $b = 0$ or $r = 0$, Assumption 1 gives a purely fractionally integrated process: the components of $\boldsymbol{y}_t = (y_{t1}, \ldots, y_{tK})'$ are fractionally integrated of order d, $\boldsymbol{y}_t \sim I(d)$ and any nontrivial linear combination is also $I(d)$. For $b > 0$ and $0 < r < K$, the system is fractionally cointegrated, $\boldsymbol{y}_t \sim CI(d, b)$. The matrix $\boldsymbol{\alpha}$ models the adjustment to equilibrium, while $\boldsymbol{\beta}$ represents the cointegrating relations.[1] We shall also discuss tests for no error correction in a given equation of the system; note that, in a single-equation framework, the cointegration rank r is restricted to be either 0 or 1,[2] and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)'$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)'$ are vectors; the required identifying restriction is typically $\beta_1 = 1$.

Let us assume for now that d is known. We discuss in Section 3.2 conditions under which d may be replaced by a consistent estimator, so building on a known fractional integration parameter is not restrictive. Equation (1) then delivers as DGP

$$\boldsymbol{x}_t = \Pi\left(\Delta_+^{-b} - 1\right)\boldsymbol{x}_t + \boldsymbol{u}_t,$$

where $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$ is zero under the null of no cointegration, and has rank r under the alternative. Under cointegration, for which b must be positive, $\left(\Delta_+^{-b} - 1\right)\boldsymbol{x}_t$ is simply a linear combination of $\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \ldots \boldsymbol{x}_1$, since

$$\Delta^{-b} - 1 = bL + \frac{b\,(b+1)\,L^2}{2!} + \frac{b\,(b+1)\,(b+2)\,L^3}{3!} + \ldots\ .$$

In other words, Equation (1) describes for any $b > 0$ a dynamic regression model. So, if b were indeed known, one could easily test either for no cointegration in the system (null hypothesis

[1]One may need to restrict the matrices in order to guarantee that the process doesn't exhibit explosive behavior when error correction is present in the system.

[2]Single-equation tests, however, are known to have power against alternatives with cointegration ranks higher that 1.

$\Pi = \mathbf{0}$) or for no error correction in the kth equation of the system (null hypothesis $\alpha_k = 0$). But the parameter b is not known in advance; moreover, b is not identified under the null of no cointegration. So we can neither regress $\boldsymbol{x}_t$ (or x_{tk} for the single-equation approach) on $\left(\Delta^{-b} - 1\right) \boldsymbol{x}_t$ to test the null of no cointegration (or no error correction), nor can we estimate b *in the same step*.

A quick fix would be to take for b some fixed value, in the manner Dolado et al. (2002) build the so-called fractional Dickey-Fuller test, or to plug in an estimate of b obtained otherwise; Avarucci and Velasco (2009) actually generalize this way the Wald test for fractional integration of Lobato and Velasco (2007) to a system fractional cointegration test based on a rescaled reduced-rank regression involving $\frac{\Delta^{-b}-1}{b}$ rather than the filter $\Delta^{-b} - 1$ imposed by the assumed autoregressive model. Another, perhaps more appealing solution has been proposed by Davies (1977, 1987). It consists of computing the relevant test statistic for all values of b within a pre-specified range and taking the largest of the resulting values as statistic; see Łasak (2010), who generalizes this way the likelihood ratio integer cointegration test of Johansen (1995). Davies' test principle is however computationally demanding, and the resulting test statistic has a nonstandard distribution; Łasak also requires $b \in [0.5 + \varepsilon, 1]$ for some $0 < \varepsilon < 0.5$. At the same time, Avarucci and Velasco (2009) impose the restriction $0 \leq b < 0.5$. For the model without short-run dynamics, Johansen and Nielsen (2012a) provide the only procedure without restrictions on either d or b.

We pursue a different strategy here, and deal with the lack of identification of b under the null using the method employed by Luukkonen et al. (1988) for testing linearity in a smooth transition autoregressive context: the parameters relevant for the nonlinear alternative are not identified under the null of linearity either. The method implies the use of a Taylor approximation for $\Delta^{-b} - 1$ around $b = 0$. This suggests that our procedures may work best under the alternative when b is small; we examine the tests' behavior for $b \in [0, 1]$ in finite samples in Section 4 and find finite-sample evidence in favor of this conjecture.

Formally, we have that

$$\frac{\mathrm{d}^i \Delta^{-b}}{\mathrm{d}b^i} = (-1)^i \Delta^{-b} \ln^i (1 - L).$$

Evaluating the derivative at $b = 0$ and using the series expansion of the logarithm (cf. Tanaka, 1999, p. 560), we obtain the first-order approximation[3]

$$\Delta^{-b} - 1 \approx b \left(L + \frac{L^2}{2} + \frac{L^3}{3} + \ldots \right). \tag{2}$$

So let $\boldsymbol{\alpha}^* = b\boldsymbol{\alpha}$ and $\Pi^* = b\Pi = \boldsymbol{\alpha}^* \boldsymbol{\beta}'$; also, with zero starting values, let

$$x^*_{t-1,k} = \left(L + \frac{L^2}{2} + \frac{L^3}{3} + \ldots + \frac{L^{t-1}}{t-1} \right) x_{tk} = \sum_{j=1}^{t-1} \frac{x_{t-j,k}}{j}$$

[3] Higher-order approximations would involve powers of b, which would take us to a restricted nonlinear LS framework, so we do not pursue the topic here.

for $k = 1, 2, \ldots, K$. We are now able to test the null hypothesis of no cointegration with the help of the multivariate regression

$$\boldsymbol{x}_t = \widehat{\Pi}^* \boldsymbol{x}^*_{t-1} + \widehat{u}_{t1}. \tag{3}$$

since the null of no (fractional) cointegration of $\boldsymbol{y}_t$ translates into the null $\Pi^* = 0$, or $\operatorname{rank} \Pi^* = 0$. With stationary regressors, the LR or the Wald approach are asymptotically equivalent. Note that the Wald approach is more easily robustified against heteroskedasticity; see Bura and Cook (2003) and Remark 1 below. Since the focus here is not on (conditional) heteroskedasticity, we build for convenience on the LR approach. Under Gaussian errors, it results in canonical correlations analysis, and the test for the null $\Pi = \mathbf{0}$ in (3) is then essentially the one proposed by Breitung and Hassler (2002).[4] Short-run dynamics make an important difference, though; see Section 2.2.

If interested in a test for no error correction in a specific equation, w.l.o.g. in the first, one may resort to the test regression

$$x_{t1} = \widehat{\alpha}^*_1 x^*_{t-1,1} + \widehat{\boldsymbol{\delta}}' \boldsymbol{z}^*_{t-1} + \widehat{u}_{t1}, \tag{4}$$

where $\boldsymbol{\delta} = b\,(\alpha_1\beta_2, \ldots \alpha_1\beta_K)'$ and $\boldsymbol{z}^*_{t-1} = \left(x^*_{t-1,2}, \ldots, x^*_{t-1,K}\right)'$. This is of course because no error correction in the first equation ($\alpha_1 = 0$) is equivalent to the null $\alpha^*_1 = 0$. Just like the system case, short-run dynamics will be dealt with in Section 2.2.

For model (4), Kremers et al. (1992) propose in the integer case a single-equation cointegration test with known (pre-specified) cointegration vector and derive its asymptotic properties under the assumption of strong exogeneity; Zivot (2000) relaxes the assumption to weak exogeneity and shows how to set up the test such that dependence on nuisance parameters is eliminated, and Banerjee et al. (1998) suggest a test requiring neither pre-specification nor pre-estimation of the cointegration vector. (Weak exogeneity, however, is still required, unless one includes *leads* of the differenced regressors in the test regression.) In fact, the single-equation test of no (integer) error-correction Banerjee et al. (1998) is formally recovered for $b = 1$ and $d = 1$.

For the single-equation framework, exogeneity—or rather lack thereof—plays an important role. There are two equivalent ways of relaxing exogeneity in such cointegrated systems. First, one can allow for non-diagonal covariance matrix Σ of the innovations, leading to the interpretation of (1) as a reduced-form model. Second, one can restrict Σ to have diagonal form and consider the simultaneous-equation model

$$\Gamma \boldsymbol{x}_t = \boldsymbol{\alpha} \left(\Delta^{-b} - 1\right) \boldsymbol{\beta}' \boldsymbol{x}_t + \boldsymbol{u}_t,$$

[4] Nielsen (2005) points out that Breitung and Hassler's fractional cointegration test, although originating in the LM framework, is not a LM test and derives the LM test for fractional cointegration.

where the matrix Γ has ones on the main diagonal. This is equivalent to including contemporaneous dependent variables x_{tk} in each equation,

$$\boldsymbol{x}_t = \boldsymbol{\alpha} \left(\Delta^{-b} - 1\right) \boldsymbol{\beta}' \boldsymbol{x}_t + (I - \Gamma) \boldsymbol{x}_t + \boldsymbol{u}_t. \tag{5}$$

Of course, one could do both, but an identifiability problem arises. While we derive two versions of the error-correction test statistic, one based on the unconditional model (1) and one based on the conditional representation (5), we stick to representation (1) as DGP.

No error-correction tests should be run against the alternative $\alpha_1^* \neq 0$. This is because α_1 (and thus α_1^*) is only restricted to be negative when there is no error-correction in the other $K-1$ equations of the system. For the actual test, one simply estimates Equation (4) by OLS and uses the t statistic of α_1^*, $t_{\alpha_1^*} = \widehat{\alpha}_1^* / \widehat{\sigma}^2_{\widehat{\alpha}_1^*}$, for inference, where $\widehat{\sigma}^2_{\widehat{\alpha}_1^*}$ is the usual regression estimate of the variance of $\widehat{\alpha}_1^*$.

If however $\alpha_1 = 0$, the test is not always consistent against cointegration; its power actually equals the size even if the (local) alternative is present in other equations. A variant of the suggested test for no error correction is obtained by testing in the so-called conditional model (5), i.e. by including contemporaneous differences in the single-equation test regression.

$$x_{t1} = \underline{\widehat{\alpha}}_1^* x_{t-1,1}^* + \underline{\widehat{\boldsymbol{\delta}}}' \boldsymbol{z}_{t-1}^* + \underline{\widehat{\boldsymbol{a}}}_0' \boldsymbol{z}_t + \underline{\widehat{u}}_{t1}, \tag{6}$$

with obvious notation $\boldsymbol{z}_t = (x_{t,2}, \ldots, x_{t,K})'$.

Comparing Equation (4) and (6) to the error-correction test of Banerjee et al. (1998), the latter's integrated regressors $\left(\Delta_+^{-1} - 1\right) \boldsymbol{x}_t$ are simply replaced by $\boldsymbol{x}_{t-1}^*$, which is analog to the relation between the Dickey-Fuller test and the regression-based LM test for fractional integration of Breitung and Hassler (2002). In contrast to the test due to Banerjee et al. (1998), our test exhibits standard asymptotics, as shall be seen in Subsection 3.

Clearly, the single-equation tests are restricted in that error-correction should be present in the test equation in order to detect cointegration. One could, of course, check several of the K equations of the system to see if the test rejects in any of these. But at the price of such an effort[5] one might just as well use a system test from the beginning. So, unless error correction is postulated in a given equation, the single-equation approach may rather help pinpoint which equations exhibit error-correction after having established the existence of cointegration via a system test.

Before proceeding to the discussion of the short-run dynamics, note that an alternative derivation of the test idea is obtained when focusing on *local* alternatives of the form

$$b = b_T = \frac{c}{\sqrt{T}} \tag{7}$$

[5]This approach is likely to lead to over-rejections, unless one adjusts the significance level for each single-equation test. Such corrections are available in the statistical literature, see e.g. Benjamini and Hochberg (1995), but are typically undersized and thus lose some power.

with $c \geq 0$: it holds according to Tanaka (1999, p. 579) that

$$\Delta^{-\frac{c}{\sqrt{T}}} - 1 = \frac{c}{\sqrt{T}} \left(L + \frac{L^2}{2} + \frac{L^3}{3} + \ldots \right) + O\left(T^{-1}\right).$$

We thus obtain the same $b \sum_{j \geq 1} j^{-1} L$ as a linear approximation for $\Delta^{-b} - 1$.

We shall take the opportunity and discuss the behavior of our test statistics under the sequence of local alternatives given by (7). Note that, in the case of tests of no integer cointegration (Johansen, 1991, e.g.), it is common practice to model both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (or rather their product) as local to zero. This would lead here to $\boldsymbol{\alpha\beta}' = AT^{-0.5}$ for some fixed $K \times K$ matrix A. But such an approach is only feasible here when b is known, so we focus on the sequence of local alternatives in Equation (7).[6]

2.2 Modelling short-run dynamics

In order to allow for short-run serial correlation in addition to fractional (co)integration of the series $\boldsymbol{y}_t$, Granger (1986) proposes the following infinite-order vector autoregressive model:

$$A(L)\,\Delta^d \boldsymbol{y}_t = \left(1 - \Delta^b\right) \boldsymbol{\alpha\beta}' \Delta^{d-b} \boldsymbol{y}_t + \boldsymbol{u}_t,$$

where $\boldsymbol{y}_t = \mathbf{0}$ for $t \leq 0$. The lag polynomial $A(L) = \sum_{j=1}^{p} A_j L^j$, with A_j $K \times K$ matrices, should have characteristic roots belonging to the stability region.

Johansen (2008) advocates however a slightly different representation,

$$A(L_b)\,\Delta^d \boldsymbol{y}_t = \left(1 - \Delta^b\right) \boldsymbol{\alpha\beta}' \Delta^{d-b} \boldsymbol{y}_t + \boldsymbol{u}_t,$$

where the polynomial $A(\cdot)$ is given in terms of the pseudo-lag operator $L_b = 1 - \Delta^b$. This representation allows for the analysis of the dynamic properties of the process $\boldsymbol{y}_t$, which is not analytically tractable with Granger's model; see Johansen (2008) and Franchi (2010). We shall argue however that Johansen's (2008) representation is not suitable for our purposes; see below.

Finally, Avarucci and Velasco (2009) allow the disturbances in the error-correction model (1) to be themselves serially correlated, leading to

$$A(L)\,\Delta^d \boldsymbol{y}_t = \left(1 - \Delta^b\right) \boldsymbol{\alpha\beta}' \Delta^{d-b} A(L)\,\boldsymbol{y}_t + \boldsymbol{u}_t.$$

For all three cases, $A(L)$ is an invertible lag polynomial and $\boldsymbol{u}_t$ is stationary white noise as specified by the following assumptions.

Assumption 2 *The pth order polynomial $A(L)$ has all roots outside the unit circle.*

Assumption 3 *The innovations $\boldsymbol{u}_t = (u_{t1}, \ldots, u_{tK})'$ are iid $(0, \Sigma)$ with $\Sigma = \{\sigma_{ij}\}_{1 \leq i,j \leq K}$ positive definite.*

[6] The local alternatives of Equation (7) are a particular case of the more general condition that $\boldsymbol{\alpha}' \left(\Delta^{-b} - 1\right) \boldsymbol{\beta}$ is formally of order $O(1/\sqrt{T})$.

The *iid* assumption can easily be relaxed, e.g. by allowing $\boldsymbol{u}_t$ to be a more general martingale difference sequence, and independence only serves to simplify the proofs; see also Remark 1 further below.

All three models have advantages and disadvantages. As mentioned, Granger (1986) is not tractable for establishing e.g. the stationarity of the process $\boldsymbol{x}_t$. But the latter two do not provide a free lunch either.

In Johansen's (2008) model, the approximation from Equation (2) would lead to $L_b \approx b\sum_{j\geq 1} j^{-1}L$. And if $A(\cdot)$ were for instance a first-order lag polynomial, the test would not be able to distinguish between fractional cointegration and lag structure, since, for $b = c/\sqrt{T}$, *both* the error-correction and the short-run components imply the presence of $\boldsymbol{x}^*_{t-1}$ on the right-hand side of test Equations (3), (4) or (6). In fact, the series y_t are fractional white noise for $b = 0$ and no short-run dynamics is allowed for under the null, which is the price to pay for identification of b. Finally, Avarucci and Velasco (2009) imply a nonlinear model.[7]

We shall therefore run with the hare and hunt with the hounds. Following the arguments of Demetrescu et al. (2008), we simply augment the relevant test regression with lags, as if Granger's model were the right one (note that, under the null of no cointegration, the true model is trivially Granger's model). At the same time, we assume the DGP to be that of Avarucci and Velasco (2009), such that we know what properties $\boldsymbol{x}_t$ has under both the null and the alternative. The main advantage of this misspecification under the alternative is that it delivers an easy-to-implement test statistic (it is entirely regression-based and conducted in one step). Moreover, under the null hypothesis, the test regression is correctly specified, while, under *local* alternatives, the test regression will be approximately correct; see e.g. the proof of Proposition 1 for details. Serendipitously, the misspecification turns out to have an advantage in the single-equation case, where it may cause nontrivial power even if no error correction is present in the test equation; see the proof of Proposition 1 further below.

For the system test, Breitung and Hassler (2002) use prewhitening of $\boldsymbol{x}_t$. In our formulation, a slightly different procedure arises, which e.g. extends the augmented LM test for fractional integration of Demetrescu et al. (2008) to a fractional cointegration test in the same manner as the Johansen (1995) (integer) cointegration rank test generalizes the Dickey-Fuller test. While both the Breitung and Hassler (2002) and our test are $\chi^2((K-r)^2)$-distributed asymptotically, ours is constructed as follows:

1. Build differences,
$$\boldsymbol{x}_t = \Delta^d_+ \boldsymbol{y}_t;$$

2. filter the differences *before* prewhitening,
$$\boldsymbol{x}^*_{t-1} = \sum_{j=1}^{t-1} j^{-1}\boldsymbol{x}_{t-j};$$

[7] In fact, Avarucci and Velasco (2009) use a prewhitening procedure to avoid nonlinear estimation; cf. the way Hassler and Breitung (2006) deal with short-run correlation.

