

# Frequency Domain Estimation of Continuous Time Cointegrated Models with Mixed Frequency and Mixed Sample Data

Marcus J. Chambers  
*University of Essex*

February 2019

## Abstract

Recent work by the author on mixed frequency data analysis has focused on the estimation of cointegrated systems in continuous time based on a fully specified dynamic system of equations, while the estimation of cointegrating vectors in a discrete time system has been approached using a semiparametric frequency domain estimator. We extend the latter approach to cover the continuous time case, establishing the asymptotic properties of the frequency domain estimator and explore, in a simulation study, the effects of misspecifying the continuous time dynamic model in discrete time compared to treating the dynamics nonparametrically. An empirical illustration is also provided.

**Keywords:** mixed frequency data; continuous time; frequency domain.

**Acknowledgements:** I am grateful to the Guest Co-Editor, Peter Zdrozny, for his support and encouragement, and to two anonymous referees for their helpful comments which have led to improvements in this paper. This research was funded in part by the Economic and Social Research Council under award number ES/M01147X/1.

**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

Time series models that embody cointegration have become a well established method for modelling nonstationary (integrated) variables that are subject to stationary long run relationships. A researcher's main focus of interest is often on the cointegrating vectors themselves which provide linear combinations of the variables that, in the most general sense, are merely assumed to be stationary. Applications of cointegration can be found in disciplines as diverse as economics, political science and climate science, and are typically motivated by some underlying theory of interest relevant to that discipline.

In addition to the stationary long run relationships attention also needs to be given to the dynamic evolution of the time series over the sample period, and there are two broad approaches to doing this. The first incorporates the long run cointegrating restrictions within a fully specified parametric dynamic model, a leading example being the cointegrated vector autoregression associated with Johansen (1991). This approach has found widespread application and is useful if modelling the dynamics alongside the cointegrating relationships is important, such as for the purposes of forecasting and impulse response analysis. An alternative, second, approach focuses solely on the cointegrating relationships and treats the system dynamics nonparametrically, either in the time domain or the frequency domain. An example of the former approach is the fully modified (FM) ordinary least squares estimator of Phillips and Hansen (1990) while an example of the latter approach is the spectral regression estimator of Phillips (1991a). A nonparametric treatment of the system dynamics is useful when the focus is purely on the cointegrating vectors and/or the researcher is unable or unwilling to specify a parametric model for the dynamics. The latter situation is not uncommon in economics in which economic theory is often informative about long run equilibrium relationships between variables but provides less (or no) guidance about the dynamic evolution towards equilibrium.

This paper is concerned with the estimation of cointegrating vectors in models formulated in continuous time and follows the second approach outlined above by treating the stationary system dynamics nonparametrically in the frequency domain. The observed sample is allowed to be comprised of a mixture of stock and flow variables as well as another feature often faced in empirical work, that of the presence of data observed at mixed frequencies. In particular it is also assumed throughout that the stock variables are observed at a smaller sampling interval (higher frequency),  $h$ , than the flow variables whose sampling interval is normalised (without loss of generality) to unity. Hence  $0 < h < 1$  and we also assume that  $k = h^{-1}$  is an integer so that there is a whole number of high frequency stock variable observations for each low frequency observation on flows. We make this assumption on the grounds that, in economics, high frequency financial variables are typically of the stock variety (e.g. stock prices, interest rates, and exchange rates) while lower frequency macroeconomic variables are often of the flow variety (e.g. consumption, income and investment). The methods and models developed below can also be extended to include high frequency flows and low frequency stocks as required.

The present paper builds on the discrete time cointegration approach with mixed sample and mixed frequency data in Chambers (2018) and also utilises the result of Chambers (2003) which shows that simple averaging of high frequency stock data to the low frequency can improve the asymptotic efficiency of cointegration estimators. The implications of more

general aggregation schemes for the estimation of cointegration vectors are obtained by Miller (2016). Alternative methods for estimating fully parametric cointegrated continuous time models with mixed sample and mixed frequency data can be found in Chambers (2016) and Thornton (2018); the present contribution can be considered as being complementary to those approaches.

The paper is organised as follows. Section 2 defines the continuous time cointegrated model which is driven by a stationary process that is required, for the subsequent purposes of deriving the asymptotic properties of the proposed estimator, to satisfy certain moment and mixing conditions. A key result (presented in Theorem 1) shows that the observed mixed frequency data, with the high frequency stocks averaged following the result in Chambers (2003), satisfy a triangular error correction model in discrete time at the low frequency. The precise relationship between the discrete time disturbances that drive this system and the underlying continuous time process is used to show that the former satisfy an invariance principle, based on the assumed properties of the latter. Two spectral regression estimators are defined in section 3, one involving a consistent first-step estimator that is used to construct a spectral density estimator using the resulting residuals, the other avoiding this first step by estimating an augmented regression involving the differences of the observed right-hand-side variables. Under appropriate conditions the estimators are shown to have familiar mixed normal limiting distributions and to belong to the class of optimal estimators as defined by Phillips (1991c). Some simulation results are presented in section 4 in which the spectral regression estimators are compared to the time domain FM estimator as well as an estimator that uses a parametric vector autoregression (VAR) for the discrete time disturbance process. Such a model is misspecified because it does not account for the temporal aggregation aspects in moving from continuous to discrete time; in effect, it approximates the true specification, which involves moving average (MA) components, with a pure VAR process. This section also includes an empirical illustration using the updated monthly stock price data and annual dividend data of Shiller (2000). The spectral regression and FM estimators, which treat the dynamics nonparametrically, provide more precise estimates of the cointegrating parameter than the estimates based on the parametric VAR approximation in that their standard errors are much smaller. Section 5 concludes, and the Appendix contains proofs of all lemmas and theorems as well as details of how the data are simulated in section 4.

The following notation is used throughout the paper. The operator  $D$  denotes the mean square differential operator such that, if  $x(t)$  is mean square differentiable, then there exists a process  $\xi(t)$  satisfying

$$\lim_{\delta \rightarrow 0} E \left( \frac{x(t + \delta) - x(t)}{\delta} - \xi(t) \right)^2 = 0,$$

in which case  $Dx(t) = \xi(t)$ . The lag operator will be denoted  $L$  such that  $L^j x_t = x_{t-j}$  for some  $j$ , not necessarily integer-valued. The  $n \times n$  identity matrix is denoted  $I_n$  while  $0_{n \times m}$  denotes a matrix of zeros of dimension  $n \times m$ . For a square matrix  $A$  the matrix exponential is defined by  $e^A = \sum_{j=0}^{\infty} A^j / j!$ ,  $\text{tr}(A)$  denotes the trace of  $A$ ,  $|A|$  denotes the determinant of  $A$ , and  $\|A\| = (\text{tr}(AA'))^{1/2}$  denotes the Euclidean norm of  $A$  where  $A'$  is the transpose of  $A$ . Finally, for a complex valued vector or matrix  $A$ ,  $A^*$  combines the operations of

transposition ( $A'$ ) and conjugation ( $\bar{A}$ ) so that  $A^* = \bar{A}'$ .

## 2. The model and mixed frequency data

We consider a model of cointegration for the  $n \times 1$  vector  $y(t) = (y_1(t)', y_2(t)')$  where  $y_1(t)$  and  $y_2(t)$  are  $n_1 \times 1$  and  $n_2 \times 1$ , respectively, and  $n_1 + n_2 = n$ . The cointegrating relationships are normalised on the  $n_1 \times 1$  vector  $y_1(t)$  so that

$$y_1(t) = Cy_2(t) + u_1(t), \quad t > 0, \quad (1)$$

$$Dy_2(t) = u_2(t), \quad t > 0, \quad (2)$$

where  $C$  is the  $n_1 \times n_2$  matrix of cointegrating vectors and  $u(t) = (u_1(t)', u_2(t)')$  is a stationary disturbance vector. In this model the cointegrating relationships are captured in (1) and the deviations from ‘equilibrium’, the elements of  $u_1(t) = y_1(t) - Cy_2(t)$ , are stationary processes. Furthermore, the stochastic trends in the system, depicted in (2), are also allowed to be driven by stationary processes which are the elements of  $u_2(t)$ . For subsequent purposes it is assumed that  $u(t)$  satisfies:

**Assumption 1.** The  $n \times 1$  vector  $u(t)$  is a strong mixing continuous time random process satisfying:

- (a)  $Eu(t) = 0$ ;
- (b)  $E|u_i(t)|^\beta < \infty$  ( $i = 1, \dots, n$ ) for some  $\beta > 2$ ;
- (c) the strong mixing coefficients satisfy  $\alpha(s) = O(s^{-\mu})$  for some  $\mu > \beta/(\beta - 2)$ , where, for positive real  $s$ ,  $\alpha(s) = \sup_t \alpha(\mathcal{F}_{-\infty}^t, \mathcal{F}_{t+s}^\infty)$ ,

$$\alpha(\mathcal{F}_{-\infty}^t, \mathcal{F}_{t+s}^\infty) = \sup_{G \in \mathcal{F}_{-\infty}^t, H \in \mathcal{F}_{t+s}^\infty} |\Pr(G \cap H) - \Pr(G)\Pr(H)|,$$

and  $\mathcal{F}_a^b$  denotes the sigma-field generated by  $u(t)$  for  $a \leq t \leq b$ ;

- (d) the spectral density matrix  $f_{uu}(\lambda)$  is Hermitian positive semi-definite and is continuous and bounded for all  $-\infty < \lambda < \infty$  with  $f_{uu}(0)$  positive definite.

Assumption 1 is stronger than is required for the derivation of the mixed frequency representation for the discrete time observations but is used in the derivation of the asymptotic properties of the spectral regression estimator in the next section. The moment and mixing conditions in (b) and (c) are used to establish that the disturbances in the discrete time mixed frequency representation satisfy a functional central limit theorem (FCLT). In particular the rate condition on  $\alpha(s)$  in (c) ensures that the mixing coefficients are integrable, this being the requirement for the continuous time process that is analogous to the summability of these coefficients for discrete time processes. Assumption 1 ensures that the disturbances in the discrete time model used for estimation are also covariance stationary.

As shown by Phillips (1991b, p.970) the vector  $y(t)$  satisfies the stochastic differential equation system

$$Dy(t) = -J Ay(t) + w(t), \quad t > 0, \quad (3)$$

where

$$A = (I_{n_1}, -C), \quad J = \begin{pmatrix} I_{n_1} \\ 0_{n_1 \times n_2} \end{pmatrix}, \quad w(t) = \begin{pmatrix} u_1(t) + Du_1(t) + Cu_2(t) \\ u_2(t) \end{pmatrix};$$

note that  $w(t)$  is a stationary continuous time process. The representation in (3) is in the form of a continuous time error correction model which is, perhaps, best seen by extracting the equation for  $y_1(t)$ , which is

$$Dy_1(t) = -[y_1(t) - Cy_2(t)] + w_1(t)$$

and which depicts  $y_1$  responding to the disequilibrium error  $y_1 - Cy_2$ .

