The Econometrics Journal

Econometrics Journal (2019), volume **0**, pp. 1
doi: 10.1093/ectj/uty002

Testing collinearity of vector time series

TUCKER S. MCELROY[†] AND AGNIESZKA JACH[‡]

[†]*U.S. Census Bureau, Associate Directorate for Research and Methodology, 4600 Silver Hill Road, Washington, DC 20233.*
Email: tucker.s.mcelroy@census.gov

[‡]*Department of Finance and Statistics, Hanken School of Economics, Helsinki, 4600 Silver Hill Road, Washington, DC 20233, Finland.*
Email: agnieszka.jach@hanken.fi

First version received: 20 August 2018; final version accepted: 18 December 2018.

Summary: We investigate the collinearity of vector time series in the frequency domain, by examining the rank of the spectral density matrix at a given frequency of interest. Rank reduction corresponds to collinearity at the given frequency. When the time series is nonstationary and has been differenced to stationarity, collinearity corresponds to co-integration at a particular frequency. We examine rank through the Schur complements of the spectral density matrix, testing for rank reduction via assessing the positivity of these Schur complements, which are obtained from a nonparametric estimator of the spectral density. New asymptotic results for the test statistics are derived under the fixed bandwidth ratio paradigm; they diverge under the alternative, but under the null hypothesis of collinearity the test statistics converge to a non-standard limiting distribution. Subsampling is used to obtain the limiting null quantiles. A simulation study and an empirical illustration for 6-variate time series data are provided.

Keywords: *trend co-integration*, *seasonal co-integration*, *Schur complement*, *spectral density rank*, *fixed-b asymptotics*, *subsampling*.

JEL codes: *C13*, *C14*.

1. INTRODUCTION

The problem we study is testing for the collinearity of vector time series, i.e., whether the multivariate spectral density of a vector time series (that has been transformed to stationarity) has reduced rank at given frequencies of interest, such as trend or seasonal frequencies. As discussed in Stock and Watson (1988), trend co-integration in an unobserved components framework corresponds to a reduced rank trend innovation covariance matrix, i.e., collinearity of the trend component's innovations. (Similarly, seasonal co-integration in such a framework corresponds to a reduced rank seasonal innovation covariance matrix.) Therefore, it is plausible to estimate the multivariate spectral density at co-integrating frequencies, and test whether the matrix is reduced rank. We propose to adopt a nonparametric approach to this problem, allowing for general frequencies in the spectral density, and we focus on determining not only the rank but also the configuration of nonzero Schur complements (as discussed further below).

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

To frame our work in terms of the relevant literature, this paper is related to the seasonal co-integration tests of Cubadda (2001) and Ahn, Cho and Seong (2004), although these papers are developed in the context of a vector autoregression. Busetti (2006) extended the common trends test statistics of Nyblom and Harvey (2000, 2001) to the seasonal case, using an unobserved components framework. In contrast to these papers, our methods are nonparametric, not requiring a specification of the process aside from a differencing polynomial. Hence our work is more closely related to Bierens (1997) and Shintani (2001), who study the eigenvalue structure of certain nonparametric statistics. However, neither of these works addresses seasonal co-integration (Bierens 1997 examines trend co-integration with a seasonal drift, but not seasonal co-integration), which is central to our own research objectives.

We propose to estimate the multivariate spectral density via tapered sample autocovariances, utilizing a fixed bandwidth ratio asymptotic theory; the invertible case has already been studied by Phillips, Sun and Jin (2006) (hereafter, PSJ). Our focus is on the noninvertible case, wherein the asymptotic theory can be quite different. Moreover, in order to study the exact rank configuration of the spectral density matrix, we will require asymptotics for determinants of submatrices. Our new results show how the limiting distribution depends on the type of kernel that is used. Another contribution of the paper is a nuanced development of collinearity through the generalized Cholesky decomposition, and its connection to frequency-specific co-integration effects. These facets are combined to yield a nonparametric test of co-integration.

In order to understand the econometric context, consider (as a specialized example) an m-variate time series $\{X_t\}$ that is composed of latent processes (Harvey, 1989) such as trend, seasonal, business cycle, and irregular. The trend could be a random walk or an integrated walk, or a more general integrated process, and if the innovation sequence driving the trend has dimension less than m (i.e., this is a white noise sequence with covariance matrix of less than full rank) then co-integration at the trend frequency is present. Similarly, nonstationary latent processes are typically formulated for the seasonal component, and innovations of reduced rank result in co-integration at some or all of the seasonal frequencies. (Seasonal frequencies are of the form $2\pi j/p$ where there are p seasons, for $1 \leq j \leq p$.) McElroy (2017) discusses the presence of seasonal co-integration in retail and construction data, and other literature includes Koopman, Ooms and Hindrayanto (2012). If the data are differenced to stationarity by applying both trend and seasonal unit root differencing polynomials, the resulting stationary series $\{\underline{X}_t\}$ will have spectral density f that has reduced rank at frequency zero and/or the seasonal frequencies.

To see the difficulty in detecting collinearity, suppose we wish to test for co-integration at some frequency λ via studying $\det f(\lambda)$. (Though $f(\lambda)$ can be complex, the matrix is always Hermitian and hence its determinant is real and non-negative.) The null hypothesis of reduced rank corresponds to $\det f(\lambda) = 0$. The estimators $\widehat{f}$ considered by PSJ satisfy

$$\det \widehat{f}(\lambda) \xrightarrow{\mathcal{D}} \det f(\lambda) \cdot \det \Xi, \tag{1.1}$$

where Ξ is a stochastic Hermitian matrix depending on bandwidth fraction (defined below). Thus, under the null hypothesis the natural test statistic $\det \widehat{f}(\lambda)$ converges to zero in probability. Clearly, a different rate of normalization is needed to ensure that $\det \widehat{f}(\lambda)$ converges to a nondegenerate distribution—such a normalization and limit theory is provided by our work.

The identification of collinearity is of intrinsic interest, as it indicates dynamic co-movements of series at particular frequencies. Furthermore, collinearity typically implies that a more parsimonious model can be used without loss of information—this is especially important for moderate- to high-dimensional data, where the parameter dimension is large unless suitable restrictions are

employed. However, the misspecification of collinearity has a deleterious impact on forecasting and signal extraction (discussed in Section 4 below), and hence it is vital to avoid wrongly enforcing spurious co-integration constraints. In this paper, the null hypothesis corresponds to such constraints being valid, and hence low p-values warrant a less parsimonious modelling of the process.

Given that one is interested in detecting co-integration, we would like to estimate the co-integrating vector and determine the structure of eigenvalues, as well as to obtain the statistical significance of results. If $\det f(\lambda) = 0$, we know only that the rank of $f(\lambda)$ is less than m. While there are some available tests in the econometric literature to determine the rank of $f(\lambda)$, for applications it is useful to know the exact rank structure, i.e., the configuration of nonzero eigenvalues. McElroy (2017) discusses a parametrization of covariance matrices in terms of a sequence of Schur decompositions associated with the generalized Cholesky decomposition. The successive Schur complements, which are unordered scalars, can be positive or zero, and the number of positive-valued Schur complements is equal to the rank. Knowing not just how many but which Schur complements are zero facilitates applications to taxonomy and signal extraction. (Whereas a reordering of the series will alter the values of the Schur complements, the number of such that are zero remains unchanged under permutation.)

With these motivations in mind, the paper proceeds to first articulate a general discussion of co-integration and rank configuration in Section 2, followed by the statistical methodology and asymptotic results in Section 3. The treatment of the Schur complement and the Cholesky decomposition involves complex matrices, and yields novel algorithms for computation. The limit distributions are nonpivotal, and are estimated via subsampling (Politis et al., 1999) in our simulations and applications (Section 4). Proofs are in the Appendix of the Supporting Information.

2. CO-INTEGRATION AND RANK CONFIGURATION

2.1. Non-invertibility arising from co-integration

2.1.1. VAR illustrations. Johansen (1988) provides an analysis of co-integration (Engle and Granger, 1987) in the context of a vector autoregression (VAR). The starting point is a VAR(1) model for the observed process, with a decomposition of the autoregressive operator into unit and stable portions. A nonstationary VAR(1) process is given by $X_t = \Phi X_{t-1} + \epsilon_t$ for $\{\epsilon_t\}$ i.i.d. with m-dimensional covariance matrix Σ. Following Gómez (2016, Section 5.7), we suppose $m - r$ of the eigenvalues of Φ equal one, so that $\Pi = \Phi - I_m$ can be written as $\alpha\,\beta'$, for $m \times r$ matrices α and β that have rank r. (Here I_m denotes an m-dimensional identity matrix.) Such a situation is described as a rank r co-integrated VAR(1); if the rank is m, then the VAR(1) is actually stationary, and we say that it is a stable VAR(1).

PROPOSITION 2.1. *Suppose that the m-variate process $\{X_t\}$ is a co-integrated VAR(1) process with coefficient matrix Φ such that $\Phi - I_m = \alpha\,\beta'$ for rank r matrices α and β. Then $Z_t = \beta' X_t$ is a stable VAR(1) process, and $\underline{X}_t = X_t - X_{t-1}$ is stationary with infinite moving average representation $\Theta(B)\epsilon_t$, with*

$$\Theta(B) = I_m + \alpha \left(I_r - [I_r + \beta'\,\alpha]\,B\right)^{-1} \beta'\,B$$

satisfying $\beta'\Theta(1) = 0$. *Also,*

$$I_m - \Phi\, B = (I_m - \Phi^*\, B)(I_m - U\, B),$$

with stable $\Phi^* = (\alpha + \beta\,(\beta'\,\beta)^{-1})\,\beta'$ *and* $U = \beta_\perp\,(\beta'_\perp\,\beta_\perp)^{-1}\,\beta'_\perp$, *where the columns of* $\beta_\perp$ *span the null space of* β'.

The columns of β are therefore the co-integrating vectors. It follows from Proposition 2.1 that $\Theta(1) = I_m - \alpha\,(\beta'\,\alpha)^{-1}\,\beta'$, and hence $f(0)$ (the spectral density of $\{\underline{X}_t\}$ at frequency zero) has rank $m - r$. Hence, this simple form of trend co-integration implies noninvertibility of the differenced time series; Proposition 2.1 furnishes a particular case of the Granger representation theorem (Johansen, 1995), whereby the moving average representation of $\underline{X}_t$ is singular at frequency zero. However, if we utilize the reduced rank (nonscalar) differencing operator $I_m - U\,B$ instead of $1 - B$, we obtain a stable VAR(1) process with coefficient matrix Φ^*.

A similar argument can be developed for the case of eigenvalues equal to negative one, or for complex unit-magnitude eigenvalues. In the latter case, we have the following extension of the frequency-zero results: if Φ has rank r with q eigenvalues equal to $e^{i\omega}$, then there must be q eigenvalues equal to $e^{-i\omega}$ as well, so that $2q \leq m - r$. Suppose that $(e^{i\omega}, v)$ is an eigen-pair of Φ; then conjugating $\Phi\, v = e^{i\omega}\, v$ yields $\Phi\, \overline{v} = e^{-i\omega}\, \overline{v}$, so that $(e^{-i\omega}, \overline{v})$ is also an eigen-pair. Hence $\Phi - e^{i\omega} I_m$ and $\Phi - e^{-i\omega}\, I_m$ both have rank at most $m - q$. Their product is $\Phi^2 - 2\,\cos(\omega)\,\Phi + I_m$, designated by Π, and (λ, v) is an eigen-pair of Φ if and only if $(\lambda^2 - 2\,\cos(\omega)\,\lambda + 1, v)$ is an eigen-pair of Π. Hence Π has $2q$ eigenvalues equal to zero, and therefore the matrix has rank at most $m - 2q$; we suppose that $q = (m - r)/2$, where r is the rank of Φ. Finally, we suppose that $\Pi = \alpha\,\beta'$ for $m \times r$ matrices α and β that have rank r.

PROPOSITION 2.2. *Suppose that the m-variate process* $\{X_t\}$ *is a co-integrated VAR(1) process with rank r coefficient matrix* Φ *having q eigenvalues equal to* $e^{i\omega}$, *for some* $\omega \in (0, \pi)$ *and for* $q = (m - r)/2$. *Also suppose that* $\Pi = \Phi^2 - 2\,\cos(\omega)\,\Phi + I_m = \alpha\,\beta'$ *for rank r matrices* α *and* β. *The* $Z_t = \beta' X_t$ *is a stable VARMA(2,1) process, and* $\underline{X}_t = X_t - 2\,\cos(\omega)\,X_{t-1} + X_{t-2}$ *is stationary with infinite moving average representation* $\Theta(B)\epsilon_t$, *with*

$$\Theta(B) = \left(I_m + \alpha\,(I_q - 2\,\cos(\omega)\,I_q\,B + (I_q - \beta'\,\alpha)\,B^2)^{-1}\,\beta'\,B^2\right)\,(I_m + (\Phi - 2\,\cos(\omega)\,I_m)\,B)$$

satisfying $\beta'\Theta(e^{\pm i\omega}) = 0$. *Also, there exists a matrix differencing operator* $D(B) = I_m - 2\,\cos(\omega)\,U\,B + U\,B^2$, *where* $U = \beta_\perp\,(\beta'_\perp\,\beta_\perp)^{-1}\,\beta'_\perp$ *(the columns of* $\beta_\perp$ *span the null space of* β'*), such that* $Y_t = D(B)X_t$ *is a stationary VARMA(2,1) process satisfying*

$$\Phi^*(B)Y_t = \Theta^*(B)\epsilon_t,$$

where

$$\Phi^*(B) = I_m - 2\,\cos(\omega)\,\beta\,(\beta'\,\beta)^{-1}\,\beta'\,B + (\beta\,(\beta'\,\beta)^{-1}\,\beta' - \Pi)\,B^2$$

$$\Theta^*(B) = I_m + (\Phi - 2\,\cos(\omega)\,I_m)\,B.$$

Hence $f(\omega)$ has rank $r = m - 2q$, indicating that application of the scalar differencing operator $1 - 2\,\cos(\omega)\,B + B^2$ reduces the process to stationarity, but also overdifferences at frequency ω, in the sense that the process is noninvertible at that frequency. So co-integration in a VAR(1) implies noninvertibility in the differenced process, although it is possible to utilize a nonscalar differencing operator $D(B)$ to yield an invertible VARMA(2,1) process.

2.1.2. Latent process illustration. The VAR formulations yield spectra for the differenced processes with collinearity at a single frequency. For illustrations of processes with more than one collinear frequency, consider a latent stochastic process structure. Suppose that the data process $\{X_t\}$ can be written as the sum of nonstationary latent processes, each of which can be differenced to stationarity via some scalar polynomial $\delta^{(\ell)}(B)$ of multiplicity q_ℓ and unit root frequency ω_ℓ. In the case that $\omega_\ell = 0$, $\delta^{(\ell)}(B) = (1-B)^{q_\ell}$, whereas $\delta^{(\ell)}(B) = (1+B)^{q_\ell}$ for $\omega_\ell = \pi$. In the case that $\omega_\ell \in (0, \pi)$, $e^{i\omega_\ell}$ is complex, and we define

$$\delta^{(\ell)}(B) = (1 - e^{i\omega_\ell}B)(1 - e^{-i\omega_\ell}B) = (1 - 2\cos(\omega_\ell)B + B^2).$$

Also, one of the latent components may be a stationary process, which will be denominated as a transient. Supposing these polynomials to be distinct (so that the latent components have distinct dynamics), the scalar differencing polynomial $\delta(B) = \prod_{\ell=1}^{p} \delta^{(\ell)}(B)$ reduces the data process to stationarity, i.e., $\underline{X}_t = \delta(B)X_t$ is stationary. Hence we have the representation

$$X_t = \sum_{\ell=1}^{p} S_t^{(\ell)} + S_t^{(0)},$$

where $\underline{S}_t^{(\ell)} = \delta^{(\ell)}(B)S_t^{(\ell)}$ is stationary for each $1 \le \ell \le p$, and $\{S_t^{(0)}\}$ is the transient. Let the reduced polynomials $\delta^{(-\ell)}(B) = \delta(B)/\delta^{(\ell)}(B)$ be defined. Then applying $\delta(B)$ to the above structural equation yields

$$\underline{X}_t = \sum_{\ell=1}^{p} \delta^{(-\ell)}(B)\, \underline{S}_t^{(\ell)} + \delta(B)\, S_t^{(0)}.$$

Suppose each stationary latent process $\underline{S}_t^{(\ell)}$ has singularities in its spectral density matrix, such that it can be represented as $\Lambda^{(\ell)}$ times some $C_t^{(\ell)}$, a stationary process of reduced dimension with spectral density matrix invertible at all frequencies. This yields a latent dynamic factor model representation of $\{\underline{X}_t\}$. Suppose each process $\{C_t^{\ell}\}$ for $0 \le \ell \le p$ is stationary with spectral density $f^{(\ell)}$. Then

$$f(\lambda) = \sum_{\ell=1}^{p} |\delta^{(-\ell)}(e^{-i\lambda})|^2\, \Lambda^{(\ell)}\, f^{(\ell)}(\lambda)\, \Lambda^{(\ell)\prime} + |\delta(e^{-i\lambda})|^2\, \Lambda^{(0)}\, f^{(0)}(\lambda)\, \Lambda^{(0)\prime},$$

and $f(\omega_k) = |\delta^{(-k)}(e^{-i\omega_k})|^2\, \Lambda^{(k)}\, f^{(k)}(\omega_k)\, \Lambda^{(k)\prime}$ for $1 \le k \le p$. Thus, f has reduced rank at frequency ω_k if $\Lambda^{(k)}$ is reduced rank. Note that if $\Lambda^{(0)}$ is full rank, then the spectrum of $\{\underline{X}_t\}$ is nonsingular except at most a finite number of frequencies.

2.2. Rank configuration via generalized Cholesky decomposition

In this paper, we suppose that the practitioner has several values of λ in mind, and will test for collinearity separately at each frequency. The previous examples give some contexts where a time series may be co-integrated at a frequency λ. In the co-integrated VAR examples, one would take either $\lambda = 0$ or $\lambda = \omega$; more generally, if $\underline{X}_t = \delta(B)\, X_t$ is obtained by differencing, then we would take λ such that $\delta(e^{-i\lambda}) = 0$. If the data is stationary, we may determine frequencies λ of interest by an initial exploratory analysis based on the autocovariance generating function—details are discussed in the Supporting Information.

We now consider the spectral density matrix f of $\{\underline{X}_t\}$ at a frequency λ. We denote this matrix $f(\lambda)$ by Σ, noting that although it can have complex entries, it is non-negative-definite (spd); it

can be decomposed into the form

$$\Sigma = L\, D\, L^*,$$

where L is unit lower-triangular (with complex entries), and D is diagonal with non-negative entries. Here * denotes the conjugate transpose. The case that D is invertible corresponds to the regular Cholesky decomposition, but when Σ is noninvertible, then some of the entries of D will be zero, and the corresponding columns of L are indeterminate. Writing $L(j)$ for the jth column of L and d_j for the jth diagonal entry of D, we have

$$\Sigma = \sum_{j=1}^{m} d_j\, L(j) L^*(j)$$

for a matrix of dimension m. In some cases we may wish to emphasize that d_j is the jth Schur complement of a particular Σ, in which case we write $d_j(\Sigma)$. Let J denote the subset of $\{1, 2, \cdots, m\}$ corresponding to nonzero d_j. We emphasize that J depends on the ordering of the time series—the interpretation of the entries of D and L, and their dependence on the ordering of elements in X_t, is discussed in McElroy (2017). However, $\text{rank}(\Sigma) = |J|$ does not depend on the ordering of the series, and we can always determine a new rank configuration after applying a permutation of the series—see the Supporting Information for further details.

In the context of co-integration testing, where it is already known via exploratory analysis that each series requires $\delta(B)$ differencing to be rendered stationary, we can safely assume that $d_1 > 0$ if λ satisfies $\delta(e^{-i\lambda}) = 0$ (see Supporting Information). Nevertheless, for stationary series that are tested for collinearity it is possible that $d_1 = 0$, so we treat this most general case below. So the spectral density can be written

$$\Sigma = \sum_{j \in J} d_j\, L(j) L^*(j), \tag{2.1}$$

and J is called the *rank configuration* of Σ. In order to compute the generalized Cholesky decomposition, a recursive algorithm is typically employed, where the later rows of L and entries of D are successively computed. Placing a j subscript on the matrices to denote their dimension, where $1 \leq j \leq m-1$, in the regular Cholesky decomposition of the matrix

$$\Sigma_{j+1} = \begin{bmatrix} \Sigma_j & \underline{\sigma}_j \\ \underline{\sigma}_j^* & s_{j+1} \end{bmatrix},$$

we have

$$\ell_{j+1} = D_j^{-1}\, L_j^{-*}\, \underline{\sigma}_j \qquad d_{j+1} = s_{j+1} - \ell_{j+1}^*\, D_j\, \ell_{j+1},$$

where $[\ell'_{j+1}, 1]$ is the last row vector of L_{j+1}. Here $-*$ is the inverse of the conjugate transpose. Because $L_j D_j \ell_{j+1} = \underline{\sigma}_j$, we see that (for $j > 0$)

$$d_{j+1} = s_{j+1} - \underline{\sigma}_j^*\, \Sigma_j^{-1}\, \underline{\sigma}_j = \frac{\det \Sigma_{j+1}}{\det \Sigma_j} \tag{2.2}$$

by a well-known matrix identity (cf., Axelsson, 1996, p.93) extended to the complex case. In other words, the various Schur complements d_j can each be represented as a ratio of (real) determinants. (When $j = 0$, in lieu of (2.2) we write $d_1 = s_1$.)

