Frequency Domain Estimation of Cointegrating Vectors with Mixed Frequency and Mixed Sample Data

Marcus J. Chambers
University of Essex

January 2018

Abstract

This paper proposes a model suitable for exploiting fully 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 finite sample performance of a spectral regression estimator in an augmented mixed frequency model is particularly encouraging as it is capable of dramatically reducing the root mean squared error obtained in an entirely low frequency model to the levels comparable to an infeasible high frequency model. The finite sample size and power properties of the Wald statistic are also found to be good. An empirical example, to stock price and dividend data, is provided to demonstrate the methods in practice.

Keywords: mixed frequency data; mixed sample data; cointegration; spectral regression.

J.E.L. classification number: C32.

Acknowledgements: I thank participants at the 10th York Econometrics Symposium, University of York, 15–16 June, 2017, for helpful comments on an earlier version of this piece of work, as well as 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. Financial support provided by the Economic and Social Research Council under grant number ES/M01147X/1 is gratefully acknowledged.

Address for Correspondence: Marcus J. Chambers, Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, England.
Tel: +44 1206 872756; fax: +44 1206 872724; e-mail: mchamb@essex.ac.uk.

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 – often termed the 'Johansen' approach following its development by Johansen (1991) – 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 VAR, which is often heavily parameterised, but may still be interested in the cointegrating vectors themselves. In such cases alternative methods are available, including, but certainly not restricted to, dynamic ordinary least squares (Stock and Watson, 1993), fully modified least squares (Phillips and Hansen, 1990), and spectral regression (Phillips, 1991a). These approaches focus on the cointegrating vectors of interest and account for the system dynamics without needing to specify a VAR. The dynamic ordinary least squares approach, for example, adds leads and lags of first-differences to the cointegrating regression; the fully modified 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 frequency. As an example, consider a macroeconomic model that contains an interest rate in addition to macroeconomic aggregates (such as output). The macroeconomic aggregates are typically available quarterly whereas the interest rate can be sampled at much higher frequencies. This means that, say, daily interest data have to be transformed into a representative quarterly figure, and different methods of doing this (such as using the end-of-quarter value, the mid-quarter value, or a quarterly average) may yield different parameter estimates and inferences.

In recent years, however, there has been a growing interest in developing methods that are capable of exploiting all the mixed frequency data that may be available, without the need for 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, Ahn and Zadrozny (2013).

It might be tempting to argue that, because cointegration describes a set of long-run/equilibrium 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. This suggests that there are limitations as to how far the high frequencies should be extended. But, used appropriately, it is possible that the additional information contained in the higher frequency data can be used advantageously 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. This is something that can be explored in appropriately-designed sampling experiments.

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 (smallest sampling interval) is taken as the fundamental frequency; an alternative continuous time approach can be found in Chambers (2017). An advantage of the spectral regression estimators is that all that is required of the model's disturbances is that they are stationary, 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 and low frequency models are, indeed, stationary.

This paper makes four 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. Although we assume, mainly for notational purposes, that the high frequency variables are stocks (which are skip-sampled) and the low frequency variables are flows (in the form of averages), the methods are easily extended to cases where there are also high frequency flows and low frequency stocks. The proposed method of dealing with the mixed frequency data turns out to be very straightforward – simply average the high frequency stock 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 it is not possible to use the high frequency observations separately, as in Foroni, Ghysels and Marcellino (2013), Foroni and Marcellino (2016) and Chambers (2016), for example, due to singularity reasons, the averaging nevertheless does use the information contained in all such 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 a high frequency stock variable and a low frequency flow variable. 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 an infeasible high frequency model (where both variables are sampled at the high frequency), a feasible low frequency model, and the mixed frequency representation. Spectral estimators based on the augmented regression are found to have good finite sample properties and the root mean squared errors obtained in the mixed frequency model are typically much smaller than those from the low frequency representation and, in some case, on a par with those obtained using the infeasible high frequency model. The performance of Wald statistics is also examined, with those obtained using the augmented spectral regression estimator in the mixed frequency model having good size and power properties.

The fourth contribution provides an empirical example to show how the methods work in practice. We follow Ghysels and Miller (2014, 2015) and use the stock price and dividend data provided in Shiller (2000), updated to 2016. Estimates of the parameter in a regression of the logarithm of the stock price (a stock variable) on the logarithm of dividends (a flow) are provided based on different detrending methods, including one based on a structural break in the trend function. In all cases the estimates are significantly different from the theoretical parameter value of unity and this null hypothesis is strongly rejected in all cases. Graphs of the detrended data, as well as the residuals from the cointegrating regression, are also provided.

The paper is organised as follows. Section 2 defines the triangular model of cointegration at the high frequency and provides feasible low and mixed frequency representations, based on the observations. Stationarity of the disturbances in these representations is also demonstrated. Issues concerning frequency domain estimation are addressed in section 3, in which the the estimators and test statistics are defined and their asymptotic properties derived. Section 4 defines the simulation experiments and reports the results obtained, while the empirical example is discussed in section 5. Concluding comments appear in section 6, and all proofs and supplementary results are presented in the Appendix.

The following notation is used throughout the paper. 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), who proposed spectral estimators in cointegrated continuous time systems, 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. 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, $\mathrm{tr}\,(A)$ denotes the trace of a square matrix A, $\|A\| = \sqrt{\mathrm{tr}\,(AA')}$ denotes the Euclidean norm of A, B^* denotes the transpose of the conjugate of a complex-valued matrix B and, for an $n \times m$ matrix C, 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 it is convenient to partition them (without loss of generality) as

$$y_1 = \begin{pmatrix} y_1^S \\ y_1^F \end{pmatrix}, \quad y_2 = \begin{pmatrix} y_2^S \\ y_2^F \end{pmatrix},$$

where y_j^S is $n_j^S \times 1$, y_j^F is $n_j^F \times 1$ and $n_j^S + n_j^F = n_j$ $(j = 1, 2)$.

We assume that the stock variables are sampled at a common high frequency corresponding to a sampling interval of length $0 < h_H = h < 1$ while flows are sampled at a common low frequency 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 observed sequences of observations on stock variables are therefore

$$\left\{y_{1,\tau h}^S\right\}_{\tau=1}^N \quad \text{and} \quad \left\{y_{2,\tau h}^S\right\}_{\tau=1}^N,$$

where N denotes the number of high frequency observations, while the observations on the flow variables are of the form

$$\left\{Y_{1t}^F\right\}_{t=1}^T \quad \text{and} \quad \left\{Y_{2t}^F\right\}_{t=1}^T,$$

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

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

i.e. the flows are averages of the (unobservable) high frequency flows $y_{j,\tau}^F$ over the low frequency observation interval $t - (k-1)h \leq \tau \leq t$.

At the high frequency the triangular cointegrated system is defined by

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

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

where $\Delta_h = 1 - L^h$ denotes the high frequency first-difference operator. The cointegration in the system is depicted by (1), while (2) denotes the n_2 unit roots/stochastic trends. With regard to the disturbance vectors, $u_{1,\tau h}$ and $u_{2,\tau h}$, we make the following assumption.

Assumption 1. The $n \times 1$ vector $u_{\tau h} = (u'_{1,\tau h}, u'_{2,\tau h})'$ is covariance stationary and has a spectral density matrix, $f_{uu}(\lambda)$ $(\pi/h < \lambda \leq \pi/h)$, that is bounded and continuous and for which $f_{uu}(0)$ is positive definite.[1] In addition, as $N \to \infty$,

$$\frac{1}{\sqrt{N}} \sum_{\tau=1}^{[Nr]} u_{\tau h} \xrightarrow{d} B_u(r), \quad 0 < r \leq 1, \tag{3}$$

where $B_u(r)$ is a Brownian motion process with covariance matrix $\Omega_u = 2\pi f_{uu}(0)$.

The covariance stationarity aspect of this assumption is sufficient for the validity of the mixed frequency data representation which requires that the disturbances in the estimating equations are stationary. The functional central limit theorem (FCLT) is used in the derivation of the asymptotic properties of the estimators. We shall refer to (1) and (2) as being the *high frequency representation.*

The triangular system, (1) and (2), also has the error correction model (ECM) representation

$$\Delta_h y_{\tau h} = -JAy_{\tau h-h} + e_{\tau h}, \quad \tau = 1, \ldots, N, \tag{4}$$

where $A = (I_{n_1}, -C)$,

$$J = \begin{pmatrix} I_{n_1} \\ 0_{n_2 \times n_1} \end{pmatrix} \quad \text{and} \quad e_{\tau h} = \begin{pmatrix} e_{1,\tau h} \\ e_{2,\tau h} \end{pmatrix} = \begin{pmatrix} u_{1,\tau h} + Cu_{2,\tau h} \\ u_{2,\tau h} \end{pmatrix}.$$

The problem with this system for the estimation of C is that the flow variables are not observed at the high frequency. However, cointegration is a property that persists at any sampling frequency, and so observations at the low frequency are also cointegrated. Rewriting (1) at the low frequency (essentially setting $t = \tau h$ and picking out the integer values for this index) yields

$$y_{1t} = Cy_{2t} + u_{1t}, \quad t = 1, \ldots, T. \tag{5}$$

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

$$s(z) = 1 + z + \ldots + z^{k-1}; \tag{6}$$

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

$$\Delta y_{2t} = w_{2t}, \quad t = 1, \ldots, T, \tag{7}$$

where $\Delta = 1 - L$ and $w_{2t} = s(L^h)u_{2,\tau h} = \sum_{l=0}^{k-1} u_{2,t-lh}$. Combining (5) and (7) results in the low frequency ECM

$$\Delta y_t = -JAy_{t-1} + v_t, \quad t = 1, \ldots, T, \tag{8}$$

[1] 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 $(-\pi, \pi]$.

where

$$v_t = \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} = \begin{pmatrix} u_{1t} + Cw_{2t} \\ w_{2t} \end{pmatrix}.$$

However, the low frequency ECM in (8) is also not directly amenable to the estimation of C because neither y_{1t}^F nor y_{2t}^F is observable, 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 in the estimation of C. It is convenient to partition the $n_1 \times n_2$ matrix C in the form

$$C = \begin{pmatrix} C_{SS} & C_{SF} \\ C_{FS} & C_{FF} \end{pmatrix},$$

where C_{SS} is $n_1^S \times n_2^S$, C_{SF} is $n_1^S \times n_2^F$, C_{FS} is $n_1^F \times n_2^S$ and C_{FF} is $n_1^F \times n_2^F$. The *mixed frequency representation* is presented in Lemma 1; it also contains a *feasible low frequency representation* in which the observable Y_{1t}^F and Y_{2t}^F replace the unobservable y_{1t}^F and y_{2t}^F in the infeasible low frequency representation of (5) and (7).