In discrete time, the vector  $y(t)$  satisfies the stochastic difference equation system

$$y(t) = e^{-JA}y(t-1) + \int_{t-1}^t e^{-JA(t-r)}w(r)dr;$$

see Phillips (1991b) and Chambers (2003). By noting that  $AJ = I_{n_1}$  it can be shown that  $e^{-JA} = I_n - (1 - e^{-1})JA$  and so the discrete time representation has the error correction form

$$\Delta y(t) = -JAy(t-1) + v(t), \tag{4}$$

where the disturbance vector is defined by

$$\begin{aligned} v(t) &= e^{-1}JAy(t-1) + \int_{t-1}^t e^{-JA(t-r)}w(r)dr \\ &= \begin{pmatrix} u_1(t) + C \int_{t-1}^t u_2(r)dr \\ \int_{t-1}^t u_2(r)dr \end{pmatrix}; \end{aligned} \tag{5}$$

the first expression for  $v(t)$  in (5) comes from equation (9) in Phillips (1991b) while the second is established in equation (5) and Lemma A1 of Chambers (2003). The discrete time representation in (4) can be adapted to any sampling frequency or any type of sampling (stock and/or flow) and underlies the mixed frequency representation derived below.

The model (3) may appear to be overly restrictive but, as shown by Phillips (1991b), it is also consistent with more general models which can include deterministic terms as well as a matrix of speed-of-adjustment coefficients. In the former case let  $\kappa(t) = \sum_{j=0}^p \kappa_j t^j$  denote a vector of deterministic terms so that (3) becomes

$$Dy(t) = \kappa(t) - JAy(t) + w(t), \quad t > 0.$$

Assuming  $C\kappa_j = 0$  ( $j = 0, \dots, p$ ), so that (1) remains valid and  $C$  annihilates both the deterministic and stochastic trends, the corresponding discrete time representation becomes

$$\Delta y(t) = k(t) - JAy(t-1) + v(t),$$

where  $k(t) = \sum_{j=0}^p k_j t^j$  and the  $k_j$  are functions of the underlying  $\kappa_j$ .<sup>1</sup> Provided the deterministic terms are removed from the data by prior regression the estimation methods in the next section remain valid with the Brownian motions in the limit distributions replaced by suitably detrended versions. When the model is formulated directly as an error correction system in continuous time with a matrix of speed-of-adjustment coefficients (say  $\Gamma$ ) we may write the system as

$$Dy(t) = \Gamma Ay(t) + e(t), \quad t > 0.$$

where  $e(t)$  is a stationary continuous time process. It is straightforward to show that this system is consistent with (3) in which the stationary disturbance vector becomes  $w(t) = e(t) + (\Gamma + J)Ay(t)$ . In this case, because our focus is on estimation of  $C$  alone, it is valid to assign any such dynamic features to the stationary disturbance,  $w(t)$ , that drives the system; these dynamic features are treated nonparametrically in the spectral regression procedures outlined in the next section.

In terms of the nature of the elements of  $y_1(t)$  and  $y_2(t)$  we assume that each vector is comprised of both stock and flow variables, so that

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

where  $y_j^S(t)$  is  $n_j^S \times 1$ ,  $y_j^F(t)$  is  $n_j^F \times 1$  and  $n_j^S + n_j^F = n_j$  ( $j = 1, 2$ ). The aim is to derive a model that is satisfied by discrete time mixed frequency data generated by the system (1) and (2). As stated in the Introduction, the stock variables are assumed to be observed at a higher frequency (corresponding to a sampling interval  $0 < h < 1$ ) than the flow variables, whose sampling interval is normalised to unity. We use  $t = 1, \dots, T$  to index the low frequency flow variables, the observations for which are therefore

$$Y_{1t}^F = \int_{t-1}^t y_1^F(r) dr, \quad Y_{2t}^F = \int_{t-1}^t y_2^F(r) dr, \quad t = 1, \dots, T.$$

The high frequency stock variables are also observed at each integer  $t$  as well as the intermediate points, at intervals of length  $h$ , between  $t$  and  $t - 1$  so that there are  $kT = T/h = N$  high frequency observations given by

$$y_{1\tau}^S = y_1^S(\tau h), \quad y_{2\tau}^S = y_2^S(\tau h), \quad \tau = 1, \dots, N.$$

High frequency observations are, therefore, available at the points  $t - (k - 1)h$  through to  $t - h$  that lie between  $t - 1$  and  $t$ , and it is these intermediate observations that can be exploited in the pursuit of improved estimation and inference procedures.

There are a number of ways in which mixed frequency data have been used in the literature. In a continuous time setting Chambers (2016) derived exact discrete models corresponding to a system of first-order stochastic differential equations<sup>2</sup> and incorporates

<sup>1</sup>See Phillips (1991b, p.978) for details.

<sup>2</sup>This approach has been extended by Thornton (2018) to a system of autoregressive moving average equations in continuous time.

the mixed frequency observations within the vectors

$$z_{1t} = \begin{pmatrix} y_{1t}^S \\ y_{1,t-h}^S \\ \vdots \\ y_{1,t-(k-1)h}^S \\ Y_{1t}^F \end{pmatrix}, \quad z_{2t} = \begin{pmatrix} y_{2t}^S \\ y_{2,t-h}^S \\ \vdots \\ y_{2,t-(k-1)h}^S \\ Y_{2t}^F \end{pmatrix}, \quad t = 1, \dots, T.$$

Both  $z_{1t}$  and  $z_{2t}$  contain observations dated  $t$  but contain, in addition, the high frequency observations at the points  $t - (k - 1)h$  through to  $t - h$ . The exact discrete models in Chambers (2016) are in the form of first-order stochastic difference equations in the vector  $z_t = (z'_{1t}, z'_{2t})'$ , which is of dimension  $(kn_1^S + n_1^F + kn_2^S + n_2^F) \times 1$ . Alternatively Ghysels (2016) defines mixed frequency VARs directly in terms of  $z_t$ , but such direct discrete time VAR specifications tend to be considerably over-parameterised compared to the continuous time approach in which the underlying model of interest is in terms of the  $n \times 1$  vector  $y(t)$  and which contains fewer unknown parameters.

The above approach based on the vectors  $z_{1t}$  and  $z_{2t}$  is not appropriate, however, in the spectral regression setting pursued here. As pointed out by Chambers (2018) the spectral density matrices of these vectors are singular. For example, consider the spectral density matrix of  $\tilde{y}_t^S = (y_{1t}^S, y_{1,t-h}^S)'$ ; if  $f_y^h(\lambda)$  denotes the spectrum of  $y_t$  (as a high frequency process) then the spectrum of  $\tilde{y}_t^S$  is given by

$$f_{y^S}(\lambda) = \begin{pmatrix} 1 \\ e^{-ih\lambda} \end{pmatrix} f_y^h(\lambda) (1 \ e^{ih\lambda}) = \begin{pmatrix} 1 & e^{ih\lambda} \\ e^{-ih\lambda} & 1 \end{pmatrix} f_y^h(\lambda), \quad -\frac{\pi}{h} < \lambda \leq \frac{\pi}{h},$$

which is clearly singular at all frequencies. To overcome this problem Chambers (2018) exploits the suggestion in Chambers (2003) to aggregate the high frequency stock variables by constructing period averages at the low frequency. This recommendation was based on asymptotic efficiency considerations concerning estimators of cointegrating vectors but the simulations reported in Chambers (2018) suggest this procedure works well in the mixed frequency setting, resulting in better finite sample performance than estimates obtained using only the low frequency observations. We adopt this aggregation approach in the current setting of a continuous time model using the low frequency averages of the high frequency stock variables. It is useful to define the operator

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

so that the aggregated/averaged variables are given by

$$Y_{jt}^S = \frac{1}{k} s(L^h) y_{jt}^S = \frac{1}{k} \sum_{l=0}^{k-1} y_{j,t-lh}^S, \quad j = 1, 2, \quad t = 1, \dots, T. \quad (6)$$

The task is to relate  $Y_{1t}^S$  and  $Y_{1t}^F$  to  $Y_{2t}^S$  and  $Y_{2t}^F$  subject to them having been generated by the system (1) and (2). A key result in facilitating this is contained in Lemma A1 in the Appendix, the use of which enables discrete time averages of the unobservable high frequency

flow process to be replaced by the observable low frequency flow variable with the stationary difference<sup>3</sup> assigned to the disturbance vector. For the presentation of the discrete time representation it is also convenient to define the vectors

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

the first two are observable, the last two are unobservable. The exact discrete time model is given in the following Theorem.

**Theorem 1.** *Let  $y_1(t)$  and  $y_2(t)$  be generated by (1) and (2). Then the observations satisfy, for  $t = 1, \dots, T$ ,*

$$Y_{1t} = CY_{2,t-1} + \xi_{1t}, \quad (7)$$

$$\Delta Y_{2t} = \xi_{2t}, \quad (8)$$

where  $\xi_{1t}$  and  $\xi_{2t}$  are stationary disturbance processes under Assumption 1.

The discrete time representation in Theorem 1 is useful from the point of view of estimation because, even with mixed frequency and mixed sample data, taking into account the temporal aggregation issues still leads to the same type of linear (in  $C$ ) system as in the direct discrete time approach. All of the associated dynamics that stem from the process  $u(t)$  in the continuous time system are assigned to the disturbance vector  $\xi_t$ , and this is a major reason why spectral regression methods are particularly useful in this setting. The following result plays an important role in deriving the properties of the spectral estimators.

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

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \xi_t \xrightarrow{d} B(r), \quad 0 < r \leq 1, \quad (9)$$

where  $B(r)$  is a Brownian motion process with covariance matrix  $\Omega$ .

The key to establishing Lemma 1 lies in utilising the precise relationship between  $\xi_t$  and  $u(t)$  (that arises in the proof of Theorem 1) and then demonstrating that the properties of  $u(t)$  in Assumption 1 ensure that  $\xi_t$  satisfies the conditions for the FCLT to hold. In what follows it is convenient to partition  $B(r)$  and  $\Omega$  in accordance with  $\xi_1$  and  $\xi_2$  in the form

$$B(r) = \begin{pmatrix} B_1(r) \\ B_2(r) \end{pmatrix}, \quad \Omega = (\Omega_1 \quad \Omega_2) = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix},$$

and to define  $\Omega_{11.2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$ .