3. prewhiten (OLS) *both* $\boldsymbol{x}_t$ and $\boldsymbol{x}^*_{t-1}$,

$$\begin{aligned}\widehat{\boldsymbol{r}}_t &= \boldsymbol{x}_t - \widehat{\Psi}\boldsymbol{X}_{t-1}\\ \widehat{\boldsymbol{r}}^*_{t-1} &= \boldsymbol{x}^*_{t-1} - \widehat{\Psi}^*\boldsymbol{X}_{t-1};\end{aligned}$$

4. compute the tr_* test based on the canonical correlations of $\widehat{\boldsymbol{r}}_t$ and $\widehat{\boldsymbol{r}}^*_{t-1}$,

$$\begin{aligned}\widetilde{S}_{00} &= \frac{1}{T}\sum_{t=p+1}^{T}\widehat{\boldsymbol{r}}_t\widehat{\boldsymbol{r}}_t'\\ \widetilde{S}_{11} &= \frac{1}{T}\sum_{t=p+1}^{T}\widehat{\boldsymbol{r}}^*_{t-1}(\widehat{\boldsymbol{r}}^*_{t-1})'\\ \widetilde{S}_{10} &= \frac{1}{T}\sum_{t=p+1}^{T}\widehat{\boldsymbol{r}}^*_{t-1}\widehat{\boldsymbol{r}}_t'\\ \mathrm{tr}_* &= T\,\mathrm{tr}\left(\widetilde{S}_{00}^{-1}\widetilde{S}_{10}'\widetilde{S}_{11}^{-1}\widetilde{S}_{10}\right)\end{aligned} \tag{8}$$

In the single-equation approach, this controlled-misspecification approach takes us e.g. to the no error correction OLS test regression

$$x_{t1} = \widehat{\alpha}^*_1 x^*_{t-1,1} + \sum_{k=2}^{K}\widehat{\alpha^*_1\beta_k}x^*_{t-1,k} + \widehat{\boldsymbol{a}}_1'\boldsymbol{x}_{t-1} + \ldots + \widehat{\boldsymbol{a}}_p'\boldsymbol{x}_{t-p} + \widehat{u}_{t1}, \tag{9}$$

where $\boldsymbol{a}_1', \ldots, \boldsymbol{a}_p'$ are the first rows of $A_1, \ldots, A_p$. The test is based on the OLS t statistic of α^*_1,

$$t_{\alpha^*_1} = \frac{\widehat{\alpha}^*_1}{\widehat{\sigma}_{\widehat{\alpha}^*_1}}. \tag{10}$$

For the no error-correction test based on the conditional form of the single-equation model, the test regression becomes with short-run dynamics

$$x_{t1} = \underline{\widehat{\alpha}}^*_1 x^*_{t-1,1} + \underline{\widehat{\boldsymbol{\delta}}}'\boldsymbol{z}^*_{t-1} + \underline{\widehat{\boldsymbol{a}}}_0'\boldsymbol{z}_t + \underline{\widehat{\boldsymbol{a}}}_1'\boldsymbol{x}_{t-1} + \ldots + \underline{\widehat{\boldsymbol{a}}}_p'\boldsymbol{x}_{t-p} + \underline{\widehat{u}}_{t1}, \tag{11}$$

with $\boldsymbol{z}_t = (x_{t2}, \ldots, x_{tK})'$.

A final note: when $K = 1$, the augmented LM test for fractional integration proposed by Demetrescu et al. (2008) is recovered from all three tests. The augmented LM test, however, is motivated by the LM principle; see Robinson (1994), Tanaka (1999), Breitung and Hassler (2002), and Demetrescu et al. (2008). The version of the test based on the conditional version of the model (11) resembles closely Hassler and Breitung's (2006) residual based LM-type fractional cointegration test; in fact, we shall compare the two in the Monte Carlo section.

3 Asymptotic results and discussion

3.1 Baseline results

We begin with the examination of the single-equation based statistic. For the test based on the reduced form, we have the following result for the t statistic for $\alpha_1^* = 0$ in (9).

Proposition 1 *Under Assumptions 1, 2 and 3 with $b = b_T$ from (7), it holds for $t_{\alpha_1^*}$ from (10) that*

$$t_{\alpha_1^*} \xrightarrow{d} \mathcal{N}\left(\frac{c\alpha_1 q}{\sigma_*}, 1\right)$$

as $T \to \infty$, where $\sigma_* = \text{plim}\,\sqrt{T}\widehat{\sigma}_{\widehat{\alpha}_1^*}$ from (10) and q is defined in the proof.

Proof: *see the Appendix.*

Although the assumptions allow for endogeneity (via contemporary correlation), the asymptotic null distribution is free of nuisance parameters irrespective of the covariance matrix of $\boldsymbol{u}_t$. In other words, the test is not affected by endogeneity under the null. Also, the asymptotic behavior does not depend on the number of regressors. This is in stark contrast to the integer case of Banerjee et al. (1998). And asymptotic normality is a nice result for practitioners: p values can easily be computed based on the standard normal cumulative distribution function.

Remark 1 *If the innovations $\boldsymbol{u}_t$ are not iid, but still have the martingale difference property, the result holds under an additional weak moment assumption when the innovations $\boldsymbol{u}_t$ are conditionally homoskedastic. In the case of conditional heteroskedasticity (e.g. when $\boldsymbol{u}_t$ follows a multivariate GARCH-type process), asymptotic normality of the test statistic is still given under a slightly stricter moment condition than finite kurtosis, but the usual standard errors are asymptotically biased. Following Demetrescu et al. (2008), this is easily accommodated by using White heteroskedasticity-consistent standard errors. Moreover, Kew and Harris (2009) show White standard errors to account for unconditional heteroskedasticity as well.*

Remark 2 *For all three proposed tests, deterministic components are easily dealt with if removed from the differenced data and not from the original observations (Robinson, 1994). E.g. for the empirically often more relevant case $d = 1$, this implies removing a non-zero mean from the differences $\boldsymbol{x}_t$ instead of detrending $\boldsymbol{y}_t$. Since the details offer no new insight we omit them.*

The no error-correction test in the conditional form is based on the t statistic of the coefficient of $x_{t-1,1}^*$ in Equation (11) above,

$$t_{\underline{\alpha}_1^*} = \frac{\widehat{\underline{\alpha}}_1^*}{\widehat{\sigma}_{\widehat{\underline{\alpha}}_1^*}}. \tag{12}$$

Our assumed model implies that $\boldsymbol{a}_0 = \boldsymbol{0}$. Still, the corresponding estimator of the coefficient vector associated to $\boldsymbol{z}_t$ is not consistent for $\boldsymbol{0}$ due to endogeneity caused by contemporaneous correlation of the elements of the innovations, hence the different notation for the parameters of Equation (11). But this does not affect the properties of the modified test under the null; see Proposition 2.

Proposition 2 *Under Assumptions 1, 2 and 3 with* $b = b_T$ *from (7), it holds for* $t_{\underline{\alpha}_1^*}$ *from (12) that*

$$t_{\underline{\alpha}_1^*} \xrightarrow{d} \mathcal{N}\left(\frac{c\underline{q}\underline{\alpha}_1}{\underline{\sigma}^*}, 1\right)$$

as $T \to \infty$, *where* $\underline{\sigma}^* = \text{plim}\,\sqrt{T}\widehat{\sigma}_{\widehat{\underline{\alpha}}_1^*}$ *from (12) and* $\underline{\alpha}_1 = \alpha_1 + \sum_{j=2}^{K} \alpha_j \underline{a}_{0j}$, *with* $\underline{q}$ *and* $\underline{\boldsymbol{a}}_0 = (\underline{a}_{02}, \ldots, \underline{a}_{0K})$ *defined in the proof.*

Proof: *see the Appendix.*

Since $\underline{\alpha}_1$ can be non-zero even when $\alpha_1 = 0$, the conditional version of the test is to be preferred: a parameter constellation under which $\alpha_1 \neq 0$ but $\underline{\alpha}_1 = 0$ is in our opinion much less plausible as one with $\alpha_1 = 0$. Still, the test will not consistently reject the false null for any DGP exhibiting (fractional) cointegration: e.g. $\underline{\alpha}_1 = 0$ if $\alpha_1 = 0$, Σ has diagonal form and there is no short-run dynamics. The test due to Hassler and Breitung (2006) suffers from the same shortcoming in spite of being residuals-based; see again the provided Monte Carlo findings.

The system test, naturally, does not face such difficulties.

Proposition 3 *Under Assumptions 1, 2 and 3 with* $b = b_T$ *from (7), it holds for* tr_* *from* (8) *that*

$$\text{tr}_* \xrightarrow{d} \chi^2_{K^2,\mu}$$

as $T \to \infty$, *where the noncentrality parameter is given by* $\mu = c^2\,\text{tr}\left(S_x^{-1} S_* \Pi' S_*^{-1} \Pi S_*\right)$ *with* $S_x = \text{E}\left(\boldsymbol{x}_t \boldsymbol{x}_t'\right)$ *and* $S_* \equiv \lim_{t\to\infty} \text{E}\left(\boldsymbol{x}_{t-1}^* (\boldsymbol{x}_{t-1}^*)'\right)$.

Proof: *see the Appendix.*

The proofs implicity assume that the true lag order is used in computing the test statistics. In practical applications, however, the order of the vector autoregressive process is not known; moreover, the order may not even be finite. Assuming a stable vector autoregressive process of infinite order with certain summability conditions for its coefficients, the short-run component can be approximated by an AR process of order growing to infinity, see Demetrescu et al. (2008) for the treatment of the augmented LM test. Note, however, that data-dependent lag selection rules fail here, the cause being the effect model selection has on subsequent inference, see Leeb and Pötscher (2005) for a review of the theoretical background, as well as Demetrescu et al. (2011) for additional experimental evidence. This may seem to be a disadvantage when p is fixed or known. But the following subsection shows that such long autoregressions can account for the use of an estimated integration parameter d.[8]

[8]While Hassler and Breitung (2006) argue that plugging in an estimate in the unknown d case, even a $\sqrt{T}$ consistent one, asymptotically distorts the test statistic, their argument is given for a *fixed* lag order p.

3.2 Unknown integration order

We have dealt sofar with the lack of identification under the null of b by building on a known d. With few exceptions, assuming d to be known is a daring thing to do; moreover, if it is the wrong value one assumes, the size control of subsequent tests is typically ruined, a misspecified d being mistaken for fractional cointegration by the typical fractional cointegration test. This subsection examines the effects on the asymptotic null distributions of the studied tests of working with some estimator of d, say $\widehat{d}$.

If plugging in $\widehat{d}$ to build the fractional differences before computing the fractional cointegration test, one ends up working with $\widetilde{\boldsymbol{x}}_t = \Delta_+^{\hat{d}} \boldsymbol{y}_t$ instead of the true differences; the two sets of differences are related,

$$\widetilde{\boldsymbol{x}}_t = \Delta_+^{\hat{d}-d} \boldsymbol{x}_t.$$

Under the null hypothesis, this implies the DGP

$$A(L)\,\Delta_+^{d-\hat{d}} \widetilde{\boldsymbol{x}}_t = \boldsymbol{u}_t,$$

so the input for the statistics from (10), (8) or (12) is actually an infinite-order vector autoregression with time-varying coefficients, since $\widehat{d}$ depends implicitly (and in a random manner) on T. The coefficients do converge to those of the model given by $A(L)$ since $\hat{d}$ is consistent. Still, one must analyze the behavior of the proposed tests under the DGP given by

$$\widetilde{\boldsymbol{x}}_t = \sum_{j=1}^{t-1} \widetilde{A}_{j,T} \widetilde{\boldsymbol{x}}_{t-j} + \boldsymbol{u}_t, \tag{13}$$

where $\widetilde{A}_{j,T}$ represents the convolution of the two filters, $A(L)$ and $\Delta_+^{d-\hat{d}}$, i.e.

$$\widetilde{A}_{j,T} = \sum_{k=0}^{j} A_k \widehat{\delta}_{j-k,T} = \sum_{k=0}^{p} A_k \widehat{\delta}_{j-k,T}$$

with $\widehat{\delta}_{j,T}$ the coefficients of $\Delta_+^{d-\hat{d}}$ and the convention that $\widehat{\delta}_{j,T} = 0$ for $j < 0$. Lemma 5 in the Appendix shows that $\left\| \widetilde{A}_{j,T} - A_j \right\| \leq C \frac{1}{j} \left| \widehat{d} - d \right| T^{\left|\widehat{d}-d\right|}$ such that $\widetilde{A}_{j,T} \xrightarrow{p} 0$ uniformly for $j > p$ and $\widetilde{A}_{j,T} \xrightarrow{p} A_j$ for $0 \leq j \leq p$.

For autoregressive modelling of order p as would be imposed by the short memory model order in Assumption 2, this implies

$$\widetilde{\boldsymbol{x}}_t = \sum_{j=1}^{p} \widetilde{A}_{j,T} \widetilde{\boldsymbol{x}}_{t-j} + \left(\boldsymbol{u}_t + \sum_{j=p+1}^{t-1} \widetilde{A}_{j,T} \widetilde{\boldsymbol{x}}_{t-j} \right)$$

so the shocks of an AR(p) model fitted to $\widetilde{\boldsymbol{x}}_t$ would only be approximately uncorrelated. Even when $\widehat{d} \xrightarrow{d} d$, this approximation has an impact on the limiting distribution of the studied test statistics: although $\widetilde{A}_{j,T} \xrightarrow{p} 0$ for $j > p$, the approximation error $\sum_{j=p+1}^{t-1} \widetilde{A}_{j,T} \widetilde{\boldsymbol{x}}_{t-j}$ does not vanish fast enough; cf. Hassler and Breitung (2006) in the univariate case and also Remark 2 of Avarucci and Velasco (2009).

In spite of this, Hassler and Meller (2013) find in Monte Carlo simulations that employing an estimated d does not appear to affect the size properties of their LM-based test for breaks in d. The cause can only be the implementation of the test, which requires p to be a deterministic function of T such that $p \to \infty$ at suitable rates to avoid problems with post-model selection inference. Loosely speaking, the serial correlation induced by the approximation error term $\sum_{j=p+1}^{t-1} \widetilde{A}_{j,T} \widetilde{\boldsymbol{x}}_{t-j}$ needs to be accounted for, and this is done by increasing the number of lags as T goes to infinity.

We formalize this intuition in Proposition 4 below, which shows that augmenting the above test regressions with p_T lagged differences instead of the true p (which is unknown anyway) and letting $p_T \to \infty$ together with $T \to \infty$ at suitable rates does account for the autocorrelation induced by plugging in an estimate of d, as long as $\widehat{d}$ converges fast enough. The proposition assumes finite 8th order moments to shorten the proofs by resorting to results of Demetrescu et al. (2008). This assumption can be relaxed at the cost of additional technicalities; the critical conditions are given below.

Assumption 4 *Let the order p_T of the approximating autoregressive polynomial $A(L)$ satisfy*

$$p_T = O(T^{\kappa_1}), \tag{14}$$

with $p_T \to \infty$ as $T \to \infty$, where $\kappa_1 \in \left(0, \frac{1}{4}\right)$.

Assumption 5 *Let $\widehat{d}$ be consistent estimator of d such that $\underline{\Delta} \leq \widehat{d} \leq \overline{\Delta}$ for some real constants $\underline{\Delta}, \overline{\Delta}$ and*

$$\widehat{d} - d = O_p\left(T^{-\kappa_2}\right),$$

with $\kappa_2 > \frac{3}{8}$.

Assumption 4 is standard; cf. Demetrescu et al. (2008). Nonparametric estimators $\widehat{d}$ such as the local Whittle or the log-periodogram estimators fulfill Assumption 5 for suitable choices of the bandwidth parameter. Assuming a compact parameter space for d is only a benign limitation, since $\underline{\Delta}, \overline{\Delta}$ may be chosen arbitrarily. We then have the following

Proposition 4 *Under Assumptions 1 through 5 and finite 8th order moments of error term, augmenting the test regressions with p_T lags, we have as $T \to \infty$ under the null of no cointegration (or no error-correction, respectively) that*

$$\widetilde{t}_{\alpha_1^*} - t_{\alpha_1^*} \xrightarrow{p} 0$$
$$\widetilde{t}_{\underline{\alpha}_1^*} - t_{\underline{\alpha}_1^*} \xrightarrow{p} 0$$
$$\widetilde{\mathrm{tr}}_* - \mathrm{tr}_* \xrightarrow{p} 0$$

where $\widetilde{\cdot}$ stands for the feasible test computed with $\widetilde{\boldsymbol{x}}_t$ instead of the (unobserved) $\boldsymbol{x}_t$.

Proof: *see the Appendix.*

Remark 3 *One important issue, sometimes sidestepped, is whether the series supposed to be cointegrated have the same degree of integration. Recently, Hualde (2013) proposed a semi-parametric test that could be employed to check this. The result in Proposition 4 helps dealing with the issue as well. Namely, applying the proposition to the case* $K = 1$*, one may conclude that, when testing some series for the null* $d = \widehat{d}$ *using the augmented LM test, the limiting distribution is still standard normal when the integration parameter is indeed* d*. This allows to check whether the persistence of one series is compatible with the estimated persistence of another series in the system, should* $\widehat{d}$ *converge fast enough as above.*

Remark 4 *For* $0 < b < 0.5$*, the results may be extended to the efficient Wald procedure of Avarucci and Velasco (2009). Examining the proof of Proposition 4, however, it becomes clear that additional restrictions may be required when* b *moves away from* 0*.*

4 Finite sample properties

In this section we examine the finite sample properties of the proposed tests by means of Monte Carlo experiments. The simulations are designed to answer several questions of interest. The first concerns the performance of the discussed test statistics relative to that of extant procedures. The second is whether the proposed single-equation fractional cointegration tests offer some advantage over system fractional cointegration tests. For this particular setup we focus on the situation where $d = 1$ and include Johansen's (1995) integer cointegration likelihood ratio test in the comparison as a benchmark, since it is widely used in practice. The third question is thus whether (or under which circumstances) fractional cointegration tests are to be preferred to the more widely spread integer cointegration test. Finally, the last question refers to the consequences of estimating d, thereby building on Section 3.2.

The Monte Carlo comparison includes the residual-based test due to Hassler and Breitung (2006) denoted by t_{HB}, the system test tr_{BH} proposed by Breitung and Hassler (2002), the tr_{AV} trace test by Avarucci and Velasco (2009), the trace test of Johansen (1995) denoted by tr_J, as well as the tests based on the $t_{\alpha_1^*}$ and $t_{\underline{\alpha}_1^*}$ statistics and their system counterpart denoted by tr_*.

