

Contents lists available at [ScienceDirect](#)

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Frequency domain estimation of cointegrating vectors with mixed frequency and mixed sample data

Marcus J. Chambers

Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, England, United Kingdom

ARTICLE INFO

Article history:

Received 14 January 2018
Received in revised form 22 May 2019
Accepted 29 October 2019
Available online xxx

JEL classification:

C32

Keywords:

Mixed frequency data
Mixed sample data
Cointegration
Spectral regression

ABSTRACT

This paper proposes a simple method for exploiting the information contained in mixed frequency and mixed sample data in the estimation of cointegrating vectors. The asymptotic properties of easy-to-compute spectral regression estimators of the cointegrating vectors are derived and these estimators are shown to belong to the class of optimal cointegration estimators. Furthermore, Wald statistics based on these estimators have asymptotic chi-square distributions which enable inferences to be made straightforwardly. Simulation experiments suggest that the spectral regression estimators considered perform well in finite samples and are at least as good as time domain fully modified estimators. The finite sample size and power properties of the spectral regression-based Wald statistic are also found to be good.

© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The concept of cointegration plays a prominent role in the analysis of multivariate time series with unit roots, and a large variety of methods is available to the applied researcher for handling such data. Prominent among these methods is the vector error correction model (VECM) which is a convenient reparameterisation of a vector autoregressive (VAR) system that accounts for the cointegration between the variables. The popularity of the fully parametric VECM approach lies in its (relative) ease of estimation and its suitability for testing for the number of cointegrating vectors that exist. The VECM method is also implemented in many econometric software packages, is amenable to use as a forecasting tool and can be subjected to the usual battery of time series specification tests.

In some circumstances, however, a researcher may be unwilling to model the system dynamics in the form of a heavily-parameterised VAR but may still be interested in the cointegrating vectors themselves. In such cases alternative methods are available, including, but certainly not restricted to, the dynamic ordinary least squares (DOLS) method of [Stock and Watson \(1993\)](#), the fully modified (FM) least squares method of [Phillips and Hansen \(1990\)](#), and the spectral regression techniques proposed by [Phillips \(1991a\)](#). These approaches focus on the cointegrating vectors of interest and account for the system dynamics without needing to specify a parametric model. The DOLS approach adds leads and lags of first differences to the cointegrating regression; the FM least squares method employs nonparametric estimates of certain covariance matrices; and the spectral regression estimator is a type of feasible generalised least squares estimator in the frequency domain.

The vast majority of the contributions to the cointegration literature, both theoretical and applied, have focused on situations in which all the variables of interest are sampled at the same frequency. In cases where the variables are sampled at different frequencies this typically amounts to converting the higher frequency series into the lowest

E-mail address: mchamb@essex.ac.uk.

<https://doi.org/10.1016/j.jeconom.2019.10.010>

0304-4076/© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

frequency. In recent years, however, there has been a growing interest in developing methods that are capable of exploiting all the mixed frequency data without converting the higher frequency data to the lowest frequency. Mixed frequency approaches applicable to testing for cointegration have been developed by Ghysels and Miller (2014, 2015) and Miller and Wang (2016), while estimation of the cointegrating vectors in regression models using mixed frequency data has been investigated by Miller (2010, 2014, 2016). It is also possible to extend the VECM approach for use with mixed frequency data; see Seong et al. (2013).

It might be tempting to argue that, because cointegration describes a set of long-run relationships between variables, the use of additional high frequency data alongside the low frequency data is unlikely to yield many benefits. Indeed, the use of very high frequency data, of the type available in finance, might introduce additional complications, such as microstructure noise. Moreover, as first emphasised by Shiller and Perron (1985), it is the span of the data, rather than the number of observations, that is important when modelling nonstationary data. This suggests that there are limitations as to how far the high frequency data should be exploited. But, used appropriately, it is possible that the additional information contained in the higher frequency data can be used to improve the properties of estimators of cointegration vectors in finite samples, even though the asymptotic properties are likely to be the same as those obtained using just the low frequency data.

In this paper we adapt the spectral regression approach of Phillips (1991a) to the estimation of cointegrating vectors using mixed frequency data. We treat the mixed frequency issue in the context of a discrete time temporal aggregation problem where the highest observed frequency is taken as the fundamental frequency; an alternative continuous time approach can be found in Chambers (2019). An advantage of the spectral regression estimators is that the model's disturbances are only required to be stationary and to satisfy a functional central limit theorem (FCLT), meaning that there is no need to assume any particular form of parametric dynamic model. By addressing the temporal aggregation directly we are able to show that the disturbances in the mixed frequency model are, indeed, stationary.

This paper makes three main contributions. The first, indicated above, is the derivation of a model that can be used with mixed frequency and mixed sample data for the estimation of cointegrating vectors. In this sense its motivation is very similar to that of Miller (2016), some of whose results are used in the proofs. The proposed method of dealing with the mixed frequency data turns out to be very straightforward – simply average the high frequency variables over the low frequency sampling interval. This, in fact, was proposed by Chambers (2003) in his study of the asymptotic efficiency of cointegration estimators under temporal aggregation. Although the high frequency observations are not used separately, the averaging nevertheless does use the information contained in all the high frequency observations.

The second main contribution is the derivation of the asymptotic properties of the spectral regression estimators of the cointegrating vectors. The estimators we consider are band limited around the zero frequency in view of cointegration being associated with this frequency. A large literature exists on the estimation of spectral density matrices but we focus on smoothed periodogram estimators in view of their relative ease of computation and analysis. It is shown that the resulting spectral regression estimators fall into the class of optimal cointegration estimators as defined by Phillips (1991c) and have the familiar mixed normal limiting distribution. We also consider a spectral estimator based on a regression that is augmented by an additional variable in first-difference form. This avoids the need for the estimation of a spectral matrix based on the residuals from an initial (consistent) estimation of the cointegrating vectors. This augmented spectral estimator possesses the same form of optimal limiting distribution. A useful feature of these limiting distributions is that Wald statistics, formed using the spectral regression estimators, have limiting chi-square distributions, thereby making inference a straightforward procedure.

The third contribution concerns some simulation evidence for the proposed methods of estimation and inference in finite samples. The simulation model involves a single cointegrating relationship between two high frequency variables. We consider the performance of the spectral regression estimator and its augmented version, based on smoothed periodogram estimators of the spectral density matrices, as well as the regression estimator based on an autoregressive spectral density estimator. We compare estimates obtained from the infeasible high frequency model (where both variables are sampled at the high frequency), a feasible low frequency model, and three mixed frequency representations. The spectral estimators are found to have good finite sample properties and compare very favourably with the time domain FM estimator. The performance of Wald statistics is also examined, with those based on the spectral regression estimator having good size and power properties.

The spectral methods proposed in this paper are related to those in Phillips (1991a,b) but are distinguished from them in the following ways. Phillips (1991a) proposed a class of spectral regression estimators for a single cointegrating vector and for time series sampled at a common frequency. Here we allow for multiple cointegrating vectors as well as data that may be sampled at different frequencies while simultaneously distinguishing between stock and flow sampling. Phillips (1991b) examined the properties of discrete time data generated from an underlying cointegrated continuous time system. His theorem and examples show conditions under which filtered series remain cointegrated, and similar arguments could be used in our discrete time setting to show that the observed mixed frequency and mixed sample data also remain cointegrated. But we go further to derive the form of a cointegrated representation for the mixed frequency observations that provides a basis for estimation and demonstrate that the disturbances in this system are stationary. Furthermore we also provide explicit proofs of the asymptotic properties of the spectral estimators.

The paper is organised as follows. Section 2 defines the triangular model of cointegration at the high frequency and provides the mixed frequency representation. Stationarity of the disturbance vector in this representation is also

demonstrated. Issues concerning frequency domain estimation are addressed in Section 3, in which the estimators and test statistics are defined and their asymptotic properties derived. Section 4 defines the simulation experiments and reports the results obtained, while some concluding comments appear in Section 5. All proofs and supplementary results are presented in Appendix.

The following notation is used. The lag operator, L , is such that, for a variable x_t , $L^h x_t = x_{t-h}$ for some real number h (not necessarily whole). Following Phillips (1991b) a variable, x_t , is $I(0)$ if it belongs to the class of covariance stationary processes that have a spectral density function, $f(\lambda)$, that is bounded and continuous and for which $f(0)$ is positive. A variable is $I(1)$ if its first difference is $I(0)$, and a vector of variables will be said to be $I(0)$ or $I(1)$ if all its elements are of the same order of integration. In the vector case it is possible that each element of the first difference is $I(0)$ by this definition but the spectral density matrix is singular at the origin; in this case the vector of variables is said to be cointegrated. Convergence in distribution is denoted \xrightarrow{d} , convergence in probability by \xrightarrow{p} and almost sure convergence by $\xrightarrow{a.s.}$. Finally, $[x]$ denotes the smallest integer less than or equal to the scalar x , I_n denotes an $n \times n$ identity matrix, $0_{n \times m}$ is an $n \times m$ matrix of zeros, \otimes denotes the Kronecker product operator, $\text{tr}(A)$ denotes the trace of a square matrix A , $\|A\| = \sqrt{\text{tr}(AA^*)}$ denotes the Euclidean norm of A , B^* is the transpose of the conjugate of a complex-valued matrix B and, for an $n \times m$ matrix C , $\text{vec}(C)$ denotes the $nm \times 1$ column vector obtained by stacking the columns of C vertically on top of each other.

2. The model and a mixed frequency representation

The model concerns the cointegration properties of the elements of an $I(1)$ vector of variables, y , of dimension $n \times 1$. It is convenient to partition y as $y = (y_1', y_2')$ where y_1 is $n_1 \times 1$, y_2 is $n_2 \times 1$ and $n_1 + n_2 = n$. The $1 \leq n_1 \leq n - 1$ cointegrating equations are normalised on the sub-vector y_1 and are expressed as linear combinations of y_2 so that $y_1 - Cy_2$ is stationary, the $n_1 \times n_2$ matrix C containing the unknown cointegrating parameters of interest (the rows denoting the cointegration vectors). In the most general setting the elements of y_1 and y_2 are allowed to comprise both stock and flow variables and to be observed at both high and low frequencies (to be defined below). It is convenient to partition these vectors (without loss of generality) as

$$y_j = \begin{pmatrix} y_j^S \\ y_j^F \end{pmatrix} \begin{pmatrix} (n_j^S \times 1) \\ (n_j^F \times 1) \end{pmatrix} = \begin{pmatrix} y_j^{SH} \\ y_j^{SL} \\ y_j^{FH} \\ y_j^{FL} \end{pmatrix} \begin{pmatrix} (n_j^{SH} \times 1) \\ (n_j^{SL} \times 1) \\ (n_j^{FH} \times 1) \\ (n_j^{FL} \times 1) \end{pmatrix}, \quad j = 1, 2,$$

where $n_j^S + n_j^F = n_j$, $n_j^{SH} + n_j^{SL} = n_j^S$ and $n_j^{FH} + n_j^{FL} = n_j^F$ ($j = 1, 2$). In the above, the first superscript denotes the type of variable (stock, S , or flow, F) while the second indicates the frequency with which the vector is sampled (high frequency, H , or low frequency, L).

We assume that the high frequency sampling corresponds to a sampling interval of length $0 < h_H = h < 1$, while the low frequency sampling interval is normalised to $h_L = 1$. We also assume that $k = h^{-1}$ is an integer so that there is a whole number of high frequency observations per low frequency observation. The variables are generated at the high frequency according to the triangular cointegrated system¹

$$y_{1,\tau h} = Cy_{2,\tau h} + u_{1,\tau h}, \quad \tau = 1, \dots, N, \tag{1}$$

$$\Delta_h y_{2,\tau h} = u_{2,\tau h}, \quad \tau = 1, \dots, N, \tag{2}$$

where $\Delta_h = 1 - L^h$ denotes the high frequency first-difference operator, N denotes the number of high frequency sampling periods, and $u_{1,\tau h}$ ($n_1 \times 1$) and $u_{2,\tau h}$ ($n_2 \times 1$) are disturbance vectors whose properties are defined below. The cointegration in the system is depicted by (1), while (2) denotes the n_2 unit roots/stochastic trends. We shall refer to (1) and (2) as being the *High Frequency Representation (HFR)*. With regard to the disturbance vector we make the following assumption.

Assumption 1. The $n \times 1$ vector $u_{\tau h} = (u'_{1,\tau h}, u'_{2,\tau h})'$ satisfies

$$u_{\tau h} = A(L^h)\epsilon_{\tau h}, \quad \tau = 1, \dots, N, \tag{3}$$

where:

(a) $A(z) = \sum_{j=0}^{\infty} A_j z^j$, $A_0 = I_n$, $A(1) \neq 0_{n \times n}$ and $\sum_{j=0}^{\infty} j \|A_j\| < \infty$;

(b) $\epsilon_{\tau h}$ is a martingale difference sequence (mds), with natural filtration $\mathcal{F}_{\tau h}$, satisfying:

¹ Intercepts and deterministic trends in the model can be handled by the methods in this paper subject to suitable demeaning and detrending of the data prior to spectral regression. Further discussion of this issue is provided in Section 5.