### 3. Estimation in the frequency domain

The model of cointegration developed in the previous section is driven by a disturbance vector,  $\xi_t$ , that is stationary under Assumption 1. As initially suggested by Phillips (1991b)

---

<sup>3</sup>This stationary difference is denoted  $\delta_t$  in Lemma A1.



a natural approach to estimating the matrix  $C$  of cointegrating vectors is, therefore, to use spectral/frequency domain regression. It is convenient to write the system of interest, defined in (7) and (8), in the form

$$Y_{0t} = JCY_{2,t-1} + \xi_t, \quad t = 1, \dots, T, \quad (10)$$

where  $Y_{0t} = (Y'_{1t}, \Delta Y'_{2,t})'$ . As outlined in Chambers (2018) for the discrete time case the spectral regression approach is based on taking discrete Fourier transforms (dFTs) in (10), yielding

$$w_0(\lambda_s) = JCw_2(\lambda_s) + w_\xi(\lambda_s), \quad s = -T/2 + 1, \dots, T/2, \quad (11)$$

where  $\{\lambda_s = 2\pi s/T; s = -T/2 + 1, \dots, T/2\}$  denotes the set of Fourier frequencies,  $T$  is assumed to be an even number for convenience,<sup>4</sup> 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  $\xi_t$ , respectively, at the Fourier frequencies.

Although the spectral regression estimator of  $C$  can be constructed using a regression across all  $T$  Fourier frequencies there are compelling reasons, in the case of cointegration, to focus solely on frequencies close to zero. From a theoretical viewpoint cointegration is a long-run phenomenon whose properties are defined in the frequency domain at the zero frequency, while simulation evidence contained in Corbae, Ouliaris and Phillips (1994) and Chambers (2018) suggest that such band-limited spectral methods work particularly well in practice. We therefore consider the symmetric set of frequencies  $\Lambda_0 = \{\lambda_s = 2\pi s/T; s = -m, \dots, m\}$  which contains the  $2m + 1$  Fourier frequencies around the origin for some integer  $m$ . A family of generalised least squares-type objective functions is then given by

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

where  $w_\xi(\lambda_s) = w_0(\lambda_s) - JCw_2(\lambda_s)$  and  $\Phi(\lambda_s)$  is a positive definite Hermitian weighting matrix. As shown 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  $\xi_t$ .

Although  $\xi_t$  is unobserved a consistent estimator of  $f_{\xi\xi}(\lambda)$  can nevertheless be obtained by using the residuals from a least squares regression of (10), denoted  $\hat{\xi}_t$ . The spectral density matrix of interest can then be estimated in a variety of ways, and the method we shall employ here is the smoothed periodogram estimator, defined by

$$\hat{f}_{\hat{\xi}\hat{\xi}}(0) = \frac{1}{2m + 1} \sum_{j=-m}^m I_{\hat{\xi}\hat{\xi}}(\lambda_j), \quad (12)$$

where  $I_{\hat{\xi}\hat{\xi}}(\lambda) = w_{\hat{\xi}}(\lambda) w_{\hat{\xi}}(\lambda)^*$  denotes the periodogram of  $\hat{\xi}_t$  and  $w_{\hat{\xi}}(\lambda)$  is the dFT of  $\hat{\xi}_t$ . The smoothed periodogram estimator is a straightforward symmetric average of  $2m + 1$

<sup>4</sup>If  $T$  is odd then we can take  $-[T/2] + 1 \leq s \leq [T/2]$ .

periodogram matrices around the frequency of interest, here being the zero frequency. More sophisticated estimates could be used but the smoothed periodogram performed well in the simulations reported in Chambers (2018). With this choice of weighting matrix the objective function becomes

$$S\left(C; \hat{f}_{\xi\xi}(0)^{-1}\right) = \frac{1}{2m+1} \sum_{s=-m}^m \text{tr} \left\{ \hat{f}_{\xi\xi}(0)^{-1} (w_0(\lambda_s) - JCw_2(\lambda_s)) (w_0(\lambda_s) - JCw_2(\lambda_s))^* \right\}. \quad (13)$$

Minimisation of (13) with respect to  $C$  results in the estimator

$$\hat{C}_0 = \left( J' \hat{f}_{\xi\xi}(0)^{-1} J \right)^{-1} J' \hat{f}_{\xi\xi}(0)^{-1} \hat{f}_{02}(0) \hat{f}_{22}(0)^{-1} \quad (14)$$

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

$$\begin{aligned} \hat{f}_{02}(0) &= \frac{1}{2m+1} \sum_{j=-m}^m I_{02}(\lambda_j), \quad I_{02}(\lambda_j) = w_0(\lambda_j) w_2(\lambda_j)^*, \\ \hat{f}_{22}(0) &= \frac{1}{2m+1} \sum_{j=-m}^m I_{22}(\lambda_j), \quad I_{22}(\lambda_j) = w_2(\lambda_j) w_2(\lambda_j)^*, \end{aligned}$$

respectively.

Although  $\hat{C}_0$  is a full system 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 (10) or (11). The augmented equation includes  $\Delta Y_{2t}$  (or its dFt) as an additional regressor vector, resulting in the time domain regression equation

$$Y_{1t} = CY_{2,t-1} + \Omega_{12}\Omega_{22}^{-1}\Delta Y_{2t} + \xi_{1.2t}, \quad t = 1, \dots, T, \quad (15)$$

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

$$w_1(\lambda_s) = Cw_2(\lambda_s) + \Omega_{12}\Omega_{22}^{-1}w_{\Delta_2}(\lambda_s) + w_{1.2}(\lambda_s), \quad s = -T/2 + 1, \dots, T/2, \quad (16)$$

where  $w_1(\lambda_s)$ ,  $w_{\Delta_2}(\lambda_s)$  and  $w_{1.2}(\lambda_s)$  are the dFts of  $Y_{1t}$ ,  $\Delta Y_{2t}$  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

$$S_A(C) = \frac{1}{2m+1} \sum_{s=-m}^m \text{tr} \{ w_{1.2}(\lambda_s) w_{1.2}(\lambda_s)^* \}, \quad (17)$$

where  $w_{1.2}(\lambda_s) = w_1(\lambda_s) - Cw_2(\lambda_s) - \Omega_{12}\Omega_{22}^{-1}w_{\Delta_2}(\lambda_s)$ . The resulting estimator can be written in the form

$$\hat{C}_0^A = \left( \hat{f}_{12}(0) - \hat{f}_{1\Delta_2}(0) \hat{f}_{\Delta_2\Delta_2}(0)^{-1} \hat{f}_{\Delta_2 1}(0) \right) \left( \hat{f}_{22}(0) - \hat{f}_{2\Delta_2}(0) \hat{f}_{\Delta_2\Delta_2}(0)^{-1} \hat{f}_{\Delta_2 2}(0) \right)^{-1}, \quad (18)$$

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

In deriving the asymptotic properties of  $\hat{C}_0$  and  $\hat{C}_0^A$  an assumption concerning the number,  $m$ , of frequencies employed in the estimation of the relevant spectral density matrices is also required, in addition to the FCLT presented in Lemma 1. For this purpose we make the following assumption.

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

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, which strengthens the moment and mixing conditions of Assumption 1, is also employed.

**Assumption 1'.** The  $n \times 1$  vector  $u(t)$  satisfies Assumption 1 but with (b) and (c) replaced by, respectively:

$$(b') \ E|u_i(t)|^{\beta'} < \infty \ (i = 1, \dots, n) \text{ for some } \beta' > 4;$$

$$(c') \ \text{the strong mixing coefficients satisfy } \alpha(s) = O(s^{-\mu'}) \text{ for some } \mu' > 3\beta' / (\beta' - 4).$$

Assumption 1' is used to establish a consistency result concerning  $\hat{f}_{\xi\xi}(0)$  which is used in the proof of Theorem 1(c) below. In particular it ensures the following result.

**Lemma 2.** Let  $\Gamma_{\xi,k} = E(\xi_t \xi_{t-k}')$  and let  $\kappa_{abcd}(0, j, k, l)$  ( $a, b, c, d = 1, \dots, n$ ) denote the fourth-order cumulant function of the elements of the vectors  $\xi_t, \xi_{t+j}, \xi_{t+k}, \xi_{t+l}$ . Then, under Assumption 1',

$$\sum_{k=-\infty}^{\infty} \|\Gamma_{\xi,k}\| < \infty$$

and

$$\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |\kappa_{abcd}(0, j, k, l)| < \infty, \quad a, b, c, d = 1, \dots, n.$$

The absolute summability of the autocovariances ensures that the spectral density of  $\xi_t$  is continuous and bounded while the condition on the fourth cumulants is used to establish a result concerning the variance of a spectral density estimator. The use of Assumptions 1, 1' and 2 enables the following result concerning the asymptotics of the smoothed periodogram estimators of spectral density matrices to be established.

**Theorem 2.** Let  $y_1(t)$  and  $y_2(t)$  be generated by (1) and (2). Then, under Assumptions 1 and 2, as  $T \rightarrow \infty$ :

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

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

If Assumptions 1' and 2 are satisfied, then

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

The first two parts of Theorem 2 follow from the results in Chambers (2018). Part (c), however, requires a slightly different method of proof owing to the different (moment and mixing) assumptions used here. The asymptotic distributions of  $\hat{C}_0$  and  $\hat{C}_0^A$  can now be stated.

**Theorem 3.** *Under Assumptions 1' and 2, as  $T \rightarrow \infty$ ,*

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

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

Both estimators therefore belong to the class of optimal estimators as defined by Phillips (1991c). These are estimators having the form of limit distribution as given in Theorem 3 i.e. mixed normal. An advantage of such optimal estimators is that their mixed normal limiting distributions enable traditional asymptotic chi-square hypothesis testing in appropriate circumstances. Such hypothesis testing requires (consistent) estimators of the covariance matrices of  $\hat{C}_0$  and  $\hat{C}_0^A$ . Define  $\hat{\gamma}_0 = \text{vec}(\hat{C}_0)$  and  $\hat{\gamma}_0^A = \text{vec}(\hat{C}_0^A)$ , where the  $\text{vec}(\cdot)$  operator stacks the columns of a matrix vertically. A consistent estimator of the covariance matrix of  $\hat{\gamma}_0$  is given by

$$\hat{V}_0 = \frac{1}{T} \left[ \hat{f}_{22}(0) \otimes J' \hat{f}_{\xi\xi}(0)^{-1} J \right]^{-1},$$

while a consistent estimator of the covariance estimator of  $\hat{\gamma}_0^A$  is given by

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

where  $\hat{f}_{11.2}(0)$  is a consistent estimator of  $f_{11.2}(0) = \Omega_{11.2}/2\pi$ , such as the smoothed periodogram estimator based on the residuals from (15). Chambers (2018) establishes the asymptotic chi-square distributions of Wald statistics based on  $\hat{C}_0$  and  $\hat{C}_0^A$ ; the same arguments apply here for Wald tests of hypotheses concerning the parameters of the continuous time cointegrating vectors.

#### 4. Simulation results and an empirical illustration

In this section we begin by exploring the finite sample properties of the spectral regression estimators and compare them to some alternative estimators of the cointegrating parameters. In particular we consider the FM estimator of Phillips and Hansen (1990), which treats the dynamics nonparametrically in the time domain, as well as VAR approximations to  $\xi_t$ . In the latter approach the discrete time model is misspecified because it is not derived from the underlying continuous time system and ignores the effects of temporal aggregation. As such, this approach attempts to approximate the dynamics parametrically compared to the spectral regression and FM approaches which do so nonparametrically.