We generate the data according to the bivariate process y_t with $y_t = 0$ for $t \leq 0$ and

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha} \left(\Delta^{-b} - 1\right) z_t + \boldsymbol{u}_t \quad (15)$$

$$z_t = \Delta y_{1t} - \Delta y_{2t} \quad (16)$$

$$\boldsymbol{u}_t = A \boldsymbol{u}_{t-1} + \boldsymbol{e}_t \quad (17)$$

for $t = 1, \ldots, T$. Thus, $\boldsymbol{y}_t \sim I(1)$ and $z_t \sim I(1-b)$. The error terms are generated independently with $\boldsymbol{e}_t \sim \mathcal{N}(0, \Sigma(\rho))$, where e_{1t} and e_{2t} has unit variance, the parameter $\rho = E(e_{1t}e_{2t})$ controls the dependence between innovations and thus, the degree of endogeneity, and is set to either 0 or 0.5. The equilibrium adjustment is controlled by the vector $\boldsymbol{\alpha}$ with $\boldsymbol{\alpha} = [-0.1, 0]$. Finally, b captures the degree of fractional cointegration. We let b take the values from {0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1} where values $0 - 0.5$ correspond to nonstationary deviations from equilibrium, and $0.6 - 1$ to stationary and short-memory deviations. Notice that the case of $b = 1$ recovers the classical unit root bivariate cointegration relation, and $b = 0$ a relationship without (fractional) cointegration, which is the null hypothesis of the corresponding testing problem.

The short-run dynamics are characterized by the matrix A, which we chose in turn as

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0.5 & 0.25 \\ 0 & 0.5 \end{bmatrix}.$$

The non-zero off-diagonal element in the last choice for A ensures that, given contemporaneous correlation, the first equation reacts to error correction in the second.

We assume a DGP without linear time trends, and we therefore take into account a constant in the cointegrating relations when computing tr_J (by means of usual demeaning). The other tests do not require adjustment since they rely on differences of order 1 of the generated series. Furthermore, if lag-augmenting, the number of lags for the Johansen trace test was selected by the Akaike information criterion, where the maximal number of lags was set according to the rule

$$p_{max} = \left[4(T/100)^{0.25}\right]. \tag{18}$$

The number of lags for the rest of the test is deterministic and equal to either zero if not lag-augmenting, or to p_{max} if augmenting; this particular choice of p_{max} was recommended by Demetrescu et al. (2008) in the univariate case. Finally, to obtain a suitable estimate of $\widehat{d}$ as well as $\widehat{b}$ for the Avarucci and Velasco (2009) statistic, (since tr_{AV} statistic relies on the consistent estimator of b) we resort to the exact local Whittle estimator applied with first-order differences, where for simplicity we only estimate d from the first series of the VAR; i.e. $\widehat{d} = \widehat{\delta}_{\Delta y_{1t}} + 1$, with bandwidth $m = T^{0.8}$. To this end, the computational routine of Shimotsu (2010) was employed.

The figures in Appendix C report the rejection frequencies in various scenarios using a nominal significance level of 5%. For the convenience of exposition and discussion each figure contains four panels: the upper-left panel reports rejection frequencies for all test statistics; the upper-right one concentrates only on the system tests while lower-left and right panels show the size/power properties of the single equation tests in the first and second equation of the considered DGP. Each result relies on 5000 Monte Carlo replications and the sample size is $T = 200$ throughout. (Results for the larger sample size $T = 500$ deliver essentially the same answers to the questions of interest and are available upon request.)

Figures 1−4 describe the benchmark scenario when d is known and set to unity, with different scenarios for short run dynamics and lag augmentation. The results reveal that all considered test statistics have similar size control. In particular, when no short run dynamics enter the DGP (see Figure 1), all tests perform close to the nominal level of 5%. Further, if the presence of short run dynamics is taken into account, see Figures 2-4, we observe marginal size distortions with rejection frequencies about $6-7\%$. Concerning the power of the tests, the Johansen procedure outperforms the rest of the tests in the benchmark scenarios without any short run dynamics for values of b between 0.6 and 1; see Figures $1-2$. This was to be expected considering that the LR cointegration test is designed for a model with $b = 1$. A nonmonotonic power curve can be observed for the single-equation tests in Figure 1, which is actually not uncommon for LM-based tests in general. This phenomenon vanishes when lag-augmenting. The single-equation tests also have marginally better power for the values of b between 0.1 and 0.5. However, an interesting observation is made when matrix A is set to nonzero values which allows for short run dynamics in our DGP, see Figure 3. While all system tests are outperformed by Johansen's test procedure for the values of b between 0.8 and 1, it is clear that the single equation tests t_{HB}, $t_{\underline{\alpha}_1}$ and $t_{\underline{\alpha}_1^*}$ reject more often under the alternative for all values of b when compared to system tests. Moreover, allowing for contemporaneous correlation between error terms e_{1t} and e_{2t} with $\rho = 0.5$, see Figure 4, further improves power properties of the proposed single equation test $t_{\underline{\alpha}_1^*}$ (see also Figures 8 and 10 where d is estimated). The latest observations verify the theoretical discussion of Sections 2 and 3 on the potential benefits of using $t_{\underline{\alpha}_1^*}$, and provide a guidance to the answer of the second question when single-equation test are favoured compare to their system counterparts.

In order to answer the third and fourth question, we turn our attention to the empirically relevant situations where parameter d is unknown and has to be estimated. Two scenarios are considered when $d = 1$, illustrated by Figures $5-7$, and $d = 0.8$, illustrated by Figures $8-10$. In the first case (Figures $5-7$), we obtain results on the consequences of estimating d, that is in line with the theoretical findings of Section 3.2. In particular, the estimation of d causes a size distortion for all fractionally based tests if not lag-augmenting, and the lag augmentation of the testing regression appears to be a reasonable remedy to this issue in small samples. Finally, Figures $8-10$ provide an illustration of the situation when parameter d is mistakenly assumed to be 1. The Johansen testing procedure is seen there to be seriously oversized, on the level of 15% of rejections, even when the test regression is augmented according to rule (18). With estimated d and lag augmentation (Figures 9 and 10), fractional cointegration tests – especially the system tests – can be oversized, but lag augmentation serves its purpose, especially considering that nonparametric estimators of d like the one used here are quite imprecise for $T = 200$. Finally, in both situations for $d = 1$ and $d = 0.8$ with $\rho = 0.5$ (see Figure 7 and 10), we also observe that the proposed conditional single-equation test $t_{\underline{\alpha}_1^*}$ remains an attractive alternative to system tests and other single-equation testing procedures.

5 Concluding remarks

The paper proposed tests for no fractional cointegration based either on checking for error correction in a single-equation framework, or on a system approach. All tests rely on a linear approximation of the fractional difference filter and have standard asymptotic limiting distributions under the null as well as under a sequence of local alternatives. The single-equation tests do not require exogeneity assumptions for validity under the null. When there is no error correction in the test equation, their power could be trivial, but may also be nontrivial, depending on the data generating process. At the same time, power can be higher when error-correction is only present in the studied equation when compared to a system test. The LM-based test proposed by Hassler and Breitung (2006) suffers from the same shortcoming; in fact the conditional test and the test proposed by Hassler and Breitung are equivalent when there is no short-run dynamics. The system test is the analog of the LR cointegration rank test due to Johansen (1995) and does not suffer from the shortcomings of the single-equation tests. The tests require as input the common fractional integration parameter d of the system, but we show them to work with an estimated d as well, provided that the number of lags goes to infinity with the sample size T at a suitable rate and the estimator $\widehat{d}$ converges fast enough. This asymptotic equivalence result can also be used to check whether the series in the system actually share a common value of d, which is a prerequisite for fractional cointegration.

The Monte Carlo analysis revealed that, for known d, the LR test for integer cointegration is to be preferred when the deviations from the long-run equilibrium are close to short memory. But if the deviations from long-run equilibrium are highly persistent, fractional cointegration tests have clearly better power properties Furthermore, they are able to deal with estimation of the degree of fractional integration of the components of the VAR system.

Appendix

When interested in the first equation of the system, we shall denote the first line of $\widetilde{A}_{j,T}$ by $\widetilde{\boldsymbol{a}}'_{j,T}$.

A Auxiliary Results

Lemma 5 *Let Assumption 5 hold true and denote* $(1-L)^{\widehat{d}-d} = \sum_{j\geq 0} \underline{\widehat{\delta}}_{j,T} L^j$, $(1-L)^{d-\widehat{d}} = \sum_{j\geq 0} \widehat{\delta}_{j,T} L^j$, *with* $\widehat{\delta}_{0,T} = 1$ *and* $\widehat{\delta}_{j,T} = \frac{1}{j}\prod_{i=0}^{j-1}\left(i - \left(d-\widehat{d}\right)\right)$ *for* $j \geq 1$.

1. *Then exists a constant* $C > 0$ *such that for all* $j \geq 1$,

$$\left|\widehat{\delta}_{j,T}\right| \leq \frac{1}{j} C \left|\widehat{d}-d\right| T^{\left|\widehat{d}-d\right|} \qquad \text{and} \qquad \left|\underline{\widehat{\delta}}_{j,T}\right| \leq \frac{1}{j} C \left|\widehat{d}-d\right| T^{\left|\widehat{d}-d\right|}, \tag{19}$$

   where $T^{\left|\widehat{d}-d\right|} \xrightarrow{p} 1$.

2. *For all* $j \geq 0$ *it holds that*

$$\left\|\widetilde{A}_{j,T} - A_j\right\| \leq \frac{1}{j} C \left|\widehat{d}-d\right| T^{\left|\widehat{d}-d\right|}, \tag{20}$$

   with $A_j = 0$ *for* $j > p$.

Proof of item 1 The proofs are essentially the same for $\widehat{\delta}_{j,T}$ or $\underline{\widehat{\delta}}_{j,T}$ so focus on $\widehat{\delta}_{j,T}$ and let $q = d - \widehat{d}$. Let furthermore j_0 be the largest integer smaller than $\max\left\{|\underline{\Delta} - d| \, ; \left|\overline{\Delta} - d\right|\right\} + 1$; note that $j_0 \geq 1$ and is fixed. Let us now examine for $j \geq 1$

$$\widehat{\delta}_{j,T} = -\frac{q}{j}\prod_{k=1}^{j-1}\frac{k-q}{k} = -\frac{q}{j}\prod_{k=1}^{j_0-1}\left(1-\frac{q}{k}\right)\prod_{k=j_0}^{j-1}\left(1-\frac{q}{k}\right).$$

(with the usual convention that $\prod_1^0 = 1$). Since j_0 is fixed the first product is bounded, and the result follows trivially for $1 \leq j \leq j_0$. To establish the desired upper bound for the modulus of $\widehat{\delta}_{j,T}$ for $j \geq j_0 + 1$, we now distinguish two cases, $q < 0$ and $q > 0$ (the desired inequality holds trivially in the case $q = 0$).

If $q > 0$, we have with $0 < q < 1$ that $0 < 1 - q/k < 1 - q/(j-1) < 1$ and thus

$$0 < \prod_{k=j_0}^{j-1}\left(1-\frac{q}{k}\right) < \left(1-\frac{q}{j-1}\right)^{j-1} < 1.$$

Hence for $q > 0$ and any $C \geq 1$,

$$\left|\widehat{\delta}_{j,T}\right| \leq \frac{|q|}{j} \leq C\frac{|q|}{j}T^{|q|},$$

since $T^{|q|} \geq 1$ for all $T \geq 2$.

If on the other hand $q < 0$, $\prod_{k=j_0}^{j-1}\left(1-\frac{q}{k}\right) = \prod_{k=j_0}^{j-1}\left(1+\frac{|q|}{k}\right)$ and the above upper bound does not apply. But we have that

$$\log\prod_{k=j_0}^{j-1}\left(1+\frac{|q|}{k}\right) = \sum_{k=j_0}^{j-1}\log\left(1+\frac{|q|}{k}\right) \leq \sum_{k=1}^{j-1}\frac{|q|}{k} < |q|\log j + C,$$

thanks to the properties of the harmonic series, so

$$\prod_{k=j_0}^{j-1}\left(1+\frac{|q|}{k}\right) \leq Cj^{|q|},$$

for some $C > 1$ and thus

$$\left|\widehat{\delta}_{j,T}\right| = \frac{|q|}{j}\prod_{k=1}^{j-1}\left(1 - \frac{q}{k}\right) \leq C\frac{|q|}{j}j^{|q|}.$$

Note that

$$0 < |q|\log j \leq |q|\log T \to 0,$$

since $q = d - \widehat{d}$ is assumed to vanish at polynomial rates.

Proof of item 2 For $j = 0$, the inequality is trivial. Let then $1 \leq j \leq p$; then

$$\widetilde{A}_{j,T} = \widehat{\delta}_{j,T}I_k + \widehat{\delta}_{j-1,T}A_1 + \ldots + A_j,$$

such that

$$\left\|\widetilde{A}_{j,T} - A_j\right\| \leq j \max_{0\leq l\leq j-1}\|A_k\| \max_{1\leq l\leq j}\left|\widehat{\delta}_{l,T}\right|.$$

With $j \leq p$ fixed and result of Lemma 1, we then obtain for suitable C

$$\left\|\widetilde{A}_{j,T} - A_j\right\| \leq C\frac{1}{j}\left|\widehat{d} - d\right|T^{\left|\widehat{d}-d\right|}$$

as required. For $j > p$,

$$\widetilde{A}_{j,T} = \widehat{\delta}_{j,T}I_k + \widehat{\delta}_{j-1,T}A_1 + \ldots + \widehat{\delta}_{j-p,T}A_p,$$

so, using the same reasoning as above,

$$\left\|\widetilde{A}_{j,T}\right\| \leq p \max_{0\leq l\leq p-1}\|A_l\| \max_{1\leq l\leq p}\left|\widehat{\delta}_{l,T}\right| \leq C\frac{1}{j}\left|\widehat{d} - d\right|T^{\left|\widehat{d}-d\right|}.$$

Lemma 6 *There exist nonnegative stochastic sequences $v_{t,T}$, $v^*_{t,T}$ such that*

$$\|\widetilde{\boldsymbol{x}}_t - \boldsymbol{x}_t\| \leq C\left|\widehat{d} - d\right|T^{\left|\widehat{d}-d\right|}v_{t,T}, \tag{21}$$

and

$$\left\|\widetilde{\boldsymbol{x}}^*_{t-1} - \boldsymbol{x}^*_{t-1}\right\| \leq C\left|\hat{d} - d\right|T^{\left|\widehat{d}-d\right|}v^*_{t-1,T}, \tag{22}$$

where $0 \leq \max_t \mathrm{E}\left(|v_{t,T}|\right) = \max_t \mathrm{E}\left(v_{t,T}\right) \leq C\,\log T$ *and* $\max_t \mathrm{E}\left(\left|v^*_{t-1,T}\right|\right) = \max_t \mathrm{E}\left(v^*_{t-1,T}\right) \leq C\,\log^2 T$.

Proof Since $\widetilde{\boldsymbol{x}}_t = \Delta^{\hat{d}-d}\boldsymbol{x}_t$ we have that

$$\widetilde{\boldsymbol{x}}_t - \boldsymbol{x}_t = \sum_{j=1}^{t-1}\underline{\widehat{\delta}}_{j,T}\boldsymbol{x}_{t-j}.$$

Using results of Lemma 5 item 1 gives

$$\|\widetilde{\boldsymbol{x}}_t - \boldsymbol{x}_t\| \leq \sum_{j=1}^{t-1}\left|\underline{\widehat{\delta}}_{j,T}\right|\|\boldsymbol{x}_{t-j}\| \leq C\left|\hat{d} - d\right|T^{\left|\hat{d}-d\right|}\sum_{j=1}^{t-1}\frac{1}{j}\|\boldsymbol{x}_{t-j}\| = C\left|\hat{d} - d\right|T^{\left|\hat{d}-d\right|}v_{t,T},$$

where

$$0 \leq v_{t,T} = \sum_{j=1}^{t-1}\frac{1}{j}\|\boldsymbol{x}_{t-j}\|,$$

such that

$$\mathrm{E}\left(|v_{t,T}|\right) = \mathrm{E}\left(\sum_{j=1}^{t-1}\frac{1}{j}\left\|\boldsymbol{x}_{t-j}\right\|\right) \le \mathrm{E}\left(\left\|\boldsymbol{x}_{t-j}\right\|\right)\sum_{j=1}^{T-1}\frac{1}{j} \le C_1\,\log T + C_2 \quad \forall t,$$

due to the logarithmic behavior of the harmonic series and the sequence $\boldsymbol{x}_t$ being uniformly L_2-bounded.

The proof of (22) is completed along the same lines:

$$\widetilde{\boldsymbol{x}}_t^* - \boldsymbol{x}_t^* = \sum_{j=1}^{t-1}\frac{1}{j}\left(\widetilde{\boldsymbol{x}}_{t-j} - \boldsymbol{x}_{t-j}\right),$$

leading to

$$\left\|\widetilde{\boldsymbol{x}}_t^* - \boldsymbol{x}_t^*\right\| \le \sum_{j=1}^{t-1}\frac{1}{j}\left\|\widetilde{\boldsymbol{x}}_{t-j} - \boldsymbol{x}_{t-j}\right\| \le C\left|\hat{d}-d\right| T^{\left|\hat{d}-d\right|}\sum_{j=1}^{T-1}\frac{1}{j}v_{t-j,T} = C\left|\hat{d}-d\right| T^{\left|\hat{d}-d\right|}v_{t-1,T}^*.$$

Using (21) and to the logarithmic behavior of the harmonic series gives

$$\mathrm{E}\left(\left|v_{t-1,T}^*\right|\right) \le \sum_{j=1}^{T-1}\frac{1}{j}\,\mathrm{E}\left(|v_{t-j,T}|\right) \le C\,\log^2 T.$$

Lemma 7 *Under Assumptions 1 through 5, we have as $T\to\infty$ that*

1. $\max_{1\le j\le T-1}\left\|\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{\boldsymbol{x}}_{t-1}^*\widetilde{\boldsymbol{x}}_{t-j}' - \frac{1}{T}\sum_{t=j+1}^{T}\boldsymbol{x}_{t-1}^*\boldsymbol{x}_{t-j}'\right\| = O_p\left(T^{-\kappa_2}\log^2 T\right)$;
2. $\mathrm{E}\left(\left\|\frac{1}{T}\sum_{t=j+1}^{T}\boldsymbol{x}_{t-1}^*\boldsymbol{x}_{t-j}'\right\|\right) \le C\left(\eta^{-j} + T^{-\frac{1}{2}}\right) \quad \forall 1\le j\le T-1$;
3. $\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}' - \frac{1}{T}\sum_{t=p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{z}}_{t-1}'\right\| = O_p\left(T^{\kappa_1-\kappa_2}\log T\right)$;
4. $\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right\| = O_p(1)$.