The generalized Cholesky decomposition replaces D_j^{-1} by its generalized inverse $D_j^{\sharp}$, defined as the diagonal matrix consisting of entries $d_j^{\sharp}$, which is either equal to d_j^{-1} or is zero if $d_j = 0$.

The new algorithm is then

$$\ell_{j+1} = D_j^{\sharp} L_j^{-*} \underline{\sigma}_j \qquad d_{j+1} = s_{j+1} - \ell_{j+1}^{*} D_j \ell_{j+1}.$$

The net effect is that the resulting matrix L_m will have lower triangular entries equal to zero in any column corresponding to a zero Schur complement. From (2.1), we see that these particular entries of L_m are indeterminate, in the sense that the decomposition of Σ_m should result in rectangular matrices L_m with columns omitted that belong to J^c. Alternatively, we can still write L_m as a square, so long as D_m is square with appropriate entries given by zeroes; then the irrelevant columns of L can be given any value, as they will be multiplied by zero—we choose arbitrarily to set such columns to have zero entries (except for the unit entry, corresponding to the diagonal of L_m). The following result gives the formula for the Schur complements in terms of ratios of determinants. We use the notation $\Sigma_j(K)$ to denote the submatrix of Σ_j where only rows and columns of indices belonging to K are retained. Let $J_j = J \cap \{1, 2, \ldots, j\}$, and $\tilde{J}_{j+1} = J_j \cup \{j+1\}$ (which can differ from J_{j+1}).

PROPOSITION 2.3. *For $j \geq 1$ the matrix $\Sigma_j(J_j)$ is positive-definite (pd), and*

$$d_{j+1} = \frac{\det \Sigma_{j+1}(\tilde{J}_{j+1})}{\det \Sigma_j(J_j)}. \tag{2.3}$$

As a result of Proposition 2.3, we can compute the diagonal entries of Σ_m by taking ratios of determinants of appropriate submatrices.

3. SPECTRAL DENSITY RANK TESTING

3.1. Testing procedure

Focusing on a frequency λ of interest, we wish to estimate the true rank configuration J (which depends on the ordering of the series) based on the nonparametric estimate $\widehat{f}(\lambda)$. A rank configuration J will be estimated, as discussed below, by testing sequentially whether each d_j is positive, for $1 \leq j \leq m$. We will test each Schur complement d_j individually; however, its distribution depends on whether Schur complements of lower index are zero. The test statistics are $\widehat{d}_1 = \widehat{\Sigma}_1$ and (for $j > 0$)

$$\widehat{d}_{j+1} = \frac{\det \widehat{\Sigma}_{j+1}(\tilde{J}_{j+1})}{\det \widehat{\Sigma}_j(J_j)}, \tag{3.1}$$

where $\widehat{\Sigma}_j$ is the estimated spectral density matrix at frequency λ, corresponding to the first j time series. For any such subseries spectral density f, the corresponding lag h autocovariance $\Gamma(h)$ is defined via the inverse Fourier Transform

$$\Gamma(h) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) e^{i\lambda h} \, d\lambda.$$

The sample autocovariance function (acf) is defined from a size T sample of the differenced time series $\{\underline{X}_t\}$ of mean μ via

$$\widehat{\Gamma}(h) = T^{-1} \sum_{t=1}^{T-h} \left(\underline{X}_{t+h} - \widehat{\mu}\right) \left(\underline{X}_t - \widehat{\mu}\right)',$$

where $h \geq 0$ and $\widehat{\mu}$ is the vector sample mean. Then given a kernel Λ, which is an even function of domain $[-1, 1]$, we set $\Lambda_b(x) = \Lambda(x/b)$ for bandwidth fraction b, and define the tapered acf spectral density estimator as

$$\widehat{f}(\lambda) = \sum_{|h|<T} \Lambda_b(h/T)\,\widehat{\Gamma}(h) e^{-i\lambda h} = T^{-1} \sum_{t,\ell=1}^{T} \Lambda_b((t-\ell)/T)\,(\underline{X}_t - \widehat{\mu})\,(\underline{X}_\ell - \widehat{\mu})^* \, e^{-i\lambda(t-\ell)}. \tag{3.2}$$

The bandwidth is bT. Note that the kernel is scalar, so that each component spectrum is tapered similarly. A fairly standard assumption is that the kernel is twice continuously differentiable, although this can be relaxed to a piecewise requirement, allowing us to include the Bartlett kernel, truncation kernel, and flat-top kernels. (McElroy and Politis (2014) treats the scalar case.)

Because the denominator of $\widehat{d}_{j+1}$ in (3.1) corresponds to a pd matrix, the theory of PSJ can be applied; the numerator, however, tends to zero if $j+1 \notin J$ (this corresponds to the null hypothesis), otherwise tending to a positive random variable. That is, if $d_{j+1} > 0$ then the spectral densities corresponding to subseries J_j and $\tilde{J}_{j+1}$ are both invertible at λ, and hence each has a spectral factorization of the form $\Psi(e^{-i\lambda}) \cdot \Psi(e^{i\lambda})'$. The key result for the tapered acf spectral density estimator is

$$\widehat{f}(\lambda) \xrightarrow{\mathcal{D}} \Psi(e^{-i\lambda}) \cdot \Xi \cdot \Psi(e^{i\lambda})'. \tag{3.3}$$

This is proved in PSJ using results in Chan and Terrin (1995), or alternatively by using the linear method of Phillips and Solo (1992). The meaning of this weak convergence of stochastic matrices is that the vec operator applied to $\widehat{f}(\lambda)$ satisfies a joint weak convergence to the vec of the right-hand side above. Hence, continuous functions of the various matrix components can be applied to the convergence; in particular, we can apply the determinant function to the convergence, obtaining (1.1). In summary, when $d_{j+1} > 0$ it can be shown that $\widehat{d}_{j+1}$ converges weakly to a positive random variable, given as the ratio of determinants of certain stochastic matrices—details are provided below.

When $d_{j+1} = 0$, we know that $\Sigma_{j+1}(\tilde{J}_{j+1})$ has a number (given by the cardinality of J_j) of positive Schur complements and a final Schur complement equal to zero; in particular, the rank equals the cardinality of J_j. The spectral density corresponding to the subseries $\tilde{J}_{j+1}$ has rank equal to the cardinality of J_j, and so the spectral factorization yields a $\Psi(B)$ of reduced rank. Further analysis (details in the theorem below) indicates that T det $\widehat{\Sigma}_{j+1}(\tilde{J}_{j+1})$ converges weakly to a nondegenerate distribution. Therefore, to test $d_{j+1} = 0$ (for $j \geq 0$) the test statistic

$$T\,\widehat{d}_{j+1} \tag{3.4}$$

is proposed; when $d_{j+1} > 0$, we know that $\widehat{d}_{j+1} = O_P(1)$ (and is not $o_P(1)$), indicating that the test is consistent. (If the kernel Λ is not positive-definite, then $\widehat{d}_{j+1}$ can take negative values; nevertheless, we are interested only in the upper one-sided alternative that $d_{j+1} > 0$.) Given $J \subset \{1, 2, \ldots, m\}$, we wish to determine J in a sequential manner, analogously to the work of Johansen (1988, 1995). For any $0 \leq j \leq m-1$ we will determine whether some $j+1 \in J$, given that we already have determined J_j as the outcome of previous tests. That is, we have already determined whether or not each k for $1 \leq k \leq j$ is in J or not. So we compute $\widehat{d}_{j+1}$ via (3.1) and test the null hypothesis that $j+1 \in J$ via the test statistic (3.4). If the null is rejected, then $J_{j+1} = J_j \cup \{j+1\}$, but otherwise $J_{j+1} = J_j$.

The key issue, then, is to estimate the limiting distribution under the hypothesis that $d_{j+1} = 0$. Because the limit is nonpivotal (it depends on the Wold decomposition of the data process) in such a way that the limiting quantiles cannot be directly computed, we propose to utilize subsampling methodology. A straightforward extension of Theorem 3.2.1 of Politis et al. (1999)—stated generally below—will provide estimates of the test statistic's quantiles. Consider a test statistic $\widehat{\theta}_T$, which for null and alternative hypotheses H_0 and H_a satisfies

ASSUMPTION 3.1. *$\tau_T \widehat{\theta}_T \xrightarrow{\mathcal{D}} W_0$ under H_0 and $\widehat{\theta}_T \xrightarrow{\mathcal{D}} W_a$ under H_a, where W_0 and W_a are continuous random variables with cdfs F_0 and F_a, respectively.*

Let $\overline{F}_0^{-1}$ and $\overline{F}_a^{-1}$ be the right-tailed quantile functions of W_0 and W_a, respectively. We denote the subsampling rate by n (as opposed to the usual notation b, which is reserved for the bandwidth fraction discussed above). The subsampling distribution estimator for the test statistic is defined via

$$L_{T,n}(x|\tau.) = \frac{1}{T-n+1} \sum_{i=1}^{T-n+1} 1_{\{\tau_n \widehat{\theta}_{n,i} \leq x\}},$$

where the dependence on the rate τ_n is explicit in the notation. Here $\widehat{\theta}_{n,i}$ denotes the evaluation of the test statistic on the subsample $\{\underline{X}_i, \ldots, \underline{X}_{i+n-1}\}$. Note that the bandwidth for the spectral density estimator based on such a subsample is bn, since the same bandwidth fraction is applied. The subsampling quantiles are denoted $L_{T,n}^{-1}(t|\tau.)$, and are given by the tth largest-order statistic of $\{\tau_n \widehat{\theta}_{n,i}\}$ (for $1 \leq i \leq T-n+1$). As discussed in Chapter 8 of Politis et al. (1999),

$$L_{T,n}^{-1}(t|\tau.) = \tau_n \cdot L_{T,n}^{-1}(t|1). \tag{3.5}$$

It is shown below that the subsampling distribution estimator $L_{T,n}(x|\tau.)$ estimates $F_0(x)$ under H_0, whereas the subsampling distribution estimator $L_{T,b}(x|1)$ estimates $F_a(x)$ under H_a.

PROPOSITION 3.1. *Let $\{\underline{X}_t\}$ be a strictly stationary time series that is strong mixing. Assume Assumption 3.1, as well as $\tau_n/\tau_T \to 0$, $n/T \to 0$, and $n \to \infty$ as $T \to \infty$. Then*

$$L_{T,n}(x|\tau.) \xrightarrow{P} F_0(x) \quad \text{if } H_0 \text{ is true}$$

$$L_{T,n}(x|1) \xrightarrow{P} F_a(x) \quad \text{if } H_a \text{ is true}$$

as $T \to \infty$. Also, as $T \to \infty$

$$\mathbb{P}[\tau_T \widehat{\theta}_T \geq L_{T,n}(1-\alpha|\tau.)^{-1}] \to \alpha$$

when H_0 is true, and converges in probability to one when H_a is true so long as W_a is positive with probability one.

Therefore, we can construct an upper one-sided subsampling critical region of asymptotic size α via

$$[L_{T,n}^{-1}(1-\alpha|\tau.), \infty).$$

This general discussion applies immediately to the collinearity test statistic of this paper, letting $\tau_T = T$ and $\widehat{\theta}_T = \widehat{d}_{j+1}$, though it remains to derive conditions under which Assumption 3.1 holds.

The entire statistical procedure amounts to a sequence of tests, whose calculation depends on the outcome of previous tests. This is not a multiple testing problem (the null hypotheses are

all different), but is rather a sequential testing problem, and there may be a concern that the random outcome of a prior test is not accounted for in the distribution theory for the current test. Chen and Fang (2018) indicate—in the related context of testing for rank, as opposed to the structure of positive Schur complements—that sequential testing ignores type I errors from previous steps and may have poor power for close alternatives. We treat the outcome of a test of $d_j = 0$ as deterministic in each stage, towards the end of approximating the distribution of the test statistic for the hypothesis $d_{j+1} = 0$. However, the subsampling methodology is conducted by applying the entire sequential testing procedure to each subsample, and hence the type I errors from previous tests (on a given subsample) are incorporated in subsequent tests. This is akin to the bootstrap procedure recommended in Chen and Fang (2018) for handling sequential rank tests.

Below, we povide the asymptotic theory for the null and alternative cases. It follows from this theory that the test statistic (3.4) grows at order T, indicating the test is consistent. However, as this is a nonparametric procedure, the test statistic has not been designed so as to maximize its power under particular parametric alternatives (cf. the procedure of Johansen, 1988, 1995). Instead, our procedure is designed so as to be broadly applicable to many types of processes—such as co-integrated VAR processes and structural time series—and hence may be useful in contexts where the practitioner is unsure about the correct parametric specification.

3.2. Asymptotic theory

Below we describe the limit distribution of the test statistic $T\,\widehat{d}_{j+1}$, after giving some preliminary results. The null hypothesis is that $d_{j+1} = 0$, such that $\Sigma_{j+1}(\tilde{J}_{j+1})$ has a number of positive Schur complements and a final Schur complement equal to zero. Fix λ and set f equal to $\Sigma_{j+1}(\tilde{J}_{j+1})$, which in turn equals $\Psi\,\Psi^*$. (Set Ψ equal to $\Psi(e^{-i\lambda})$, to simplify notation.) When $j > 0$, by partitioning Ψ into $\Psi_\sharp$ and $\Psi_\flat$ among the first J_j series and the last $(j+1)$th series, we can write

$$f = \begin{bmatrix} \Psi_\sharp\,\Psi_\sharp^* & \Psi_\sharp\,\Psi_\flat^* \\ \Psi_\flat\,\Psi_\sharp^* & \Psi_\flat\,\Psi_\flat^* \end{bmatrix}, \tag{3.6}$$

and obtain the following result.

PROPOSITION 3.2. *Given the partition (3.6) of $f = \Sigma_{j+1}(\tilde{J}_{j+1})$ when $j > 0$, and defining*

$$v = \left[\Psi_\sharp\,\Psi_\sharp^*\right]^{-1}\Psi_\sharp\,\Psi_\flat^* \tag{3.7}$$

and $w^* = [v^*, -1]$, the last Schur complement can be written

$$d_{j+1} = \Psi_\flat\,\Psi_\flat^* - \Psi_\flat\,\Psi_\sharp^*\left[\Psi_\sharp\,\Psi_\sharp^*\right]^{-1}\Psi_\sharp\,\Psi_\flat^* = w^*\,f\,w.$$

Under the null hypothesis, $\Psi_\sharp^*\,v = \Psi_\flat^*$.

Note that because the expression for d_{j+1} in Proposition 3.2 is derived algebraically, it also holds true for the estimators:

$$\widehat{d}_{j+1} = \widehat{w}^*\,\widehat{f}\,\widehat{w}.$$

Also, the term $\Psi_\sharp\,\Psi_\sharp^*$ occurring in (3.7) is the spectral density of the first J_j series; denote this spectrum by g, and denote the cross-spectrum of the $(j+1)$th series with the first J_j series by k. So $g = \Psi_\sharp\,\Psi_\sharp^*$ and $k = \Psi_\flat\,\Psi_\sharp^*$, and $v = g^{-1}\,k^*$. The corresponding estimator is $\widehat{v} = \widehat{g}^{-1}\,\widehat{k}^*$.

Next, define the discrete Fourier transform (DFT) process for any given series $\{W_t\}$ via

$$d_W^s(\lambda) = T^{-1/2} \sum_{t=1}^{[Ts]} W_t e^{-i\lambda t} \tag{3.8}$$

for $s \in [0, 1]$ and $\lambda \in [-\pi, \pi]$. Under fairly broad conditions, the DFT process satisfies a functional central limit theorem; sufficient conditions are given in PSJ. We provide a concrete scenario in which this result is obtained.

THEOREM 3.1. *Suppose that $\{W_t\}$ is a strictly stationary process whose kth-order cumulants exist and are summable for all $k \geq 1$. Also suppose that for some $\delta > 0$ the jth component of the DFT (3.8) satisfies*

$$\mathbb{E}[|\mathcal{R}d_{W_j}^s(\lambda)|^{2+\delta}] = O(1) \qquad \mathbb{E}[|\mathcal{I}d_{W_j}^s(\lambda)|^{2+\delta}] = O(1)$$

uniformly in λ and s. Then the DFT satisfies a functional central limit theorem for each λ:

$$d_W(\lambda) \xrightarrow{\mathcal{D}} \Psi(e^{-i\lambda})\,(B_R - iB_I)\,, \tag{3.9}$$

where B_R and B_I are independent vector Brownian motions with covariance matrix

$$Cov[B_R(s), B_R(t)] = Cov[B_I(s), B_I(t)] = \min\{s, t\}\, I_m/2$$

when $\lambda \in (0, \pi)$, and when $\lambda = 0, \pi$ they have structure $B_I \equiv 0$ and

$$Cov[B_R(s), B_R(t)] = \min\{s, t\}\, I_m.$$

It is possible to establish limit theorems such as Theorem 3.1 using other assumptions, such as assuming that $\{W_t\}$ is a linear process, but the cumulant conditions are sufficient for our purposes. Our main result below requires that such a functional limit theorem holds, and provides the asymptotic theory for $\widehat{d}_{j+1}$ supposing a wide class of kernels. We assume that Λ is piecewise-smooth and continuous, with possible kinks at a finite collection of point x_k. A kink is defined by where the left-hand $\dot{\Lambda}^-(x_k)$ and right-hand $\dot{\Lambda}^+(x_k)$ derivatives do not coincide. Let K denote the number of such kinks, and observe that [0, 1] is thereby partitioned into open intervals C_0, C_1, ..., C_K, where $C_0 = [0, x_1)$, $C_1 = (x_1, x_2)$, and so forth, with $C_K = (x_K, 1]$. In the case that $x_1 = 0$ (e.g., with the Bartlett kernel), C_0 is omitted, with a similar convention for C_K when $x_K = 1$ (e.g., for the truncation kernel).

Let B denote a complex vector Brownian motion satisfying $B = B_R - iB_I$, with B_R and B_I as defined in Theorem 3.1. To handle the effect of the sample mean, we introduce a Brownian bridge as $\underline{B}(r) = B(r) - rB(1)$ when $\lambda = 0$, but $\underline{B}(r) = B(r)$ when $\lambda \neq 0$. Furthermore, we have the following modifications of the kernel featured in the limit: $\underline{\underline{\Lambda}}_b(r, s)$ equals $\Lambda_b(r - s)$ if $\lambda = 0$, and equals

$$\Lambda_b(r-s) - \int_0^1 \Lambda_b(x-s)\,dx - \int_0^1 \Lambda_b(x-r)\,dx + \int_0^1\int_0^1 \Lambda_b(x-y)\,dxdy$$

if $\lambda \neq 0$. Also, $\underline{\Lambda}_b(r, s)$ equals $\Lambda_b(r - s)$ if $\lambda \neq 0$, and equals

$$\Lambda_b(r-s) - \int_0^1 \Lambda_b(x-s)\,dx \cdot 1_{[0,1]}(r)$$

if $\lambda \neq 0$, where $1_{[0,1]}(r)$ denotes the indicator function. Similarly, let $\underline{\dot{\Lambda}}_b(r, s)$ be defined by simply substituting $\dot{\Lambda}_b$ for Λ_b in the above formulas. Suppose the $\tilde{J}_{j+1}$ variables are denoted by $\{\underline{X}_t\}$,

where $\underline{X}_t = \mu + \Psi(B)\,\epsilon_t$ for white noise $\{\epsilon_t\}$ of covariance matrix equal to the identity. Then the Beveridge–Nelson factorization yields $\Psi(B) = \Psi(e^{-i\lambda}) + (1 - e^{i\lambda}B)\,\Theta(B)$, where

$$\Theta(B) = \frac{\Psi(B) - \Psi(e^{-i\lambda})}{1 - e^{i\lambda}B} = -\sum_{j\geq 1} \Psi_j e^{-i\lambda j} \left(\sum_{k=0}^{j-1} B^k e^{i\lambda k} \right).$$

This filter $\Theta(B)$ involves complex coefficients when $\lambda \neq 0, \pi$. Set $Z_t(\lambda) = e^{-i\lambda t}\,\Theta(B)\epsilon_t$, which is a complex-valued stationary process. Let $Z_\infty(\lambda)$ be an independent copy of $Z_0(\lambda) = \Theta(B)\epsilon_0$.