The simulations are based on a bivariate cointegrated system in continuous time of the form given in (1) and (2) with cointegrating parameter  $C = 1$ . It is assumed that  $y_1$  is observed as a high frequency stock variable and  $y_2$  as a low frequency flow variable which

results in the sequences of observations of the form

$$\{y_{1\tau} = y_1(\tau h); \tau = 1, \dots, N\}, \quad \left\{ Y_{2t} = \int_{t-1}^t y_2(r) dr; t = 1, \dots, T \right\},$$

where  $T = Nh$ . The high frequency observations are aggregated as in section 2 to yield the sequence

$$\left\{ Y_{1t} = \frac{1}{k} \sum_{l=0}^{k-1} y_{1,t-lh}; t = 1, \dots, T \right\}.$$

We set the span  $T = 100$  and the high frequency sampling interval to be  $h = 1/12$  (so that  $k = 12$ ) which implies that there are  $N = 1200$  high frequency observations. This sampling scheme would therefore correspond to 100 years of monthly data and is motivated by the empirical example that follows which is based on monthly stock price data and yearly dividend data. A total of 10,000 replications of each experiment are conducted.

The model specification is completed by a particular form of dynamic model for the stationary bivariate continuous time vector  $u(t) = (u_1(t), u_2(t))'$ . The first specification is that  $u(t)$  is white noise<sup>5</sup> while the second is that  $u(t)$  satisfies the stationary stochastic differential equation system

$$du(t) = \Phi u(t) dt + \zeta(dt), \quad t > 0, \quad (19)$$

subject to the initial condition  $u(0) = (0, 0)'$ , where the roots of the equation  $|zI_2 - \Phi| = 0$  have negative real parts to ensure stationarity and  $\zeta(dt)$  is a bivariate uncorrelated random measure with covariance matrix  $\Sigma dt$ ; see Bergstrom (1984) for a definition of random measures and their properties. We set  $\Sigma = I_2$  and  $\Phi = \phi I_2$  with  $\phi \in \{-5, -10\}$  so that  $u_1$  and  $u_2$  are independent autoregressive (AR) processes each with AR coefficient  $e^{\phi h} \in \{0.6592, 0.4346\}$  at the high frequency. The data are generated at the high frequency and then aggregated appropriately; details are given in the appendix. To summarise,  $\xi_t$  is a bivariate MA(1) process when  $u(t)$  is white noise (owing to the temporal aggregation) and a bivariate ARMA(1,1) process when  $u(t)$  is a continuous time autoregression.

The estimators we consider are the OLS estimator of  $C$  in (7); the spectral regression estimators  $\hat{C}_0$  and  $\hat{C}_0^A$ , denoted FD and FDA, respectively; the FM estimator of  $C$ ; and frequency domain estimators of  $C$  based on a VAR specification for  $\xi_t$  (which is misspecified). The FD and FDA estimators are obtained using  $m = [T^{0.7}] = 25$  periodogram ordinates in estimating the spectral density matrices; other values were explored but this rule worked well in accordance with the simulation results in Chambers (2018). The FM estimator uses the Parzen kernel and the automatic bandwidth selection procedure of Andrews (1991) and is denoted FM(Auto). The approximating VAR( $p$ ) models are of the form

$$G(L)\xi_t = e_t, \quad t = 1, \dots, T,$$

where  $G(z) = I_2 - \sum_{j=1}^p G_j z^j$  and  $e_t$  is white noise with variance matrix  $\Sigma_e$ . Three values of  $p - 1, 2$  and  $4$  – are considered, the resulting estimators being denoted VAR(1), VAR(2) and VAR(4). Let  $I_{\xi\xi}(\lambda) = w_\xi(\lambda)w_\xi(\lambda)^*$  where  $w_\xi(\lambda)$  is defined in section 3 and is a function of

<sup>5</sup>In this case it is perhaps more natural to write the system in the form  $(y_1(t) - Cy_2(t))dt = \zeta_1(dt)$  and  $dy_2(t) = \zeta_2(dt)$ , where  $\zeta_1(dt)$  and  $\zeta_2(dt)$  are uncorrelated random measures.

$C$ . Then, following Dunsmuir (1979), we use the following version of the frequency domain (Whittle) Gaussian likelihood function:<sup>6</sup>

$$L_1(C, \theta_G, \Sigma_e) = \log |\Sigma_e| + \frac{1}{T} \sum_{j \in \Lambda} \text{tr} \{ f_{\xi\xi}(\lambda_j)^{-1} I_{\xi\xi}(\lambda_j) \},$$

where  $\Lambda = \{\lambda_s = 2\pi s/T; s = -T/2 + 1, \dots, T/2\}$  denotes the set of Fourier frequencies,  $\theta_G$  denotes the vector of unknown parameters in  $G(z)$ , and  $f_{\xi\xi}(\lambda)$  is the spectral density matrix of  $\xi_t$  given by

$$f_{\xi\xi}(\lambda) = \frac{1}{2\pi} G(e^{-i\lambda})^{-1} \Sigma_e \left[ G(e^{i\lambda})' \right]^{-1}, \quad -\pi < \lambda \leq \pi;$$

a similar form of objective function was used by Chambers and McCrorie (2007) in their frequency domain approach to the estimation of fully parametric continuous time cointegrated systems. Significant computational advantages can be gained by concentrating  $\Sigma_e$  out of the likelihood function, resulting in (ignoring a constant)

$$L(C, \theta_G) = L_1(C, \theta_G, \hat{\Sigma}_e) = \log \left| \frac{2\pi}{T} \sum_{j \in \Lambda} G(e^{-i\lambda_j}) I_{\xi\xi}(\lambda_j) G(e^{i\lambda})' \right|,$$

where  $\hat{\Sigma}_e$  (as a function of  $C$  and  $\Theta_G$ ) denotes the matrix that minimises  $L_1(\cdot)$  with respect to  $\Sigma_e$ . The estimators of  $C$  and  $\theta_G$  are those that minimise  $L(\cdot)$ . In the case of the VAR(1) use of the concentrated function reduces the number of unknown parameters from eight to five while in the cases of the VAR(2) and VAR(4) the reductions are from twelve to nine and from twenty to seventeen, respectively.

The bias and root mean squared error (RMSE) of each of the estimators obtained in the simulations are reported in Table I. All estimators provide substantial reductions over OLS in terms of bias and RMSE when  $u(t)$  is white noise. When  $u(t)$  is a continuous time AR(1) process the bias of the all estimators is larger (in absolute value) than that of OLS but they all have lower RMSE compared with OLS. In all cases the spectral estimators produce the smallest RMSE.<sup>7</sup> For each estimator the bias and RMSE are both seen to increase as  $u(t)$  moves from being white noise to a continuous time AR process with  $\phi = -5$  to a continuous time AR process with  $\phi = -10$ .

We conclude this section with an empirical illustration of the methods derived in this paper. The illustration uses the extended data set on US stock prices and dividends based on Shiller (2000).<sup>8</sup> We treat the monthly stock price data as a stock variable<sup>9</sup> and the yearly dividends as a flow variable. The sample begins in January 1871 and ends in December 2016 which gives  $T = 146$  yearly observations on dividends and  $N = 1752$  monthly observations on stock prices; the high frequency sampling interval is  $h = 1/12$ . Ghysels and Miller (2015) have also examined this data set in the context of a mixed frequency cointegration model and provide arguments as to why cointegration may not hold between these variables. We nevertheless proceed assuming that cointegration holds and use demeaned data, as in Ghysels

<sup>6</sup>This function is actually an approximation to  $-2/T$  times the exact likelihood function.

<sup>7</sup>The results are reported to four decimal places so the actual differences in bias in some of the cases are only apparent at the fifth decimal place.

<sup>8</sup>The data can be downloaded from <http://aida.econ.yale.edu/~shiller/data.htm>.

<sup>9</sup>Note, though, that the data represent the average daily closing price over the month.

and Miller (2015). The model of interest, based on (1) and (2), is given by

$$\begin{aligned}\log P(t) &= C \log R(t) + u_P(t), \quad t > 0, \\ D \log R(t) &= u_R(t), \quad t > 0,\end{aligned}$$

where  $P(t)$  denotes the stock price at time  $t$  and  $R(t)$  denotes the rate of flow of dividends at time  $t$ . The cointegrating parameter is  $C$  and  $u(t) = (u_P(t), u_R(t))'$  is assumed to be a stationary continuous time process. The monthly stock price data are averaged into yearly equivalents as defined by (6) (with  $k = 12$ ).

The estimation results are given in Table II. Estimates of  $C$  obtained by OLS, FD, FDA, FM and VAR are provided. The spectral regression estimators are based on  $m = 4$ ,  $m = 12$  and  $m = 32$  averaged periodograms, these values corresponding to  $m = \lceil T^\delta \rceil$  for  $\delta \in \{0.3, 0.5, 0.7\}$ . The headings for these entries in the Table are of the form  $\text{FD}(m)$  and  $\text{FDA}(m)$ . The same values, as well as the automatic bandwidth value, were used for the FM estimators, being denoted  $\text{FM}(m)$  and  $\text{FM}(\text{Auto})$ , respectively. The frequency domain  $\text{VAR}(p)$ -based estimates of  $C$  are given for  $p \in \{1, 2, 4\}$ . The OLS, spectral regression and FM estimates of  $C$  are all in the vicinity of 0.52 and show little variation as  $m$  is varied. The standard errors of the spectral estimators are lower than for the OLS and FM estimators and show a small increase as  $m$  increases. Estimates of  $C$  obtained using the VAR approximations are slightly larger, ranging from 0.60 and 0.63, but have much larger standard errors than the OLS, spectral and FM estimators. The estimates of all the other VAR coefficients (not reported) have large standard errors relative to the estimates themselves, suggesting that the parameters are insignificantly different from zero. However, interpreting the coefficients in this case is problematical if the continuous time model is taken seriously in view of the VARs being misspecified in this case. Nevertheless Table II also reports likelihood ratio test statistics based on the function  $L(\cdot)$  for testing the VAR orders 2 and 4 against  $p = 1$ . It can be seen that the test results in the non-rejection of the null hypothesis that the additional parameters in the  $\text{VAR}(2)$  and  $\text{VAR}(4)$  are zero when compared with the  $\text{VAR}(1)$ , despite all of the coefficients in the  $\text{VAR}(1)$  being individually insignificant.

## 5. Conclusion

This paper has been concerned with the estimation of continuous time cointegrating vectors when the data are a mixture of stocks and flows observed at different frequencies. A spectral regression estimator based on Chambers (2018), which treats the system dynamics nonparametrically, was proposed and was found to perform well in a simulation study. An illustration of the methods using the updated mixed frequency stock price and dividend data of Shiller (2000) was also provided.