Lemma 1. *Let y_1 and y_2 satisfy the high frequency cointegrated system in* (1) *and* (2)*. Then:*
(a) Mixed Frequency Representation. Define the observable aggregated stock variables

$$Y_{1t}^S = \frac{1}{k}\sum_{l=0}^{k-1} y_{1,t-lh}^S, \quad Y_{2t}^S = \frac{1}{k}\sum_{l=0}^{k-1} y_{2,t-lh}^S, \quad t = 1, \ldots, T.$$

Then the mixed frequency observations satisfy, for $t = 2, \ldots, T$,

$$Y_{1t}^S = C_{SS} Y_{2,t-1}^S + C_{SF} Y_{2,t-1}^F + \xi_{1t}^S, \tag{9}$$

$$Y_{1t}^F = C_{FS} Y_{2,t-1}^S + C_{FF} Y_{2,t-1}^F + \xi_{1t}^F, \tag{10}$$

$$\Delta Y_{2t}^S = \xi_{2t}^S, \tag{11}$$

$$\Delta Y_{2t}^F = \xi_{2t}^F, \tag{12}$$

where the disturbance vector $\xi_t = \left(\xi_{1,t}^{S\prime}, \xi_{1t}^{F\prime}, \xi_{2t}^{S\prime}, \xi_{2t}^{F\prime}\right)'$ is I(0) under Assumption 1.
(b) Feasible Low Frequency Representation. The observed low frequency observations satisfy, for $t = 2, \ldots, T$,

$$y_{1t}^S = C_{SS} y_{2,t-1}^S + C_{SF} Y_{2,t-1}^F + \zeta_{1t}^S, \tag{13}$$

$$Y_{1t}^F = C_{FS} y_{2,t-1}^S + C_{FF} Y_{2,t-1}^F + \zeta_{1t}^F, \tag{14}$$

$$\Delta y_{2t}^S = \zeta_{2t}^S, \tag{15}$$

$$\Delta Y_{2t}^F = \zeta_{2t}^F, \tag{16}$$

where the disturbance vector $\zeta_t = \left(\zeta_{1,t}^{S\prime}, \zeta_{1t}^{F\prime}, \zeta_{2t}^{S\prime}, \zeta_{2t}^{F\prime}\right)'$ is I(0) under Assumption 1.

Both representations in Lemma 1 provide a basis for the estimation of the matrix of cointegration vectors, C. The mixed frequency representation is based on the entire sample of mixed frequency observable variables even though the high frequency stocks are not included separately at each high frequency time point but are aggregated to, in effect, mimic the observed flow variables. In fact, it is precisely this form of aggregation of stock variables that is 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). The feasible low frequency representation, on the other hand, skip-samples the high frequency stocks at the low frequency, thereby discarding entirely all the information contained in the observations at the intermediate points.

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 mixed frequency representation 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. Approaches that use the high frequency observations separately have been shown to be possible in some circumstances. For example, Ghysels (2016) deals with a vector autoregressive (VAR) representation for the mixed frequency vector of the form (in our notation)

$$z_t = \left(y_{1t}^{S\prime}, y_{1,t-h}^{S\prime}, \ldots, y_{1,t-(k-1)h}^{S\prime}, Y_{1t}^{F\prime}, y_{2t}^{S\prime}, y_{2,t-h}^{S\prime}, \ldots, y_{2,t-(k-1)h}^{S\prime}, Y_{2t}^{F\prime} \right)^{\prime},$$

thereby including the intermediate high frequency observations on the stocks. A similar approach is followed for a continuous time system by Chambers (2016) but is more parsimonious because the restrictions on the discrete time representation arising from the temporal aggregation are explicitly taken into account. It would also be possible to derive a representation for this vector in the cointegrated system considered here but would result in $kn_1^S + n_1F$ cointegration equations and $kn_2^S + n_2^F$ stochastic trends. The resulting vector of disturbances – an expanded version of ξ_t defined in Lemma 1 – then has a singular spectral density matrix. The reason for this is that the expanded ξ_t, say $\tilde{\xi}_t$ (which contains $n_k = kn_1^S + n_1F + kn_2^S + n_2^F$ 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 $n_k \times n$ matrix whose elements are polynomials that depict the way u_t and its high frequency lags feed into $\tilde{\xi}_t$. If $f_{uu}(\lambda)$ denotes the spectral density matrix of u_t then $H(e^{i\lambda})f_{uu}(\lambda)H(e^{-i\lambda})'$ is the spectral density matrix of $\tilde{\xi}_t$, which is 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.

The feasible low frequency representation in Lemma 1 can be regarded as the typical approach in time series in which data are reduced to the lowest frequency. The representation does not arise simply by choosing integer values of τh in the high frequency model because, as has been shown in the infeasible low frequency representation, this results in the unobservable flows y_{1t}^F and y_{2t}^F. The feasible representation replaces unobservables by observables and assigns the differences, such as $Y_{2,t-1}^S - y_{2,t-1}^S$ and $y_{2,t-1}^F - Y_{2,t-1}^F$, to the disturbances. These terms are stationary under Assumption 1; see Lemma A1 in the Appendix.

The two representations in Lemma 1 also have equivalent ECM forms. In order to demonstrate this it is convenient to define, for $t = 1, \ldots, T$, the vectors

$$Y_{1t} = \begin{pmatrix} Y_{1t}^S \\ Y_{1t}^F \end{pmatrix}, \quad Y_{2t} = \begin{pmatrix} Y_{2t}^S \\ Y_{2t}^F \end{pmatrix}, \quad \xi_{1t} = \begin{pmatrix} \xi_{1t}^S \\ \xi_{1t}^F \end{pmatrix}, \quad \xi_{2t} = \begin{pmatrix} \xi_{2t}^S \\ \xi_{2t}^F \end{pmatrix},$$

$$z_{1t} = \begin{pmatrix} y_{1t}^S \\ Y_{1t}^F \end{pmatrix}, \quad z_{2t} = \begin{pmatrix} y_{2t}^S \\ Y_{2t}^F \end{pmatrix}, \quad \zeta_{1t} = \begin{pmatrix} \zeta_{1t}^S \\ \zeta_{1t}^F \end{pmatrix}, \quad \zeta_{2t} = \begin{pmatrix} \zeta_{2t}^S \\ \zeta_{2t}^F \end{pmatrix},$$

as well as the stacked vectors $Y_t = (Y_{1t}', Y_{2t}')'$, $\xi_t = (\xi_{1t}', \xi_{2t}')'$, $z_t = (z_{1t}', z_{2t}')'$ and $\zeta_t = (\zeta_{1t}', \zeta_{2t}')'$. The mixed frequency ECM representation can then be written

$$\Delta Y_t = -JAY_{t-1} + \xi_t, \quad t = 1, \ldots, T, \tag{17}$$

while the feasible low frequency ECM representation is given by

$$\Delta z_t = -JAz_{t-1} + \zeta_t, \quad t = 1, \ldots, T. \tag{18}$$

Both the triangular representations in Lemma 1 as well as the ECM representations in (17) and (18) provide a suitable basis for the estimation of the parameters of the matrix C. We now turn to the analysis of a frequency domain-based estimator that rests only on weak assumptions concerning the disturbances in the high frequency model.

3. Estimation in the frequency domain

We shall focus initially on the mixed frequency representation and subsequently demonstrate how the results can be applied to the feasible low frequency representation.

3.1. The mixed frequency model

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 merely stationary under Assumption 1 rather than having any specific (parametric) dynamic structure, a natural approach to estimating the matrix C of cointegrating vectors is to use spectral/frequency domain regression. Based, then, on the mixed frequency representation in Lemma 1 we can write the system of interest as

$$Y_{0t} = JCY_{2,t-1} + \xi_t, \quad t = 1, \ldots, T, \tag{19}$$

where $Y_{0t} = (Y_{1t}', \Delta Y_{2,t}')'$.[2] The spectral regression approach is based on taking discrete Fourier transforms (dFts) in (19), yielding

$$w_0(\lambda_s) = JCw_2(\lambda_s) + w_\xi(\lambda_s), \quad s = -T/2 + 1, \ldots, T/2, \tag{20}$$

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

[2]This representation can also be obtained from the mixed frequency ECM in (17) by adding $Y_{1,t-1}$ to both sides of the equation. We also assume, for convenience, that observations for $t = 1, \ldots, T$ are available, rather than just $t = 2, \ldots, T$.

assumed to be an even number for convenience,[3] and

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

denote the dfTs of Y_{0t}, $Y_{2,t-1}$ and ξ_t, respectively, at the Fourier frequencies.

In cases where C is unrestricted – as is the case here – a simple least squares-type of spectral regression estimator can be obtained by choosing C so as to minimise an objective function of the form

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

where $w_\xi(\lambda_s) = w_0(\lambda_s) - JCw_2(\lambda_s)$, 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 the set 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 the property that both λ_s and $-\lambda_s$ 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, 1991b) as well as the simulation results reported in Corbae, Ouliaris and Phillips (1994). We therefore consider the symmetric set of frequencies $\Lambda_0 = \{\lambda_s = 2\pi s/T; s = -m, \ldots, m\}$ which contains the $2m + 1$ Fourier frequencies around the origin for some integer m. We also generalise the objective function by incorporating a positive definite Hermitian weighting matrix, $\Phi(\lambda)$, resulting in a generalised least squares-type of objective function of the form

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

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 nevertheless be obtained by using the residuals from a least squares regression of (19). These residuals – denoted $\hat{\xi}_t$ – can then be used in a variety of ways to estimate the spectral density matrix of interest.

The method we shall employ here to estimate $f_{\xi\xi}(0)$ is the smoothed periodogram estimator, defined by

$$\hat{f}_{\hat{\xi}\hat{\xi}}(0) = \frac{1}{2m+1} \sum_{j=-m}^{m} I_{\hat{\xi}\hat{\xi}}(\lambda_j), \tag{21}$$

where $I_{\hat{\xi}\hat{\xi}}(\lambda) = w_{\hat{\xi}}(\lambda) w_{\hat{\xi}}(\lambda)^*$ and $w_{\hat{\xi}}(\lambda)$ is the dFt of $\hat{\xi}_t$. The smoothed periodogram estimator is a straightforward symmetric average of $2m+1$ periodogram matrices around the frequency of interest (the frequency of interest here being zero). More sophisticated estimates could be used but the smoothed periodogram has performed well in the simulations that are reported

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

in the next section. With this choice of weighting matrix the objective function becomes

$$S(C) = \frac{1}{2m+1}\sum_{s=-m}^{m} \mathrm{tr}\left\{\hat{f}_{\hat{\xi}\hat{\xi}}(0)^{-1}\left(w_0(\lambda_s) - JCw_2(\lambda_s)\right)\left(w_0(\lambda_s) - JCw_2(\lambda_s)\right)^*\right\}. \tag{22}$$

Minimisation of (22) with respect to C results in the estimator

$$\hat{C}_0 = \left(J'\hat{f}_{\hat{\xi}\hat{\xi}}(0)^{-1}J\right)^{-1} J'\hat{f}_{\hat{\xi}\hat{\xi}}(0)^{-1}\hat{f}_{02}(0)\hat{f}_{22}(0)^{-1} \tag{23}$$

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

$$\hat{f}_{22}(0) = \frac{1}{2m+1}\sum_{j=-m}^{m} I_{22}(\lambda_j), \quad I_{22}(\lambda_j) = w_2(\lambda_j)w_2(\lambda_j)^*,$$

$$\hat{f}_{02}(0) = \frac{1}{2m+1}\sum_{j=-m}^{m} I_{02}(\lambda_j), \quad I_{02}(\lambda_j) = w_0(\lambda_j)w_2(\lambda_j)^*,$$

respectively. By noting that $w_0(\lambda) = JCw_2(\lambda) + w_\xi(\lambda)$ it follows that $\hat{f}_{02}(0) = JC\hat{f}_{22}(0) + \hat{f}_{\xi 2}(0)$, and making this substitution in (23) it is possible to express $\hat{C}_0$ directly in terms of C itself:

$$\hat{C}_0 = C + \left(J'\hat{f}_{\hat{\xi}\hat{\xi}}(0)^{-1}J\right)^{-1} J'\hat{f}_{\hat{\xi}\hat{\xi}}(0)^{-1}\hat{f}_{\xi 2}(0)\hat{f}_{22}(0)^{-1}. \tag{24}$$

Although this expression is not used for computation it is used for analytical purposes to derive the limiting distribution of the estimator. Also of use in some cases are the (column) vectorised versions of (23) and (24), given by

$$\hat{\gamma}_0 = \left(\hat{f}_{22}(0)^{-1} \otimes \left(J'\hat{f}_{\hat{\xi}\hat{\xi}}(0)^{-1}J\right)^{-1} J'\hat{f}_{\hat{\xi}\hat{\xi}}(0)^{-1}\right)\mathrm{vec}\left(\hat{f}_{02}(0)\right) \tag{25}$$

and

$$\hat{\gamma}_0 = \gamma + \left(\hat{f}_{22}(0)^{-1} \otimes \left(J'\hat{f}_{\hat{\xi}\hat{\xi}}(0)^{-1}J\right)^{-1} J'\hat{f}_{\hat{\xi}\hat{\xi}}(0)^{-1}\right)\mathrm{vec}\left(\hat{f}_{\xi 2}(0)\right), \tag{26}$$

respectively, where $\gamma = \mathrm{vec}(C)$ and $\hat{\gamma}_0 = \mathrm{vec}(\hat{C}_0)$. Similar expressions for spectral estimators can be found in Robinson (1972) and Phillips (1991b), any differences arising from the ordering of matrices under the trace operator and the use of row, rather than column, vectorisation.

