University of Essex

Department of Economics

Discussion Paper Series

The Estimation of Continuous Time Models with Mixed Frequency Data

Marcus Chambers

The Estimation of Continuous Time Models

with Mixed Frequency Data

Marcus J. Chambers
University of Essex

January 2016

Abstract

This paper derives exact representations for discrete time mixed frequency data generated by an underlying multivariate continuous time model. Allowance is made for different combinations of stock and flow variables as well as deterministic trends, and the variables themselves may be stationary or nonstationary (and possibly cointegrated). The resulting discrete time representations allow for the information contained in high frequency data to be utilised alongside the low frequency data in the estimation of the parameters of the continuous time model. Monte Carlo simulations explore the finite sample performance of the maximum likelihood estimator of the continuous time system parameters based on mixed frequency data, and a comparison with extant methods of using data only at the lowest frequency is provided. An empirical application demonstrates the methods developed in the paper and it concludes with a discussion of further ways in which the present analysis can be extended and refined.

Keywords. Continuous time; mixed frequency data; exact discrete time models; stock and flow variables.

J.E.L. classification number. C32.

Address for Correspondence: Marcus J. Chambers, Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, England.
Tel: +44 (0)1206 872756; fax: +44 (0)1206 872724; e-mail: mchamb@essex.ac.uk.

1. Introduction

In multivariate models it is not uncommon to find that the variables of interest are observed at different frequencies. A leading example where this arises is in models containing both financial and macroeconomic variables, the former being observable at high frequencies (e.g. daily, hourly or minute-by-minute) and the latter at lower frequencies (often monthly or quarterly). The default method in such cases is usually to express all data in terms of a common frequency, typically by aggregating the high frequency data to the lowest frequency, which inevitably results in a lot of information contained in the higher frequency data being discarded. If utilising this information can lead to better inferences about the relationships between variables, then it is important to develop methods that enable the information contained in the high frequency data to be retained alongside the low frequency data.

In recent years a variety of methods have been proposed to deal with mixed frequency data. Prominent among these in a regression framework has been the Mixed Data Sampling (MIDAS) approach proposed by Ghysels, Santa-Clara and Valkanov (2002, 2006). A typical model in this framework contains a dependent variable that is observed less frequently than the explanatory variables, and the MIDAS method makes use of a polynomial weighting scheme to aggregate the high frequency regressors. A key feature of this approach is that the weighting scheme depends only on a small number of unknown parameters (possibly only one or two) which makes this a parsimonious method of aggregating the high frequency data. The basic MIDAS approach has been refined further by Ghysels, Sinko and Valkanov (2007), Andreou, Ghysels and Kourtellos (2010) and Foroni, Marcellino and Schumacher (2015), among others.

Another area in which progress has been made recently in the analysis of mixed frequency data is that of cointegration. Miller (2010) considers a general cointegrating regression in which the integrated regressors may be mismeasured, missing, observed at different frequencies or have certain other types of irregularity, and derives the asymptotic properties of least squares and related inferential techniques. Issues of testing for the presence of cointegration amongst mixed frequency (and temporally aggregated) data series have been explored by Ghysels and Miller (2014a, 2014b) while Seong, Ahn and Zadrozny (2013) consider estimation of vector error correction models with mixed frequency data via state space representations and the Kalman filter. A common implication of this body of work is that incorporating the additional information in high frequency data, rather than discarding it, can lead to better properties of estimators and inferential procedures and to improved forecasts.

To date, the analysis of mixed frequency data has, with one notable exception discussed below, been conducted firmly within the realm of discrete time models. A common approach is to consider a vector of variables observed at two different frequencies, one being high (e.g. weekly), say y^H, and one being low (e.g. quarterly), say y^L. Suppose the low frequency variables are observed once every k (high frequency) time periods, where k is an integer. Then the aim is to specify a model for the vector comprising y_{kt}^L, y_{kt}^H, $y_{kt-1}^H, \ldots, y_{kt-(k-1)}^H$ ($t = 1, \ldots, T_L$), where T_L denotes the number of low frequency observations. Recent work in this vein has focused on vector autoregressive (VAR) models, variants of which are in widespread use in empirical macroeconomics and finance where data are often available at different frequencies. Examples of this approach include Anderson *et al.* (2015), Foroni, Ghysels and Marcellino (2013), Foroni and Marcellino (2013, 2014a, 2014b), and Ghysels

(2014), which collectively cover issues of identification, estimation, inference, and impulse response analysis in the context of regular and structural VARs.[1] A common feature of this approach is that the high frequency time scale is, in effect, assumed to be the fundamental time scale determining the dynamic relationships between the variables and, hence, the model specification is tied to the (arbitrary) highest sampling frequency.

The notable exception referred to at the start of the last paragraph is Zadrozny (1988) who considers a continuous time autoregressive moving average (ARMA) model[2] with mixed quarterly and annual data for a set of stock and flow variables. The model is cast in state space form and Kalman filter recursions are used to compute the Gaussian likelihood function. An advantage of the continuous time specification is that the model of interest, whose parameters are estimated, is not tied to the (highest) sampling frequency. Instead, the dynamic process of interest operates more frequently (i.e. continuously) than the highest observation frequency. This is a more realistic setting for many financial variables, such as financial assets, that are traded (nearly) continuously. It is also, arguably, more relevant for macroeconomic aggregates which, although only observed at low frequencies, may be subject to changes, in response to stimuli, at *any* point within the sampling interval.[3]

In this paper we consider the estimation of continuous time models formulated as a system of stochastic differential equations when the observed data are recorded at different frequencies. The approach is structural in the sense that the parameters of interest are those of the continuous time model which govern the (unobservable) dynamics of the observed variables. The temporal aggregation of stock and flow variables is taken into account in the derivation of exact discrete models which have the property that data generated by the continuous time system satisfy these discrete time representations exactly – there is no approximation error involved in these representations of the discrete time data. We treat the cases of common sampling (i.e. all stocks or all flows) and mixed sampling (a combination of stocks and flows), both with mixed frequency data. The discrete time representations have wide applicability because very few restrictions are placed on the underlying continuous time system, allowing for nonstationary (including cointegrated) series, as well as stationary series.