THEOREM 3.2. *Let $W_t = \underline{X}_t - \mu$, and suppose that $\{W_t\}$ has DFT process $d_W^s(\lambda)$ satisfying a functional central limit theorem (3.9). Let f denote the spectral density of the components of the process with indices in $\tilde{J}_{j+1}$, and evaluated at λ. For $j \geq 0$, if $d_{j+1} > 0$ then*

$$\widehat{d}_{j+1} \xrightarrow{\mathcal{D}} d_{j+1} \left[\Psi(e^{-i\lambda})\,\Xi\,\Psi(e^{i\lambda})' \right],$$

i.e., the last Schur complement of the stochastic matrix $\Psi(e^{-i\lambda})\,\Xi\,\Psi(e^{i\lambda})'$, where

$$\begin{aligned}
\Xi = &-\sum_{k=0}^{K} \int_{|r-s|\in C_k} \ddot{\Lambda}_b(r-s)\,\underline{B}(r)\,\underline{B}(s)^*\,dr\,ds \\
&+ \sum_{k=1}^{K} \left(\dot{\Lambda}_b^-(x_k) - \dot{\Lambda}_b^+(x_k) \right) \int_0^{1-x_k b} [\underline{B}(r)\,\underline{B}(r+x_k b)^* + \underline{B}(r+x_k b)\,\underline{B}(r)^*]\,dr \\
&+ \int_0^1 \dot{\Lambda}_b(1-r)\,[\underline{B}(1)\,\underline{B}(r)^* + \underline{B}(r)\,\underline{B}(1)^*]\,dr + \Lambda_b(0)\,\underline{B}(1)\,\underline{B}(1)^*.
\end{aligned}$$

If $d_{j+1} = 0$ and $j > 0$, then

$$T\,\widehat{d}_{j+1} \xrightarrow{\mathcal{D}} w^*\,Q\,w - w^*\,R\,\Psi_\sharp^* \left[\Psi_\sharp\,\Xi\,\Psi_\sharp^* \right]^{-1} \Psi_\sharp\,R^*\,w,$$

where

$$Q = -2\,\dot{\Lambda}_b^+(0)\,Cov[Z_0(\lambda)] + [Z_\infty(\lambda), Z_0(\lambda)] \begin{bmatrix} \underline{\underline{\Lambda}}_b(1,1) & -\underline{\underline{\Lambda}}_b(1,0) \\ -\underline{\underline{\Lambda}}_b(0,1) & \underline{\underline{\Lambda}}_b(0,0) \end{bmatrix} \begin{bmatrix} Z_\infty(\lambda)^* \\ Z_0(\lambda)^* \end{bmatrix}$$

$$R = [Z_\infty(\lambda), Z_0(\lambda)] \begin{bmatrix} \int_0^1 \dot{\underline{\Lambda}}_b(1,s)\,\underline{B}(s)^*\,ds + \underline{\Lambda}_b(1,1)\,\underline{B}(1)^* \\ -\int_0^1 \dot{\underline{\Lambda}}_b(0,s)\,\underline{B}(s)^*\,ds - \underline{\Lambda}_b(0,1)\,\underline{B}(1)^* \end{bmatrix}.$$

Also, if $d_1 = 0$ then $T\,\widehat{d}_1 \xrightarrow{\mathcal{D}} Q$.

REMARK 3.1. The alternative hypothesis case of Theorem 3.2 generalizes the results of PSJ to non-smooth kernels. As in that work, the choice of kernel has no impact on rate of convergence, but does influence the distribution of the limiting random variable. The null hypothesis case is novel, and has a very different rate of convergence. It is shown in the proof that the limiting distribution is continuous. In either case, the limit distribution could potentially have positive probability on negative values, depending on the kernel that is utilized.

From Theorems 3.1 and 3.2 it is immediate that Assumption 3.1 holds, and therefore Proposition 3.1 can be applied.

COROLLARY 3.1. *Suppose $\{W_t\}$ is strictly stationary and strong mixing, such that the assumptions of Theorem 3.1 hold. If n is chosen such that $n/T + 1/n \to 0$ as $T \to \infty$, then the conclusions of Proposition 3.1 hold, as applied with $\widehat{\theta}_T = \widehat{d}_{j+1}$ and $\tau_T = T$.*

4. APPLICATIONS

Throughout this section we consider co-integration testing for series that all require some $\delta(B)$ differencing to be rendered stationary, and with a λ satisfying $\delta(e^{-i\lambda}) = 0$. From the discussion in the Supporting Information, we can without loss of generality take $d_1 > 0$.

4.1. Simulation studies

Consider a bivariate ($m = 2$) process $\{X_t\}$ involving latent components for trend $\{\mu_t\}$, seasonality $\{\xi_t\}$, and irregular $\{\iota_t\}$

$$X_t = \mu_t + \xi_t + \iota_t,$$

where the vector white noise irregular has a covariance matrix Σ^ι. The trend component satisfies

$$(1 - B)\mu_t = \varepsilon_t^\mu,$$

for a vector white noise process $\{\varepsilon_t^\mu\}$ of covariance matrix Σ^μ. The seasonal component, with the seasonal period p (without the loss of generality we assume that p is even), is given as

$$\xi_t = \sum_{j=1}^{p/2} \xi_t^{(j)},$$

where the so-called atomic seasonal processes $\{\xi_t^{(j)}\}$ are defined through

$$\overbrace{(1 - 2\cos(2\pi j/p)B + B^2)}^{\delta^{(j)}(B)} \xi_t^{(j)} = \varepsilon_t^{(j)}, \quad 1 \le j \le p/2 - 1,$$
$$\overbrace{(1 + B)}^{\delta^{(p/2)}(B)} \xi_t^{(p/2)} = \varepsilon_t^{(p/2)},$$

for vector white noise processes $\{\varepsilon_t^{(j)}\}$, $1 \le j \le p/2$, with covariance matrices $\Sigma^{(j)}$. Vector white noise processes driving the trend, seasonal, and irregular components are independent. The application of the operator $U(B) = \prod_{j=1}^{p/2} \delta^{(j)}(B)$ to $\{\xi_t\}$ reduces it to stationarity. We apply our test to the differenced data $\{\underline{X}_t\}$, which has the form

$$\begin{aligned}\underline{X}_t &= (1 - B)U(B)X_t \\ &= U(B)\varepsilon_t^\mu + \sum_{j=1}^{p/2}(1 - B)\prod_{k\neq j}\delta^{(k)}(B)\varepsilon_t^{(j)} + (1 - B)U(B)\iota_t,\end{aligned}$$

and whose spectral density (with $z = \mathrm{e}^{-i\lambda}$) is given by

$$f(\lambda) = |U(z)|^2\Sigma^\mu + \sum_{j=1}^{p/2}|(1 - z)\prod_{k\neq j}\delta^{(k)}(z)|^2\Sigma^{(j)} + |(1 - z)U(z)|^2\Sigma^\iota.$$

In the first phase, we perform size and power studies involving trend and irregular. Trend co-integration (frequency $\lambda = 0$) occurs when Σ^{μ} has reduced rank, while Σ^{ι} is of full rank. We compare our test with the test of Nyblom and Harvey (2000, Section 4), which is applied to un-differenced data. We reiterate that our procedure also yields a rank configuration (not just the matrix rank), which is something that available parametric tests cannot do.

In the second phase, we examine the three-component model. Seasonal co-integration (at λ_j) takes place when one of the covariance matrices $\Sigma^{(j)}$ has reduced rank, while the other covariance matrices are of full rank. We focus on $p = 4$ (quarterly data), leading to two atomic seasonal processes with operators $\delta^{(1)}(B) = (1 + B^2)$ (frequency $\lambda_1 = \pi/2$, the fundamental frequency) and $\delta^{(2)}(B) = 1 + B$ (frequency $\lambda_2 = \pi$). We take $\Sigma^{(1)}$ of reduced rank, while keeping all the other covariance matrices of full rank. We compare our test with the test of Busetti (2006), which is closely related to the work of Nyblom and Harvey (2000).

All vector white noise processes are generated in the R package MASS from a mean-zero, multivariate normal distribution. We consider sample sizes $T \in \{1{,}000,\ 3{,}000,\ 5{,}000\}$ (after discarding the initial 300 observations) and the number of replicates is $R = 1{,}000$. The empirical size and power refer to the proportion of rejections of the null hypothesis (trend or seasonal co-integration) in R replicates. The nominal size of the test is $\in \{0.01, 0.05, 0.10\}$. In the estimation of the second Schur complement (of the spectral density matrix at the relevant frequency), we use the fixed-b parameter $\in \{0.1, 0.3, 0.5\}$ and two kernels, Bartlett and Parzen. In the subsampling, we consider subsample size chosen adaptively according to the methodology outlined in Götze and Rackauskas (2001) and Bickel and Sakov (2008), which is designed to improve the test's size in finite-sample. Namely, among the candidate subsample sizes n of the form $n_k = \lfloor q^k T \rfloor$ for $k = 1, 2, \ldots$ and $0 < q < 1$, we calculate the Kolmogorov–Smirnov distance between two consecutive subsampling distribution functions, for n_1 and n_2, for n_2 and n_3, and so on. We focus on the pair that minimizes this distance, and choose the smaller of the two values of n as the optimal one. In our study $q = 0.75$ and we consider block sizes n that constitute approximately 3–20% of the sample size T.

Although the competing tests of Nyblom and Harvey (2000) for trend co-integration and of Busetti (2006) for seasonal co-integration are applied to $\{X_t\}$, in the latter case one can make adjustments for stochastic trend, serial correlation and unattended unit roots (which we did). This involves pre-filtering the data with $(1 - B)(1 + B)$ and employing a Newey–West kernel with $m = \lceil T^{0.25} \rceil$ (see Section 5 of Busetti (2006) for details).

For a generic 2×2 spectral density matrix $f(\lambda)$ with coherence $\kappa_{21}(\lambda) = f_{21}(\lambda)/(\sqrt{f_{11}(\lambda)}\sqrt{f_{22}(\lambda)})$, $\kappa_{21}^*(\lambda) = \kappa_{12}(\lambda)$, we have the following generalized Cholesky decomposition :

$$f(\lambda) = \begin{bmatrix} f_{11}(\lambda) & \sqrt{f_{11}(\lambda)}\sqrt{f_{22}(\lambda)}\kappa_{12}(\lambda) \\ \sqrt{f_{11}(\lambda)}\sqrt{f_{22}(\lambda)}\kappa_{21}(\lambda) & f_{22}(\lambda) \end{bmatrix}$$

$$= \overbrace{\begin{bmatrix} 1 & 0 \\ \frac{\sqrt{f_{22}(\lambda)}}{\sqrt{f_{11}(\lambda)}}\kappa_{21}(\lambda) & 1 \end{bmatrix}}^{L} \overbrace{\begin{bmatrix} f_{11}(\lambda) & 0 \\ 0 & f_{22}(\lambda)(1 - \kappa_{12}(\lambda)\kappa_{21}(\lambda)) \end{bmatrix}}^{D} \overbrace{\begin{bmatrix} 1 & \frac{\sqrt{f_{22}(\lambda)}}{\sqrt{f_{11}(\lambda)}}\kappa_{12}(\lambda) \\ 0 & 1 \end{bmatrix}}^{L^*},$$

and hence the second Schur complement of $f(\lambda)$ is $d_2 = f_{22}(\lambda)(1 - \kappa_{12}(\lambda)\kappa_{21}(\lambda))$.

Likewise, for a generic 2×2 covariance matrix Σ, with variances σ_1^2, σ_2^2 and correlation coefficient ρ, we have the following (generalized) Cholesky decomposition:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix} = \overbrace{\begin{bmatrix} 1 & 0 \\ \frac{\sigma_2}{\sigma_1}\rho & 1 \end{bmatrix}}^{L} \overbrace{\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2(1-\rho^2) \end{bmatrix}}^{D} \overbrace{\begin{bmatrix} 1 & \frac{\sigma_2}{\sigma_1}\rho \\ 0 & 1 \end{bmatrix}}^{L^*},$$

and hence the second Schur complement of Σ is $d_2 = \sigma_2^2(1-\rho^2)$.

4.1.1. Trend and irregular. We set

$$\Sigma^{\iota} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \Sigma^{\mu} = q^{\mu} \begin{bmatrix} 1 & \rho^{\mu} \\ \rho^{\mu} & 1 \end{bmatrix},$$

where q^{μ} is defined as the *signal-to-noise ratio*. We consider $q^{\mu} \in \{1, 3, 5\}$. When ρ^{μ} equals 1, the white noise process $\{\varepsilon_t^{\mu}\}$ is collinear, and $\{X_t\}$ is trend co-integrated; otherwise, Σ^{μ} is of full rank. Hence in the size study we put $\rho^{\mu} = 1$, and when examining the power we set $\rho^{\mu} \in \{0.99, 0.98, \ldots, 0.95\}$.

4.1.2. Trend, seasonal, and irregular. When $p = 4$, we set

$$\Sigma^{\iota} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \Sigma^{\mu} = q^{\mu} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \Sigma^{(1)} = q^{(1)} \begin{bmatrix} 1 & \rho^{(1)} \\ \rho^{(1)} & 1 \end{bmatrix} \qquad \Sigma^{(2)} = q^{(2)} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We set the signal-to-noise ratios equal to $q^{\mu} = q^{(2)} = 1$ and $q^{(1)} \in \{1, 3, 5\}$. In the size study we set $\rho^{(1)} = 1$, and in the power study $\rho^{(1)} \in \{0.99, 0.98, \ldots, 0.95\}$.

Before assessing the finite-sample performance of the subsampling test, we examine large-sample distributions of $\widehat{d}_2$ and of $T\,\widehat{d}_2$ by simulating 1,000 replicates of the bivariate process for sample sizes $T = 1{,}000, 1{,}500, \ldots, 10{,}000$, under the null and alternative hypotheses. When the data is generated under the null, $\widehat{d}_2$ has a degenerate limit, but the limit of $T\,\widehat{d}_2$ is random (convergence for the Bartlett-based statistic is faster than for the Parzen-based statistic). When the simulated data falls under the alternative hypothesis, $\widehat{d}_2$ has a non-degenerate limiting distribution, and $T\,\widehat{d}_2$ diverges. These simulations corroborate in finite samples the asymptotic theory established above.

In addition, we perform simulations with four time series ($m = 4$), manipulating the spectral density matrix in such a way that the second and third time series can be co-integrated (trend or seasonal) or not. All the other parameters are set as in the bivariate case. We use only one sample size, $T = 1{,}000$, one fixed-b ratio ($b = 0.3$), and begin with $J = \{1\}$. Focusing on a given frequency, we start the recursive procedure by testing $H_0\colon\ d_{j+1} = 0$ versus $H_1\colon\ d_{j+1} > 0$ when $j{+}1 = 2$ with $J_1 = \{1\}$ and $\tilde{J}_2 = \{1, 2\}$. If we reject the null hypothesis (at a 5% level), we obtain $J_2 = \{1, 2\}$ and $\tilde{J}_3 = \{1, 2, 3\}$ in the next step, and test $H_0\colon\ d_{j+1} = 0$ versus $H_1\colon\ d_{j+1} > 0$ when $j{+}1 = 3$. If on this occasion the series are collinear (we fail to reject the null), then J_3 remains the same as J_2, and $\tilde{J}_4$ becomes $\{1, 2, 4\}$. The last test is $H_0\colon\ d_{j+1} = 0$ versus $H_1\colon\ d_{j+1} > 0$ for $j{+}1 = 4$.

4.1.3. Summary of the simulation results. In general, small size and power distortions can be observed in the subsampling test results. A small value of the fixed-b ratio parameter should not be used. This is consistent with the recommendation of PSJ, according to which one should not use a large exponent when exponentiating kernels (a small bandwidth ratio b corresponds to a

large exponent). The value of 0.3 offers an acceptable trade-off between size and power. The use of the Bartlett kernel causes the test to be slightly oversized (Tables C1 and C7 of the Supporting Information), especially in the three-component model. The performance of the test improves with increasing sample size and increasing signal-to-noise ratio, but large values of the sample size and signal-to-noise ratio are needed to obtain empirical rates that are close to the nominal levels. When the Parzen kernel is utilized, the test is somewhat conservative (Tables C2 and C8 of the Supporting Information). The test of Nyblom and Harvey (2000) has good size properties (cf. Table C3 of the Supporting Information); recall that this is a parametric test whose power is maximized against the homogeneous alternatives considered in this paper. Similarly, the test of Busetti (2006) also has good size properties (cf. Table C9 of the Supporting Information).

With regard to the empirical power, the subsampling test with the Bartlett kernel (Tables C4 and C10 of the Supporting Information) is more powerful than the one with the Parzen kernel (Tables C5 and C11 of the Supporting Information), which is to be expected considering that the former has a somewhat inflated size. The power of both tests increases with decreasing correlation, increasing sample size, and increasing signal-to-noise ratio. The test of Nyblom and Harvey (2000) has excellent power properties (cf. Table C6 of the Supporting Information). The same is true for the test of Busetti (2006)—see Table C12.

The properties of the subsampling test pertaining to the 4-variate case are in good agreement with the bivariate results just discussed. The results for the trend and irregular model are in Table C13 for the Bartlett kernel (empirical size and power) and in Table C14 for the Parzen kernel (empirical size and power). For the model with all three components, the respective results are in Tables C15 and C16 of the Supporting Information.

Our nonparametric test has been compared with competing procedures, which essentially test whether $\min_j d_j = 0$; in contrast, our test examines whether each d_j is zero. In the case of a bivariate process (wherein $d_1 = 0$ is ruled out) the procedures are equivalent, and for the trend–irregular case the Nyblom and Harvey (2000) method is superior in terms of size and power, as is the Busetti (2006) method for the trend–seasonal–irregular case. However, when the dimension is greater than two, our method yields additional results on rank configuration; we remark that it is important to have a large sample available and to avoid low bandwidth. For all the methods, a low signal content (relative to the noise) implies the need for a larger sample to obtain adequate size and power.

4.2. *Immigration data*

The data under consideration is daily, 6-variate New Zealand immigration data recorded between September 1, 1997, and July 31, 2012. The first, third, and fifth series pertain to arrivals, while the second, fourth, and sixth series measure departures. The first and second series count the category of a temporary visit, while the third and fourth series count visitors (short term), and the last two series measure permanent and long-term visits. All data were log-transformed, so that seasonal amplitude is not dependent on trend level. To difference the data to stationarity we consider the following differencing polynomials: $(1 - B)$ for the trend and $1 - 2\cos(2\pi \cdot 1/7)B + B^2$, $1 - 2\cos(2\pi \cdot 2/7)B + B^2$, $1 - 2\cos(2\pi \cdot 3/7)B + B^2$ for each harmonic of the weekly seasonality. The product of these polynomials leads to the operator $1 - B^7$ with which we filter the data. The data and the differenced data (5,441 observations) are shown in Figures 1 and 2, respectively, while the sample acf of the latter can be found in Figure 3.

The frequencies of interest are: 0, $2\pi \cdot 1/7$, $2\pi \cdot 2/7$, $2\pi \cdot 3/7$, and we also include $2\pi/365$, which corresponds to the annual seasonality, though in this case collinearity does not imply

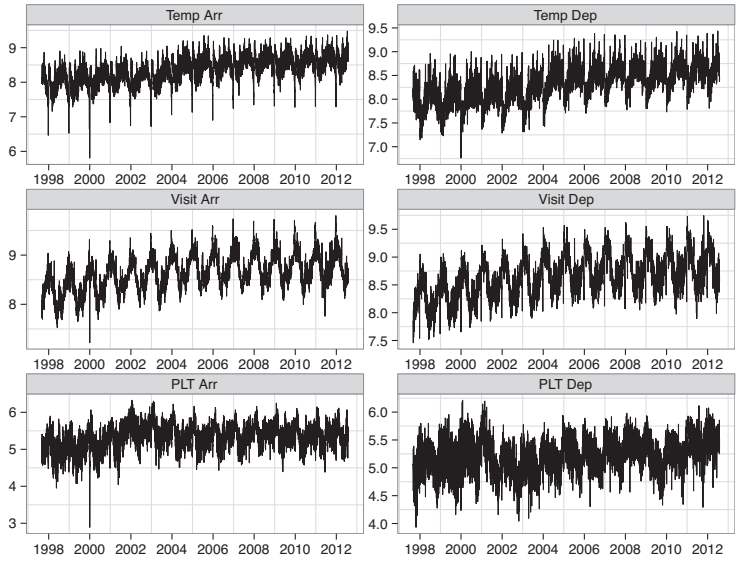

Figure 1. Six-variate (log-transformed) New Zealand immigration data.

co-integration. (Annual seasonality is arguably a stationary effect in the data, as evidence of the autocorrelation plot indicates a decay across annual lags towards zero.) At each frequency and for the initial $J = \{1\}$, the subsampling test is applied recursively to test $H_0\colon\ d_{j+1} = 0$ versus $H_1\colon\ d_{j+1} > 0$ starting with the index sets $J_1 = \{1\}$ and $\tilde{J}_2 = \{1, 2\}$ for $j{+}1 = 2$ and then modifying the sets accordingly, depending on whether the null is rejected or not (at a 5% level). We use 0.3 for the fixed-b ratio, the adaptive rule to choose the subsample size (with $q = 0.9$ instead of $q = 0.75$, i.e., a finer grid of candidate block sizes than in Section 4.1), and the two kernels employed previously.

The p-values (recall the discussion at the end of Section 3.1, regarding multiple testing) of the subsampling test are given in Table 1 (top half for the Bartlett kernel and bottom half for Parzen), together with the corresponding rank configurations. Bartlett-based p-values indicate that there are two common trends, as well as weekly seasonal co-integration effects: the first weekly seasonal is driven by a single process, whereas the second and third weekly seasonals have ranks two and three, respectively. There is also evidence of collinearity at the annual frequency (the rank is five). Results for the Parzen kernel tend to yield lower ranks, which is consistent with the simulation results (too few rejections of the null hypothesis). Again, should the six series be reordered then we cannot expect the rank configurations J to remain unchanged, although the cardinality of J would be unaltered (because matrix rank is invariant to permutation).