The derivation of the asymptotic properties of the spectral density matrix estimators and, hence, of $\hat{\gamma}_0$, relies on an FCLT for the normalised partial sums of ξ_t. Based on (3) in Assumption 1 it is possible to derive an appropriate FCLT for the partial sums of ξ_t, which is a function of $u_{\tau h}$. This is presented below.

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

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} \xi_t \xrightarrow{d} B(r), \quad 0 < r \leq 1, \tag{27}$$

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

$$G = \left(\begin{array}{cc} hI_{n_1} & C \\ 0_{n_2 \times n_1} & I_{n_2} \end{array} \right).$$

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 partial sum of interest in Lemma 2 to the one in Assumption 1. 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 the Appendix.

The derivation of the asymptotic properties of $\hat{C}_0$ 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. $\dfrac{m}{T} + \dfrac{1}{m} \to 0$ as $T \to \infty$.

Hence m is required to grow with T but at a slower rate, which is a common assumption in the literature on spectral density estimation; see, for example, Brockwell and Davis (1991, p.351). A further assumption concerning the rate of growth of sums of autocovariances of $u_{\tau h}$ is employed.

Assumption 3. Let $\Gamma_{u,lh} = E(u_{\tau h} u'_{\tau h - lh})$. Then

$$\sum_{l=-N}^{N} |l| \|\Gamma_{u,lh}\| = O(N^{1/2}).$$

This assumption is satisfied if, for example, $u_{\tau h}$ is the linear process

$$u_{\tau h} = \sum_{j=0}^{\infty} A_j e_{\tau h - jh},$$

where $e_{\tau h} \sim \text{iid}(0, \Sigma)$ and $\sum_{j=0}^{\infty} j^{1/2} \|A_j\| < \infty$; see, for example, Fuller (1996, p.367) for a demonstration of this result. Assumption 3 enables a similar condition on the rate of growth of sums of autocovariances of the disturbances in the mixed frequency representation (ξ_t) to be established.

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

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

Lemma 3 is required to establish a consistency result concerning $\hat{f}_{\hat{\xi}\hat{\xi}}(0)$ which is provided in Theorem 1(c) below. The use of Assumptions 1–3 enables the following result concerning the asymptotics 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 $T \to \infty$:*

(a) $\dfrac{2m+1}{T^2}\hat{f}_{22}(0) \xrightarrow{d} \dfrac{1}{\pi}\displaystyle\int_0^1 B_2 B_2'$;

(b) $\dfrac{2m+1}{T}\hat{f}_{\xi 2}(0) \xrightarrow{d} \dfrac{1}{\pi}\displaystyle\int_0^1 dBB_2' + \dfrac{1}{2\pi}\Omega_2$, *where* $\Omega_2 = \displaystyle\sum_{j=-\infty}^{\infty} E(\xi_{t+j}\xi_{2t}')$.

If, in addition, Assumption 3 is satisfied, then

(c) $\hat{f}_{\hat{\xi}\hat{\xi}}(0) = f_{\xi\xi}(0) + o_p(1)$.

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

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

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 distribution of $\hat{C}_0$ can now be stated.

Theorem 2. *Under Assumptions 1–3, as $T \to \infty$,*

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

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

The estimator $\hat{C}_0$ therefore belongs to the class of optimal estimators as defined by Phillips (1991c). These are estimators having the form of limit distribution as given in Theorem 2 i.e. mixed normal. Although the optimality has been achieved using a system-wide estimator, Phillips (1991a, 1991b, 1991c) showed that equivalent asymptotic efficiency can be achieved using an augmented (frequency domain) regression estimator based on only the first n_1 equations of the system (19) or (20). The augmented equation includes ΔY_{2t} (or its dFt) as an additional regressor vector, resulting in the time domain regression equation

$$Y_{1t} = CY_{2,t-1} + F\Delta Y_{2t} + \xi_{1.2t}, \quad t = 1, \ldots, T, \tag{28}$$

where $F = \Omega_{12}\Omega_{22}^{-1}$ and $\xi_{1.2t} = \xi_{1t} - F\xi_{2t}$. In the frequency domain the relevant equation is

$$w_1(\lambda_s) = Cw_2(\lambda_s) + Fw_{\Delta_2}(\lambda_s) + w_{1.2}(\lambda_s), \quad s = -T/2+1, \ldots, T/2, \tag{29}$$

where $w_1(\lambda_s)$, $w_{\Delta_2}(\lambda_s)$ and $w_{1.2}(\lambda_s)$ are the dFts of Y_{1t}, ΔY_{2t} and $\xi_{1.2t}$, respectively. One advantage of this approach is that it is not necessary to construct an estimator of the disturbance spectral density matrix using an initial consistent estimator. The band-limited estimator of C based on the augmented equation is obtained by minimising the least-squares objective function

$$S_A(C) = \frac{1}{2m+1}\sum_{s=-m}^{m} \text{tr}\left\{w_{1.2}(\lambda_s)w_{1.2}(\lambda_s)^*\right\}, \tag{30}$$

where $w_{1.2}(\lambda_s) = w_1(\lambda_s) - Cw_2(\lambda_s) - Fw_{\Delta_2}(\lambda_s)$. The resulting estimator can be written in the form

$$\hat{C}_0^A = \left(\hat{f}_{12}(0) - \hat{f}_{1\Delta_2}(0)\hat{f}_{\Delta_2\Delta_2}(0)^{-1}\hat{f}_{\Delta_2 1}(0)\right)\left(\hat{f}_{22}(0) - \hat{f}_{2\Delta_2}(0)\hat{f}_{\Delta_2\Delta_2}(0)^{-1}\hat{f}_{\Delta_2 2}(0)\right)^{-1}, \quad (31)$$

where the $\hat{f}(0)$ are the smoothed periodogram estimators using the relevant variables. Using the results in Theorem 1[4] we find that

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

$$\frac{2m+1}{T^2}\left(\hat{f}_{22}(0) - \hat{f}_{2\Delta_2}(0)\hat{f}_{\Delta_2\Delta_2}(0)^{-1}\hat{f}_{\Delta_2 2}(0)\right) \xrightarrow{d} \frac{1}{\pi}\int_0^1 B_2B_2',$$

results which imply that

$$T(\hat{C}_0^A - C) \xrightarrow{d} \int_0^1 dB_{1.2}B_2'\left(\int_0^1 B_2B_2'\right)^{-1}.$$

Hence $\hat{C}_0^A$ shares the optimality properties of $\hat{C}_0$ but has potential computational advantages.

An 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 $q < n_1n_2$ possibly non-linear restrictions on the elements of C, represented by the null hypothesis

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

where $r(\cdot)$ 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 $q \times n_1n_2$ 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

$$W_0 = \frac{2m+1}{2}r(\hat{\gamma}_0)'\left(R(\hat{\gamma}_0)\hat{V}_0^{-1}R(\hat{\gamma}_0)'\right)^{-1}r(\hat{\gamma}_0), \quad (32)$$

where

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

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

$$W_0^A = \frac{2m+1}{2}r(\hat{\gamma}_0^A)'\left(R(\hat{\gamma}_0^A)(\hat{V}_0^A)^{-1}R(\hat{\gamma}_0^A)'\right)^{-1}r(\hat{\gamma}_0^A), \quad (33)$$

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

[4]Note that Theorem 1(c) also applies to $\hat{f}_{\Delta_2\Delta_2}(0)$ in this case.

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 1–3.[5] The limiting distributions of these Wald statistics are given below.

Theorem 3. *Under Assumptions 1–3, as* $T \to \infty$, $W_0 \xrightarrow{d} \chi^2_q$ *and* $W_0^A \xrightarrow{d} \chi^2_q$ *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.

3.2. The feasible low frequency model

The results obtained for the mixed frequency model have parallels in the feasible low frequency framework, albeit with appropriate modifications. In place of (19) we now have

$$z_{0t} = JCz_{2,t-1} + \zeta_t, \quad t = 1, \ldots, T,$$

where $z_{0t} = (z'_{1t}, \Delta z'_{2t})'$. Smoothed periodogram estimators can be formed using the dFts of the appropriate elements of z_{0t} and $z_{2,t-1}$; we shall denote these as $\tilde{f}_{22}(\lambda)$ etc., resulting in the estimator

$$\tilde{C}_0 = \left(J'\tilde{f}_{\tilde{\zeta}\tilde{\zeta}}(0)^{-1}J\right)^{-1} J'\tilde{f}_{\tilde{\zeta}\tilde{\zeta}}(0)^{-1}\tilde{f}_{02}(0)\tilde{f}_{22}(0)^{-1}$$

where $\tilde{\zeta}$ denotes the residuals obtained from an initial consistent estimator of C. It is also possible to consider an augmented regression of the form

$$z_{1t} = Cz_{2,t-1} + F\Delta z_{2t} + \zeta_{1.2t}, \quad t = 1, \ldots, T,$$

resulting in the estimator

$$\tilde{C}_0^A = \left(\tilde{f}_{12}(0) - \tilde{f}_{1\Delta_2}(0)\tilde{f}_{\Delta_2\Delta_2}(0)^{-1}\tilde{f}_{\Delta_2 1}(0)\right)\left(\tilde{f}_{22}(0) - \tilde{f}_{2\Delta_2}(0)\tilde{f}_{\Delta_2\Delta_2}(0)^{-1}\tilde{f}_{\Delta_2 2}(0)\right)^{-1}.$$

Wald statistics, having the same form as in (32) and (33), can be constructed based on $\tilde{C}_0$ and $\tilde{C}_0^A$ and will be denoted $\tilde{W}_0$ and $\tilde{W}_0^A$, respectively. Results analogous to Lemmas 2 and 3 and Theorems 1–3 are contained in the following proposition.

