Recursive Adjustment for General Deterministic Components and Improved Cointegration Rank Tests*

Benjamin Born
University of Mannheim

Matei Demetrescu†
University of Bonn

This version: September 19, 2013

Abstract

The paper discusses tests for the cointegration rank of integrated vector autoregressions when the series are recursively adjusted for deterministic components. To this end, the asymptotic properties of recursive, or adaptive, procedures for the removal of general additive deterministic components are analyzed in two different, complementary, situations. When the stochastic component of the examined time series is weakly stationary (as would be the equilibrium errors), the effect of recursive adjustment vanishes with increasing sample size. When the suitably normalized stochastic component converges weakly to some limiting continuous-time process with integrable paths (as would be the case with the common stochastic trends), recursive adjustment has a permanent effect even asymptotically: the normalized recursively adjusted process converges weakly to a recursively adjusted version of the limiting process. The null limiting distributions of the cointegration rank tests can be expressed in terms of recursively adjusted Brownian motions. Moreover, the finite-sample properties of the cointegration rank tests with recursive adjustment are examined in cases of empirical relevance: the considered deterministic components are a constant, and a constant and a linear trend, respectively. Compared to the likelihood ratio tests or the tests with GLS adjustment, improvements in terms of empirical rejection frequencies under the null are found in finite samples; improvements are found under the alternative as well, with the likelihood ratio test performing increasingly better as the magnitude of the initial condition increases. Regarding rank selection, a very simple combination of the three testing procedures with different adjustments performs best.

Key words

Adaptive detrending, Vector autoregressive process, Long-run relation, LR-type test

JEL classification

C12 (Hypothesis Testing), C32 (Time-Series Models)

---

*The authors would like to thank Christoph Hanck, Helmut Lütkepohl, Michalis Stamatogiannis and Carsten Trenkler for very helpful comments and suggestions.

†**Corresponding Author.** Address: Hausdorff Center for Mathematics and Department of Economics, University of Bonn, Adenauerallee 24-42, D-53113 Bonn, Germany. E-mail address: `matei.demetrescu@uni-bonn.de`.

1 Introduction

The question of how to account for deterministic components in the data is perhaps as old as time series econometrics itself. The more popular method to separate the stochastic component from the deterministic one appears to still be adjustment using ordinary least squares [OLS], even in cases where generalized least squares [GLS] may be the efficient solution.

Recursive removal of deterministic trend components, or recursive adjustment,[1] was introduced as an alternative by So and Shin (1999) for demeaning in the context of estimating of covariances and autoregressive parameters. Recursive demeaning amounts to subtracting from each observation y_t the mean $\overline{y}_t = t^{-1} \sum_{j=1}^{t} y_j$; being adaptive, it has e.g. in autoregressive models the advantage of preserving the martingale difference property of the product of regressor and disturbance, unlike the usual OLS demeaning. This way, the bias, if not the variance, of statistics computed with recursive demeaning can be reduced. See So and Shin (1999) for details. The bias reduction is very useful for working with possibly integrated time series. In this context, recursive demeaning has been used with the Dickey-Fuller test for a unit root by Shin and So (2001); Taylor (2002) accounts for a linear trend in a recursive manner in the more general setup of seasonal unit root tests; see also Rodrigues (2006) for unit root tests with recursive trend removal and Kuzin (2005) for the issue of recursive deseasonalizing. Since the estimation bias of the autoregressive parameters—which is substantial in the case of autoregressive roots close to unity—is reduced, more powerful unit root tests can be obtained. This is documented by Leybourne et al. (2005) who show that tests using recursive adjustment are comparable with unit root tests using local-to-unity GLS adjustment (Elliott et al., 1996). At the same time, the recursive procedure keeps the simplicity inherent to the OLS approach – recall that GLS adjustment requires specification of fine-tuning parameters. But unit root testing is not the only field of application of recursive adjustment: Sul et al. (2005) use recursive demeaning to reduce the bias of autoregressive estimates for long-run variance estimation with prewhitening.

In the context of multivariate nonstationary models, it is only natural to ask whether recursive adjustment brings advantages over alternative adjustment methods when testing for (no) cointegration or for the cointegration rank.

So this paper's contribution is twofold. First, it examines recursive adjustment for general deterministic components. This is a necessary step in analyzing cointegration rank tests with recursive adjustment; but the topic is of its own interest and we treat it under more general conditions than required for the analysis of vector autoregressive [VAR] processes. Concretely, two possible situations for the stochastic component are

[1] The term "recursive detrending" has been used to denote the recursive removal of deterministic components as well; "recursive adjustment" is preferred here, as it leaves little space for confusion with the particular case of removing a linear trend.

distinguished. In the first, weak stationarity is assumed – a condition satisfied, for instance, by the equilibrium error of a cointegrated VAR. In the second, nonstationary, case, a limiting continuous-time process with integrable paths should exist when a suitable normalization is used; this applies to the stochastic trends of the VAR in question. But nonstationary fractionally integrated processes can also be accommodated by this assumption, the weak limit being a fractional Brownian motion. The effect on the recursively adjusted series is different in the two cases, and precise results in this direction are derived. In the stationary case, the effect vanishes as the time index grows to infinity, but remains non-negligible at the beginning of the sample. In the nonstationary case, the limiting behavior is a continuous-time recursively adjusted process.

The second contribution is to use recursive adjustment when dealing with the deterministic component of cointegrated VARs for the purpose of testing hypotheses about the cointegration rank. The procedure for estimating the cointegration rank of integrated VARs proposed by Johansen (1995) has proved to be an extremely valuable tool in specifying cointegrated VAR models for subsequent inference.[2] But when the data generating process exhibits deterministic components, there is no gold standard for how to remove them from the data. Johansen (1995) suggests to remain in the likelihood ratio [LR] framework; Saikkonen and Lütkepohl (2000a), Lütkepohl and Saikkonen (2000), and Saikkonen and Lütkepohl (2000b) on the other hand suggest to remove the constant or the linear trend by GLS adjustment prior to applying the LR or Lagrange multiplier tests. Neither procedure uniformly dominates the other. Since the LR rank test can be seen as a generalization of the Dickey-Fuller test, it is expected that recursive adjustment performs well in the multivariate setting, too. The paper examines the asymptotic and small-sample properties of the LR-type rank test applied to recursively adjusted data. The asymptotic distributions depend, not surprisingly, on recursively adjusted Wiener processes. In small samples, recursive adjustment is found to perform better in most cases in terms of empirical rejection frequencies under both the null and the alternative; the initial condition also plays a role. When focusing on the selection of a cointegration rank rather than on the test of a specific null (e.g. no cointegration), a simple combination of the tests with recursive, OLS-type, and GLS adjustment appears to be the better choice.

Before proceeding to the main part of the paper, let us introduce some notation. Let boldface symbols denote column vectors, "$\Rightarrow$" stand for weak convergence in a space of càdlàg vector functions defined on $[0, 1]$, space endowed with a suitable norm, and "I(0)" and "I(1)" stand for integration of order 0 and 1, respectively. Furthermore, $\|\cdot\|$ denotes the Euclidean vector norm and the corresponding induced matrix norm, and the Landau symbols $O(\cdot)$ and $O_p(\cdot)$ have their usual meaning. Let $\operatorname{diag}\mathbf{a} = \operatorname{diag}(a_1, \ldots, a_m)'$ denote an $m \times m$ diagonal matrix having the elements a_i on the main diagonal. Finally, C^* is a

[2]Sequential testing is not the only way to infer about the cointegration rank: Poskitt and Lütkepohl (1995) suggest using specific information criteria. See also the more recent Cheng and Phillips (2009).

constant whose value may differ at different occurrences.

2 Model and test

2.1 Assumptions

Let the observed series be generated according to

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{d}_t, \quad t = 1, \dots, T, \tag{1}$$

where $\mathbf{x}_t$ is a K-dimensional integrated VAR$(p+1)$ process, possibly cointegrated, and the vector $\mathbf{d}_t = (d_{1t}, \dots, d_{Kt})'$ is a deterministic function of t.

Let the error-correction representation of $\mathbf{x}_t$ be

$$\Delta \mathbf{x}_t = \Pi \mathbf{x}_{t-1} + \sum_{j=1}^{p} A_j \Delta \mathbf{x}_{t-j} + \boldsymbol{\varepsilon}_t, \tag{2}$$

where $\mathbf{x}_0$ is fixed and the innovations $\boldsymbol{\varepsilon}_t$ are uncorrelated. When $\Pi = 0$, $\mathbf{x}_t$ is integrated, and, when Π has full rank, $\mathbf{x}_t$ is stationary. When $r = \operatorname{rank} \Pi$ is nonzero, but less than K, the matrix can be decomposed as $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ being both $K \times r$ matrices. Denote by $A_\perp$ the orthogonal complement w.r.t. $\mathbb{R}^K$ of a matrix A with K rows and $0 < r < K$ columns. The following assumption ensures that $\mathbf{x}_t$ is I(1) with cointegration rank r.

Assumption 1 *Let the characteristic equation*

$$\det\left((1-z)I_K - \Pi z - \sum_{j=1}^{p} A_j (1-z) z^j\right) = 0$$

have no roots outside the unit circle and $K - r$ roots on the unit circle, and, if $r > 0$, let the matrix $\boldsymbol{\alpha}'_\perp \left(I_K - \sum_{j=1}^{p} A_j\right) \boldsymbol{\beta}_\perp$ have full rank.

The innovations $\boldsymbol{\varepsilon}_t$ are taken to satisfy conditions standard in the cointegration literature:

Assumption 2 *Let $\boldsymbol{\varepsilon}_t$ be either an iid sequence with finite variance or a weakly stationary martingale difference sequence with uniformly bounded 4th-order moments. Denote by $\Sigma_{\boldsymbol{\varepsilon}}$ the covariance matrix of $\boldsymbol{\varepsilon}_t$, and let $\Sigma_{\boldsymbol{\varepsilon}}$ be positive definite.*

The Granger representation theorem implies for $\mathbf{x}_t$,

$$\mathbf{x}_t = \Xi \sum_{j=1}^{t} \boldsymbol{\varepsilon}_j + O_p(1), \tag{3}$$

where the matrix Ξ is given as $\Xi = \boldsymbol{\beta}_\perp \left(\boldsymbol{\alpha}'_\perp \left(I_K - \sum_{j=1}^p A_j\right) \boldsymbol{\beta}_\perp\right)^{-1} \boldsymbol{\alpha}'_\perp$. Given Assumption 2, it follows immediately that

$$\frac{1}{\sqrt{T}}\, \mathbf{x}_{[sT]} \Rightarrow \mathbf{B}(s),$$

where the covariance matrix of the K-dimensional Brownian motion $\mathbf{B}(s)$ is $\Xi\Sigma_{\boldsymbol{\varepsilon}}\Xi'$ and satisfies $\operatorname{rank} \Xi\Sigma_{\boldsymbol{\varepsilon}}\Xi' = \operatorname{rank} \Xi = K - r$.

In what concerns the stochastic part of the model, the assumptions are in fact quite standard. See Johansen (1995) and Lütkepohl (2005). It is the deterministic component that receives more attention in this paper.

Let each of the K deterministic components d_{it} be a linear combination of simple trend functions, which are the same for all K elements of $\mathbf{y}_t$,

$$d_{it} = \sum_{l=1}^{L} c_{il} f_l(t). \tag{4}$$

Further, let the trend functions $f_l(\cdot)$ be known. Naturally, the coefficients may differ across the K elements of $\mathbf{y}_t$; in particular, some may be 0, thus accommodating the situation where some of the trend functions are specific to certain elements of $\mathbf{y}_t$.

For the asymptotic analysis, an assumption about the behavior of the trend function $\mathbf{f}(t)$ as T grows to infinity is required.

Assumption 3 *Let $D_T = \operatorname{diag}(T^{\alpha_1}, \ldots, T^{\alpha_L})$ for some real constants α_l, $l = 1, \ldots, L$; the trend functions $f_l(t)$, $l = 1, \ldots, L$, are linearly independent and satisfy*

$$\mathbf{f}(t) = D_T\ \boldsymbol{\tau}\left(\frac{t}{T}\right), \quad t = 1, \ldots, T,$$

where $\boldsymbol{\tau}(s) = (\tau_1(s), \ldots, \tau_L(s))'$ is a deterministic vector function on $[0, 1]$, piecewise smooth with bounded derivative such that $\int_0^s \boldsymbol{\tau}(r)\boldsymbol{\tau}(r)'\mathrm{d}r$ has full rank $\forall 0 < s \leq 1$.

The condition of a full rank in the limit is not equivalent to the linear independence of the trend functions and is thus required. The assumption allows e.g. for a constant (where $\alpha_1 = 0$ and $\tau_1 = 1$), for polynomial trends (where $\alpha_1 = k$ and $\tau_1 = s^k$), or for smooth transitions (where $\boldsymbol{\tau}$ contains so-called sigmoid functions such as the hyperbolic tangent). Piecewise linear trends or polynomial splines are covered by the assumption as well. Assume at this point that a constant is always included in the set of deterministic functions: a linear trend, for instance, is not plausible at all without a nonzero intercept.[3] Actually, a quadratic trend would not be plausible without a linear one either, and so on.

[3] In the direction of the stochastic trends, the constant is not distinguishable from the initial value; adjusting for deterministics with a constant included, as assumed, renders the statistics pivotal.

Up to the mentioned inclusion of a constant, we leave the choice of (not) considering all powers in a higher-order polynomial trend function to the practitioner: the assumption encompasses both restricted and unrestricted cases, and allows one to work with recursive adjustment as one considers fit for the data set at hand.

Rodrigues (2013) analyzes unit root tests with recursive adjustment for intercept, trend, and breaks. Dealing with structural breaks at a fixed position in the sample is easily accommodated via interaction terms involving the usual dummy variables. In terms of Assumption 3, a break behaves almost like a constant, having $\alpha_1 = 0$; and $\tau_1 = \mathbf{1}(s > \lambda)$ or $\tau_1 = \mathbf{1}(s < \lambda)$, with $\mathbf{1}(\cdot)$ the indicator function, for some (relative) break time $0 < \lambda < 1$.