We also applied the methods of Nyblom and Harvey (2000) (frequency 0) and Busetti (2006) (frequencies $2\pi \cdot 1/7$, $2\pi \cdot 2/7$, and $2\pi \cdot 3/7$) to test the null hypothesis that the number of common components is 5 versus the alternative hypothesis that the number of common components is 6 (i.e., the time series are not co-integrated). For Nyblom and Harvey's test, the data was pre-filtered with $1 + B + \ldots + B^6$ to remove seasonality, while for Busetti's test we applied the polynomials $(1 - B)\,\Pi_{k \neq j}(1 - 2\cos(2\pi \cdot k/7)B + B^2)$ for each $j = 1, 2, 3$. The test statistics were 0.0307 for trend co-integration (the 5% critical value is 0.042) and 0.1250, 0.2695, and

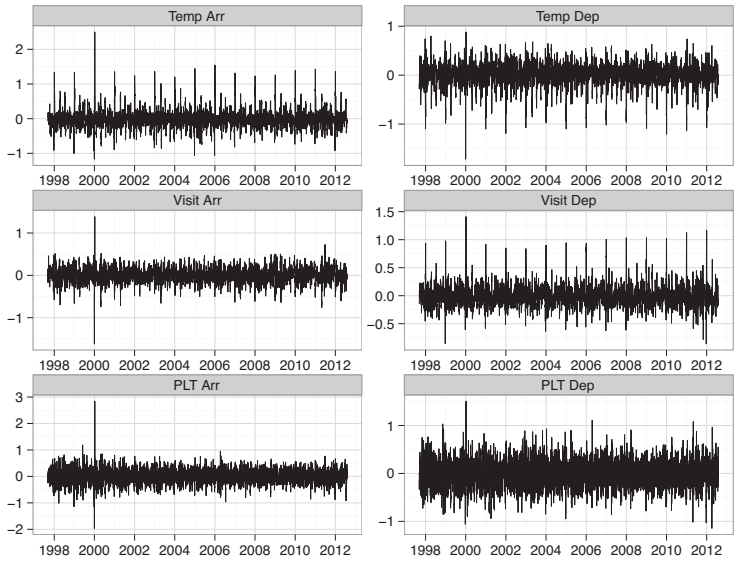

Figure 2. New Zealand immigration data, log-transformed and differenced via $1 - B^7$.

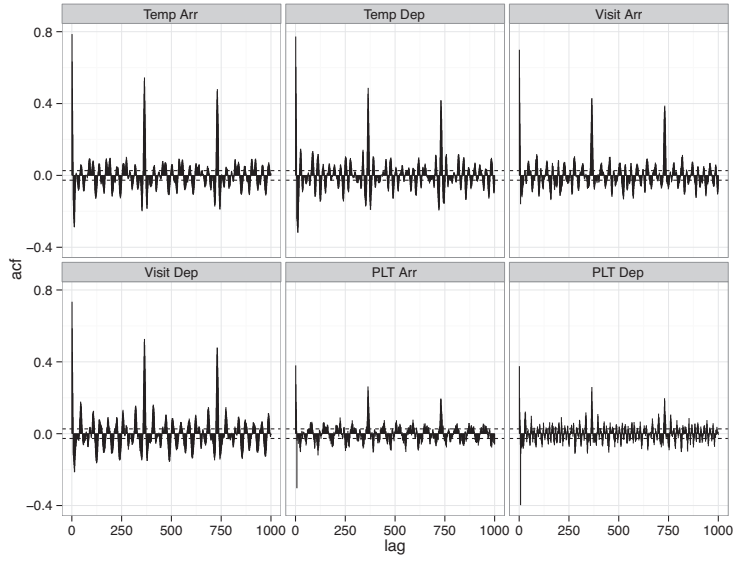

Figure 3. Sample acf of the differenced (via $1 - B^7$) log-transformed New Zealand immigration data.

0.6842 for seasonal co-integration at the three weekly frequencies, respectively (the 5% critical value is 0.157). Hence there is evidence that the six time series are not seasonally co-integrated at the second and third weekly frequencies. The discrepancy in the results between our method (which yield rank configuration) and the parametric methods (which yields the overall rank) may be due to the departure of the immigration data from an exact structural model.

Table 1. New Zealand immigration results.

$\lambda \backslash j+1$	2	3	4	5	6	J
			Bartlett kernel			
0	0.6108	0.2246	0.1908	0.0296	0.1331	{1, 5}
$2\pi \cdot 1/365$	⋆	0.0329	0.2913	⋆	0.0065	{1, 2, 3, 5, 6}
$2\pi \cdot 1/7$	0.2499	0.1901	0.2407	0.3176	0.4882	{1}
$2\pi \cdot 2/7$	0.1342	0.3110	0.0770	0.2669	0.0082	{1, 6}
$2\pi \cdot 3/7$	0.0486	0.0216	0.2277	0.2208	0.4120	{1, 2, 3}
			Parzen kernel			
0	1.0000	1.0000	1.0000	1.0000	0.8756	{1}
$2\pi \cdot 1/365$	1.0000	0.9894	1.0000	1.0000	0.0099	{1, 6}
$2\pi \cdot 1/7$	1.0000	0.7634	0.9673	0.9969	0.9886	{1}
$2\pi \cdot 2/7$	0.2160	0.9543	0.4268	0.8842	0.0119	{1, 6}
$2\pi \cdot 3/7$	0.1013	0.7414	0.6231	0.3605	0.6575	{1}

Note: p-values of the subsampling test H_0: $d_{j+1} = 0$ versus H_1: $d_{j+1} > 0$ are displayed for the New Zealand immigration data (log-transformed and differenced via $1 - B^7$), along with the final rank configuration J, using fixed-b ratio 0.3, subsample size n selected adaptively, and nominal size 5%. The ⋆ marks a p-value less than 0.0001.

Ultimately, we care about the rank configurations in order to achieve particular applications—the correct identification of collinearity, or co-integration, is extremely important for such applications as forecasting and signal extraction. In other work with the immigration data, the authors found that mis-specification of the rank configuration of any component can cause leakage of the signal during extraction. In essence, a component that is not co-integrated, but is wrongly identified as such, will be insufficiently suppressed if it is featured as noise in an extraction problem. For instance: if we desire to extract the trend and we indicate that the third weekly seasonal has rank 1 (as indicated by the Parzen kernel) when actually the rank is 3 (as indicated by the Bartlett kernel), then there may be insufficient suppression of dynamics at frequency $6\pi/7$, with the result that these oscillatory effects are found in the extracted trend. In such an application, the results from the Bartlett kernel would be preferred, given its superior size and power performance.

REFERENCES

Ahn, S. K., S. Cho and B. C. Seong (2004). Inference of seasonal cointegration: Gaussian reduced rank estimation and tests for various types of cointegration. *Oxford Bulletin of Economics and Statistics 66*, 261–84.

Axelsson, O. (1996). *Iterative Solution Methods*, Cambridge University Press, New York, NY.

Bickel, P. J. and A. Sakov (2008). On the choice of *m* in the *m* out of *n* bootstrap and confidence bounds for extrema. *Statistica Sinica 18*, 967–85.

Bierens, H. J. (1997). Nonparametric cointegration analysis. *Journal of Econometrics 77*, 379–404.

Busetti, F. (2006). Tests of seasonal integration and cointegration in multivariate unobserved component models. *Journal of Applied Econometrics 21*, 419–38.

Chan, N. H. and N. Terrin (1995). Inference for unstable long-memory fractional unit root autoregressions. *Annals of Statistics 23*, 1662–83.

Chen, Q. and Z. Fang (2018). *Improved inference on the rank of a matrix*. https://sites.google.com/site/zfangmetrics/research, arXiv preprint arXiv:181202337

Cubadda, G. (2001). Complex reduced rank models for seasonally cointegrated time series. *Oxford Bulletin of Economics and Statistics 63*, 497–511.

Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica 55*, 251–76.

Gómez, V. (2016). *Multivariate Time Series with Linear State Space Structure*. Springer, New York, NY.

Götze, F. and A. Rackauskas (2001). Adaptive choice of bootstrap sample sizes. In van der Aad, V., G. Mathisca de and K. Chris (Eds.), *State of the Art in Probability and Statistics, IMS Lecture Notes Monograph Series, 36*, 286–309. Cambridge University Press, Cambridge.

Harvey, A. (1989). *Forecasting, Structural Time Series Models, and the Kalman Filter*. Cambridge University Press, Cambridge.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control 12*, 231–54.

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

Koopman, S.J., M. Ooms and I. Hindrayanto (2012). A multivariate periodic unobserved components time series analysis for sectoral U.S. employment. In Bell, W., S. Holan and T. McElroy (Eds.), *Economic Time Series: Modeling and Seasonality*. Chapman and Hall, New York, NY.

McElroy, T. (2017). Multivariate seasonal adjustment, economic identities, and seasonal taxonomy. *Journal of Business and Economics Statistics 35*, 511–25.

McElroy, T. and D. Politis (2014). Spectral density and spectral distribution inference for long memory time series via fixed-b asymptotics. *Journal of Econometrics 182*, 211–25.

Nyblom, J. and A. Harvey (2000). Tests of common stochastic trends. *Econometric Theory 16*, 176–99.

Nyblom, J. and A. Harvey (2001). Testing against smooth stochastic trends. *Journal of Applied Econometrics 16*, 415–29.

Phillips, P. C. and V. Solo (1992). Asymptotics for linear processes. *Annals of Statistics 20*, 971–1001.

Phillips, P. C., Y. Sun and S. Jin (2006). Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. *International Economic Review 47*, 837–94.

Politis, D., J. Romano and M. Wolf (1999). *Subsampling*. Spinger, New York, NY.

Shintani, M. (2001). A simple cointegrating rank test without vector autoregression. *Journal of Econometrics 105*, 337–62.

Stock, J. H. and M. W. Watson (1988). Testing for common trends. *Journal of the American Statistical Association 83*, 1097–107.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online appendix of this article at the publisher's website:

Online Appendix
Replication Package

Co-editor Michael Jansson handled this manuscript.

APPENDIX A: SUPPLEMENTARY RESULTS

A.1. Determining collinear frequencies from exploratory analysis

In some cases we may want a method to identify frequencies λ where collinearity may be occurring, as it is too expensive to apply collinearity testing for all $\lambda \in [0, \pi]$. We here describe an algorithm for determining such λ from the autocovariance sequence—in empirical applications, sample autocovariances (possibly tapered) could be used instead. Let the autocovariance generating function be denoted $G(z) = \sum_{h\in\mathbb{Z}} \Gamma(h)\, z^h$, so that $f(\lambda) = G(e^{-i\lambda})$. Hence, $f(\lambda)$ is reduced rank if and only if $\det f(\lambda) = 0$, which holds if and only if $\det G(z)$ has a root $z = e^{-i\lambda}$. If the autocovariances truncate at some lag q, then $g(z) = \det[z^q\, G(z)]$ is a degree $2qm$ polynomial. If working with the tapered sample autocovariances $\Lambda_b(h/T)\,\widehat{\Gamma}(h)$, then $q = \lceil bT \rceil$. Now

$$
\begin{aligned}
z^q\, G(z) &= \sum_{k=0}^{2q} \Gamma(k-q)\, z^k = \Gamma(-q) \left(\sum_{k=0}^{2q} \Gamma(-q)^{-1}\, \Gamma(k-q)\, z^k \right) \\
&= \Gamma(-q) \left(I - \sum_{k=1}^{2q} A_k\, z^k \right),
\end{aligned}
$$

where $A_k = -\Gamma(-q)^{-1}\, \Gamma(k-q)$ for $1 \le k \le 2q$. The companion matrix corresponding to these coefficient matrices is

$$
A = \begin{bmatrix}
A_1 & \dots & A_{2q-1} & A_{2q} \\
I_m & \dots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
0 & \dots & I_m & 0
\end{bmatrix},
$$

and its $2qm$ eigenvalues are the reciprocals of the roots of

$$
\det[I - \sum_{k=1}^{2q} A_k\, z^k] = -\det \Gamma(q)\, g(z).
$$

In other words, we obtain the roots of $\det G(z)$ from the reciprocal eigenvalues of A, and check whether any of the roots lie on the unit circle. If they do, we take λ equal to the angular portion.

A.2. Permuting series and rank configuration

A rank configuration $J(\Sigma)$ can be determined from Σ by computing the generalized complex Cholesky decomposition $\Sigma = L\, D\, L^*$. A permutation Π applied to the series yields a new spectral density matrix $\Pi\, \Sigma\, \Pi'$, and we can obtain a new decomposition of the form $\widetilde{L}\, \widetilde{D}\, \widetilde{L}^*$. The new rank configuration $\widetilde{J}$ can be different from J, but $|\widetilde{J}| = |J|$ because they are both equal to the rank of Σ, which is unaffected by permutation. We provide an example with $m = 3$. With the entries of L denoted by ℓ_{ij}, we find that

$$
\Sigma = \begin{bmatrix}
d_1 & \ell_{21}\, d_1 & \ell_{31}\, d_1 \\
\ell_{21}\, d_1 & \ell_{21}^2\, d_1 + d_2 & \ell_{21}\, \ell_{31}\, d_1 + \ell_{32}\, d_2 \\
\ell_{31}\, d_1 & \ell_{21}\, \ell_{31}\, d_1 + \ell_{32}\, d_2 & \ell_{31}^2\, d_1 + \ell_{32}^2\, d_2 + d_3
\end{bmatrix}.
$$

Suppose that Π swaps the first two series, leaving the third series in place. Then we can solve for the components of the new Cholesky decomposition in terms of $d_1, d_2, d_3, \ell_{21}, \ell_{31}$, and ℓ_{32}:

$$\widetilde{d}_1 = \ell_{21}^2\, d_1 + d_2$$
$$\widetilde{\ell}_{21} = \frac{\ell_{21}\, d_1}{\ell_{21}^2\, d_1 + d_2}$$
$$\widetilde{d}_2 = \frac{d_1\, d_2}{\ell_{21}^2\, d_1 + d_2}$$
$$\widetilde{\ell}_{31} = \frac{\ell_{21}\, \ell_{31}\, d_1 + \ell_{32}\, d_2}{\ell_{21}^2\, d_1 + d_2}$$
$$\widetilde{\ell}_{32} = \ell_{31} - \ell_{21}\, \ell_{32}$$
$$\widetilde{d}_3 = d_3.$$

Hence, if $d_1 = 0$ (and $d_2, d_3 > 0$) then $\widetilde{d}_2 = 0$, and the rank configuration $J = \{2, 3\}$ gets transformed to $\widetilde{J} = \{1, 3\}$. However, if instead only d_2 equals zero, then $\widetilde{d}_2 = 0$ and $J = \{1, 3\}$ is left unchanged by the permutation.

In summary, rank configuration can be altered by permutation, but rank is not. In applications, an ordering of series is selected according to convenience, and thereafter the corresponding rank configuration can be determined.

A.3. Discussion of $d_1 = 0$

It can happen that $d_1 = 0$ for some λ, but this indicates that the spectral density for the first series $\{X_t^{(1)}\}$ is zero at frequency λ, and this is typically an indication of over-differencing. As this implies that the first series is noninvertible, and hence cannot be forecasted, it is preferable to avoid differencing that causes such a situation. If the data has not been differenced (i.e., it is stationary to begin with), then we can test for $d_1 = 0$ and proceed to d_2, etc. If all the series require $\delta(B)$ differencing to be rendered stationary, then $d_1 > 0$ for all λ such that $\delta(e^{-i\lambda}) = 0$, which is proved as follows. Because $e^{-i\lambda}$ is a root, we can write $\delta(B) = \widetilde{\delta}(B)\, \eta(B)$ with $\eta(B)$ given (cf. Section 2.1.2) as follows:

$$\eta(B) = \begin{cases} (1 - e^{i\lambda}\, B)\, (1 - e^{-i\lambda}\, B) & \text{if } \lambda \in (0, \pi) \\ 1 - B \quad \text{if } \lambda = 0 \\ 1 + B \quad \text{if } \lambda = \pi. \end{cases}$$

If $d_1 = 0$ at λ, then $f_{11}(\lambda) = 0$ and there exists a real-valued function $g(\omega)$ that is positive, and such that

$$f_{11}(\omega) = g(\omega)\, |\eta(e^{-i\omega})|^2.$$

Therefore $\widetilde{\delta}(B) X_t^{(1)}$ is stationary with spectral density g, which contradicts the assumption that $\delta(B)$ is the minimal differencing polynomial for the first series. Thus we must have $d_1 > 0$.

APPENDIX B: PROOFS OF RESULTS

Proof of Proposition 2.1. The result is essentially known (Gómez, 2016), but we provide some additional details. The r-dimensional process $\{Z_t\}$ satisfies

$$Z_t = \beta'\, X_t = \beta'\, \Phi\, X_{t-1} + \beta'\, \epsilon_t = (I_r + \beta'\, \alpha)\, Z_{t-1} + \beta'\, \epsilon_t. \tag{B.1}$$

This has the form of a VAR(1) of coefficient matrix $I_r + \beta'\, \alpha$ driven by white noise of variance $\beta'\, \Sigma\, \beta$. We show that $I_r + \beta'\, \alpha$ is stable, i.e., has all eigenvalues less than one in magnitude. First, observe that there

exists an $m \times (m-r)$-dimensional matrix $\beta_\perp$ whose columns span the null space of β, so that $\beta' \beta_\perp = 0$. By assumption, $m-r$ eigenvalues of Φ equal one, and r eigenvalues have magnitude less than one. Clearly, (θ, w) is an eigen-pair of $\beta' \alpha$ if and only if $(\theta+1, w)$ is an eigen-pair of $I_r + \beta' \alpha$, which holds if and only if $(\theta+1, \alpha w)$ is an eigen-pair of Φ (because $\Phi = I_m + \alpha \beta'$). Since $\beta' \alpha$ is full rank, $\theta \neq 0$. On the other hand, $\theta + 1$ is an eigenvalue of Φ only if either $|\theta+1| < 1$ or $\theta + 1 = 1$, but this latter is impossible, because $\theta \neq 0$. Hence the eigenvalues of $I_r + \beta' \alpha$ have magnitude less than one, and $\{Z_t\}$ is stationary.

For the differences, we have

$$\underline{X}_t = \Phi X_{t-1} - X_{t-1} + \epsilon_t = \Pi X_{t-1} + \epsilon_t = \alpha Z_{t-1} + \epsilon_t.$$

We already know that $\{Z_t\}$ is stationary, and hence so is $\{\underline{X}_t\}$. Using the representation (B.1), we have

$$Z_t = (I_r - [I_r + \beta' \alpha]B)^{-1} \beta' \epsilon_t$$

$$\underline{X}_t = \left(I_m + \alpha (I_r - [I_r + \beta' \alpha]B)^{-1} \beta' B\right) \epsilon_t.$$

This derives the stated formula for $\Theta(B)$, and

$$\beta' \Theta(B) = (I_r - I_r B)(I_r - [I_r + \beta' \alpha]B)^{-1} \beta',$$

so that $\beta' \Theta(1) = 0$. Hence the spectral density of $\{\underline{X}_t\}$, namely $\Theta(e^{-i\lambda}) \Sigma \Theta(e^{i\lambda})'$, has rank $m-r$ at frequency $\lambda = 0$. Next, letting $V = \beta (\beta' \beta)^{-1} \beta'$, we see that $U V = 0$, and both U and V are idempotent. Moreover, $\Pi U = 0$, and $U + V = I_m$. Hence $(I_m - U B)(I_m - V B) = I_m - I_m B$, and

$$\begin{aligned} I_m - \Phi B &= I_m - I_m B - \Pi B = (I_m - U B)(I_m - V B) - \Pi (I_m - U B) B \\ &= [I_m - (V + \Pi) B] (I_m - U B). \end{aligned}$$

Setting $\Phi^* = V + \Pi$ yields the stated factorization. To prove that it is stable, observe that (λ, v) is an eigen-pair of Φ^* if and only if either $\lambda = 0$ with v in the range space of $\beta_\perp$, or (λ, v) is an eigen-pair of Φ with v in the range space of β. In the latter case, either $|\lambda| < 1$ or $\lambda = 1$; but $(1, v)$ is an eigen-pair of Φ if and only if $(0, v)$ is an eigen-pair of Π, which holds if and only if v is in the range space of $\beta_\perp$. Therefore $|\lambda| < 1$. □

Proof of Proposition 2.2. The r-dimensional process $\{Z_t\}$ satisfies

$$\begin{aligned} Z_t &= \beta' \Phi X_{t-1} + \beta' \epsilon_t \\ &= \beta' \Phi^2 X_{t-2} + \beta' \Phi \epsilon_{t-1} + \beta' \epsilon_t \\ &= \beta' \left(\alpha \beta' + 2 \cos(\omega) \Phi - I_m\right) X_{t-2} + \beta' \Phi \epsilon_{t-1} + \beta' \epsilon_t \\ &= (\beta' \alpha - I_r) Z_{t-2} + 2 \cos(\omega) \beta' (X_{t-1} - \epsilon_{t-1}) + \beta' \Phi \epsilon_{t-1} + \beta' \epsilon_t \\ &= 2 \cos(\omega) Z_{t-1} + (\beta' \alpha - I_r) Z_{t-2} + (\beta' \Phi - 2 \cos(\omega) \beta') \epsilon_{t-1} + \beta' \epsilon_t. \end{aligned}$$

This is re-expressed as

$$(I_r - 2 \cos(\omega) I_r B + (I_r - \beta' \alpha) B^2) Z_t = \beta' (I_m + (\Phi - 2 \cos(\omega) I_m) B) \epsilon_t.$$

Next, the matrix polynomial $I_r - 2 \cos(\omega) I_r B + (I_r - \beta' \alpha) B^2$ is stable if and only if the matrix

$$A = \begin{bmatrix} 2 \cos(\omega) I_r & \beta' \alpha - I_r \\ I_r & 0 \end{bmatrix}$$

has eigenvalues of magnitude less than one. If (λ, v) is an eigen-pair of A, and v has component vectors v_1 and v_2, then $v_1 = \lambda v_2$ and $2 \cos(\omega) v_1 + (\beta' \alpha - I_r) v_2 = \lambda v_1$, which implies that $(\lambda^2 - 2 \cos(\omega) \lambda + 1) v_2 = \beta' \alpha v_2$, indicating that $(\lambda^2 - 2 \cos(\omega) \lambda + 1, v_2)$ is an eigen-pair of $\beta' \alpha$. As argued in the proof of Proposition 2.1, we can show that (θ, w) is an eigen-pair of $\beta' \alpha$ if and only if $(\theta, \alpha w)$ is an eigen-pair of Π, which holds if and only if $(\phi, \alpha w)$ is an eigen-pair of Φ, such that $\theta = \phi^2 - 2 \cos(\omega) \phi + 1$. Thus, any

eigenvalue λ of A is an eigenvalue of Φ; hence either $|\lambda| < 1$ or $\lambda = e^{\pm i\omega}$, but in this case we would have the corresponding eigenvalue θ of $\beta' \alpha$ equal to zero, and this matrix is full rank. Hence A is stable. This means that $\{Z_t\}$ has a stable VARMA(2,1) representation.