Proposition 1. *(a) Under Assumptions 1, as* $T \to \infty$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} \zeta_t \xrightarrow{d} b(r), \quad 0 < r \leq 1, \tag{34}$$

where $b(r)$ *is a Brownian motion process with covariance matrix* $P = h^{-1}G_\zeta \Omega_u G'_\zeta$,

$$G_\zeta = \begin{pmatrix} hI_{n_1} & C_k \\ 0_{n_2\times n_1} & I_{n_2} \end{pmatrix}, \quad C_k = \begin{pmatrix} C_{SS} & c_1 C_{SF} \\ c_2 C_{FS} & C_{FF} \end{pmatrix},$$

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

$c_1 = (k+1)/2 = (h+1)/2h$ *and* $c_2 = (3-k)/2 = (3h-1)/2h$.
(b) Let $\Gamma_{\zeta,l} = E(\zeta_t \zeta_{t-l}')$. *Then, under Assumption 3,*

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

(c) Let $b(r) = (b_1(r)', b_2(r)')'$. *Then, under Assumptions 1 and 2, as* $T \to \infty$*:*

(i) $\dfrac{2m+1}{T^2} \tilde{f}_{22}(0) \xrightarrow{d} \dfrac{1}{2\pi} \int_0^1 b_2 b_2'$;

(ii) $\dfrac{2m+1}{T} \tilde{f}_{\zeta 2}(0) \xrightarrow{d} \dfrac{1}{2\pi} \int_0^1 db b_2' + \dfrac{1}{2\pi} P_2$, *where* $P_2 = \sum_{j=1}^{\infty} E(\zeta_{t+j} \zeta_{2t}')$.

If, in addition, Assumption 3 is satisfied, then

(iii) $\tilde{f}_{\tilde{\zeta}\tilde{\zeta}}(0) = f_{\zeta\zeta}(0) + o_p(1)$.

(d) Under Assumptions 1–3, as $T \to \infty$,

$$T(\tilde{C}_0 - C) \xrightarrow{d} \int_0^1 db_{1.2} b_2' \left(\int_0^1 b_2 b_2' \right)^{-1}$$

where $b_{1.2}(r)$ *is a Brownian motion process with covariance matrix* $P_{11.2} = P_{11} - P_{12} P_{22}^{-1} P_{21}$. $T(\tilde{C}_0^A - C)$ *also converges to the same limiting distribution.*
(e) Under Assumptions 1–3, as $T \to \infty$, $\tilde{W}_0 \xrightarrow{d} \chi_q^2$ *and* $\tilde{W}_0^A \xrightarrow{d} \chi_q^2$ *under* H_0.

Proposition 1 demonstrates the validity of the spectral regression methods for the feasible low frequency model. The finite sample performance of these methods in this aggregated model are explored in the next section and compared with those based on the mixed frequency representation.

4. Simulation results

In this section we explore the finite sample properties of the spectral regression estimators and the Wald statistics. Our focus is on the case where there is a high frequency stock variable, y_1, and a low frequency flow variable, y_2, that are related 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 infeasible case, where both variables are observed at the highest frequency, as well as the feasible low frequency and mixed frequency cases of interest.

The simulation model is motivated by the empirical relationship between stock prices and dividends that has been the focus of much research. Ghysels and Miller (2015) have tested for cointegration between these variables using mixed frequency techniques based on an extended version of the data in Shiller (2000). The updated data are now available (at the time of writing) from January 1871 to August 2017, a span of 147 years. The stock price data are monthly while dividends are observed annually although an interpolated monthly dividend series is also available. In accordance with this type of data availability the simulations take the data span to be $T = 100$ and the high frequency sampling interval to be $h = 1/12$, which leads to $N = 1200$ high frequency observations. Data are generated at the highest frequency and then aggregated as required. We therefore generate $y_{1,\tau h}$ and $y_{2,\tau h}$ ($\tau = 1, \ldots, N$) but only the former is assumed to be observed by the econometrician. The latter (flow) variable is 'observed' at the low frequency in the form $Y_{2t} = (1/k)\sum_{l=0}^{k-1} y_{2,t-lh}$ ($t = 1, \ldots, T$) and for the mixed frequency representation we can aggregate the 'observed' y_1 to produce $Y_{1t} = (1/k)\sum_{l=0}^{k-1} y_{1,t-lh}$. The high frequency bivariate innovations satisfy a first-order vector autoregression of the form

$$u_{\tau h} = \Phi u_{\tau h-h} + \epsilon_{\tau h}, \quad \tau = 1, \ldots, N,$$

where $\epsilon_{\tau h}$ is an uncorrelated $N(0, I_2)$ process and the following autoregressive matrices were used: $\Phi = 0_{2\times 2}$ (so that $u_{\tau h}$ is Gaussian white noise) and $\Phi = \Phi_j$ ($j = 1, \ldots, 6$) where

$$\Phi_1 = \begin{pmatrix} 0.8 & 0 \\ 0 & 0.8 \end{pmatrix}, \quad \Phi_2 = \begin{pmatrix} 0.8 & 0 \\ 0.5 & 0.8 \end{pmatrix}, \quad \Phi_3 = \begin{pmatrix} 0.8 & 0 \\ -0.5 & 0.8 \end{pmatrix},$$

$$\Phi_4 = \begin{pmatrix} 0.8 & 0.5 \\ -0.5 & 0.8 \end{pmatrix}, \quad \Phi_5 = \begin{pmatrix} 0.8 & -0.5 \\ 0.5 & 0.8 \end{pmatrix}, \quad \Phi_6 = \begin{pmatrix} 0.95 & 0 \\ 0 & 0.95 \end{pmatrix}.$$

These specifications allow for the presence of positive and negative feedback to/from u_1 and u_2 while Φ_6 has roots closer to the boundary of the stationary region. The matrices Φ_1 to Φ_3 have repeated roots equal to 1.25; Φ_4 and Φ_5 have roots of $0.8989 \pm 0.56i$ with modulus 1.06; and Φ_6 has repeated roots of 1.0526. In all cases $u_0 = (0, 0)'$.

In the simulations we consider three different estimation models, as follows:

Model 1 ("High"). This is the infeasible model where high frequency observations on both variables are used. The model estimated for $\tau = 1, \ldots, N$ is therefore, from (4),

$$\begin{aligned} y_{1,\tau h} &= C y_{2,\tau h-h} + e_{1,\tau h}, \\ y_{2,\tau h} &= y_{2,\tau h-h} + e_{2,\tau h}. \end{aligned}$$

Model 2 ("Low"). This is the feasible low frequency model which has the form

$$\begin{aligned} y_{1t} &= C Y_{2,t-1} + \zeta_{1t}, \\ Y_{2t} &= Y_{2,t-1} + \zeta_{2t}, \end{aligned}$$

where $t = 1, \ldots, T$; see Lemma 1(b).

Model 3 ("Mixed"). The mixed frequency representation based on Lemma 1(a) is of the form (for $t = 1, \ldots, T$)

$$\begin{aligned} Y_{1t} &= C Y_{2,t-1} + \xi_{1t}, \\ Y_{2t} &= Y_{2,t-1} + \xi_{2t}. \end{aligned}$$

A total of 10,000 replications for each VAR model for $u_{\tau h}$ were carried out and estimates of each of the three models were computed. In addition to the ordinary least squares (OLS) estimator of C we also consider three different spectral regression estimators, each using three different values of the bandwidth parameter m, resulting in ten different estimates of C in each replication. The first spectral regression estimator is $\hat{C}_0$, defined in (23), in which the smoothed periodogram estimator $\hat{f}_{\hat{\xi}\hat{\xi}}(0)$ is based on a set of OLS residuals, $\hat{\xi}_t$; this estimator is denoted FD in what follows. The second is the augmented estimator $\hat{C}_0^A$, defined in (31), 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.[6] The ASDE of $f_{\xi\xi}(0)$ first fits a first-order VAR to the OLS residuals of the form

$$\hat{\xi}_t = \hat{K}\hat{\xi}_{t-1} + \hat{v}_t,$$

and then computes the estimator of the spectral density matrix at the origin using

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

where $\hat{\Sigma}_v$ is the estimated covariance matrix of the residuals from the VAR. The choice of m is required to satisfy Assumption 2 and so we take $m = [T^\delta]$ in the low and mixed frequency cases with $\delta \in \{0.3, 0.5, 0.7\}$; for $T = 180$ this results in $m \in \{3, 10, 25\}$. In the infeasible high frequency case we scale the low and mixed frequency values by k leading to $m \in \{36, 120, 300\}$. 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 error (RMSE) of the estimators, multiplied by 10^4 (hence the entry in the Table of 19.75 for the estimator FDA1 in Model 1 under white noise, for example, is to be interpreted as an actual RMSE of 0.001975). Beginning with the case of white noise disturbances in the high frequency model it is apparent that all the spectral estimators produce uniformly smaller RMSEs than OLS across all three models. Estimation of the low frequency model, which does not utilise the full mixed frequency sample data, leads to an approximate six-fold increase in the RMSE values (and even larger for OLS). The mixed model, on the other hand, results in a large reduction in the RMSEs, with those of the FDA estimators almost achieving the values of the infeasible high frequency model.

The remaining entries in Table 1 are based on data in which the high frequency innovations, $u_{\tau h}$, are generated by a first-order VAR process. Broadly speaking a similar qualitative pattern emerges in the VAR cases as in the white noise case, although the magnitudes of the RMSEs are somewhat different. In four of the cases (Φ_2, Φ_3, Φ_4 and Φ_5) the RMSE of the FDA1 estimator in the mixed frequency model is actually smaller than in the high frequency model; these are the cases in which feedback is allowed between u_1 and u_2. When the feedback is purely from u_1 to u_2 (i.e. Φ_2 and Φ_3) the RMSEs are typically smaller than in the white noise case but when additional feedback from u_2 to u_1 is allowed the RMSEs

[6] Smoothed periodogram estimators are used to estimate the remaining spectral density matrices.

are generally higher. A comparison of the results for Φ_1 with those for Φ_6 suggests that allowing the roots to move closer to unity has little impact on the FD and FDA estimators but there is a large increase in the RMSEs for the ASD estimator in the mixed frequency model, which is presumably due to difficulties in accurately estimating the spectral density matrix when the roots are close to unity.

The general picture to emerge from Table 1 is that utilising all the high frequency data in the mixed frequency model provides substantial improvements over discarding such data and using only the low frequency observations. Moreover spectral regression estimators appear to provide a useful method for the estimation of the cointegration parameter in such settings.[7] It is also of interest to investigate the finite sample properties of the Wald statistics based on such 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 the white noise case for $u_{\tau h}$ as well as the three VAR processes using Φ_2, Φ_4 and Φ_6. In addition to the OLS-based test we report results for the spectral regression estimators using the fewest periodogram ordinates i.e. FD1, FDA1 and ASD1. The entries in Table 2 are percentages and those for power are not size-adjusted; the nominal size of the tests is 5 percent.

The OLS-based Wald tests show the greatest size distortions in the VAR cases while all tests have size distortions in the infeasible high frequency model; the sizes of the FDA1-based tests are closest to the nominal size in the low and mixed frequency models. In terms of power the FDA1-based tests also dominate and achieve high power even for the relatively close values of C under H_1 to its value under H_0. This feature, combined with the good performance of the FDA estimators in terms of RMSE, suggest that spectral regression of the augmented regression model using the mixed frequency model provides a good platform for estimation and inference in cointegrated models.

5. An empirical example

In this section we investigate the relationship between US stock prices and dividends using the extended data set based on Shiller (2000).[8] The stock price data are available monthly (being the average daily closing price during the month) while the dividend data are yearly. We treat the monthly price data as a stock variable (despite the averaging within the month) and the dividend data are clearly in the form of a flow. The sample begins in January 1871 and we use data through to December 2016, yielding $T = 146$ observations on dividends and $N = 1752$ observations on stock prices (with $h = 1/12$). As in Ghysels and Miller (2015) we can consider beginning-of-period (BOP) sampling for prices, taking the January value each year, end-of-period (EOP) sampling, in which the December value is chosen, or yearly averaging of the twelve months each year. The first two sampling methods enable the feasible low frequency model to be estimated while the latter enables the mixed frequency model to be estimated.