In the presence of breaks, there actually is an interesting computational aspect to the recursive adjustment procedure: some algebra shows that recursively adjusting (as described in the following subsection) for the trend components and their interactions with a dummy variable is equivalent to simply adjusting the pre- and post-break subsamples separately whenever all trend components exhibit a break at the same time. This gives a straightforward solution to an issue with Assumption 3, which is not fulfilled by step dummies defined as $\tau_1 = \mathbf{1}(s > \lambda)$ violating the full-rank requirement on $\int_0^s \boldsymbol{\tau}(r)\boldsymbol{\tau}(r)'\mathrm{d}r$ for $s < \lambda$. Another way of dealing with the issue is to use dummies defined as $\tau_1 = \mathbf{1}(s < \lambda)$.

The assumption implies furthermore that the break time is known; but it is a plausible conjecture that the break time can be endogenized in a straightforward manner. In the context of unit root tests, for instance, breaks at unknown time can be accounted for by using inf-type statistics as those introduced by Zivot and Andrews (1992); see Lütkepohl et al. (2004) for the corresponding discussion of cointegration-rank tests.[4]

Note that Assumption 3 does not directly cover seasonal deterministics; but the work of Taylor (2002) and Kuzin (2005) indicates how seasonal components can be reduced to the corresponding nonseasonal case. Finally, there are cases where $\int_0^s \boldsymbol{\tau}(r)\boldsymbol{\tau}(r)'\mathrm{d}r$ is not a proper integral and different methods apply; see the discussion following Proposition 1.

2.2 Recursive adjustment

So far, only recursive demeaning and detrending have been considered in the literature. But the extension to general additive trend components is straightforward. The procedure is OLS-based as well, but the unknown coefficients are not estimated from the entire sample; rather, one estimates for each t the unknown trend parameters based on the sample $1, \ldots, t$, and uses them subsequently to adjust the observation y_t. Recursive adjustment of a multivariate time series is understood to be element-wise.

[4]Establishing rigorous asymptotics for inf-type statistics with recursive adjustment for structural breaks is beyond the scope of this paper. See Rodrigues (2013) for the case of unit root tests and the problems associated with this procedure.

Denoting by C the matrix of coefficients $\{c_{il}\}_{i=1,\ldots,K,l=1,\ldots,L}$ and by $\mathbf{f}(t)$ the vector $(f_1(t), \ldots, f_L(t))'$, one has in matrix notation

$$\mathbf{d}_t = C\mathbf{f}(t).$$

Let us now turn our attention to the corresponding procedures of recursively adjusting the series. Formally, there are $T - L + 1$ least-squares regressions, one for each $t = L, \ldots, T$:

$$\mathbf{y}_j = \widehat{C}^{(t)}\mathbf{f}(j) + \widehat{\mathbf{x}}_j^{(t)}, \quad j = 1, \ldots, t,$$

giving

$$\widehat{C}^{(t)} = \left(\left(\sum_{j=1}^{t} \mathbf{f}(j)\mathbf{f}(j)'\right)^{-1} \left(\sum_{j=1}^{t} \mathbf{f}(j)\mathbf{y}_j'\right)\right)'.$$

Denote by $\widetilde{\mathbf{y}}_t$ the recursively adjusted series, for which the above definition delivers

$$\widetilde{\mathbf{y}}_t = \mathbf{y}_t - \left(\left(\sum_{j=1}^{t} \mathbf{f}(j)\mathbf{f}(j)'\right)^{-1} \left(\sum_{j=1}^{t} \mathbf{f}(j)\mathbf{y}_j'\right)\right)' \mathbf{f}(t) \tag{5}$$

for $t = L, \ldots, T$. At the beginning of the sample, one obtains $\widetilde{\mathbf{y}}_L = \mathbf{0}$; set $\widetilde{\mathbf{y}}_t = \mathbf{0}$ as well for $t = 1, \ldots, L - 1$.

Fitting an OLS regression for each t between L and T may be computationally more expensive than simple OLS adjustment; but simple closed-form expressions for $\widetilde{\mathbf{y}}_t$ are available e.g. in the case of recursive demeaning or recursive detrending (see Example 3), and one can make use of the Sherman-Morrisson formula to simplify the computation of $\widehat{C}^{(t)}$ by updating $\widehat{C}^{(t-1)}$.

Equation (5) implies for $t \geq L$ that

$$\widetilde{\mathbf{y}}_t = \mathbf{x}_t - \left(\left(\sum_{j=1}^{t} \mathbf{f}(j)\mathbf{f}(j)'\right)^{-1} \left(\sum_{j=1}^{t} \mathbf{f}(j)\mathbf{x}_j'\right)\right)' \mathbf{f}(t),$$

also, $\widetilde{\mathbf{d}}_t = \mathbf{0}$ for all t. More generally, recursive adjustment is—just like the usual OLS procedure—linear:

$$A\widetilde{\mathbf{y}}_t = \widetilde{A\mathbf{y}}_t \tag{6}$$

for conformable matrices A, so recursively adjusting $\mathbf{y}_t$ leaves the potential cointegration structure unaffected. Concretely, if $\mathbf{a}'\mathbf{x}_t$ is an I(0) linear combination, $\widetilde{\mathbf{a}'\mathbf{x}}_t$ behaves asymptotically stationary, cf. Proposition 1 below, and, as a consequence of Equation (6), so will $\mathbf{a}'\widetilde{\mathbf{x}}_t$. If on the other hand $\mathbf{a}'\mathbf{x}_t$ is an I(1) linear combination, $\mathbf{a}'\widetilde{\mathbf{x}}_t$ is stochastically trending too.

2.3 The test with recursive adjustment

The invariance of the cointegration property to recursive adjustment suggests that such adjustment can be used with Johansen's (1995) test for the cointegration rank. Even when the innovations are Gaussian, the result is not an LR test as the trend parameters are not concentrated out of the likelihood function.

Without deterministics, the LR test can be interpreted as a problem of canonical correlations of the differences $\Delta\mathbf{y}_t$ and lagged levels $\mathbf{y}_{t-1}$ (with prewhitening if short-run dynamics are to be accounted for). This subsection thus discusses how $\mathbf{y}_{t-1}$ and $\Delta\mathbf{y}_t$ can be adjusted for their respective deterministic components.

Let $Rec^{\boldsymbol{\tau}}$ denote the LR-type test statistic with recursive adjustment. To compute it, the lagged levels $\mathbf{y}_{t-1}$ are dealt with as indicated by Equation (5), delivering the recursively adjusted $\widetilde{\mathbf{y}}_{t-1}$.

Should the deterministic component only include an intercept, for instance, the differenced series $\Delta\mathbf{y}_t$ do not require adjustment at all. But this is not the case in general; moreover, the deterministic component of the differenced series, $\mathbf{d}_t - \mathbf{d}_{t-1}$, is typically different from the deterministic component of the levels; see the following lemma.

Lemma 1 *Denote by $\mathcal{S}$ the set of points where $\boldsymbol{\tau}$ or its derivative have jump discontinuities. Let $\dot{\boldsymbol{\tau}}(s) = \left(\frac{\partial f_1}{\partial s}, \ldots, \frac{\partial f_L}{\partial s}\right)'$ for $s \in [0,1] \setminus \mathcal{S}$, and let for convenience $\mathbf{f}(-1) = \mathbf{f}(0)$. Under the conditions of Assumption 3, it holds true that*

$$\sup_{s\in[0,1]\setminus\mathcal{S}} \left| TD_T^{-1}\left(\mathbf{f}([sT]) - \mathbf{f}([sT]-1)\right) - \dot{\boldsymbol{\tau}}(s)\right| \to 0\,.$$

Proof: *Obvious and omitted.*

The discontinuity points in $\mathcal{S}$ correspond to additive outliers in the differences. Their effect on the test statistics studied here depends on their magnitude: a differenced shift in the mean is asymptotically negligible; but a break in the coefficient of a linear trend without a simultaneous shift in the mean canceling the discontinuity is not, and has to be taken into account when adjusting the differences by means of an impulse dummy. A discontinuity of such magnitude is not plausible, though. So, also considering the situation of a constant differenced away, let us convene that the vector $\dot{\boldsymbol{\tau}}(s)$ does not contain zero elements or "outliers".[5] The conditions of the lemma exclude for instance trend functions such as t^p for $0 < p < 1$, since its derivative has a pole; the reason for the exclusion is technical, as unbounded derivatives of $\boldsymbol{\tau}$ might lead to indeterminate forms under the supremum and the result would not follow immediately.

[5] Alternatively, one could have required the differenced deterministic component $\Delta\mathbf{d}_t$ to obey Assumption 3 without the smoothness condition; the levels $\mathbf{d}_t$ would have had as limiting functions the corresponding indefinite integrals and potentially an additional constant. Again, special care would have been required when modeling discontinuities in the levels.

Lemma 1 makes clear that the differences $\Delta \mathbf{y}_t$ require appropriate adjustment, except for the case of a (piecewise) constant deterministic component. In the context of recursive detrending, Rodrigues (2006) lists three possibilities to do so. Two adjust the levels $\mathbf{y}_t$ and the lagged levels $\mathbf{y}_{t-1}$ separately, and the sample mean of the differences $\Delta \mathbf{y}_t$ is subtracted from the adjusted levels. This is equivalent to OLS adjusting the differences $\Delta \mathbf{y}_t$ directly. The third, put forward by Taylor (2002), ultimately amounts to using differenced adjusted series. One may also consider recursively adjusting the differences for the differenced deterministics. Either way, the interaction of the adjustment procedures for the the differences and the lagged levels leaves non-negligible terms, as can be seen in the following example.

Example 1 *Consider for simplicity the univariate case $y_t = c_0 + c_1 t + x_t$ where x_t is an integrated process. The recursively detrended levels are given by*

$$\widetilde{y}_t = x_t + \frac{2}{t}\sum_{j=1}^{t} x_j - \frac{6}{t(t+1)}\sum_{j=1}^{t} j\, x_j,$$

and their differences are given by

$$\Delta\widetilde{y}_t = \Delta x_t - \frac{4t-2}{t(t+1)}x_t - \frac{2}{t(t-1)}\sum_{j=1}^{t-1} x_j + \frac{12}{t(t^2-1)}\sum_{j=1}^{t-1} j\, x_j.$$

For comparison, the recursively demeaned differences are given by

$$\widetilde{\Delta y}_t = \Delta x_t - \frac{1}{t}x_t.$$

With x_t an integrated process, it follows immediately that e.g. $\frac{1}{t}x_t = O_p(t^{-0.5})$, and the other terms such as $\frac{2}{t(t-1)}\sum_{j=1}^{t-1} x_j$ can be shown to have the same order. Examine now the sample cross-product moment of $\widetilde{y}_{t-1}$ and $\widetilde{\Delta y}_t$ (or $\Delta\widetilde{y}_t$), as it would appear in a Dickey-Fuller type statistic: while the terms resulting from the interaction of the adjustment procedures, e.g. $\sum_{t=2}^{T}\frac{1}{t}x_t\widetilde{x}_{t-1}$, can be shown to be nonnegligible, establishing their limiting behavior is not without challenge.

We opt for OLS adjustment of the differenced series for their own deterministic component. This presents two advantages over the alternatives. On the one hand, the complexity of the interaction terms is kept under control; see Proposition 3 for the consequences in the limit. On the other hand, it makes lag augmentation easier to deal with from a technical point of view (uniformity in t of the adjustment term does have its advantages), since it is not clear whether simply lagging e.g. recursively adjusted differences would lead to a pivotal test statistic, or whether different adjustment schemes for each lagged difference are not necessary after all.

Denote thus by $\breve{\Delta\mathbf{y}}_t$ the series $\Delta\mathbf{y}_t$ with OLS adjustment, i.e.

$$\breve{\Delta\mathbf{y}}_t = \Delta\mathbf{y}_t - \left(\sum_{t=2}^{T}\Delta\mathbf{y}_t\Delta\mathbf{f}(t)'\right)\left(\sum_{t=2}^{T}\Delta\mathbf{f}(t)\Delta\mathbf{f}(t)'\right)^{-1}\Delta\mathbf{f}(t)\,, \tag{7}$$

where $\Delta\mathbf{f}(t)$ contains only the deterministic components relevant to the differences. Although $\sqrt{T}$-consistent, this step does affect the limiting distributions; see Proposition 3 below.

After adjusting the levels and the differences for their specific deterministic components, one regresses $\breve{\Delta\mathbf{y}}_t$ from (7) and $\widetilde{\mathbf{y}}_{t-1}$ from (5) on the lags $\breve{\mathbf{z}}_{t-1} = \left(\left(\breve{\Delta\mathbf{y}}_{t-1}\right)', \ldots, \left(\breve{\Delta\mathbf{y}}_{t-p}\right)'\right)'$ to obtain the prewhitened residuals $\mathbf{r}_{0t}$ and $\mathbf{r}_{1t}$. Using them, the usual moment matrices are built,

$$S_{ij} = \frac{1}{T}\sum_{t=p+1}^{T}\mathbf{r}_{it}\mathbf{r}_{jt}',$$

$i, j = 0, 1$. Denote by $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_K$ the decreasingly ordered solutions of the equation of real variable λ

$$\left|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}\right| = 0,$$

i.e. the eigenvalues of $S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}$. The (trace) $Rec^{\boldsymbol{\tau}}$ statistic is

$$Rec^{\boldsymbol{\tau}}(r) = T\sum_{i=r+1}^{K}\widehat{\lambda}_i. \tag{8}$$

Alternatively, one could use the maximum eigenvalue test,

$$Rec^{\boldsymbol{\tau}}_{\max}(r) = \widehat{\lambda}_{r+1}. \tag{9}$$

The following section discusses the limiting behavior of the trace and max eigenvalue statistics. Along the way, we analyze the effect of recursive adjustment on stationary series (as would be the equilibrium error under nonzero cointegration rank) as well as on nonstationary series (as would be the stochastic trends).

3 Main results

We begin by examining the effect of recursive removal of the deterministic trend component $\mathbf{d}_t$ from the observed series. It is found to depend on the properties of the stochastic component. We thus distinguish two different situations. First, we work with weakly stationary processes. Second, the suitably normalized process is taken to possess a limiting continuous-time process with a.s. integrable paths. Based on these results, we then provide the asymptotics for the cointegration rank tests with recursive adjustment.

3.1 Recursive adjustment under stationarity

Let us first deal with the case of stationarity. The sequence of equilibrium errors falls under this category, being in fact I(0) in our model. One can establish the following result extending the work of So and Shin (1999); while it is formulated slightly more generally than required, the additional cost in terms of proof complexity is negligible.