Next, for the differences we have

$$\begin{aligned}\underline{X}_t &= X_t - 2\cos(\omega) X_{t-1} + X_{t-2} \\ &= \Phi^2 X_{t-2} + \Phi \epsilon_{t-1} + \epsilon_t - 2\cos(\omega)(\Phi X_{t-2} + \epsilon_{t-1}) + X_{t-2} \\ &= \Pi X_{t-2} + (\Phi - 2\cos(\omega) I_m)\epsilon_{t-1} + \epsilon_t,\end{aligned}$$

and $\Pi X_{t-2} = \alpha Z_{t-2}$, so that $\{\underline{X}_t\}$ is stationary. Putting this together with the VARMA(2,1) representation for $\{Z_t\}$ yields $\underline{X}_t = \Theta(B)\epsilon_t$ as stated. It follows that

$$\beta' \Theta(B) = (I_r - 2\cos(\omega) I_r B + I_r B^2)\left(I_r - 2\cos(\omega) I_r B + (I_r - \beta'\alpha) B^2\right)^{-1} \beta'(I_m + (\Phi - 2\cos(\omega) I_m) B),$$

so that $\beta' \Theta(e^{\pm i\omega}) = 0$, and the spectral density has rank $m - q$ at frequencies $\pm\omega$. For the factorization, let U and V be defined as in the proof of Proposition 2.1, and observe that

$$(I_m - U e^{-i\omega} B)(I_m - V e^{-i\omega} B) = (I_m - V e^{-i\omega} B)(I_m - U e^{-i\omega} B) = I_m - I_m e^{-i\omega} B,$$

and the same result holds for $e^{i\omega}$ as well. Also, the differencing operator $D(B) = (I_m - U e^{-i\omega} B)(I_m - U e^{i\omega} B)$, so that $\Phi^*(B) D(B) = \delta(B) - \Pi B^2$, where $\delta(B) = I_m(1 - 2\cos(\omega)B + B^2)$. Finally, $I_m + (\Phi - 2\cos(\omega) I_m) B$ times $I_m - \Phi B$ equals $\delta(B) - \Pi B^2$, or in other words $\Theta^*(B)(I_m - \Phi B) = \Phi^*(B) D(B)$. Hence the process $\{Y_t\}$ satisfies

$$\Phi^*(B) Y_t = \Phi^*(B) D(B) X_t = \Theta^*(B)(I_m - \Phi B) X_t = \Theta^*(B)\epsilon_t.$$

To show that $\Phi^*(B)$ is stable, it suffices to determine the eigenvalues of the matrix

$$C = \begin{bmatrix} 2\cos(\omega) V & \Pi - V \\ I_m & 0 \end{bmatrix}.$$

We find that an eigen-pair (λ, x) of C must satisfy $\lambda^2 x_2 = 2\cos(\omega)\lambda V x_2 + (\Pi - V) x_2$, where x consists of components x_1 and x_2. Because x_2 must belong to either the range of $\beta_\perp$ or the range of β, we find in the former case that $\lambda = 0$ (since $V \beta_\perp = \Pi \beta_\perp = 0$). In the latter case, we have $x_2 = \beta y$ for some y, and find that $(\lambda^2 - 2\cos(\omega)\lambda + 1, \beta y)$ is an eigen-pair of Π. This implies that $(\lambda, \beta y)$ is an eigen-pair of Φ, so either $|\lambda| < 1$ or $\lambda = e^{\pm i\omega}$. The latter case entails a zero eigenvalue of Π, which is full rank, so we conclude that C is stable. Hence $\Phi^*(B)$ is invertible, and $\{Y_t\}$ is a VARM(2,1) process. □

Proof of Proposition 2.3. For any $j \geq 1$ the matrix $\Sigma_j(J_j)$ can be represented as $H \Sigma_j K'$, where H is a selection matrix with rows given by unit vectors e_k such that $k \in J_j$. In other words, H looks like an identity matrix with rows corresponding to indices in J_j^c removed. From (2.2) we obtain

$$\Sigma_j(J_j) = \sum_{k \in J_j} d_k (HL(k))(HL(k))^*.$$

Now each vector $HL(k)$ is given by $L(k)$ with components removed corresponding to indices in J_j^c, and hence the square matrix $\widetilde{L}$ with columns given by each $HL(k)$ for $k \in J_j$ is unit lower triangular. Thus, $\Sigma_j(J_j) = \widetilde{L}\,\widetilde{D}\,\widetilde{L}^*$, where $\widetilde{D}$ is diagonal consisting of d_k with $k \in J_j$; hence $\Sigma_j(J_j)$ is Hermitian pd. Next, for the $j = 1$ case we have $J_1 = \{1\}$ by assumption, and direct calculation yields $\det \Sigma_2(\{1, 2\}) / \det \Sigma_1(\{1\}) = (d_1 d_2)/d_1 = d_2$. For the general case of (2.4), consider the matrix

$$\Sigma_{j+1}(\tilde{J}_{j+1}) = \begin{bmatrix} \Sigma_j(J_j) & \underline{\sigma}_j(J_j) \\ \underline{\sigma}_j^*(J_j) & s_{j+1} \end{bmatrix} = H \Sigma_{j+1} H',$$

where the selection matrix H consists of unit vectors e_k such that $k \in \tilde{J}_{j+1}$. Again using (2.2), it is clear that $\Sigma_{j+1}(\tilde{J}_{j+1})$ is Hermitian non-negative definite, with Schur complements d_k for $k \in \tilde{J}_{j+1}$; the last Schur complement d_{j+1} may possibly be zero, but the others are all positive. Using the matrix identity of Axelsson (1996, p.93),

$$\det \Sigma_{j+1}(\tilde{J}_{j+1}) = \det \Sigma_j(J_j) \cdot \left(s_{j+1} - \underline{\sigma}_j^*(J_j)\, \Sigma_j(J_j)^{-1}\, \underline{\sigma}_j(J_j)\right).$$

Hence, $s_{j+1} - \underline{\sigma}_j^*(J_j)\, \Sigma_j(J_j)^{-1}\, \underline{\sigma}_j(J_j)$ is equal to the ratio of determinants stated in (2.4), and is also by definition the Schur complement of $\Sigma_j(J_j)$ in $\Sigma_{j+1}(\tilde{J}_{j+1})$, which must in turn be equal to the final d_k in the generalized Cholesky decomposition of $\Sigma_{j+1}(\tilde{J}_{j+1})$, namely d_{j+1}. □

Proof of Proposition 3.1. To prove the convergences, we can apply Theorem 3.2.1 of Politis, Romano, and Wolf (1999) to either the H_0 or H_a cases, noting that $L_{T,n}(x|1)$ is the subsampling distribution estimator corresponding to $\widehat{\theta}_T$ (with no normalizing rate). The consistency of the test is proved by using (3.9):

$$\begin{aligned}
&\mathbb{P}\left[\ \tau_T\, \widehat{\theta}_T \geq L_{T,n}^{-1}(1-\alpha|\tau_\cdot)|H_a\right] \\
&\quad = \mathbb{P}[\widehat{\theta}_T \geq (\tau_n/\tau_T)\, L_{T,n}^{-1}(1-\alpha|1)|H_a] \\
&\quad \approx \mathbb{P}[W_a \geq (\tau_n/\tau_T)\, \overline{F}_a^{-1}(\alpha)] \to 1,
\end{aligned}$$

so long as W_a is non-negative, and $\tau_n/\tau_T \to 0$. □

Proof of Proposition 3.2. The first formula for d_{j+1} follows from application of (2.3) to (3.10), and this equals $w^* f w$ by direct calculation. Because $\Psi^*\, w$ is a vector of length $\sqrt{w^*\, f\, w} = \sqrt{d_{j+1}}$, which equals zero under the null hypothesis, we have $\Psi^*\, w = 0$ and hence $\Psi_\flat = v^*\, \Psi_\sharp$ under the null hypothesis. □

Proof of Theorem 3.1. We will prove a joint weak convergence for the cosine and sine transforms, defined as the real and imaginary parts respectively of $d_W^s(\lambda)$, where $W_t = \underline{X}_t - \mu$. First we focus on the case that $\lambda \neq 0, \pi$. Let η and τ be real vectors, and consider weak convergence of the scalar process

$$\eta'\, \mathcal{R}[d_W^s(\lambda)] + \tau'\, \mathcal{I}[d_W^s(\lambda)] = T^{-1/2} \sum_{t=1}^{[Ts]} \left(\eta' W_t \cos(\lambda t) - \tau' W_t \sin(\lambda t)\right).$$

To do so, we consider the finite-dimensional distributions; for any n, and any real scalars $\nu_1, \ldots, \nu_n$, consider the convergence of

$$\sum_{k=1}^{n} \nu_k\, \left(\eta'\, \mathcal{R}[d_W^{s_k}(\lambda)] + \tau'\, \mathcal{I}[d_W^{s_k}(\lambda)]\right). \tag{B.2}$$

Proceeding by the method of cumulants, we know that the mean is zero, and the variance is

$$\begin{aligned}
&\textstyle\sum_{k_1,k_2=1}^{n} \nu_{k_1} \nu_{k_2}\, \mathrm{Cov}\left(\eta'\, \mathcal{R}[d_W^{s_{k_1}}(\lambda)] + \tau'\, \mathcal{I}[d_W^{s_{k_2}}(\lambda)],\, \eta'\, \mathcal{R}[d_W^{s_{k_2}}(\lambda)] + \tau'\, \mathcal{I}[d_W^{s_{k_2}}(\lambda)]\right) \\
&\quad = \textstyle\sum_{k_1,k_2=1}^{n} \nu_{k_1} \nu_{k_2}\, \mathrm{Cov}\left(\eta'\, \mathcal{R}[d_W^{s_{k_1}}(\lambda)],\, \eta'\, \mathcal{R}[d_W^{s_{k_2}}(\lambda)]\right) \\
&\quad + \textstyle\sum_{k_1,k_2=1}^{n} \nu_{k_1} \nu_{k_2}\, \mathrm{Cov}\left(\eta'\, \mathcal{R}[d_W^{s_{k_1}}(\lambda)],\, \tau'\, \mathcal{I}[d_W^{s_{k_2}}(\lambda)]\right) \\
&\quad + \textstyle\sum_{k_1,k_2=1}^{n} \nu_{k_1} \nu_{k_2}\, \mathrm{Cov}\left(\tau'\, \mathcal{I}[d_W^{s_{k_1}}(\lambda)],\, \eta'\, \mathcal{R}[d_W^{s_{k_2}}(\lambda)]\right) \\
&\quad + \textstyle\sum_{k_1,k_2=1}^{n} \nu_{k_1} \nu_{k_2}\, \mathrm{Cov}\left(\tau'\, \mathcal{I}[d_W^{s_{k_1}}(\lambda)],\, \tau'\, \mathcal{I}[d_W^{s_{k_2}}(\lambda)]\right).
\end{aligned}$$

Each of these covariances involves a sum of $[Ts_{k_1}]$ random variables with a sum of $[Ts_{k_2}]$ random variables, so that when $k_1 \neq k_2$ the number of summands differs. However, using stationarity it is possible to reduce

the calculation to covariances of sums with an equal number of summands. As an intermediary result, for random variables $\{U_t\}$ and $\{V_t\}$ that are jointly weakly stationary we have

$$\begin{aligned}\mathrm{Cov}\left(\sum_{t=1}^{N+L} U_t, \sum_{t=1}^{N+L} V_t\right) &= \mathrm{Cov}\left(\sum_{t=1}^{N} U_t, \sum_{t=1}^{N} V_t\right) + \mathrm{Cov}\left(\sum_{t=N+1}^{N+L} U_t, \sum_{t=N+1}^{N+L} V_t\right) \\ &+ \mathrm{Cov}\left(\sum_{t=1}^{N} U_t, \sum_{t=N+1}^{N+L} V_t\right) + \mathrm{Cov}\left(\sum_{t=N+1}^{N+L} U_t, \sum_{t=1}^{N} V_t\right) \\ &= -\mathrm{Cov}\left(\sum_{t=1}^{N} U_t, \sum_{t=1}^{N} V_t\right) + \mathrm{Cov}\left(\sum_{t=1}^{L} U_t, \sum_{t=1}^{L} V_t\right) \\ &+ \mathrm{Cov}\left(\sum_{t=1}^{N} U_t, \sum_{t=1}^{N+L} V_t\right) + \mathrm{Cov}\left(\sum_{t=1}^{N+L} U_t, \sum_{t=1}^{N} V_t\right),\end{aligned}$$

which implies

$$\begin{aligned}&\mathrm{Cov}\left(\sum_{t=1}^{N} U_t, \sum_{t=1}^{N+L} V_t\right) + \mathrm{Cov}\left(\sum_{t=1}^{N+L} U_t, \sum_{t=1}^{N} V_t\right) \\ &= \mathrm{Cov}\left(\sum_{t=1}^{N+L} U_t, \sum_{t=1}^{N+L} V_t\right) + \mathrm{Cov}\left(\sum_{t=1}^{N} U_t, \sum_{t=1}^{N} V_t\right) - \mathrm{Cov}\left(\sum_{t=1}^{L} U_t, \sum_{t=1}^{L} V_t\right).\end{aligned} \tag{B.3}$$

Letting A^s and B^s denote either of the terms $\eta' \mathcal{R}[d_W^s(\lambda)]$ and $\tau' \mathcal{I}[d_W^s(\lambda)]$, we see from (B.3) that

$$\begin{aligned}&\sum_{k_1,k_2=1}^{n} \nu_{k_1}\nu_{k_2}\,\mathrm{Cov}\,(A^{s_{k_1}}, B^{s_{k_2}}) \\ &= \sum_{k=1}^{n} \nu_k^2\,\mathrm{Cov}\,(A^{s_k}, B^{s_k}) + \sum_{k_1<k_2} \nu_{k_1}\nu_{k_2}\left[\mathrm{Cov}\,(A^{s_{k_1}}, B^{s_{k_2}}) + \mathrm{Cov}\,(A^{s_{k_2}}, B^{s_{k_1}})\right] \\ &= \sum_{k=1}^{n} \nu_k^2\,\mathrm{Cov}\,(A^{s_k}, B^{s_k}) + \sum_{k_1<k_2} \nu_{k_1}\nu_{k_2}\left[\mathrm{Cov}\,(A^{s_{k_2}}, B^{s_{k_2}}) + \mathrm{Cov}\,(A^{s_{k_1}}, B^{s_{k_1}}) - \mathrm{Cov}\left(A^{s_{k_2}-s_{k_1}}, B^{s_{k_2}-s_{k_1}}\right)\right].\end{aligned}$$

To analyse the covariance of sums with an equal number of summands, we present some preliminary results: for any s_k, it follows from results in Brillinger (2001) that (suppressing s_k and λ in the notation for the DFT)

$$\mathrm{Cov}\left(\eta' d_W, \eta' \overline{d_W}\right) = \sum_{h_1,h_2=1}^{n} \eta_{h_1}\eta_{h_2}\,\mathrm{Cov}\left(d_{W_{h_1}}, \overline{d_{W_{h_2}}}\right) = O(T^{-1}) + s_k\,\eta' f(\lambda)\,\eta. \tag{B.4}$$

However, the covariance of the DFT with itself (without conjugation) is just $O(T^{-1})$. From (B.4), it follows that

$$\begin{aligned}\mathrm{Cov}\left(\eta' \mathcal{R}[d_W], \eta' \mathcal{R}[d_W]\right) &= O(T^{-1}) + s_k\,\eta' f(\lambda)\,\eta/2 \\ \mathrm{Cov}\left(\tau' \mathcal{I}[d_W], \tau' \mathcal{I}[d_W]\right) &= O(T^{-1}) + s_k\,\tau' f(\lambda)\,\tau/2 \\ \mathrm{Cov}\left(\eta' \mathcal{R}[d_W], \tau' \mathcal{I}[d_W]\right) &= O(T^{-1}) + s_k\,\frac{i}{4}\left(\eta' f(\lambda)\,\tau - \tau' f(\lambda)\,\eta\right).\end{aligned}$$

Applying these results in turn to the four terms of the variance of (B.2), in conjunction with (B.3) yields that the variance is

$$O(T^{-1}) + \left(\sum_{k=1}^{n} \nu_k^2 s_k + 2 \sum_{k_1<k_2} \nu_{k_1} \nu_{k_2} s_{k_1} \right) \left[\eta' f(\lambda) \eta + \tau' f(\lambda) \tau + i(\eta' f(\lambda) \tau - \tau' f(\lambda) \eta) \right] /2.$$

The expression in parentheses is the variance of $\sum_{k=1}^{n} \nu_k B(s_k)$, where B is a standard Brownian motion. Moreover, the variance of

$$\eta' \mathcal{R}[\Psi \sum_{k=1}^{n} \nu_k (B_R(s_k) - i B_I(s_k))] + \tau' \mathcal{I}[\Psi \sum_{k=1}^{n} \nu_k (B_R(s_k) - i B_I(s_k))]$$

equals the asymptotic variance; this is because the variance equals

$$\sum_{k_1,k_2=1}^{n} \nu_{k_1} \nu_{k_2} \left(\eta' \mathcal{R}\Psi + \tau' \mathcal{I}\Psi \right) \mathrm{Cov}[B_R(s_{k_1}), B_R(s_{k_2})] \left(\eta' \mathcal{R}\Psi + \tau' \mathcal{I}\Psi \right)'$$

$$+ \sum_{k_1,k_2=1}^{n} \nu_{k_1} \nu_{k_2} \left(\tau' \mathcal{R}\Psi - \eta' \mathcal{I}\Psi \right) \mathrm{Cov}[B_I(s_{k_1}), B_I(s_{k_2})] \left(\tau' \mathcal{R}\Psi - \eta' \mathcal{I}\Psi \right)'$$

$$= \frac{1}{2} \sum_{k_1,k_2=1}^{n} \nu_{k_1} \nu_{k_2} \left(\sum_{k=1}^{n} \nu_k^2 s_k + 2 \sum_{k_1<k_2} \nu_{k_1} \nu_{k_2} s_{k_1} \right) \left[\eta' \mathcal{R}\Psi \mathcal{R}\Psi' \eta + \tau' \mathcal{I}\Psi \mathcal{I}\Psi' \tau + 2 \eta' \mathcal{R}\Psi \mathcal{I}\Psi' \tau \right]$$

$$+ \frac{1}{2} \sum_{k_1,k_2=1}^{n} \nu_{k_1} \nu_{k_2} \left(\sum_{k=1}^{n} \nu_k^2 s_k + 2 \sum_{k_1<k_2} \nu_{k_1} \nu_{k_2} s_{k_1} \right) \left[\tau' \mathcal{R}\Psi \mathcal{R}\Psi' \tau + \eta' \mathcal{I}\Psi \mathcal{I}\Psi' \eta - 2 \eta' \mathcal{I}\Psi \mathcal{R}\Psi' \tau \right],$$

and f satisfies

$$f = \Psi \Psi^* = \left(\mathcal{R}\Psi \mathcal{R}\Psi' + \mathcal{I}\Psi \mathcal{I}\Psi' \right) + i \left(\mathcal{I}\Psi \mathcal{R}\Psi' - \mathcal{R}\Psi \mathcal{I}\Psi' \right),$$

thereby implying that

$$\eta' f \eta = \eta' \left(\mathcal{R}\Psi \mathcal{R}\Psi' + \mathcal{I}\Psi \mathcal{I}\Psi' \right) \eta$$
$$\tau' f \tau = \tau' \left(\mathcal{R}\Psi \mathcal{R}\Psi' + \mathcal{I}\Psi \mathcal{I}\Psi' \right) \tau$$
$$\eta' f \tau - \tau' f \eta = 2i \left(\eta' \mathcal{I}\Psi \mathcal{R}\Psi' \tau - \eta' \mathcal{R}\Psi \mathcal{I}\Psi' \tau \right).$$

Hence, the first and second cumulants of d_W^s converge jointly to those of $\Psi(B_R(s) - iB_I(s))$. The summability conditions on cumulants, together with results from Brillinger (2001), ensure that higher-order cumulants tend to zero. Next, by substituting a linearly interpolated process for the DFT, as in the proof of Theorem 1 of McElroy and Politis (2014), we can establish tightness via Problem 4.11 of Karatzas and Shreve (1991) and the uniform integrability condition. This completes the case $\lambda \in (0, \pi)$. If $\lambda = 0, \pi$, the imaginary part of the DFT is zero, and the DFT equals its real part. This entails that

$$\mathrm{Cov}\left(\eta' \mathcal{R}[d_W], \eta' \mathcal{R}[d_W] \right) = O(T^{-1}) + s_k \, \eta' f(\lambda) \eta,$$

and hence the limiting variance of (B.2) is

$$O(T^{-1}) + \left(\sum_{k=1}^{n} \nu_k^2 s_k + 2 \sum_{k_1<k_2} \nu_{k_1} \nu_{k_2} s_{k_1} \right) \left[\eta' f(\lambda) \eta \right].$$

This is the limiting variance of $\eta' \Psi \sum_{k=1}^{n} \nu_k B_R(s_k)$, using the fact that Ψ and f are real-valued when $\lambda = 0, \pi$. This completes the proof. □

The following two lemmas follow immediately from summation by parts.