[7] These results are robust to replacing $\epsilon_{\tau h}$ in the VAR specification for $u_{\tau h}$ by an ARCH process and by a Gamma distributed process although we do not report the results here in order to economise on space.

[8] The data can be downloaded from `http://aida.econ.yale.edu/~shiller/data.htm`.

Logarithms of the yearly dividend data and the three types of sampled stock price data are presented in Figure 1.[9] The different types of sampling of the stock prices appear to make very little difference to the properties of the resulting series, while both variables display upward trends over the sample period. Ghysels and Miller (2015) use demeaned data in their cointegration analysis, arguing that any trends in the series should be the same and the drifts will cancel out. This does not seem to be supported by inspection of Figure 1 in which the slopes appear to be significantly different. They also argue that cointegration is not to be expected owing to the increasing proportion of tech companies, many of which do not pay dividends, since the 1990's leading to a structural break or a breakdown in the relationship between these variables. In the analysis below we use both demeaned and detrended data and proceed on the basis of cointegration. We also examine the residual plots from the cointegrating regressions for any obvious evidence of nonstationarity.

We begin with the demeaned data, which are depicted in Figure 2. The upward trends clearly remain in the demeaned series, as would be expected. The dividend data lie below the price data in the first half of the sample and then rise and remain above the prices in the second half of the sample, the cross-over point being around 1950. We compute OLS as well as the spectral and augmented spectral regression estimators using $m = [T^{\delta}]$ for $\delta = \{0.3, 0.5, 07\}$ (the values that were used in the simulations), yielding $m = \{4, 12, 32\}$ for $T = 146$. The underlying regression of interest is of the form

$$\log P_{\tau h} = C \log D_{\tau h} + u_{\tau h}, \quad \tau = 1, \ldots, N,$$

where P denotes stock price and D dividends. In view of unit roots in $\log P$ and $\log D$, stationarity of the log price-dividend ratio, $\log(P/D)$, suggests that the cointegrating parameter should be equal to unity. We therefore test $H_0 : C = 1$ against $H_1 : C \neq 1$ using the Wald statistics proposed in section 3. Results using the demeaned data are presented in Table 3. As can be seen from Table 3, the estimates of C are stable at around 0.52 with small standard errors, suggesting that the data are sufficiently informative to reject the hypothesis that $C = 1$. Indeed, the Wald statistics are highly significant. The residuals from the cointegrating regression using the augmented spectral regression estimator with $m = 4$ and averaged price data (in which the estimated coefficient is 0.5251) are graphed in Figure 3. The residuals are reasonably stable although there is evidence of trending towards the end of the sample period.

A plot of the detrended data is given in Figure 4. Unlike the demeaned data there is much more variation in the series with multiple crossing points. The results obtained with the detrended data are given in Table 4. The OLS estimates of C are very similar to those obtained with the demeaned data but the spectral estimators are uniformly larger at roughly 0.57; the standard errors are also larger than those computed with the demeaned data. The Wald statistics are also lower than those with the demeaned data although all remain highly significant (the largest significance is obtained with the spectral estimator using $m = 4$ with the end-of-period price data, although the value is only 0.001). Figure 5 plots the residuals from the cointegrating regression using the augmented spectral regression estimator with $m = 4$ and averaged price data (in which the estimated coefficient is 0.5727). The residuals

[9] To be consistent with the theory in section 2, the averaged stock price data are the yearly averages of the logarithms of the monthly prices, not the logarithm of the average monthly price.

display the same pattern as those obtained with the detrended data.

The detrended data in Figure 4 are suggestive of a possible trend break around the middle of the sample. Detrending the two sub-periods, 1871–1942 and 1943–2016, separately, yields the detrended data in Figure 6, in which there is no evidence of any remaining trends. Using the break-detrended data for estimation has a significant impact on the results, which are reported in Table 5. Estimates of C are much lower than those reported in Tables 3 and 4 and the standard errors are larger. The choice of m also has a greater impact on the spectral regression estimates than previously, ranging from roughly 0.10 with $m = 4$ to 0.22 with $m = 12$ and between 0.11 and 0.18 with $m = 32$. In all cases the null hypothesis $C = 1$ is convincingly rejected. The residuals from the cointegrating regression using the augmented spectral regression estimator with $m = 4$ and averaged price data (in which the estimated coefficient is 0.0965) are displayed in Figure 7. These residuals show slightly less dispersion than those reported earlier although there are a couple of noticeable spikes.

6. Concluding comments

This paper has proposed a model suitable for 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 a spectral regression estimator in an augmented mixed frequency model is particularly encouraging as it is capable of dramatically reducing the RMSE obtained in an entirely low frequency model to the levels comparable to an infeasible high frequency model. The size and power properties of the associated Wald statistic are also good. An empirical example, to stock price and dividend data, was also provided to demonstrate the methods in practice.

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 – such as is carried out in the empirical example – is a valid approach in which 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, Ouliaris and Phillips (2002) in band limited spectral regression in models in which the parameters are frequency-dependent. Alternatively the intercept and trend could be estimated as part of the spectral regression procedure. An assessment of the finite sample effects of these alternative approaches would constitute an interesting research exercise.

Appendix

Proofs of Lemmas and Theorems

Proof of Lemma 1. (a) We begin the derivation of the mixed frequency representation by selecting the first n_1 equations from (8) relating to y_{1t}, which are given by

$$y_{1t} = Cy_{2,t-1} + v_{1t}, \quad t = 1, \ldots, T. \tag{35}$$

The first n_1^S equations relating to y_{1t}^S are

$$y_{1t}^S = C_{SS}y_{2,t-1}^S + C_{SF}y_{2,t-1}^F + v_{1t}^S, \quad t = 1, \ldots, T, \tag{36}$$

where v_{1t} has been partitioned as $v_{1t} = \left(v_{1t}^{S\prime}, v_{1t}^{F\prime}\right)'$. In view of Y_{2t}^F being observed as an average of the unobservable y_{2t}^F between $t-1$ and t it makes sense to aggregate this equation in the same way. As $Y_{2t}^F = k^{-1}s(L^h)y_{2t}^F$ we apply the operator $k^{-1}s(L^h)$ to (36) to obtain (9) with $\xi_{1t}^S = k^{-1}\sum_{l=0}^{k-1} v_{1,t-lh}^S$. The representation for Y_{1t}^F is obtained in the same way by applying the same filter to the last n_1^F equations of (35), which are

$$y_{1t}^F = C_{FS}y_{2,t-1}^S + C_{FF}y_{2,t-1}^F + v_{1t}^F, \quad t = 1, \ldots, T. \tag{37}$$

The result is (10) with $\xi_{1t}^F = k^{-1}\sum_{l=0}^{k-1} v_{1,t-lh}^F$. Finally, the stochastic trends for Y_{2t}^S and Y_{2t}^F are obtained by applying the same filter again, this time to (7), resulting in (11) and (12) with $\xi_{2t}^S = k^{-1}\sum_{l=0}^{k-1} w_{2,t-lh}^S$ and $\xi_{2t}^F = k^{-1}\sum_{l=0}^{k-1} w_{2,t-lh}^F$.

(b) The objective in the feasible low frequency representation is to skip-sample the high frequency stock variables at integer values of τh and relate them to the observed low frequency flows. The equation for y_{1t}^S is obtained from (36) as follows:

$$\begin{aligned} y_{1t}^S &= C_{SS}y_{2,t-1}^S + C_{SF}y_{2,t-1}^F + v_{1t}^S \\ &= C_{SS}y_{2,t-1}^S + C_{SF}Y_{2,t-1}^F + v_{1t}^S + C_{SF}\left(y_{2,t-1}^F - Y_{2,t-1}^F\right) \\ &= C_{SS}y_{2,t-1}^S + C_{SF}Y_{2,t-1}^F + \zeta_{1t}^S, \end{aligned}$$

where $\zeta_{1t}^S = v_{1t}^S + C_{SF}\delta_{2,t-1}^F$ and $\delta_{2,t-1}^F = y_{2,t-1}^F - Y_{2,t-1}^F$ is I(0) using Lemma A1 with $j = 0$. For Y_{1t}^F a similar procedure can be carried out using (10):

$$\begin{aligned} Y_{1t}^F &= C_{FS}Y_{2,t-1}^S + C_{FF}Y_{2,t-1}^F + \xi_{1t}^F \\ &= C_{FS}y_{2,t-1}^S + C_{FF}Y_{2,t-1}^F + \xi_{1t}^F + C_{FS}\left(Y_{2,t-1}^S - y_{2,t-1}^S\right) \\ &= C_{FS}y_{2,t-1}^S + C_{FF}Y_{2,t-1}^F + \zeta_{1t}^F, \end{aligned}$$

where $\zeta_{1t}^F = \xi_{1t}^F - C_{FS}\delta_{2,t-1}^S$ is I(0) using Lemma A1. Finally, the stochastic trends for y_{2t}^S come directly from the first n_2^S equations of (7), so that $\zeta_{2t}^S = w_{2t}^S$, while those for Y_{2t}^F are simply (12), so that $\zeta_{2t}^F = \xi_{2t}^F$. □

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

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

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

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

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. □

Proof of Lemma 3. Let $M_u(z) = \sum_{l=-\infty}^{\infty} \Gamma_{u,lh} z^{lh}$ denote the autocovariance generating function (AGF) of $u_{\tau h}$. 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 for ξ_t at the high frequency. 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 $k-1$:

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

where the s_{1l} 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 terms involving, in addition to

$s(z)s(z^{-1})$ itself,

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

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

$$\begin{aligned} \sum_{l=-T}^{T} |l| \|\Gamma_{\xi,l}\| &= \sum_{l=-T}^{T} |l| \left\| \sum_{m=-p}^{p} a_m \Gamma_{u,lh-mh} \right\| \\ &\leq \sum_{m=-p}^{p} |a_m| \sum_{l=-T}^{T} |l| \|\Gamma_{u,lh-mh}\| = O(T^{1/2}) \end{aligned}$$

which implies the required result as p is finite and independent of T. □

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

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

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

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

We are then led to consider

$$\begin{aligned} \frac{2m+1}{T^2}\hat{f}_{22}(0) &= \frac{1}{2\pi T}\sum_{k=-T+1}^{T-1}\left(\frac{1}{T}\hat{\Gamma}_{22,k}\right)w_k \\ &\overset{d}{\to} \frac{1}{2\pi}\int_0^1 B_2B_2' \times \lim_{T\to\infty}\frac{1}{T}\sum_{k=-T+1}^{T-1} w_k \\ &= \frac{1}{\pi}\int_0^1 B_2B_2' \end{aligned}$$

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

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

We are then led to consider

$$\begin{aligned} \frac{2m+1}{T}\hat{f}_{\xi 2}(0) &= \frac{1}{2\pi T}\sum_{k=-T+1}^{T-1}\left(\frac{1}{T}\hat{\Gamma}_{\xi 2,k}\right)w_k \\ &\overset{d}{\to} \frac{1}{2\pi}\int_0^1 dBB_2' \times \lim_{T\to\infty}\frac{1}{T}\sum_{k=-T+1}^{T-1} w_k + \lim_{T\to\infty}\frac{1}{2\pi T}\sum_{k=-T+1}^{T-1} S_{2,k+1}w_k \\ &= \frac{1}{\pi}\int_0^1 dBB_2' + \lim_{T\to\infty}\frac{1}{2\pi T}\sum_{k=-T+1}^{T-1} S_{2,k+1}w_k \end{aligned}$$

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 $S_{2,T} \to 0$ as $T \to \infty$ while, from Lemma A4(a),