Proposition 1 *Let Assumptions 3 hold true and assume* $\mathbf{w}_t$ *to be some zero-mean, weakly stationary, process with autocovariance matrices* Γ_h *such that, as* $T \to \infty$, $\phi(T) = \frac{1}{T}\sum_{h=1}^{T} \|\Gamma_h\| \to 0$. *Then* $\exists C^* \in \mathbb{R}^+$ *not depending on* t *such that*

$$\mathrm{E}\left(\|\widetilde{\mathbf{w}}_t - \mathbf{w}_t\|^2\right) \leq C^* \, \phi(t).$$

Proof: *See the Appendix.*

The assumptions of the proposition allow for what is usually understood under short memory, and apply thus (even if indirectly) to $\mathbf{w}_t = \boldsymbol{\beta}'\mathbf{x}_{t-1}$, i.e. the equilibrium errors. But departures from weak stationarity could be allowed for, as long as the variance of the sample mean vanishes as $T \to \infty$, cf. Phillips (1987); the topic is not pursued here, however, just as the case of processes with infinite variance is not addressed in this subsection.[6] Note that weakly stationary processes with long-memory are covered as well; this would allow one to use recursive adjustment for processes fractionally integrated of order d, as long as $d < 0.5$.

Remark 1 *For fixed* t *at the beginning of the sample, the recursive adjustment scheme does not consistently filter the stochastic component, but only unbiasedly; this is no surprise as it only uses the limited information available at time* t. *As* t *grows to infinity (e.g.* $t = [sT]$ *for some fixed* $0 < s < 1$*), one has that* $\widetilde{\mathbf{w}}_t - \mathbf{w}_t \xrightarrow{p} 0$. *Chebyshev's inequality indicates, in a perhaps less rigorous notation, that* $\widetilde{\mathbf{w}}_t - \mathbf{w}_t = O_p\left(\sqrt{\phi(t)}\right)$ *uniformly in* t.

Remark 2 *The result requires in any case finiteness of the expectation of* $\mathbf{w}_t$*: it does not hold if the sample mean is not consistent, even for an iid sequence. The rate at which* $\widetilde{\mathbf{w}}_t$ *converges to* $\mathbf{w}_t$ *is related to the convergence rate of the sample mean* $\overline{\mathbf{w}}$*, and, more generally, to the amount of serial dependence of the series* $\mathbf{w}_t$.

If α_l is negative for some $1 \leq l \leq L$, the trend function $f_l(t)$ vanishes as $T \to \infty$. Such trend functions often do not affect the asymptotic behavior of statistics based on the data *without* accounting for this trend component (see also Example 2 below). In small samples, however, even such trend functions can make a difference, so it is advisable to take

[6] In the stationary case with infinite variances, the OLS estimators of the coefficients c_{il} may be inconsistent and Least Absolute Deviations fitting called for; see Remark 2. Dealing with infinite variances is much easier in the nonstationary case; see the following subsection for details.

them into consideration. In this case, their coefficient may not be consistently estimated, but the nuisance trend parameters are still eliminated. The invertibility requirement in Assumption 3 may also impose conditions on α_l; for instance, if $f_l(t) = t^{\alpha_l}$, it is required that $\alpha_l > -0.5$.

For cases such as $f_l(t) = t^{\alpha_l}$ with $\alpha_l < -0.5$, one must resort to different methods. This is because $\int_0^s \boldsymbol{\tau}(r)\boldsymbol{\tau}(r)'\mathrm{d}r$ is not a proper integral. Which approach should be used when analyzing such situations depends on the particularities of the respective data generating process. Assume, for instance, the vector $\mathbf{f}(t)$ to have absolutely summable elements, $\sum_{j\geq 1} |f_l(j)| < \infty \;\forall\, l = 1, \ldots, L$, and the limit $\sum_{t\geq 1} \mathbf{f}(t)\mathbf{f}(t)'$ to be for simplicity nonsingular. Then, $\sum_{j=1}^{t} \mathbf{w}_j \mathbf{f}(j)'$ has uniformly bounded variance. Hence, $\widehat{C}^{(t)} - C$ is uniformly bounded in probability, and $\left(\widehat{C}^{(t)} - C\right)\mathbf{f}(t)$ will vanish as required when $t \to \infty$ since $\mathbf{f}(t) \to \mathbf{0}$.

The fact that recursive deterministic components are not "completely" removed at the beginning of the sample, does not hinder the use of the recursively adjusted data: for instance, covariance matrices are consistently estimated, as shown in the example below.

Example 2 *Assume the process* $\mathbf{w}_t$ *to be covariance-ergodic,* $T^{-1}\sum_{h+1}^{T} \mathbf{w}_t\mathbf{w}_{t-h}' \xrightarrow{p} \Gamma_h$. *Denote by* $\widehat{\Gamma}_h$ *the sample autocovariances with recursive adjustment,* $\widehat{\Gamma}_h = \frac{1}{T}\sum_{t=h+1}^{T} \widetilde{\mathbf{w}}_t\widetilde{\mathbf{w}}_{t-h}'$ *and by* $\|\cdot\|_1$ *the city-block norm. Then,*

$$\begin{aligned}
&\mathrm{E}\left(\left\|\mathrm{vec}\left(\widehat{\Gamma}_h - \frac{1}{T}\sum_{t=h+1}^{T}\mathbf{w}_t\mathbf{w}_{t-h}'\right)\right\|_1\right) \\
&\leq \frac{1}{T}\sum_{t=h+1}^{T}\mathrm{E}\left(\left\|\mathrm{vec}\left(\mathbf{w}_t\left(\widetilde{\mathbf{w}}_{t-h} - \mathbf{w}_{t-h}\right)'\right)\right\|_1\right) \\
&\quad + \frac{1}{T}\sum_{t=h+1}^{T}\mathrm{E}\left(\left\|\mathrm{vec}\left(\left(\widetilde{\mathbf{w}}_t - \mathbf{w}_t\right)\mathbf{w}_{t-h}'\right)\right\|_1\right) \\
&\quad + \frac{1}{T}\sum_{t=h+1}^{T}\mathrm{E}\left(\left\|\mathrm{vec}\left(\left(\widetilde{\mathbf{w}}_t - \mathbf{w}_t\right)\left(\widetilde{\mathbf{w}}_{t-h} - \mathbf{w}_{t-h}\right)'\right)\right\|_1\right) \\
&\leq \frac{C^*K^2}{T}\left(\sum_{t=h+1}^{T}\phi(t-h) + \sum_{t=h+1}^{T}\phi(t) + \sum_{t=h+1}^{T}\phi(t)\phi(t-h)\right).
\end{aligned}$$

(To establish the second inequality, the Cauchy-Schwarz inequality has been applied element-wise, and the fact that $\mathbf{w}_t$ *has uniformly bounded second-order moments was exploited.) Since* $\phi(t) \to 0$, *convergence follows. Note that the convergence rate of* $\widehat{\Gamma}_h$ *is*

$$\widehat{\Gamma}_h - \Gamma_h = O_p\left(\max\left\{\frac{1}{T}\sum_{t=1}^{T}\phi(t), \mathrm{E}\left(\left\|\mathrm{vec}\left(\Gamma_h - \frac{1}{T}\sum_{t=h+1}^{T}\mathbf{w}_t\mathbf{w}_{t-h}'\right)\right\|_1\right)\right\}\right).$$

The effect of recursive adjustment on the common trends of the (co-)integrated VAR is quite different. We study it in a more general setup in the following subsection.

3.2 Recursive adjustment under nonstationarity

The behavior of the adjusted series changes in the nonstationary case, say when working in the space of the common trends $\mathbf{w}_t = \boldsymbol{\beta}_\perp' \mathbf{x}_t$. Namely, the difference between $\widetilde{\mathbf{w}}_t$ and $\mathbf{w}_t$ does not vanish asymptotically, not even at the end of the sample. To make the term nonstationarity more precise, we assume that there exists a limiting continuous-time process for the suitably normalized and interpolated process $\mathbf{w}_t$ in question. So let

$$\mathrm{diag}\left(\boldsymbol{\nu}_T^{-1}\right) \mathbf{w}_{[sT]} \Rightarrow \mathbf{V}(s), \quad s \in [0,1] \tag{10}$$

for some K-dimensional vector $\boldsymbol{\nu}_T$ of normalizing functions such that $\boldsymbol{\nu}_T \to \infty$ as $T \to \infty$, where the K-dimensional process $\mathbf{V}(s)$ is pathwise integrable and almost surely continuous at the origin, $\lim_{s\to 0} \mathbf{V}(s) = \mathbf{V}(0)$ a.s. To formulate the result, define the recursively adjusted limiting process,

$$\widetilde{\mathbf{V}}^{\boldsymbol{\tau}}(s) = \mathbf{V}(s) - \mathbf{V}(0) - \left(\int_0^s (\mathbf{V}(r) - \mathbf{V}(0))\, \boldsymbol{\tau}(r)'\mathrm{d}r\right)\left(\int_0^s \boldsymbol{\tau}(r)\boldsymbol{\tau}(r)'\mathrm{d}r\right)^{-1} \boldsymbol{\tau}(s)$$

for $s \in (0,1]$. Subtracting $\mathbf{V}(0)$ in inconsequential, since $\widetilde{\mathbf{w}}_t$ is invariant to a constant shift, having convened in Section 2.1 that a constant is always to be accounted for. At the same time, this ensures almost sure continuity of $\widetilde{\mathbf{V}}^{\boldsymbol{\tau}}(\cdot)$ at the origin; see the proof of Proposition 2 below. We are now in the position to prove weak convergence of the suitably normalized process $\widetilde{\mathbf{y}}_t$ to the recursively adjusted process $\widetilde{\mathbf{V}}^{\boldsymbol{\tau}}(s)$.

Proposition 2 *Let Assumption 3 hold true jointly with convergence in (10). Then*

$$\mathrm{diag}\left(\boldsymbol{\nu}_T^{-1}\right) \widetilde{\mathbf{w}}_{[sT]} \Rightarrow \widetilde{\mathbf{V}}^{\boldsymbol{\tau}}(s)$$

and $\widetilde{\mathbf{V}}^{\boldsymbol{\tau}}(s)$ *is almost surely continuous at* $s = 0$.

Proof: *See the Appendix.*

The assumptions of the proposition allow for instance $\mathbf{w}_t$ to be an I(1) process, converging weakly to a multivariate Brownian motion (possibly with singular covariance matrix) with $\boldsymbol{\nu}_T = (T^{0.5}, \ldots, T^{0.5})'$, but also for I(d) processes converging to fractional Brownian motion (of type I or II) with a normalizing factor $\boldsymbol{\nu}_T$ depending on the fractional integration parameter d. Processes converging to α-stable Lévy motions are also allowed – hence not requiring $\mathbf{w}_t$ to have finite variance; naturally, $\boldsymbol{\nu}_T$ depends on the index α in this case. Nonzero starting values are covered, too; the key requirement is continuity of the paths at $s = 0$.

Equation (6) applies of course in the nonstationary case as well, leading to

$$\operatorname{diag}\left(\boldsymbol{\nu}_T^{-1}\right)\widetilde{A\mathbf{w}}_{[sT]} \Rightarrow A\widetilde{\mathbf{V}}^{\boldsymbol{\tau}}(s)$$

and to

$$\operatorname{diag}\left(\boldsymbol{\nu}_T^{-1}\right)A\widetilde{\mathbf{w}}_{[sT]} \Rightarrow \widetilde{A\mathbf{V}}^{\boldsymbol{\tau}}(s)$$

for some conformable matrix A.

Example 3 *Let $K = 1$. If $T^{-0.5}\, w_{[sT]}$ converges weakly to a Brownian motion $B(s)$ and one recursively removes a constant ($L = 1$ with $f_1(t) = 1$ and $\alpha_1 = 0$), or a constant and a linear trend ($L = 2$, with $f_1(t) = 1$ and $\alpha_1 = 0$, as well as $f_2(t) = t$ and $\alpha_2 = 1$), two particular cases previously discussed in the literature are recovered. In the constant-only case, it holds that*

$$\widetilde{w}_t = w_t - \frac{1}{t}\sum_{j=1}^{t} w_j; \tag{11}$$

in the constant-and-linear-trend case it holds that

$$\widetilde{w}_t = w_t + \frac{2}{t}\sum_{j=1}^{t} w_j - \frac{6}{t(t+1)}\sum_{j=1}^{t} j\, w_j; \tag{12}$$

$\widetilde{w}_{[sT]}$ converges weakly after normalization to the recursively demeaned Brownian motion $\widetilde{B}^1(s)$,

$$\widetilde{B}^1(s) = \begin{cases} B(s) - \frac{1}{s}\int_0^s B(r)\mathrm{d}r\,, & s > 0 \\ 0 \;\; a.s.\,, & s = 0 \end{cases},$$

and the recursively detrended Brownian motion $\widetilde{B}^{1;s}(s)$,

$$\widetilde{B}^{1;s}(s) = \begin{cases} B(s) + \frac{2}{s}\int_0^s B(r)\mathrm{d}r - \frac{6}{s^2}\int_0^s r\,B(r)\mathrm{d}r, & s > 0 \\ 0 \;\; a.s.\,, & s = 0 \end{cases},$$

respectively. See Shin and So (2001), Taylor (2002), or Chang (2002).

3.3 Limiting distribution

We are now in the position to give the asymptotic distributions of the trace and maximum eigenvalue test statistics; see the following proposition. As was to be expected, they have the typical structure of a multivariate Dickey-Fuller distribution and depend on the deterministic trend term assumed.