The methods in this paper are straightforward to implement and are particularly useful in circumstances where the focus is on the cointegrating parameters themselves and the researcher does not wish to specify a parametric model for the system dynamics. In circumstances where it is important to explicitly model such dynamics – for example, for the purposes of forecasting or impulse response analysis – alternative techniques for handling parametric cointegrated continuous time systems with mixed sample and mixed frequency data are available and can be found in Chambers (2016) and Thornton (2018).

## Data availability statement

The data used in the application in section 4 are freely available for download at <http://aida.econ.yale.edu/~shiller/data.htm> and are discussed in Shiller (2000).

## Appendix

A key result in proving Theorem 1 is the following:

**Lemma A1.** *Let  $y(t)$  be a continuous time process satisfying  $Dy(t) = u(t)$  where  $u(t)$  is a stationary continuous time process satisfying Assumption 1. Then*

$$\delta_t = \frac{1}{k} \sum_{l=0}^{k-1} y(t-lh) - \int_{t-1}^t y(r) dr = \sum_{l=0}^{k-1} \int_{t-lh-h}^{t-lh} \left( \frac{l+1}{k} - (t-s) \right) u(s) ds$$

is also a stationary process.

**Proof of Lemma A1.** We begin by noting that, for all  $t > 0$ ,

$$y(t) = y(0) + \int_0^t u(s) ds,$$

which implies that

$$\begin{aligned} \int_{t-1}^t y(r) dr &= y(0) + \int_{t-1}^t \int_0^r u(s) ds dr \\ &= y(0) + \int_0^{t-1} u(s) ds + \int_{t-1}^t (t-s) u(s) ds, \end{aligned} \quad (20)$$

the latter expression using the decomposition of double integrals into single integrals along the lines of Bergstrom (1997) and McCrorie (2000). Furthermore,

$$\begin{aligned} \frac{1}{k} \sum_{l=0}^{k-1} y(t-lh) &= \frac{1}{k} \sum_{l=0}^{k-1} \left( y(0) + \int_0^{t-lh} u(s) ds \right) \\ &= y(0) + \frac{1}{k} \sum_{l=0}^{k-1} \int_0^{t-lh} u(s) ds. \end{aligned}$$

Now, for  $0 \leq l \leq k-1$ ,

$$\int_0^{t-lh} u(s) ds = \int_0^{t-1} u(s) ds + \int_{t-1}^{t-lh} u(s) ds$$

and so

$$\frac{1}{k} \sum_{l=0}^{k-1} y(t-lh) = y(0) + \int_0^{t-1} u(s) ds + \frac{1}{k} \sum_{l=0}^{k-1} \int_{t-1}^{t-lh} u(s) ds. \quad (21)$$



From (20) and (21) it follows that

$$\delta_t = \frac{1}{k} \sum_{l=0}^{k-1} \int_{t-1}^{t-lh} u(s) ds - \int_{t-1}^t (t-s)u(s) ds$$

and the expression for  $\delta_t$  in the Lemma is a consequence of calculating the contribution of each sub-integral of length  $h$  over the interval  $(t-1, t]$ .  $\square$

**Proof of Theorem 1.** It is convenient to partition  $C$  in the form

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

where the first subscript of each sub-matrix corresponds to the appropriate sub-vector of  $y_1$  and the second to the appropriate sub-vector of  $y_2$ ; for example,  $C_S$  is of dimension  $n_1^S \times n$  and  $C_{FS}$  is of dimension  $n_1^F \times n_2^S$ . We begin with the equation for  $Y_{1t}^S$ . The first  $n_1^S$  equations of (4) give (noting that  $y_{1t}^S = y_1^S(t)$ )

$$y_{1t}^S = C_{SS}y_{2,t-1}^S + C_{SF}y_2^F(t-1) + v_1^S(t),$$

where

$$v_1^S(t) = u_1^S(t) + C_S \int_{t-1}^t u_2(r) dr.$$

Applying the operator  $k^{-1}s(L^h)$  to this equation yields

$$Y_{1t}^S = C_{SS}Y_{2,t-1}^S + C_{SF} \frac{1}{k} \sum_{l=0}^{k-1} y_2^F(t-1-lh) + \frac{1}{k} \sum_{l=0}^{k-1} v_1^S(t-lh).$$

Replacing the unobservable component involving  $y_2^F$  with the observable  $Y_{2,t-1}^F$  and using Lemma A1 results in

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

where, noting the definition of  $v_1^S(t)$ ,

$$\xi_{1t}^S = \frac{1}{k} \sum_{l=0}^{k-1} u_1^S(t-lh) + C_S \frac{1}{k} \sum_{l=0}^{k-1} \int_{t-1-lh}^{t-lh} u_2(r) dr + C_{SF}\delta_{2,t-1}^F.$$

Taking equations  $n_1^S + 1$  to  $n_1$  of (4) we obtain

$$y_1^F(t) = C_{FS}y_2^S(t-1) + C_{FF}y_2^F(t-1) + v_1^F(t),$$

where

$$v_1^F(t) = u_1^F(t) + C_F \int_{t-1}^t u_2(r) dr.$$

Integrating over  $(t-1, t]$  yields

$$Y_{1t}^F = C_{FS} \int_{t-2}^{t-1} y_2^S(r) dr + C_{FF}Y_{2,t-1}^F + \int_{t-1}^t v_1^F(r) dr.$$

Proceeding as before – replacing the unobservable component, this time involving  $y_2^S$ , with  $Y_{2,t-1}^S$  – and utilising Lemma A1 and the definition of  $v_1^F(t)$ , results in

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

where

$$\xi_{1t}^F = \int_{t-1}^t u_1^F(r)dr + C_F \int_{t-1}^t \int_{r-1}^r u_2(s)dsdr - C_{FS}\delta_{2,t-1}^S.$$

Turning to the components of  $y_2$ , from (4), the equations relating to  $y_2^S$  are simply

$$\Delta y_2^S(t) = \int_{t-1}^t u_2^S(r)dr,$$

and so applying the filter  $k^{-1}s(L^h)$  results in

$$\Delta Y_{2t}^S = \xi_{2t}^S, \quad \xi_{2t}^S = \frac{1}{k} \sum_{l=0}^{k-1} \int_{t-1-lh}^{t-lh} u_2^S(r)dr.$$

Also from (4) the equations relating to  $y_2^F$  are

$$\Delta y_2^F(t) = \int_{t-1}^t u_2^F(r)dr,$$

which can be integrated over the interval  $(t-1, t]$  to give

$$\Delta Y_{2t}^F = \xi_{2t}^F, \quad \xi_{2t}^F = \int_{t-1}^t \int_{r-1}^r u_2^F(s)dsdr.$$

The representation in the Theorem is obtained by combining the equations for  $Y_{1t}^S$  and  $Y_{1t}^F$  to give (7) with  $\xi_{1t} = (\xi_{1t}^S, \xi_{1t}^F)'$ , and (8) results by combining the equations for  $\Delta Y_{2t}^S$  and  $\Delta Y_{2t}^F$  with  $\xi_{2t} = (\xi_{2t}^S, \xi_{2t}^F)'$ .  $\square$

**Proof of Lemma 1.** We demonstrate the validity of the FCLT by verifying that the conditions of Corollary 2.2 of Phillips and Durlauf (1986) are satisfied. The requirements are: (i)  $E(\xi_t) = 0$ ; (ii)  $E|\xi_{it}|^\beta < \infty$  ( $i = 1, \dots, n$ ) for some  $\beta > 2$ ; and (iii)  $\alpha_s = O(s^{-\mu})$  for some  $\mu > \beta/(\beta - 2)$ , where  $\alpha_s$  denote the strong mixing coefficients of  $\xi_t$ . Condition (i) is clearly satisfied. Turning to (ii) we have, from the definition of  $\xi_{1t}^S$  in the proof of Theorem 1 and of  $\delta_t$  in Lemma A1,

$$\begin{aligned} \xi_{1i,t}^S &= \frac{1}{k} \sum_{l=0}^{k-1} \left( u_{1i}^S(t-lh) + \sum_{j=1}^{n_2} C_{S,ij} \int_{t-1-lh}^{t-lh} u_{2j}^S(r)dr \right. \\ &\quad \left. + \sum_{j=1}^{n_2^F} C_{SF,ij} \int_{t-1-h-lh}^{t-1-lh} [l+1-k(t-1-r)]u_{2j}^F(r)dr \right), \quad i = 1, \dots, n_1^S, \end{aligned}$$

and so application of Loève's  $c_r$  inequality (see, for example, Davidson, 1994, p.140) yields

$$\begin{aligned}
E |\xi_{1i,t}^S|^\beta &\leq \frac{1}{k} \sum_{l=0}^{k-1} E \left| u_{1i}^S(t-lh) + \sum_{j=1}^{n_2} C_{S,ij} \int_{t-1-lh}^{t-lh} u_{2j}(r) dr \right. \\
&\quad \left. + \sum_{j=1}^{n_2^F} C_{SF,ij} \int_{t-1-h-lh}^{t-1-lh} [l+1-k(t-1-r)] u_{2j}^F(r) dr \right|^\beta \\
&\leq \frac{1}{k} \sum_{l=0}^{k-1} E |\phi_{itl}^S|^\beta,
\end{aligned}$$

where  $\phi_{itl}^S$  is implicitly defined. A further application of the  $c_r$  inequality to the summands of interest yields

$$\begin{aligned}
E |\phi_{itl}^S|^\beta &\leq (1+n_2+n_2^F)^{\beta-1} \left( E |u_{1i}^S(t-lh)|^\beta + \sum_{j=1}^{n_2} E \left| C_{S,ij} \int_{t-1-lh}^{t-lh} u_{2j}(r) dr \right|^\beta \right. \\
&\quad \left. + \sum_{j=1}^{n_2^F} E \left| C_{SF,ij} \int_{t-1-h-lh}^{t-1-lh} [l+1-k(t-1-r)] u_{2j}^F(r) dr \right|^\beta \right).
\end{aligned}$$

Now  $E |u_{1i}^S(t-lh)|^\beta = E |u_{1i}^S(t)|^\beta < \infty$  by Assumption 1(b) while, using Lemma A3 of Chambers (2003),

$$\begin{aligned}
E \left| C_{S,ij} \int_{t-1-lh}^{t-lh} u_{2j}(r) dr \right|^\beta &\leq |C_{S,ij}| E |u_{2j}(t)|^\beta < \infty, \\
E \left| C_{SF,ij} \int_{t-1-h-lh}^{t-1-lh} [l+1-k(t-1-r)] u_{2j}^F(r) dr \right|^\beta \\
&\leq |C_{SF,ij}|^\beta \left| \int_{t-1-lh-h}^{t-1-lh} (l+1-k(t-1-r)) dr \right|^\beta E |u_{2j}^F(t)|^\beta \\
&= |C_{SF,ij}|^\beta \left| h(l+1) - \frac{k}{2}(2l+1)h^2 \right|^\beta E |u_{2j}^F(t)|^\beta < \infty.
\end{aligned}$$

Hence  $E |\phi_{itl}^S|^\beta < \infty$  implying  $E |\xi_{1i,t}^S|^\beta < \infty$  also.