Proof of item 1

Write first

$$\begin{aligned}\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{\boldsymbol{x}}_{t-1}^*\widetilde{\boldsymbol{x}}_{t-j}' &= \frac{1}{T}\sum_{t=j+1}^{T}\left(\widetilde{\boldsymbol{x}}_{t-1}^*-\boldsymbol{x}_{t-1}^*\right)\boldsymbol{x}_{t-j}' + \frac{1}{T}\sum_{t=j+1}^{T}\left(\widetilde{\boldsymbol{x}}_{t-1}^*-\boldsymbol{x}_{t-1}^*\right)\left(\widetilde{\boldsymbol{x}}_{t-j}'-\boldsymbol{x}_{t-j}'\right)\\ &\quad + \frac{1}{T}\sum_{t=j+1}^{T}\boldsymbol{x}_{t-1}^*\left(\widetilde{\boldsymbol{x}}_{t-j}'-\boldsymbol{x}_{t-j}'\right) + \frac{1}{T}\sum_{t=j+1}^{T}\boldsymbol{x}_{t-1}^*\boldsymbol{x}_{t-j}'.\end{aligned}$$

Applying inequality the Cauchy-Schwarz inequality to the first term on the r.h.s. yields that

$$\left\|\frac{1}{T}\sum_{t=j+1}^{T}\left(\widetilde{\boldsymbol{x}}_{t-1}^*-\boldsymbol{x}_{t-1}^*\right)\boldsymbol{x}_{t-j}'\right\| \le \sqrt{\frac{1}{T}\sum_{t=j+1}^{T}\left\|\widetilde{\boldsymbol{x}}_{t-1}^*-\boldsymbol{x}_{t-1}^*\right\|^2\frac{1}{T}\sum_{t=j+1}^{T}\left\|\boldsymbol{x}_{t-j}\right\|^2},$$

where $\frac{1}{T}\sum_{t=j+1}^{T}\left\|\boldsymbol{x}_{t-j}\right\|^2 \le \frac{1}{T}\sum_{t=1}^{T}\left\|\boldsymbol{x}_t\right\|^2$ for all j (so $\max_{1\le j\le T-1}\frac{1}{T}\sum_{t=j+1}^{T}\left\|\boldsymbol{x}_{t-j}\right\|^2 \le \frac{1}{T}\sum_{t=1}^{T}\left\|\boldsymbol{x}_t\right\|^2$) and

$$\frac{1}{T}\sum_{t=j+1}^{T}\left\|\widetilde{\boldsymbol{x}}_{t-1}^*-\boldsymbol{x}_{t-1}^*\right\|^2 \le \frac{1}{T}\sum_{t=j+1}^{T}C^2\left|\widehat{d}-d\right|^2 T^{2\left|\widehat{d}-d\right|}\left(v_{t-1,T}^*\right)^2 = C\frac{1}{T}\left|\widehat{d}-d\right|^2 T^{2\left|\widehat{d}-d\right|}\sum_{t=j+1}^{T}\left(v_{t-1,T}^*\right)^2,$$

with $v_{t-1,T}^*$ defined in Lemma 6. Therefore, given that $\sum_{t=j+1}^{T}\left(v_{t-1,T}^*\right)^2 \le \sum_{t=1}^{T}\left(v_{t-1,T}^*\right)^2$, it follows that

$$\max_{1\le j\le T-1}\left\|\frac{1}{T}\sum_{t=j+1}^{T}\left(\widetilde{\boldsymbol{x}}_{t-1}^*-\boldsymbol{x}_{t-1}^*\right)\boldsymbol{x}_{t-j}'\right\| \le C\left|\widehat{d}-d\right| T^{\left|\widehat{d}-d\right|}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left\|\boldsymbol{x}_t\right\|^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(v_{t-1,T}^*\right)^2}.$$

Moreover, Minkowski's inequality indicates that

$$\mathrm{E}\left(\left(v_{t-1,T}^{*}\right)^{2}\right) \leq \left(\sum_{j=1}^{t-1}\frac{1}{j}\sqrt{\mathrm{E}\left(v_{t-j,T}^{2}\right)}\right)^{2},$$

where

$$\begin{aligned}\mathrm{E}\left(v_{t-j,T}^{2}\right) &= \mathrm{E}\left(\left(\sum_{j=1}^{t-1}\frac{1}{j}\left\|\boldsymbol{x}_{t-j}\right\|\right)^{2}\right) = \sum_{j=1}^{t-1}\sum_{k=1}^{t-1}\frac{1}{j}\frac{1}{k}\mathrm{E}\left(\left\|\boldsymbol{x}_{t-j}\right\|\left\|\boldsymbol{x}_{t-k}\right\|\right),\\ &\leq C\log^{2}T\end{aligned}$$

since $\mathrm{E}\left(\left\|\boldsymbol{x}_{t-j}\right\|\left\|\boldsymbol{x}_{t-k}\right\|\right) \leq \sqrt{\mathrm{E}\left(\left\|\boldsymbol{x}_{t-j}\right\|^{2}\right)\mathrm{E}\left(\left\|\boldsymbol{x}_{t-k}\right\|^{2}\right)}$ is uniformly bounded under our assumption and $\sum_{k=1}^{t-1}\frac{1}{k} \leq C\log T$ as before. Then,

$$\mathrm{E}\left(\left(v_{t-1,T}^{*}\right)^{2}\right) \leq C\log^{4}T,$$

this implies in turn that $\frac{1}{T}\sum_{t=1}^{T}\left(v_{t-1,T}^{*}\right)^{2} = O_{p}\left(\log^{4}T\right)$ thanks to Markov's inequality and, with $\frac{1}{T}\sum_{t=1}^{T}\left\|\boldsymbol{x}_{t}\right\|^{2} = O_{p}\left(1\right)$, we obtain that

$$\max_{1\leq j\leq T-1}\left\|\frac{1}{T}\sum_{t=j+1}^{T}\left(\widetilde{\boldsymbol{x}}_{t-1}^{*}-\boldsymbol{x}_{t-1}^{*}\right)\boldsymbol{x}_{t-j}'\right\| = O_{p}\left(T^{-\kappa_{2}}\log^{2}T\right).$$

The second and third terms on the r.h.s. of the equation decomposing $\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{\boldsymbol{x}}_{t-1}^{*}\widetilde{\boldsymbol{x}}_{t-j}'$ have the same magnitude order, as is easily checked using similar arguments.

Proof of item 2 Recall that the induced norm is bounded above by the Euclidean norm, so applying Jensen's inequality takes us to

$$\begin{aligned}\mathrm{E}\left(\left\|\frac{1}{T}\sum_{t=j+1}^{T}\boldsymbol{x}_{t-1}^{*}\boldsymbol{x}_{t-j}'\right\|\right) &= \mathrm{E}\left(\sqrt{\sum_{k=1}^{K}\sum_{l=1}^{K}\left(\frac{1}{T}\sum_{t=j+1}^{T}x_{t-1,k}^{*}x_{t-j,l}\right)^{2}}\right) \leq \sqrt{\sum_{k=1}^{K}\sum_{l=1}^{K}\mathrm{E}\left(\frac{1}{T}\sum_{t=j+1}^{T}x_{t-1,k}^{*}x_{t-j,l}\right)^{2}}\\ &\leq K^{2}\sqrt{\max_{1\leq k,l\leq K}\mathrm{E}\left(\frac{1}{T}\sum_{t=j+1}^{T}x_{t-1,k}^{*}x_{t-j,l}\right)^{2}}.\end{aligned}$$

Then,

$$\mathrm{E}\left(\left|\frac{1}{T}\sum_{t=j+1}^{T}x_{t-1,k}^{*}x_{t-j,l}\right|^{2}\right) = \frac{1}{T^{2}}\sum_{t=2}^{T}\sum_{s=2}^{T}\mathrm{E}\left(x_{t-1,k}^{*}x_{t-j,l}x_{s-1,k}^{*}x_{s-j,l}\right).$$

In turn,

$$\mathrm{E}\left(x_{t-1,k}^{*}x_{t-j,l}x_{s-1,k}^{*}x_{s-j,l}\right) = \sum_{m_{1}=1}^{t-1}\sum_{m_{2}=1}^{s-1}\frac{1}{m_{1}}\frac{1}{m_{2}}\mathrm{E}\left(x_{t-1-m_{1},k}x_{t-j,l}x_{s-1-m_{2},k}x_{s-j,l}\right),$$

and we analyze the expectation on the r.h.s.,

$$\begin{aligned}&\mathrm{E}\left(x_{t-1-m_{1},k}x_{t-j,l}x_{s-1-m_{2},k}x_{s-j,l}\right) =\\ &= \sum_{n_{1}\geq0}\sum_{n_{2}\geq0}\sum_{n_{3}\geq0}\sum_{n_{4}\geq0}\mathrm{E}\left(\boldsymbol{b}_{n_{1},k}\boldsymbol{u}_{t-1-m_{1}-n_{1}}\boldsymbol{b}_{n_{2},l}\boldsymbol{u}_{t-j-n_{2}}\boldsymbol{b}_{n_{3},k}\boldsymbol{u}_{s-1-m_{2}-n_{3}}\boldsymbol{b}_{n_{4},l}\boldsymbol{u}_{s-j-n_{4}}\right),\end{aligned}$$

where the MA coefficients $\boldsymbol{b}_{n,k}$ (the kth line of the nth Wold coefficients matrix of $\boldsymbol{x}_t$) have exponential decay, i.e. $\|\boldsymbol{b}_{n,k}\| \leq C\eta^{-n}$ for some $C > 0$ and $\eta > 1$. Note that the above expectations are nonzero only if the four indices of $\boldsymbol{u}$ are either all equal or pairwise equal, and are uniformly bounded when nonzero. Summing up, we obtain that

$$|\mathrm{E}\left(x_{t-1-m_1,k}x_{t-j,l}x_{s-1-m_2,k}x_{s-j,l}\right)| \leq C\left(\eta^{-|j-m_1|}\eta^{-|j-m_2|} + \eta^{-|t-s-(m_1-m_2)|} + \eta^{-|t-s-(m_2-m_1)|}\right),$$

which leads to

$$\begin{aligned}\left|\mathrm{E}\left(x^*_{t-1,k}x_{t-j,l}x^*_{s-1,k}x_{s-j,l}\right)\right| &= \left|\sum_{m_1=1}^{t-1}\sum_{m_2=1}^{s-1}\frac{1}{m_1}\frac{1}{m_2}\mathrm{E}\left(x_{t-1-m_1,k}x_{t-j,l}x_{s-1-m_2,k}x_{s-j,l}\right)\right| \\ &\leq C\left(\eta^{-j} + \eta^{-|t-s|}\right),\end{aligned}$$

and thus to

$$\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=2}^{T}\mathrm{E}\left(x^*_{t-1,k}x_{t-j,l}x^*_{s-1,k}x_{s-j,l}\right) \leq C\left(\eta^{-j} + \frac{1}{T}\right)$$

as required for the result.

Proof of item 3

Note that

$$\begin{aligned}\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}'_{t-1} - \frac{1}{T}\sum_{t=p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{z}}'_{t-1} &= \frac{1}{T}\sum_{t=p_T+1}^{T}\left(\widetilde{\boldsymbol{z}}_{t-1} - \boldsymbol{z}_{t-1}\right)\left(\boldsymbol{z}_{t-1} - \widetilde{\boldsymbol{z}}_{t-1}\right)' \\ &+\frac{1}{T}\sum_{t=p_T+1}^{T}\left(\boldsymbol{z}_{t-1} - \widetilde{\boldsymbol{z}}_{t-1}\right)\boldsymbol{z}'_{t-1} + \frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\left(\boldsymbol{z}_{t-1} - \widetilde{\boldsymbol{z}}_{t-1}\right)'.\end{aligned}$$

Moreover,

$$\|\boldsymbol{z}_{t-1} - \widetilde{\boldsymbol{z}}_{t-1}\|^2 = \sum_{j=1}^{p_T}\|\boldsymbol{x}_{t-j} - \widetilde{\boldsymbol{x}}_{t-j}\|^2 = \sum_{j=1}^{p_T}C\left|\widehat{d} - d\right|^2 T^{2\left|\widehat{d}-d\right|}v^2_{t-1,T},$$

where $\mathbb{E}\left(v^2_{t-1,T}\right) \leq C\ln^2 T$. Hence,

$$\mathbb{E}\|\boldsymbol{z}_{t-1} - \widetilde{\boldsymbol{z}}_{t-1}\|^2 = O_p\left(T^{\kappa_1-2\kappa_2}\log^2 T\right),$$

with $\frac{1}{T}\sum_{t=p_T+1}^{T}\|\boldsymbol{z}_{t-1}\|^2 = O_p\left(T^{\kappa_1}\right)$ in turn applies that

$$\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\left(\boldsymbol{z}_{t-1} - \widetilde{\boldsymbol{z}}_{t-1}\right)\boldsymbol{z}'_{t-1}\right\| = O_p\left(T^{\kappa_1-\kappa_2}\log T\right),$$

and

$$\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\left(\widetilde{\boldsymbol{z}}_{t-1} - \boldsymbol{z}_{t-1}\right)\left(\boldsymbol{z}_{t-1} - \widetilde{\boldsymbol{z}}_{t-1}\right)'\right\| = O_p\left(T^{\kappa_1-2\kappa_2}\log^2 T\right).$$

The result follows with $T^{-\kappa_2} \to 0$.

Proof of item 4

This is a multivariate extension of Lemma 7 in of Demetrescu (2009).

Lemma 8 *Under Assumptions 1 through 5, the following properties hold true:*

1. $\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\boldsymbol{x}^*_{t-1}u_{t1} - \frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{x}}^*_{t-1}\widetilde{u}^{(p_T)}_{t1} = O_p\left(T^{\frac{1}{2}-2\kappa_2}\log^3 T\right)$
2. $\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{x}^*_{t-1}\boldsymbol{z}'_{t-1} - \frac{1}{T}\sum_{t=p_T+1}^{T}\widetilde{\boldsymbol{x}}^*_{t-1}\widetilde{\boldsymbol{z}}'_{t-1}\right\| = O_p\left(T^{\frac{\kappa_1}{2}-\kappa_2}\ln^2 T\right)$

3. $\left\|\left(\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right)^{-1}\right\| = O_p(1)$

4. $\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}u_{t1}\right\| = O_p\left(T^{\frac{\kappa_1}{2}}\right)$

5. $\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\widetilde{\boldsymbol{x}}_{t-1}^{*}\widetilde{\boldsymbol{z}}_{t-1}'\right\| = O_p(1)$

6. $\left\|\left(\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right)^{-1} - \left(\frac{1}{T}\sum_{t=p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{z}}_{t-1}'\right)^{-1}\right\| = O_p(T^{\kappa_1-\kappa_2}\log T)$

7. $\left\|\left(\frac{1}{T}\sum_{t=p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{z}}_{t-1}'\right)^{-1}\right\| = O_p(1)$

8. $\left\|\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{u}_{t1}^{(p_T)} - \frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\boldsymbol{z}_{t-1}u_{t1}\right\| = O_p\left(T^{\frac{1}{2}-2\kappa_2+\frac{\kappa_1}{2}}\log^2 T\right)$

9. $\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{x}_{t-1}^{*}\boldsymbol{x}_{t-1}^{*\prime} - \frac{1}{T}\sum_{t=p_T+1}^{T}\widetilde{\boldsymbol{x}}_{t-1}^{*}\widetilde{\boldsymbol{x}}_{t-1}^{*\prime}\right\| = O_p\left(T^{-\kappa_2}\log^3 T\right)$

Proof of item 1 We have that

$$\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\boldsymbol{x}_{t-1}^{*}u_{t1} - \frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{x}}_{t-1}^{*}\widetilde{u}_{t1}^{(p_T)} = \frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\left(\boldsymbol{x}_{t-1}^{*} - \widetilde{\boldsymbol{x}}_{t-1}^{*}\right)u_{t1} + \frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{x}}_{t-1}^{*}\sum_{j=p_T+1}^{t-1}\widetilde{\boldsymbol{a}}_j\widetilde{\boldsymbol{x}}_{t-j}. \quad (23)$$

To examine the first term on the r.h.s. of (23), $\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\left(\boldsymbol{x}_{t-1}^{*} - \widetilde{\boldsymbol{x}}_{t-1}^{*}\right)u_{t1}$, we start with a discussion of the the simpler expression $\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\left(\boldsymbol{x}_{t-1} - \widetilde{\boldsymbol{x}}_{t-1}\right)u_{t1}$, for which we have elementwise that

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=p_T+1}^{T}\left(x_{t-1,k} - \widetilde{x}_{t-1,k}\right)u_{t1} &= \frac{1}{\sqrt{T}}\sum_{t=p_T+1}^{T}\left(\sum_{j=1}^{t-1}\widehat{\underline{\delta}}_{j,T}x_{t-1-j,k}\right)u_{t1} \\ &= \frac{1}{\sqrt{T}}\sum_{j=p_T+1}^{T-1}\widehat{\underline{\delta}}_{j,T}\left(\sum_{t=j+1}^{T}x_{t-1-j,k}u_{t1}\right).\end{aligned}$$

Putting together the fact that $\mathrm{E}\left|\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}x_{t-1-j,k}u_{t1}\right|$ is uniformly bounded because $x_{t-1-j,k}u_{t1}$ is a uniformly L_2-bounded martingale difference sequence, results of Lemma 6 and logarithmic behaviour of harmonic series gives the following

$$\left|\frac{1}{\sqrt{T}}\sum_{t=p_T+1}^{T}\left(x_{t-1,k} - \widetilde{x}_{t-1,k}\right)u_{t1}\right| = O\left(T^{-\kappa_2}\log T\right).$$

Returning to $\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\left(\boldsymbol{x}_{t-1}^{*} - \widetilde{\boldsymbol{x}}_{t-1}^{*}\right)u_{t1}$, note that the reasoning is the same, but the coefficients of the sum terms $\left(\sum_{t=j+1}^{T}x_{t-1-j,k}u_{t1}\right)$ after rearranging decay at a slower rate, i.e.,

$$\frac{1}{\sqrt{T}}\sum_{t=p_T+1}^{T}\left(\boldsymbol{x}_{t-1}^{*} - \widetilde{\boldsymbol{x}}_{t-1}^{*}\right)u_{t1} = \frac{1}{\sqrt{T}}\sum_{j=p_T+1}^{T-1}\frac{\widehat{\underline{\delta}}_{j,T}}{j}\left(\sum_{t=j+1}^{T}\boldsymbol{x}_{t-1-j}u_{t1}\right),$$

where $\sum_{j=p_T+1}^{T-1}\frac{\widehat{\underline{\delta}}_{j,T}}{j}$ is of order $O_p\left(T^{-\kappa_2}\log^2 T\right)$; cf. Lemma (5).