- (i) $E(\epsilon_{\tau h} | \mathcal{F}_{\tau h-h}) = 0$;
- (ii) $\frac{1}{N} \sum_{\tau=1}^N E(\epsilon_{\tau h} \epsilon'_{\tau h} | \mathcal{F}_{\tau h-h}) \xrightarrow{a.s.} E(\epsilon_{\tau h} \epsilon'_{\tau h}) = \Sigma_{\epsilon}$ as $N \rightarrow \infty$, where Σ_{ϵ} is a symmetric positive definite matrix;
- (iii) $E(\|\epsilon_{\tau h}\|^4 | \mathcal{F}_{\tau h-h}) \leq K < \infty$ a.s. for all $\tau = 1, \dots, N$.

Assumption 1 defines $u_{\tau h}$ as a linear process driven by an mds disturbance vector; it therefore allows certain types of conditional heteroskedasticity (such as autoregressive conditional heteroskedasticity, or ARCH) provided condition (b)(ii) holds. In view of the constant unconditional variance it follows that the spectral density matrix of $u_{\tau h}$ is given by

$$f_{uu}(\lambda) = \frac{1}{2\pi} A(e^{-i\lambda}) \Sigma_{\epsilon} A(e^{i\lambda})', \quad -\pi/h < \lambda \leq \pi/h; \tag{4}$$

the condition on $A(z)$ in **Assumption 1(a)** ensures that $f_{uu}(0) = (1/2\pi)A(1)\Sigma_{\epsilon}A(1)'$ is positive definite.² An implication of **Assumption 1** is that $u_{\tau h}$ satisfies the FCLT

$$\frac{1}{\sqrt{N}} \sum_{\tau=1}^{[Nr]} u_{\tau h} \xrightarrow{d} B_u(r) \text{ as } N \rightarrow \infty, \quad 0 < r \leq 1, \tag{5}$$

where $B_u(r)$ is a Brownian motion process with covariance matrix $\Omega_u = 2\pi f_{uu}(0)$. The FCLT is used in the derivation of the asymptotic properties of the estimators.

The problem with the system (1) and (2) for the estimation of C is that the low frequency variables are not observed at the high frequency. To be precise, the high frequency observations are given by

$$\{y_{1,\tau h}^{SH}\}_{\tau=1}^N, \quad \{y_{1,\tau h}^{FH}\}_{\tau=1}^N, \quad \{y_{2,\tau h}^{SH}\}_{\tau=1}^N, \quad \{y_{2,\tau h}^{FH}\}_{\tau=1}^N,$$

while the low frequency observations are of the form

$$\{y_{1t}^{SL}\}_{t=1}^T, \quad \{y_{1t}^{FL}\}_{t=1}^T, \quad \{y_{2t}^{SL}\}_{t=1}^T, \quad \{y_{2t}^{FL}\}_{t=1}^T,$$

where T denotes the number of low frequency observations (and is also the time span that the data cover); in fact, $T = Nh$. The low frequency flow variables are of the form

$$y_{jt}^{FL} = \frac{1}{k} \sum_{l=0}^{k-1} y_{j,t-lh}^{FL}, \quad j = 1, 2, \quad t = 1, \dots, T,$$

i.e. the low frequency flows are averages of the (unobservable) high frequency flows $y_{j,\tau h}^{FL}$ over the low frequency observation interval $t - (k - 1)h \leq \tau \leq t$.³ However, cointegration is a property that persists at any sampling frequency, and so observations at the low frequency are also cointegrated. Re-writing (1) at the low frequency (essentially setting $t = \tau h$ and picking out the integer values for this index) yields

$$y_{1t} = Cy_{2t} + u_{1t}, \quad t = 1, \dots, T. \tag{6}$$

The corresponding stochastic trends in (2) can be transformed to the low frequency by the application of the filter $s(L^h)$ where

$$s(z) = 1 + z + \dots + z^{k-1}; \tag{7}$$

noting that $s(L^h)\Delta_h y_{2,\tau h} = y_{2,\tau h} - y_{2,\tau h-kh} = y_{2t} - y_{2,t-1}$ we obtain

$$\Delta y_{2t} = w_{2t}, \quad t = 1, \dots, T, \tag{8}$$

where $\Delta = 1 - L$ and $w_{2t} = s(L^h)u_{2t} = \sum_{l=0}^{k-1} u_{2,t-lh}$. However, the low frequency representation in (6) and (8) is also not directly amenable to the estimation of C because neither y_{1t}^{FL} nor y_{2t}^{FL} is observable (the aggregates, Y_{1t}^{FL} and Y_{2t}^{FL} , are observable instead) and hence it can be regarded as an *Infeasible Low Frequency Representation*. The challenge is to utilise the low and high frequency representations so that all of the information contained in the observed sample – at both the high and low frequencies – can be used to estimate C .

A simple, but effective, method of combining the mixed frequency data at the low frequency is to simply average the high frequency stock and flow variables. Such a procedure was suggested by Chambers (2003), albeit in a continuous time setting, for improving the asymptotic efficiency of cointegration estimators when observations on higher frequency stock

² Note that the frequency range is $(-\pi/h, \pi/h]$ because $u_{\tau h}$ is defined at the high frequency and the frequency range for low frequency data is normalised to be $(-\pi, \pi]$.

³ Chysels and Miller (2015) and Miller (2016) consider more general types of aggregation of the form $\sum_{l=0}^{k-1} \omega_{l+1} y_{j,t-lh}^{FL}$ where the aggregation weights sum to unity and may be known or unknown; implicitly we are setting $\omega_{i+1} = k^{-1}$ ($i = 1, \dots, k$). The assumption of known weights here avoids the need for employing a method such as MIDAS.

variables were available; similar efficiency arguments could be shown to hold in the discrete time setting employed here. We therefore construct low frequency averages of the form

$$Y_{jt}^{iH} = \frac{1}{k} \sum_{l=0}^{k-1} y_{j,t-lh}^{iH}, \quad i = S, F, \quad j = 1, 2, \quad t = 1, \dots, T,$$

resulting in the averaged variables Y_{1t}^{SH} , Y_{1t}^{FH} , Y_{2t}^{SH} and Y_{2t}^{FH} . We shall refer to the resulting model as the *Mixed Frequency Representation (MFR)* and its form is presented below.

Lemma 1. Let y_1 and y_2 satisfy the high frequency cointegrated system in (1) and (2), and define the observable averaged vectors

$$x_{1t} = \begin{pmatrix} Y_{1t}^{SH} \\ y_{1t}^{SL} \\ Y_{1t}^{FH} \\ Y_{1t}^{FL} \end{pmatrix}, \quad x_{2t} = \begin{pmatrix} Y_{2t}^{SH} \\ y_{2t}^{SL} \\ Y_{2t}^{FH} \\ Y_{2t}^{FL} \end{pmatrix}, \quad t = 1, \dots, T.$$

Then the Mixed Frequency Representation is, for $t = 2, \dots, T$,

$$x_{1t} = Cx_{2,t-1} + \xi_{1t}, \tag{9}$$

$$\Delta x_{2t} = \xi_{2t}, \tag{10}$$

where the disturbance vector $\xi_t = (\xi'_{1t}, \xi'_{2t})'$ is $I(0)$ under Assumption 1.

The representation in Lemma 1 provides a basis for the estimation of the matrix of cointegration vectors, C . Its derivation is based on the infeasible low frequency representation in (6) and (8) in which deviations between observable (possibly averaged) and unobservable components, e.g. $y_{2,t-1}^{SH} - Y_{2,t-1}^{SH}$, are assigned to the disturbance vector, ξ_t . Such quantities can be shown to be stationary; see Lemma A1 in Appendix. The MFR is, therefore, based on the entire sample of mixed frequency observable variables even though the high frequency variables are not included separately at each high frequency time point but are averaged to, in effect, mimic the observed flow variables. It is precisely this form of averaging of (stock) variables that was proposed by Chambers (2003) to improve the efficiency of the estimation of cointegration vectors when the stocks are available at a higher frequency than the flows; it is also nested within the aggregation schemes considered in Miller (2016).⁴

It might be tempting to argue that the mixed frequency representation is discarding data by aggregating the high frequency stocks rather than including them separately. However, an important feature of the MFR in Lemma 1 to note is that it retains the n_1 cointegration equations and the n_2 stochastic trends of the underlying high frequency model. Alternative approaches that use the high frequency observations separately have been shown to be possible in some circumstances. For example, consider the vector $z_t = (z'_{1t}, z'_{2t})'$, where

$$z_{jt} = (y_{jt}^{SH'}, \dots, y_{j,t-(k-1)h}^{SH'}, y_{jt}^{FH'}, \dots, y_{j,t-(k-1)h}^{FH'}, y_{jt}^{SL'}, y_{jt}^{FL'})', \quad j = 1, 2;$$

each z_{jt} contains $n_{zj} = k(n_j^{SH} + n_j^{FH}) + n_j^{SL} + n_j^{FL}$ elements, so that z_t is of dimension $n_z \times 1$, where $n_z = n_{z1} + n_{z2}$. Ghysels (2016) considers a VAR representation in the expanded vector z_t while Chambers (2016) and Thornton (2019) examine continuous time systems and exploit the restrictions on the discrete time representation arising from temporal aggregation which results in a more parsimonious system than an unrestricted VAR in z_t .

It would, of course, also be possible to specify a similar VAR representation in the cointegrated system considered here but the resulting vector of disturbances – an expanded version of ξ_t defined in Lemma 1 – would then have a singular spectral density matrix. The reason for this is that the expanded ξ_t , say $\tilde{\xi}_t$ (which contains $n_z > n$ elements), is a function of only n underlying random variables contained in the vector u_t . In other words, we can write $\tilde{\xi}_t = H(L)u_t$ where $H(z)$ is an $n_z \times n$ matrix whose elements are polynomials that depict the way u_t and its high frequency lags feed into $\tilde{\xi}_t$. The spectral density matrix of $\tilde{\xi}_t$, given by $H(e^{i\lambda})f_{uu}(\lambda)H(e^{-i\lambda})'$, is then singular. In particular, the inverse of this matrix at the origin ($\lambda = 0$) characterises the limiting distribution of the spectral regression estimator, and therefore causes a degeneracy in this expanded system. We now turn to the analysis of a frequency domain-based estimator of C that rests only on the weak Assumption 1 concerning the disturbances in the HFR.

3. Estimation in the frequency domain

In view of the level of generality associated with the model of cointegration developed in the previous section, in which the disturbance vector, ξ_t , is a linear process under Assumption 1 rather than having any specific (parametric)

⁴ It is also possible to derive a *Feasible Low Frequency Representation* that skip-samples the high frequency variables at the low frequency, thereby discarding entirely all the information contained in the observations at the intermediate points. This representation is sampling scheme 2 in the simulation study in Section 4.

dynamic structure, a natural approach to estimating the matrix C of cointegrating vectors is to use spectral/frequency domain regression. Let $x_{0t} = (x'_{1t}, \Delta x'_{2t})'$ and define $J = (I_{n_1}, 0_{n_1 \times n_2})'$. Then, based on the MFR in Lemma 1, we can write the system of interest as⁵

$$x_{0t} = J C x_{2,t-1} + \xi_t, \quad t = 1, \dots, T. \tag{11}$$

Taking discrete Fourier transforms (dFts) in (11) and assuming T to be even⁶ yields

$$w_0(\lambda_s) = J C w_2(\lambda_s) + w_\xi(\lambda_s), \quad s = -T/2 + 1, \dots, T/2, \tag{12}$$

where $\{\lambda_s = 2\pi s/T; s = -T/2 + 1, \dots, T/2\}$ denotes the set of Fourier frequencies and

$$w_0(\lambda_s) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T x_{0t} e^{it\lambda_s}, \quad w_2(\lambda_s) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T x_{2,t-1} e^{it\lambda_s}, \quad w_\xi(\lambda_s) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T \xi_t e^{it\lambda_s},$$

denote the dFts of x_{0t} , $x_{2,t-1}$ and ξ_t , respectively, at the Fourier frequencies. When C is unrestricted a simple spectral regression estimator can be obtained by minimising

$$S_1(C) = \frac{1}{\#(\Lambda)} \sum_{\lambda_s \in \Lambda} \text{tr} \{ w_\xi(\lambda_s) w_\xi(\lambda_s)^* \},$$

where $w_\xi(\lambda_s) = w_0(\lambda_s) - J C w_2(\lambda_s)$, Λ denotes the set of frequencies over which the estimator is to be determined, and $\#(\Lambda)$ denotes the number of frequencies in Λ . In the most general case Λ consists of all the Fourier frequencies in the interval $(-\pi, \pi]$; however, if the model is believed to hold only over a subset of $(-\pi, \pi]$ then Λ can be restricted accordingly, resulting in a band-limited estimator. In all situations we require both λ_s and $-\lambda_s$ to belong to Λ .