LEMMA 0.1. *Given complex sequences $y_0, y_1, \ldots$ and $x_0, x_1, \ldots$ and a function $A(h)$ for $h \in \mathbb{Z}$,*

$$\sum_{\ell,t=1}^{T} A(t-\ell)\,[y_t - y_{t-1}]\,x_\ell^* = \textstyle\sum_{\ell=1}^{T}\sum_{t=1}^{T-1}[A(t-\ell) - A(t+1-\ell)]\,y_t\,x_\ell^* + \sum_{\ell=1}^{T} A(T-\ell)\,y_T\,x_\ell^* - \sum_{\ell=1}^{T} A(1-\ell)\,y_0\,x_\ell^*.$$

LEMMA 0.2. *Given complex sequences $y_0, y_1, \ldots$ and $x_0, x_1, \ldots$ and a symmetric function $A(h)$ for $h \in \mathbb{Z}$,*

$$\begin{aligned}\sum_{t,\ell=1}^{T} A(t-\ell)\,[y_t - y_{t-1}]\,[x_\ell - x_{\ell-1}]^* = &\sum_{t,\ell=1}^{T-1}[2\,A(t-\ell) - A(t+1-\ell) - A(t-1-\ell)]\,y_t\,x_\ell^* \\ &+ \sum_{t=1}^{T-1}[A(t) - A(t-1)]\,[y_0\,x_t^* + y_t\,x_0^*] \\ &+ \sum_{t=1}^{T-1}[A(t-T) - A(t+1-T)]\,[y_T\,x_t^* + y_t\,x_T^*] \\ &+ A(0)\,[y_0\,x_0^* + y_T\,x_T^*] - A(T-1)\,[y_0\,x_T^* + y_T\,x_0^*].\end{aligned}$$

Proof of Theorem 3.2. Letting $Y_t = \underline{X}_t - \widehat{\mu}$, we obtain $Y_t = (\mu - \widehat{\mu}) + \Psi(B)\epsilon_t$, and therefore if $\lambda \neq 0$

$$d_Y^s(\lambda) = -(\widehat{\mu} - \mu)\,T^{-1/2}\sum_{t=1}^{[Ts]} e^{-i\lambda t} + \Psi(e^{-i\lambda})\,d_\epsilon^s(\lambda) + T^{-1/2}\left(Z_{[Ts]}(\lambda) - Z_0(\lambda)\right).$$

In the first term, the DFT is $O(T^{-1/2})$ because the sum of complex exponentials is bounded for all T, whereas $\widehat{\mu} - \mu = O_P(T^{-1/2})$. The second term is bounded in probability because $d_\epsilon^s(\lambda) = O_P(1)$. The third term is $O_P(T^{-1/2})$, because $Z_{[Ts]}(\lambda) = O_P(1)$ for all T. Under H_a, $\Psi(e^{-i\lambda})$ is full rank and the second term is leading. Under H_0, $\Psi(e^{-i\lambda})$ is reduced rank, such that the second and third terms are leading. Either way, the first term, which is $O_P(T^{-1})$, is negligible. On the other hand, if $\lambda = 0$ we have the expansion

$$d_Y^s(0) = \Psi(1)\,d_{\overline{\epsilon}}^s(0) + T^{-1/2}\left(Z_{[Ts]}(0) - Z_0(0) - \frac{[Ts]}{T}\,[Z_T(0) - Z_0(0)]\right)$$

with $d_{\overline{\epsilon}}^s(0) = T^{-1/2}\sum_{t=1}^{[Ts]}(\epsilon_t - \overline{\epsilon})$. Evidently, the sample mean has been incorporated into the two terms. Therefore, no matter the value of λ we have

$$d_Y^s(\lambda) = O_P(T^{-1}) + \Psi(e^{-i\lambda})\,d_{\tilde{\epsilon}}^s(\lambda) + d_{\tilde{Z}}^s(\lambda), \tag{B.5}$$

where $d_{\tilde{\epsilon}}^s(\lambda)$ equals either $d_\epsilon^s(\lambda)$ or $d_{\overline{\epsilon}}^s(0)$ depending on whether $\lambda \neq 0$ or $\lambda = 0$. Similarly, we define $\tilde{Z}_t = \overline{Z}_t(\lambda) - \overline{Z}_{t-1}(\lambda)$, where $\overline{Z}_t(\lambda)$ equals either $Z_t(\lambda)$ or $Z_t(0) - tT^{-1}Z_T(0)$ depending on whether $\lambda \neq 0$ or $\lambda = 0$. Next, using $Y_t e^{-i\lambda t} = T^{1/2}\,(d_Y^{t/T}(\lambda) - d_Y^{(t-1)/T}(\lambda))$ and applying Lemma 0.2 to the expression for the spectral density estimator (3.6), we obtain

$$\begin{aligned}\widehat{f}(\lambda) &= T^{-1}\sum_{t,\ell=1}^{T}\Lambda_b\left(\frac{t-\ell}{T}\right)Y_t\,Y_\ell'\,e^{-i\lambda(t-\ell)} \qquad\qquad (B.6)\\ &= \sum_{t,\ell=1}^{T-1}\Delta_{\Lambda_b}^2\left(\frac{t-\ell}{T}\right)d_Y^{t/T}(\lambda)\,d_Y^{\ell/T}(\lambda)^* \\ &\quad + \sum_{t=1}^{T-1}\Delta_{\Lambda_b}^1\left(\frac{t-T}{T}\right)\left(d_Y^1(\lambda)\,d_Y^{\ell/T}(\lambda)^* + d_Y^{t/T}(\lambda)\,d_Y^1(\lambda)^*\right) + \Lambda_b(0)\,d_Y^1(\lambda)\,d_Y^1(\lambda)^*.\end{aligned}$$

Here $\Delta^2_{\Lambda_b}(x) = 2\Lambda_b(x) - \Lambda_b(x - 1/T) - \Lambda_b(x + 1/T)$ and $\Delta^1_{\Lambda_b}(x) = \Lambda_b(x) - \Lambda_b(x + 1/T)$. Observe that this expression for $\widehat{f}$ depends on $\{Y_t\}$ only through the DFT $d^s_Y(\lambda)$, and we can denote it via $\widehat{f}_{Y,Y}(\lambda)$ for short. Applying (B.5) to each term (and suppressing λ in the notation) yields

$$\widehat{f}_{Y,Y} = O_P(T^{-1}) + \Psi\, \widehat{f}_{\tilde{\epsilon},\tilde{\epsilon}} \Psi^* + \widehat{f}_{\tilde{Z},\tilde{\epsilon}}\, \Psi^* + \Psi\, \widehat{f}_{\tilde{\epsilon},\tilde{Z}} + \widehat{f}_{\tilde{Z},\tilde{Z}}, \tag{B.7}$$

where $\widehat{f}_{\tilde{\epsilon},\tilde{\epsilon}}$ and so forth are obtained by swapping in the DFTs of $\tilde{\epsilon}$ and/or $\tilde{Z}$ for Y. When H_a holds, only the second term of (B.7) is pertinent, the others being of lower order. We proceed to analyse this case first by applying (B.6), replacing d_Y by $d_{\tilde{\epsilon}}$.

Suppose that x is a kink point in Λ (which means that bx is a kink of Λ_b). If t and ℓ are such that $(t - \ell)/T = x$, then

$$\begin{aligned}
\Delta^2_{\Lambda_b}\left(\frac{t-\ell}{T}\right) &= \left[\Lambda_b(x) - \Lambda_b(x + T^{-1})\right] + \left[\Lambda_b(x) - \Lambda_b(x - T^{-1})\right] \\
&\approx -T^{-1}\, \dot{\Lambda}^+_b(x) + T^{-1}\, \dot{\Lambda}^-_b(x) \\
\Delta^1_{\Lambda_b}\left(\frac{t-\ell}{T}\right) &= \left[\Lambda_b(x) - \Lambda_b(x + T^{-1})\right] \approx -T^{-1}\, \dot{\Lambda}^+_b(x).
\end{aligned}$$

Now with $t = \ell + Tx$ for $\ell = 1, 2, \ldots, [T(1 - x)]$, we see that for each $1 \le t \le T$ there exists a series of ℓ values such that $(t - \ell)/T = x$, when Tx is an integer. Hence the double sum over $1 \le t, \ell \le T$ in the expansion for $\widehat{f}_{\tilde{\epsilon},\tilde{\epsilon}}$ can be broken down into ranges of values falling in the intervals of smoothness C_k, plus the expressions at the kinks. Because $\Delta^1_{\Lambda_b}(x) = -T^{-1}\, \dot{\Lambda}_b(x)$ and $\Delta^2_{\Lambda_b}(x) = -T^{-2}\, \ddot{\Lambda}_b(x)$ uniformly in x belonging to a region of smoothness C_k, we can apply Theorem 3.1 to find that the first term in the expansion of $\widehat{f}_{\tilde{\epsilon},\tilde{\epsilon}}$ converges weakly to the first two terms of Ξ stated in Theorem 3.2. The second term in the expansion of $\widehat{f}_{\tilde{\epsilon},\tilde{\epsilon}}$ involves only a single summation over $1 \le t \le T - 1$, and so there are only a finite number of indices t for which $(t - T)/T$ equals a given kink—hence the kink contribution is negligible. Again using Theorem 3.1, this second term converges weakly to the third term of Ξ stated in Theorem 3.2. The last term of Ξ follows at once from the third term of $\widehat{f}_{\tilde{\epsilon},\tilde{\epsilon}}$. This establishes the weak convergence of $\widehat{f}$ to $\Psi\, \Xi\, \Psi^*$. Furthermore, using (B.7) we find (when $j > 0$) that

$$\begin{aligned}
\widehat{g} &= O_P(T^{-1}) + \Psi_\sharp\, \widehat{f}_{\tilde{\epsilon},\tilde{\epsilon}}\, \Psi^*_\sharp \\
\widehat{k} &= O_P(T^{-1}) + \Psi_\flat\, \widehat{f}_{\tilde{\epsilon},\tilde{\epsilon}}\, \Psi^*_\sharp.
\end{aligned}$$

Therefore

$$\widehat{v} = \widehat{g}^{-1}\, \widehat{k} \xrightarrow{\mathcal{D}} \left[\Psi_\sharp\, \Xi\, \Psi^*_\sharp\right]^{-1} \Psi_\sharp\, \Xi\, \Psi^*_\flat,$$

and using $\widehat{d}_{j+1} = \widehat{w}^*\, \widehat{f}\, \widehat{w}$ the stated result now follows. When $j = 0$, $\widehat{d}_1 \xrightarrow{\mathcal{D}} \Psi_\sharp\, \Xi\, \Psi^*_\sharp$, where $\Psi_\sharp$ denotes the first row of Ψ. Now we consider the case that $d_{j+1} = 0$, where $j > 0$. The expressions for $\widehat{g}$ and $\widehat{k}$ are now

$$\begin{aligned}
\widehat{g} &= O_P(T^{-1}) + \Psi_\sharp\, \widehat{f}_{\tilde{\epsilon},\tilde{\epsilon}}\, \Psi^*_\sharp + \Psi_\sharp\, \widehat{f}_{\tilde{\epsilon},\tilde{Z}}\, [I\ 0]' + [I\ 0]\, \widehat{f}_{\tilde{Z},\tilde{\epsilon}}\, \Psi^*_\sharp + [I\ 0]\, \widehat{f}_{\tilde{Z},\tilde{Z}}\, [I\ 0]' \\
\widehat{k} &= O_P(T^{-1}) + \Psi_\flat\, \widehat{f}_{\tilde{\epsilon},\tilde{\epsilon}}\, \Psi^*_\sharp + \Psi_\flat\, \widehat{f}_{\tilde{\epsilon},\tilde{Z}}\, [I\ 0]' + [0\ I]\, \widehat{f}_{\tilde{Z},\tilde{\epsilon}}\, \Psi^*_\sharp + [0\ I]\, \widehat{f}_{\tilde{Z},\tilde{Z}}\, [I\ 0]'
\end{aligned}$$

and it follows that

$$\begin{aligned}\widehat{v} &= \widehat{g}^{-1}\,[I\,0]\,\widehat{f}_{Y,Y}\,[0\,I]' \\ &= O_P(T^{-1}) + \widehat{g}^{-1}\left(\left[\Psi_\sharp\,\widehat{f}_{\bar\epsilon,\bar\epsilon} + [I\,0]\,\widehat{f}_{\bar Z,\bar\epsilon}\right]\Psi_\flat^* + \left[\Psi_\sharp\,\widehat{f}_{\bar\epsilon,\bar Z} + [I\,0]\,\widehat{f}_{\bar Z,\bar Z}\right][0\,I]'\right) \\ &= \widehat{g}^{-1}\left(\left[\Psi_\sharp\,\widehat{f}_{\bar\epsilon,\bar\epsilon} + [I\,0]\,\widehat{f}_{\bar Z,\bar\epsilon}\right]\Psi_\sharp^*\,v + \left[\Psi_\sharp\,\widehat{f}_{\bar\epsilon,\bar Z} + [I\,0]\,\widehat{f}_{\bar Z,\bar Z}\right][0\,I]'\right) \\ &= \widehat{g}^{-1}\left(\left[\widehat{g} - \Psi_\sharp\,\widehat{f}_{\bar\epsilon,\bar Z}\,[I\,0]' - [I\,0]\,\widehat{f}_{\bar Z,\bar Z}\,[I\,0]'\right]v + \left[\Psi_\sharp\,\widehat{f}_{\bar\epsilon,\bar Z} + [I\,0]\,\widehat{f}_{\bar Z,\bar Z}\right][0\,I]'\right) \\ &= v - \widehat{g}^{-1}\left[\Psi_\sharp\,\widehat{f}_{\bar\epsilon,\bar Z} + [I\,0]\,\widehat{f}_{\bar Z,\bar Z}\right]w,\end{aligned} \tag{B.8}$$

because $w = [I\,0]'\,v - [0\,I]'$. Set $B = \Psi_\sharp\,\widehat{f}_{\bar\epsilon,\bar Z} + [I\,0]\,\widehat{f}_{\bar Z,\bar Z}$, so that

$$\widehat{v} - v = O_P(T^{-1}) - \widehat{g}^{-1}\,B\,w. \tag{B.9}$$

Hence $\widehat{w} = w + [I\,0]'\,[\widehat{v} - v] = O_P(T^{-1}) + w - [I\,0]'\,\widehat{g}^{-1}\,B\,w$, so that by (B.7)

$$T\,\widehat{d}_{j+1} = o_P(1) + T\,w^*\,\widehat{f}_{\bar Z,\bar Z}\,w - T\,w^*\,B^*\,\widehat{g}^{-1}\,B\,w,$$

using the null hypothesis and $[I\,0]\,\widehat{f}\,w = O_P(T^{-1}) + B\,w$. Because g is pd, we can apply the H_a case above to obtain $\widehat{g} \xrightarrow{\mathcal{D}} \Psi_\sharp\,\Xi\,\Psi_\sharp^*$. Therefore the limit of $\widehat{d}_{j+1}$ is determined by the terms $\widehat{f}_{\bar Z,\bar Z}$ and $\widehat{f}_{\bar\epsilon,\bar Z}$, which we describe next. Then we claim the following two convergences hold:

$$T\,\widehat{f}_{\bar Z,\bar Z} \xrightarrow{\mathcal{D}} Q \tag{B.10}$$

$$T^{1/2}\,\widehat{f}_{\bar Z,\bar\epsilon} \xrightarrow{\mathcal{D}} R, \tag{B.11}$$

where the limits Q and R are given in the statement of the theorem. To establish (B.10), first take the case that $\lambda \neq 0$. Suppressing λ in the notation for $Z_t(\lambda)$, we apply Lemma 0.2 to obtain

$$\begin{aligned}T\,\widehat{f}_{\bar Z,\bar Z} = &\sum_{t,\ell=1}^{T-1}\Delta_{\Lambda_b}^2\left(\frac{t-\ell}{T}\right)Z_t\,Z_\ell^* \\ &- \sum_{t=1}^{T-1}\Delta_{\Lambda_b}^1\left(\frac{t-1}{T}\right)\left(Z_t\,Z_0^* + Z_0\,Z_t^*\right) + \sum_{t=1}^{T-1}\Delta_{\Lambda_b}^1\left(\frac{t-T}{T}\right)\left(Z_T\,Z_t^* + Z_t\,Z_T^*\right) \\ &+ \Lambda_b(0)\left(Z_0\,Z_0^* + Z_T\,Z_T^*\right) - \Lambda_b(1-T^{-1})\left(Z_0\,Z_T^* + Z_T\,Z_0^*\right).\end{aligned}$$

The second and third terms are $O_P(T^{-1/2})$, because $\{Z_t\}$ is stationary. The sum of the fourth and fifth terms converge in probability to

$$\Lambda_b(0)\left(Z_0\,Z_0^* + Z_\infty\,Z_\infty^*\right) - \Lambda_b(1)\left(Z_0\,Z_\infty^* + Z_\infty\,Z_0^*\right),$$

using the fact that $Z_T \xrightarrow{P} Z_\infty$. This limit is independent of Z_0, using the representation of $Z_t(\lambda)$. The first term of $T\,\widehat{f}_{\bar Z,\bar Z}$ is $(T-1)\sum_{|h|<T-1}\Delta_{\Lambda_b}^2(h/(T-1))\,\widehat{\Gamma}_Z(h)$, where the sample autocovariance based on sample size $T-1$ is $\widehat{\Gamma}_Z(h) = (T-1)^{-1}\sum_{t=1}^{T-1-|h|}Z_t\,Z_{t+h}^*$. Now if x is a kink point of Λ and $h/(T-1) = x$, then the corresponding contribution to the double sum is approximately $-[\dot\Lambda_b^-(x) - \dot\Lambda_b^+(x)]\,\widehat{\Gamma}((T-1)x)$. This has mean and variance both tending to zero as $T \to \infty$, so long as $x \neq 0$. In an interval of smoothness C_k, the contribution is approximately

$$-T^{-1}\sum_{|h/T|\in C_k}\ddot\Lambda_b(h/T)\,\widehat{\Gamma}_Z(h),$$

which has mean of order T^{-1} and vanishing variance (because the acf of $\{Z_t\}$ is summable). The final case occurs when there is a kink at $x = 0$, in which case we have the asymptotic contribution

$$-\left[\dot{\Lambda}_b^-(0) - \dot{\Lambda}_b^+(0)\right]\Gamma(0) = -2\,\dot{\Lambda}_b^+(0)\,\mathrm{Cov}(Z).$$