Proposition 3 *Let $\mathbf{W}(s)$ be a $(K-r)$-dimensional vector of independent standard Wiener processes and $\widetilde{\mathbf{W}}^{\boldsymbol{\tau}}(s)$ the corresponding recursively adjusted version, and $\mathrm{d}\breve{\mathbf{W}}^{\dot{\boldsymbol{\tau}}}(s)$ denote* *the adjusted differential* $\mathrm{d}\mathbf{W}(s) - \left(\int_0^1 \mathrm{d}\mathbf{W}\dot{\boldsymbol{\tau}}' \left(\int_0^1 \dot{\boldsymbol{\tau}}\dot{\boldsymbol{\tau}}'\mathrm{d}r\right)^{-1}\right) \dot{\boldsymbol{\tau}}(s)\ \mathrm{d}s$ *where* $\dot{\boldsymbol{\tau}}(s) = \frac{\mathrm{d}}{\mathrm{d}s}\boldsymbol{\tau}(s)$. *Under Assumptions 1, 2, and 3 and* $T \to \infty$,

1. *it holds for the* $Rec^{\boldsymbol{\tau}}$ *statistics from (8) computed with* $\mathbf{y}_t$ *from (1) that*

$$Rec^{\boldsymbol{\tau}}(r) \xrightarrow{d} \operatorname{tr}\left[\int_0^1 \mathrm{d}\breve{\mathbf{W}}^{\dot{\boldsymbol{\tau}}} \left(\widetilde{\mathbf{W}}^{\boldsymbol{\tau}}\right)' \left(\int_0^1 \widetilde{\mathbf{W}}^{\boldsymbol{\tau}} \left(\widetilde{\mathbf{W}}^{\boldsymbol{\tau}}\right)' \mathrm{d}s\right)^{-1} \int_0^1 \widetilde{\mathbf{W}}^{\boldsymbol{\tau}} \left(\mathrm{d}\breve{\mathbf{W}}^{\dot{\boldsymbol{\tau}}}\right)'\right],$$

2. *and* $Rec^{\boldsymbol{\tau}}_{\max}(r)$ *from (9) converges in distribution to the maximum eigenvalue of the matrix in brackets.*

Proof: See the Appendix.

The consistency of tests for $r = r_0$ against $r > r_0$ based on $Rec^{\boldsymbol{\tau}}(r_0)$ or $Rec^{\boldsymbol{\tau}}_{\max}(r_0)$ is given: it is argued in the proof that the r non-zero eigenvalues have *the same* limit as when computing them with the unobservable $\mathbf{x}_t$ and not adjusting for a trend. Building on the proof of Proposition 3, the power against sequences of local alternatives can be analyzed. While the topic is not pursued here, it is obvious from the proof that the power is nontrivial in $1/T$ neighbourhoods of the null.

Remark 3 *If the hypothesized cointegration rank,* r_0*, is larger than the actual one, the distribution of* $Rec^{\boldsymbol{\tau}}(r_0)$ *is not given as above, but by the sum of the* $K - r$ *smallest solutions of the eigenvalue problem*

$$\left|\rho \int_0^1 \widetilde{\mathbf{W}}^{\boldsymbol{\tau}} \left(\widetilde{\mathbf{W}}^{\boldsymbol{\tau}}\right)' \mathrm{d}s - \int_0^1 \widetilde{\mathbf{W}}^{\boldsymbol{\tau}} \left(\mathrm{d}\breve{\mathbf{W}}^{\dot{\boldsymbol{\tau}}}\right)' \int_0^1 \mathrm{d}\breve{\mathbf{W}}^{\dot{\boldsymbol{\tau}}} \left(\widetilde{\mathbf{W}}^{\boldsymbol{\tau}}\right)'\right| = 0.$$

See Johansen (1995), top of p. 158.

Should some of the deterministic components in the differences be orthogonal to the cointegrating relations, Proposition 3 still applies, but the deterministic components would be overspecified – a case in which losses in terms of power against sequences of local alternatives can appear, in comparison with the correctly specified model; see Doornik et al. (1998) or Saikkonen and Lütkepohl (1999) for the case of a linear trend. This could be offset by constructing tests that account only for the deterministics present in the cointegrating relations.[7] The drawback is that such tests are only asymptotically similar; see Nielsen and Rahbek (2000) for a discussion of the implications. Moreover, dealing with the uncertainty about such orthogonality is not trivial, as pointed out e.g. in Demetrescu et al. (2009). This paper takes the easy way out and only discusses test statistics that are similar.

[7]It is tempting to conjecture what their asymptotic behavior would be: the unaccounted trend component, if "trending enough," will dominate one of the stochastic trends and the limiting distribution changes accordingly.

4 Comparison with LR and GLS-adjusted tests

This subsection presents results from an examination of the small-sample properties of tests for the cointegration rank based on recursively adjusted data. The two most common cases in empirical work were considered: a constant, as well as a constant and a linear trend. All simulations were conducted in *Matlab*.[8] We compared the Rec^{τ} test with the original LR test of Johansen (1995) and the tests with GLS adjustment (denoted here by SL); see Saikkonen and Luukkonen (1997) and Saikkonen and Lütkepohl (1999) for the mean-adjusted version, and Saikkonen and Lütkepohl (2000b) and Lütkepohl et al. (2001) for the trend-adjusted version.

The test statistics with recursive adjustment were computed as indicated in Section 2.3. In particular, adjusting for deterministics was conducted as follows. The lagged levels $\mathbf{y}_{t-1}$ were recursively adjusted (element-wise) as in Equations (11) and (12). The differences $\Delta\mathbf{y}_t$ required no adjustment in the case of a constant. In the case of a linear trend, they were adjusted for a non-zero mean by subtracting their sample average.

But before proceeding to the examination of the small-sample properties of the test for the cointegration rank with recursive removal of deterministics, critical values are required. These were obtained by discrete approximation of the relevant stochastic functionals using 100 000 replications and 400 integration points.

	trace test						maximum eigenvalue test					
	constant			constant & trend			constant			constant & trend		
$K-r$	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
1	2.78	3.95	6.77	2.29	3.30	5.89	2.76	3.89	6.63	2.28	3.28	5.79
2	9.46	11.28	15.32	7.66	9.36	13.18	8.54	10.28	14.13	6.95	8.57	12.07
3	19.37	21.95	27.26	15.56	17.87	22.74	14.36	16.51	20.99	12.02	13.97	18.28
4	32.73	35.85	42.34	26.27	29.07	35.26	20.27	22.63	27.47	17.35	19.67	24.28
5	50.14	53.99	61.36	40.20	43.74	50.93	26.19	28.70	34.00	22.78	25.29	30.26
6	71.27	75.71	84.59	57.61	61.80	70.30	32.06	34.78	40.31	28.23	30.87	36.45
7	96.14	101.2	111.2	78.17	82.97	92.35	38.07	40.95	46.85	33.83	36.67	42.39
8	124.9	130.7	142.4	102.4	107.9	118.9	43.98	46.97	53.32	39.44	42.44	48.39
9	158.0	164.6	177.1	130.3	136.3	148.6	50.02	53.17	59.79	45.20	48.29	54.86
10	194.9	202.0	215.8	161.5	168.3	181.7	55.96	59.26	66.12	50.95	54.22	60.86

Table 1: Critical values of recursively demeaned and recursively detrended tests

The critical values for the trace and max eigenvalue statistics are given in Table 1 for $K-r$ ranging from 1 to 10; but only results for the trace test are presented in the following small-sample comparison, as it appears to be the more popular variant in applied work. The critical values for the LR tests were taken from Johansen (1995), and those for the SL tests from Trenkler (2003).

Having obtained the needed critical values, one can now examine the small-sample properties of the Rec^{τ} test. As the asymptotic analysis focuses on the similar test statistics, the Monte Carlo samples were generated without deterministic components, $\mathbf{y}_t = \mathbf{x}_t$.

[8] Also, an EViews add-in for the cointegration rank test with recursive adjustment is available from the authors: `http://www.ect.uni-bonn.de/mitarbeiter/prof.-dr.-matei-demetrescu/reclr`.

The stochastic component was generated as follows:

$$\mathbf{x}_t = \begin{bmatrix} \psi I_r & 0 \\ 0 & I_{K-r} \end{bmatrix} \mathbf{x}_{t-1} + \mathbf{u}_t. \tag{13}$$

For $\psi = 1$ (or $r = 0$), the process $\mathbf{x}_t$ is integrated but not cointegrated; for $|\psi| < 1$ and $0 < r < K$, full-rank linear transformations of $\mathbf{x}_t$ are cointegrated with cointegration rank r. We generated the series with initial value

$$\mathbf{x}_0 = \begin{bmatrix} \gamma/\sqrt{1-\psi^2}\; I_r & 0 \\ 0 & \mathbf{0}_{K-r} \end{bmatrix}, \tag{14}$$

where γ was varied to investigate the impact of the initial condition on the behavior of the tests. The initial condition is known to affect the performance of unit root tests; see Müller and Elliott (2003) and Elliott and Müller (2006). By extension, initial values are expected to have an effect on tests for the cointegration rank as well. E.g. Ahlgren and Juselius (2012) quantify the different reaction of the LR and SL tests to the initial value.

The K elements of the process $\mathbf{u}_t$ driving the (co)integrated VAR from (13) were taken to be stable univariate AR(1) processes,

$$\mathbf{u}_t = \phi \mathbf{u}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{15}$$

with $|\phi| < 1$ and *iid* innovations $\boldsymbol{\varepsilon}_t$; more precisely,

$$\boldsymbol{\varepsilon}_t \sim \mathcal{N}\left(0, \Sigma_{\boldsymbol{\varepsilon}} = \begin{bmatrix} I_r & \Theta \\ \Theta' & I_{K-r} \end{bmatrix}\right). \tag{16}$$

Using the VAR(2) data generating process in (13)–(15) allows for a straightforward distinction between the long-run and the short-run properties. By setting $\phi = 0$, $\mathbf{x}_t$ becomes a particular VAR(1) process, a data generating process also used by Toda (1994) and subsequently in a number of other simulation studies where properties of cointegration rank tests were explored; see among others Demetrescu et al. (2009). In Toda's view, this data generating process is particularly useful for investigating the properties of tests for the cointegration rank, since other VAR(1) processes can be obtained from it by full-rank linear transformations to which the tests are invariant. At the same time, it is a parsimonious parametrization which allows for a reduced complexity of the Monte Carlo experiments. Only results for $K = 3$ are reported; simulations were also run with $K = 5$, but the results are similar and we do not report them here to save space.

The matrix Θ is an $r \times (K-r)$ matrix. This matrix was either zero,

$$\Theta = \begin{cases} (0.8, 0.4) & \text{for } K=3, r=1 \\ (0.8, 0.4)' & \text{for } K=3, r=2. \end{cases},$$

or

$$\Theta = \begin{cases} (-0.8, 0.4) & \text{for } K=3, r=1, \\ (-0.8, 0.4)' & \text{for } K=3, r=2. \end{cases}$$

The parameter ψ was set to 0.5. We investigated the impact of the initial condition by setting $\gamma = 0$ and $\gamma = 5$. The generated samples were of size $T = 100$. The simulations were run with 10 000 replications for each case.

For each case, both the relative rejection frequencies of the test based on the Rec^{τ}, LR, and SL statistics, and the relative frequencies of ranks selected by the sequential testing procedure suggested by Johansen (1995) are indicated.

In addition, we report for each case the relative frequencies of ranks selected by a combination procedure, choosing as estimate for the cointegration rank the rounded up average of the ranks estimated using each of the three tests. The reason for using a combination procedure, apart from the fact that it delivers an estimate of the cointegration rank (which is the key input for estimation of vector error-correction models), is as follows. The Monte Carlo experiments in this section show that there are some differences in the behavior of the compared tests, especially in terms of rejection frequencies under the alternative. But the differences depend on the data generating process, so it is not known in advance which of the three tests would perform better. Thus, it appears to be a superior strategy to combine the three tests, following e.g. Harvey et al. (2009) or Bayer and Hanck (2013). There is, however, a serious drawback to combining the tests directly (say, via unions of rejections) because critical values have to be obtained specifically for each combination method. Opposed to that, it is straightforward to combine the estimates of the cointegration rank of the analyzed VAR as outlined above.

The small-sample examination was focused on two cases. First, ϕ was set to 0, and the order of the VAR(1) process $\mathbf{x}_t$ was assumed to be known. This served the purpose of separating the problems of testing for the cointegration rank, and of choosing an appropriate lag order. Second, ϕ was set to -0.5, and the lag order for the unrestricted VAR was chosen by means of the Akaike information criterion, as it is closer to what might be done in real-life applications.

The discussion focusses here on the specification with short-run dynamics, i.e. $\phi = -0.5$.[9] The results for the case $\gamma = 0$ are reported in the upper halves of Tables 2–4. In the case where $\Theta = 0$, the recursively adjusted test has good control with respect to the empirical rejection frequencies under the null; while deviating from the nominal 5%, the

[9]The results for the case without short-run dynamics are available on the corresponding author's homepage. The results are broadly consistent to those with short-run dynamics.

empirical rejection frequency is in all but one case between 4.4% and 5.9%. In contrast, in most of the cases the LR test rejects a bit more often under the null.[10] Compared to the LR test, the SL test is closer to the nominal 5%, especially when detrending, but is not as good as the Rec^{τ} test when demeaning. Moreover, when demeaning, the LR and SL tests tend to have lower empirical rejection frequencies under the alternative in spite of having higher empirical rejection frequencies under the null; in particular the SL test always rejects less, while the LR test is ranking in the middle this time. When detrending, the LR test in some cases rejects more often than the Rec^{τ} test but this can be attributed to the higher empirical rejection frequencies under the null. When looking at the relative frequencies of chosen ranks for $\Theta = 0$, in all but one case is the Rec^{τ} test better than the two other single tests; the combination procedure is marginally better than all three single tests. In the case where Θ is a nonzero matrix as indicated above, the overall image changes, but not essentially. Now, the superiority of the Rec^{τ} test is not so clear-cut anymore. When focusing on rank selection, the Rec^{τ} test is better than the other two single tests for the case with a constant, but the LR beats the Rec^{τ} test in the case of a constant and a trend, at least for $r = 1$ and $r = 2$. The combination procedure is always better in selecting the correct rank. The choice of nonzero Θ does not seem crucial as the results are very similar for both specifications.

Let us now examine the case with nonzero initial value, i.e. $\gamma = 5$ in our simulations. For the empirical rejection frequencies under the null, the picture is similar to the case with $\gamma = 0$. Concerning the empirical rejection frequencies under the alternative, the SL test is the most affected of the three single tests. There is a large drop in its empirical rejection frequency under the alternative and its ability to select the correct cointegration rank for r=1 and r=2. By construction, the combination procedure is affected when one of the three tests shows a decrease in its ability to select the correct cointegration rank and is therefore not always superior to the three single tests in this specification. But it still is a valid choice if the underlying data generating process is not known. Consistent with the findings of Meng et al. (2013) for the univariate case, the power of our recursively adjusted procedure compared to the LR test drops when the initial value is large.[11] This is especially the case for $\Theta = 0$. When Θ is a nonzero matrix, the difference is much smaller, which can be explained by the fact that the overall power is different for different Θ.