Turning to  $\xi_{1t}^F$  we have, from its definition in the proof of Theorem 1,

$$\begin{aligned}
\xi_{1i,t}^F &= \int_{t-1}^t u_{1i}^F(r) dr + \sum_{j=1}^{n_2} C_{F,ij} \int_{t-1}^t \int_{r-1}^r u_{2j}(s) ds dr \\
&\quad - \frac{1}{k} \sum_{l=0}^{k-1} \sum_{j=1}^{n_2^S} C_{FS,ij} \int_{t-1-h-lh}^{t-1-lh} [l+1-k(t-1-r)] u_{2j}^S(r) dr, \quad i = 1, \dots, n_1^F.
\end{aligned}$$

Applying the  $c_r$  inequality yields

$$E|\xi_{1i,t}^F|^\beta \leq (1 + n_2 + n_2^S)^{\beta-1} \left( E \left| \int_{t-1}^t u_{1i}^F(r) dr \right|^\beta + \sum_{j=1}^{n_2} E \left| C_{F,ij} \int_{t-1}^t \int_{r-1}^r u_{2j}(s) ds dr \right|^\beta \right. \\ \left. + \sum_{j=1}^{n_2^S} E \left| C_{FS,ij} \frac{1}{k} \sum_{l=0}^{k-1} \int_{t-1-h-lh}^{t-1-lh} [l+1 - k(t-1-r)] u_{2j}^S(r) dr \right|^\beta \right).$$

Use of Lemma A3 in Chambers (2003) yields, for the first two components,

$$E \left| \int_{t-1}^t u_{1i}^F(r) dr \right|^\beta \leq E |u_{1i}^F(t)|^\beta < \infty,$$

$$E \left| C_{F,ij} \int_{t-1}^t \int_{r-1}^r u_{2j}(s) ds dr \right|^\beta \leq |C_{F,ij}|^\beta E |u_{2j}(t)|^\beta < \infty,$$

while applying the  $c_r$  inequality again to the third components gives

$$E \left| C_{FS,ij} \frac{1}{k} \sum_{l=0}^{k-1} \int_{t-1-h-lh}^{t-1-lh} [l+1 - k(t-1-r)] u_{2j}^S(r) dr \right|^\beta \\ \leq |C_{SF,ij}|^\beta \frac{1}{k} \sum_{l=0}^{k-1} E \left| \int_{t-1-h-lh}^{t-1-lh} [l+1 - k(t-1-r)] u_{2j}^S(r) dr \right|^\beta \\ \leq \frac{1}{k} |C_{SF,ij}|^\beta E |u_{2j}^S(t)|^\beta \sum_{l=0}^{k-1} \left| h(l+1) - \frac{k}{2}(2l+1)h^2 \right|^\beta < \infty.$$

Combining these results establishes that  $E|\xi_{1i,t}^F|^\beta < \infty$ . Finally, from their definitions in the proof of Theorem 1,

$$E|\xi_{2i,t}^S|^\beta \leq \frac{1}{k} \sum_{l=0}^{k-1} E \left| \int_{t-lh}^{t-lh} u_{2i}^S(r) dr \right|^\beta \leq E |u_{2i}^S(t)|^\beta < \infty, \quad i = 1, \dots, n_2^S,$$

$$E|\xi_{2i,t}^F|^\beta \leq E |u_{2i}^F(t)|^\beta < \infty, \quad i = 1, \dots, n_2^F,$$

which make use of the  $c_r$  inequality and Lemma A3 of Chambers (2003), respectively.

Finally it is necessary to show that  $\xi_t$  satisfies the appropriate mixing coefficient decay rate. This follows because  $\xi_t$  is a measurable function of  $u(t)$  over a finite interval and so inherits the same mixing properties; see, for example, Theorem 14.1 of Davidson (1994). This completes the proof.  $\square$

**Proof of Lemma 2.** The arguments used in the proof of Lemma 1 ensure that  $E|\xi_{it}|^{\beta'} < \infty$  and that  $\xi_t$  inherits the same mixing properties as  $u(t)$ . The conditions of Lemma 1 of Andrews (1991) are then satisfied by  $\xi_t$  and the stated summability properties of the autocovariance matrices and fourth-order cumulants follow from Andrews' Lemma 1 and Assumption A.  $\square$

**Proof of Theorem 2.** Parts (a) and (b) follow from Chambers (2018). For part (c) we

begin with the decomposition

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

and then proceed to show that each of the two terms in parentheses is  $o_p(1)$ . The first term is  $O_p(1/m) = o_p(1)$  under Assumption 2 using the arguments in Chambers (2018). For the second term consider

$$\begin{aligned} E\hat{f}_{\xi\xi}(0) - f_{\xi\xi}(0) &= \frac{1}{2\pi} \sum_{|k|<T} \left(1 - \frac{|k|}{T}\right) w_k \Gamma_{\xi,k} - \frac{1}{2\pi} \sum_{-\infty}^{\infty} \Gamma_{\xi,k} \\ &= \frac{1}{2\pi} \sum_{|k|<T} \left(1 - \frac{|k|}{T}\right) (w_k - 1) \Gamma_{\xi,k} - \frac{1}{2\pi} \sum_{|k|<T} \frac{|k|}{T} \Gamma_{\xi,k} - \frac{1}{2\pi} \sum_{|k|\geq T} \Gamma_{\xi,k}, \end{aligned}$$

where  $w_k = (2m+1)^{-1} \sum_{s=-m}^m e^{-ik\lambda_s}$ ; cf. Rosenblatt (1984, p.1171). Then

$$\left\| E\hat{f}_{\xi\xi}(0) - f_{\xi\xi}(0) \right\| \leq \frac{1}{2\pi} \max_{|k|<T} |w_k - 1| \sum_{|k|<T} \|\Gamma_{\xi,k}\| + \frac{1}{2\pi} \sum_{|k|<T} \frac{|k|}{T} \|\Gamma_{\xi,k}\| + \frac{1}{2\pi} \sum_{|k|\geq T} \|\Gamma_{\xi,k}\|.$$

The first term converges to zero by the absolute summability of the autocovariances and the fact that  $w_k \rightarrow 1$  for all  $k$ , while the second and third terms also converge to zero from the absolute summability of the autocovariances and using Kronecker's Lemma for the second. Hence

$$\left\| E\hat{f}_{\xi\xi}(0) - f_{\xi\xi}(0) \right\| \rightarrow 0$$

as  $T \rightarrow \infty$ . Next consider an element of  $\hat{f}_{\xi\xi}(0)$ , denoted  $\hat{f}_{\xi\xi,ab}(0)$  ( $a, b, = 1, \dots, n$ ). Then

$$\hat{f}_{\xi\xi,ab}(0) = \frac{1}{2m+1} \sum_{s=-m}^m I_{\xi\xi,ab}(\lambda_s), \quad a, b, = 1, \dots, n.$$

It follows that

$$\begin{aligned} \text{var} \left( \hat{f}_{\xi\xi,ab}(0) \right) &= \frac{1}{(2m+1)^2} \sum_{s=-m}^m \text{var} \left( I_{\xi\xi,ab}(\lambda_s) \right) \\ &\quad + \frac{1}{(2m+1)^2} \sum_{r \neq s} \sum_{s=-m}^m \text{cov} \left( I_{\xi\xi,ab}(\lambda_r), \overline{I_{\xi\xi,ab}(\lambda_s)} \right). \end{aligned}$$

Using Lemma 2 we can apply the results of Anderson (1971, chapter 8) to show that the variances and covariances of elements of  $I_{\xi\xi}(\lambda)$  are  $O(1/T)$  and hence (using  $K_1$  and  $K_2$  as generic constants)

$$\begin{aligned} \text{var} \left( \hat{f}_{\xi\xi,ab}(0) \right) &\leq \frac{1}{(2m+1)^2} \left( \frac{K_1(2m+1)}{T} + \frac{K_2 2m(2m+1)}{T} \right) \\ &= O \left( \frac{1}{mT} \right) + O \left( \frac{1}{T} \right) = o(1) \end{aligned}$$

under Assumption 2. Hence  $\hat{f}_{\xi\xi}(0) \xrightarrow{p} f_{\xi\xi}(0)$  and the proof of (c) is complete.  $\square$

**Proof of Theorem 3.** Given the validity of Theorem 2 the proof follows from Chambers (2018).  $\square$

### Generation of discrete time data

The discrete time data are required to satisfy the underlying continuous time system in (1) and (2). Stock ( $y_1$ ) and flow ( $y_2$ ) data are generated at the high frequency and then aggregated to construct the low frequency variables. Denoting the high frequency time index by  $\tau = 1, \dots, N = T/h$  and the low frequency index by  $t = 1, \dots, T$ , we have

$$y_{1,\tau h} = y_1(\tau h) = C y_2(\tau h) + u_1(\tau h), \quad \tau = 1, \dots, N,$$

where  $y_2(\tau h)$  denotes the ‘unobserved’ value of  $y_2$ . As for  $y_2$ , integrating (2) once yields

$$\Delta_h y_2(\tau h) = \int_{\tau h-h}^{\tau h} u_2(r) dr, \quad \tau = 1, \dots, N,$$

while a further integration results in

$$\Delta_h y_{2,\tau h} = \Delta_h \int_{\tau h-h}^{\tau h} y_2(r) dr = \int_{\tau h-h}^{\tau h} \int_{r-h}^r u_2(s) ds dr, \quad \tau = 1, \dots, N.$$

The aggregated stock variable is obtained from the high frequency observations using

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

while the observed flow variables are obtained using

$$Y_{2t} = \sum_{l=0}^{k-1} y_{2,t-lh} = \int_{t-1}^t y_2(r) dr, \quad t = 1, \dots, T.$$

Given the assumed dynamic specification of  $u(t)$  it is therefore necessary to generate the following quantities:

$$u(\tau h), \quad u_{\tau h} = \int_{\tau h-h}^{\tau h} u(r) dr, \quad U_{\tau h} = \int_{\tau h-h}^{\tau h} \int_{r-h}^r u(s) ds dr, \quad \tau = 1, \dots, N.$$

We consider the following two specifications for  $u(t)$ :

(i)  $u(t)$  is white noise

In this case we assume that  $E(u(t)u(s)') = \Sigma$  if  $t = s$  and  $= 0$  otherwise. Clearly  $u(\tau h)$  is also white noise with covariance matrix  $\Sigma$ . Furthermore integrals of  $u(t)$  over non-overlapping intervals will also be uncorrelated so that  $u_{\tau h}$  is also white noise with covariance matrix

$$\Sigma_u = \int_{\tau h-h}^{\tau h} \Sigma dr = h\Sigma.$$

Turning to  $U_{\tau h}$  we can write

$$\begin{aligned} U_{\tau h} &= \int_{\tau h-h}^{\tau h} \left( \int_r^{\tau h} ds \right) u(r) dr + \int_{\tau h-2h}^{\tau h-h} \left( \int_{\tau h-h}^{r+h} ds \right) u(r) dr \\ &= \int_{\tau h-h}^{\tau h} (\tau h - r) u(r) dr - \int_{\tau h-2h}^{\tau h-h} (\tau h - 2h - r) u(r) dr. \end{aligned}$$

The two integrals are uncorrelated so we find that

$$E(U_{\tau h} U'_{\tau h}) = \left( \int_{\tau h-h}^{\tau h} (\tau h - r)^2 dr + \int_{\tau h-2h}^{\tau h-h} (\tau h - 2h - r)^2 dr \right) \Sigma = \frac{2}{3} h^3 \Sigma$$

while the first-order autocovariance is given by

$$E(U_{\tau h} U'_{\tau h-h}) = - \int_{\tau h-2h}^{\tau h-h} (\tau h - 2h - r) (\tau h - h - r) dr \Sigma = \frac{1}{6} h^3 \Sigma.$$

All other autocovariances are zero and so  $U_{\tau h}$  is a vector MA(1) process.