Note that this quantity is zero if the taper is smooth at zero, but is nonzero in the case of a Bartlett taper ($\dot{\Lambda}_b^+(0) = -1/b$). Hence we obtain (B.10) when $\underline{\underline{\Lambda}}_b(r,s)$ is given by $\Lambda_b(r-s)$. Next, if $\lambda = 0$ we need to modify the above calculations with the inclusion of $T^{-1}(Z_T(0) - Z_0(0))$. Again writing Z_t for $Z_t(0)$, to the expression for $T\,\widehat{f}_{\tilde{Z},\tilde{Z}}$ are added the terms

$$\begin{aligned}
&-T^{-1}\textstyle\sum_{t,\ell=1}^{T}\Lambda_b\left(\frac{t-\ell}{T}\right)(Z_t - Z_{t-1})(Z_T - Z_0)^*\\
&-T^{-1}\textstyle\sum_{t,\ell=1}^{T}\Lambda_b\left(\frac{t-\ell}{T}\right)(Z_T - Z_0)(Z_\ell - Z_{\ell-1})^*\\
&+T^{-2}\textstyle\sum_{t,\ell=1}^{T}\Lambda_b\left(\frac{t-\ell}{T}\right)(Z_T - Z_0)(Z_T - Z_0)^*.
\end{aligned}$$

Whereas the third term converges to

$$\int_0^1\int_0^1 \Lambda_b(r-s)\,dr ds\,(Z_\infty - Z_0)(Z_\infty - Z_0)^*,$$

to the first term (the second term is the transpose) we apply Lemma 0.1, obtaining

$$\begin{aligned}
&-T^{-1}\left(\sum_{\ell=1}^{T}\sum_{t=1}^{T-1}\Delta^1_{\Lambda_b}\left(\frac{t-\ell}{T}\right)Z_t + \sum_{\ell=1}^{T}\left[\Lambda_b\left(\frac{T-\ell}{T}\right)Z_T - \Lambda_b\left(\frac{1-\ell}{T}\right)Z_0\right]\right)(Z_T - Z_0)^*\\
&= O_P(T^{-1/2}) + \left(\int_0^1 \Lambda_b(x)\,dx\,Z_0 - \int_0^1 \Lambda_b(1-x)\,dx\,Z_\infty\right)(Z_\infty - Z_0)^*.
\end{aligned}$$

Adding these limit results to the $\lambda \neq 0$ limit, and simplifying, yields (B.10). To prove (B.11), we first consider $\lambda \neq 0$. Applying Lemma 0.1,

$$T\,\widehat{f}_{\tilde{Z},\tilde{\epsilon}} = \sum_{\ell=1}^{T}\sum_{t=1}^{T-1}\Delta^1_{\Lambda_b}\left(\frac{t-\ell}{T}\right)Z_t\,\epsilon'_\ell e^{i\lambda\ell} + \sum_{\ell=1}^{T}\Lambda_b\left(\frac{T-\ell}{T}\right)Z_T\,\epsilon'_\ell e^{i\lambda\ell} - \sum_{\ell=1}^{T}\Lambda_b\left(\frac{1-\ell}{T}\right)Z_0\,\epsilon'_\ell e^{i\lambda\ell}. \tag{B.12}$$

Whereas (following previous arguments to handle kinks) the first term is $O_P(1)$, the second term multiplied by $T^{-1/2}$ is (by Lemma 0.1)

$$Z_T\left(-\sum_{\ell=1}^{T-1}\Delta^1_{\Lambda_b}\left(\frac{T-\ell-1}{T}\right)d_\epsilon^{\ell/T*} + \Lambda_b(0)\,d_\epsilon^{1*}\right) \xrightarrow{P} Z_\infty\left(\int_0^1 \dot{\Lambda}_b(1-r)\,B(r)^*\,dx + \Lambda_b(0)\,B(1)^*\right).$$

The third term multiplied by $T^{-1/2}$ is

$$Z_0\left(-\sum_{\ell=1}^{T-1}\Delta^1_{\Lambda_b}\left(\frac{\ell-1}{T}\right)d_\epsilon^{\ell/T*} - \Lambda_b\left(\frac{T-1}{T}\right)d_\epsilon^{1*}\right) \xrightarrow{P} Z_0\left(\int_0^1 \dot{\Lambda}_b(r)B(r)^*\,dx - \Lambda_b(1)\,B(1)^*\right).$$

Combining these two expressions, and simplifying, yields (B.11) when $\underline{\Lambda}_b = \Lambda_b$. When $\lambda = 0$, the expressions involving ϵ_t are modified by the sample mean, and the limits of $d_\epsilon^{t/T}$ are that of a Brownian bridge $\underline{B}$.

Also, we modify the difference of $Z_t(0)$ by subtracting $T^{-1}(Z_T(0) - Z_0(0))$, so that from (B.12) we subtract

$$T^{-1}(Z_T - Z_0)\sum_{t=1}^{T}\sum_{\ell=1}^{T}\Lambda_b\left(\frac{t-\ell}{T}\right)\epsilon'_\ell e^{i\lambda\ell}$$

$$= T^{-1/2}(Z_T - Z_0)\left(\sum_{t=1}^{T}\sum_{\ell=1}^{T-1}\Delta^1_{\Lambda_b}\left(\frac{t-\ell}{T}\right)d^{\ell/T*}_\epsilon + \sum_{t=1}^{T}\Lambda_b\left(\frac{T-t}{T}\right)d^{1*}_\epsilon\right),$$

via Lemma 0.1. Here the DFT of ϵ is centred by its sample mean, and hence will converge to a Brownian bridge process. The first term in parentheses is both $O_P T$ (the second term is zero, because $d^1_\epsilon = 0$), and hence multiplying the whole expression by $T^{-1/2}$ yields a convergence in probability to

$$(Z_\infty - Z_0)\left(\int_0^1\int_0^1 \dot{\Lambda}_b(r-s)\underline{B}^*(r)\,dr\,ds\right).$$

In the case that Λ is only piecewise-smooth, the above integrals of $\dot{\Lambda}$ are interpreted as the sum over the smooth regions C_k where the derivatives are defined. (The kinks, however, will not contribute to the integral.) Combining and simplifying the limiting random variables yields (B.11). We conclude that $T^{1/2}B \xrightarrow{\mathcal{D}} \Psi_\sharp R^*$, which proves the theorem. Finally, letting $c^* = w^*[Z_\infty, Z_0]$ we find that the limit can be written

$$-2\,\dot{\Lambda}^+_b(0)\,w^*\mathrm{Cov}(Z)\,w + c^*\left(\begin{bmatrix}\underline{\underline{\Lambda}}_b(1,1) & -\underline{\underline{\Lambda}}_b(1,0)\\ -\underline{\underline{\Lambda}}_b(0,1) & \underline{\underline{\Lambda}}_b(0,0)\end{bmatrix} - A\,\Psi^*_\sharp\left[\Psi_\sharp\,\Xi\,\Psi^*_\sharp\right]^{-1}\Psi_\sharp\,A^*\right)c$$

with

$$A = \begin{bmatrix}\int_0^1 \dot{\underline{\Lambda}}_b(1,s)\,\underline{B}(s)^*\,ds + \underline{\Lambda}_b(1,1)\,\underline{B}(1)^*\\ -\int_0^1 \dot{\underline{\Lambda}}_b(0,s)\,\underline{B}(s)^*\,ds - \underline{\Lambda}_b(0,1)\,\underline{B}(1)^*\end{bmatrix}.$$

Because c has a continuous distribution (as Z_∞ and Z_0 are independent and continuous) and Ξ has a continuous distribution, the limit distribution does not have point masses. Finally, when $j = 0$ we have $\Psi = 0$ under H_0, and hence (B.7) reduces to

$$\widehat{f}_{Y,Y} = O_P(T^{-1}) + \widehat{f}_{\tilde{Z},\tilde{Z}}.$$

Therefore by (B.10), $T\,\widehat{f}_{YY} \xrightarrow{\mathcal{D}} Q$. □

REFERENCES

Axelsson, O. (1996). *Iterative Solution Methods*. New York, NY: Cambridge University Press.

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory*. San Francisco, CA: Siam.

Busetti, F. (2006). Tests of seasonal integration and cointegration in multivariate unobserved component models. *Journal of Applied Econometrics 21*, 419–38.

Gómez, V. (2016). *Multivariate Time Series with Linear State Space Structure*. New York, NY: Springer.

Karatzas, I. and S. Shreve (1991). *Brownian Motion and Stochastic Calculus*. New York, NY: Springer.

McElroy, T. and D. Politis (2014). Spectral density and spectral distribution inference for long memory time series via fixed-b asymptotics. *Journal of Econometrics 182*, 211–25.

Nyblom, J. and A. Harvey (2000). Tests of common stochastic trends. *Econometric Theory 16*, 176–99.

Politis, D., J. Romano, and M. Wolf (1999). *Subsampling*. New York, NY: Springer.

APPENDIX C: SUPPLEMENTARY TABLES

Table C1. Trend and irregular size (Bartlett).

					Nominal level		
					0.01	0.05	0.10
T	q^{μ}	ρ^{μ}	b	Average test stat.		Empirical size	
1000	1	1	0.1	43.154	0.024	0.087	0.210
1000	1	1	0.3	16.629	0.020	0.075	0.135
1000	1	1	0.5	10.861	0.018	0.071	0.137
1000	3	1	0.1	43.636	0.005	0.026	0.066
1000	3	1	0.3	16.692	0.008	0.057	0.102
1000	3	1	0.5	10.901	0.012	0.061	0.104
1000	5	1	0.1	43.446	0.003	0.018	0.046
1000	5	1	0.3	16.570	0.006	0.046	0.093
1000	5	1	0.5	10.771	0.010	0.046	0.099
3000	1	1	0.1	43.485	0.012	0.062	0.117
3000	1	1	0.3	16.643	0.013	0.061	0.110
3000	1	1	0.5	10.847	0.012	0.061	0.108
5000	1	1	0.1	43.841	0.018	0.069	0.135
5000	1	1	0.3	16.873	0.017	0.073	0.127
5000	1	1	0.5	11.026	0.015	0.067	0.126

Note: trend and irregular case, empirical size of the subsampling test ($R = 1000$ replicates, Bartlett taper, subsample size n selected adaptively) and the average test statistic.

Table C2. Trend and irregular size (Parzen).

					Nominal level		
					0.01	0.05	0.10
T	q^{μ}	ρ^{μ}	b	Average test stat.		Empirical size	
1000	1	1	0.1	6.229	0	0	0
1000	1	1	0.3	3.778	0.005	0.029	0.057
1000	1	1	0.5	3.175	0.008	0.036	0.081
1000	3	1	0.1	6.120	0	0	0
1000	3	1	0.3	3.752	0.006	0.026	0.060
1000	3	1	0.5	3.175	0.008	0.048	0.084
1000	5	1	0.1	6.157	0	0	0
1000	5	1	0.3	3.755	0.006	0.031	0.057
1000	5	1	0.5	3.157	0.009	0.038	0.074
3000	1	1	0.1	4.725	0.003	0.010	0.018
3000	1	1	0.3	3.649	0.005	0.037	0.086
3000	1	1	0.5	3.122	0.005	0.036	0.094
5000	1	1	0.1	4.136	0.005	0.017	0.029
5000	1	1	0.3	3.401	0.009	0.034	0.064
5000	1	1	0.5	2.944	0.009	0.034	0.072

Note: trend and irregular case, empirical size of the subsampling test ($R = 1000$ replicates, Parzen taper, subsample size n selected adaptively) and the average test statistic.

Table C3. Trend and irregular size (Nyblom–Harvey).

				Nominal level		
T	q^{μ}	ρ^{μ}	Average test stat.	0.01	0.05 Empirical size	0.10
1000	1	1	0.086	0.012	0.057	0.111
1000	3	1	0.086	0.014	0.045	0.096
1000	5	1	0.083	0.007	0.047	0.090
3000	1	1	0.086	0.009	0.052	0.112
3000	3	1	0.087	0.013	0.050	0.097
3000	5	1	0.092	0.013	0.060	0.105
5000	1	1	0.086	0.013	0.046	0.101
5000	3	1	0.087	0.010	0.053	0.107
5000	5	1	0.089	0.014	0.060	0.113

Note: trend and irregular case, empirical size of Nyblom and Harvey (2000) test ($R = 1000$ replicates) and the average test statistic.

Table C4. Trend and irregular power (Bartlett).

					Nominal level		
T	q^{μ}	ρ^{μ}	b	Average test stat	0.01	0.05 Empirical power	0.10
1000	1	0.95	0.1	122.221	0.998	1	1
1000	1	0.95	0.3	74.201	0.892	0.974	0.990
1000	1	0.95	0.5	52.794	0.722	0.898	0.949
1000	1	0.97	0.1	92.404	0.986	0.997	0.999
1000	1	0.97	0.3	52.582	0.785	0.919	0.960
1000	1	0.97	0.5	36.632	0.593	0.820	0.901
1000	1	0.99	0.1	59.907	0.567	0.847	0.938
1000	1	0.99	0.3	28.890	0.361	0.633	0.772
1000	1	0.99	0.5	19.724	0.250	0.461	0.633
1000	3	0.95	0.1	288.933	1	1	1
1000	3	0.95	0.3	190.011	0.963	0.985	0.994
1000	3	0.95	0.5	138.365	0.886	0.954	0.985
1000	3	0.97	0.1	188.673	0.998	1	1
1000	3	0.97	0.3	120.725	0.954	0.986	0.993
1000	3	0.97	0.5	85.931	0.850	0.949	0.981
1000	3	0.99	0.1	94.335	0.928	0.972	0.994
1000	3	0.99	0.3	52.543	0.766	0.909	0.946
1000	3	0.99	0.5	36.687	0.577	0.789	0.890
1000	5	0.95	0.1	453.318	1	1	1
1000	5	0.95	0.3	307.936	0.965	0.984	0.997
1000	5	0.95	0.5	218.452	0.911	0.964	0.987
1000	5	0.97	0.1	291.420	1	1	1
1000	5	0.97	0.3	193.402	0.963	0.989	0.994

Table C4. Continued

					Nominal level		
					0.01	0.05	0.10
T	q^{μ}	ρ^{μ}	b	Average test stat	Empirical power		
1000	5	0.97	0.5	138.581	0.898	0.967	0.985
1000	5	0.99	0.1	128.346	0.990	1	1
1000	5	0.99	0.3	77.095	0.879	0.954	0.980
1000	5	0.99	0.5	54.726	0.727	0.879	0.944
3000	1	0.95	0.1	288.835	1	1	1
3000	1	0.95	0.3	193.449	0.958	0.981	0.991
3000	1	0.95	0.5	139.641	0.885	0.954	0.973
3000	1	0.97	0.1	193.665	0.999	0.999	1
3000	1	0.97	0.3	126.137	0.962	0.985	0.993
3000	1	0.97	0.5	89.447	0.866	0.949	0.971
3000	1	0.99	0.1	94.450	0.986	0.997	0.999
3000	1	0.99	0.3	53.341	0.784	0.938	0.974
3000	1	0.99	0.5	37.593	0.581	0.821	0.907
5000	1	0.95	0.1	445.503	0.999	1	1
5000	1	0.95	0.3	298.338	0.963	0.980	0.991
5000	1	0.95	0.5	217.649	0.907	0.962	0.982
5000	1	0.97	0.1	294.690	1	1	1
5000	1	0.97	0.3	194.816	0.963	0.984	0.993
5000	1	0.97	0.5	139.974	0.906	0.962	0.980
5000	1	0.99	0.1	128.608	0.998	1	1
5000	1	0.99	0.3	77.434	0.915	0.970	0.991
5000	1	0.99	0.5	54.160	0.755	0.909	0.955

Note: trend and irregular case, empirical power of the subsampling test ($R = 1000$ replicates, Bartlett taper, subsample size n selected adaptively) and the average test statistic.

Table C5. Trend and irregular power (Parzen).

					Nominal level		
					0.01	0.05	0.10
T	q^{μ}	ρ^{μ}	b	Average test stat.	Empirical power		
1000	1	0.95	0.1	92.619	0.882	0.936	0.958
1000	1	0.95	0.3	69.306	0.876	0.943	0.966
1000	1	0.95	0.5	52.266	0.720	0.830	0.881
1000	1	0.97	0.1	57.447	0.508	0.632	0.712
1000	1	0.97	0.3	42.722	0.703	0.854	0.907

Table C5. Continued

					Nominal level		
					0.01	0.05	0.10
T	q^{μ}	ρ^{μ}	b	Average test stat.	Empirical power		
1000	1	0.97	0.5	32.062	0.586	0.734	0.818
1000	1	0.99	0.1	23.842	0.012	0.031	0.044
1000	1	0.99	0.3	16.837	0.266	0.497	0.618
1000	1	0.99	0.5	12.861	0.239	0.431	0.540
1000	3	0.95	0.1	260.907	0.998	0.999	1
1000	3	0.95	0.3	191.049	0.969	0.988	0.993
1000	3	0.95	0.5	137.103	0.871	0.922	0.943
1000	3	0.97	0.1	161.799	0.976	0.986	0.995
1000	3	0.97	0.3	119.704	0.929	0.969	0.986
1000	3	0.97	0.5	89.079	0.830	0.910	0.938
1000	3	0.99	0.1	57.284	0.511	0.640	0.702
1000	3	0.99	0.3	40.973	0.691	0.850	0.906
1000	3	0.99	0.5	30.713	0.566	0.731	0.805
1000	5	0.95	0.1	434.482	1	1	1
1000	5	0.95	0.3	320.242	0.980	0.992	0.996
1000	5	0.95	0.5	237.973	0.879	0.935	0.963
1000	5	0.97	0.1	261.015	0.994	0.998	0.999
1000	5	0.97	0.3	199.236	0.975	0.987	0.994
1000	5	0.97	0.5	148.289	0.863	0.922	0.954
1000	5	0.99	0.1	93.948	0.866	0.918	0.943
1000	5	0.99	0.3	69.527	0.849	0.918	0.946
1000	5	0.99	0.5	51.133	0.708	0.825	0.871
3000	1	0.95	0.1	261.394	1	1	1
3000	1	0.95	0.3	194.977	0.969	0.985	0.989
3000	1	0.95	0.5	145.746	0.843	0.908	0.935
3000	1	0.97	0.1	159.875	1	1	1
3000	1	0.97	0.3	124.32	0.968	0.986	0.991
3000	1	0.97	0.5	94.059	0.850	0.920	0.940
3000	1	0.99	0.1	57.109	0.919	0.980	0.992
3000	1	0.99	0.3	43.714	0.799	0.910	0.943
3000	1	0.99	0.5	32.984	0.632	0.783	0.846
5000	1	0.95	0.1	440.744	1	1	1
5000	1	0.95	0.3	338.476	0.975	0.983	0.989
5000	1	0.95	0.5	254.120	0.889	0.943	0.958
5000	1	0.97	0.1	261.587	1	1	1
5000	1	0.97	0.3	193.668	0.975	0.986	0.991
5000	1	0.97	0.5	142.806	0.877	0.931	0.949
5000	1	0.99	0.1	91.108	0.999	1	1
5000	1	0.99	0.3	68.700	0.929	0.974	0.983
5000	1	0.99	0.5	51.315	0.737	0.874	0.914

Note: trend and irregular case, empirical power of the subsampling test ($R = 1000$ replicates, Parzen taper, subsample size n selected adaptively) and the average test statistic.

Table C6. Trend and irregular power (Nyblom–Harvey).

				Nominal level		
T	q^{μ}	ρ^{μ}	Average test stat.	0.01	0.05 Empirical power	0.10
1000	1	0.95	8.646	1	1	1
1000	1	0.97	8.442	1	1	1
1000	1	0.99	6.404	0.998	1	1
1000	3	0.95	9.746	1	1	1
1000	3	0.97	9.589	1	1	1
1000	3	0.99	8.355	1	1	1
1000	5	0.95	9.963	1	1	1
1000	5	0.97	9.865	1	1	1
1000	5	0.99	9.098	1	1	1
3000	1	0.95	29.920	1	1	1
3000	1	0.97	27.764	1	1	1
3000	1	0.99	25.679	1	1	1
3000	3	0.95	30.212	1	1	1
3000	3	0.97	30.820	1	1	1
3000	3	0.99	28.174	1	1	1
3000	5	0.95	31.505	1	1	1
3000	5	0.97	30.933	1	1	1
3000	5	0.99	29.338	1	1	1
5000	1	0.95	50.829	1	1	1
5000	1	0.97	47.935	1	1	1
5000	1	0.99	45.281	1	1	1
5000	3	0.95	51.815	1	1	1
5000	3	0.97	49.786	1	1	1
5000	3	0.99	49.473	1	1	1
5000	5	0.95	50.558	1	1	1
5000	5	0.97	51.790	1	1	1
5000	5	0.99	50.584	1	1	1

Note: trend and irregular case, empirical power of Nyblom and Harvey (2000) test ($R = 1000$ replicates) and the average test statistic.

Table C7. Trend, seasonal, and irregular size (Bartlett).

					Nominal level		
T	$q^{(1)}$	$\rho^{(1)}$	b	Average test stat.	0.01	0.05 Empirical size	0.10
1000	1	1	0.1	507.145	0.272	0.585	0.786
1000	1	1	0.3	199.855	0.035	0.122	0.226
1000	1	1	0.5	134.539	0.026	0.083	0.166
1000	3	1	0.1	517.047	0.027	0.145	0.320
1000	3	1	0.3	200.572	0.013	0.066	0.142
1000	3	1	0.5	134.428	0.013	0.060	0.119
1000	5	1	0.1	523.195	0.014	0.077	0.183
1000	5	1	0.3	203.850	0.015	0.064	0.135
1000	5	1	0.5	137.107	0.013	0.066	0.130
3000	1	1	0.1	517.816	0.046	0.178	0.344
3000	1	1	0.3	200.544	0.012	0.057	0.131
3000	1	1	0.5	134.063	0.006	0.046	0.114
5000	1	1	0.1	520.216	0.022	0.094	0.207
5000	1	1	0.3	200.723	0.011	0.054	0.108
5000	1	1	0.5	134.347	0.012	0.047	0.099

Note: trend, seasonal, and irregular case, empirical size of the subsampling test ($R = 1000$ replicates, Bartlett taper, subsample size n selected adaptively) and the average test statistic.