To sum up our findings, the Rec^{τ} test is superior to its competitors in terms of empirical rejection frequencies under both the null and the alternative in the majority of the studied cases. Similarly, the combination procedure to estimate the cointegration rank performs well. In most cases, the LR test performs worse than the Rec^{τ} test, but it seems

[10] Such behavior of the LR test is not surprising: it has been signaled in the literature before.

[11] It may be worth mentioning that the extent of this finding also depends on the short-run dynamics. In particular, for $\phi = 0$, the power of the recursively adjusted procedure falls much slower and is still superior to that of the LR test for the parameter values of this Monte Carlo study.

Table 2: Relative rejection frequencies and frequencies of ranks selected for $\tau = 1$ and $\tau = (1, s)'$

			Rejection						Choice							
			constant			constant & trend			constant				constant & trend			
		r_0	Rec^τ	LR	SL	Rec^τ	LR	SL	Rec^τ	LR	SL	$Comb$	Rec^τ	LR	SL	$Comb$
$\gamma = 0$	r=0	0	7.1	8.8	7.3	5.8	9.3	7.0	92.9	91.3	92.7	93.4	94.2	90.7	93.0	94.5
		1	-	-	-	-	-	-	6.5	7.7	6.5	6.1	5.5	8.6	6.5	5.4
		2	-	-	-	-	-	-	0.5	0.9	0.7	0.5	0.3	0.7	0.4	0.2
	r=1	0	94.3	90.5	82.1	82.9	84.9	76.4	5.7	9.5	17.9	7.2	17.1	15.1	23.6	15.1
		1	5.8	6.6	6.1	4.4	6.3	5.5	88.5	83.9	76.2	87.6	78.5	78.6	71.1	81.4
		2	-	-	-	-	-	-	5.2	6.1	5.2	4.9	4.0	5.9	5.0	3.3
	r=2	0	99.9	99.9	99.6	99.3	99.8	98.5	0.1	0.1	0.4	0.1	0.7	0.2	1.5	0.4
		1	98.0	95.5	84.2	88.6	89.0	78.8	1.9	4.4	15.5	2.9	10.7	10.8	20.1	10.1
		2	5.9	5.5	6.0	4.6	5.4	5.4	92.1	90.0	78.7	92.2	84.0	83.6	73.3	86.1
$\gamma = 5$	r=0	0	7.1	8.8	7.3	5.8	9.3	7.0	92.9	91.3	92.7	93.4	94.2	90.7	93.0	94.5
		1	-	-	-	-	-	-	6.5	7.7	6.5	6.1	5.5	8.6	6.5	5.4
		2	-	-	-	-	-	-	0.5	0.9	0.7	0.5	0.3	0.7	0.4	0.2
	r=1	0	79.9	94.4	26.2	72.4	90.6	29.4	20.1	5.6	73.8	20.9	27.6	9.4	70.6	26.8
		1	5.1	6.8	6.4	4.0	6.7	5.2	74.8	87.5	22.8	74.0	68.5	83.9	25.9	69.7
		2	-	-	-	-	-	-	4.5	6.2	3.2	4.7	3.6	6.3	3.3	3.4
	r=2	0	99.7	100.0	95.9	98.2	99.9	85.7	0.3	0.0	4.1	0.3	1.8	0.1	14.4	1.3
		1	83.7	96.8	42.6	79.2	92.6	43.6	16.0	3.2	54.9	14.3	18.9	7.3	45.8	17.4
		2	5.2	5.7	5.8	4.1	5.3	5.0	78.5	91.0	38.1	81.2	75.2	87.3	36.6	78.2

Note: The data generating process is given by equations (11)–(14) in the paper, with $\psi = 0.5$, Θ a conformable matrix with zero elements, and $\phi = -0.5$. The series are prewhitened using a number of lagged differences indicated by the AIC. *LR* denotes the test with adjustment for an unrestricted constant, or for an unrestricted constant and an unrestricted linear trend. *SL* denotes the test with GLS adjustment for an unrestricted constant, or or for an unrestricted constant and an unrestricted linear trend. All tests are of $r = r_0$ against $r > r_0$. *Comb* denotes the combination procedures for the mean-adjusted or mean and trend-adjusted case. For further details, see the paper.

Table 3: Relative rejection frequencies and frequencies of ranks selected for $\tau = 1$ and $\tau = (1, s)'$, $\Theta = (0.8, 0.4)$

			Rejection						Choice							
			constant			constant & trend			constant				constant & trend			
		r_0	Rec^τ	LR	SL	Rec^τ	LR	SL	Rec^τ	LR	SL	$Comb$	Rec^τ	LR	SL	$Comb$
$\gamma = 0$	r=0	0	7.1	8.8	7.3	5.8	9.3	7.0	92.9	91.3	92.7	93.4	94.2	90.7	93.0	94.5
		1	-	-	-	-	-	-	6.5	7.7	6.5	6.1	5.5	8.6	6.5	5.4
		2	-	-	-	-	-	-	0.5	0.9	0.7	0.5	0.3	0.7	0.4	0.2
	r=1	0	99.7	99.9	97.7	96.6	99.8	97.3	0.3	0.1	2.3	0.2	3.4	0.2	2.7	0.8
		1	6.3	7.6	6.7	5.3	7.5	5.4	93.4	92.3	91.0	94.2	91.3	92.3	91.9	94.9
		2	-	-	-	-	-	-	5.7	6.9	6.1	5.3	5.0	7.1	5.2	4.2
	r=2	0	100.0	100.0	99.9	99.2	100.0	99.7	0.0	0.0	0.1	0.0	0.8	0.0	0.3	0.1
		1	98.9	98.2	87.4	93.4	95.8	85.3	1.0	1.8	12.5	1.3	5.9	4.2	14.4	4.6
		2	5.6	5.7	5.6	4.4	5.5	4.7	93.3	92.5	82.3	94.5	89.0	90.3	80.9	92.4
$\gamma = 5$	r=0	0	7.1	8.8	7.3	5.8	9.3	7.0	92.9	91.3	92.7	93.4	94.2	90.7	93.0	94.5
		1	-	-	-	-	-	-	6.5	7.7	6.5	6.1	5.5	8.6	6.5	5.4
		2	-	-	-	-	-	-	0.5	0.9	0.7	0.5	0.3	0.7	0.4	0.2
	r=1	0	96.1	99.9	69.2	92.0	99.8	74.3	3.9	0.1	30.9	2.8	8.1	0.1	25.7	3.8
		1	5.7	7.8	6.7	4.3	7.8	5.4	90.3	92.1	63.4	91.9	87.7	92.0	69.4	92.4
		2	-	-	-	-	-	-	5.2	7.2	5.2	4.9	4.1	7.4	4.7	3.8
	r=2	0	99.5	100.0	99.2	97.7	100.0	96.5	0.5	0.0	0.8	0.2	2.3	0.0	3.5	0.7
		1	92.0	98.2	74.3	86.2	96.1	73.8	7.5	1.8	25.4	4.5	11.5	3.9	23.8	8.0
		2	5.1	5.7	5.6	3.6	5.8	4.2	86.9	92.5	69.6	91.9	82.6	90.3	69.1	88.7

Note: See Table 2.

Table 4: Relative rejection frequencies and frequencies of ranks selected for $\tau = 1$ and $\tau = (1, s)'$, $\Theta = (-0.8, 0.4)$

			Rejection						Choice							
			constant			constant & trend			constant				constant & trend			
		r_0	Rec^τ	LR	SL	Rec^τ	LR	SL	Rec^τ	LR	SL	$Comb$	Rec^τ	LR	SL	$Comb$
		0	7.1	8.8	7.3	5.8	9.3	7.0	92.9	91.3	92.7	93.4	94.2	90.7	93.0	94.5
	r=0	1	-	-	-	-	-	-	6.5	7.7	6.5	6.1	5.5	8.6	6.5	5.4
		2	-	-	-	-	-	-	0.5	0.9	0.7	0.5	0.3	0.7	0.4	0.2
		0	99.7	99.9	97.5	96.6	99.8	97.5	0.3	0.1	2.5	0.2	3.4	0.2	2.5	0.8
$\gamma = 0$	r=1	1	6.2	8.1	6.8	5.3	7.9	5.6	93.5	91.8	90.9	94.1	91.4	91.8	91.8	94.9
		2	-	-	-	-	-	-	5.6	7.3	6.1	5.3	4.9	7.3	5.4	4.3
		0	100.0	100.0	99.9	99.3	100.0	99.8	0.0	0.0	0.1	0.0	0.8	0.0	0.2	0.1
	r=2	1	99.1	98.3	87.7	93.3	96.1	85.3	0.9	1.7	12.2	1.2	5.9	3.9	14.5	4.5
		2	5.2	5.9	5.6	4.8	5.7	4.9	93.9	92.4	82.5	94.5	88.5	90.4	80.6	92.0
		0	7.1	8.8	7.3	5.8	9.3	7.0	92.9	91.3	92.7	93.4	94.2	90.7	93.0	94.5
	r=0	1	-	-	-	-	-	-	6.5	7.7	6.5	6.1	5.5	8.6	6.5	5.4
		2	-	-	-	-	-	-	0.5	0.9	0.7	0.5	0.3	0.7	0.4	0.2
		0	96.3	99.9	68.7	92.2	99.8	74.0	3.7	0.1	31.3	2.8	7.8	0.2	26.0	4.0
$\gamma = 5$	r=1	1	5.3	8.0	7.1	4.3	8.1	5.8	91.0	91.9	62.7	92.0	87.8	91.7	68.6	91.9
		2	-	-	-	-	-	-	4.8	7.2	5.5	4.9	4.0	7.5	5.1	4.0
		0	100.0	100.0	99.9	98.7	100.0	98.7	0.1	0.0	0.1	0.0	1.3	0.0	1.3	0.3
	r=2	1	86.8	99.3	18.1	83.7	98.3	20.9	13.2	0.7	81.8	12.9	15.0	1.7	78.0	14.7
		2	4.8	6.1	5.5	4.4	6.2	4.3	82.0	93.2	16.9	83.5	79.4	92.1	19.1	82.2

Note: See Table 2.

to be more robust to an increase in the initial value, at least in terms of power. The SL test is the most affected one given an increase of the initial value: in some specifications it shows a dramatic drop in empirical rejection frequency under the alternative for $\gamma = 5$. In a nutshell, the Rec^{τ} test is superior to the LR and SL tests in terms of empirical rejection frequencies both the null and the alternative in many cases, but not in all. No single test dominates the other two uniformly. The combination procedure consisting of choosing as cointegration rank the rank indicated by the average of chosen ranks[12] appears to be the better alternative when estimating the cointegration rank.

5 Conclusions

Recursive demeaning or detrending have been successfully used in time series inference, e.g. to reduce the bias of autoregressive parameter estimates or to increase the power of unit root tests. Moving on to multivariate, potentially nonstationary time series, we addressed the question of whether such improvements are achievable when inferencing about the cointegrating rank as well.

The paper therefore analyzed the asymptotic and small-sample behavior of so-called Rec^{τ} statistics for testing the cointegration rank of cointegrated VARs. They arise by recursively adjusting the lagged levels for deterministic components prior to the application of the test.

To do so, the paper also analyzed the effect of recursive removal of general trend components (recursive adjustment) on time series. Two cases were discussed: first, the series were assumed to be weakly stationary (corresponding to the equilibrium error), and second, the series were assumed to be nonstationary and to converge weakly to a continuous-time process when suitably normalized (corresponding to the common trends). In the first case, the effect of recursive adjustment was shown to vanish asymptotically, while, in the second, the limiting process was shown to be affected in a predictable manner. Thus, the null limiting distributions of the cointegration rank tests can be given in terms of recursively adjusted Brownian motions.

Focusing on the empirically relevant cases of a constant and a trend, it was found by Monte Carlo simulations that the proposed Rec^{τ} test procedure performs well in finite samples, and is often superior to its competitors in terms of empirical rejection frequencies under the null and under the alternative. No test was uniformly better, however. When focusing on the estimation of the cointegration rank, a very simple procedure combining the outcomes of the Rec^{τ}, the LR and the SL tests performs very promising.

[12] In our experience, this was practically always a two-out-of-three decision.

Appendix

Proof of Proposition 1

To derive the asymptotic behavior of the difference between $\widetilde{\mathbf{w}}_t$ and $\mathbf{w}_t$, note that

$$\widetilde{\mathbf{w}}_t = \mathbf{w}_t - \left(\frac{1}{t}\sum_{j=1}^{t}\mathbf{w}_j\left(D_t^{-1}\mathbf{f}(j)\right)'\right)\left(\frac{1}{t}\sum_{j=1}^{t}\left(D_t^{-1}\mathbf{f}(j)\right)\left(D_t^{-1}\mathbf{f}(j)\right)'\right)^{-1} D_t^{-1}\mathbf{f}(t)$$

with D_t^{-1} defined in Assumption 3. Let then

$$s_{j,t} = \left(D_t^{-1}\mathbf{f}(j)\right)'\left(\frac{1}{t}\sum_{j=1}^{t}\left(D_t^{-1}\mathbf{f}(j)\right)\left(D_t^{-1}\mathbf{f}(j)\right)'\right)^{-1} D_t^{-1}\mathbf{f}(t)$$

and note that, for $s \in [0,1]$,

$$s_{[st],t} \Rightarrow \boldsymbol{\tau}(s)'\left(\int_0^1 \boldsymbol{\tau}(s)\boldsymbol{\tau}(s)'\mathrm{d}s\right)^{-1}\boldsymbol{\tau}(1)$$

as $t \to \infty$ such that

$$\sup_{s\in[0,1]}\left|s_{[st],t} - \boldsymbol{\tau}(s)'\left(\int_0^1 \boldsymbol{\tau}(s)\boldsymbol{\tau}(s)'\mathrm{d}s\right)^{-1}\boldsymbol{\tau}(1)\right| \to 0;$$

the sequence of suprema being convergent, it must be bounded. Moreover, the matrix $\int_0^1 \boldsymbol{\tau}(s)\boldsymbol{\tau}(s)'\mathrm{d}s$ is invertible according to Assumption 3. Hence $s_{[st],t}$ converges as $t \to \infty$ to a continuous function, which, being defined on the compact interval $[0,1]$, is also bounded. This implies that $s_{j,t}$ is itself bounded (uniformly in s):

$$\begin{aligned}
\sup_{s\in[0,1]}\left|s_{[st],t}\right| &\le \sup_{s\in[0,1]}\left|s_{[st],t} - \boldsymbol{\tau}(s)'\left(\int_0^1 \boldsymbol{\tau}(s)\boldsymbol{\tau}(s)'\mathrm{d}s\right)^{-1}\boldsymbol{\tau}(1)\right| \\
&\quad + \sup_{s\in[0,1]}\left|\boldsymbol{\tau}(s)'\left(\int_0^1 \boldsymbol{\tau}(s)\boldsymbol{\tau}(s)'\mathrm{d}s\right)^{-1}\boldsymbol{\tau}(1)\right| \\
&\le C^*
\end{aligned}$$