The paper is organised as follows. Section 2 introduces the mixed frequency sampling framework and the continuous time system under consideration is defined. It also motivates the idea that discarding information by aggregating to a lower frequency has adverse impacts on the properties of estimators by reporting some simulation results based on a simple univariate continuous time model. The more general sampling framework is introduced and the continuous time system under consideration is defined. Section 3 is concerned with the derivation of exact discrete time models in the case of common sampling with mixed frequencies, results being reported for the cases of stock sampling (Theorem 1) and flow sampling (Theorem 2). Section 4 considers the more complicated situation in which there is a mixed sample of stock and flow variables, the exact representation being reported in Theorem 3. Estimation, based on the Gaussian likelihood, is discussed in section 5, which also covers some computational issues and reports the results of a simulation exercise

---

[1]Foroni and Marcellino (2014b) also consider DSGE models under mixed frequency sampling.

[2]Zadrozny (1988) also extends the ARMA model to allow for exogenous variables.

[3]This argument has been made most eloquently by Bergstrom (1990).

involving stationary as well as cointegrated stock variables observed at mixed frequencies. Use of the exact discrete time model for mixed frequency data is shown to result in estimators with smaller bias and root mean square error (RMSE) than estimators based on extant methods of aggregating all data to the lowest frequency. An empirical application is provided in section 6 in which monthly price indices for the UK and US are combined with daily, weekly and monthly exchange rate data in an investigation of long run purchasing power parity (PPP) relationships. Results based on the lower frequency exchange rate data do not reject the PPP restrictions but this appears to be due to the estimated parameters having large standard errors – they are not sufficiently informative to be able to reject the restrictions. Use of the daily exchange rate data, however, results in parameter estimates with much smaller standard errors which result in a strong rejection of the PPP restrictions. Section 7 concludes the paper and points to some directions for further research, and proofs of the three theorems, as well as some additional results, are provided in the Appendix.

2. Temporal aggregation and data sampling: the modelling framework

The general framework is concerned with the $n \times 1$ vector $y(t) = [y_1(t)', y_2(t)']'$ where $y_i(t)$ is $n_i \times 1$ $(i = 1, 2)$ and $n_1 + n_2 = n$. Attention is given to the situation where the variables are observed at two different sampling intervals (or frequencies), although the methods can be generalised to account for more than two frequencies. The sub-vector y_1 contains the variables observed at the highest frequency while y_2 contains the variables observed at the lowest frequency; these shall be referred to as the high frequency variables and the low frequency variables, respectively. For convenience (and without loss of generality) the time index shall be normalised to unity to correspond to the low frequency sampling interval. The low frequency observations are therefore, for stock variables, of the form

$$\{y_{2t}\}_{t=0}^{T} = \{y_{20}, y_{21}, \ldots, y_{2T}\} = \{y_2(0), y_2(1), \ldots, y_2(T)\},$$

while for flow variables the observed vector sequence is

$$\{Y_{2t}\}_{t=0}^{T} = \{Y_{20}, Y_{21}, \ldots, Y_{2T}\} = \left\{ \int_{-1}^{0} y_2(r)dr, \int_{0}^{1} y_2(r)dr, \ldots, \int_{T-1}^{T} y_2(r)dr \right\}.$$

In both cases T is regarded as being the effective sample size in view of the dynamic models derived in subsequent sections being conditioned on an initial value (either y_{20} or Y_{20}). In this framework the effective span is also equal to T in view of the low frequency sampling interval being normalised to unity. The sampling interval for the high frequency variable will be denoted h, so that for stock variables the observations are of the form

$$\{y_{1,\tau h}\}_{\tau=0}^{N} = \{y_{10}, y_{1h}, \ldots, y_{1,Nh}\} = \{y_1(0), y_1(h), \ldots, y_1(Nh)\},$$

while for flow variables the sequence is

$$\{Y_{1,\tau h}\}_{\tau=0}^{N} = \{Y_{10}, Y_{1h}, \ldots, Y_{1,Nh}\} = \left\{ \frac{1}{h} \int_{-h}^{0} y_1(r)dr, \frac{1}{h} \int_{0}^{h} y_1(r)dr, \ldots, \frac{1}{h} \int_{Nh-h}^{Nh} y_1(r)dr \right\}.$$

For the high frequency variables, therefore, N denotes the (effective) number of observations,

3

where $Nh = T$ (implying that $N = T/h$). It is also convenient to assume that the high sampling frequency (the inverse of the sampling interval) is an integer; this will be denoted k and clearly $k = h^{-1}$. For example, if the low frequency data are observed annually and the high frequency data monthly, then $h = \frac{1}{12}$ and $k = 12$, reflecting the fact that the high frequency data are sampled twelve times more frequently than the low frequency data. Note, too, that the high frequency flow variables have been normalised by dividing by h which expresses the observed flow in terms of the low frequency equivalent. For example, if y_1 represents the rate of flow of consumers' income and it is observed quarterly, so that $h = \frac{1}{4}$, then dividing by h expresses the observed quarterly flow as an annual equivalent. Although this normalisation is not essential in stationary models it does have some importance in nonstationary/cointegrated systems; see, for example, Chambers (2011, p.160). Figure 1 illustrates the relationship between the high and low frequency sampling schemes; note that $kh = 1$, and in effect $\tau = tk$ or $\tau h = tkh = t$.