In the case of cointegration there are compelling reasons to limit Λ to a set of frequencies around the origin based on the theoretical arguments in Phillips (1991a,b) as well as the simulation results reported in Corbae et al. (1994). We therefore consider the symmetric set of frequencies $\Lambda_0 = \{\lambda_s = 2\pi s/T; s = -m, \dots, m\}$ which contains the $2m + 1$ Fourier frequencies around the origin for some integer m . As in Robinson (1972) we also generalise the objective function by incorporating a positive definite Hermitian weighting matrix, $\Phi(\lambda)$, resulting in an objective function of the form

$$S_2(C) = \frac{1}{2m + 1} \sum_{\lambda_s \in \Lambda_0} \text{tr} \{ \Phi(\lambda_s) w_\xi(\lambda_s) w_\xi(\lambda_s)^* \}.$$

However, as argued by Phillips (1991a), the choice of the weighting matrix $\Phi(\lambda)$ is critical when spectral regression is applied using $I(1)$ time series. For reasons of efficiency we require $\Phi(\lambda)$ to be proportional to $f_{\xi\xi}(\lambda)^{-1}$, the inverse of the spectral density matrix of the unobservable disturbance vector ξ_t . Although ξ_t is unobserved a consistent estimator of $f_{\xi\xi}(\lambda)$ can be obtained by using the residuals, $\hat{\xi}_t$, from a least squares regression of (11).

Our estimator of $f_{\xi\xi}(0)$ is the smoothed periodogram estimator, defined by

$$\hat{f}_{\hat{\xi}\hat{\xi}}(0) = \frac{1}{2m + 1} \sum_{j=-m}^m I_{\hat{\xi}\hat{\xi}}(\lambda_j), \tag{13}$$

where $I_{\hat{\xi}\hat{\xi}}(\lambda) = w_{\hat{\xi}}(\lambda) w_{\hat{\xi}}(\lambda)^*$ and $w_{\hat{\xi}}(\lambda)$ is the dFt of $\hat{\xi}_t$; it is therefore a straightforward symmetric average of $2m + 1$ periodogram matrices around the zero frequency. More sophisticated estimates could be used but the smoothed periodogram has performed well in the simulations that are reported in the next section, despite its simplicity. With this choice of weighting matrix the objective function becomes

$$S(C) = \frac{1}{2m + 1} \sum_{s=-m}^m \text{tr} \left\{ \hat{f}_{\hat{\xi}\hat{\xi}}(0)^{-1} (w_0(\lambda_s) - J C w_2(\lambda_s)) (w_0(\lambda_s) - J C w_2(\lambda_s))^* \right\}. \tag{14}$$

Minimisation of (14) with respect to C results in the estimator

$$\hat{C}_0 = \left(J' \hat{f}_{\hat{\xi}\hat{\xi}}(0)^{-1} J \right)^{-1} J' \hat{f}_{\hat{\xi}\hat{\xi}}(0)^{-1} \hat{f}_{02}(0) \hat{f}_{22}(0)^{-1} \tag{15}$$

where the spectral density estimators $\hat{f}_{02}(0)$ and $\hat{f}_{22}(0)$ are defined by

$$\hat{f}_{22}(0) = \frac{1}{2m + 1} \sum_{j=-m}^m I_{22}(\lambda_j), \quad \hat{f}_{02}(0) = \frac{1}{2m + 1} \sum_{j=-m}^m I_{02}(\lambda_j),$$

and where $I_{22}(\lambda_j) = w_2(\lambda_j) w_2(\lambda_j)^*$ and $I_{02}(\lambda_j) = w_0(\lambda_j) w_2(\lambda_j)^*$.

⁵ We assume, for convenience, that x_{0t} and $x_{2,t-1}$ are observed for $t = 1, \dots, T$ rather than just $t = 2, \dots, T$.

⁶ If T is odd then we can take $-[T/2] + 1 \leq s \leq [T/2]$.

An alternative, asymptotically equivalent, estimator of C , that avoids estimation of the spectral matrix $f_{\xi\xi}(0)$, can be obtained by augmenting (9) with Δx_{2t} as an additional regressor, leading to consideration of

$$x_{1t} = Cx_{2,t-1} + F\Delta x_{2t} + \xi_{1,2t}, \quad t = 1, \dots, T, \tag{16}$$

where $F = \Omega_{12}\Omega_{22}^{-1}$ and $\xi_{1,2t} = \xi_{1t} - F\xi_{2t}$. In the frequency domain the relevant equation is

$$w_1(\lambda_s) = Cw_2(\lambda_s) + Fw_{\Delta_2}(\lambda_s) + w_{1,2}(\lambda_s), \quad s = -T/2 + 1, \dots, T/2, \tag{17}$$

where $w_1(\lambda_s)$, $w_{\Delta_2}(\lambda_s)$ and $w_{1,2}(\lambda_s)$ are the dFts of x_{1t} , Δx_{2t} and $\xi_{1,2t}$, respectively. The band-limited estimator of C based on the augmented equation is obtained by minimising

$$S_A(C) = \frac{1}{2m+1} \sum_{s=-m}^m \text{tr} \{w_{1,2}(\lambda_s)w_{1,2}(\lambda_s)^*\}, \tag{18}$$

where $w_{1,2}(\lambda_s) = w_1(\lambda_s) - Cw_2(\lambda_s) - Fw_{\Delta_2}(\lambda_s)$. The resulting estimator is given by

$$\hat{C}_0^A = \left(\hat{f}_{12}(0) - \hat{f}_{1\Delta_2}(0)\hat{f}_{\Delta_2\Delta_2}(0)^{-1}\hat{f}_{\Delta_22}(0)\right) \left(\hat{f}_{22}(0) - \hat{f}_{2\Delta_2}(0)\hat{f}_{\Delta_2\Delta_2}(0)^{-1}\hat{f}_{\Delta_22}(0)\right)^{-1}, \tag{19}$$

where the $\hat{f}_{ab}(0)$ are the smoothed periodogram estimators using the relevant variables.

In order to derive the asymptotic properties of the estimators it is convenient to relate them directly to the matrix C . By noting that $w_0(\lambda) = JCw_2(\lambda) + w_\xi(\lambda)$ it follows that $\hat{f}_{02}(0) = J\hat{C}\hat{f}_{22}(0) + \hat{f}_{\xi 2}(0)$, and making this substitution in (15) we obtain

$$\hat{C}_0 = C + \left(J'\hat{f}_{\xi\xi}(0)^{-1}J\right)^{-1} J'\hat{f}_{\xi\xi}(0)^{-1}\hat{f}_{\xi 2}(0)\hat{f}_{22}(0)^{-1}. \tag{20}$$

Adopting a similar procedure for \hat{C}_0^A , using $w_1(\lambda) = Cw_2(\lambda) + Fw_{\Delta_2}(\lambda) + w_{1,2}(\lambda)$ and making the appropriate substitutions, we find that

$$\hat{C}_0^A = C + \left(\hat{f}_{1,2,2}(0) - \hat{f}_{1,2,\Delta_2}(0)\hat{f}_{\Delta_2\Delta_2}(0)^{-1}\hat{f}_{\Delta_22}(0)\right) \left(\hat{f}_{22}(0) - \hat{f}_{2\Delta_2}(0)\hat{f}_{\Delta_2\Delta_2}(0)^{-1}\hat{f}_{\Delta_22}(0)\right)^{-1}, \tag{21}$$

where the subscript 1.2 denotes a quantity based on the dft of $\xi_{1,2t}$.⁷

The derivation of the asymptotic properties of \hat{C}_0 and \hat{C}_0^A clearly relies on the asymptotic properties of various spectral density matrix estimators, which in turn will be driven by an FCLT for the normalised partial sums of ξ_t . Based on (5) – the FCLT for partial sums of $u_{\tau h}$ that holds under Assumption 1 – it is possible to derive an appropriate FCLT for the partial sums of ξ_t , which are functions of the partial sums of $u_{\tau h}$. This is presented below.

Lemma 2. Under Assumption 1, as $T \rightarrow \infty$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \xi_t \xrightarrow{d} B(r), \quad 0 < r \leq 1, \tag{22}$$

where $B(r)$ is a Brownian motion process with covariance matrix $\Omega = h^{-1}G\Omega_u G'$ and

$$G = \begin{pmatrix} hI_{n_1} & C \\ \mathbf{0}_{n_2 \times n_1} & I_{n_2} \end{pmatrix}.$$

The key to establishing Lemma 2 lies in utilising the precise relationship between ξ_t and $u_{\tau h}$ (that arises in the proof of Lemma 1) and then relating the two partial sum processes. The matrix G arises through use of a Beveridge–Nelson-type of decomposition of the matrix filter linking ξ_t and $u_{\tau h}$; details of this filter can be found in Lemma A2 in Appendix. In addition to Lemma 2 a result concerning the growth rate of autocovariances of ξ_t is useful in establishing the asymptotic properties of \hat{C}_0 and \hat{C}_0^A , and is given below.

Lemma 3. Let $\Gamma_{\xi,l} = E(\xi_t \xi_{t-l}')$. Then, under Assumption 1,

$$\sum_{l=-T}^T |l| \|\Gamma_{\xi,l}\| = O(T^{1/2}).$$

⁷ Note that such dft's are used purely as theoretical quantities in the proofs and are in no way used to compute the estimator itself.

Lemma 3 is used to establish a consistency result concerning $\hat{f}_{\xi\xi}(0)$ which is provided in **Theorem 1(c)**. The derivation of the asymptotic properties of \hat{C}_0 and \hat{C}_0^A also requires an assumption concerning the number, m , of frequencies employed in the estimation of the relevant spectral density matrices. To this end we make the following assumption.

Assumption 2. $\frac{m}{T} + \frac{1}{m} \rightarrow 0$ as $T \rightarrow \infty$.

Assumption 2 is common in the literature on spectral density estimation; see, for example, **Brockwell and Davis (1991, p.351)**. The use of **Assumptions 1 and 2** enables the following result concerning the asymptotic properties of the smoothed periodogram estimators of spectral density matrices to be established.

Theorem 1. Let $B(r) = (B_1(r)', B_2(r)')$. Then, under **Assumptions 1 and 2**, as $T \rightarrow \infty$:

- (a) $\frac{2m+1}{T^2} \hat{f}_{22}(0) \xrightarrow{d} \frac{1}{\pi} \int_0^1 B_2 B_2'$;
- (b) $\frac{2m+1}{T} \hat{f}_{\xi 2}(0) \xrightarrow{d} \frac{1}{\pi} \int_0^1 dB B_2' + \frac{1}{2\pi} \Omega_2$, where $\Omega_2 = \sum_{j=-\infty}^{\infty} E(\xi_{t+j} \xi_{2t}')$;
- (c) $\hat{f}_{\xi\xi}(0) = f_{\xi\xi}(0) + o_p(1)$.

It is convenient to partition Ω conformably with $B_1(r)$ and $B_2(r)$ in the form

$$\Omega = (\Omega_1 \ \Omega_2) = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$$

and to define $\Omega_{11.2} = \Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21}$. Note that the $n \times n_2$ matrix Ω_2 is the same matrix that appears in **Theorem 1(b)**. The asymptotic distributions of \hat{C}_0 and \hat{C}_0^A can now be stated.

Theorem 2. Under **Assumptions 1 and 2**, as $T \rightarrow \infty$,

$$T(\hat{C}_0 - C), T(\hat{C}_0^A - C) \xrightarrow{d} \int_0^1 dB_{1.2} B_2' \left(\int_0^1 B_2 B_2' \right)^{-1}$$

where $B_{1.2}(r)$ is a Brownian motion process with covariance matrix $\Omega_{11.2}$.