For the second term on the r.h.s. of (23), rearrange terms to obtain elementwise

$$\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\widetilde{x}_{t-1,k}^{*}\sum_{j=p_T+1}^{t-1}\widetilde{\boldsymbol{a}}_{j,T}'\widetilde{\boldsymbol{x}}_{t-j} = \sum_{j=p_T+1}^{T-1}\widetilde{\boldsymbol{a}}_{j,T}'\left(\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\widetilde{x}_{t-1,k}^{*}\widetilde{\boldsymbol{x}}_{t-j}\right),$$

leading to

$$\left\|\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{x}}_{t-1}^{*}\sum_{j=p_T+1}^{t-1}\widetilde{\boldsymbol{a}}_{j,T}'\widetilde{\boldsymbol{x}}_{t-j}\right\| \leq C\sqrt{T}\sum_{j=p_T+1}^{T-1}\left\|\widetilde{\boldsymbol{a}}_{j,T}'\right\|\left\|\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{\boldsymbol{x}}_{t-1}^{*}\widetilde{\boldsymbol{x}}_{t-j}'\right\|.$$

We have by Lemma 5 item 2 that

$$\left\| \frac{1}{\sqrt{T}} \sum_{p_T+1}^{T} \widetilde{\boldsymbol{x}}_{t-1}^{*} \sum_{j=p_T+1}^{t-1} \widetilde{\boldsymbol{a}}_{j,T}' \widetilde{\boldsymbol{x}}_{t-j} \right\| \leq C\sqrt{T} \left| \widehat{d} - d \right| T^{\left|\widehat{d}-d\right|} \sum_{j=p_T+1}^{T-1} \frac{1}{j} \left\| \frac{1}{T} \sum_{t=j+1}^{T} \widetilde{\boldsymbol{x}}_{t-1}^{*} \widetilde{\boldsymbol{x}}_{t-j}' \right\|.$$

Now,

$$\begin{aligned} \sum_{j=p_T+1}^{T-1} \frac{1}{j} \left\| \frac{1}{T} \sum_{t=j+1}^{T} \widetilde{\boldsymbol{x}}_{t-1}^{*} \widetilde{\boldsymbol{x}}_{t-j}' \right\| &\leq \sum_{j=p_T+1}^{T-1} \frac{1}{j} \left\| \frac{1}{T} \sum_{t=j+1}^{T} \widetilde{\boldsymbol{x}}_{t-1}^{*} \widetilde{\boldsymbol{x}}_{t-j}' - \frac{1}{T} \sum_{t=j+1}^{T} \boldsymbol{x}_{t-1}^{*} \boldsymbol{x}_{t-j}' \right\| \\ &\quad + \sum_{j=p_T+1}^{T-1} \frac{1}{j} \left\| \frac{1}{T} \sum_{t=j+1}^{T} \boldsymbol{x}_{t-1}^{*} \boldsymbol{x}_{t-j}' \right\|. \end{aligned}$$

Using item 1 of Lemma 7,

$$\begin{aligned} \sum_{j=p_T+1}^{T-1} \frac{1}{j} \left\| \frac{1}{T} \sum_{t=j+1}^{T} \widetilde{\boldsymbol{x}}_{t-1}^{*} \widetilde{\boldsymbol{x}}_{t-j}' - \frac{1}{T} \sum_{t=j+1}^{T} \boldsymbol{x}_{t-1}^{*} \boldsymbol{x}_{t-j}' \right\| &\leq \sum_{j=p_T+1}^{T-1} \frac{1}{j} \max_{p_T<j<T} \left\| \frac{1}{T} \sum_{t=j+1}^{T} \widetilde{\boldsymbol{x}}_{t-1}^{*} \widetilde{\boldsymbol{x}}_{t-j}' - \frac{1}{T} \sum_{t=j+1}^{T} \boldsymbol{x}_{t-1}^{*} \boldsymbol{x}_{t-j}' \right\| \\ &= O_p\left(T^{-\kappa_2} \log^3 T\right); \end{aligned}$$

furthermore, item 2 of Lemma 7 takes us to

$$\begin{aligned} \mathrm{E}\left(\left| \sum_{j=p_T+1}^{T-1} \frac{1}{j} \left\| \frac{1}{T} \sum_{t=j+1}^{T} \boldsymbol{x}_{t-1}^{*} \boldsymbol{x}_{t-j}' \right\| \right| \right) &\leq \sum_{j=p_T+1}^{T-1} \frac{1}{j} \mathrm{E}\left(\left\| \frac{1}{T} \sum_{t=j+1}^{T} \boldsymbol{x}_{t-1}^{*} \boldsymbol{x}_{t-j}' \right\| \right) \\ &\leq C \sum_{j=p_T+1}^{T-1} \frac{1}{j} \left(\eta^{-j} + T^{-0.5} \right) \\ &= O\left(T^{-\frac{1}{2}} \log T \right) \end{aligned}$$

since $\sum_{j=p_T+1}^{T-1} \eta^{-j} \leq C\eta^{-p_T} = o\left(T^{-0.5}\right)$, so we have indeed

$$\begin{aligned} \left\| \frac{1}{\sqrt{T}} \sum_{p_T+1}^{T} \widetilde{\boldsymbol{x}}_{t-1}^{*} \sum_{j=p_T+1}^{t-1} \widetilde{\boldsymbol{a}}_{j,T}' \widetilde{\boldsymbol{x}}_{t-j} \right\| &= O_p\left(\max\left(T^{\frac{1}{2}-2\kappa_2} \log^3 T, T^{-\kappa_2} \log T \right) \right) \\ &= O_p\left(T^{\frac{1}{2}-2\kappa_2} \log^3 T \right). \end{aligned}$$

as required.

Proof of item 2 The triangle inequality gives

$$\begin{aligned} \left\| \frac{1}{T} \sum_{t=p_T+1}^{T} \boldsymbol{x}_{t-1}^{*} \boldsymbol{z}_{t-1}' - \frac{1}{T} \sum_{t=p_T+1}^{T} \widetilde{\boldsymbol{x}}_{t-1}^{*} \widetilde{\boldsymbol{z}}_{t-1}' \right\| &\leq \left\| \frac{1}{T} \sum_{t=p_T+1}^{T} \boldsymbol{x}_{t-1}^{*} \left(\boldsymbol{z}_{t-1} - \widetilde{\boldsymbol{z}}_{t-1} \right)' \right\| \\ &\quad + \left\| \frac{1}{T} \sum_{t=p_T+1}^{T} \left(\boldsymbol{x}_{t-1}^{*} - \widetilde{\boldsymbol{x}}_{t-1}^{*} \right) \boldsymbol{z}_{t-1}' \right\|. \end{aligned} \tag{24}$$

For the first term we have that

$$\left\| \frac{1}{T} \sum_{t=p_T+1}^{T} \boldsymbol{x}_{t-1}^{*} \left(\boldsymbol{z}_{t-1} - \widetilde{\boldsymbol{z}}_{t-1} \right)' \right\| \leq \sqrt{ \frac{1}{T} \sum_{t=p_T+1}^{T} \left\| \boldsymbol{z}_{t-1} - \widetilde{\boldsymbol{z}}_{t-1} \right\|^2 \frac{1}{T} \sum_{t=p_T+1}^{T} \left\| \boldsymbol{x}_{t-1}^{*} \right\|^2 },$$

where $\frac{1}{T}\sum_{t=p_T+1}^{T}\|\boldsymbol{z}_{t-1}-\widetilde{\boldsymbol{z}}_{t-1}\|^2 = O_p\left(T^{\kappa_1-2\kappa_2}\ln^2 T\right)$ (see the proof of item 3 of Lemma 7) and $\frac{1}{T}\sum_{t=p_T+1}^{T}\left\|\boldsymbol{x}_{t-1}^*\right\|^2$ is $O_p\left(\ln^2 T\right)$. For the second term, the Cauchy-Schwarz inequality for matrix norms yields

$$\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\left(\boldsymbol{x}_{t-1}^*-\widetilde{\boldsymbol{x}}_{t-1}^*\right)\boldsymbol{z}_{t-1}'\right\| \leq \sqrt{\frac{1}{T}\sum_{t=p_T+1}^{T}\left\|\left(\boldsymbol{x}_{t-1}^*-\widetilde{\boldsymbol{x}}_{t-1}^*\right)\right\|^2\frac{1}{T}\sum_{t=p_T+1}^{T}\|\boldsymbol{z}_{t-1}\|^2}.$$

Taking into account that $\frac{1}{T}\sum_{t=p_T+1}^{T}\left\|\left(\boldsymbol{x}_{t-1}^*-\widetilde{\boldsymbol{x}}_{t-1}^*\right)\right\|^2 = O_p\left(T^{-2\kappa_2}\ln^4 T\right)$, and $\frac{1}{T}\sum_{t=p_T+1}^{T}\left\|\boldsymbol{z}_{t-1}'\right\|^2 = O_p\left(T^{\kappa_1}\right)$, we have that both terms on the r.h.s. of Equation (24) are $O_p(T^{\frac{\kappa_1}{2}-\kappa_2}\ln^2 T)$.

Proof of item 3 This is a multivariate extension of Lemma 7 in Demetrescu (2009).

Proof of item 4 Follows by noting that $x_{t-j,k}u_{t1}$ is a martingale difference sequence, uniformly L_2-bounded in t and j.

Proof of item 5 Using the triangle inequality we obtain that

$$\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{x}_{t-1}^*\boldsymbol{z}_{t-1}'\right\| \leq \sum_{j=1}^{p_T}\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{x}_{t-1}^*\boldsymbol{x}_{t-j}'\right\|. \tag{25}$$

Resorting then to item (i) of this Lemma, the result follows.

Proof of item 6 It follows from Lütkepohl (1996, Section 8.4.1, (11c)) that

$$\begin{aligned}&\left\|\left(\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right)^{-1}-\left(\frac{1}{T}\sum_{t=p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{z}}_{t-1}'\right)^{-1}\right\|\\ &\quad\leq\frac{\left\|\left(\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right)^{-1}\right\|^2\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'-\frac{1}{T}\sum_{t=p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{z}}_{t-1}'\right\|}{1-\left\|\left(\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right)^{-1}\right\|\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'-\frac{1}{T}\sum_{t=p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{z}}_{t-1}'\right\|}\end{aligned}$$

if

$$\left\|\left(\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right)^{-1}\right\|\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'-\frac{1}{T}\sum_{t=p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{z}}_{t-1}'\right\|<1$$

and

$$\left\|\left(\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right)^{-1}\left(\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'-\frac{1}{T}\sum_{t=p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{z}}_{t-1}'\right)\right\|<1.$$

The induced matrix norm $\|\cdot\|$ used here is submultiplicative, so the former condition implies the latter; moreover, $\left\|\left(\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right)^{-1}\right\|$ is bounded (see item 3 of this lemma) and we may thus conclude that $\left\|\left(\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right)^{-1}\right\|\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'-\frac{1}{T}\sum_{t=p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{z}}_{t-1}'\right\|<1$ with probability approaching 1 when $\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'-\frac{1}{T}\sum_{t=p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{z}}_{t-1}'\right\|\xrightarrow{p}0$. The result follows.

Proof of item 7 Direct consequence of items 3 and 6 of this Lemma.

Proof of item 8 We have that

$$\begin{aligned}&\left\|\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{u}_{t1}^{(p_T)}-\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\boldsymbol{z}_{t-1}u_{t1}\right\|\\&\leq\left\|\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\left(\widetilde{\boldsymbol{z}}_{t-1}-\boldsymbol{z}_{t-1}\right)u_{t1}\right\|+\left\|\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\sum_{j=p_T+1}^{t-1}\widetilde{\boldsymbol{a}}_j\widetilde{\boldsymbol{x}}_{t-j}\right\|.\end{aligned}\tag{26}$$

Analogously to the first part of the proof of item 1 of this Lemma, the first term on the r.h.s. of (26) bahaves as

$$\left\|\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\left(\boldsymbol{z}_{t-1}-\widetilde{\boldsymbol{z}}_{t-1}\right)u_{t1}\right\|=O_p\left(T^{\frac{\kappa_1}{2}-\kappa_2}\log T\right),$$

and for the second term it holds using Lemma 5 that

$$\left\|\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\sum_{j=p_T+1}^{t-1}\widetilde{\boldsymbol{a}}_{j,T}'\widetilde{\boldsymbol{x}}_{t-j}\right\|\leq C\sqrt{T}\left|\widehat{d}-d\right|T^{\left|\widehat{d}-d\right|}\sum_{j=p_T+1}^{T-1}\frac{1}{j}\left\|\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{x}}_{t-j}'\right\|.$$

Now, for all j,

$$\left\|\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{x}}_{t-j}'\right\|\leq\left\|\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{x}}_{t-j}'-\frac{1}{T}\sum_{t=j+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{x}_{t-j}'\right\|+\left\|\frac{1}{T}\sum_{t=j+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{x}_{t-j}'\right\|,$$

where with same arguments as in item 1 of Lemma 7

$$\left\|\frac{1}{T}\sum_{t=j+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{x}}_{t-j}'-\frac{1}{T}\sum_{t=j+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{x}_{t-j}'\right\|=O_p\left(T^{\frac{\kappa_1}{2}-\kappa_2}\log T\right),$$

and by same arguments as in item 2 of Lemma 7

$$\left\|\frac{1}{T}\sum_{t=j+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{x}_{t-j}'\right\|=O_p\left(T^{\frac{\kappa_1}{2}-\frac{1}{2}}\right)+O_p\left(\eta^{-j}\right).$$

Hence, the order of the second term on the r.h.s. of (26) is given as $O_p\left(T^{\frac{1}{2}-2\kappa_2+\frac{\kappa_1}{2}}\log^2 T\right)$, and the result follows.

Proof of item 9 Note that the difference can be rewritten as

$$\frac{1}{T}\sum_{t=p_T+1}^{T}\left(\boldsymbol{x}_{t-1}^*-\widetilde{\boldsymbol{x}}_{t-1}^*\right)\boldsymbol{x}_{t-1}^{*\prime}+\frac{1}{T}\sum_{t=p_T+1}^{T}\left(\boldsymbol{x}_{t-1}^*-\widetilde{\boldsymbol{x}}_{t-1}^*\right)\left(\boldsymbol{x}_{t-1}^*-\widetilde{\boldsymbol{x}}_{t-1}^*\right)'+\sum_{t=p_T+1}^{T}\boldsymbol{x}_{t-1}^*\left(\boldsymbol{x}_{t-1}^*-\widetilde{\boldsymbol{x}}_{t-1}^*\right)',\tag{27}$$

where the first and the third terms have identical asymptotic behavior. In particular, applying the Cauchy-Schwarz inequality yields that

$$\left\|\frac{1}{T}\sum_{t=p_T+1}^{T}\left(\boldsymbol{x}_{t-1}^*-\widetilde{\boldsymbol{x}}_{t-1}^*\right)\boldsymbol{x}_{t-1}^{*\prime}\right\|\leq\sqrt{\frac{1}{T}\sum_{t=p_T+1}^{T}\left\|\boldsymbol{x}_{t-1}^*-\widetilde{\boldsymbol{x}}_{t-1}^*\right\|^2\frac{1}{T}\sum_{t=p_T+1}^{T}\left\|\boldsymbol{x}_{t-1}^*\right\|^2}.$$

By Lemma 6 and item 1 of Lemma 7,

$$\frac{1}{T}\sum_{t=p_T+1}^{T}\left\|\boldsymbol{x}_{t-1}^*-\widetilde{\boldsymbol{x}}_{t-1}^*\right\|^2=C\left|\widehat{d}-d\right|^2T^{2\left|\widehat{d}-d\right|}\frac{1}{T}\sum_{t=p_T+1}^{T}\left(v_{t-1,T}^*\right)^2=O_p\left(T^{-2\kappa_2}\log^4 T\right).\tag{28}$$

Moreover, due the logarithmic behaviour of the harmonic series and the uniform L_2 boundness of $\boldsymbol{x}_t$,

$$\mathrm{E}\left(\left\|\boldsymbol{x}^*_{t-1}\right\|^2\right) \leq \mathrm{E}\left(\left\|\sum_{j=1}^{t-1} \frac{\boldsymbol{x}_{t-j}}{j}\right\|^2\right) \leq \mathrm{E}\left(\left\|\boldsymbol{x}_{t-j}\right\|^2\right)\left(\sum_{j=1}^{t-1} \frac{1}{j}\right)^2 \leq C \log^2 T. \tag{29}$$

Putting together (28) and (29) gives that

$$\left\|\frac{1}{T} \sum_{t=p_T+1}^{T} \left(\boldsymbol{x}^*_{t-1} - \widetilde{\boldsymbol{x}}^*_{t-1}\right) \boldsymbol{x}^{*\prime}_{t-1}\right\| = O_p\left(T^{-\kappa_2} \log^3 T\right).$$

The second summand of Equation (27) has the same magnitude order, which is easily checked using similar arguments.

B Proofs of the main results

All sums run over $p+1,\ldots,T$, unless specified otherwise.