High (τh)   0 h $2h$ kh $2kh$ Nh

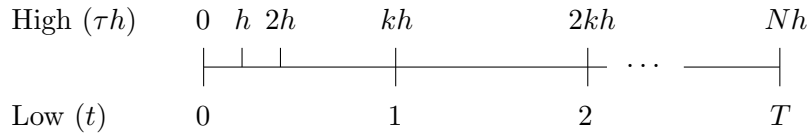Low (t)   0   1   2 T

Figure 1. High and low frequency sampling schemes.

At this stage it is, perhaps, useful to give some indication as to the importance of using higher frequency data when available, and to do this a simple univariate example is provided. Suppose $y(t)$ evolves according to the stochastic differential equation

$$dy(t) = ay(t)dt + \sigma dW(t), \quad t > 0, \tag{1}$$

where $W(t)$ denotes a Wiener process, $y(0) = 0$ and $a < 0$ to ensure stationarity. Furthermore, in accordance with the framework defined above, suppose that $y(t)$ is a stock variable observed at the points $h, 2h, \ldots, Nh = T$. Three sampling intervals are considered, $h = 1, \frac{1}{4}, \frac{1}{12}$, as well as two spans, $T = 25, 100$, and two values for the autoregressive parameter, $a = -5, -1$. One interpretation of these values is that if $h = 1$ corresponds to a year, then $h = \frac{1}{4}$ and $h = \frac{1}{12}$ correspond, respectively, to quarterly and monthly sampling intervals, while the two spans correspond to 25 and 100 years. The observed data can be shown to satisfy

$$y(th) = e^{ah}y(th - h) + \epsilon(th), \quad t = 1, \ldots, N,$$

where

$$\epsilon(th) = \sigma \int_{th-h}^{th} e^{a(th-s)}dW(s)$$

is an independent $N(0, \sigma_\epsilon^2)$ sequence with $\sigma_\epsilon^2 = \sigma^2(e^{2ah} - 1)/2a$.

The results from 10,000 replications appear in Table 1, in which the bias and root mean square error (RMSE) of the maximum likelihood estimator of a are reported, the likelihood function having been concentrated with respect to the parameter σ^2. For all four combinations of a and T it can be seen that increasing the sampling interval h (i.e. decreasing the sampling frequency) leads to an increase in both the bias and RMSE of the estimator. Furthermore, for a given value of a, increasing the span T for a given frequency (and thereby increasing sample size N) also reduces bias and RMSE. Also, for given span T, the bias and RMSE are smaller the closer is a to zero. These results suggest that it is desirable to use the highest frequency data available if at all possible.[4]

Returning to the more general framework it is assumed in what follows that the $n \times 1$ vector $y(t)$ satisfies the stochastic differential equation system

$$dy(t) = [\mu + \gamma t + Ay(t)]dt + \zeta(dt), \quad t > 0, \tag{2}$$

where μ and γ are $n \times 1$ vectors of intercept and trend parameters, A is an $n \times n$ matrix of coefficients, and $\zeta(dt)$ is an $n \times 1$ vector of (white noise) random measures satisfying $E[\zeta(dt)] = 0$, $E[(\zeta(dt)\zeta(dt)'] = \Sigma dt$ (where Σ is a symmetric positive definite matrix), and $E[\zeta(\Delta_1)\zeta(\Delta_2)'] = 0$ for non-overlapping time intervals Δ_1 and Δ_2; see, for example, Bergstrom (1984) for details of random measures and their use in econometrics. The aim is to estimate the vectors μ and γ and the matrices A and Σ. The elements of these vectors and matrices may be entirely unrestricted (apart from ensuring the symmetry and positive definiteness of Σ) or they may be known functions of an underlying parameter vector θ. Either way it is necessary to relate the parameters of the system (2) to the observed data.

The system of equations in (2) can be regarded as a continuous time VAR(1) with a deterministic trend, and the focus is on deriving exact discrete models that have the property that data generated by (2) satisfy these discrete time representations exactly and, moreover, are expressed in terms of both the high and low frequency variables. An alternative approach, adopted by Zadrozny (1988), is to write the system in state space form and to use the Kalman filter to derive the Gaussian likelihood function. Zadrozny (1988) deals with a more general continuous time dynamic system, namely a continuous time ARMA(p, q) (without deterministics), but the precise relationships between the discrete time observations are not needed in his approach. It can be of interest, however, to derive explicitly these exact discrete time models and to compare them with the existing mixed frequency VARs that have appeared in the literature, and this partly motivates the approach followed here. It is nevertheless possible, in principle, to derive discrete time models for more general continuous time ARMA(p, q) systems with mixed frequency data by extending the methods of Chambers and Thornton (2012), albeit at the cost of additional notational complexity; such extensions are left for future work with the current contribution focusing on the principles of dealing with mixed frequency data in the setting of (2).

It is perhaps worth reiterating that the usual approach to estimating the parameters of (2) with data observed at mixed frequencies is to aggregate the high frequency data to coincide with the lowest frequency, thereby discarding information contained in the high

---

[4]Abstracting, of course, from other complications that may arise when sampling more frequently, such as seasonality and microstructure noise.

frequency data. The approach adopted here, however, deals with the solution to (2) in terms of the high frequency timescale and then manipulates the resulting expressions so that no data are discarded. The precise formulations are given in the following sections.

3. Models with common data sampling

In this section discrete time models are derived for the situations in which both frequency variables are either all stocks or all flows. The next section deals with the mixed sample case. The starting point is the solution to (2) which is given by

$$y(t) = e^{At}y(0) + \int_0^t e^{A(t-s)}(\mu + \gamma s)ds + \int_0^t e^{A(t-s)}\zeta(ds), \quad t > 0, \tag{3}$$

where the matrix exponential is defined as

$$e^{At} = \sum_{j=0}^\infty \frac{(At)^j}{j!} = I_n + At + \frac{1}{2!}A^2t^2 + \frac{1}{3!}A^3t^3 + \dots, \tag{4}$$

where I_n denotes the $n \times n$ identity matrix. The above solution is unique in a mean square sense and the definition of the integral with respect to the random measure can be found in Bergstrom (1984).