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

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,

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

as required.
(c) We begin by using the decomposition

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

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

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

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

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

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 can be shown to be $o_p(1)$ if Assumption 3 holds in addition to Assumptions 1 and 2 as Lemma 3 can be used to control the rate of growth of the autocovariances of ξ_t. This second term is a consistency result for the infeasible smoothed periodogram estimator based on the unobservable ξ_t and follows, for example, from results in Hannan (1970) and Fuller (1996). □

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, from (24), we are led to consider

$$\begin{aligned}T(\hat{C}_0-C) &= \left(J'f_{\xi\xi}(0)^{-1}J\right)^{-1}J'f_{\xi\xi}(0)^{-1}\left(\frac{2m+1}{T}\hat{f}_{\xi 2}(0)\right)\left(\frac{2m+1}{T^2}\hat{f}_{22}(0)\right)^{-1}+o_p(1) \\ &\xrightarrow{d} \left(J'f_{\xi\xi}(0)^{-1}J\right)^{-1}J'f_{\xi\xi}(0)^{-1}\left(\frac{1}{\pi}\int_0^1 dBB_2'+\frac{1}{2\pi}\Omega_2\right)\left(\frac{1}{\pi}\int_0^1 B_2B_2'\right)^{-1} \\ &= \left(J'f_{\xi\xi}(0)^{-1}J\right)^{-1}J'f_{\xi\xi}(0)^{-1}\int_0^1 dBB_2'\left(\int_0^1 B_2B_2'\right)^{-1} \\ &\quad +\frac{1}{2}\left(J'f_{\xi\xi}(0)^{-1}J\right)^{-1}J'f_{\xi\xi}(0)^{-1}\Omega_2\left(\int_0^1 B_2B_2'\right)^{-1}.\end{aligned}$$

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

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

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 $B_{1.2}=\left(I_{n_1}:-\Omega_{12}\Omega_{22}^{-1}\right)B=B_1-\Omega_{12}\Omega_{22}^{-1}B_2$. For the second term we begin by noting that

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

and, hence, it follows that

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

This demonstrates that the second term is zero and the limiting distribution is defined as in the Theorem. □

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),

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

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

$$Tr(\hat{\gamma}_0) = R(\bar{\gamma})T(\hat{\gamma}_0 - \gamma) \xrightarrow{d} R_0 \left[\left(\int_0^1 B_2B_2'\right)^{-1} \otimes I_{n_1}\right] \int_0^1 (B_2 \otimes dB_{1.2})\, R_0'.$$

This limiting distribution, conditional on B_2, is $N(0, R_0QR_0')$ 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 $N(0, R_0QR_0')$ random variables weighted by the matrix $(R_0QR_0')^{-1}$, 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. □

Proof of Proposition 1. (a) The proof follows that of Lemma 2 based on

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

Following the same steps we find that the limit Brownian motion is $b(r) = (1/\sqrt{h})G_\zeta(1)B_u(r)$ which has the stated covariance matrix.
(b) The proof follows Lemma 3 with appropriate modifications.
(c) The proof follows from Theorem 1.
(d) The proof is based on Theorem 2.
(e) The proof uses the same steps as the proof of Theorem 3. □

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 $0 \le j \le k-1$ 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 (5). The spectral density matrix of δ_{t-jh} is then given by

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

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

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

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

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

and so $f_{\delta,j}(\lambda)$ is positive and bounded for $\lambda \ne 0$ 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 $\lambda \to 0$). □

Lemma A2. *The disturbances in the mixed frequency and feasible low frequency representations, ξ_t and ζ_t, respectively, are related to those in the high frequency representation, $u_{\tau h}$,* *as follows:*

$$\xi_t = G(L^h)s(L^h)u_t, \quad t = 1, \ldots, T, \tag{38}$$

$$\zeta_t = G_\zeta(L^h)s(L^h)u_t, \quad t = 1, \ldots, T, \tag{39}$$

where $s(z)$ is defined in (6)*,*

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

$$G_\zeta(z) = \begin{pmatrix} s(z)^{-1}I_{n_1^S} & 0_{n_1^S \times n_1^F} & C_{SS} & (1 + z^k s_2(z))C_{SF} \\ 0_{n_1^F \times n_1^S} & hI_{n_1^F} & (hs(z) - z^k s_2(z))C_{FS} & hs(z)C_{FF} \\ 0_{n_2^S \times n_1^S} & 0_{n_2^S \times n_1^F} & I_{n_2^S} & 0_{n_2^S \times n_2^F} \\ 0_{n_2^F \times n_1^S} & 0_{n_2^F \times n_1^F} & 0_{n_2^F \times n_2^S} & hs(z)I_{n_2^F} \end{pmatrix},$$

and $s_2(z) = s(z) - h\sum_{l=0}^{k-1}(l+1)z^l$.

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 equation following (8), $v_{1t} = u_{1t} + Cw_{2t}$ and from (7), $w_{2t} = s(L^h)u_{2t}$. 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 (38).

Turning to ζ_t we again use the definitions from the proof of Lemma 1 which give $\zeta_{1t}^S = v_{1t}^S + C_{SF}\delta_{2,t-1}^F$, $\zeta_{1t}^F = \xi_{1t}^F - C_{FS}\delta_{2,t-1}^S$, $\zeta_{2t}^S = w_{2t}^S$ and $\zeta_{2t}^F = \xi_{2t}^F$. It is therefore necessary to find a term relating $\delta_{2,t-1}$ to u_t. We begin with the definition of δ_{2t} which, from Lemma A1, is

$$\delta_{2t} = \begin{pmatrix} \delta_{2t}^S \\ \delta_{2t}^F \end{pmatrix}' = y_{2t} - \frac{1}{k}\sum_{l=0}^{k-1} y_{2,t-lh}.$$

Now, because $\Delta y_{2t} = w_{2t}$ we have

$$y_{2t} = y_{2,t-1} + \sum_{p=0}^{k-1} w_{2,t-ph} \quad \text{and} \quad y_{2,t-lh} = y_{2,t-1} + \sum_{p=l}^{k-1} w_{2,t-ph}, \quad l = 1, \ldots, k-1,$$

from which it follows that

$$\delta_{2t} = \sum_{p=0}^{k-1} w_{2,t-ph} - \frac{1}{k}\sum_{l=0}^{k-1}\sum_{p=l}^{k-1} w_{2,t-ph} = s_2(L^h)w_{2t}$$

where $s_2(z)$ is defined in the statement of the Lemma and noting that $k^{-1} = h$ and

$$\sum_{l=0}^{k-1}\sum_{p=l}^{k-1} z^p = \sum_{l=0}^{k-1}(l+1)z^l.$$

Making these substitutions we obtain

$$\begin{aligned}
\zeta_{1t}^S &= u_{1t}^S + C_{SS}s(L^h)u_{2t}^S + \Big(1 + Ls_2(L^h)\Big)\, C_{SF}s(L^h)u_{2t}^F, \\
\zeta_{1t}^F &= hs(L^h)s(L^h)u_{1t}^F + \Big(hs(L^h) - Ls_2(L^h)\Big)\, C_{FS}s(L^h)u_{2t}^S + hC_{FF}s(L^h)^2u_{2t}^F,
\end{aligned}$$

while $\zeta_{2t}^S = s(L^h)u_{2t}^S$ and $\zeta_{2t}^F = hs(L^h)^2u_{2t}^F$. The representation stated in the Lemma then follows immediately. □

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

(a) $\frac{1}{T^2}\sum_{t=1}^{T} Y_{2,t-1}Y_{2,t-1-k}' \xrightarrow{d} \int_0^1 B_2B_2'$;

(b) $\frac{1}{T}\sum_{t=1}^{T} \xi_t Y_{2,t-1-k}' \xrightarrow{d} \int_0^1 dBB_2' + S_{2,k+1}$, *where* $S_{2,k} = \sum_{j=k}^{\infty} E(\xi_{t+j}\xi_{2t}')$;

(c) $\frac{1}{T^2}\sum_{t=1}^{T} z_{2,t-1}z_{2,t-1-k}' \xrightarrow{d} \int_0^1 b_2b_2'$;

(d) $\frac{1}{T}\sum_{t=1}^{T} \zeta_t z_{2,t-1-k}' \xrightarrow{d} \int_0^1 dbb_2' + S_{2,k+1}^{\zeta}$, *where* $S_{2,k}^{\zeta} = \sum_{j=k}^{\infty} E(\zeta_{t+j}\zeta_{2t}')$.

Proof of Lemma A3. The proofs are standard and follow from the FCLTs for ξ_t and ζ_t in Lemma 2 and Proposition 1; see, for example, Phillips and Durlauf (1986) for details. □

Lemma A4. *Let*

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

where $\lambda_j = 2\pi j/T$ $(j = -m, \ldots, m)$. *Then, under Assumption 2:*

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

(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

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

as $T\to\infty$ 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 $k>0$ we have, by similar steps as above,

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

and so it follows that

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

as $T\to\infty$ under Assumption 2. For $k=0$ we need to consider the additional term

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

as $T\to\infty$ 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

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

as $T\to\infty$ 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. □

References

Brockwell, P.J., Davis, R.A., 1991. *Time Series: Theory and Methods* (Second Edition). 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. *Journal of Econometrics* 193, 390–404.

Chambers, M.J., 2017. The estimation of continuous time cointegrated models with mixed frequency and mixed sample data. In preparation.

Corbae, D., Ouliaris, S., Phillips, P.C.B., 1994. A reexamination of the consumption function using frequency domain regressions. *Empirical Economics* 19, 595–609.

Corbae, D., Ouliaris, S., Phillips, P.C.B., 2002. Band spectral regression with trending data. *Econometrica* 70, 1067–1109.

Foroni, C., Ghysels, E., Marcellino, M., 2013. Mixed frequency vector autoregressive models. In: Fomby, T.B., Kilian, L., Murphy, A. (Eds.), *VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims, Advances in Econometrics, Volume 32*. Emerald Group Publishing Limited, Bingley.

Foroni, C., Marcellino, M., 2016. Mixed frequency structural VARs. *Journal of the Royal Statistical Society, Series A* 179, 403–425.

Fuller, W.A., 1996. *Introduction to Statistical Time Series* (Second Edition). Wiley, New York.

Ghysels, E., 2016. Macroeconomics and the reality of mixed frequency data. *Journal of 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.), *Essays in Honor of Peter C.B. Phillips, Advances in Econometrics, Volume 33*. Emerald Group Publishing Limited, Bingley.

Ghysels, E., Miller, J.I., 2015. Testing for cointegration with temporally aggregated and mixed-frequency time series. *Journal of Time Series Analysis* 36, 797–816.

Hamilton, J.D., 1994. *Time Series Analysis*. Princeton University Press, Princeton.

Hannan, E.J., 1970. *Multiple Time Series*. Wiley, New York.

Johansen, J., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.

Miller, J.I., 2010. Cointegrating regressions with messy regressors and an application to mixed-frequency series. *Journal of Time Series Analysis* 31, 255–277.

Miller, J.I., 2014. Mixed-frequency cointegrating regressions with parsimonious distributed lag structures. *Journal of Financial Econometrics* 12, 584–615.

Miller, J.I., 2016. Conditionally efficient estimation of long-run relationships using mixed-frequency time series. *Econometric Reviews* 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. *Journal of Time Series Analysis* 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.

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. *Review of Economic Studies* 53, 473–495.

Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99–125.

Robinson, P.M., 1972. Non-linear regression for multiple time-series. *Journal of Applied Probability* 9, 758–768.