Now, $s_{j,t}$ is a scalar, so it holds that

$$\mathrm{Cov}\left(\widetilde{\mathbf{w}}_t - \mathbf{w}_t\right) = \mathrm{Cov}\left(\frac{1}{t}\sum_{j=1}^{t}\mathbf{w}_j s_{j,t}\right) = \frac{1}{t^2}\sum_{i=1}^{t}\sum_{j=1}^{t} s_{i,t}s_{j,t}\,\mathrm{E}\left(\mathbf{w}_i\mathbf{w}_j'\right)$$

and thus that

$$\|\mathrm{Cov}\left(\widetilde{\mathbf{w}}_t - \mathbf{w}_t\right)\| \leq \frac{1}{t^2}\sum_{i=1}^{t}\sum_{j=1}^{t}|s_{i,t}s_{j,t}|\,\|\Gamma_{i-j}\| \leq \sup_{t\leq T, j\leq t} s_{j,t}^2 \frac{1}{t^2}\sum_{i=1}^{t}\sum_{j=1}^{t}\|\Gamma_{i-j}\|\,.$$

By using the fact that the double sum on the r.h.s. is bounded by the square of the sum in the assumptions of the proposition, one obtains

$$\|\mathrm{Cov}\left(\widetilde{\mathbf{w}}_t - \mathbf{w}_t\right)\| \leq (\phi(t))^2 \leq C^*\phi(t)$$

for some C^* not depending on t (we used the fact that $\phi(t) \to 0$ implying that there exists a finite t_0 for which $|\phi(t)| < 1$ for all $t > t_0$, and hence $(\phi(t))^2 \leq \phi(t)$ for all $t > t_0$). Note further that, due to the Cauchy-Schwarz inequality, it holds that

$$\mathrm{E}\left(\|\widetilde{\mathbf{w}}_t - \mathbf{w}_t\|^2\right) = \mathrm{tr}\,\mathrm{Cov}\left(\widetilde{\mathbf{w}}_t - \mathbf{w}_t\right) \leq K\,\|\mathrm{Cov}\left(\widetilde{\mathbf{w}}_t - \mathbf{w}_t\right)\|\,\|I_K\|\,,$$

which establishes the result.

Proof of Proposition 2

We begin by establishing continuity at $s = 0$ of $\widetilde{\mathbf{V}}^{\boldsymbol{\tau}}(s)$. Let $D_{\boldsymbol{\tau}} = \mathrm{diag}(\boldsymbol{\tau}(s))$ and denote by $\boldsymbol{\iota}$ a vector of ones. Write now

$$\begin{aligned}
&\left(\int_0^s \left(\mathbf{V}(r) - \mathbf{V}(0)\right)\boldsymbol{\tau}(r)'\mathrm{d}r\right)\left(\int_0^s \boldsymbol{\tau}(r)\boldsymbol{\tau}(r)'\mathrm{d}r\right)^{-1}\boldsymbol{\tau}(s)\\
&= \left(\frac{1}{s}\int_0^s \left(\mathbf{V}(r) - \mathbf{V}(0)\right)\left(D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\tau}(r)\right)'\mathrm{d}r\right)\left(\frac{1}{s}\int_0^s D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\tau}(r)\left(D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\tau}(r)\right)'\mathrm{d}r\right)^{-1}D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\tau}(s).
\end{aligned}$$

Since $\boldsymbol{\tau}$ is continuous on $[0,1]$, $D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\tau}(r) \to \boldsymbol{\iota}$ as $s \to 0$ $\forall 0 < r < s$. Hence, $D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\tau}(r) \to \mathbf{1} = O(\boldsymbol{\iota})$. Moreover, $\int_0^s D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\tau}(r)\left(D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\tau}(r)\right)'\mathrm{d}r$ is invertible for any $s > 0$ and, using the first Mean Value Theorem for integration element-wise, it can be seen that $\frac{1}{s}\int_0^s D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\tau}(r)\left(D_{\boldsymbol{\tau}}^{-1}\boldsymbol{\tau}(r)\right)'\mathrm{d}r$ is bounded away from 0. Summing up,

$$\left(\int_0^s \left(\mathbf{V}(r) - \mathbf{V}(0)\right)\boldsymbol{\tau}(r)'\mathrm{d}r\right)\left(\int_0^s \boldsymbol{\tau}(r)\boldsymbol{\tau}(r)'\mathrm{d}r\right)^{-1}\boldsymbol{\tau}(s) = O\left(\frac{1}{s}\int_0^s \left(\mathbf{V}(r) - \mathbf{V}(0)\right)\mathrm{d}r\right)$$

along each path of $\mathbf{V}(s)$. Applying the first Mean Value Theorem for integration again, it follows for each element $1 \leq k \leq K$ that $\exists 0 < \xi_k < s$ such that

$$\frac{1}{s}\int_0^s \left(V_k(r) - V_k(0)\right)\boldsymbol{\iota}'\mathrm{d}r = \left(V_k(\xi_k) - V_k(0)\right)\boldsymbol{\iota}'.$$

With the continuity of $\mathbf{V}(s)$ at the origin, one has as $s \to 0$ (and thus $\xi_k \to 0$) that

$$\widetilde{\mathbf{V}}^{\boldsymbol{\tau}}(s) \to \mathbf{0}$$

almost surely as required.

To establish the desired weak convergence, assume w.l.o.g. that $\tau_l \geq 0\ \forall l$ and consider the functional

$$\mathcal{F}\left[\mathbf{V}\right] = \begin{cases} \left(\mathbf{V}(s) - \mathbf{V}(0)\right) - \left(\int_0^s \left(\mathbf{V}(r) - \mathbf{V}(0)\right) \boldsymbol{\tau}\left(r\right)' \mathrm{d}r\right) \left(\int_0^s \boldsymbol{\tau}\left(r\right) \boldsymbol{\tau}\left(r\right)' \mathrm{d}r\right)^{-1} \boldsymbol{\tau}\left(s\right), & s > 0 \\ \mathbf{0} & s = 0. \end{cases}$$

The result follows with the continuous mapping theorem [CMT] if $\mathcal{F}\left[\cdot\right]$ is a continuous functional, i.e.

$$\sup_{s\in[0,1]} \left\|\mathbf{V}_1\left(s\right) - \mathbf{V}_2\left(s\right)\right\| \to 0 \quad \text{implies} \quad \sup_{s\in[0,1]} \left\|\mathcal{F}\left[\mathbf{V}_1\right]\left(s\right) - \mathcal{F}\left[\mathbf{V}_2\right]\left(s\right)\right\| \to 0.$$

Write with $\mathbf{V}(0)$ cancelling out

$$\begin{aligned} \mathcal{F}\left[\mathbf{V}_1\right] - \mathcal{F}\left[\mathbf{V}_2\right] = {} & \mathbf{V}_1\left(s\right) - \mathbf{V}_2\left(s\right) \\ & - \left(\int_0^s \left(\mathbf{V}_1\left(s\right) - \mathbf{V}_2\left(s\right)\right) \boldsymbol{\tau}\left(r\right)' \mathrm{d}r\right) \left(\int_0^s \boldsymbol{\tau}\left(r\right) \boldsymbol{\tau}\left(r\right)' \mathrm{d}r\right)^{-1} \boldsymbol{\tau}\left(s\right). \end{aligned}$$

Then, $\sup_{s\in[0,1]} \left\|\mathcal{F}\left[\mathbf{V}_1\right]\left(s\right) - \mathcal{F}\left[\mathbf{V}_2\right]\left(s\right)\right\|$ is bounded by

$$\begin{aligned} & \sup_{s\in[0,1]} \left\|\mathbf{V}_1\left(s\right) - \mathbf{V}_2\left(s\right)\right\| \\ & + \left(\int_0^s \sup_{s\in[0,1]} \left|\mathbf{V}_1\left(s\right) - \mathbf{V}_2\left(s\right)\right| \boldsymbol{\tau}\left(r\right)' \mathrm{d}r\right) \left(\int_0^s \boldsymbol{\tau}\left(r\right) \boldsymbol{\tau}\left(r\right)' \mathrm{d}r\right)^{-1} \boldsymbol{\tau}\left(s\right), \end{aligned}$$

where $\sup_{s\in[0,1]} \left|\mathbf{V}_1\left(s\right) - \mathbf{V}_2\left(s\right)\right|$ is the vector $\left\{\sup_{s\in[0,1]} \left|V_{1,l}\left(s\right) - V_{2,l}\left(s\right)\right|\right\}_{1\leq k\leq K}$. For any $s \in \left[q, 1\right]$, the second summand on the r.h.s. vanishes when

$$\sup_{s\in[q,1]} \left|V_{1,l}\left(s\right) - V_{2,l}\left(s\right)\right| \leq \sup_{s\in[q,1]} \left\|\mathbf{V}_1\left(s\right) - \mathbf{V}_2\left(s\right)\right\| \leq \sup_{s\in[0,1]} \left\|\mathbf{V}_1\left(s\right) - \mathbf{V}_2\left(s\right)\right\| \to \mathbf{0}.$$

So we only have to show that the property extends to $s \in \left[0, q\right]$. This is indeed the case if

$$\left(\int_0^s \boldsymbol{\iota}\boldsymbol{\tau}\left(r\right)' \mathrm{d}r\right) \left(\int_0^s \boldsymbol{\tau}\left(r\right) \boldsymbol{\tau}\left(r\right)' \mathrm{d}r\right)^{-1} \boldsymbol{\tau}\left(s\right) = O\left(1\right)$$

as $s \to 0$. In turn, the boundedness is implicit from the proof of continuity of $\mathcal{F}\left[\mathbf{V}\right]\left(s\right)$ at 0 and the result follows.

We now state and prove two preliminary lemmas required for the proof of Proposition 3.

Lemma 2

1. *Under Assumptions 1, 2, and 3, it holds as* $T \to \infty$ *that* $\sup_{t \le T} \left| \breve{\Delta \mathbf{y}}_t - \Delta \mathbf{x}_t \right| = \sup_{t \le T} \left| \breve{\Delta \mathbf{x}}_t - \Delta \mathbf{x}_t \right| = O_p(T^{-0.5})$.
2. *Let* $\mathbf{v}_t$ *be some I(1) vector process, possibly cointegrated, such that* $T^{-0.5}\mathbf{v}_{[sT]} \Rightarrow \mathbf{B}^*(s)$, $s \in [0,1]$, *for some Brownian motion* $\mathbf{B}^*$ *with suitable covariance matrix. Then,* $\sum_{t=p+1}^{T} \widetilde{\mathbf{v}}_{t-1} \Delta \mathbf{v}_{t-j}' = O_p(T)$.

Proof of Lemma 2

1. Under Assumptions 1, 2, and 3,

$$\sqrt{T}\left(\sum_{t=2}^{T} \Delta \mathbf{y}_t \Delta \mathbf{f}(t)'\right)\left(\sum_{t=2}^{T} \Delta \mathbf{f}(t) \Delta \mathbf{f}(t)'\right)^{-1} \Delta \mathbf{f}(t)$$
$$\Rightarrow \left(\int_0^1 \mathrm{d}\mathbf{V}(r)\, \dot{\boldsymbol{\tau}}(r)'\right)\left(\int_0^1 \dot{\boldsymbol{\tau}}(r)\dot{\boldsymbol{\tau}}(r)'\mathrm{d}r\right)^{-1} \dot{\boldsymbol{\tau}}(s)$$

(see the arguments leading to Equation (19) in the proof of Lemma 3.5 below). Since $\dot{\boldsymbol{\tau}}(s)$ is bounded on $[0,1]$, the r.h.s. is $O_p(1)$ uniformly, and we have that

$$\sup_{t \le T} \left| \breve{\Delta \mathbf{y}}_t - \Delta \mathbf{x}_t \right| = \sup_{t \le T} \left| \left(\sum_{t=2}^{T} \Delta \mathbf{y}_t \Delta \mathbf{f}(t)'\right)\left(\sum_{t=2}^{T} \Delta \mathbf{f}(t) \Delta \mathbf{f}(t)'\right)^{-1} \Delta \mathbf{f}(t) \right| = O_p(T^{-0.5})$$

as required.

2. The behavior of sample cross-product moments of lagged levels and lagged differences is well-understood in the case of no detrending; see e.g. Phillips and Durlauf (1986) or Phillips (1988). Concretely, $\sum_{t=p+1}^{T} \mathbf{v}_{t-1} \Delta \mathbf{v}_{t-j}' = O_p(T)$, $1 \le j \le p$ under fairly general conditions covering our assumptions. For recursively adjusted levels, it holds that

$$\sum_{t=p+1}^{T} \widetilde{\mathbf{v}}_{t-1} \Delta \mathbf{v}_{t-j}' = \sum_{t=p+1}^{T} \mathbf{v}_t \Delta \mathbf{v}_{t-j}' - \sum_{t=p+1}^{T} \left(\sum_{j=1}^{t} \mathbf{v}_j \mathbf{f}(j)'\right)\left(\sum_{j=1}^{t} \mathbf{f}(j)\mathbf{f}(j)'\right)^{-1} \mathbf{f}(t) \Delta \mathbf{v}_{t-j}'.$$

The first summand on the r.h.s. is $O_p(1)$; by using the fact that $\mathbf{v}_{t-1} = O_p(T^{0.5})$ uniformly together with arguments similar to those used in the proof of Propositions 1 and 2, one can conclude that the second summand on the r.h.s. behaves like $T^{0.5} \sum_{t=p+1}^{T} \boldsymbol{\tau}([sT]) \Delta \mathbf{v}_{t-j}$.

But $T^{-0.5}\sum_{t=p+1}^{T}\boldsymbol{\tau}([sT])\Delta\mathbf{v}_{t-j} \Rightarrow \int_0^1 \boldsymbol{\tau}(s)\mathrm{d}\mathbf{B}^*$, so

$$\sum_{t=p+1}^{T}\widetilde{\mathbf{v}}_{t-1}\Delta\mathbf{v}'_{t-j} = O_p(T)$$

as required.