3.1. Stock variables

When both y_1 and y_2 are stock variables the objective is to derive a discrete time model that simultaneously incorporates the low frequency observations $y_{2t} = y_2(t)$ $(t = 0, \dots, T)$ and the high frequency observations $y_{1,\tau h} = y_1(\tau h)$ $(\tau = 0, \dots, N)$. What this effectively reduces to is finding a representation that holds for the points $t = 1, \dots, T$, given the value for $t = 0$, but which also contains the intermediate points $t - h, t - 2h, \dots, t - 1 + h$ between each t and $t-1$; these intermediate points correspond to the observations on y_1. The solution (3) can be used to relate $y(t)$ to $y(t - h)$ in the form[5]

$$y(t) = c(t) + Fy(t - h) + \epsilon(t), \tag{5}$$

where $F = e^{Ah}$ and the deterministic vector $c(t)$ and random disturbance vector $\epsilon(t)$ are defined by

$$c(t) = \int_{t-h}^t e^{A(t-s)}(\mu + \gamma s)ds, \tag{6}$$

$$\epsilon(t) = \int_{t-h}^t e^{A(t-s)}\zeta(ds). \tag{7}$$

For the purpose of deriving discrete time representations for the observed mixed frequency data, it is convenient to partition the system (5) in accordance with y_1 and y_2 as follows:

$$
\begin{aligned}
y_1(t) &= c_1(t) + F_{11}y_1(t - h) + F_{12}y_2(t - h) + \epsilon_1(t), &(8)\\
y_2(t) &= c_2(t) + F_{21}y_1(t - h) + F_{22}y_2(t - h) + \epsilon_2(t), &(9)
\end{aligned}
$$

---

[5]In (5), t can take any value corresponding to τh.

in which

$$c(t) = \begin{bmatrix} c_1(t) \\ c_2(t) \end{bmatrix}, \quad F = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}, \quad \epsilon(t) = \begin{bmatrix} \epsilon_1(t) \\ \epsilon_2(t) \end{bmatrix}.$$

This autoregressive representation depicts the law of motion for both frequency variables but the problem in using it as a basis for estimation is that the observations on $y_2(t)$ are not observed at intervals of length h; this variable is only observed when t is an integer. Nevertheless, (8) and (9) form the basis for a discrete time representation which is presented in Theorem 1 below.

Theorem 1. *Let $y(t)$ be generated by (2) with y_1 and y_2 consisting of stock variables which are observed as $y_{1,\tau h} = y_1(\tau h)$ ($\tau = 0, \ldots, N$) and $y_{2t} = y_2(t)$ ($t = 0, \ldots, T$). Then the observations satisfy, for $t = 1, \ldots, T$,*

$$y_{1t} = b_{1t} + \sum_{j=1}^{k} B_{11,j} y_{1,t-jh} + B_{12,0} y_{2,t-1} + \eta_{1t}, \tag{10}$$

$$y_{1,t-h} = b_{1,t-h} + \sum_{j=1}^{k-1} B_{11,j} y_{1,t-h-jh} + B_{12,1} y_{2,t-1} + \eta_{1,t-h}, \tag{11}$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$y_{1,t-(k-1)h} = b_{1,t-(k-1)h} + B_{11,1} y_{1,t-1} + B_{12,k-1} y_{2,t-1} + \eta_{1,t-(k-1)h}, \tag{12}$$

$$y_{2t} = b_{2t} + \sum_{j=1}^{k} B_{21,j} y_{1,t-jh} + B_{22} y_{2,t-1} + \eta_{2t}, \tag{13}$$

where

$$B_{11,j} = \begin{cases} F_{11}, & j = 1, \\ F_{12} F_{22}^{j-2} F_{21}, & j = 2, \ldots, k, \end{cases} \qquad B_{12,j} = F_{12} F_{22}^{k-j-1}, \quad j = 0, \ldots, k-1,$$

$$B_{21,j} = F_{22}^{j-1} F_{21}, \quad j = 1, \ldots, k, \qquad\qquad B_{22} = F_{22}^k,$$

$$b_{1,t-jh} = c_1(t - jh) + F_{12} \sum_{l=1}^{k-j-1} F_{22}^{l-1} c_2(t - jh - lh), \quad j = 0, \ldots, k-1,$$

$$b_{2t} = \sum_{j=0}^{k-1} F_{22}^j c_2(t - jh).$$

Furthermore, $\eta_t = [\eta'_{1t}, \eta'_{1,t-h}, \ldots, \eta'_{1,t-(k-1)h}, \eta'_{2t}]'$ is a vector white noise process with co-variance matrix

$$\Omega_\eta = E(\eta_t \eta'_t) = \sum_{j=0}^{k-1} F_{\eta,j} \Omega_\epsilon F'_{\eta,j},$$

where $\Omega_\epsilon = E(\epsilon(t)\epsilon(t)') = \int_0^h e^{Ar} \Sigma e^{A'r} dr$ and the $F_{\eta,j}$ ($j = 0, \ldots, k-1$) are defined in the proof of the Theorem.

Remark 1. A key feature of the discrete time representation in Theorem 1 is that information contained in the high frequency series, y_1, is not discarded, because the intermediate observations on y_1 between each $t-1$ and t, given by $y_{1,t-kh}, y_{1,t-1+h}, y_{1,t-1+2h}, \ldots, y_{t-h}$, are included on the right-hand sides of (10)–(13). This contrasts with extant methods based on aggregating the high frequency observations to the lowest frequency; such a system would then be given by (8) and (9) with $h = 1$.