Robinson, P.M., Marinucci, D., 2003. Semiparametric frequency domain analysis of fractional cointegration. In: Robinson, P.M. (Ed.), *Time Series with Long Memory.* Oxford University Press, Oxford.

Seong, B., Ahn, S.K., Zadrozny, P.A., 2013. Estimation of vector error correction models with mixed-frequency data. *Journal of Time Series Analysis* 34, 194–205.

Shiller, R.J., 2000. *Irrational Exuberance.* Princeton University Press, Princeton.

Stock, J.H., Watson, M.W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.

Table 1

RMSE ($\times 10^4$) of frequency domain estimators in VAR(1) model for $u_{\tau h}$, $T = 100$, $h = 1/12$

Model	OLS	FD1	FD2	FD3	FDA1	FDA2	FDA3	ASD1	ASD2	ASD3
$\Phi = 0_{2\times 2}$ (white noise)										
High	35.41	19.74	19.62	19.57	19.75	19.62	19.57	19.70	19.62	19.56
Low	402.64	119.56	111.21	139.45	113.71	109.73	135.83	109.31	109.75	131.81
Mixed	233.97	39.13	24.00	21.54	23.19	20.80	20.30	34.58	32.02	30.94
$\Phi = \Phi_1$										
High	22.28	19.82	19.73	19.71	19.82	19.73	19.71	19.83	19.75	19.74
Low	335.78	62.64	60.60	97.30	42.69	57.52	93.70	42.25	58.34	90.25
Mixed	190.83	40.87	24.51	21.45	23.22	20.78	20.29	47.45	43.72	41.24
$\Phi = \Phi_2$										
High	13.81	5.54	6.69	6.70	5.57	6.71	6.71	5.60	6.84	6.88
Low	308.54	52.53	54.82	94.28	21.21	51.50	91.37	24.01	52.55	87.31
Mixed	176.59	35.42	14.43	11.41	3.49	4.70	7.23	119.22	110.37	104.36
$\Phi = \Phi_3$										
High	11.25	5.71	6.82	6.82	5.65	6.79	6.80	5.76	7.00	7.05
Low	292.98	51.64	46.97	77.74	22.92	43.05	74.79	25.99	44.68	73.24
Mixed	161.39	34.40	14.43	8.77	3.57	4.81	7.30	55.63	52.28	53.79
$\Phi = \Phi_4$										
High	201.00	25.50	239.83	275.49	25.24	240.09	276.08	25.19	239.70	275.32
Low	638.15	552.00	494.56	487.55	568.99	494.85	486.33	545.23	494.37	482.29
Mixed	292.44	37.78	24.33	22.32	23.77	21.77	23.03	31.94	31.03	39.40
$\Phi = \Phi_5$										
High	344.76	25.04	238.03	274.06	24.62	237.40	273.51	24.59	237.75	273.66
Low	464.90	195.84	179.93	201.18	196.84	179.19	198.34	187.60	178.87	190.28
Mixed	286.98	38.40	24.01	25.39	23.61	21.22	22.56	28.97	26.84	28.17
$\Phi = \Phi_6$										
High	23.27	20.32	20.27	20.26	20.30	20.25	20.24	20.53	20.50	20.49
Low	204.75	63.33	48.63	57.45	29.06	42.90	54.91	42.10	51.13	60.07
Mixed	129.93	47.36	27.79	24.12	23.36	20.96	20.60	163.07	148.37	142.74

Table 2

Size and power of Wald tests in VAR(1) model for $u_{\tau h}$, $T = 100$, $h = 1/12$

Model	OLS	FD1	FDA1	ASD1	Model	OLS	FD1	FDA1	ASD1
		$\Phi = 0$ (white noise)					$\Phi = \Phi_2$		
		$C = 0.95$					$C = 0.95$		
High	100.00	100.00	100.00	100.00	High	100.00	100.00	100.00	100.00
Low	87.92	82.30	94.01	90.20	Low	91.03	92.20	100.00	99.28
Mixed	97.64	94.97	100.00	97.17	Mixed	97.23	97.91	100.00	91.45
		$C = 0.99$					$C = 0.99$		
High	99.34	92.31	92.73	92.50	High	99.98	100.00	100.00	100.00
Low	9.60	12.86	28.51	15.84	Low	7.16	37.18	99.37	76.82
Mixed	3.81	45.01	92.00	60.33	Mixed	2.27	52.09	100.00	57.47
		$C = 1.00$ (size)					$C = 1.00$ (size)		
High	5.30	0.75	0.81	0.58	High	38.88	1.08	1.32	7.06
Low	9.14	3.59	6.67	0.76	Low	10.66	1.60	7.21	0.13
Mixed	5.20	2.19	6.67	0.36	Mixed	8.48	1.49	7.07	3.57
		$C = 1.01$					$C = 1.01$		
High	81.25	92.71	93.07	92.92	High	99.99	100.00	100.00	100.00
Low	15.68	18.44	24.65	15.02	Low	20.65	43.92	92.69	71.53
Mixed	18.02	58.15	92.00	67.25	Mixed	25.48	66.20	100.00	62.25
		$C = 1.05$					$C = 1.05$		
High	99.99	100.00	100.00	100.00	High	100.00	100.00	100.00	100.00
Low	50.68	78.55	90.72	86.16	Low	59.77	89.56	100.00	98.58
Mixed	67.10	93.24	100.00	96.22	Mixed	76.59	97.21	100.00	93.53
		$\Phi = \Phi_4$					$\Phi = \Phi_6$		
		$C = 0.95$					$C = 0.95$		
High	92.84	99.98	100.00	100.00	High	100.00	100.00	100.00	100.00
Low	53.31	28.12	29.09	21.62	Low	91.22	97.30	100.00	99.51
Mixed	97.22	91.29	100.00	95.85	Mixed	97.51	99.39	100.00	92.63
		$C = 0.99$					$C = 0.99$		
High	51.76	73.16	84.10	84.74	High	99.53	97.79	97.81	91.85
Low	17.61	6.57	7.93	2.46	Low	7.29	39.35	93.00	76.37
Mixed	4.82	41.29	90.58	55.07	Mixed	4.77	51.18	92.03	43.62
		$C = 1.00$ (size)					$C = 1.00$ (size)		
High	80.18	0.16	1.06	1.11	High	72.75	11.15	11.04	1.50
Low	16.39	6.01	7.13	1.78	Low	19.26	2.25	7.46	4.14
Mixed	3.99	1.92	6.71	0.29	Mixed	20.41	2.50	7.13	1.26
		$C = 1.01$					$C = 1.01$		
High	100.00	86.32	96.66	96.91	High	99.89	98.00	97.99	92.22
Low	18.90	7.46	7.98	1.97	Low	35.48	45.00	83.21	75.90
Mixed	12.58	52.20	91.52	63.12	Mixed	43.98	69.05	92.21	50.60
		$C = 1.05$					$C = 1.05$		
High	100.00	99.99	100.00	100.00	High	100.00	100.00	100.00	100.00
Low	40.08	27.34	26.11	17.38	Low	78.93	97.40	99.97	99.73
Mixed	57.77	89.32	100.00	94.90	Mixed	91.94	99.77	99.99	98.97

Table 3

Empirical results using demeaned data, 1871–2016

	OLS	Spectral estimators			Augmented spectral estimators		
		$m = 4$	$m = 12$	$m = 32$	$m = 4$	$m = 12$	$m = 32$
Beginning-of-period stock price data							
$\hat{C}$	0.5179	0.5219	0.5187	0.5169	0.5224	0.5188	0.5169
	(0.0070)	(0.0056)	(0.0059)	(0.0066)	(0.0059)	(0.0062)	(0.0067)
Wald	4785.2169	228.4292	566.3292	1182.5369	201.9959	513.0848	1158.6059
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
End-of-period stock price data							
$\hat{C}$	0.5236	0.5269	0.5232	0.5228	0.5273	0.5232	0.5228
	(0.0076)	(0.0055)	(0.0059)	(0.0073)	(0.0058)	(0.0062)	(0.0073)
Wald	3935.4756	232.2694	565.3833	960.1324	205.2335	512.4292	940.6256
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Averaged stock price data							
$\hat{C}$	0.5212	0.5247	0.5217	0.5200	0.5251	0.5217	0.5200
	(0.0070)	(0.0054)	(0.0058)	(0.0067)	(0.0058)	(0.0061)	(0.0068)
Wald	4623.3407	236.8964	585.3597	1125.7540	209.3646	530.3569	1103.0700
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]

NB: standard errors in parentheses; marginal p-values in square brackets.

Table 4

Empirical results using detrended data, 1871–2016

	OLS	Spectral estimators			Augmented spectral estimators		
		$m = 4$	$m = 12$	$m = 32$	$m = 4$	$m = 12$	$m = 32$
Beginning-of-period stock price data							
$\hat{C}$	0.5314	0.5720	0.5655	0.5375	0.5726	0.5659	0.5377
	(0.0254)	(0.0201)	(0.0216)	(0.0240)	(0.0199)	(0.0216)	(0.0242)
Wald	340.6945	14.0226	34.6383	82.6871	14.1974	34.7241	81.0518
	[0.0000]	[0.0009]	[0.0000]	[0.0000]	[0.0008]	[0.0000]	[0.0000]
End-of-period stock price data							
$\hat{C}$	0.5166	0.5752	0.5668	0.5229	0.5760	0.5674	0.5230
	(0.0277)	(0.0201)	(0.0218)	(0.0264)	(0.0196)	(0.0215)	(0.0266)
Wald	304.0481	13.7937	33.8112	72.8259	14.4043	34.5473	71.3816
	[0.0000]	[0.0010]	[0.0000]	[0.0000]	[0.0007]	[0.0000]	[0.0000]
Averaged stock price data							
$\hat{C}$	0.5254	0.5720	0.5653	0.5314	0.5727	0.5658	0.5316
	(0.0257)	(0.0197)	(0.0213)	(0.0245)	(0.0194)	(0.0211)	(0.0247)
Wald	341.4540	14.5462	35.7437	81.4669	14.8820	36.0851	79.8545
	[0.0000]	[0.0007]	[0.0000]	[0.0000]	[0.0006]	[0.0000]	[0.0000]

NB: standard errors in parentheses; marginal p-values in square brackets.

Table 5

Empirical results using detrended data, 1871–2016, with break in 1942

		Spectral estimators			Augmented spectral estimators		
	OLS	$m = 4$	$m = 12$	$m = 32$	$m = 4$	$m = 12$	$m = 32$
Beginning-of-period stock price data							
$\hat{C}$	0.1961	0.0810	0.2189	0.1853	0.0795	0.2194	0.1846
	(0.0437)	(0.0418)	(0.0443)	(0.0405)	(0.0409)	(0.0448)	(0.0419)
Wald	338.2393	14.9317	26.5672	89.8763	15.6142	25.9452	84.4451
	[0.0000]	[0.0006]	[0.0000]	[0.0000]	[0.0004]	[0.0000]	[0.0000]
End-of-period stock price data							
$\hat{C}$	0.1239	0.1046	0.2171	0.1136	0.1044	0.2194	0.1129
	(0.0458)	(0.0417)	(0.0452)	(0.0430)	(0.0420)	(0.0451)	(0.0444)
Wald	365.1547	14.1929	25.6917	94.5647	14.0428	25.7014	88.8211
	[0.0000]	[0.0008]	[0.0000]	[0.0000]	[0.0009]	[0.0000]	[0.0000]
Averaged stock price data							
$\hat{C}$	0.1725	0.0975	0.2247	0.1622	0.0965	0.2259	0.1615
	(0.0433)	(0.0407)	(0.0438)	(0.0408)	(0.0405)	(0.0441)	(0.0421)
Wald	365.2956	15.1418	26.8000	94.0066	15.3443	26.3375	88.3082
	[0.0000]	[0.0005]	[0.0000]	[0.0000]	[0.0005]	[0.0000]	[0.0000]

NB: standard errors in parentheses; marginal p-values in square brackets.