Lemma 3 *(This is the analog of Lemma 10.3 of Johansen, 1995.) Let $\boldsymbol{\beta}_\perp$ be an orthogonal complement of $\boldsymbol{\beta}$, $\overline{\boldsymbol{\gamma}} = \boldsymbol{\beta}_\perp\left(\boldsymbol{\beta}'_\perp\boldsymbol{\beta}_\perp\right)^{-1}$, as well as $B_T = \overline{\boldsymbol{\gamma}}$, to maintain the original notation of Johansen. Define also*

$$\mathrm{Cov}\left(\begin{array}{c}\Delta\mathbf{x}_t\\ \boldsymbol{\beta}'\mathbf{x}_{t-1}\end{array}\middle|\Delta\mathbf{x}_{t-1},\ldots,\Delta\mathbf{x}_{t-p}\right) = \left(\begin{array}{cc}\Sigma_{00} & \Sigma_{0\boldsymbol{\beta}}\\ \Sigma_{\beta 0} & \Sigma_{\boldsymbol{\beta}\boldsymbol{\beta}}\end{array}\right).$$

It then holds as $T\to\infty$ that

1. $S_{00} \xrightarrow{p} \Sigma_{00}$,
2. $\boldsymbol{\beta}'S_{11}\boldsymbol{\beta} \xrightarrow{p} \Sigma_{\boldsymbol{\beta}\boldsymbol{\beta}}$,
3. $\boldsymbol{\beta}'S_{10} \xrightarrow{p} \Sigma_{\boldsymbol{\beta}0}$,
4. $T^{-1}B'_T S_{11}B_T \xrightarrow{d} \int_0^1 \widetilde{\mathbf{B}}^{\boldsymbol{\tau}}(s)\left(\widetilde{\mathbf{B}}^{\boldsymbol{\tau}}(s)\right)'\mathrm{d}s$,
5. $B'_T S_{10}\boldsymbol{\alpha}_\perp \xrightarrow{d} \int_0^1 \widetilde{\mathbf{B}}^{\boldsymbol{\tau}}(s)\left(\mathrm{d}\breve{\mathbf{V}}^{\dot{\boldsymbol{\tau}}}(s)\right)'\boldsymbol{\alpha}_\perp$,
6. $B'_T S_{11}\boldsymbol{\beta} = O_p(1)$,

where $\mathbf{V}$ is a K-dimensional Brownian motion with covariance matrix $\mathrm{Cov}(\mathbf{V}) = \Sigma_{\boldsymbol{\varepsilon}}$, $\mathrm{d}\breve{\mathbf{V}}^{\dot{\boldsymbol{\tau}}}$ is the differential defined in Proposition 3, and $\widetilde{\mathbf{B}}^{\boldsymbol{\tau}}$ is the recursively adjusted $(K-r)$-dimensional Brownian motion $\mathbf{B}$ for which $\mathrm{Cov}(\mathbf{B}) = \overline{\boldsymbol{\gamma}}\Xi\Sigma_{\boldsymbol{\varepsilon}}\Xi'\overline{\boldsymbol{\gamma}}'$.

Proof of Lemma 3

1. Let $\mathbf{z}_{t-1} = \left((\Delta\mathbf{x}_{t-1})',\ldots,(\Delta\mathbf{x}_{t-p})'\right)'$. It holds thanks to standard OLS algebra that

$$\mathbf{r}_{0t} = \breve{\Delta\mathbf{y}}_t - \frac{1}{T}\sum_{t=p+1}^{T}\breve{\Delta\mathbf{y}}_t\breve{\mathbf{z}}'_{t-1}\left(\frac{1}{T}\sum_{t=p+1}^{T}\breve{\mathbf{z}}_{t-1}\breve{\mathbf{z}}'_{t-1}\right)^{-1}\breve{\mathbf{z}}_{t-1}.$$

Recall, $\sup_{t\le T}\left\|\breve{\Delta\mathbf{y}}_t - \Delta\mathbf{x}_t\right\| = O_p\left(T^{-0.5}\right)$ so it follows that $\sup_{t\le T}\left\|\breve{\mathbf{z}}_{t-1} - \mathbf{z}_{t-1}\right\| = O_p\left(T^{-0.5}\right)$.

Then,

$$\begin{aligned}
\frac{1}{T}\sum_{t=p+1}^{T}\breve{\mathbf{z}}_{t-1}\breve{\mathbf{z}}_{t-1}' &= \frac{1}{T}\sum_{t=p+1}^{T}\left(\mathbf{z}_{t-1}+\breve{\mathbf{z}}_{t-1}-\mathbf{z}_{t-1}\right)\left(\mathbf{z}_{t-1}+\breve{\mathbf{z}}_{t-1}-\mathbf{z}_{t-1}\right)' \\
&= \frac{1}{T}\sum_{t=p+1}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}' + \frac{1}{T}\sum_{t=p+1}^{T}\mathbf{z}_{t-1}\left(\breve{\mathbf{z}}_{t-1}-\mathbf{z}_{t-1}\right)' \\
&\quad + \frac{1}{T}\sum_{t=p+1}^{T}\left(\breve{\mathbf{z}}_{t-1}-\mathbf{z}_{t-1}\right)\mathbf{z}_{t-1}' + \frac{1}{T}\sum_{t=p+1}^{T}\left(\breve{\mathbf{z}}_{t-1}-\mathbf{z}_{t-1}\right)\left(\breve{\mathbf{z}}_{t-1}-\mathbf{z}_{t-1}\right)'
\end{aligned}$$

so, for a suitable constant C^{*},

$$\begin{aligned}
&\left\|\frac{1}{T}\sum_{t=p+1}^{T}\breve{\mathbf{z}}_{t-1}\breve{\mathbf{z}}_{t-1}' - \frac{1}{T}\sum_{t=p+1}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\right\| \\
&\qquad \leq C^{*}\left(\sup_{t}\left\|\breve{\mathbf{z}}_{t-1}-\mathbf{z}_{t-1}\right\|\cdot\frac{1}{T}\sum_{t=p+1}^{T}\left\|\mathbf{z}_{t-1}\right\| + \left(\sup_{t}\left\|\breve{\mathbf{z}}_{t-1}-\mathbf{z}_{t-1}\right\|\right)^{2}\right).
\end{aligned}$$

Given the moment conditions on $\mathbf{z}_{t-1}$, $\frac{1}{T}\sum_{t=p+1}^{T}\left\|\mathbf{z}_{t-1}\right\| = O_p(1)$ so

$$\frac{1}{T}\sum_{t=p+1}^{T}\breve{\mathbf{z}}_{t-1}\breve{\mathbf{z}}_{t-1}' = \frac{1}{T}\sum_{t=p+1}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}' + O_p\left(T^{-0.5}\right)$$

Similarly,

$$\frac{1}{T}\sum_{t=p+1}^{T}\breve{\Delta\mathbf{y}}_{t-1}\breve{\mathbf{z}}_{t-1}' = \frac{1}{T}\sum_{t=p+1}^{T}\Delta\mathbf{x}_{t-1}\mathbf{z}_{t-1}' + O_p\left(T^{-0.5}\right).$$

Hence, using $\sup_{t\leq T}\left\|\breve{\mathbf{z}}_{t-1}-\mathbf{z}_{t-1}\right\| = O_p\left(T^{-0.5}\right)$ again, one immediately has that

$$\sup_{t\leq T}\left\|\mathbf{r}_{0t} - \Delta\mathbf{x}_t - \frac{1}{T}\sum_{t=p+1}^{T}\Delta\mathbf{x}_t\mathbf{z}_{t-1}'\left(\frac{1}{T}\sum_{t=p+1}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\right)^{-1}\mathbf{z}_{t-1}\right\| = O_p(T^{-0.5}),$$

or

$$\mathbf{r}_{0t} = \Delta\mathbf{x}_t - \frac{1}{T}\sum_{t=p+1}^{T}\Delta\mathbf{x}_t\mathbf{z}_{t-1}'\left(\frac{1}{T}\sum_{t=p+1}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\right)^{-1}\mathbf{z}_{t-1} + O_p(T^{-0.5})$$

uniformly. We then have using uniformity of the $O_p(T^{-0.5})$ term that

$$\begin{aligned}
S_{00} &= \frac{1}{T}\sum\Delta\mathbf{x}_{t-1}\Delta\mathbf{x}_{t-1}' \\
&\quad - \frac{1}{T}\sum\Delta\mathbf{x}_{t-1}\mathbf{z}_{t-1}'\left(\frac{1}{T}\sum\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\right)^{-1}\frac{1}{T}\sum\mathbf{z}_{t-1}\Delta\mathbf{x}_{t-1}' + o_p(1)
\end{aligned}$$

such that

$$\begin{aligned} S_{00} &\xrightarrow{p} \mathrm{Cov}\left(\Delta\mathbf{x}_{t-1}\right) - \mathrm{Cov}\left(\Delta\mathbf{x}_{t-1}, \mathbf{z}'_{t-1}\right)\left(\mathrm{Cov}\left(\mathbf{z}_{t-1}\right)\right)^{-1}\mathrm{Cov}\left(\mathbf{z}_{t-1}, \Delta\mathbf{x}'_{t-1}\right) \\ &= \mathrm{Cov}\left(\Delta\mathbf{x}_{t-1}|\mathbf{z}_{t-1}\right) = \Sigma_{00}. \end{aligned}$$

2. We have that

$$\mathbf{r}_{1t} = \widetilde{\mathbf{y}}_{t-1} - \frac{1}{T}\sum_{t=p+1}^{T}\widetilde{\mathbf{y}}_{t-1}\breve{\mathbf{z}}'_{t-1}\left(\frac{1}{T}\sum_{t=p+1}^{T}\breve{\mathbf{z}}_{t-1}\breve{\mathbf{z}}'_{t-1}\right)^{-1}\breve{\mathbf{z}}_{t-1}$$

with $\widetilde{\mathbf{y}}_{t-1} = \widetilde{\mathbf{x}}_{t-1}$. In the direction of $\boldsymbol{\beta}$, $\mathbf{x}_{t-1}$ is I(0), so the assumed moment conditions indicate that $\frac{1}{T}\sum_{t=p+1}^{T}\|\boldsymbol{\beta}'\widetilde{\mathbf{x}}_{t-1}\| = O_p(1)$. It then follows along the lines of the proof of Lemma 2.1 that

$$\boldsymbol{\beta}'\mathbf{r}_{1t} = \boldsymbol{\beta}'\widetilde{\mathbf{x}}_{t-1} - \frac{1}{T}\sum_{t=p+1}^{T}\boldsymbol{\beta}'\widetilde{\mathbf{x}}_{t-1}\mathbf{z}'_{t-1}\left(\frac{1}{T}\sum_{t=p+1}^{T}\mathbf{z}_{t-1}\mathbf{z}'_{t-1}\right)^{-1}\mathbf{z}_{t-1} + O_p(T^{-0.5})$$

uniformly. As a consequence of the uniformity of the $O_p(T^{-0.5})$ term, the sample covariance of $\boldsymbol{\beta}'\mathbf{r}_{1t}$ equals, up to a vanishing term, the sample covariance of $\boldsymbol{\beta}'\widetilde{\mathbf{x}}_{t-1} - \frac{1}{T}\sum_{t=p+1}^{T}\boldsymbol{\beta}'\widetilde{\mathbf{x}}_{t-1}\mathbf{z}'_{t-1}\left(\frac{1}{T}\sum_{t=p+1}^{T}\mathbf{z}_{t-1}\mathbf{z}'_{t-1}\right)^{-1}\mathbf{z}_{t-1}$. So $\boldsymbol{\beta}'S_{11}\boldsymbol{\beta}$ is, up to a vanishing term, the sample covariance of a recursively adjusted I(0) process and it holds along the lines of Example 2 that

$$\begin{aligned} \boldsymbol{\beta}'S_{11}\boldsymbol{\beta} &\xrightarrow{p} \mathrm{Cov}\left(\boldsymbol{\beta}'\mathbf{x}_{t-1}\right) \\ &\qquad - \mathrm{Cov}\left(\boldsymbol{\beta}'\mathbf{x}_{t-1}, \mathbf{z}'_{t-1}\right)\left(\mathrm{Cov}\left(\mathbf{z}_{t-1}\right)\right)^{-1}\mathrm{Cov}\left(\mathbf{z}_{t-1}, \mathbf{x}'_{t-1}\boldsymbol{\beta}\right) \\ &= \mathrm{Cov}\left(\boldsymbol{\beta}'\mathbf{x}_{t-1}|\mathbf{z}_{t-1}\right) \\ &= \Sigma_{\boldsymbol{\beta}\boldsymbol{\beta}}. \end{aligned}$$

3. Analogous to Lemma 3.1 and Lemma 3.2 and omitted.

4. Along $\overline{\boldsymbol{\gamma}}$, $\mathbf{x}_{t-1}$ is I(1); make use of the CMT to establish that

$$\frac{1}{T^{1.5}}\sum_{t=p+1}^{T}\widetilde{\overline{\boldsymbol{\gamma}}'\mathbf{x}}_{t-1} \xrightarrow{d} \int_0^1 \widetilde{\mathbf{B}}^{\boldsymbol{\tau}}(s)\,\mathrm{d}s, \tag{17}$$

and exploit again the fact that $\sup_{t\leq T}\|\breve{\mathbf{z}}_{t-1} - \mathbf{z}_{t-1}\| = O_p(T^{-0.5})$ to conclude that

$$\overline{\boldsymbol{\gamma}}'\mathbf{r}_{1t} = \widetilde{\overline{\boldsymbol{\gamma}}'\mathbf{x}}_{t-1} - \frac{1}{T}\sum_{t=p+1}^{T}\widetilde{\overline{\boldsymbol{\gamma}}'\mathbf{x}}_{t-1}\mathbf{z}'_{t-1}\left(\frac{1}{T}\sum_{t=p+1}^{T}\mathbf{z}_{t-1}\mathbf{z}'_{t-1}\right)^{-1}\mathbf{z}_{t-1} + O_p(1)$$

uniformly. Use now Lemma 2.2 to conclude that

$$\frac{1}{T}\sum_{t=p+1}^{T}\widetilde{\overline{\boldsymbol{\gamma}}'\mathbf{x}}_{t-1}\mathbf{z}'_{t-1} = O_p(1),$$

leading to

$$\sup_{t\leq T}\left\|\overline{\boldsymbol{\gamma}}'\mathbf{r}_{1t} - \widetilde{\overline{\boldsymbol{\gamma}}'\mathbf{x}}_{t-1}\right\| = O_p(1),$$

as well as

$$\sup_{t\leq T}\left\|\frac{1}{\sqrt{T}}\overline{\boldsymbol{\gamma}}'\mathbf{r}_{1t}\frac{1}{\sqrt{T}}\widetilde{\overline{\boldsymbol{\gamma}}'\mathbf{x}}_{t-1}\right\| = O_p(T^{-0.5}). \tag{18}$$

The limit of $T^{-1}B'_TS_{11}B_T$ follows with Equation (18), Proposition 2, and the CMT.