Remark 2. It is, perhaps, of use to consider a particular example to illustrate the nature of the discrete time representation in Theorem 1. Setting $h = \frac{1}{3}$ and taking time units to correspond to a quarter of a year implies that the high frequency variables are observed monthly and the low frequency variables quarterly.[6] Then $k = h^{-1} = 3$ and the resulting system of equations is

$$
\begin{aligned}
y_{1t} &= b_{1t} &+ B_{11,1}y_{1,t-h} &+ B_{11,2}y_{1,t-2h} &+ B_{11,3}y_{1,t-3h} &+ B_{12,0}y_{2,t-1} \\
&&&&&+ \eta_{1t}, \\
y_{1,t-h} &= b_{1,t-h} & &+ B_{11,1}y_{1,t-2h} &+ B_{11,2}y_{1,t-3h} &+ B_{12,1}y_{2,t-1} \\
&&&&&+ \eta_{1,t-h}, \\
y_{1,t-2h} &= b_{1,t-2h} & & &+ B_{11,1}y_{1,t-3h} &+ B_{12,2}y_{2,t-1} \\
&&&&&+ \eta_{1,t-2h}, \\
y_{2t} &= b_{2t} &+ B_{21,1}y_{1,t-h} &+ B_{21,2}y_{1,t-2h} &+ B_{21,3}y_{1,t-3h} &+ B_{22}y_{2,t-1} \\
&&&&&+ \eta_{2t}.
\end{aligned}
$$

Note that $y_{1,t-3h} = y_{1,t-1}$ and that, for each quarter, there is an equation for y_2 plus three equations for y_1, corresponding to each month in the quarter. None of these equations contains lags beyond those dated $t-1$, and all of the monthly observations on y_1 are included in the low frequency equation for y_2, thereby incorporating all the available intermediate information on the high frequency variable.

Remark 3. Following on from the example above it is worth commenting that the coefficient vectors and matrices are solely functions of the parameters of the continuous time system, even though the discrete time system is in the form of a mixed frequency VAR; see, for example, the representations in Foroni, Ghysels and Marcellino (2013). To emphasise this feature, the continuous time system contains $2n$ trend parameters (μ and γ), n^2 autoregressive parameters (A), and $n(n+1)/2$ covariance parameters (Σ). By way of contrast the unrestricted discrete time mixed frequency VAR contains $6n_1 + 2n_2$ trend parameters (b_{jt}), $6n_1^2 + 6n_1n_2 + n_2^2$ autoregressive parameters ($B_{ij,k}$ and B_{22}), and $(9n_1^2 + 6n_1n_2 + n_2^2 + 3n_1 + n_2)/2$ covariance parameters. A simple calculation reveals that the process of temporal aggregation imposes a total of $9n_1^2 + 6n_1n_2 + 5n_1$ restrictions. The parsimony achieved by taking the temporal aggregation into account is even greater with mixed frequency data than with data observed at a common frequency, and greater still with flow data, as will be seen below.

Remark 4. Even if the underlying continuous time system is ignored, and (8) and (9) are regarded as the VAR of interest that operates at the highest sampling interval h, Theorem 1 still provides the correct form of aggregated system for stock variables. This is because the b_{jt}, $B_{ij,k}$ and B_{22} are expressed as functions of $c(t)$ and F, which again provides parsimony over a mixed frequency VAR written directly in terms of the discrete time observations,

---

[6]For convenience it is assumed that all quarters are of equal length. Although each quarter contains three months not all months have the same number of days.

as in (10)–(13). Moreover, the representation takes into account the restrictions on the covariance matrix that arise from temporal aggregation, a feature that is often ignored in standard treatments of mixed frequency data.

Remark 5. The discrete time representation in (10)–(13) holds for both stationary and non-stationary (including cointegrated) continuous time systems in view of no restrictions having been placed on the continuous time system matrix A. For example, neither A, F nor any of their sub-matrices is required to be non-singular.

Remark 6. Theorem 1 shows, for stock variables, that a continuous time autoregressive model of order one translates into a discrete time autoregressive model of order one with mixed frequency data, as in the case where a common sampling frequency exists. The difference here is that accounting for mixed frequencies results in a more complicated pattern of restrictions on the discrete time data, and the covariance matrix Ω_η reflects the presence of the higher frequency components in its construction.

3.2. Flow variables

When both high and low frequency variables are observed as flows it is necessary to integrate the system to express it in terms of the observations. This could be achieved by integrating (5) directly or by integrating the discrete time representation in Theorem 1 itself, and it is the latter approach that is followed here. Although the low frequency variable, y_2, is observed as an integral over $(t-1, t]$, it is actually convenient to integrate instead over the interval $(t - h, t]$ that corresponds to the high frequency observations, as in the case of stock variables above. In manipulating these systems to eliminate unobservables (in particular the integral of the low frequency variable over intervals of length h) frequent use is made of the filter function $s(L^h)$ where

$$s(z) = 1 + z + z^2 + \ldots + z^{k-1} = \sum_{j=0}^{k-1} z^j \tag{14}$$

and L denotes the lag operator.[7] This is because the observable integral of y_2, denoted Y_{2t}, can be regarded as the sum of the unobservable integrals over intervals of length h, denoted $Y_{2t}^u = \int_{t-h}^t y_2(r)dr$, as follows:

$$Y_{2t} = \int_{t-1}^t y_2(r)dr = \sum_{j=0}^{k-1} \int_{t-jh-h}^{t-jh} y_2(r)dr = \sum_{j=0}^{k-1} Y_{2,t-jh}^u = s(L^h)Y_{2t}^u.$$

Some properties of a convolution of $s(z)$ with another matrix filter are given in Lemma 1 in the Appendix and are used in the proof of Theorem 2 below.