The form of limit distribution in **Theorem 2** shows that \hat{C}_0 and \hat{C}_0^A belong to the class of optimal estimators of **Phillips (1991c)** and that such optimality can be achieved either through system-wide estimation (yielding \hat{C}_0) or through augmentation of the cointegration equations (yielding \hat{C}_0^A). A particular advantage of optimal estimators is that their mixed normal limiting distributions enable traditional asymptotic chi-square hypothesis testing in appropriate circumstances. Suppose that interest centres on a set of $q < n_1 n_2$ possibly non-linear restrictions on the elements of C , represented by the null hypothesis

$$H_0 : r(\gamma) = 0,$$

where $\gamma = \text{vec}(C)$ and $r(\cdot)$ is a $q \times 1$ vector whose elements are twice continuously differentiable functions of γ . Let $R(\gamma) = \partial r(\gamma) / \partial \gamma'$ be the $q \times n_1 n_2$ matrix of first derivatives, assumed to be of rank q . Then a Wald statistic for testing H_0 based on \hat{C}_0 against the alternative $H_1 : r(\gamma) \neq 0$ is given by

$$W_0 = \frac{2m+1}{2} r(\hat{\gamma}_0)' \left(R(\hat{\gamma}_0) \hat{V}_0^{-1} R(\hat{\gamma}_0)' \right)^{-1} r(\hat{\gamma}_0), \tag{23}$$

where $\hat{\gamma}_0 = \text{vec}(\hat{C}_0)$ and

$$\hat{V}_0 = \hat{f}_{22}(0) \otimes J' \hat{f}_{\xi\xi}(0)^{-1} J.$$

A Wald statistic can also be defined using \hat{C}_0^A ; it is given by

$$W_0^A = \frac{2m+1}{2} r(\hat{\gamma}_0^A)' \left(R(\hat{\gamma}_0^A) (\hat{V}_0^A)^{-1} R(\hat{\gamma}_0^A)' \right)^{-1} r(\hat{\gamma}_0^A), \tag{24}$$

where $\hat{\gamma}_0^A = \text{vec}(\hat{C}_0^A)$ and

$$\hat{V}_0^A = \left(\hat{f}_{22}(0) - \hat{f}_{2\Delta_2}(0) \hat{f}_{\Delta_2\Delta_2}(0)^{-1} \hat{f}_{\Delta_2 2}(0) \right) \otimes \hat{f}_{11.2}(0)^{-1}.$$

For \hat{V}_0^A we require $\hat{f}_{11.2}(0)$ to be a consistent estimator of $f_{11.2}(0)$; this can be achieved using the smoothed periodogram estimator

$$\hat{f}_{11.2}(0) = \frac{1}{2m+1} \sum_{j=-m}^m \hat{w}_{1.2}(\lambda_j) \hat{w}_{1.2}(\lambda_j)^*$$

where $\hat{w}_{1,2}(\lambda_j) = w_1(\lambda_j) - \hat{C}_0^A w_2(\lambda_j) - \hat{F}_0^A w_{\Delta_2}(\lambda_j)$, which is consistent under [Assumptions 1](#) and [2](#).⁸ The limiting distributions of these Wald statistics are given below.

Theorem 3. Under [Assumptions 1](#) and [2](#), as $T \rightarrow \infty$, $W_0, W_0^A \xrightarrow{d} \chi_q^2$ under H_0 .

Asymptotic chi-square inference can therefore be conducted based on both band-limited spectral regression estimators. A simulation analysis of the finite sample properties of the estimators and Wald tests is provided in [Section 4](#).

4. Simulation results

In this section we explore the finite sample properties of the spectral regression estimators and Wald statistics. Our focus is on a bivariate model with cointegrating parameter $C = 1$ so that $y_1 - y_2$ is stationary. One advantage of a simulation exercise is that data can be generated at any chosen frequency and aggregated as required. We can therefore investigate the properties of the estimators and tests in the following five sampling schemes:

Scheme 1 is the infeasible case where high frequency observations on both variables are available i.e. $y_{1,\tau h}$ and $y_{2,\tau h}$ for $\tau = 1, \dots, N$. The variables could be stocks or flows but their nature is irrelevant in this case as they satisfy the HFR.

Scheme 2 uses only low frequency observations of the variables that are generated by the HFR; in terms of the MFR in [Lemma 1](#) we have $x_{1t} = y_{1t}$ and $x_{2t} = y_{2t}$ for $t = 1, \dots, T$. This would correspond to a situation where only low frequency observations on two stock variables were available despite them actually being generated at the high frequency.

Schemes 3–5 use combinations of averaged and/or low frequency observations. In scheme 3 we have $x_{1t} = y_{1t}$ and $x_{2t} = Y_{2t}$; scheme 4 uses $x_{1t} = Y_{1t}$ and $x_{2t} = y_{2t}$; and scheme 5 has $x_{1t} = Y_{1t}$ and $x_{2t} = Y_{2t}$. These can correspond to various types of stock and flow sampling; for example, scheme 5 represents averaging a high frequency stock combined with a low frequency flow, or various sampling of flow variables with the high frequency observations being averaged.

The simulations take the data span to be $T = 100$ and the high frequency sampling interval to be $h = 1/3$, which leads to $N = 300$ high frequency observations. Data are generated at the highest frequency (i.e. $y_{1,\tau h}$ and $y_{2,\tau h}$ for $\tau = 1, \dots, N$) and then aggregated as required. The high frequency bivariate innovations satisfy a first-order vector autoregression of the form

$$u_{\tau h} = \Psi u_{\tau h-h} + \epsilon_{\tau h}, \quad \tau = 1, \dots, N,$$

where $\epsilon_{\tau h}$ is a Gaussian white noise process with an identity covariance matrix and the following autoregressive matrices were used: $\Psi_0 = 0_{2 \times 2}$ (so that $u_{\tau h}$ is also Gaussian white noise), $\Psi_1 = 0.4I_2$ and $\Psi_2 = 0.8I_2$, the latter two having repeated roots of 2.5 and 1.25, respectively. We also consider ARCH innovations when the VAR matrices are Ψ_0 and Ψ_2 , in which case each $\epsilon_{j,\tau h}$ ($j = 1, 2$) is Gaussian with variance $\sigma_{j,\tau h}^2$ and $\sigma_{j,\tau h} = \alpha \epsilon_{j,\tau h-h}^2$ with $\alpha = 0.9$. There are, therefore, five different generating schemes under consideration and five different sampling schemes for each one. In all cases $u_0 = (0, 0)'$.

A total of 10,000 replications of each model for $u_{\tau h}$ were conducted and estimates under each of the five sampling schemes were computed. In addition to the ordinary least squares (OLS) estimator of C we also consider three different spectral regression estimators as well as the time domain fully-modified (FM) OLS estimator, each using three different values of the bandwidth parameter m , resulting in thirteen different estimates of C in each replication.⁹ The first spectral regression estimator is \hat{C}_0 , defined in [\(15\)](#), in which the smoothed periodogram estimator $\hat{f}_{\xi\xi}(0)$ is based on a set of OLS residuals, $\hat{\xi}_i$; this estimator is denoted FD in what follows. The second is the augmented estimator \hat{C}_0^A , defined in [\(19\)](#), and is denoted FDA. The third estimator is \hat{C}_0 but is based on an autoregressive spectral density estimator (ASDE) of $f_{\xi\xi}(0)$ rather than a smoothed periodogram estimator; this is denoted ASD.¹⁰ The ASDE of $f_{\xi\xi}(0)$ first fits a first-order VAR to the OLS residuals and then computes the estimator of the spectral density matrix at the origin using the expression

$$\hat{f}_{\xi\xi}^{ASDE}(0) = (I_2 - \hat{K})^{-1} \hat{\Sigma}_v (I_2 - \hat{K}')^{-1},$$

where \hat{K} and $\hat{\Sigma}_v$ denote the estimated autoregressive and covariance matrices in the VAR, respectively. The choice of m is required to satisfy [Assumption 2](#) and so we take $m = [T^\delta]$ in schemes 2–5 with $\delta \in \{0.3, 0.5, 0.7\}$; for $T = 100$ this results in $m \in \{3, 10, 25\}$. In the infeasible high frequency case (scheme 1) we scale these values by k leading to $m \in \{9, 30, 75\}$. The estimators based on each choice of δ are distinguished by appending 1, 2 or 3 to their abbreviated name, corresponding to the three values of δ in increasing order. Hence FD1 refers to \hat{C}_0 using $\delta = 0.3$, FD2 refers to \hat{C}_0 based on $\delta = 0.5$, and so on.

The simulation results concerning the performance of the estimators of C are presented in [Table 1](#). In view of the well-known trade-off between bias and variance in spectral density estimation the Table reports the root mean squared

⁸ Here, \hat{F}_0^A is the band-limited estimator of F in the augmented regression.

⁹ We use the same number of estimates of the autocovariances in the time domain FM estimators as the number of frequencies in the spectral estimators.

¹⁰ Smoothed periodogram estimators are used to estimate the remaining spectral density matrices.

Table 1
Simulated RMSE ($\times 10^4$) of frequency domain and fully modified estimators, $T = 100$, $h = 1/3$.

Model	Sampling	OLS	FD1	FD2	FD3	FDA1	FDA2	FDA3	ASD1	ASD2	ASD3	FM1	FM2	FM3
White noise ($\Psi = 0_{2 \times 2}$)	1	144.61	82.04	79.43	78.53	82.27	79.44	78.49	81.54	79.34	78.50	80.14	82.48	84.53
	2	378.28	161.06	142.88	139.80	162.98	143.01	140.04	156.85	142.69	140.54	140.37	143.58	147.20
	3	396.59	179.24	161.22	167.78	179.76	160.85	166.01	173.09	160.52	165.42	156.83	160.53	165.13
	4	257.59	119.29	106.91	104.29	121.20	106.92	104.16	116.32	106.63	104.64	107.59	119.91	142.58
	5	264.53	95.94	82.81	80.90	93.50	82.27	80.59	90.37	82.52	81.34	80.95	83.36	85.38
VAR(1) ($\Psi = 0.4I_2$)	1	96.62	82.26	79.60	78.82	82.43	79.62	78.82	81.82	79.60	78.89	80.43	83.11	85.30
	2	268.19	114.71	99.73	97.59	113.87	99.41	97.55	110.05	99.64	98.20	97.98	100.83	103.49
	3	328.08	124.16	109.54	119.65	120.30	108.60	117.74	116.18	108.80	117.14	105.10	107.24	110.85
	4	172.76	99.86	89.41	88.30	100.52	89.32	88.70	96.89	89.21	88.62	88.07	95.64	108.92
	5	220.77	96.59	83.11	80.96	93.62	82.43	80.72	90.88	83.05	81.68	81.17	83.72	85.80
VAR(1) ($\Psi = 0.8I_2$)	1	91.35	83.21	81.15	80.83	83.24	81.12	80.80	83.30	81.70	81.43	81.82	85.04	88.18
	2	159.03	101.10	85.76	83.62	96.00	84.34	83.01	96.83	88.27	86.95	83.58	86.63	89.49
	3	208.07	107.37	90.82	92.11	97.49	88.72	91.05	97.79	91.86	93.56	86.74	88.26	92.53
	4	119.13	95.82	84.70	83.56	94.54	84.04	83.27	93.04	85.79	85.17	83.09	87.13	94.13
	5	156.71	99.99	84.89	82.74	94.37	83.35	82.03	97.28	89.08	87.71	82.57	85.54	88.33
Pure ARCH(1) ($\Psi = 0_{2 \times 2}$) ($\alpha = 0.9$)	1	194.99	117.12	109.50	108.43	117.45	109.46	108.41	116.25	109.70	108.91	111.35	114.83	115.60
	2	462.58	226.51	196.53	191.99	231.36	196.19	191.08	217.07	196.85	192.75	191.67	196.66	204.55
	3	447.26	243.42	215.17	222.60	248.27	214.22	219.54	232.50	214.55	220.94	209.80	216.15	227.78
	4	329.67	154.53	139.48	143.63	156.28	139.21	142.43	149.41	139.07	146.26	143.54	157.11	180.24
	5	296.79	133.31	116.45	112.84	133.18	115.50	111.98	129.17	116.90	116.14	112.67	115.79	116.53
VAR-ARCH(1, 1) ($\Psi = 0.8I_2$) ($\alpha = 0.9$)	1	121.73	119.12	113.80	113.30	119.33	113.74	113.23	119.67	115.55	115.17	115.15	119.17	120.77
	2	184.59	138.64	121.07	117.58	136.07	119.28	116.62	143.47	132.05	130.40	117.62	120.97	122.43
	3	230.14	145.21	125.13	124.09	137.49	122.52	126.62	140.33	131.42	132.15	120.18	123.00	126.11
	4	146.33	132.69	118.70	116.32	135.01	117.93	115.87	132.44	121.18	119.76	116.54	120.96	125.66
	5	180.77	137.25	119.51	116.14	135.12	117.74	115.21	181.46	150.43	147.64	116.21	119.38	120.69

error (RMSE) of the estimators, multiplied by 10^4 (hence the entry in the Table of 82.27 for the estimator FDA1 in scheme 1 under white noise, for example, is to be interpreted as an actual RMSE of 0.008227). Some general features emerge from Table 1. First, the RMSEs of the spectral estimators tend to be decreasing in m with the exception of scheme 3 where the RMSE is U-shaped in m ; this contrasts with the FM estimators in which the RMSE is increasing in m . Scheme 3, in which x_{2t} is the averaged variable, also has the highest RMSE values for each estimator across the five sampling schemes, while the RMSEs in scheme 5 (in which both x_{1t} and x_{2t} are averaged variables) are virtually on a par with the values obtained in the infeasible HFR (scheme 1). The effect of increasing dependence in the VAR – moving from Ψ_0 (white noise) to Ψ_1 to Ψ_2 – leads to small increases in RMSE in schemes 1 and 5 and reductions in RMSE in schemes 2–4. The addition of ARCH effects tends to increase the RMSEs across all five sampling schemes. Overall the results in Table 1 indicate that the spectral estimators typically perform at least as well as, and sometimes better than, the FM estimators in the majority of cases, subject to m being chosen large enough. Among the class of spectral estimators the ASD variants tend to have higher RMSE than the FD and FDA versions.¹¹

It is also of interest to investigate the finite sample properties of Wald statistics based on the spectral and FM estimators. To do this we examine the size properties of the tests under the null hypothesis $H_0 : C = 1$ and the power properties under the alternative $H_1 : C \neq 1$ using the four fixed alternatives $C = \{0.95, 0.99, 1.01, 1.05\}$. The results are presented in Table 2 for $h = 1/3$ and the VAR model using Ψ_2 as well as its ARCH variant. In addition to the OLS-based test we report results for the spectral regression estimators using the largest number of periodogram ordinates (FD3, FDA3 and ASD3) as well as the FM estimator based on the smallest number of autocovariances (FM1) – these correspond to the estimators that typically have the lowest RMSE. The entries in Table 2 are percentages and those for power are size-adjusted; the nominal size of the tests is 5 percent.