Proof of Proposition 1

Since $\boldsymbol{u}_t$ is a linear process with *iid* innovations and absolutely summable coefficients, the properties of $\boldsymbol{x}_t$ hinge on the error-correction model (1). It follows e.g. from Lemma 1 in Avarucci and Velasco (2009) that, for any $0 \leq b < 0.5$, $\boldsymbol{x}_t$ is a (truncated) linear process with square summable coefficients and as such asymptotically stationary. This applies for the local alternatives studied here as well. Let then $\boldsymbol{x}_{t-1}^{**} = \sum_{j=1}^{\infty} \frac{\boldsymbol{x}_{t-j}}{j}$ and $\boldsymbol{u}_{t-1}^{**} = \sum_{j=1}^{\infty} \frac{\boldsymbol{u}_{t-j}}{j}$; using $\boldsymbol{x}_{t-1}^{**}$ instead of $\boldsymbol{x}_{t-1}^{*}$ does not affect the analysis; the equivalence is established here in a very similar way to Lemma 2 in Demetrescu et al. (2008) and relies on the fact that the Wold coefficients of $\boldsymbol{x}_t$ vanish fast enough; we do not give a proof to save space. Hence, the initial test is asymptotically equivalent to

$$x_{t1} = \widehat{\alpha}_1^{*} x_{t-1,1}^{**} + \widehat{\boldsymbol{\delta}}' \boldsymbol{z}_{t-1}^{**} + \widehat{\boldsymbol{a}}_1' \boldsymbol{x}_{t-1} + \ldots + \widehat{\boldsymbol{a}}_p' \boldsymbol{x}_{t-p} + \widehat{u}_{t1},$$

Recall that, under the local alternative, we have

$$\Delta^{-b} - 1 = \frac{c}{\sqrt{T}} \left(L + \frac{L^2}{2} + \frac{L^3}{3} + \ldots \right) + O\left(T^{-1}\right);$$

furthermore, $A(L)\boldsymbol{x}_{t-1}^{**} = \frac{c}{\sqrt{T}} \left(\boldsymbol{\alpha\beta}' A(L) \boldsymbol{x}_{t-2}^{**}\right)^{**} + \boldsymbol{u}_{t-1}^{**} = \boldsymbol{u}_{t-1}^{**} + O_p(T^{-0.5})$ such that

$$\begin{aligned}
\boldsymbol{x}_t &= \frac{c}{\sqrt{T}} \boldsymbol{\alpha\beta}' A(L) \boldsymbol{x}_{t-1}^{**} + A_1 \boldsymbol{x}_{t-1} + \ldots + A_p \boldsymbol{x}_{t-p} + \boldsymbol{u}_t + O_p(T^{-1}) \\
&= \frac{c}{\sqrt{T}} \boldsymbol{\alpha\beta}' \boldsymbol{u}_{t-1}^{**} + A_1 \boldsymbol{x}_{t-1} + \ldots + A_p \boldsymbol{x}_{t-p} + \boldsymbol{u}_t + O_p(T^{-1}),
\end{aligned}$$

with an $O\left(T^{-1}\right)$ term that can be safely ignored.

Denote by $\boldsymbol{w}_t$ the vector of regressors, $\boldsymbol{w}_t = \left(\left(\boldsymbol{u}_{t-1}^{**}\right)', \boldsymbol{x}_{t-1}', \ldots, \boldsymbol{x}_{t-p}'\right)'$, and let also $\boldsymbol{w}_t^0 = \left(\left(\boldsymbol{u}_{t-1}^{**}\right)', \boldsymbol{x}_{t-1}', \ldots, \boldsymbol{x}_{t-p}'\right)'$. Note that $\mathrm{E}\left(\boldsymbol{w}_t\right) = 0$ and let $\Sigma_{\boldsymbol{w}} = \mathrm{Cov}\left(\boldsymbol{w}_t\right)$. The processes $\boldsymbol{w}_t$ and $\boldsymbol{w}_t^0$ are is strictly stationary and ergodic due to square summability of the involved filters. Let $\boldsymbol{\gamma} = \left(\boldsymbol{0}', \boldsymbol{a}_1', \ldots, \boldsymbol{a}_p'\right)'$, be the true values of the parameters in test Equation (4). One has as OLS estimators

$$\widehat{\boldsymbol{\gamma}} = \left(\frac{1}{T} \sum \boldsymbol{w}_t \boldsymbol{w}_t'\right)^{-1} \left(\frac{1}{T} \sum \boldsymbol{w}_t x_{t1}\right).$$

Since $\boldsymbol{w}_t$ is strictly stationary and ergodic with zero mean, it holds that

$$\frac{1}{T} \sum \boldsymbol{w}_t \boldsymbol{w}_t' \stackrel{p}{\to} \Sigma_{\boldsymbol{w}},$$

and, due to the obvious nonsingularity of $\Sigma_{\boldsymbol{w}}$, the convergence also holds for the inverse matrices.

Given the zero-mean and the *iid* property of $\boldsymbol{u}_t$, the sequence $\boldsymbol{w}_{t-1}\boldsymbol{u}_t'$ has the martingale difference property; it also holds

$$\mathrm{Cov}\left(\boldsymbol{w}_t u_{t1}\right) = \mathrm{E}\left(\boldsymbol{w}_t u_{t1}^2 \boldsymbol{w}_t'\right) = \sigma_{11} \Sigma_{\boldsymbol{w}},$$

which is again nonsingular. Since, up to an $O_p(T^{-1})$ term,

$$x_{t1} = (\boldsymbol{w}_t^0)' \boldsymbol{\gamma} + \frac{c\alpha_1}{\sqrt{T}} \boldsymbol{\beta}' \boldsymbol{u}_{t-1}^{**} + u_{t1},$$

we have that

$$\frac{1}{T} \sum \boldsymbol{w}_t x_{t1} = \frac{1}{T} \sum \boldsymbol{w}_t (\boldsymbol{w}_t^0)' \boldsymbol{\gamma} + \frac{c\alpha_1}{\sqrt{T}} \frac{1}{T} \sum \boldsymbol{w}_t \boldsymbol{\beta}' \boldsymbol{u}_{t-1}^{**} + \frac{1}{T} \sum \boldsymbol{w}_t u_{t1}.$$

Moreover,

$$\frac{1}{T}\sum \boldsymbol{w}_t(\boldsymbol{w}_t^0)'\boldsymbol{\gamma} = \frac{1}{T}\sum \boldsymbol{w}_t\boldsymbol{w}_t'\boldsymbol{\gamma} + \frac{1}{T}\sum \boldsymbol{w}_t(\boldsymbol{w}_t^0 - \boldsymbol{w}_t)'\boldsymbol{\gamma}$$

where the latter average is by zero construction, given that $\boldsymbol{w}_t$ differs from $\boldsymbol{w}_t^0$ only for the first K elements (which are zero for $\boldsymbol{\gamma}$). Note that the orthogonality is only given for local alternatives, since, with $b > 0$ fixed, the probability limit of $\widehat{\boldsymbol{\gamma}}$ may be nonzero. This implies that the unconditional test may have power even if $\alpha_1 = 0$.

Thus, with $\boldsymbol{w}_t\boldsymbol{\beta}'\boldsymbol{u}_{t-1}^{**} = O_p(1)$, and thus $c\alpha_1 T^{-1.5}\sum \boldsymbol{w}_t\boldsymbol{\beta}'\boldsymbol{u}_{t-1}^{**} = o_p(1)$, consistency of $\widehat{\boldsymbol{\gamma}}$ follows due to a Law of Large Numbers for finite-variance martingale differences. This implies consistent estimation of the residual variance $\widehat{\sigma}_{11}$. Further, it holds that

$$\sqrt{T}\left(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) = \left(\frac{1}{T}\sum \boldsymbol{w}_t\boldsymbol{w}_t'\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum \boldsymbol{w}_t u_{t1} + \frac{c\alpha_1}{T}\sum \boldsymbol{w}_t\boldsymbol{\beta}'\boldsymbol{u}_{t-1}^{**}\right).$$

and, thanks to a suitable multivariate martingale central limit theorem (White, 2001, Theorem 5.25), we obtain that

$$\frac{1}{\sqrt{T}}\sum \boldsymbol{w}_t u_{t1} \xrightarrow{d} N\left(0, \sigma_{11}\Sigma_{\boldsymbol{w}}\right).$$

Then,

$$\frac{c\alpha_1}{T}\sum \boldsymbol{w}_t\boldsymbol{\beta}'\boldsymbol{u}_{t-1}^{**} \xrightarrow{p} c\alpha_1 \operatorname{E}\left(\boldsymbol{w}_t\boldsymbol{\beta}'\boldsymbol{u}_{t-1}^{**}\right),$$

so

$$\sqrt{T}\left(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) \xrightarrow{d} \mathcal{N}\left(c\,\alpha_1\Sigma_{\boldsymbol{w}}^{-1}\operatorname{E}\left(\boldsymbol{w}_t\boldsymbol{\beta}'\boldsymbol{u}_{t-1}^{**}\right), \sigma_{11}\Sigma_{\boldsymbol{w}}^{-1}\right).$$

Dividing $\widehat{\gamma}_1 - 0$ (i.e. $\widehat{\alpha}_1^* - 0$) by $\frac{1}{\sqrt{T}}\sqrt{\widehat{\sigma}_{11}\widehat{\Sigma}_{\boldsymbol{w}[1,1]}^{-1}}$ and letting

$$s = \operatorname{plim} c\,\alpha_1 \cdot \frac{\left(\Sigma_{\boldsymbol{w}}^{-1}\operatorname{E}\left(\boldsymbol{w}_t\boldsymbol{\beta}'\boldsymbol{u}_{t-1}^{**}\right)\right)_{[1,1]}}{\sqrt{\widehat{\sigma}_{11}\widehat{\Sigma}_{\boldsymbol{w}[1,1]}^{-1}}}$$

leads to

$$t_{\alpha_1^*} \xrightarrow{d} \mathcal{N}\left(s, 1\right).$$

To establish the desired result, note that $\boldsymbol{\beta}'\boldsymbol{u}_{t-1}^{**}$ is a scalar and thus $\boldsymbol{\beta}'\boldsymbol{u}_{t-1}^{**} = \left(\boldsymbol{u}_{t-1}^{**}\right)'\boldsymbol{\beta}$; then examine

$$c\,\alpha_1 \cdot \frac{\left(\Sigma_{\boldsymbol{w}}^{-1}E\left(\boldsymbol{w}_t\boldsymbol{\beta}'\boldsymbol{u}_{t-1}^{**}\right)\right)_{[1,1]}}{\sqrt{\widehat{\sigma}_{11}\widehat{\Sigma}_{\boldsymbol{w}[1,1]}^{-1}}} = \frac{c\,\alpha_1}{\sqrt{\widehat{\sigma}_{11}\widehat{\Sigma}_{\boldsymbol{w}[1,1]}^{-1}}} \cdot \left(\Sigma_{\boldsymbol{w}}^{-1}\operatorname{E}\left(\boldsymbol{w}_t\left(\boldsymbol{u}_{t-1}^{**}\right)'\right)\boldsymbol{\beta}\right)_{[1,1]}.$$

Letting $q = \Sigma_{\boldsymbol{w}}^{-1}\operatorname{E}\left(\boldsymbol{w}_t\left(\boldsymbol{u}_{t-1}^{**}\right)'\right)\boldsymbol{\beta}$, the result follows with $\sigma_*^2 = \operatorname{plim}\widehat{\sigma}_{11}\widehat{\Sigma}_{\boldsymbol{w}[1,1]}^{-1}$ which obviously exists; cf. Demetrescu et al. (2008, Proposition 3).

Proof of Proposition 2

Redefine the vector of regressors, $\boldsymbol{w}_t = \left(\left(\boldsymbol{u}_{t-1}^{**}\right)', \boldsymbol{z}_t', \boldsymbol{x}_{t-1}', \ldots, \boldsymbol{x}_{t-p}'\right)'$, corresponding to the regression with contemporaneous differences, and let $\underline{\widehat{\boldsymbol{\gamma}}}$ denote the expanded vector of parameter estimators. Like in the proof of Proposition 1, $\boldsymbol{w}_t$ is an ergodic and stationary process with mean zero, and $\underline{\widehat{\boldsymbol{\gamma}}}$ will have a proper probability limit. The limit will not exhibit the true values studied there, but rather

$$\underline{\boldsymbol{\gamma}} = \operatorname{plim}\underline{\widehat{\boldsymbol{\gamma}}} = \Sigma_{\boldsymbol{w}}^{-1}\operatorname{E}\left(\boldsymbol{w}_t x_{t1}\right).$$

This only equals $\left(\mathbf{0}', \mathbf{0}', \boldsymbol{a}_1', \ldots, \boldsymbol{a}_p'\right)'$ if $\boldsymbol{u}_t$ is not contemporaneously correlated (i.e. no regressor endogeneity in the test equation). However, we do not need $\underline{\widehat{\boldsymbol{\gamma}}}$ to converge to $\left(\mathbf{0}', \mathbf{0}', \boldsymbol{a}_1', \ldots, \boldsymbol{a}_p'\right)'$: the analysis can proceed considering as 'true' values the probability limit of $\widehat{\underline{\gamma}}$. Let thus $\underline{\gamma} = \text{plim}\,\widehat{\underline{\gamma}} = \left(\mathbf{0}', \underline{\boldsymbol{a}}_0', \underline{\boldsymbol{a}}_1', \ldots, \underline{\boldsymbol{a}}_p'\right)'$.

To make use, as before, of a central limit theorem for martingale differences, we need to orthogonalize $\boldsymbol{u}_t$ such that the resulting innovation $\underline{u}_{t1}$ is orthogonal to the regressors.

This can be accomplished by employing a suitable decomposition of Σ. Let C be a $K \times K$ matrix such that its first row is $(1, \underline{\boldsymbol{a}}_0')$ and $C\Sigma C'$ has diagonal form (but is not necessarily the identity matrix). The matrix C thus has $K^2 - K$ free parameters, whereas the condition that the matrix $C\Sigma C'$ (which is symmetric by construction) has diagonal form only imposes $K(K-1)/2$ restrictions; thus, C exists and is not even unique.

So let $\boldsymbol{v}_t = C\boldsymbol{u}_t$ and $\underline{\sigma}_{11} = \text{Var}\,(v_{t1})$. Pre-multiply the vector autoregression equation with C and add on both sides $(I - C)\,\boldsymbol{x}_t$ to arrive at the equivalent DGP

$$\boldsymbol{x}_t = \frac{c}{\sqrt{T}} C\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{u}_{t-1}^{**} + (I - C)\,\boldsymbol{x}_t + CA_1\boldsymbol{x}_{t-1} + \ldots + CA_p\boldsymbol{x}_{t-p} + \boldsymbol{v}_t,$$

where $\boldsymbol{v}_t$ is a *iid* sequence of contemporarily uncorrelated variables. (For this reason, v_{t1} is now orthogonal to $\boldsymbol{z}_t$ only containing v_{t2} through v_{tK} contemporaneously.) So we work now with

$$x_{t1} = \boldsymbol{w}_t'\underline{\gamma} + \frac{c}{\sqrt{T}}\left(C\boldsymbol{\alpha}\boldsymbol{\beta}'\right)_{[1,.]}\boldsymbol{x}_{t-1}^{**} + v_{t1},$$

and the sequence $\boldsymbol{w}_t v_{t1}$ is a martingale difference sequence with variance

$$\text{E}\left(\boldsymbol{w}_t v_{t1}^2 \boldsymbol{w}_t'\right) = \underline{\sigma}_{11}\Sigma_{\boldsymbol{w}},$$

with $\underline{\sigma}_{11} = \text{Var}\,(v_{t1})$. (The simplifying assumption that v_{t1} is conditionally homoskedastic was made.[9])

The vector of adjustment speed coefficients for this DGP is $\underline{\boldsymbol{\alpha}} = C\boldsymbol{\alpha} = (1, \underline{\boldsymbol{a}}_0)'\,\boldsymbol{\alpha}$; it does not depend on the particular decomposition of Σ. The proof continues along the lines of the proof of Proposition 1, one just replaces α_1 with $\underline{\alpha}_1 = \alpha_1 + \sum_{j=2}^{K}\alpha_j\underline{a}_{0j}$. The quantity $\underline{q}$ is defined analogously.

Proof of Proposition 3

The tr_* trace test has the following shape

$$\text{tr}_* = T\,\text{tr}\left(\widetilde{S}_{00}^{-1}\widetilde{S}_{10}'\widetilde{S}_{11}^{-1}\widetilde{S}_{10}\right), \tag{30}$$

where $\widetilde{S}_{00} = \frac{1}{T}\sum_{t=p+1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t'$, $\widetilde{S}_{11} = \frac{1}{T}\sum_{t=p+1}^{T}\boldsymbol{x}_{t-1}^*(\boldsymbol{x}_{t-1}^*)'$, and $\widetilde{S}_{10} = \frac{1}{T}\sum_{t=p+1}^{T}\boldsymbol{x}_{t-1}^*\boldsymbol{x}_t'$. Further, the test regression under the local alternatives takes the following shape

$$\boldsymbol{x}_t = \frac{c}{\sqrt{T}}\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}_{t-1}^* + \boldsymbol{u}_t. \tag{31}$$

Note that the system tr_* test given in (30) can be rewritten as

$$\text{tr}_* = T\,\text{vec}\left(\widetilde{S}_{10}\widetilde{S}_{00}^{-1/2}\right)'\left[\text{I}_K \otimes \widetilde{S}_{11}^{-1}\right]\text{vec}\left(\widetilde{S}_{10}\widetilde{S}_{00}^{-1/2}\right), \tag{32}$$

where $\text{vec}(\cdot)$ denote the standard vectorization transformation of the matrix and $\otimes$ stands for the Kronecker product. Further decomposition of the $\text{vec}\left(\widetilde{S}_{10}\widetilde{S}_{00}^{-1/2}\right)$ is convenient at this point to derive the asymptotic

[9] This implies that White standard errors may be necessary for the conditional regression even if $\boldsymbol{u}_t$ is conditionally homoskedastic.

behaviour of the tr_* test, i.e.,

$$\begin{aligned}\sqrt{T}\,\mathrm{vec}\left(\widetilde{S}_{10}\widetilde{S}_{00}^{-1/2}\right) &= \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\mathrm{vec}\left(\boldsymbol{x}_t(\boldsymbol{x}^*_{t-1})'\widetilde{S}_{00}^{-1/2}\right)\\ &= \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\mathrm{vec}\left(\boldsymbol{u}_t(\boldsymbol{x}^*_{t-1})'\widetilde{S}_{00}^{-1/2}\right)\\ &+ \frac{c}{T}\sum_{t=2}^{T}\mathrm{vec}\left(\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}^*_{t-1}(\boldsymbol{x}^*_{t-1})'\widetilde{S}_{00}^{-1/2}\right). \end{aligned} \quad (33)$$

Note that

$$\begin{aligned} &\text{(i)} \quad \widetilde{S}_{00} \xrightarrow{p} S_x\,\mathrm{E}\left(\boldsymbol{x}_t\boldsymbol{x}_t'\right);\\ &\text{(ii)} \quad \widetilde{S}_{11} \xrightarrow{p} S_* \equiv \lim_{t\to\infty}\mathrm{E}\left(\boldsymbol{x}^*_{t-1}(\boldsymbol{x}^*_{t-1})'\right);\\ &\text{(iii)} \quad \widetilde{\psi}_t \equiv \mathrm{vec}\left(\boldsymbol{u}_t(\boldsymbol{x}^*_{t-1})'\widetilde{S}_{00}^{-1/2}\right) \text{ is a martingale difference sequence.}\end{aligned}$$

Therefore, the central limit theorem for mds (see, e.g., Corollary 5.26 in White, 2001) can be applied to the first term on the r.h.s of (33), i.e.,

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\widetilde{\psi}_t \xrightarrow{d} N\left(0, \mathrm{I}_K\otimes S_*\right). \quad (34)$$

With the same arguments LLN for mds applies to the second term on the r.h.s of (33), i.e.,

$$\frac{c}{T}\sum_{t=2}^{T}\mathrm{vec}\left(\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}^*_{t-1}(\boldsymbol{x}^*_{t-1})'\widetilde{S}_{00}^{-1/2}\right) \xrightarrow{p} c\,\mathrm{vec}\left(\boldsymbol{\alpha}\boldsymbol{\beta}'S_*S_x^{-1/2}\right), \quad (35)$$

Recall that $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$ then it follows that tr_* has a χ^2 limiting distribution with K^2 degrees of freedom and noncentrality parameter μ given as

$$\begin{aligned}\mu &= c^2\,\mathrm{vec}\left(\Pi S_* S_x^{-1/2}\right)'\left[\mathrm{I}_K\otimes S_*^{-1}\right]\mathrm{vec}\left(\Pi S_* S_x^{-1/2}\right)\\ &= c^2\,\mathrm{tr}\left(S_x^{-1}S_*\Pi' S_*^{-1}\Pi S_*\right).\end{aligned} \quad (36)$$

Proof of Proposition 4

Note that, following Demetrescu et al. (2008), the t statistic based on the true $\boldsymbol{x}_t$s is asymptotically standard normally distributed. (This is a straightforward multivariate extension of their Proposition 2).