Theorem 2. *Let $y(t)$ be generated by (2) with y_1 and y_2 consisting of flow variables which are observed as*

$$Y_{1,\tau h} = \frac{1}{h} \int_{\tau h - h}^{\tau h} y_1(r)dr, \quad \tau = 0, \ldots, N, \quad Y_{2t} = \int_{t-1}^t y_2(r)dr, \quad t = 0, \ldots, T.$$

---

[7] The lag operator satisfies $L^j x_t = x_{t-j}$ where j need not be an integer (as is the case here when considering L^h).

Furthermore, define the aggregated high-frequency flow variables

$$Y^s_{1,\tau h} = s(L^h)Y_{1,\tau h} = \sum_{j=0}^{k-1} Y_{1,\tau h-jh}, \quad \tau = k, \ldots, N.$$

Then Y^s_1 and Y_2 satisfy, for $t = 1, \ldots, T$,

$$Y^s_{1t} = c_{1t} + \sum_{j=1}^{k} C_{11,j}Y^s_{1,t-jh} + C_{12,0}Y_{2,t-1} + u_{1t}, \tag{15}$$

$$Y^s_{1,t-h} = c_{1,t-h} + \sum_{j=1}^{k-1} C_{11,j}Y^s_{1,t-h-jh} + C_{12,1}Y_{2,t-1} + u_{1,t-h}, \tag{16}$$

$$\vdots \qquad\qquad \vdots$$

$$Y^s_{1,t-(k-1)h} = c_{1,t-(k-1)h} + C_{11,1}Y^s_{1,t-1} + C_{12,k-1}Y_{2,t-1} + u_{1,t-(k-1)h}, \tag{17}$$

$$Y_{2t} = c_{2t} + \sum_{j=1}^{k} C_{21,j}Y^s_{1,t-jh} + C_{22}Y_{2,t-1} + u_{2t}, \quad t = 2, \ldots, T, \tag{18}$$

where $c_{1t} = h^{-1}s(L^h)b^*_{1t}$, $c_{2t} = s(L^h)b^*_{2t}$, $b^*_{jt} = \int_{t-h}^{t} b_{jr}dr$ $(j = 1, 2)$,

$$C_{11,j} = B_{11,j}, \quad j = 1, \ldots, k, \quad C_{12,j} = h^{-1}B_{12,j}, \quad j = 0, \ldots, k-1,$$

$$C_{21,j} = hB_{21,j}, \quad j = 1, \ldots, k, \quad C_{22} = B_{22},$$

and the b_{jt} $(j = 1, 2)$, $B_{11,j}$, $B_{21,j}$ $(j = 1, \ldots, k)$, $B_{12,j}$ $(j = 0, \ldots, k-1)$ and B_{22} are defined in Theorem 1. Furthermore, $u_t = [u'_{1t}, u'_{1,t-h}, \ldots, u'_{1,t-(k-1)h}, u'_{2t}]'$ is a vector MA(1) process with

$$\Omega_{u0} = E(u_t u'_t) = K_h \left(\sum_{j=0}^{2k-2} H_j \Omega_{\epsilon 0} H'_j + \sum_{j=1}^{2k-1} \left(H_j \Omega'_{\epsilon 1} H'_{j-1} + H_{j-1}\Omega_{\epsilon 1} H'_j \right) \right) K_h,$$

$$\Omega_{u1} = E(u_t u'_{t-1}) = K_h \left(\sum_{j=k}^{2k-2} H_j \Omega_{\epsilon 0} H'_{j-k} + \sum_{j=k-1}^{2k-2} H_j \Omega_{\epsilon 1} H'_{j-k+1} + \sum_{j=k+1}^{2k-2} H_j \Omega'_{\epsilon 1} H'_{j-k-1} \right) K_h,$$

and where the remaining components are defined in the proof of the Theorem.

Remark 7. The discrete time representation in Theorem 2 contains aggregated versions of the high frequency flows. It would be possible, if desired, to express the equations in terms of Y_{1t} and its lags directly, although there is nothing to be gained from doing so for estimation purposes.[8] One application where this would be necessary is forecasting, but all that is really needed is to express Y^s_{1t} in the form

$$Y^s_{1t} = Y_{1t} + \sum_{l=1}^{k-1} Y_{1,t-lh}$$

---

[8]Such a representation would appear less parsimonious but would emphasise the fact that up to $2k - 1$ lags of the observed high frequency flow appear in the discrete time model.

and to incorporate the lags on the right-hand side of (15).

Remark 8. The discrete time representations in (15)–(18) contain additional lags of the high frequency variable y_1 owing to the fact that additional backward substitution is required to eliminate all unobservable components (which arise because of the nature of flow variables). The resulting discrete time equations implicitly contain all observations on the high frequency variable over the interval $(t-2, t]$ that provide additional information regarding the dynamics of the system and which may potentially result in better estimates of the underlying continuous time parameters.

Remark 9. As in the case of stock variables no assumptions have been made concerning the continuous time parameter matrix A that governs stationarity. As a result the discrete time representation in Theorem 2 is also applicable to both stationary and non-stationary (including cointegrated) systems.

Remark 10. A key difference of the discrete model for flows when compared to that for stocks is the presence of a moving average component in the disturbance vector. This is a common feature when flow variables are concerned and arises due to the integration involved in determining the observations; see Working (1960).

Remark 11. The covariance matrices, $\Omega_{\epsilon 0}$ and $\Omega_{\epsilon 1}$, appearing in the definitions of the covariance matrices of u_t, Ω_{u0} and Ω_{u1}, correspond to $E(\epsilon_t \epsilon_t')$ and $E(\epsilon_t \epsilon_{t-h}')$, respectively, where $\epsilon_t = \int_{t-h}^{t} \epsilon(r) dr$.