5. Denote by $\breve{\mathbf{x}}_t$ the series $\mathbf{x}_t$ adjusted the usual way for the deterministic components of the differences, by $\breve{\mathbf{r}}_{1t}$ the (unfeasible) residuals from the projection of $\breve{\mathbf{x}}_t$ on the first p lags of $\breve{\Delta\mathbf{y}}_t$, and let $\breve{S}_{11} = \frac{1}{T}\sum_{p+1}^{T}\mathbf{r}_{1t}\breve{\mathbf{r}}'_{1t}$. Then, with $\boldsymbol{\alpha}\boldsymbol{\alpha}'_\perp = 0$,

$$\begin{aligned} B'_TS_{10}\boldsymbol{\alpha}_\perp &= B'_T\left(S_{10} - \breve{S}_{11}\boldsymbol{\beta}\boldsymbol{\alpha}'\right)\boldsymbol{\alpha}_\perp \\ &= \overline{\boldsymbol{\gamma}}'\frac{1}{T}\sum_{p+1}^{T}\mathbf{r}_{1t}\left(\mathbf{r}_{0t} - \boldsymbol{\alpha}\boldsymbol{\beta}'\breve{\mathbf{r}}_{1t}\right)'\boldsymbol{\alpha}_\perp. \end{aligned}$$

Now, $\sup_{t\leq T}\|(\mathbf{r}_{0t} - \boldsymbol{\alpha}\boldsymbol{\beta}'\breve{\mathbf{r}}_{1t}) - \breve{\boldsymbol{\varepsilon}}_t\| = O_p(T^{-0.5})$ since $\mathbf{r}_{0t}$ and $\boldsymbol{\alpha}\boldsymbol{\beta}'\breve{\mathbf{r}}_{1t}$ have been adjusted for deterministics *in the same way*; cf. the proof of Lemma 2.1. Using the argument of Johansen (1995), 2nd display on p. 148, it follows that the $o_p(1)$ term does not affect the asymptotic behavior of $B'_TS_{10}\boldsymbol{\alpha}_\perp$,

$$B'_TS_{10}\boldsymbol{\alpha}_\perp = \frac{1}{T}\sum_{t=p+1}^{T}\widetilde{\overline{\boldsymbol{\gamma}}'\mathbf{x}}_{t-1}\breve{\boldsymbol{\varepsilon}}'_t\boldsymbol{\alpha}_\perp + o_p(1),$$

where $\breve{\boldsymbol{\varepsilon}}_t = \boldsymbol{\varepsilon}_t - \left(\sum_{t=p+1}^{T}\boldsymbol{\varepsilon}_t\Delta\mathbf{f}(t)'\right)\left(\sum_{t=p+1}^{T}\Delta\mathbf{f}(t)\Delta\mathbf{f}(t)'\right)^{-1}\Delta\mathbf{f}(t)$, where $\Delta\mathbf{f}(t) = \mathbf{f}(t) - \mathbf{f}(t-1)$. Hence, $\frac{1}{T}\sum_{t=p+1}^{T}\widetilde{\overline{\boldsymbol{\gamma}}'\mathbf{x}}_{t-1}\breve{\boldsymbol{\varepsilon}}'_t$ can be written as the difference of two terms,

$$\frac{1}{T}\sum_{t=p+1}^{T}\widetilde{\overline{\boldsymbol{\gamma}}'\mathbf{x}}_{t-1}\boldsymbol{\varepsilon}'_t$$

and

$$\begin{aligned} &\frac{1}{T^{1.5}}\sum_{t=p+1}^{T}\widetilde{\overline{\boldsymbol{\gamma}}'\mathbf{x}}_{t-1}\left(TD_T^{-1}\Delta\mathbf{f}(t)\right)'\left(\frac{1}{T}\sum_{t=2}^{T}\left(TD_T^{-1}\Delta\mathbf{f}(t)\right)\left(TD_T^{-1}\Delta\mathbf{f}(t)\right)'\right)^{-1} \\ &\qquad\times\left(\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\left(TD_T^{-1}\Delta\mathbf{f}(t)\right)\boldsymbol{\varepsilon}'_t\right). \end{aligned}$$

For the first term, one makes use of Theorem 2.2 in Kurtz and Protter (1991), whose conditions hold with $\boldsymbol{\varepsilon}_t$ from Assumption 2 (in particular the martingale difference property of $\widetilde{\overline{\boldsymbol{\gamma}}'\mathbf{x}}_{t-1}\boldsymbol{\varepsilon}_t'$ and Equation (17)) leading to

$$\frac{1}{T}\sum_{t=p+1}^{T}\widetilde{\overline{\boldsymbol{\gamma}}'\mathbf{x}}_{t-1}\boldsymbol{\varepsilon}_t' \xrightarrow{d} \int_0^1 \widetilde{\mathbf{B}}^{\boldsymbol{\tau}}(s)\left(\mathrm{d}\mathbf{V}(s)\right)'.$$

For the second term, use Equation (17), the CMT, and Lemma 1 to arrive at

$$\int_0^1 \widetilde{\mathbf{B}}^{\boldsymbol{\tau}}(s)\left(\left(\int_0^1 \mathrm{d}\mathbf{V}(r)\,\dot{\boldsymbol{\tau}}(r)'\left(\int_0^1 \dot{\boldsymbol{\tau}}(r)\dot{\boldsymbol{\tau}}(r)'\mathrm{d}r\right)^{-1}\right)\dot{\boldsymbol{\tau}}(s)\mathrm{d}s\right)' \tag{19}$$

as required.

6. Note that $B_T'S_{11}\boldsymbol{\beta}$ is the sample cross-product moment of a recursively adjusted I(1) process with a recursively adjusted I(0) process. The derivation of the behavior of $\overline{\boldsymbol{\gamma}}'\mathbf{r}_{1t}$ and Example 2 show that the recursive adjustment of the I(0) process has only an $O_p(1)$ effect on $B_T'S_{11}\boldsymbol{\beta}$, and Lemma 2.2 can be applied to obtain the desired magnitude order.

Proof of Proposition 3

The proposition is proved for the trace statistic only; the result for the maximum eigenvalue test statistic uses the same arguments.

The arguments of the proof of Theorem 11.1 of Johansen (1995) apply in a one-to-one manner in our situation: the only difference consists in the behavior of the sample cross-product moment matrices S_{00}, S_{01} and S_{11} under recursive adjustment (more precisely, S_{00}, $\boldsymbol{\beta}'S_{11}\boldsymbol{\beta}$, $\boldsymbol{\beta}'S_{10}$, $T^{-1}B_T'S_{11}B_T$, $B_T'S_{10}\boldsymbol{\alpha}_\perp$, and $B_T'S_{11}$). The required results are derived in Lemma 3 above; compare with Lemma 10.3 of Johansen (1995).

Following the argument leading to Equation 11.16 of Johansen (1995), the largest r eigenvalues $\widehat{\lambda}_1,\ldots,\widehat{\lambda}_r$ are seen to converge in probability to the solutions of the determinant equation

$$\left|\lambda\Sigma_{\boldsymbol{\beta\beta}} - \Sigma_{\boldsymbol{\beta}0}\Sigma_{00}^{-1}\Sigma_{0\boldsymbol{\beta}}\right| = 0,$$

and the remaining $K-r$ eigenvalues converge in probability to 0. The arguments on p. 159 of the reference, together with the preliminary results of Lemma 3, establish the asymptotic null distribution of the $Rec^{\boldsymbol{\tau}}$ statistic, thus completing the proof.

References

Ahlgren, N. and M. Juselius (2012). Tests for Cointegration Rank and the Initial Condition. *Empirical Economics 42*(3), 667–691.

Bayer, C. and C. Hanck (2013, 01). Combining Non-Cointegration Tests. *Journal of Time Series Analysis 34*(1), 83–95.

Chang, Y. (2002). Nonlinear IV Unit Root Tests in Panels With Cross-Sectional Dependency. *Journal of Econometrics 110*(2), 261–292.

Cheng, X. and P. C. B. Phillips (2009). Semiparametric Cointegrating Rank Selection. *The Econometrics Journal 12*(Supplement S1), S83–S104.

Demetrescu, M., H. Lütkepohl, and P. Saikkonen (2009). Testing for the Cointegrating Rank of a Vector Autoregressive Process with Uncertain Deterministic Trend Term. *The Econometrics Journal 12*(3), 414–435.

Doornik, J. A., D. F. Hendry, and B. Nielsen (1998). Inference in Cointegrating Models: UK M1 Revisited. *Journal of Economic Surveys 12*(5), 533–572.

Elliott, G. and U. K. Müller (2006). Minimizing the Impact of the Initial Condition on Testing for Unit Roots. *Journal of Econometrics 135*, 285–310.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica 64*(4), 813–836.

Harvey, D. I., S. J. Leybourne, and A. M. R. Taylor (2009). Unit Root Testing in Practice: Dealing with uncertainty over the trend and initial condition. *Econometric Theory 25*(3), 587–636.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

Kurtz, T. G. and P. Protter (1991). Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations. *Annals of Probability 19*(3), 1035–1070.

Kuzin, V. (2005). Recursive Demeaning and Deterministic Seasonality. *Statistics & Probability Letters 72*(3), 195–204.

Leybourne, S., T. H. Kim, and P. Newbold (2005). Examination of Some More Powerful Modifications of the Dickey-Fuller Test. *Journal of Time Series Analysis 26*(3), 355–369.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.

Lütkepohl, H. and P. Saikkonen (2000). Testing for the Cointegrating Rank of a VAR Process With a Time Trend. *Journal of Econometrics 95*(1), 177–198.

Lütkepohl, H., P. Saikkonen, and C. Trenkler (2001). Maximum Eigenvalue versus Trace Tests for the Cointegrating Rank of a VAR process. *Econometrics Journal 4*(2), 8.

Lütkepohl, H., P. Saikkonen, and C. Trenkler (2004). Testing for the Cointegrating Rank of a VAR Process with Level Shift at Unknown Time. *Econometrica 72*(2), 647–662.

Meng, M., H. Lee, M. H. Cho, and J. Lee (2013). Impacts of the Initial Observation on Unit Root Tests Using Recursive Demeaning and Detrending Procedures. *Economics Letters 120*(2), 195 – 199.

Müller, U. K. and G. Elliott (2003). Tests for Unit Roots and the Initial Condition. *Econometrica 71*(4), 1269–1286.

Nielsen, B. and A. Rahbek (2000). Similarity Issues in Cointegration Analysis. *Oxford Bulletin of Economics and Statistics 62*(1), 5–22.

Phillips, P. C. B. (1987). Time Series Regression with a Unit Root. *Econometrica 55*(2), 277–301.

Phillips, P. C. B. (1988). Weak Convergence of Sample Covariance Matrices to Stochastic Integrals via Martingale Approximations. *Econometric Theory 4*(3), 528–533.

Phillips, P. C. B. and S. N. Durlauf (1986). Multiple Time Series Regression with Integrated Processes. *The Review of Economic Studies 53*(4), 473–495.

Poskitt, D. S. and H. Lütkepohl (1995). Consistent Estimation of The Number of Cointegration Relations in a Vector Autoregressive Model. In R. Galata and H. Küchenhoff (Eds.), *Econometrics in theory and practice. Festschrift for Hans Schneeweiß*, pp. 87–100. Physica.

Rodrigues, P. M. M. (2006). Properties of Recursive Trend-Adjusted Unit Root Tests. *Economics Letters 91*(3), 413–419.

Rodrigues, P. M. M. (2013). Recursive Adjustment, Unit Root Tests and Structural Breaks. *Journal of Time Series Analysis 34*(1), 62–82.

Saikkonen, P. and H. Lütkepohl (1999). Local Power of Likelihood Ratio Tests for The Cointegrating Rank of a VAR Process. *Econometric Theory 15*(1), 50–78.

Saikkonen, P. and H. Lütkepohl (2000a). Testing for the Cointegrating Rank of a VAR Process With an Intercept. *Econometric Theory 16*(3), 373–406.

Saikkonen, P. and H. Lütkepohl (2000b). Trend Adjustment Prior to Testing for the Cointegrating Rank of a Vector Autoregressive Process. *Journal of Time Series Analysis 21*(4), 435–456.

Saikkonen, P. and R. Luukkonen (1997). Testing Cointegration in Infinite Order Vector Autoregressive Processes. *Journal of Econometrics 81*(1), 93–126.

Shin, D. W. and B. S. So (2001). Recursive Mean Adjustment for Unit Root Tests. *Journal of Time Series Analysis 22*(5), 595–612.

So, B. S. and D. W. Shin (1999). Recursive Mean Adjustment in Time-Series Inferences. *Statistics & Probability Letters 43*(1), 65–73.

Sul, D., P. C. B. Phillips, and C. Y. Choi (2005). Prewhitening Bias in HAC Estimation. *Oxford Bulletin of Economics & Statistics 67*(4), 517–546.

Taylor, A. M. R. (2002). Regression-Based Unit Root Tests With Recursive Mean Adjustment for Seasonal and Nonseasonal Time Series. *Journal of Business & Economic Statistics 20*(2), 269–281.

Toda, H. Y. (1994). Finite Sample Properties of Likelihood Ratio Tests for Cointegrating Ranks when Linear Trends are Present. *Review of Economics and Statistics 76*(1), 66–79.

Trenkler, C. (2003). A New Set of Critical Values for Systems Cointegration Tests with a Prior Adjustment for Deterministic Terms. *Economics Bulletin 3*(11), 1–9.

Zivot, E. and D. W. K. Andrews (1992). Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis. *Journal of Business & Economic Statistics 10*(3), 251–270.