Let now $\widehat{\cdot}$ denote infeasible OLS estimates based on the true $\boldsymbol{x}_t$s and $\widetilde{\cdot}$ the OLS estimates based on the feasible $\widetilde{\boldsymbol{x}}_t = \Delta_+^{\widehat{d}}\boldsymbol{y}_t$ instead of $\boldsymbol{x}_t = \Delta_+^{d}\boldsymbol{y}_t$. Partition the parameter vector $\boldsymbol{\gamma}$ as $\boldsymbol{\gamma} = \left(\alpha_1, \boldsymbol{\delta}_1'\right)'$; what we need to show for the propositions to continue to hold if working with an estimate $\widehat{d}$ of d is that

$$\sqrt{T}\left(\widehat{\gamma} - \widetilde{\gamma}\right) \xrightarrow{p} 0$$

and

$$\sqrt{T}\left(s.e.\left(\widehat{\alpha}_1\right) - s.e.\left(\widetilde{\alpha}_1\right)\right) \xrightarrow{p} 0.$$

With $\boldsymbol{z}_{t-1}$ the stacked lags of $\boldsymbol{x}_t$ (of dimension $K\cdot p_T$), $\widetilde{\boldsymbol{z}}_{t-1}$ the corresponding stacked lags of $\widetilde{\boldsymbol{x}}_t$, and $\widetilde{u}_{t1}{}^{(p_T)} = u_{t1} + \sum_{j=p_T+1}^{t-1}\widetilde{\boldsymbol{a}}_{j,T}'\widetilde{\boldsymbol{x}}_{t-j}$, this amounts to showing that the differences

1. between

$$\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\boldsymbol{x}^*_{t-1}u_{t1} - \frac{1}{T}\sum_{p_T+1}^{T}\boldsymbol{x}^*_{t-1}\boldsymbol{z}_{t-1}'\left(\frac{1}{T}\sum_{p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\boldsymbol{z}_{t-1}u_{t1}$$

and

$$\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{x}}_{t-1}^{*}\widetilde{u}_{t1}^{(p_T)}-\frac{1}{T}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{x}}_{t-1}^{*}\widetilde{\boldsymbol{z}}_{t-1}'\left(\frac{1}{T}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{z}}_{t-1}'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{u}_{t1}^{(p_T)},$$

2. and between

$$\frac{1}{T}\sum_{p_T+1}^{T}\boldsymbol{x}_{t-1}^{*}\left(\boldsymbol{x}_{t-1}^{*}\right)'-\frac{1}{T}\sum_{p_T+1}^{T}\boldsymbol{x}_{t-1}^{*}\boldsymbol{z}_{t-1}'\left(\frac{1}{T}\sum_{p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right)^{-1}\frac{1}{T}\sum_{p_T+1}^{T}\boldsymbol{z}_{t-1}\left(\boldsymbol{x}_{t-1}^{*}\right)'$$

and

$$\frac{1}{T}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{x}}_{t-1}^{*}\left(\widetilde{\boldsymbol{x}}_{t-1}^{*}\right)'-\frac{1}{T}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{x}}_{t-1}^{*}\widetilde{\boldsymbol{z}}_{t-1}'\left(\frac{1}{T}\sum_{p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right)^{-1}\frac{1}{T}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\left(\widetilde{\boldsymbol{x}}_{t-1}^{*}\right)',$$

vanish.

It is now clear why the augmentation with $p_T\to\infty$ lags is required: without it, $\widetilde{u}_{t1}{}^{(p_T)}$ would not be asymptotically iid fast enough, even for $\widehat{d}-d=O_p\left(T^{-1/2}\right)$.

Now, since $\kappa_2>1/4$, Lemma 8 item 1 indicates that it suffices to show that

$$\frac{1}{T}\sum_{p_T+1}^{T}\boldsymbol{x}_{t-1}^{*}\boldsymbol{z}_{t-1}'\left(\frac{1}{T}\sum_{p_T+1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right)\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\boldsymbol{z}_{t-1}u_{t1}$$
$$-\frac{1}{T}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{x}}_{t-1}^{*}\widetilde{\boldsymbol{z}}_{t-1}'\left(\frac{1}{T}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{\boldsymbol{z}}_{t-1}'\right)\frac{1}{\sqrt{T}}\sum_{p_T+1}^{T}\widetilde{\boldsymbol{z}}_{t-1}\widetilde{u}_{t1}^{(p_T)}\stackrel{p}{\to}0,$$

or, with obvious notation, that

$$AB^{-1}C-\widetilde{A}\widetilde{B}^{-1}\widetilde{C}\stackrel{p}{\to}0.$$

We have furthermore that

$$\begin{aligned}\widetilde{A}\widetilde{B}^{-1}\widetilde{C} &= \widetilde{A}\widetilde{B}^{-1}C+\widetilde{A}\widetilde{B}^{-1}\left(\widetilde{C}-C\right)\\ &= \widetilde{A}B^{-1}C+\widetilde{A}\left(\widetilde{B}^{-1}-B^{-1}\right)C+\widetilde{A}\widetilde{B}^{-1}\left(\widetilde{C}-C\right)\\ &= AB^{-1}C+\left(\widetilde{A}-A\right)B^{-1}C+\widetilde{A}\left(\widetilde{B}^{-1}-B^{-1}\right)C+\widetilde{A}\widetilde{B}^{-1}\left(\widetilde{C}-C\right)\end{aligned}$$

so this amounts to showing that

$$\left\|\widetilde{A}-A\right\|\left\|B^{-1}\right\|\left\|C\right\|+\left\|\widetilde{A}\right\|\left\|\widetilde{B}^{-1}-B^{-1}\right\|\left\|C\right\|+\left\|\widetilde{A}\right\|\left\|\widetilde{B}^{-1}\right\|\left\|\widetilde{C}-C\right\|\stackrel{p}{\to}0$$

where $\|\cdot\|$ stands for the Euclidean vector norm and the corresponding induced matrix norm. All three summands above vanish as $T\to\infty$, as implied by Lemma 8 items 2, 3 and 4, 5, 6 and 4, and 5, 7 and 8.

The same reasoning shows that Lemma 8 item 9, together with items 2, 3 and 5, 5, 6 and 2, and 5, 7 and 2, implies the second difference to vanish, thus concluding the proof.

C Appendix: Figures

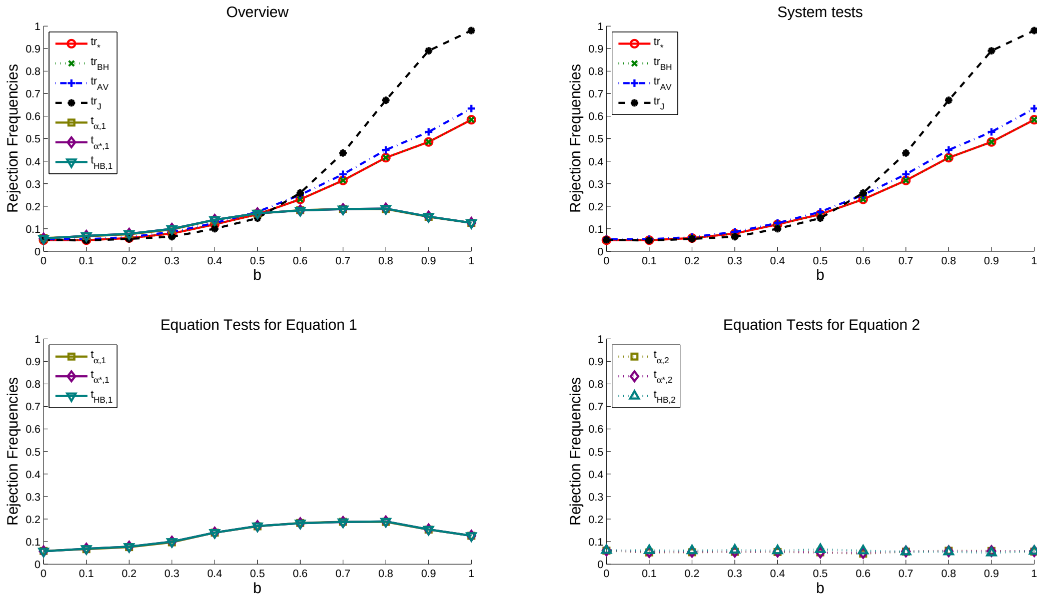

Figure 1: Size/power performance of the different test statistics. Sample size $T = 200$, $d = 1$ and is known, no short run dynamics (i.e., $A = 0$), $\rho = 0$ and number of augmentations p_{max} is set to zero.

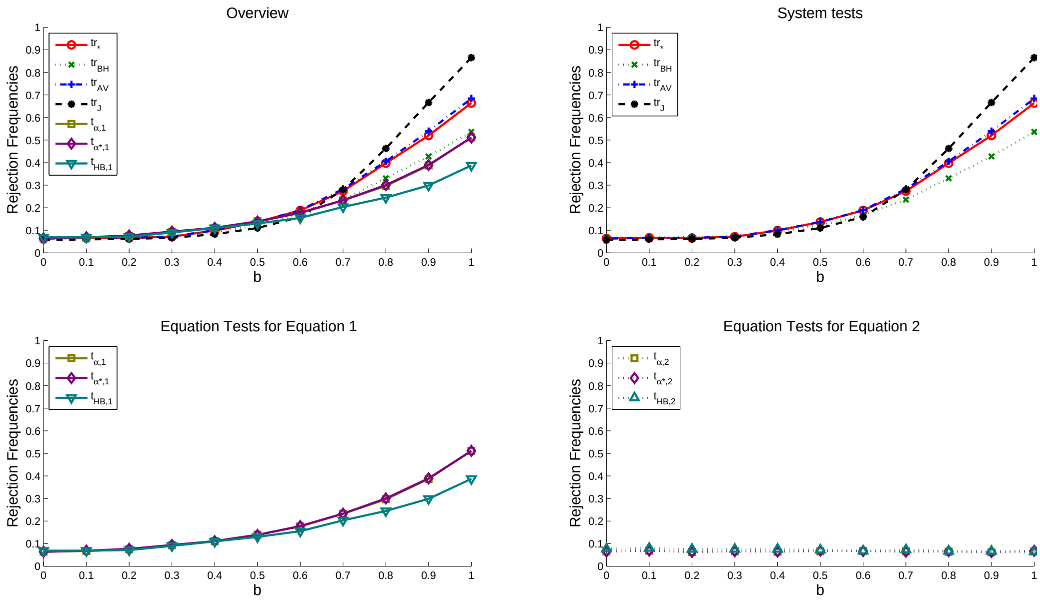

Figure 2: Size/power performance of the different test statistics. Sample size $T = 200$, $d = 1$ and is known, no short run dynamics (i.e., $A = 0$), $\rho = 0$ and number of augmentations p_{max} is set to $4(T/100)^{0.25}$.

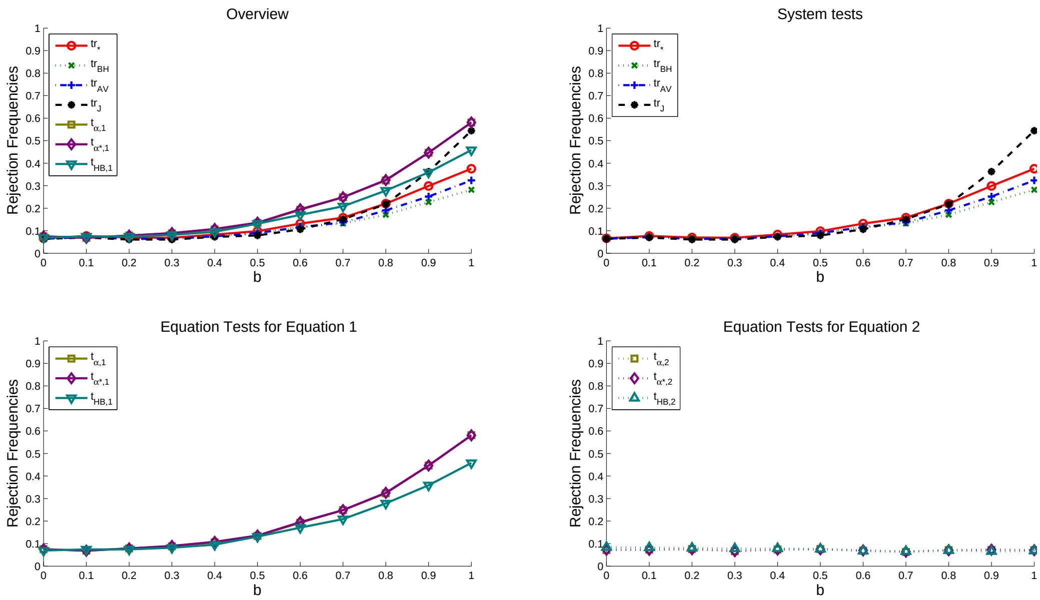

Figure 3: Size/power performance of the different test statistics. Sample size $T = 200$, $d = 1$ and is known, with short run dynamics (i.e., $A \neq 0$), $\rho = 0$ and number of augmentations p_{max} is set to $4(T/100)^{0.25}$.

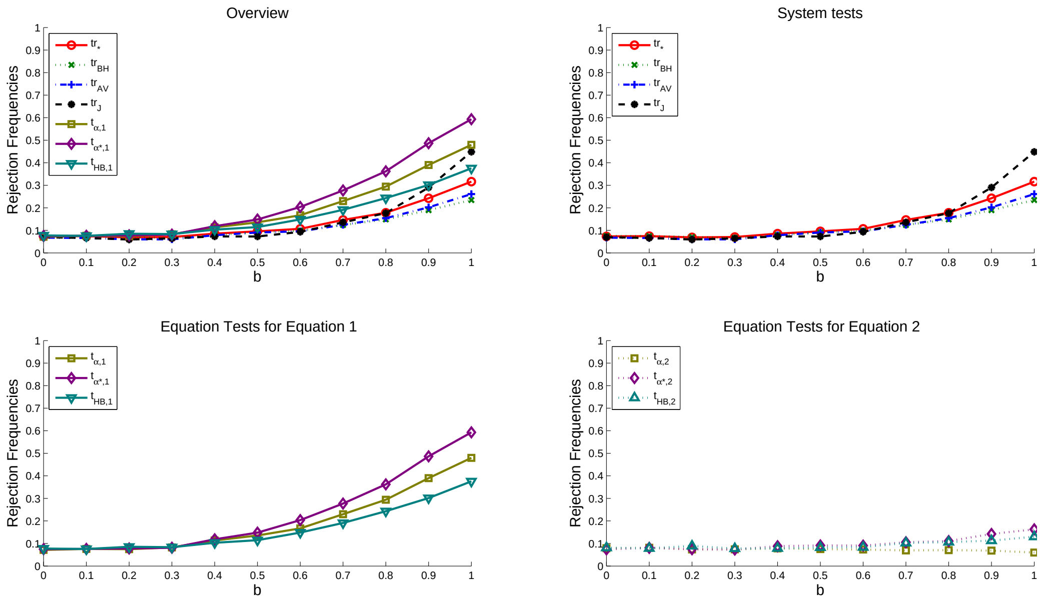

Figure 4: Size/power performance of the different test statistics. Sample size $T = 200$, $d = 1$ and is known, with short run dynamics (i.e., $A \neq 0$), $\rho = 0.5$ and number of augmentations p_{max} is set to $4(T/100)^{0.25}$.