Remark 12. Remarks 3 and 4, relating to the parsimony of the continuous time approach, apply equally, if not more so, in the case of flow variables. In fact, the disturbance vector being MA(1) introduces an additional $(kn_1 + n_2)^2$ parameters into the discrete time system via the first-order autocovariance matrix, yet all of the discrete time parameters remain functions of the same number of parameters in the underlying continuous time system (2).

Attention now turns to situations in which the variables are a mixture of stocks and flows as well as the observations being available at different sampling frequencies.

4. Models with mixed data sampling

In many applications the variables of interest are a mixture of stocks and flows, and it is therefore of practical importance to extend the discrete time representations in section 3 to allow for such circumstances. In the most general scenario both the high and low frequency observations would consist of stocks and flows, but in order to avoid unnecessary complication the model considered is one where the high frequency variables are stocks and the low frequency variables are flows. While this distinction is somewhat arbitrary it does serve to highlight the issues involved in treating a mixed sample. It also has the advantage, however, of encapsulating situations where the high frequency variables are financial variables such as asset prices, exchange rates and interest rates (observed as stocks) and the low frequency variables are macroeconomic aggregates such as income/output, consumption, and investment expenditures (observed as flows).

In order to derive an exact discrete model, the following assumption is made concerning a sub-matrix of A in (2).

Assumption 1. The $n_1 \times n_1$ sub-matrix A_{11} of A is non-singular.

Remark 13. The matrix A_{11} governs the response of $dy_1(t)$ to the level of $y_1(t)$ in the continuous time system (2) and its invertibility enables the unobservable variable $\int_{t-h}^{t} y_1(r)dr$ to be expressed in terms of the observable $y_1(t) - y_1(t-h)$ and $\int_{t-h}^{t} y_2(r)dr$, the latter being observable once the operator $s(L^h)$ has been applied. This type of assumption was also made concerning stock variables by Agbeyegbe (1987, 1988) (Assumption 2 in both cases) and Simos (1996) (Assumption 3) although both authors also made the assumption that the entire matrix A was nonsingular – this additional assumption is not required here. The assumption does, however, rule out the possibility of cointegration between a set of nonstationary stock variables.[9]

In presenting the exact discrete model for the mixed sample case it is convenient to partition the matrix functions e^{Ar}, $P_0(r)$ and $P_1(r)$ in accordance with y_1 and y_2 as follows:

$$F(r) = e^{Ar} = \begin{bmatrix} F_{11}(r) & F_{12}(r) \\ F_{21}(r) & F_{22}(r) \end{bmatrix}, \quad P_j(r) = \begin{bmatrix} P_{j,11}(r) & P_{j,12}(r) \\ P_{j,21}(r) & P_{j,22}(r) \end{bmatrix}, \quad j = 0, 1.$$

These functions are used as weights in integrals with respect to the random measure $\zeta(dt)$ that arise in deriving the exact discrete model. For example, $F(r)$ is the weight function in the definition of the random vector $\epsilon(t)$ in (7). The presentation of the results in Theorem 3 is also aided by grouping some of the definitions together in Table 2.

Theorem 3. *Let $y(t)$ be generated by (2) with y_1 consisting of stock variables, which are observed as $y_{1,\tau h} = y_1(\tau h)$ ($\tau = 0, \ldots, N$), and y_2 consisting of flow variables, which are observed as*

$$Y_{2t} = \int_{t-1}^{t} y_2(r)dr, \quad t = 0, \ldots, T.$$

Furthermore, define the aggregated high frequency stock variables

$$y_{1,\tau h}^{s} = s(L^h)y_{1,\tau h} = \sum_{j=0}^{k-1} y_{1,\tau h-jh}, \quad \tau = k, \ldots, N.$$

---

[9]Subsequent work will explicitly investigate the case of cointegrated continuous time systems with mixed frequency data in more detail, thereby extending existing results in Chambers (2009, 2011) to the mixed frequency setting.

Then, under Assumption 1, the observations satisfy, for $t = 1, \ldots, T$,

$$\Delta_h y_{1t}^s = \phi_{1t} + \sum_{j=1}^{k} \Phi_{11,j} \Delta_h y_{1,t-jh}^s + \Phi_{12,0} Y_{2,t-1} + \xi_{1t}, \tag{19}$$

$$\Delta_h y_{1,t-h}^s = \phi_{1,t-h} + \sum_{j=1}^{k-1} \Phi_{11,j} \Delta_h y_{1,t-jh}^s + \Phi_{12,1} Y_{2,t-1} + \xi_{1,t-h}, \tag{20}$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$\Delta_h y_{1,t-(k-1)h}^s = \phi_{1,t-(k-1)h} + \Phi_{11,1} \Delta_h y_{1,t-1}^s + \Phi_{12,k-1} Y_{2,t-1} + \xi_{1,t-(k-1)h}, \tag{21}$$

$$Y_{2t} = \phi_{2t} + \sum_{j=1}^{k} \Phi_{21,j} \Delta_h y_{1,t-jh}^s + \Phi_{22} Y_{2,t-1} + \xi_{2t}, \tag{22}$$

where $\Delta_h = 1 - L^h$,

$$\phi_{1,t-jh} = s(L^h) \left(g_{1,t-jh} + G_{12} \sum_{l=1}^{k-j-1} G_{22}^{l-1} g_{2,t-jh-lh} \right), \quad \phi_{2t} = s(L^h) \sum_{j=0}^{k-1} G_{22}^{j} g_{2,t-jh},$$

$$\Phi_{11,j} = \begin{cases} G_{11}, & j = 1, \\ G_{12} G_{22}^{j-2} G_{21}, & j = 2, \ldots, k, \end{cases} \qquad \Phi_{12,j} = G_{12} G_{22}^{k-j-1}, \quad j = 0, \ldots, k-1,$$