(ii)  $u(t)$  is continuous time AR(1)

We shall proceed assuming that  $u(t) = [u_1(t), u_2(t)]'$  satisfies

$$du(t) = \Phi u(t) dt + \zeta(dt), \quad t > 0, \quad (22)$$

where the roots of the equation  $|zI - \Phi| = 0$  have negative real parts (to ensure stationarity) and  $\zeta(dt)$  is a vector random measure satisfying  $E\zeta(dt) = 0$ ,  $E\zeta(dt)\zeta(dt)' = \Sigma dt$  (where  $\Sigma$  is a symmetric, positive definite matrix) and  $E\zeta(S_1)\zeta(S_2)' = 0$  for non-overlapping sub-intervals,  $S_1$  and  $S_2$ , of the real line; see Bergstrom (1984) for a detailed treatment of random measures as applied to stochastic differential equations systems. The solution to (22), which is unique in a mean square sense, is given by

$$u(\tau h) = e^{\tau h \Phi} u(0) + \int_0^{\tau h} e^{(\tau h - r) \Phi} \zeta(dr), \quad \tau > 0,$$

and can be used to show that  $u(\tau h)$  satisfies the stochastic difference equation

$$u(\tau h) = F u(\tau h - h) + \epsilon_{\tau h}, \quad \tau = 1, \dots, N, \quad (23)$$

where, defining  $F(r) = e^{r\Phi}$ ,  $F = F(h) = e^{h\Phi}$  and

$$\epsilon_{\tau h} = \int_{\tau h-h}^{\tau h} F(\tau h - r) \zeta(dr)$$

is a vector white noise process with covariance matrix

$$E(\epsilon_{\tau h} \epsilon'_{\tau h}) = \Omega_{\epsilon\epsilon} = \int_0^h F(r) \Sigma F(r)' dr.$$

Discrete time data for  $u(\tau h)$  can be generated using (23) given an initial value  $u(0)$  and a sequence of white noise innovation vectors,  $\epsilon_{\tau h}$ , with covariance  $\Omega_{\epsilon\epsilon}$ .

Turning to integrals of  $u$  it would be possible to integrate (23) over  $(\tau h - h, \tau h]$  to derive an appropriate law of motion. Here we follow an alternative approach based on Zadlozny (1988). Note that, for  $s \geq 0$ ,

$$u(\tau h - h + s) = F(s)u(\tau h - h) + \int_{\tau h - h}^{\tau h - h + s} F(\tau h - h + s - r)\zeta(dr);$$

this representation follows from (23). It is convenient for what follows to use the change of variable  $w = r - (\tau h - h)$  in the integral with respect to the random measure; this yields

$$u(\tau h - h + s) = F(s)u(\tau h - h) + \int_0^s F(s - w)\zeta(\tau h - h + dw).$$

This equation can be used to link

$$u_{\tau h} = \int_{\tau h - h}^{\tau h} u(r)dr = \int_0^h u(\tau h - h + s)ds$$

to  $u(\tau h - h)$ , as follows, by integrating:

$$u_{\tau h} = \int_0^h F(s)dsu(\tau h - h) + \int_0^h \int_0^s F(s - w)\zeta(\tau h - h + dw)ds.$$

Let  $\Phi_1(x) = \int_0^x F(s)ds$ . Then the above equation is of the form

$$u_{\tau h} = \Phi_1 u(\tau h - h) + \rho_{\tau h}, \quad (24)$$

where  $\Phi_1 = \Phi_1(h)$  and

$$\rho_{\tau h} = \int_0^h \int_0^s F(s - w)\zeta(\tau h - h + dw)ds.$$

The limits of integration in  $\rho_{\tau h}$  can be changed to give

$$\rho_{\tau h} = \int_0^h \left( \int_w^h F(s - w)ds \right) \zeta(\tau h - h + dw);$$

see, for example, McCrorie (2000) for details. But  $\int_w^h F(s - w)ds = \int_w^h e^{(s-w)\Phi} ds = \int_0^{(h-w)} e^{r\Phi} dr = \Phi_1(h - w)$  and so

$$\rho_{\tau h} = \int_0^h \Phi_1(h - w)\zeta(\tau h - h + dw) = \int_{\tau h - h}^{\tau h} \Phi_1(\tau h - r)\zeta(dr).$$

Note that  $\rho_{\tau h}$  is vector white noise with covariance matrix

$$E(\rho_{\tau h}\rho'_{\tau h}) = \Omega_{\rho\rho} = \int_0^h \Phi_1(r)\Sigma\Phi_1(r)'dr.$$

Hence  $u_{\tau h}$  can be generated from (24) given an initial value  $u(0)$  and a sequence  $\rho_{\tau h}$  with appropriate covariance properties.



We also need the law of motion for  $U_{\tau h}$ . Using (24) we have

$$\int_{s-h}^s u(r)dr = \Phi_1 u(s-h) + \int_{s-h}^s \Phi_1(s-r)\zeta(dr),$$

from which a further integration yields

$$U_{\tau h} = \Phi_1 \int_{\tau h-h}^{\tau h} u(s-h)ds + \int_{\tau h-h}^{\tau h} \int_{s-h}^s \Phi_1(s-r)\zeta(dr)ds,$$

which is of the form

$$U_{\tau h} = \Phi_1 u_{\tau h-h} + \phi_{\tau h}, \quad (25)$$

where the disturbance vector is

$$\phi_{\tau h} = \int_{\tau h-h}^{\tau h} \int_{s-h}^s \Phi_1(s-r)\zeta(dr)ds.$$

A change of limits of integration means that  $\phi_{\tau h}$  can be written in the form

$$\phi_{\tau h} = \int_{\tau h-h}^{\tau h} \left( \int_r^{\tau h} \Phi_1(s-r)ds \right) \zeta(dr) + \int_{\tau h-2h}^{\tau h-h} \left( \int_{\tau h-h}^{r+h} \Phi_1(s-r)ds \right) \zeta(dr).$$

Let  $\Phi_2(x) = \int_0^x \Phi_1(r)dr$ . Then we obtain

$$\int_r^{\tau h} \Phi_1(s-r)ds = \int_0^{\tau h-r} \Phi_1(v)dv = \Phi_2(\tau h-r),$$

$$\int_{\tau h-h}^{r+h} \Phi_1(s-r)ds = \int_{\tau h-h-r}^h \Phi_1(v)dv = - \int_h^{\tau h-h-r} \Phi_1(v)dv = -(\Phi_2(\tau h-h-r) - \Phi_2(h))$$

which means that  $\phi_{\tau h}$  can be written

$$\phi_{\tau h} = \int_{\tau h-h}^{\tau h} \Phi_2(\tau h-r)\zeta(dr) - \int_{\tau h-2h}^{\tau h-h} (\Phi_2(\tau h-h-r) - \Phi_2(h))\zeta(dr).$$

Note that  $\phi_{\tau h}$  is an MA(1) process with  $E(\phi_{\tau h}\phi'_{\tau h}) = \Omega_{\phi\phi}$  and  $E(\phi_{\tau h}\phi'_{\tau h-h}) = \Omega_{1,\phi\phi}$  where

$$\Omega_{\phi\phi} = \int_0^h \Phi_2(r)\Sigma\Phi_2(r)'dr + \int_0^h (\Phi_2(r) - \Phi_2(h))\Sigma(\Phi_2(r) - \Phi_2(h))'dr,$$

$$\Omega_{1,\phi\phi} = - \int_0^h (\Phi_2(r) - \Phi_2(h))\Sigma\Phi_2(r)'dr.$$

Discrete time data for  $u(\tau h)$ ,  $u_{\tau h}$  and  $U_{\tau h}$  can, therefore, be generated using the stochastic difference equations (23), (24) and (25), respectively. In effect, given initial values for each of these series, what this requires is the generation of two vector white noise sequences,  $\epsilon_{\tau h}$  and  $\rho_{\tau h}$ , and a vector MA(1) sequence,  $\phi_{\tau h}$ . Data are generated at the high frequency and then aggregated appropriately. Let  $e_h, \dots, e_{Nh}$  denote a sequence of  $N$  i.i.d. random variables of dimension  $2 \times 1$ , each having mean zero and covariance matrix  $I_2$ . Let  $M_{\epsilon\epsilon}$  and  $M_{\rho\rho}$  denote the  $2 \times 2$  Cholesky factorisations of the covariance matrices  $\Omega_{\epsilon\epsilon}$  and  $\Omega_{\rho\rho}$ , respectively, so that  $M_{\epsilon\epsilon}$  and  $M_{\rho\rho}$  are lower triangular matrices that satisfy  $M_{\epsilon\epsilon}M'_{\epsilon\epsilon} = \Omega_{\epsilon\epsilon}$

and  $M_{\rho\rho}M'_{\rho\rho} = \Omega_{\rho\rho}$ . Then the sequences  $\epsilon_{\tau h} = M_{\epsilon\epsilon}e_{\tau h}$  and  $\rho_{\tau h} = M_{\rho\rho}e_{\tau h}$  ( $\tau = 1, \dots, N$ ) have the required covariance properties.