It can be seen from Table 2 that the OLS-based Wald tests show the greatest size distortions closely followed by the FM-based tests. The largest size distortions for the spectral-based tests occur (perhaps surprisingly) in the infeasible high frequency sampling scheme 1 while the size is much closer to the nominal level in schemes 2–5 (although the ASD3-based tests are notably under-sized). The size-adjusted power is good for the spectral- and FM-based tests across all sampling schemes and dominates that of the OLS-based tests substantially in schemes 2–5. These features, combined with the good performance of the spectral estimators in terms of RMSE, suggest that spectral regression provides a good platform for estimation and inference in cointegrated models with mixed frequency and mixed sample data using the simple averaging of high frequency data.

¹¹ Results obtained when changing the level of aggregation to $h = 1/12$ show reductions in RMSE in the majority of cases when compared to Table 1. This is presumably due to the corresponding increase in the number of high frequency observations that are feeding into the averaged data in the MFR.

Table 2
Simulated size and size-adjusted power of Wald tests, $T = 100$, $h = 1/3$.

C	OLS	FD3	FDA3	ASD3	FM1	C	OLS	FD3	FDA3	ASD3	FM1
VAR(1) ($\Psi = 0.8I_2$)						VAR-ARCH(1, 1) ($\Psi = 0.8I_2, \alpha = 0.9$)					
Sampling scheme 1						Sampling scheme 1					
0.95	96.88	98.86	98.86	98.28	98.54	0.95	94.21	96.04	96.03	94.49	95.98
0.99	15.20	41.52	42.29	37.60	40.71	0.99	16.72	43.64	44.18	39.84	43.90
1.00	46.90	21.76	21.53	1.47	18.43	1.00	46.24	21.54	21.70	1.69	18.56
1.01	39.48	43.43	42.76	41.34	41.88	1.01	39.77	44.71	44.21	42.01	43.74
1.05	96.28	98.43	98.43	98.06	98.18	1.05	94.36	96.15	96.13	94.80	95.91
Sampling scheme 2						Sampling scheme 2					
0.95	77.06	98.51	98.45	97.81	98.31	0.95	76.85	95.75	95.62	94.10	95.62
0.99	0.51	36.89	41.10	30.87	40.20	0.99	1.14	40.52	43.09	33.65	43.20
1.00	17.94	4.93	5.82	1.23	15.87	1.00	17.24	4.55	5.56	1.10	16.21
1.01	19.73	42.49	41.91	39.48	41.03	1.01	19.74	44.14	43.07	40.35	43.22
1.05	72.04	98.16	98.17	97.71	98.05	1.05	72.66	96.00	95.71	94.94	95.76
Sampling scheme 3						Sampling scheme 3					
0.95	57.61	99.23	99.20	98.39	98.69	0.95	60.43	96.75	96.56	94.27	96.07
0.99	0.71	39.77	44.93	36.59	41.66	0.99	1.18	42.59	46.91	38.75	43.95
1.00	18.84	4.06	5.22	2.99	14.14	1.00	17.83	3.73	4.83	2.99	13.80
1.01	15.11	34.21	33.36	31.47	37.39	1.01	15.16	36.33	35.36	32.65	39.31
1.05	59.47	96.35	96.31	95.78	97.34	1.05	60.94	94.20	93.76	92.47	95.00
Sampling scheme 4						Sampling scheme 4					
0.95	90.44	97.96	97.92	97.19	98.34	0.95	88.74	95.55	95.47	93.56	95.93
0.99	1.59	35.42	37.98	30.23	40.57	0.99	2.54	38.91	40.50	33.96	43.59
1.00	19.73	6.20	6.54	1.51	15.99	1.00	19.54	5.62	6.02	1.62	15.59
1.01	26.65	44.84	44.10	42.29	41.97	1.01	26.68	46.76	45.80	43.34	43.62
1.05	86.23	98.62	98.65	98.32	98.28	1.05	85.23	96.38	96.35	95.16	96.00
Sampling scheme 5						Sampling scheme 5					
0.95	74.90	98.64	98.62	97.63	98.41	0.95	75.23	96.11	95.91	94.08	95.79
0.99	0.46	36.91	41.10	29.73	40.48	0.99	1.10	40.97	43.68	32.84	43.75
1.00	19.76	5.64	6.85	0.72	17.09	1.00	18.91	5.24	6.36	0.75	17.35
1.01	19.32	42.82	41.98	39.34	41.56	1.01	19.70	44.69	43.66	40.27	43.55
1.05	71.91	98.24	98.21	97.77	98.11	1.05	72.99	96.13	95.91	95.15	95.82

5. Concluding comments

This paper has proposed a simple method capable of exploiting fully the information contained in mixed frequency and mixed sample data in the estimation of cointegrating vectors. The properties of easy-to-compute spectral regression estimators of the cointegrating parameters have been derived, these being in the form of theoretical asymptotic properties as well as simulated finite sample properties. The proposed estimators belong to the class of optimal cointegration estimators defined by Phillips (1991c) and Wald statistics based on these estimators have asymptotic chi-square distributions. The finite sample performance of spectral regression estimators have good RMSE properties which are at least as small as those of a time domain fully modified estimator. The size and size-adjusted power properties of associated Wald statistics are also good.

The model analysed contains no deterministic components but it is a straightforward matter to deal with an intercept and time trend, for example. Demeaning and detrending the data by regression methods prior to the application of the frequency domain regression to estimate the cointegration parameters is valid in this framework and the limiting distributions are defined in terms of demeaned and detrended Brownian motion processes. Such an approach is valid because the cointegration parameters are assumed to be fixed, thereby avoiding the problems highlighted by Corbae et al. (2002) in band limited spectral regression in models in which the parameters are frequency-dependent. Alternatively the intercept and trend could also be estimated as part of the spectral regression procedure albeit at the cost of an increase in the dimension of the parameter vector.

Acknowledgements

I thank the Co-Editor, Oliver Linton, an anonymous Associate Editor and three anonymous referees for their helpful feedback on the original version of this paper. I am also grateful to Zack Miller for his insightful comments, as well as participants at the 10th York Econometrics Symposium, University of York, 15–16 June, 2017, and attendees at the session on ‘Co-integration, trend breaks, and mixed frequency data’ at the 11th International Conference on Computational and Financial Econometrics, Senate House, University of London, 16–18 December, 2017, for their helpful comments. Financial support provided by the Economic and Social Research Council, United Kingdom under Grant Number ES/M01147X/1 is gratefully acknowledged.

Appendix. Proofs of lemmas and theorems

Proof of Lemma 1. We begin by noting that we can relate x_{1t} and x_{2t} to y_{1t} and y_{2t} , respectively, using the matrix filter relationships

$$x_{1t} = S_1(L^h)y_{1t}, \quad x_{2t} = S_2(L^h)y_{2t}, \quad t = 1, \dots, T,$$

where the matrix filter functions are defined as

$$S_j(z) = \begin{pmatrix} k^{-1}s(z)I_{n_j^{SH}} & 0 & 0 & 0 \\ 0 & I_{n_j^{SL}} & 0 & 0 \\ 0 & 0 & k^{-1}s(z)I_{n_j^{FH}} & 0 \\ 0 & 0 & 0 & k^{-1}s(z)I_{n_j^{FL}} \end{pmatrix}, \quad j = 1, 2.$$

Re-writing (6) in the form $y_{1t} = Cy_{2,t-1} + v_{1t}$, where $v_{1t} = u_{1t} + Cw_{2t}$ (a stationary vector), we can pre-multiply by $S_1(L^h)$ to obtain

$$x_{1t} = S_1(L^h)CS_2(L^h)^{-1}x_{2,t-1} + S_1(L^h)v_{1t}.$$

Inspection of the individual sub-matrices of $S_1(z)CS_2(z)^{-1}$ shows that we can write

$$S_1(z)CS_2(z)^{-1} = C + K(z)$$

where

$$K(z) = \begin{pmatrix} 0 & \phi_1(z)C_{SS}^{HL} & 0 & 0 \\ \phi_2(z)C_{SS}^{LH} & 0 & \phi_2(z)C_{SF}^{LH} & \phi_2(z)C_{SF}^{LL} \\ 0 & \phi_1(z)C_{FS}^{HL} & 0 & 0 \\ 0 & \phi_1(z)C_{FS}^{LL} & 0 & 0 \end{pmatrix},$$

$\phi_1(z) = k^{-1}s(z) - 1$, $\phi_2(z) = ks(z)^{-1} - 1$ and the C_{ij}^{kl} ($i, j = F, S$; $k, l = H, L$) are sub-matrices of C of dimension $n_1^{ik} \times n_2^{jl}$ (for example, C_{SF}^{LH} is $n_1^{SL} \times n_2^{FH}$). Hence

$$\begin{aligned} x_{1t} &= Cx_{2,t-1} + \xi_{1t}, \\ \xi_{1t} &= S_1(L^h)v_{1t} + K(L^h)x_{2,t-1}, \end{aligned}$$

and we need to demonstrate that ξ_{1t} is stationary. From the definitions of $S_1(z)$ and v_{1t} the first component is clearly stationary. In the second, note that $\phi_1(L^h)$ operates on $x_{2,t-1}^{SL} = y_{2,t-1}^{SL}$; we then have

$$\phi_1(L^h)y_{2,t-1}^{SL} = k^{-1}s(L^h)y_{2,t-1}^{SL} - y_{2,t-1}^{SL} = -\delta_{2,t-1}^{SL},$$

which is stationary by Lemma A1. Also, $\phi_2(L^h)$ operates on the remaining (aggregated) elements; for example,

$$\begin{aligned} \phi_2(L^h)y_{2,t-1}^{SH} &= ks(L^h)^{-1}Y_{2,t-1}^{SH} - Y_{2,t-1}^{SH} \\ &= y_{2,t-1}^{SH} - Y_{2,t-1}^{SH} = \delta_{2,t-1}^{SH} \end{aligned}$$

which is also stationary by Lemma A1. Similar results apply to $Y_{2,t-1}^{FH}$ and $Y_{2,t-1}^{FL}$, so that

$$K(L^h)x_{2,t-1} = \begin{pmatrix} -C_{SS}^{HL}\delta_{2,t-1}^{SL} \\ C_{SS}^{LH}\delta_{2,t-1}^{SH} + C_{SF}^{LH}\delta_{2,t-1}^{FH} + C_{SF}^{LL}\delta_{2,t-1}^{FL} \\ -C_{FS}^{HL}\delta_{2,t-1}^{SL} \\ -C_{FS}^{LL}\delta_{2,t-1}^{SL} \end{pmatrix}$$

is a vector of stationary components, implying that ξ_{1t} is also stationary. Finally, (10) is obtained by applying the operator $S_2(L^h)$ to (8) which results in

$$\xi_{2t} = S_2(L^h)w_{2t}$$

which is also clearly stationary. \square

Proof of Lemma 2. From Lemma A2 we can relate the partial sum of interest to that of u_t as follows:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \xi_t = G(L^h)s(L^h) \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t.$$

The task is then to relate the partial sums involving fractions of T to the FCLT in [Assumption 1](#) which deals with the high frequency process and partial sums involving a fraction of N . Following the proof of [Lemma 1](#) of [Miller \(2016\)](#) we can write

$$\begin{aligned} \sum_{\tau=1}^{[Nr]} u_{\tau h} &= \sum_{t=1}^{[Tr]} \sum_{l=0}^{k-1} u_{t-lh} + \sum_{l=[Tr]/h+1}^{[Nr]} u_{lh} \\ &= s(L^h) \sum_{t=1}^{[Tr]} u_t + \sum_{l=[Tr]/h+1}^{[Nr]} u_{lh}. \end{aligned}$$

from which we obtain

$$\begin{aligned} s(L^h) \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t &= \frac{1}{\sqrt{T}} \sum_{\tau=1}^{[Nr]} u_{\tau h} - \frac{1}{\sqrt{T}} \sum_{l=[Tr]/h+1}^{[Nr]} u_{lh} \\ &= \frac{1}{\sqrt{T}} \sum_{\tau=1}^{[Nr]} u_{\tau h} + o_p(1), \end{aligned}$$

the last quantity being asymptotically negligible owing to the summation being over a finite interval and hence will converge to zero, as shown in [Miller \(2016\)](#). Now, the elements of $G(z)$ are polynomials of order no greater than $k - 1$ so we can use [Lemma 2.1](#) of [Phillips and Solo \(1992\)](#) to write

$$G(z) = G(1) - (1 - z)\tilde{G}(z)$$

where the elements of $\tilde{G}(z)$ are polynomials of order no greater than $k - 2$. We can then write, using $T = hN$,