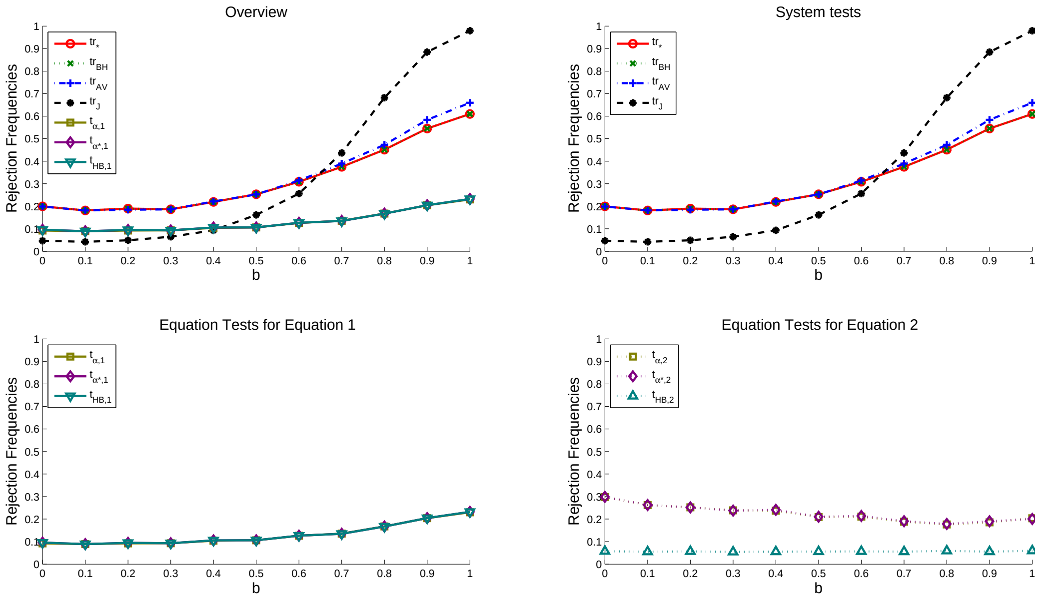

Figure 5: Size/power performance of the different test statistics. Sample size $T = 200$, $d = 1$ and is estimated with ELW estimator, no short run dynamics (i.e., $A = 0$), $\rho = 0$ and number of augmentations p_{max} is set to zero.

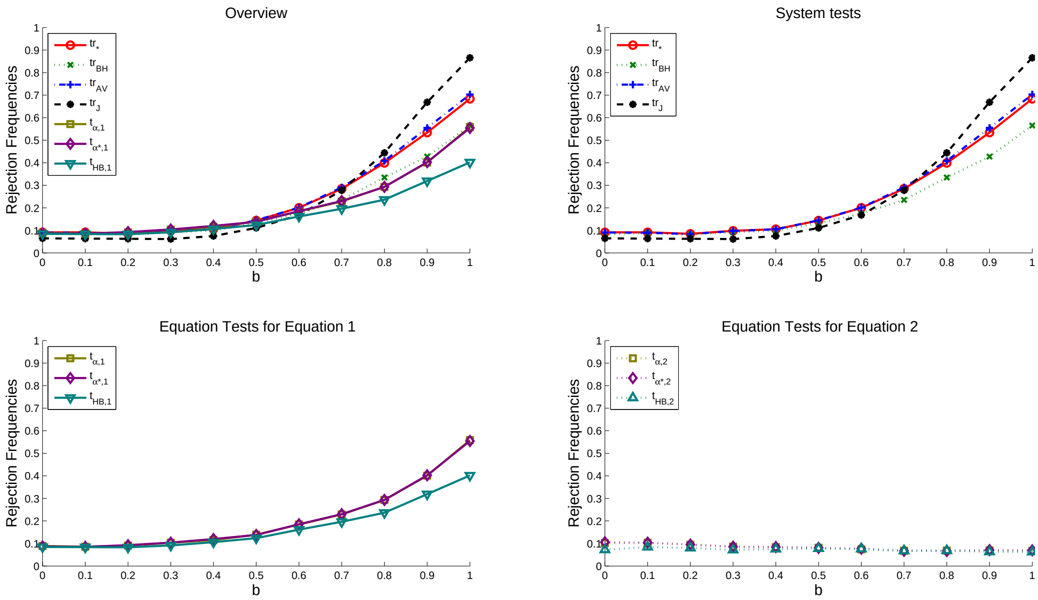

Figure 6: Size/power performance of the different test statistics. Sample size $T = 200$, $d = 1$ and is estimated with ELW estimator, no short run dynamics (i.e., $A = 0$), $\rho = 0$ and number of augmentations p_{max} is set to $4(T/100)^{0.25}$.

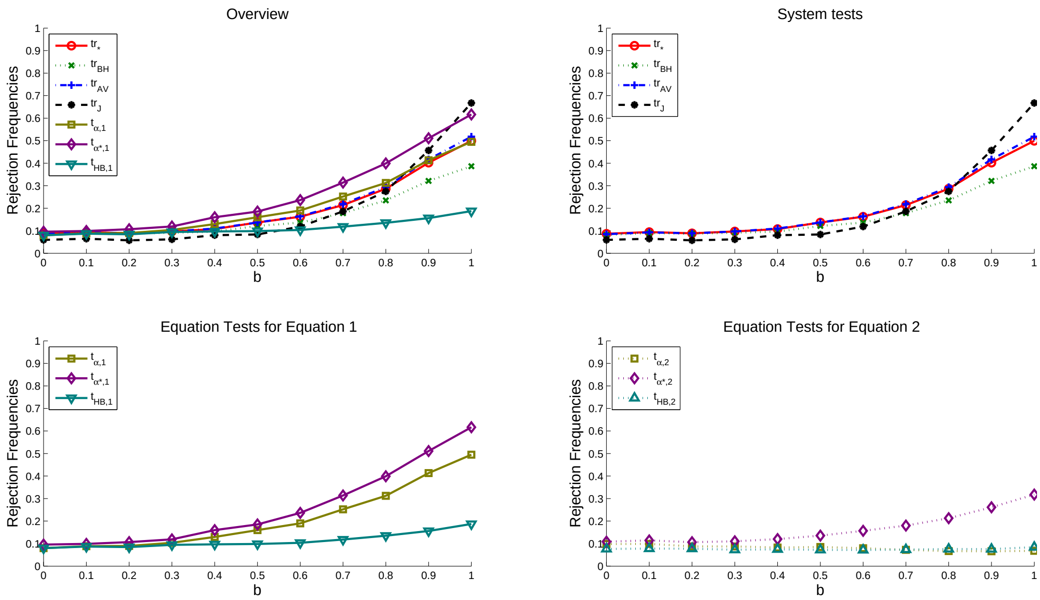

Figure 7: Size/power performance of the different test statistics. Sample size $T = 200$, $d = 1$ and is estimated with ELW estimator, no short run dynamics (i.e., $A = 0$), $\rho = 0.5$ and number of augmentations p_{max} is set to $4(T/100)^{0.25}$.

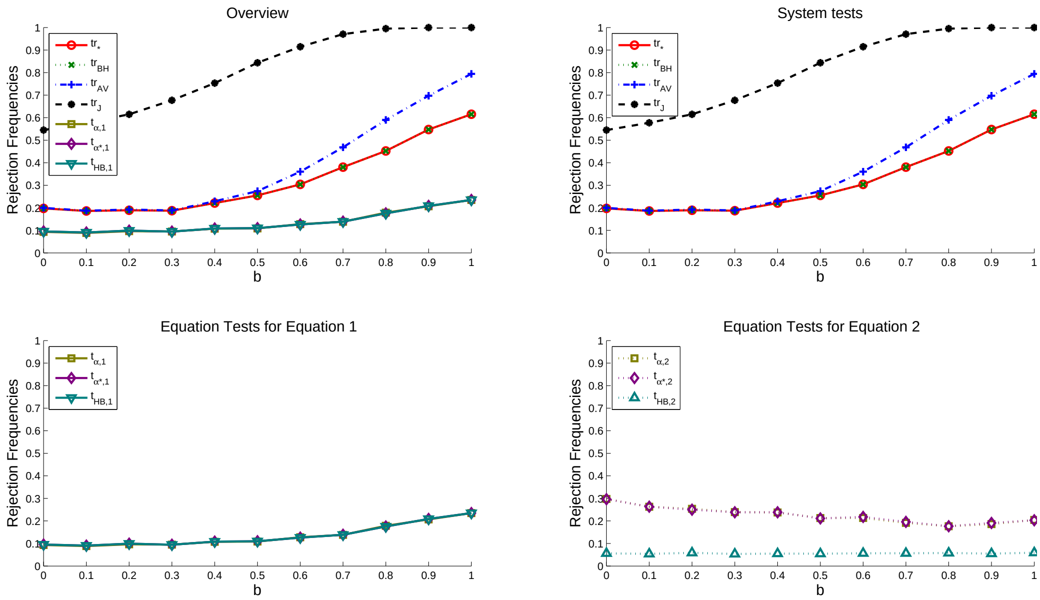

Figure 8: Size/power performance of the different test statistics. Sample size $T = 200$, $d = 0.8$ and is estimated with ELW estimator, no short run dynamics (i.e., $A = 0$), $\rho = 0$ and number of augmentations p_{max} is set to zero.

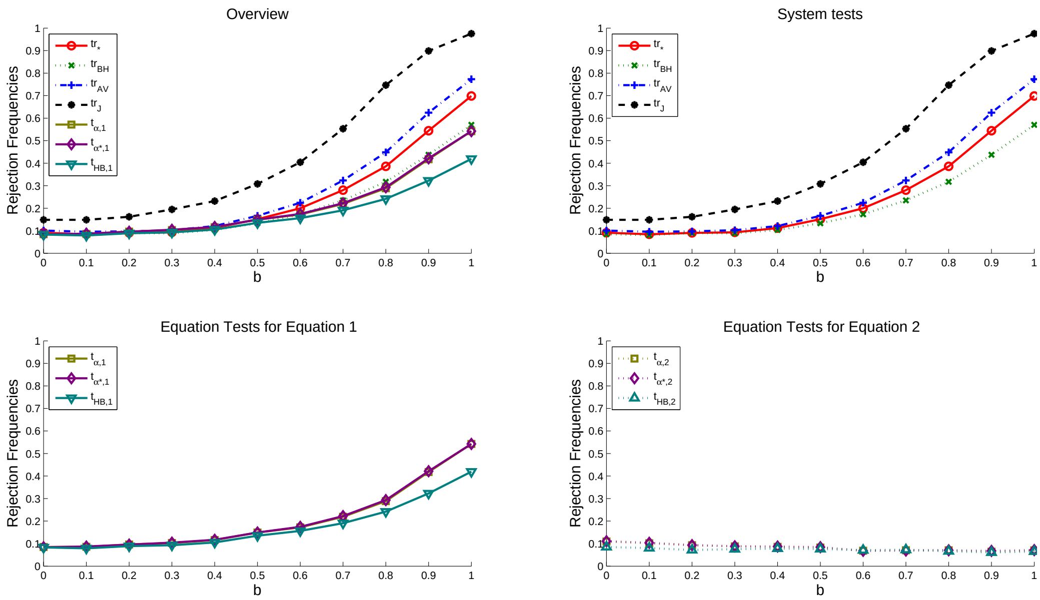

Figure 9: Size/power performance of the different test statistics. Sample size $T = 200$, $d = 0.8$ and is estimated with ELW estimator, no short run dynamics (i.e., $A = 0$), $\rho = 0$ and number of augmentations p_{max} is set to $4(T/100)^{0.25}$.

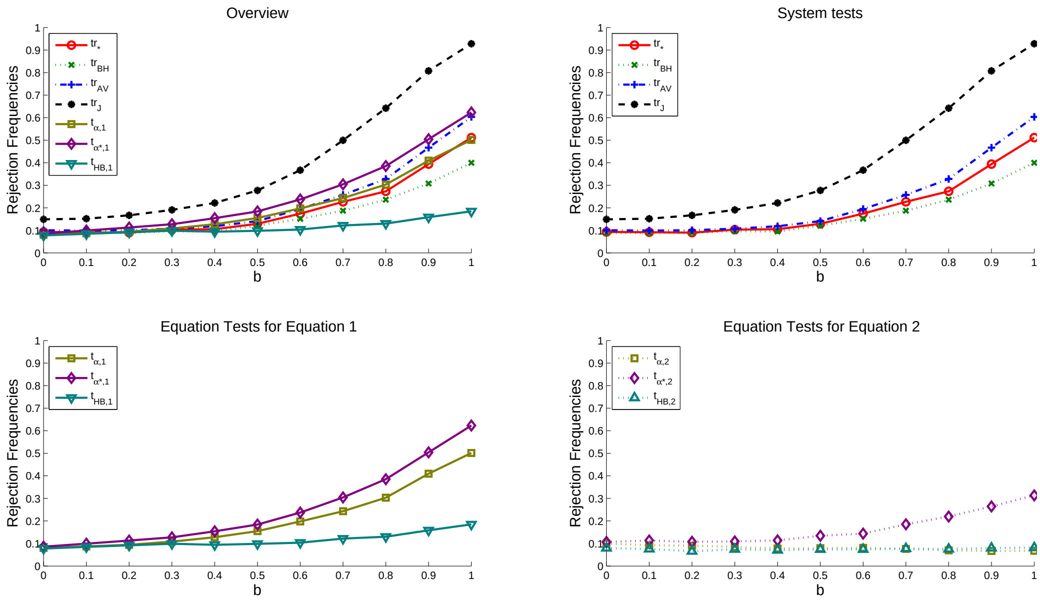

Figure 10: Size/power performance of the different test statistics. Sample size $T = 200$, $d = 0.8$ and is estimated with ELW estimator, no short run dynamics (i.e., $A = 0$), $\rho = 0.5$ and number of augmentations p_{max} is set to $4(T/100)^{0.25}$.

References

Avarucci, M. and C. Velasco (2009). A Wald Test for the Cointegration Rank in Nonstationary Fractional Systems. *Journal of Econometrics 151*(2), 178–189.

Banerjee, A., J. J. Dolado, and R. Mestre (1998). Error-Correction Mechanism Tests for Cointegration in a Single-Equation Framework. *Journal of Time Series Analysis 19*(3), 267–283.

Benjamini, Y. and Y. Hochberg (1995). Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological) 57*(1), 289–300.

Breitung, J. and U. Hassler (2002). Inference on the Cointegration Rank in Fractionally Integrated Processes. *Journal of Econometrics 110*(2), 167–185.

Bura, E. and R. D. Cook (2003). Rank Estimation in Reduced-Rank Regression. *Journal of Multivariate Analysis 87*(1), 159–176.

Davies, R. B. (1977). Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternative. *Biometrika 64*(2), 247–254.

Davies, R. B. (1987). Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternative. *Biometrika 74*(1), 33–43.

Demetrescu, M. (2009). Panel Unit Root Testing with Nonlinear Instruments for Infinite-Order Autoregressive Processes. *Journal of Time Series Econometrics 1*(2), article 3.

Demetrescu, M., U. Hassler, and V. Kuzin (2011). Pitfalls of Post-Model Selection Testing: Experimental quantification. *Empirical Economics 40*(2), 359–372.

Demetrescu, M., V. Kuzin, and U. Hassler (2008). Long Memory Testing in the Time Domain. *Econometric Theory 24*(1), 176–215.

Dolado, J. J., J. Gonzalo, and L. Mayoral (2002). A Fractional Dickey-Fuller Test for Unit Roots. *Econometrica 70*(5), 1963–2006.

Engle, R. F., D. F. Hendry, and J. F. Richard (1983). Exogeneity. *Econometrica 51*(2), 277–304.

Franchi, M. (2010). A Representation Theory for Polynomial Cofractionality in Vector Autoregressive Models. *Econometric Theory 26*(4), 1201–1217.

Granger, C. W. J. (1986). Developments in the Study of Cointegrated Economic Variables. *Oxford Bulletin of Economics and Statistics 48*(3), 213–228.

Hassler, U. and J. Breitung (2006). A Residual-Based LM-Type Test against Fractional Cointegration. *Econometric Theory 22*(6), 1091–1111.

Hassler, U. and B. Meller (2013). Detecting Multiple Breaks in Long Memory the Case of US Inflation. *Empirical Economics 46*, 653–680.

Hualde, J. (2013). A Simple Test for the Equality of Integration Orders . *Economics Letters 119*(3), 233–237.

Johansen, S. (1991). The Power Function of the Likelihood Ratio Test for Cointegration. In J. Gruber (Ed.), *Econometric decision models: New methods of modeling and applications*, pp. 323–335. Springer Verlag.

Johansen, S. (1992). Cointegration in Partial Systems and the Efficiency of Single-Equation Analysis. *Journal of Econometrics 52*(3), 389–402.

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

Johansen, S. (2008). A Representation Theory for a Class of Vector Autoregressive Models for Fractional Processes. *Econometric Theory 24*(3), 651–676.

Johansen, S. and M. Ø. Nielsen (2012a). Likelihood Inference for a Fractionally Cointegrated Vector Autoregressive Model. *Econometrica 80*(6), 2667–2732.

Johansen, S. and M. Ø. Nielsen (2012b). The Role of Initial Values in Nonstationary Fractional Time Series Models. Working Paper 12-18, University of Copenhagen Department of Economics.

Kew, H. and D. Harris (2009). Heteroskedasticity-Robust Testing for a Fractional Unit Root. *Econometric Theory 25*(6), 1734–1753.

Kremers, J. J. M., N. R. Ericsson, and J. J. Dolado (1992). The Power of Cointegration Tests. *Oxford Bulletin of Economics and Statistics 54*(3), 325–348.

Łasak, K. (2010). Likelihood Based Testing for no Fractional Cointegration. *Journal of Econometrics 158*(1), 67–77.

Leeb, H. and B. Pötscher (2005). Model Selection and Inference: Facts and fiction. *Econometric Theory 21*(1), 21–59.

Lobato, I. N. and C. Velasco (2007). Efficient Wald Tests for Fractional Unit Roots. *Econometrica 75*(2), 575–589.

Lütkepohl, H. (1996). *Handbook of Matrices.* Academic Press.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988). Testing Linearity Against Smooth Transition Autoregressive Models. *Biometrika 75*(3), 491–499.

Nielsen, M. Ø. (2005). Multivariate Lagrange Multiplier Tests for Fractional Integration. *Journal of Financial Econometrics 3*(3), 372–398.

Robinson, P. M. (1994). Efficient Tests of Nonstationary Hypotheses. *Journal of the American Statistical Association 89*(428), 1420–1437.

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

Tanaka, K. (1999). The Nonstationary Fractional Unit Root. *Econometric Theory 15*(4), 549–582.

Urbain, J. P. (1993). *Exogeneity in error correction models.* Springer.

White, H. (2001). *Asymptotic theory for econometricians.* Academic Press.

Zivot, E. (2000). The Power of Single Equation Tests for Cointegration when the Cointegrating Vector is Prespecified. *Econometric Theory 16*(03), 407–439.