The generation of the vector MA(1) process  $\phi_{\tau h}$  is also based on the same set of underlying i.i.d. random vectors  $e_{\tau h}$ . Let  $e = (e'_h, \dots, e'_{Nh})'$  denote the  $2N \times 1$  vector of i.i.d. random variates having mean vector zero and covariance matrix  $I_{2N}$ , and let

$$V_{\phi\phi} = \begin{pmatrix} \Omega_{\phi\phi} & \Omega'_{1,\phi\phi} & 0 & 0 & \dots & 0 & 0 & 0 \\ \Omega_{1,\phi\phi} & \Omega_{\phi\phi} & \Omega'_{1,\phi\phi} & 0 & \dots & 0 & 0 & 0 \\ 0 & \Omega_{1,\phi\phi} & \Omega_{\phi\phi} & \Omega'_{1,\phi\phi} & \dots & 0 & 0 & 0 \\ \vdots & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & \dots & \Omega_{\phi\phi} & \Omega'_{1,\phi\phi} & 0 \\ 0 & 0 & 0 & 0 & \dots & \Omega_{1,\phi\phi} & \Omega_{\phi\phi} & \Omega'_{1,\phi\phi} \\ 0 & 0 & 0 & 0 & \dots & 0 & \Omega_{1,\phi\phi} & \Omega_{\phi\phi} \end{pmatrix}$$

denote the  $2N \times 2N$  covariance matrix that the required  $2N \times 1$  vector  $\phi = (\phi'_h, \dots, \phi'_{Nh})'$  is required to satisfy. Furthermore, denote the  $2N \times 2N$  lower triangular Cholesky factorisation of  $V_{\phi\phi}$  by  $M_{\phi\phi}$  so that  $M_{\phi\phi}M'_{\phi\phi} = V_{\phi\phi}$ . Then  $\phi = M_{\phi\phi}e$  satisfies the required MA(1) autocovariance structure because

$$E(\phi\phi') = E(M_{\phi\phi}ee'M'_{\phi\phi}) = M_{\phi\phi}E(ee')M'_{\phi\phi} = M_{\phi\phi}M'_{\phi\phi} = V_{\phi\phi}.$$

In fact, the sparse nature of  $V_{\phi\phi}$  is reflected in  $M_{\phi\phi}$  as well, meaning that an efficient recursive algorithm can be employed to compute the elements of  $\phi$ . Moreover, the  $2 \times 2$  non-zero sub-matrices of  $M_{\phi\phi}$  converge rapidly to fixed matrices as one moves further into the array, meaning that computational storage is also much reduced; see Bergstrom (1990, chapter 7) for a proof of this result and for details of how the procedure operates in the case of an MA(2) vector sequence.

In all cases the underlying continuous time process is assumed to be stationary with an invertible autoregressive matrix  $\Phi$ . As a result, closed form solutions are available for the computation of the relevant (auto-)covariance matrices. In the case of  $\Omega_{\epsilon\epsilon}$ , recall that

$$\Omega_{\epsilon\epsilon} = \int_0^h e^{r\Phi} \Sigma e^{r\Phi'} dr.$$

As in Phillips (1973) it is possible to use integration by parts to derive the following formula:

$$\text{vec}(\Omega_{\epsilon\epsilon}) = [(I_n \otimes \Phi) + (\Phi \otimes I_n)]^{-1} \left[ (e^{h\Phi} \otimes e^{h\Phi}) - I_{n^2} \right] \text{vec}(\Sigma),$$

where the  $\text{vec}(\cdot)$  operator stacks the columns of an  $n \times n$  matrix into an  $n^2 \times 1$  column vector and  $\otimes$  denotes the Kronecker product operator. For  $\Omega_{\rho\rho}$  recall that

$$\Omega_{\rho\rho} = \int_0^h \Phi_1(r) \Sigma \Phi_1(r)' dr.$$

Under the invertibility of  $\Phi$  we have

$$\Phi_1(r) = \int_0^r e^{s\Phi} ds = \Phi^{-1} (e^{r\Phi} - I_n).$$

Substituting this expression into  $\Omega_{\rho\rho}$  it can be shown that

$$\Omega_{\rho\rho} = \Phi^{-1} (\Omega_{\epsilon\epsilon} - \Sigma\Phi'_1 - \Phi_1\Sigma + h\Sigma) (\Phi^{-1})',$$

where  $\Phi_1 = \Phi_1(h) = \Phi^{-1}(e^{h\Phi} - I_n)$ . Turning to  $\Omega_{\phi\phi}$  we require evaluation of

$$\Omega_{\phi\phi} = 2 \int_0^h \Phi_2(r)\Sigma\Phi_2(r)'dr - \Phi_2(h)\Sigma \int_0^h \Phi_2(r)'dr - \int_0^h \Phi_2(r)dr\Sigma\Phi_2(h)' + h\Phi_2(h)\Sigma\Phi_2(h)'$$

Recall that

$$\Phi_2(r) = \int_0^r \Phi_1(s)ds = \int_0^r \Phi^{-1}(e^{s\Phi} - I_n) ds = \Phi^{-1}(\Phi_1(r) - rI_n) = \Phi^{-2}(e^{r\Phi} - I_n) - r\Phi^{-1}$$

which immediately gives  $\Phi_2(h) = \Phi^{-1}(\Phi_1 - hI_n)$ . Using this expression we find that

$$\begin{aligned} \int_0^h \Phi_2(r)\Sigma\Phi_2(r)'dr &= \Phi^{-1} \left( \int_0^h \Phi_1(r)\Sigma\Phi_1(r)'dr - \Sigma \int_0^h r\Phi_1(r)'dr \right. \\ &\quad \left. - \int_0^h r\Phi_1(r)dr\Sigma + \frac{h^3}{3}\Sigma \right) (\Phi^{-1})'. \end{aligned}$$

The first component in parentheses is simply  $\Omega_{\rho\rho}$ . Now let

$$\Phi_5 = \int_0^h r\Phi_1(r)dr = \Phi^{-1} \int_0^h r(e^{r\Phi} - I_n) dr = \Phi^{-1} \left( \Phi_4 - \frac{h^2}{2}I_n \right)$$

where

$$\Phi_4 = \int_0^h re^{r\Phi} dr = (h\Phi^{-1} - \Phi^{-2})e^{h\Phi} + \Phi^{-2} = \Phi^{-1}(he^{h\Phi} - \Phi_1).$$

Hence

$$\Phi_6 = \int_0^h \Phi_2(r)\Sigma\Phi_2(r)'dr = \Phi^{-1} \left( \Omega_{\rho\rho} - \Sigma\Phi'_5 - \Phi_5\Sigma + \frac{h^3}{3}\Sigma \right) (\Phi^{-1})'.$$

Next we consider

$$\Phi_7 = \int_0^h \Phi_2(r)dr = \int_0^h (\Phi^{-2}(e^{r\Phi} - I_n) - r\Phi^{-1}) dr = \Phi^{-2}(\Phi_1 - hI_n) - \frac{h^2}{2}\Phi^{-1}.$$

Combining these results we find that

$$\Omega_{\phi\phi} = 2\Phi_6 - \Phi_2\Sigma\Phi'_7 - \Phi_7\Sigma\Phi'_2 + h\Phi_2\Sigma\Phi'_2,$$

where  $\Phi_2 = \Phi_2(h) = \Phi^{-1}(\Phi_1 - hI_n)$ . Finally, the autocovariance matrix of  $\phi_{\tau h}$  is

$$\Omega_{1,\phi\phi} = - \int_0^h (\Phi_2(r) - \Phi_2(h)) \Sigma\Phi_2(r)'dr = \Phi_2\Sigma\Phi'_7 - \Phi_6.$$

## References

- Anderson, T.W., 1971. *The Statistical Analysis of Time Series*. Wiley, New York.
- Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Bergstrom, A.R., 1984. Continuous time stochastic models and issues of aggregation over time. In: Griliches, Z. and Intriligator, M.D. (Eds.), *Handbook of Econometrics*, vol. 2. North-Holland, Amsterdam.
- Bergstrom, A.R., 1990. *Continuous Time Econometric Modelling*. Oxford University Press, Oxford.
- Bergstrom, A.R., 1997. Gaussian estimation of mixed-order continuous-time dynamic models with unobservable stochastic trends from mixed stock and flow data. *Econometric Theory* 13, 467–505.
- 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., 2018. Frequency domain estimation of cointegrating vectors with mixed frequency and mixed sample data. *Economics Discussion Paper*, University of Essex.
- Chambers, M.J., McCrorie, J.R., 2007. Frequency domain estimation of temporally aggregated Gaussian cointegrated systems. *Journal of Econometrics* 136, 1–29.
- Corbae, D., Ouliaris, S., Phillips, P.C.B., 1994. A reexamination of the consumption function using frequency domain regressions. *Empirical Economics* 19, 595–609.
- Davidson, J., 1994. *Stochastic Limit Theory*. Oxford University Press, Oxford.
- Dunsmuir, W., 1979. A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. *Annals of Statistics* 7, 490–506.
- Ghysels, E., 2016. Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics* 193, 294–314.
- 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.
- Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.
- McCrorie, J.R., 2000. Deriving the exact discrete analog of a continuous time system. *Econometric Theory* 16, 998–1015.

- Miller, J.I., 2016. Conditionally efficient estimation of long-run relationships using mixed-frequency time series. *Econometric Reviews* 35, 1142–1171.
- Phillips, P.C.B., 1973. The problem of identification in finite parameter continuous time models. *Journal of Econometrics* 1, 351–362.
- 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, 967–80.
- Phillips, P.C.B., 1991c. Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.
- Phillips, P.C.B., Durlauf, S.N., 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.
- Rosenblatt, M., 1984. Asymptotic normality, strong mixing and spectral density estimates. *Annals of Statistics* 12, 1167–1180.
- Shiller, R.J., 2000. *Irrational Exuberance*. Princeton University Press, Princeton.
- Thornton, M.A., 2018. Exact observable representations of linear continuous time models with mixed frequency data. University of York.
- Zadrozny, P., 1988. Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies. *Econometric Theory* 4, 108–124.

**Table I** Bias and RMSE of Estimators

Estimator	Specification of $u(t)$					
	White Noise		AR $\phi = -5$		AR $\phi = -10$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
OLS	-0.0083	0.0248	-0.0014	0.0320	0.0099	0.0344
FD	-0.0001	0.0071	0.0052	0.0246	0.0177	0.0300
FDA	-0.0001	0.0070	0.0052	0.0245	0.0177	0.0300
FM(Auto)	0.0001	0.0071	0.0054	0.0249	0.0179	0.0303
VAR(1)	0.0001	0.0100	0.0057	0.0251	0.0180	0.0304
VAR(2)	0.0001	0.0075	0.0055	0.0270	0.0180	0.0320
VAR(4)	0.0001	0.0072	0.0054	0.0288	0.0179	0.0335

**Table II** Estimates of  $C$  in Stock Price-Dividend Model

Estimator	Estimate	Std. Err.	Estimator	Estimate	Std. Err.
OLS	0.5212	0.0070	FM(Auto)	0.5253	0.0161
FD(4)	0.5247	0.0054	FM(4)	0.5189	0.0110
FD(12)	0.5217	0.0058	FM(12)	0.5253	0.0152
FD(32)	0.5200	0.0067	FM(32)	0.5258	0.0171
FDA(4)	0.5251	0.0058	VAR(1)	0.6009	0.3653
FDA(12)	0.5217	0.0061	VAR(2)	0.6290	0.3648
FDA(32)	0.5200	0.0068	VAR(4)	0.6292	0.3480

VAR Likelihood Ratio Tests

Test	Statistic	$p$ -value	Test	Statistic	$p$ -value
1 vs. 2	2.9377	0.5683	1 vs. 4	4.2928	0.9776