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \xi_t &= \frac{1}{\sqrt{h}} G(1) \frac{1}{\sqrt{N}} \sum_{\tau=1}^{[Nr]} u_{\tau h} - \frac{1}{\sqrt{h}} \tilde{G}(L^h) \frac{1}{\sqrt{N}} \sum_{\tau=1}^{[Nr]} \Delta_h u_{\tau h} + o_p(1) \\ &= \frac{1}{\sqrt{h}} G(1) \frac{1}{\sqrt{N}} \sum_{\tau=1}^{[Nr]} u_{\tau h} + o_p(1) \end{aligned}$$

because

$$\frac{1}{\sqrt{N}} \sum_{\tau=1}^{[Nr]} \Delta_h u_{\tau h} = \frac{1}{\sqrt{N}} (u_{[Nr]h} - u_0) = o_p(1).$$

It follows that, as $T \rightarrow \infty$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \xi_t \xrightarrow{d} B(r)$$

where $B(r) = (1/\sqrt{h})G(1)B_u(r)$ is a Brownian motion process with covariance matrix Ω as defined in the Lemma. \square

Proof of Lemma 3. Let $M_u(z) = \sum_{l=-\infty}^{\infty} \Gamma_{u, lh} z^{lh}$ denote the autocovariance generating function (AGF) of $u_{\tau h}$, where $\Gamma_{u, lh} = E(u_{\tau h} u'_{\tau h-lh})$. Then, from [Hamilton \(1994, p.268\)](#), the AGF of ξ_t , measured in high frequency time units, is given by

$$M^H(z) = G(z^h) s(z^h) M_u(z) s(z^{-h}) G(z^{-h})' = \sum_{l=-\infty}^{\infty} \Gamma_{\xi, lh} z^{lh},$$

where $\Gamma_{\xi, lh} = E(\xi_t \xi'_{t-lh})$ is the high frequency autocovariance matrix at lag lh . To convert this to the low frequency time units we take integer values (setting $m = lh$) to give

$$M(z) = \sum_{m=-\infty}^{\infty} \Gamma_{\xi, m} z^m.$$

In case where the limits in $M^H(z)$ are finite, such as for a finite-order moving average, an appropriate adjustment needs to be made to the limits in $M(z)$ i.e. if $M^H(z) = \sum_{l=-K}^K \Gamma_{\xi, lh} z^{lh}$ then $M(z) = \sum_{m=-[Kh]}^{[Kh]} \Gamma_{\xi, m} z^m$. The aim is to first relate $\Gamma_{\xi, lh}$ to $\Gamma_{u, lh}$. The product $s(z)s(z^{-1})$ is a two-sided scalar polynomial of order $k - 1$:

$$s(z)s(z^{-1}) = \sum_{l=0}^{k-1} z^l \sum_{m=0}^{k-1} z^{-m} = \sum_{l=-(k-1)}^{k-1} s_1 z^l$$

where the s_{lh} coefficients are implicitly defined. Next, let

$$\Gamma_{u,lh} = \begin{pmatrix} \Gamma_{u,lh}^{11} & \Gamma_{u,lh}^{12} \\ \Gamma_{u,lh}^{21} & \Gamma_{u,lh}^{22} \end{pmatrix}.$$

Then, from the form of $G(z)$ in Lemma A2, we find that

$$G(z)M_u(z)G(z^{-1})' = h^2 \sum_{l=-\infty}^{\infty} \begin{pmatrix} C_{lh}^{11} & C_{lh}^{12} \\ C_{lh}^{21} & C_{lh}^{22} \end{pmatrix},$$

where

$$\begin{aligned} C_{lh}^{11} &= \Gamma_{u,lh}^{11} + s(z)C\Gamma_{u,lh}^{21} + s(z^{-1})\Gamma_{u,lh}^{12}C' + s(z)s(z^{-1})C\Gamma_{u,lh}^{22}C', \\ C_{lh}^{12} &= s(z^{-1})\Gamma_{u,lh}^{12} + s(z)s(z^{-1})C\Gamma_{u,lh}^{22}, \\ C_{lh}^{21} &= s(z)\Gamma_{u,lh}^{21} + s(z)s(z^{-1})\Gamma_{u,lh}^{22}C', \\ C_{lh}^{22} &= s(z)s(z^{-1})\Gamma_{u,lh}^{22}. \end{aligned}$$

When multiplied by $s(z)s(z^{-1})$ these matrices will have additional terms involving

$$\begin{aligned} s(z)^2s(z^{-1}) &= \sum_{m=0}^{k-1} z^m \sum_{l=-(k-1)}^{k-1} s_{1l}z^{-l} = \sum_{l=-(k-1)}^{2k-2} s_{2l}z^l, \\ s(z)s(z^{-1})^2 &= \sum_{l=-(k-1)}^{k-1} s_{1l}z^l \sum_{m=0}^{k-1} z^{-m} = \sum_{l=-(2k-2)}^{k-1} s_{3l}z^l, \\ s(z)^2s(z^{-1})^2 &= \sum_{l=-(k-1)}^{k-1} s_{1l}z^l \sum_{m=-(k-1)}^{k-1} s_{1l}z^{-m} = \sum_{l=-(2k-2)}^{2k-2} s_{4l}z^l, \end{aligned}$$

where the coefficients are again implicitly defined. Hence each summand of interest, $\Gamma_{u,lh}$, is multiplied by a finite-order scalar polynomial in z^h of order $2k - 2$ at most. We therefore need to consider quantities of the form (with $p = 2k - 2$)

$$\sum_{m=-p}^p a_m z^{mh} \sum_{l=-\infty}^{\infty} \Gamma_{u,lh} z^{lh} = \sum_{l=-\infty}^{\infty} \left(\sum_{m=-p}^p a_m \Gamma_{u,lh-mh} \right) z^{lh},$$

which implies that $\Gamma_{\xi,lh} = \sum_{m=-p}^p a_m \Gamma_{u,lh-mh}$. Taking integer values of lh we obtain

$$\begin{aligned} \sum_{l=-T}^T |l| \|\Gamma_{\xi,l}\| &= \sum_{l=-T}^T |l| \left\| \sum_{m=-p}^p a_m \Gamma_{u,lh-mh} \right\| \\ &\leq \sum_{m=-p}^p |a_m| \sum_{l=-T}^T |l| \|\Gamma_{u,lh-mh}\|. \end{aligned}$$

But p is finite and independent of T , so that $\sum_{m=-p}^p |a_m| = O(1)$, while a sufficient condition¹² for $\sum_{l=-T}^T |l| \|\Gamma_{u,lh-mh}\| = O(T^{1/2})$ is that $\sum_{j=0}^{\infty} j^{1/2} \|A_j\| < \infty$, which is implied by Assumption 1(a). Hence $\sum_{l=-T}^T |l| \|\Gamma_{\xi,l}\| = O(T^{1/2})$ as required. \square

Proof of Theorem 1. (a) We begin by noting that we can write

$$\begin{aligned} \hat{f}_{22}(0) &= \frac{1}{2m+1} \sum_{j=-m}^m I_{22}(\lambda_j) \\ &= \frac{1}{2m+1} \sum_{j=-m}^m \left(\frac{1}{2\pi} \sum_{k=-T+1}^{T-1} \hat{f}_{22,k} e^{-ik\lambda_j} \right) \\ &= \frac{1}{2\pi(2m+1)} \sum_{k=-T+1}^{T-1} \hat{f}_{22,k} w_k \end{aligned}$$

¹² See, for example, Fuller (1996, p.367) for a demonstration of this result.

where $w_k = \sum_{j=-m}^m e^{-ik\lambda_j}$ and

$$\hat{\Gamma}_{22,k} = \begin{cases} \frac{1}{T} \sum_{t=k+2}^T x_{2,t-1} x'_{2,t-1-k}, & k \geq 0, \\ \frac{1}{T} \sum_{t=2}^{T+k} x_{2,t-1} x'_{2,t-1-k}, & k < 0. \end{cases}$$

We are then led to consider

$$\begin{aligned} \frac{2m+1}{T^2} \hat{f}_{22}(0) &= \frac{1}{2\pi T} \sum_{k=-T+1}^{T-1} \left(\frac{1}{T} \hat{\Gamma}_{22,k} \right) w_k \\ &\xrightarrow{d} \frac{1}{2\pi} \int_0^1 B_2 B_2' \times \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=-T+1}^{T-1} w_k \\ &= \frac{1}{\pi} \int_0^1 B_2 B_2' \end{aligned}$$

using Lemma A3(a) and as the limit involving the sum of w_k is equal to 2; see Lemma A4(a).

(b) Proceeding in a similar way as in part (a) we find that

$$\begin{aligned} \hat{f}_{\xi 2}(0) &= \frac{1}{2m+1} \sum_{j=-m}^m I_{\xi 2}(\lambda_j) \\ &= \frac{1}{2\pi(2m+1)} \sum_{k=-T+1}^{T-1} \hat{\Gamma}_{\xi 2,k} w_k \end{aligned}$$

where w_k is as previously defined and

$$\hat{\Gamma}_{\xi 2,k} = \begin{cases} \frac{1}{T} \sum_{t=k+2}^T \xi_t x'_{2,t-1-k}, & k \geq 0, \\ \frac{1}{T} \sum_{t=2}^{T+k} \xi_t x'_{2,t-1-k}, & k < 0. \end{cases}$$

We are then led to consider

$$\begin{aligned} \frac{2m+1}{T} \hat{f}_{\xi 2}(0) &= \frac{1}{2\pi T} \sum_{k=-T+1}^{T-1} \left(\frac{1}{T} \hat{\Gamma}_{\xi 2,k} \right) w_k \\ &\xrightarrow{d} \frac{1}{2\pi} \int_0^1 dB B_2' \times \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=-T+1}^{T-1} w_k + \lim_{T \rightarrow \infty} \frac{1}{2\pi T} \sum_{k=-T+1}^{T-1} S_{2,k+1} w_k \\ &= \frac{1}{\pi} \int_0^1 dB B_2' + \lim_{T \rightarrow \infty} \frac{1}{2\pi T} \sum_{k=-T+1}^{T-1} S_{2,k+1} w_k \end{aligned}$$

using Lemma A3(b) and Lemma A4(a) and where $S_{2,k} = \sum_{l=k}^{\infty} \Gamma_{\xi 2,l}$. Using summation-by-parts the second term can be written

$$\begin{aligned} \frac{1}{T} \sum_{k=-T+1}^{T-1} S_{2,k+1} w_k &= S_{2,T} \left(\frac{1}{T} \sum_{k=-T+1}^{T-1} w_k \right) + \sum_{k=-T+1}^{T-2} \left(\frac{1}{T} \sum_{l=-T+1}^k w_l \right) (S_{2,k+1} - S_{2,k+2}) \\ &= S_{2,T} \left(\frac{1}{T} \sum_{k=-T+1}^{T-1} w_k \right) + \sum_{k=-T+1}^{T-2} \left(\frac{1}{T} \sum_{l=-T+1}^k w_l \right) \Gamma_{\xi 2,k+1} \end{aligned}$$

because $S_{2,k+1} - S_{2,k+2} = \Gamma_{\xi 2,k+1}$. Now $S_{2,T} \rightarrow 0$ as $T \rightarrow \infty$ while, from Lemma A4(a),

$$\frac{1}{T} \sum_{k=-T+1}^{T-1} w_k \rightarrow 2,$$

hence the first term converges to zero. As for the second term we have, from Lemma A4(b),

$$\frac{1}{T} \sum_{l=-T+1}^k w_l = 1 + O\left(\frac{m}{T}\right)$$

for all k , and so we deduce that, under Assumption 2,

$$\lim_{T \rightarrow \infty} \frac{1}{2\pi T} \sum_{k=-T+1}^{T-1} S_{2,k+1} w_k = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \Gamma_{\xi 2,k}$$

as required.