Table C8. Trend, seasonal, and irregular size (Parzen).

					Nominal level		
T	$q^{(1)}$	$\rho^{(1)}$	b	Average test stat.	0.01	0.05 Empirical size	0.10
1000	1	1	0.1	116.435	0	0	0
1000	1	1	0.3	50.701	0	0	0.002
1000	1	1	0.5	42.988	0.003	0.02	0.036
1000	3	1	0.1	119.039	0	0	0
1000	3	1	0.3	51.777	0	0.001	0.002
1000	3	1	0.5	44.036	0.004	0.014	0.035
1000	5	1	0.1	116.375	0	0	0
1000	5	1	0.3	49.594	0	0	0.002
1000	5	1	0.5	41.747	0	0.01	0.027
3000	1	1	0.1	70.465	0	0	0
3000	1	1	0.3	46.126	0.007	0.025	0.052
3000	1	1	0.5	41.666	0.010	0.037	0.080
5000	1	1	0.1	62.783	0	0	0.002
5000	1	1	0.3	46.909	0.006	0.037	0.078
5000	1	1	0.5	43.193	0.006	0.054	0.103

Note: trend, seasonal, and irregular case, empirical size of the subsampling test ($R = 1000$ replicates, Parzen taper, subsample size n selected adaptively) and the average test statistic.

Table C9. Trend, seasonal, and irregular size (Busetti).

				Nominal level		
T	$q^{(1)}$	$\rho^{(1)}$	Average test stat.	0.01	0.05 Empirical size	0.10
1000	1	1	0.256	0.012	0.056	0.107
1000	3	1	0.255	0.008	0.054	0.110
1000	5	1	0.259	0.013	0.053	0.117
3000	1	1	0.256	0.017	0.058	0.120
3000	3	1	0.257	0.016	0.058	0.113
3000	5	1	0.259	0.010	0.052	0.109
5000	1	1	0.262	0.016	0.055	0.125
5000	3	1	0.258	0.009	0.064	0.112
5000	5	1	0.262	0.018	0.071	0.123

Note: trend, seasonal, and irregular case, empirical size of Busetti (2006) test ($R = 1000$ replicates) and the average test statistic.

Table C10. Trend, seasonal, and irregular power (Bartlett).

					Nominal level		
T	$q^{(1)}$	$\rho^{(1)}$	b	Average test stat.	0.01	0.05 Empirical power	0.10
1000	1	0.95	0.1	844.529	0.997	0.999	0.999
1000	1	0.95	0.3	507.585	0.945	0.990	0.996
1000	1	0.95	0.5	409.646	0.873	0.956	0.975
1000	1	0.97	0.1	712.689	0.977	0.996	1
1000	1	0.97	0.3	386.815	0.821	0.945	0.978
1000	1	0.97	0.5	301.339	0.698	0.852	0.914
1000	1	0.99	0.1	578.401	0.713	0.915	0.976
1000	1	0.99	0.3	265.030	0.329	0.621	0.781
1000	1	0.99	0.5	191.936	0.245	0.499	0.639
1000	3	0.95	0.1	1595.570	1	1	1
1000	3	0.95	0.3	1177.967	0.997	1	1
1000	3	0.95	0.5	999.805	0.984	0.994	0.996
1000	3	0.97	0.1	1155.152	0.998	1	1
1000	3	0.97	0.3	778.467	0.989	0.998	1
1000	3	0.97	0.5	643.060	0.950	0.987	0.996
1000	3	0.99	0.1	729.369	0.821	0.952	0.989
1000	3	0.99	0.3	391.694	0.740	0.901	0.946
1000	3	0.99	0.5	303.333	0.633	0.819	0.897

Table C10. Continued

					Nominal level		
T	$q^{(1)}$	$\rho^{(1)}$	b	Average test stat.	0.01	0.05 Empirical power	0.10
1000	5	0.95	0.1	2297.171	1	1	1
1000	5	0.95	0.3	1776.290	0.999	1	1
1000	5	0.95	0.5	1517.357	0.989	0.997	0.998
1000	5	0.97	0.1	1601.939	1	1	1
1000	5	0.97	0.3	1154.964	0.997	1	1
1000	5	0.97	0.5	981.672	0.982	0.995	0.999
1000	5	0.99	0.1	886.094	0.955	0.996	1
1000	5	0.99	0.3	530.615	0.919	0.978	0.995
1000	5	0.99	0.5	430.819	0.843	0.949	0.969
3000	1	0.95	0.1	1570.029	1	1	1
3000	1	0.95	0.3	1133.445	0.996	1	1
3000	1	0.95	0.5	957.164	0.971	0.987	0.991
3000	1	0.97	0.1	1166.588	1	1	1
3000	1	0.97	0.3	787.718	0.991	0.999	0.999
3000	1	0.97	0.5	655.850	0.959	0.987	0.991
3000	1	0.99	0.1	739.682	0.974	0.992	0.998
3000	1	0.99	0.3	401.767	0.800	0.936	0.973
3000	1	0.99	0.5	315.983	0.645	0.841	0.923
5000	1	0.95	0.1	2309.570	1	1	1
5000	1	0.95	0.3	1792.660	1	1	1
5000	1	0.95	0.5	1549.784	0.984	0.997	1
5000	1	0.97	0.1	1611.832	1	1	1
5000	1	0.97	0.3	1189.512	0.999	1	1
5000	1	0.97	0.5	1016.570	0.986	0.997	0.999
5000	1	0.99	0.1	892.700	0.995	1	1
5000	1	0.99	0.3	530.795	0.942	0.985	0.994
5000	1	0.99	0.5	426.782	0.846	0.953	0.977

Note: trend, seasonal, and irregular case, empirical power of the subsampling test ($R = 1000$ replicates, Bartlett taper, subsample size n selected adaptively) and the average test statistic.

Table C11. Trend, seasonal, and irregular power (Parzen).

					Nominal level		
T	$q^{(1)}$	$\rho^{(1)}$	b	Average test stat.	0.01	0.05 Empirical power	0.10
1000	1	0.95	0.1	486.997	0.078	0.162	0.217
1000	1	0.95	0.3	380.945	0.736	0.865	0.910
1000	1	0.95	0.5	338.612	0.779	0.879	0.919
1000	1	0.97	0.1	333.833	0.002	0.004	0.004
1000	1	0.97	0.3	248.039	0.396	0.575	0.675
1000	1	0.97	0.5	219.768	0.523	0.713	0.792
1000	1	0.99	0.1	193.195	0	0	0

Table C11. Continued

					Nominal level		
T	$q^{(1)}$	$\rho^{(1)}$	b	Average test stat.	0.01	0.05 Empirical power	0.10
1000	1	0.99	0.3	120.420	0.026	0.073	0.114
1000	1	0.99	0.5	106.101	0.109	0.253	0.362
1000	3	0.95	0.1	1221.469	0.916	0.958	0.970
1000	3	0.95	0.3	1044.155	0.979	0.992	0.995
1000	3	0.95	0.5	936.247	0.962	0.983	0.991
1000	3	0.97	0.1	784.278	0.414	0.567	0.664
1000	3	0.97	0.3	646.214	0.922	0.962	0.976
1000	3	0.97	0.5	571.558	0.907	0.950	0.976
1000	3	0.99	0.1	341.196	0.001	0.001	0.001
1000	3	0.99	0.3	257.347	0.393	0.564	0.648
1000	3	0.99	0.5	229.678	0.521	0.704	0.776
1000	5	0.95	0.1	1949.538	0.998	1	1
1000	5	0.95	0.3	1685.357	0.998	1	1
1000	5	0.95	0.5	1504.030	0.986	0.997	0.998
1000	5	0.97	0.1	1229.433	0.895	0.946	0.969
1000	5	0.97	0.3	1033.793	0.979	0.995	0.997
1000	5	0.97	0.5	920.610	0.960	0.984	0.994
1000	5	0.99	0.1	497.009	0.004	0.015	0.040
1000	5	0.99	0.3	394.153	0.704	0.826	0.879
1000	5	0.99	0.5	353.321	0.753	0.850	0.900
3000	1	0.95	0.1	1181.823	0.997	1	1
3000	1	0.95	0.3	1043.287	0.996	1	1
3000	1	0.95	0.5	936.829	0.984	0.997	0.998
3000	1	0.97	0.1	748.234	0.904	0.949	0.962
3000	1	0.97	0.3	645.084	0.988	0.996	0.998
3000	1	0.97	0.5	569.218	0.960	0.988	0.993
3000	1	0.99	0.1	302.595	0.027	0.061	0.081
3000	1	0.99	0.3	252.463	0.649	0.840	0.920
3000	1	0.99	0.5	227.233	0.625	0.795	0.878
5000	1	0.95	0.1	1878.759	1	1	1
5000	1	0.95	0.3	1685.618	1	1	1
5000	1	0.95	0.5	1506.755	0.99	0.995	0.997
5000	1	0.97	0.1	1184.723	1	1	1
5000	1	0.97	0.3	1062.522	0.998	1	1
5000	1	0.97	0.5	965.024	0.988	0.997	0.999
5000	1	0.99	0.1	441.559	0.853	0.951	0.967
5000	1	0.99	0.3	387.050	0.907	0.969	0.984
5000	1	0.99	0.5	349.111	0.833	0.936	0.969

Note: trend, seasonal, and irregular case, empirical power of the subsampling test ($R = 1000$ replicates, Parzen taper, subsample size n selected adaptively) and the average test statistic.

Table C12. Trend, seasonal, and irregular power (Busetti).

				Nominal level		
T	$q^{(1)}$	$\rho^{(1)}$	Average test stat.	0.01	0.05 Empirical power	0.10
1000	1	0.95	6.358	0.998	1	1
1000	1	0.97	6.117	0.999	1	1
1000	1	0.99	4.967	0.981	0.997	0.998
1000	3	0.95	6.704	1	1	1
1000	3	0.97	6.848	1	1	1
1000	3	0.99	5.933	0.998	0.999	1
1000	5	0.95	6.827	1	1	1
1000	5	0.97	6.895	1	1	1
1000	5	0.99	6.585	0.999	0.999	1
3000	1	0.95	15.981	1	1	1
3000	1	0.97	15.494	1	1	1
3000	1	0.99	14.449	1	1	1
3000	3	0.95	16.304	1	1	1
3000	3	0.97	15.686	1	1	1
3000	3	0.99	15.937	1	1	1
3000	5	0.95	16.597	1	1	1
3000	5	0.97	15.812	1	1	1
3000	5	0.99	16.663	1	1	1
5000	1	0.95	24.238	1	1	1
5000	1	0.97	24.046	1	1	1
5000	1	0.99	22.996	1	1	1
5000	3	0.95	24.997	1	1	1
5000	3	0.97	24.242	1	1	1
5000	3	0.99	24.301	1	1	1
5000	5	0.95	24.927	1	1	1
5000	5	0.97	25.080	1	1	1
5000	5	0.99	24.252	1	1	1

Note: trend, seasonal, and irregular case, empirical power of Busetti (2006) test ($R = 1000$ replicates) and the average test statistic.

Table C13. Trend and irregular size/power (Bartlett).

T	q^{μ}	ρ^{μ}	b	j+1	Average test stat.	Empirical size/power
1000	1	0.95	0.3	2	598.406	0.993
1000	1	0.95	0.3	3	64.256	0.945
1000	1	0.95	0.3	4	428.462	0.989
1000	1	0.97	0.3	2	603.895	0.988
1000	1	0.97	0.3	3	45.128	0.898
1000	1	0.97	0.3	4	431.411	0.987
1000	1	0.99	0.3	2	616.232	0.993
1000	1	0.99	0.3	3	26.832	0.581
1000	1	0.99	0.3	4	459.297	0.987
1000	1	1	0.3	2	595.783	0.991
1000	1	1	0.3	3	17.435	0.077
1000	1	1	0.3	4	512.357	0.987
1000	3	0.95	0.3	2	1815.703	0.990
1000	3	0.95	0.3	3	162.368	0.983
1000	3	0.95	0.3	4	1251.024	0.991
1000	3	0.97	0.3	2	1806.621	0.988
1000	3	0.97	0.3	3	107.909	0.976
1000	3	0.97	0.3	4	1224.213	0.990
1000	3	0.99	0.3	2	1832.353	0.992
1000	3	0.99	0.3	3	50.356	0.889
1000	3	0.99	0.3	4	1342.165	0.990
1000	3	1	0.3	2	1737.463	0.985
1000	3	1	0.3	3	23.601	0.059
1000	3	1	0.3	4	1473.580	0.989
1000	5	0.95	0.3	2	2989.559	0.988
1000	5	0.95	0.3	3	273.518	0.984
1000	5	0.95	0.3	4	2017.171	0.993
1000	5	0.97	0.3	2	2996.955	0.988
1000	5	0.97	0.3	3	169.112	0.989
1000	5	0.97	0.3	4	2007.379	0.991
1000	5	0.99	0.3	2	2922.443	0.983
1000	5	0.99	0.3	3	77.428	0.938
1000	5	0.99	0.3	4	2177.390	0.989
1000	5	1	0.3	2	2994.401	0.991
1000	5	1	0.3	3	25.495	0.047
1000	5	1	0.3	4	2454.164	0.991

Note: trend and irregular case, proportion of rejections of $H_0: d_{j+1} = 0$ versus $H_1: d_{j+1} > 0$ in the subsampling test at a 5% nominal level ($R = 1000$ replicates of 4-variate time series, Bartlett taper, subsample size n selected adaptively) and the average test statistic.

Table C14. Trend and irregular size/power (Parzen).

T	q^{μ}	ρ^{μ}	b	j+1	Average test stat.	Empirical size/power
1000	1	0.95	0.3	2	662.834	0.993
1000	1	0.95	0.3	3	57.509	0.909
1000	1	0.95	0.3	4	433.41	0.972
1000	1	0.97	0.3	2	674.793	0.996
1000	1	0.97	0.3	3	37.098	0.803
1000	1	0.97	0.3	4	473.165	0.962
1000	1	0.99	0.3	2	636.274	0.984
1000	1	0.99	0.3	3	16.501	0.440
1000	1	0.99	0.3	4	488.768	0.974
1000	1	1	0.3	2	656.764	0.985
1000	1	1	0.3	3	5.866	0.030
1000	1	1	0.3	4	535.262	0.981
1000	3	0.95	0.3	2	1957.244	0.984
1000	3	0.95	0.3	3	166.894	0.980
1000	3	0.95	0.3	4	1333.822	0.971
1000	3	0.97	0.3	2	2021.529	0.984
1000	3	0.97	0.3	3	105.361	0.933
1000	3	0.97	0.3	4	1286.798	0.971
1000	3	0.99	0.3	2	2043.859	0.985
1000	3	0.99	0.3	3	41.738	0.793
1000	3	0.99	0.3	4	1353.229	0.951
1000	3	1	0.3	2	1993.391	0.985
1000	3	1	0.3	3	10.234	0.031
1000	3	1	0.3	4	1612.487	0.972
1000	5	0.95	0.3	2	3273.152	0.986
1000	5	0.95	0.3	3	285.903	0.991
1000	5	0.95	0.3	4	2186.997	0.968
1000	5	0.97	0.3	2	3253.737	0.982
1000	5	0.97	0.3	3	178.159	0.976
1000	5	0.97	0.3	4	2180.046	0.966
1000	5	0.99	0.3	2	3233.297	0.984
1000	5	0.99	0.3	3	68.911	0.895
1000	5	0.99	0.3	4	2258.041	0.966
1000	5	1	0.3	2	3228.706	0.981
1000	5	1	0.3	3	14.976	0.026
1000	5	1	0.3	4	2844.034	0.979

Note: trend and irregular case, proportion of rejections of $H_0: d_{j+1} = 0$ versus $H_1: d_{j+1} > 0$ in the subsampling test at a 5% nominal level ($R = 1000$ replicates of 4-variate time series, Parzen taper, subsample size n selected adaptively) and the average test statistic.

Table C15. Trend, seasonal, and irregular size/power (Bartlett).

T	$q^{(1)}$	$\rho^{(1)}$	b	$j+1$	Average test stat.	Empirical size/power
1000	1	0.95	0.3	2	3431.320	1
1000	1	0.95	0.3	3	447.375	0.981
1000	1	0.95	0.3	4	2450.678	1
1000	1	0.97	0.3	2	3354.192	1
1000	1	0.97	0.3	3	349.027	0.939
1000	1	0.97	0.3	4	2514.043	1
1000	1	0.99	0.3	2	3383.906	1
1000	1	0.99	0.3	3	245.551	0.616
1000	1	0.99	0.3	4	2640.491	0.999
1000	1	1	0.3	2	3276.443	1
1000	1	1	0.3	3	192.328	0.169
1000	1	1	0.3	4	2763.396	1
1000	3	0.95	0.3	2	9977.77	1
1000	3	0.95	0.3	3	955.626	0.999
1000	3	0.95	0.3	4	6794.931	0.999
1000	3	0.97	0.3	2	9995.539	1
1000	3	0.97	0.3	3	666.103	0.996
1000	3	0.97	0.3	4	6884.675	1
1000	3	0.99	0.3	2	10131.859	1
1000	3	0.99	0.3	3	349.482	0.890
1000	3	0.99	0.3	4	7262.340	1
1000	3	1	0.3	2	9842.764	1
1000	3	1	0.3	3	195.161	0.082
1000	3	1	0.3	4	8122.924	0.999
1000	5	0.95	0.3	2	16323.456	1
1000	5	0.95	0.3	3	1488.596	0.999
1000	5	0.95	0.3	4	11238.458	0.999
1000	5	0.97	0.3	2	16084.533	0.999
1000	5	0.97	0.3	3	1003.036	0.999
1000	5	0.97	0.3	4	11006.001	0.999
1000	5	0.99	0.3	2	16246.002	1
1000	5	0.99	0.3	3	462.494	0.967
1000	5	0.99	0.3	4	11669.805	1
1000	5	1	0.3	2	16618.797	1
1000	5	1	0.3	3	196.877	0.080
1000	5	1	0.3	4	13005.614	1

Note: trend, seasonal, and irregular case, proportion of rejections of H_0: $d_{j+1} = 0$ versus H_1: $d_{j+1} > 0$ in the subsampling test at a 5% nominal level ($R = 1000$ replicates of 4-variate time series, Bartlett taper, subsample size n selected adaptively) and the average test statistic.

Table C16. Trend, seasonal, and irregular size/power (Parzen).

T	$q^{(1)}$	$\rho^{(1)}$	b	$j+1$	Average test stat.	Empirical size/power
1000	1	0.95	0.3	2	3454.147	1
1000	1	0.95	0.3	3	315.317	0.791
1000	1	0.95	0.3	4	2438.745	0.999
1000	1	0.97	0.3	2	3414.047	1
1000	1	0.97	0.3	3	208.865	0.504
1000	1	0.97	0.3	4	2542.215	0.998
1000	1	0.99	0.3	2	3440.203	0.999
1000	1	0.99	0.3	3	104.355	0.051
1000	1	0.99	0.3	4	2847.3	1
1000	1	1	0.3	2	3481.681	1
1000	1	1	0.3	3	45.377	0
1000	1	1	0.3	4	2863.417	1
1000	3	0.95	0.3	2	10159.124	1
1000	3	0.95	0.3	3	869.335	0.989
1000	3	0.95	0.3	4	6927.782	1
1000	3	0.97	0.3	2	10155.594	1
1000	3	0.97	0.3	3	543.721	0.931
1000	3	0.97	0.3	4	7042.252	0.997
1000	3	0.99	0.3	2	10308.321	1
1000	3	0.99	0.3	3	217.452	0.491
1000	3	0.99	0.3	4	8015.146	1
1000	3	1	0.3	2	10261.554	1
1000	3	1	0.3	3	47.314	0
1000	3	1	0.3	4	8578.804	1
1000	5	0.95	0.3	2	16931.012	0.999
1000	5	0.95	0.3	3	1442.593	0.997
1000	5	0.95	0.3	4	11250.759	0.998
1000	5	0.97	0.3	2	16954.534	1
1000	5	0.97	0.3	3	899.47	0.989
1000	5	0.97	0.3	4	11666.628	1
1000	5	0.99	0.3	2	17056.249	1
1000	5	0.99	0.3	3	334.333	0.786
1000	5	0.99	0.3	4	12066.504	1
1000	5	1	0.3	2	17158.759	1
1000	5	1	0.3	3	48.339	0.003
1000	5	1	0.3	4	14099.387	0.999

Note: trend, seasonal, and irregular case, proportion of rejections of H_0: $d_{j+1} = 0$ versus H_1: $d_{j+1} > 0$ in the subsampling test at a 5% nominal level ($R = 1000$ replicates of 4-variate time series, Parzen taper, subsample size n selected adaptively) and the average test statistic.