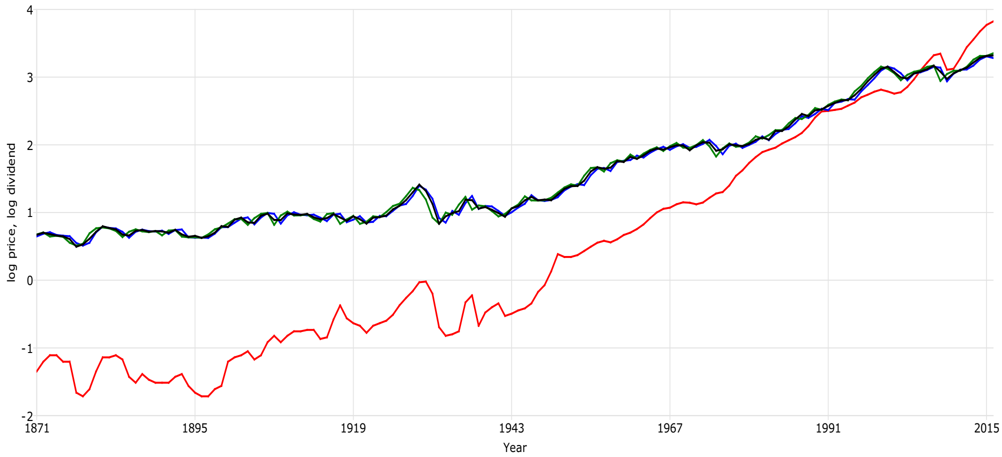

Figure 1. Logarithms of dividends and sampled stock prices, 1871–2016.

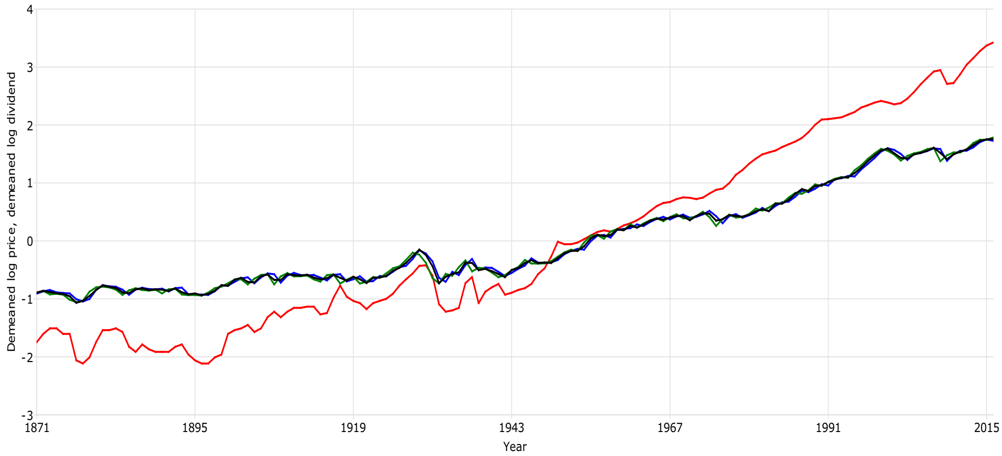

Figure 2. Demeaned logarithms of dividends and sampled stock prices, 1871–2016.

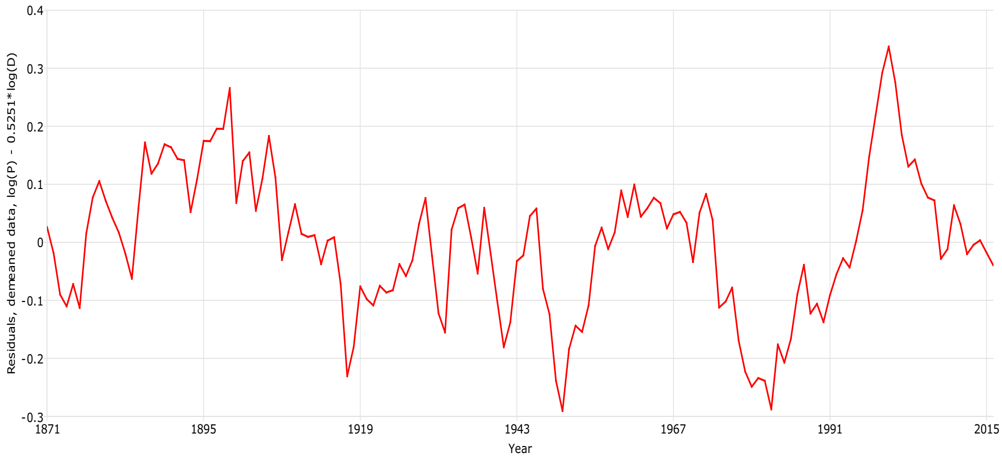

Figure 3. Residuals from cointegrating regression using demeaned data, 1871–2016.

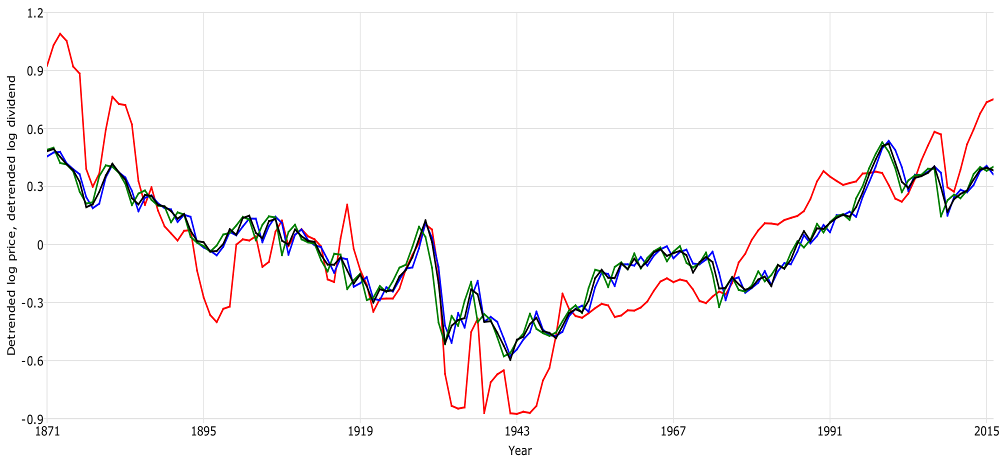

Figure 4. Detrended logarithms of dividends and sampled stock prices, 1871–2016.

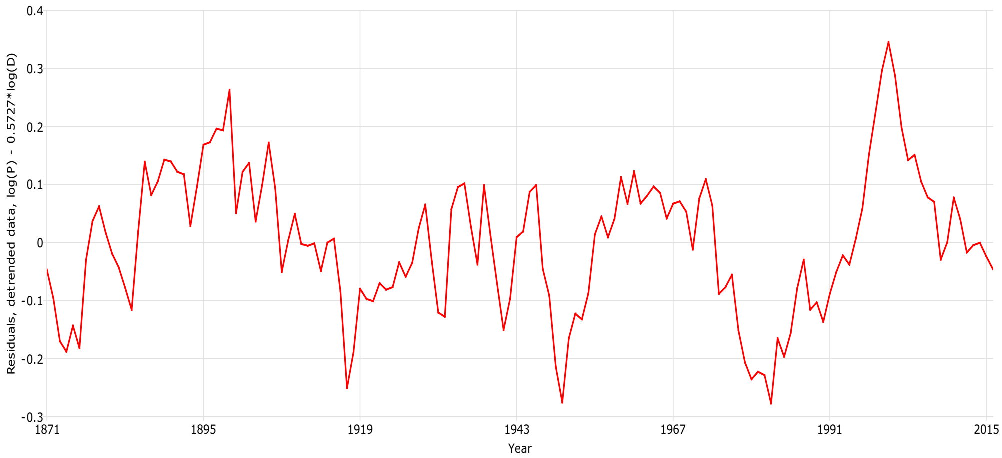

Figure 5. Residuals from cointegrating regression using detrended data, 1871–2016.

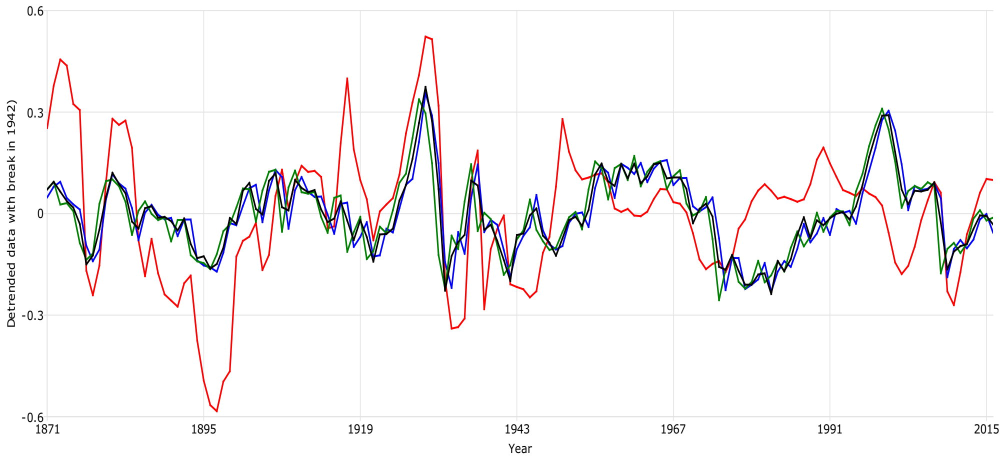

Figure 6. Detrended logarithms of dividends and sampled stock prices, 1871–2016, break in 1942.

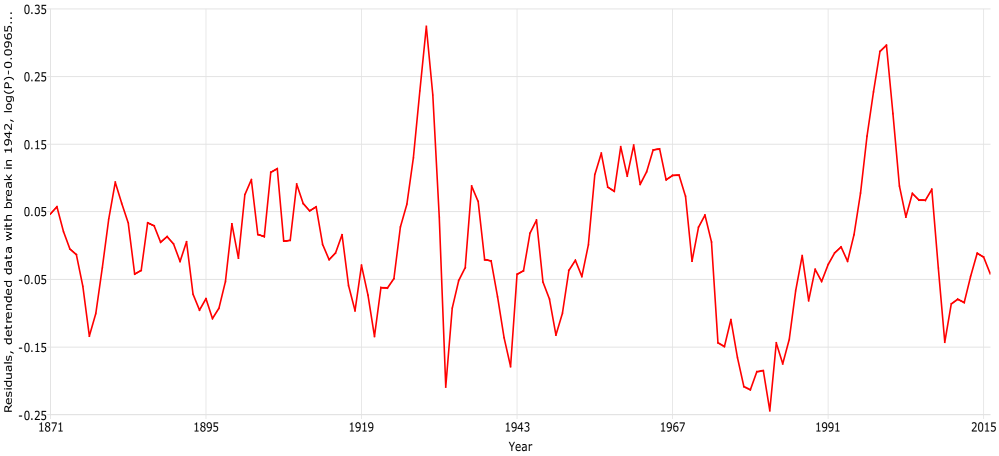

Figure 7. Residuals from cointegrating regression using detrended data, 1871–2016, break in 1942.