(c) We begin by using the decomposition

$$\hat{f}_{\xi\xi}(0) - f_{\xi\xi}(0) = \left(\hat{f}_{\xi\xi}(0) - \hat{f}_{\xi\xi}(0)\right) + \left(\hat{f}_{\xi\xi}(0) - f_{\xi\xi}(0)\right)$$

and then proceed to show that each of the two terms in parentheses is $o_p(1)$. Note that

$$\hat{f}_{\xi\xi}(0) = \frac{1}{2m+1} \sum_{j=-m}^m w_{\xi}(\lambda_j) w_{\xi}(\lambda_j)^*$$

and that $\hat{\xi}_t = Y_{0t} - J\hat{C}Y_{2,t-1}$ where \hat{C} is an initial estimator of C such that $T(\hat{C} - C) = O_p(1)$. Substituting for Y_{0t} we obtain $\hat{\xi}_t = \xi_t - J(\hat{C} - C)Y_{2,t-1}$ which implies that

$$w_{\hat{\xi}}(\lambda_j) = w_{\xi}(\lambda_j) - J(\hat{C} - C)w_2(\lambda_j).$$

It then follows that

$$\begin{aligned} I_{\hat{\xi}\hat{\xi}}(\lambda_j) &= \left(w_{\hat{\xi}}(\lambda_j) - J(\hat{C} - C)w_2(\lambda_j)\right) \left(w_{\hat{\xi}}(\lambda_j) - J(\hat{C} - C)w_2(\lambda_j)\right)^* \\ &= I_{\xi\xi}(\lambda_j) + J(\hat{C} - C)I_{22}(\lambda_j)(\hat{C} - C)'J' - J(\hat{C} - C)I_{2\xi}(\lambda_j) - I_{\xi 2}(\lambda_j)(\hat{C} - C)'J' \end{aligned}$$

and so the quantity of interest is

$$\begin{aligned} \hat{f}_{\hat{\xi}\hat{\xi}}(0) - \hat{f}_{\xi\xi}(0) &= \frac{1}{2m+1} \sum_{j=-m}^m \left(I_{\hat{\xi}\hat{\xi}}(\lambda_j) - I_{\xi\xi}(\lambda_j)\right) \\ &= J(\hat{C} - C)\hat{f}_{22}(0)(\hat{C} - C)'J' - J(\hat{C} - C)\hat{f}_{2\xi}(0) - \hat{f}_{\xi 2}(0)(\hat{C} - C)'J'. \end{aligned}$$

Using the stochastic orders of magnitude already established we obtain

$$\begin{aligned} (2m+1) \left(\hat{f}_{\hat{\xi}\hat{\xi}}(0) - \hat{f}_{\xi\xi}(0)\right) &= JT(\hat{C} - C)\frac{2m+1}{T^2}\hat{f}_{22}(0)T(\hat{C} - C)'J' \\ &\quad - JT(\hat{C} - C)\frac{2m+1}{T}\hat{f}_{2\xi}(0) - \frac{2m+1}{T}\hat{f}_{\xi 2}(0)T(\hat{C} - C)'J' \\ &= O_p(1) \end{aligned}$$

and so $\hat{f}_{\hat{\xi}\hat{\xi}}(0) - \hat{f}_{\xi\xi}(0) = O_p(1/m) = o_p(1)$ under Assumption 2. The second term of interest is $o_p(1)$ because Lemma 3 controls the rate of growth of the autocovariances of ξ_t . This second term is a consistency result for the infeasible estimator based on the unobservable ξ_t and follows, for example, from results in Hannan (1970) and Fuller (1996). \square

Proof of Theorem 2. From Theorem 1(c) we can replace $\hat{f}_{\hat{\xi}\hat{\xi}}(0)$ with $f_{\xi\xi}(0)$ and so

$$\begin{aligned} T(\hat{C}_0 - C) &= (J'f_{\xi\xi}(0)^{-1}J)^{-1}J'f_{\xi\xi}(0)^{-1} \left(\frac{2m+1}{T}\hat{f}_{\xi 2}(0)\right) \left(\frac{2m+1}{T^2}\hat{f}_{22}(0)\right)^{-1} + o_p(1) \\ &\stackrel{d}{\rightarrow} (J'f_{\xi\xi}(0)^{-1}J)^{-1}J'f_{\xi\xi}(0)^{-1} \left(\frac{1}{\pi} \int_0^1 dB B_2' + \frac{1}{2\pi} \Omega_2\right) \left(\frac{1}{\pi} \int_0^1 B_2 B_2'\right)^{-1} \\ &= (J'f_{\xi\xi}(0)^{-1}J)^{-1}J'f_{\xi\xi}(0)^{-1} \int_0^1 dB B_2' \left(\int_0^1 B_2 B_2'\right)^{-1} \\ &\quad + \frac{1}{2} (J'f_{\xi\xi}(0)^{-1}J)^{-1}J'f_{\xi\xi}(0)^{-1} \Omega_2 \left(\int_0^1 B_2 B_2'\right)^{-1}. \end{aligned}$$

Using the definitions $2\pi f_{\xi\xi}(0) = \Omega$ and $\Omega_{11.2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$ it can be shown that

$$J'f_{\xi\xi}(0)^{-1}J = 2\pi \Omega_{11.2}^{-1} \text{ and } J'f_{\xi\xi}(0)^{-1} = 2\pi \left(\Omega_{11.2}^{-1} : -\Omega_{11.2}^{-1}\Omega_{12}\Omega_{22}^{-1}\right),$$

results which imply that

$$(J'f_{\xi\xi}(0)^{-1}J)^{-1}J'f_{\xi\xi}(0)^{-1} = (I_{n_1} : -\Omega_{12}\Omega_{22}^{-1}).$$

Hence the first term in the limiting distribution can be written

$$\int_0^1 dB_{1.2}B_2' \left(\int_0^1 B_2B_2' \right)^{-1}$$

where $B_{1.2} = (I_{n_1} : -\Omega_{12}\Omega_{22}^{-1})B = B_1 - \Omega_{12}\Omega_{22}^{-1}B_2$. For the second term, note that

$$\Omega_2 = \Omega \begin{pmatrix} 0_{n_1 \times n_2} \\ I_{n_2} \end{pmatrix}$$

and so it follows that

$$J'\Omega^{-1}\Omega_2 = (I_{n_1} : 0_{n_1 \times n_2})\Omega^{-1}\Omega \begin{pmatrix} 0_{n_1 \times n_2} \\ I_{n_2} \end{pmatrix} = 0.$$

Hence the second term is zero and the limiting distribution is defined as in the Theorem.

The analysis of \hat{C}_0^A proceeds in a similar fashion. Using the arguments in the second part of the proof of Theorem 1(c), Lemma 3 ensures that

$$\hat{f}_{1.2,\Delta_2}(0) = f_{1.2,\Delta_2}(0) + o_p(1) \text{ and } \hat{f}_{\Delta_2\Delta_2}(0) = f_{\Delta_2\Delta_2}(0) + o_p(1).$$

Then, based on Theorem 1(c),

$$\begin{aligned} & \frac{2m+1}{T} \hat{f}_{1.2,2}(0) - \hat{f}_{1.2,\Delta_2}(0) \hat{f}_{\Delta_2\Delta_2}(0)^{-1} \left(\frac{2m+1}{T} \hat{f}_{\Delta_22}(0) \right) \\ & \xrightarrow{d} \frac{1}{\pi} \int_0^1 dB_{1.2}B_2' + \frac{1}{2\pi} S_3 - f_{1.2,\Delta_2}(0) f_{\Delta_2\Delta_2}(0)^{-1} \left(\frac{1}{\pi} \int_0^1 dB_2B_2' + \frac{1}{2\pi} S_4 \right) \end{aligned}$$

where $S_3 = \sum_{j=-\infty}^{\infty} E(\xi_{1.2,t+j}\xi'_{2t}) = 2\pi f_{1.2,2}(0)$ and $S_4 = \sum_{j=-\infty}^{\infty} E(\xi_{2,t+j}\xi'_{2t}) = 2\pi f_{22}(0)$. But S_3 is just the covariance between $B_{1.2}$ and B_2 , which is zero, because

$$E(B_{1.2}B_2') = E(B_1 - \Omega_{12}\Omega_{22}^{-1}B_2)B_2' = \Omega_{12} - \Omega_{12}\Omega_{22}^{-1}\Omega_{22} = 0.$$

Hence $f_{1.2,2}(0) = 0$ and $f_{1.2,\Delta_2}(0) = 0$ which implies that

$$\frac{2m+1}{T} \hat{f}_{1.2,2}(0) - \hat{f}_{1.2,\Delta_2}(0) \hat{f}_{\Delta_2\Delta_2}(0)^{-1} \left(\frac{2m+1}{T} \hat{f}_{\Delta_22}(0) \right) \xrightarrow{d} \frac{1}{\pi} \int_0^1 dB_{1.2}B_2'.$$

Similarly we find that

$$\begin{aligned} & \frac{2m+1}{T^2} \hat{f}_{22}(0) - \frac{1}{2m+1} \left(\frac{2m+1}{T} \hat{f}_{2\Delta_2}(0) \hat{f}_{\Delta_2\Delta_2}(0)^{-1} \frac{2m+1}{T} \hat{f}_{\Delta_22}(0) \right) \\ & = \frac{2m+1}{T^2} \hat{f}_{22}(0) + o_p(1) \xrightarrow{d} \frac{1}{\pi} \int_0^1 B_2B_2' \end{aligned}$$

using Theorem 1(a). Combining these limits yields the stated distribution for \hat{C}_0^A . \square

Proof of Theorem 3. We begin with W_0 and note that the limiting distribution of $\hat{\gamma}_0$ has the representation

$$T(\hat{\gamma}_0 - \gamma) \xrightarrow{d} \left[\left(\int_0^1 B_2B_2' \right)^{-1} \otimes I_{n_1} \right] \int_0^1 (B_2 \otimes dB_{1.2}).$$

Let $M_{22} = \int_0^1 B_2B_2'$. Then, from the proof of Lemma 5.1 in Park and Phillips (1988),

$$\int_0^1 (B_2 \otimes dB_{1.2}) \Big|_{B_2} \sim N(0, M_{22} \otimes \Omega_{11.2})$$

in view of B_2 and $B_{1.2}$ being independent. It then follows that the limiting distribution of $T(\hat{\gamma}_0 - \gamma)$, conditional on B_2 , is $N(0, M_{22}^{-1} \otimes \Omega_{11.2})$. Now consider

$$r(\hat{\gamma}_0) = r(\gamma) + R(\bar{\gamma})(\hat{\gamma}_0 - \gamma),$$

where the elements of $\bar{\gamma}$ lie on the line segment between $\hat{\gamma}_0$ and γ . Under H_0 , $r(\gamma) = 0$ while the consistency of $\hat{\gamma}_0$ ensures that $R(\bar{\gamma}) \xrightarrow{p} R(\gamma) = R_0$. Then it follows that

$$Tr(\hat{\gamma}_0) = R(\bar{\gamma})T(\hat{\gamma}_0 - \gamma) \xrightarrow{d} R_0 \left[\left(\int_0^1 B_2 B_2' \right)^{-1} \otimes I_{n_1} \right] \int_0^1 (B_2 \otimes dB_{1,2}) R_0'$$

This limiting distribution, conditional on B_2 , is $N(0, R_0 Q R_0')$ where $Q = M_{22}^{-1} \otimes \Omega_{11,2}$. Theorem 1 implies that

$$\frac{2m+1}{2T^2} \hat{V}_0 \xrightarrow{d} Q^{-1}$$

and hence we are led to consider

$$W_0 = Tr(\hat{\gamma}_0)' \left[R(\hat{\gamma}_0) \left(\frac{2m+1}{2T^2} \hat{V}_0 \right)^{-1} R(\hat{\gamma}_0)' \right]^{-1} Tr(\hat{\gamma}_0).$$

The limiting distribution of this quantity, conditional on B_2 , involves a quadratic form in $N(0, R_0 Q R_0')$ random variables weighted by the matrix $(R_0 Q R_0')^{-1}$, and hence is χ_q^2 . But because this does not depend on B_2 it is also the unconditional distribution. Similar arguments apply to W_0^A . □

Supplementary Lemmas

The following Lemma is used in the proof of Lemma 1. It is more general than is actually required in the proof of Lemma 1 (which uses the result for $j = 0$) but the additional cost of showing that it holds for $j = 1, \dots, k - 1$ i.e. at any point in the interval over which the aggregation takes place, is minimal.

Lemma A1. Let $\Delta_h y_{\tau h}$ ($\tau = 1, \dots, N$) be an $I(0)$ process, where $0 < h < 1$ denotes the sampling interval, and let $k = h^{-1}$ be an integer. Then, for $0 \leq j \leq k - 1$ and $t = 1, \dots, T$,

$$\delta_{t-jh} = y_{t-jh} - \frac{1}{k} \sum_{l=0}^{k-1} y_{t-lh}$$

is an $I(0)$ process.

Proof of Lemma A1. We first write $\delta_{t-jh} = s_{\delta,j}(L^h)y_t$ where $s_{\delta,j}(z) = z^j - k^{-1}s(z)$ and $s(z)$ is defined following (6). The spectral density matrix of δ_{t-jh} is then given by

$$f_{\delta,j}(\lambda) = |s_{\delta,j}(e^{ih\lambda})|^2 f_y(\lambda), \quad -\frac{\pi}{h} < \lambda \leq \frac{\pi}{h}.$$

where $f_y(\lambda)$ is the pseudo-spectrum of y_t satisfying $f_y(\lambda) = O(\lambda^{-2})$ as $\lambda \rightarrow 0$. Now

$$|s_{\delta,j}(e^{ih\lambda})|^2 = \left| e^{ijh\lambda} - \frac{1}{k} \sum_{l=0}^{k-1} e^{ilh\lambda} \right|^2 = \left| ijh\lambda - \frac{1}{k} \sum_{l=0}^{k-1} ilh\lambda + O(\lambda^2) \right|^2,$$

the leading term of which is

$$ijh\lambda - \frac{1}{k} \sum_{l=0}^{k-1} ilh\lambda = ih\lambda \left(j - \frac{1}{k} \sum_{l=0}^{k-1} l \right) = ih\lambda \left(j - \frac{(k-1)}{2} \right).$$

It follows that

$$|s_{\delta,j}(e^{ih\lambda})|^2 = h^2 \lambda^2 \left(j - \frac{(k-1)}{2} \right)^2 + O(\lambda^4)$$

and so $f_{\delta,j}(\lambda)$ is positive and bounded for $\lambda \neq 0$ while $f_{\delta,j}(0) = Ch^2(j - (k - 1)/2)^2 > 0$ (where we have taken $f_y(\lambda) \sim C\lambda^{-2}$ as $\lambda \rightarrow 0$). □

Lemma A2. The disturbances in the MFR, ξ_t , are related to those in the HFR, $u_{\tau h}$, by

$$\xi_t = G(L^h)s(L^h)u_t, \quad t = 1, \dots, T, \tag{25}$$

where $s(z)$ is defined in (7) and

$$G(z) = h \begin{pmatrix} I_{n_1} & Cs(z) \\ 0_{n_2 \times n_1} & s(z)I_{n_2} \end{pmatrix}.$$

Proof of Lemma A2. From the proof of Lemma 1 we find that $\xi_{1t} = k^{-1}s(L^h)v_{1t}$ and $\xi_{2t} = k^{-1}s(L^h)w_{2t}$. Furthermore, from the proof of Lemma 1, $v_{1t} = u_{1t} + Cw_{2t}$ and from (8), $w_{2t} = s(L^h)u_{2t}$. Combining this information yields $\xi_{1t} = k^{-1}s(L^h)u_{1t} + k^{-1}Cs(L^h)^2u_{2t}$ and $\xi_{2t} = k^{-1}s(L^h)^2u_{2t}$ which results in (25). \square

Lemma A3. Under Assumption 1, as $T \rightarrow \infty$:

$$(a) \frac{1}{T^2} \sum_{t=1}^T x_{2,t-1}x'_{2,t-1-k} \xrightarrow{d} \int_0^1 B_2B'_2;$$

$$(b) \frac{1}{T} \sum_{t=1}^T \xi_t x'_{2,t-1-k} \xrightarrow{d} \int_0^1 dB B'_2 + S_{2,k+1}, \text{ where } S_{2,k} = \sum_{j=k}^{\infty} E(\xi_{t+j}\xi'_{2t}).$$

Proof of Lemma A3. The proofs are standard and follow from the FCLT for ξ_t in Lemma 2; see, for example, Phillips and Durlauf (1986) for details. \square

Lemma A4. Let $w_k = \sum_{j=-m}^m e^{-ik\lambda_j}$, where $\lambda_j = 2\pi j/T$ ($j = -m, \dots, m$). Then, under Assumption 2:

$$(a) \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=-T+1}^{T-1} w_k = 2;$$

$$(b) \frac{1}{T} \sum_{l=-T+1}^k w_l = 1 + O\left(\frac{m}{T}\right).$$

Proof of Lemma A4. (a) We first note that $\sum_{j=-m}^m e^{-ijx} = 1 + 2 \sum_{j=1}^m \cos jx$ and so

$$\begin{aligned} \frac{1}{T} \sum_{k=-T+1}^{T-1} w_k &= \frac{1}{T} \sum_{k=-T+1}^{T-1} \left(1 + 2 \sum_{j=1}^m \cos \frac{2\pi kj}{T} \right) \\ &= \frac{2T-1}{T} + \frac{2}{T} \sum_{j=1}^m \sum_{k=-T+1}^{T-1} \cos \frac{2\pi kj}{T}. \end{aligned}$$

The sum over the index k can be decomposed as

$$\begin{aligned} \sum_{k=-T+1}^{T-1} \cos \frac{2\pi kj}{T} &= \sum_{k=1}^T \cos \frac{2\pi kj}{T} - \cos 2\pi j + \cos 0 + \sum_{k=-T}^{-1} \cos \frac{2\pi kj}{T} - \cos(-2\pi j) \\ &= 2 \sum_{k=1}^T \cos \frac{2\pi kj}{T} - 2 \cos 2\pi j + 1 \quad (\text{as } \cos(-x) = \cos x \text{ and } \cos 0 = 1) \\ &= -1 \end{aligned}$$

because $\sum_{k=1}^T \cos 2\pi kj/T = 0$ and $\cos 2\pi j = 1$ for all integer j . Hence

$$\frac{1}{T} \sum_{k=-T+1}^{T-1} w_k = 2 - \frac{1}{T} - \frac{2m}{T} \rightarrow 2$$

as $T \rightarrow \infty$ under Assumption 2.

(b) Note that

$$\frac{1}{T} \sum_{l=-T+1}^k w_l = \begin{cases} \frac{1}{T} \sum_{l=0}^k w_l + \frac{1}{T} \sum_{l=-T+1}^{-1} w_l, & k > 0, \\ \frac{1}{T} w_0 + \frac{1}{T} \sum_{l=-T+1}^{-1} w_l, & k = 0, \\ \frac{1}{T} \sum_{l=-T+1}^{-1} w_l - \frac{1}{T} \sum_{l=k+1}^{-1} w_l, & k < 0. \end{cases}$$

All cases involve the common component

$$\begin{aligned} \frac{1}{T} \sum_{l=-T+1}^{-1} w_l &= \frac{1}{T} \sum_{l=-T+1}^{-1} \left(1 + 2 \sum_{j=1}^m \cos \frac{2\pi lj}{T} \right) \\ &= \frac{T-1}{T} + \frac{2}{T} \sum_{l=1}^{T-1} \sum_{j=1}^m \cos \frac{2\pi lj}{T} \end{aligned}$$

using $\cos(-x) = \cos x$. But

$$\sum_{l=1}^{T-1} \cos \frac{2\pi lj}{T} = \sum_{l=1}^T \cos \frac{2\pi lj}{T} - \cos 2\pi j = -1$$

for the reasons used in part (a), and so

$$\frac{1}{T} \sum_{l=-T+1}^{-1} w_l = 1 - \frac{1}{T} - \frac{2m}{T} = 1 - \frac{(2m+1)}{T}.$$

We now need to consider the additional terms that depend on the sign of k . For $k > 0$

$$\frac{1}{T} \sum_{l=0}^k w_l = \frac{1}{T} \sum_{l=0}^k \left(1 + 2 \sum_{j=1}^m \cos \frac{2\pi lj}{T} \right) = \frac{k+1}{T} + \frac{2}{T} \sum_{l=0}^k \sum_{j=1}^m \cos \frac{2\pi lj}{T}$$

and so it follows that

$$\begin{aligned} \left| \frac{1}{T} \sum_{l=0}^k w_l \right| &\leq \frac{k+1}{T} + \frac{2}{T} \sum_{l=0}^k \sum_{j=1}^m \left| \cos \frac{2\pi lj}{T} \right| \\ &= \frac{k+1}{T} + \frac{2(k+1)m}{T} \rightarrow 0 \end{aligned}$$

as $T \rightarrow \infty$ under [Assumption 2](#). For $k = 0$ we need to consider the additional term

$$\frac{1}{T} w_0 = \frac{2m+1}{T} \rightarrow 0$$

as $T \rightarrow \infty$ under [Assumption 2](#) while for $k < 0$ the additional term is

$$\frac{1}{T} \sum_{l=k+1}^{-1} w_l = \frac{1}{T} \sum_{l=k+1}^{-1} \left(1 + 2 \sum_{j=1}^m \cos \frac{2\pi lj}{T} \right) = \frac{|k|-1}{T} + \frac{2}{T} \sum_{l=1}^{|k|-1} \sum_{j=1}^m \cos \frac{2\pi lj}{T}$$

from which it follows, as in the $k > 0$ case, that

$$\left| \frac{1}{T} \sum_{l=k+1}^{-1} w_l \right| \leq \frac{(|k|-1)(2m+1)}{T} \rightarrow 0$$

as $T \rightarrow \infty$ under [Assumption 2](#). Hence, for all fixed k ,

$$\frac{1}{T} \sum_{l=-T+1}^k w_l = 1 + O\left(\frac{m}{T}\right)$$

under [Assumption 2](#). \square

References

Brockwell, P.J., Davis, R.A., 1991. *Time Series: Theory and Methods*, second ed. Springer-Verlag, New York.
 Chambers, M.J., 2003. The asymptotic efficiency of cointegration estimators under temporal aggregation. *Econometric Theory* 19, 49–77.
 Chambers, M.J., 2016. The estimation of continuous time models with mixed frequency data. *J. Econometrics* 193, 390–404.
 Chambers, M.J., 2019. The estimation of continuous time cointegrated models with mixed frequency and mixed sample data. *J. Time Series Anal.* 40, 887–913.
 Corbae, D., Ouliaris, S., Phillips, P.C.B., 1994. A reexamination of the consumption function using frequency domain regressions. *Empir. Econom.* 19, 595–609.
 Corbae, D., Ouliaris, S., Phillips, P.C.B., 2002. Band spectral regression with trending data. *Econometrica* 70, 1067–1109.
 Fuller, W.A., 1996. *Introduction to Statistical Time Series*, second ed. Wiley, New York.
 Ghysels, E., 2016. Macroeconomics and the reality of mixed frequency data. *J. Econometrics* 193, 294–314.

- Ghysels, E., Miller, J.I., 2014. On the size distortion from linearly interpolating low-frequency series for cointegration tests. In: Fomby, T.B., Park, J.Y., Chang, Y. (Eds.), In: *Essays in Honor of Peter C.B. Phillips*, *Advances in Econometrics*, vol. 33, Emerald Group Publishing Limited, Bingley pp. 93–122.
- Ghysels, E., Miller, J.I., 2015. Testing for cointegration with temporally aggregated and mixed-frequency time series. *J. Time Series Anal.* 36, 797–816.
- Hamilton, J.D., 1994. *Time Series Analysis*. Princeton University Press, Princeton.
- Hannan, E.J., 1970. *Multiple Time Series*. Wiley, New York.
- Miller, J.I., 2010. Cointegrating regressions with messy regressors and an application to mixed-frequency series. *J. Time Series Anal.* 31, 255–277.
- Miller, J.I., 2014. Mixed-frequency cointegrating regressions with parsimonious distributed lag structures. *J. Financ. Econ.* 12, 584–615.
- Miller, J.I., 2016. Conditionally efficient estimation of long-run relationships using mixed-frequency time series. *Econometric Rev.* 35, 1142–1171.
- Miller, J.I., Wang, X., 2016. Implementing residual-based KPSS tests for cointegration with data subject to temporal aggregation and mixed sampling frequencies. *J. Time Series Anal.* 37, 810–824.
- Park, J.Y., Phillips, P.C.B., 1988. Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* 4, 468–497.
- Phillips, P.C.B., 1991a. Spectral regression for cointegrated time series. In: Barnett, W.A., Powell, J., Tauchen, G. (Eds.), *Nonparametric and Semiparametric Methods in Economics and Statistics*. Cambridge University Press, Cambridge, pp. 413–435.
- Phillips, P.C.B., 1991b. Error correction and long run equilibrium in continuous time. *Econometrica* 59, 967–980.
- Phillips, P.C.B., 1991c. Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.
- Phillips, P.C.B., Durlauf, S., 1986. Multiple time series regression with integrated processes. *Rev. Econom. Stud.* 53, 473–495.
- Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with I(1) processes. *Rev. Econom. Stud.* 57, 99–125.
- Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. *Ann. Statist.* 20, 971–1001.
- Robinson, P.M., 1972. Non-linear regression for multiple time-series. *J. Appl. Probab.* 9, 758–768.
- Seong, B., Ahn, S.K., Zadrozny, P.A., 2013. Estimation of vector error correction models with mixed-frequency data. *J. Time Series Anal.* 34, 194–205.
- Shiller, R.J., Perron, P., 1985. Testing the random walk hypothesis: power versus frequency of observation. *Econom. Lett.* 18, 381–386.
- Stock, J.H., Watson, M.W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.
- Thornton, M.A., 2019. Exact discrete representations of linear continuous time models with mixed frequency data. *J. Time Series Anal.* 40, 951–967.