$$\Phi_{21,j} = G_{22}^{j-1} G_{21}, \quad j = 1, \ldots, k, \qquad\qquad \Phi_{22} = G_{22}^{k},$$

and the remaining components are defined in Table 2. Furthermore,

$$\xi_t = [\xi_{1t}', \xi_{1,t-h}', \ldots, \xi_{1,t-(k-1)h}', \xi_{2t}']'$$

is a vector MA(1) process with

$$\Omega_{\xi 0} = E(\xi_t \xi_t') = k \Omega_{\bar{\xi} 0} + \sum_{j=1}^{k-1} (k-j) \left(\Omega_{\bar{\xi} j} + \Omega_{\bar{\xi} j}' \right),$$

$$\Omega_{\xi 1} = E(\xi_t \xi_{t-1}') = k \Omega_{\bar{\xi} k} + \sum_{j=1}^{k-1} (k-j) \Omega_{\bar{\xi}, k-j},$$

and where the components determining these matrices are defined in Table 2.

Remark 14. The form of the exact discrete model in the mixed sample case is more complicated than in the case of common data sampling owing to the fact that more operations are required in order to eliminate unobservable components from the system and replace them with observable variables. The key component in this process is the integration of the system (2) over the interval $(t - h, t]$ and then solving for the integral of y_1 in terms of its first difference, which is where Assumption 1 is utilised.

Remark 15. The $(kn_1 + n_2) \times 1$ disturbance vector ξ_t is an MA(1) process as is common in discrete time representations of first-order stochastic differential equations containing flow variables. The autocovariance matrices, $\Omega_{\xi 0}$ and $\Omega_{\xi 1}$, are more complicated than in the case of pure flow variables reported in Theorem 2 owing to the mixed sampling characteristics of

y_1 (stocks) and y_2 (flows) in addition to the mixed observation frequencies.

Remark 16. Although Theorem 3 has focused on the case of high frequency stock variables and low frequency flow variables it would be possible, in principle, to derive a discrete time representation for the case in which the high frequency variables are flows and the low frequency variables are stocks. Should such a scenario arise then a discrete time model could be derived based on the methods utilised in the proof of Theorem 3.

Attention is now turned in the following sections to issues of estimation (including computational considerations) as well as simulation evidence and an empirical example.

5. Model estimation, computation, and simulation evidence

5.1. Estimation

A natural approach to the estimation of the parameters of the continuous time system (2) is to maximise the likelihood function based on the exact discrete time representations presented in Theorems 1–3. Assuming that the random measure disturbance vector $\zeta(dt)$ in (2) is Gaussian (which would be equivalent to it being the increment of a Brownian motion process with covariance matrix Σdt) results in the discrete time disturbances also being Gaussian. Consider, first, the case of stock variables, in which the relevant vector of disturbances is η_t defined in Theorem 1; this vector has dimension $n^* \times 1$, where $n^* = n_1 k + n_2$, and is known to be vector white noise with covariance matrix Ω_η. Let θ denote the vector of unknown parameters in the model, which will consist of the elements of the vectors μ and γ, the matrix A and the unique elements of the covariance matrix Σ.[10] In this case the Gaussian log-likelihood, based on a sample of size T, is given by

$$\ln L_\eta(\theta) = -\frac{n^* T}{2} \ln 2\pi - \frac{T}{2} \ln |\Omega_\eta| - \frac{1}{2} \sum_{t=1}^{T} \eta_t' \Omega_\eta^{-1} \eta_t, \tag{23}$$

where $|\cdot|$ denotes the determinant of a matrix. In cases where there are flow variables present the log-likelihood function is of a more complicated form, reflecting the MA(1) nature of the disturbances. In the case of a mixed sample, the relevant vector of disturbances is ξ_t, defined in Theorem 3. Let $\xi = (\xi_1', \ldots, \xi_T')'$ denote the $n^* T \times 1$ vector of all sample disturbances stacked vertically. The covariance matrix of ξ, denoted $\Omega_\xi = E(\xi \xi')$, is of block Toeplitz form, with typical $n^* \times n^* T$ band given by

$$[0, \ldots, 0, \Omega_{\xi 1}, \Omega_{\xi 0}, \Omega_{\xi 1}', 0, \ldots, 0],$$

where each 0 above is an $n^* \times n^*$ matrix of zeros. The log-likelihood function is then

$$\ln L_\xi(\theta) = -\frac{n^* T}{2} \ln 2\pi - \frac{1}{2} \ln |\Omega_\xi| - \frac{1}{2} \xi' \Omega_\xi^{-1} \xi. \tag{24}$$

Clearly this function poses more computational challenges than does $\ln L_\eta$, due to the re-

---

[10]In the simulations reported below and in the empirical application in the next section the elements of the Cholesky factorisation of Σ, denoted M, rather than of Σ itself, are estimated in order to ensure that the covariance matrix is positive definite. Both M and Σ have the same number of unknown elements, $n(n+1)/2$.

quired calculation of the determinant and inverse of the $n^*T \times n^*T$ covariance matrix Ω_ξ. The same would be true in the case of flow variables where ξ and Ω_ξ would be replaced by u and Ω_u, respectively, using the relevant information in Theorem 2. However, the computational difficulties associated with log-likelihood functions of the form in (24) arising from continuous time systems were addressed in many of the articles in Bergstrom (1990), whose proposed procedure is to implement a Cholesky factorisation of the entire matrix Ω_ξ which is also able to exploit its sparse nature.

Alternative approaches could also be used to compute the likelihood function. Robinson (1993) and Chambers (1998), for example, employ a frequency domain (Whittle) approximation to the likelihood and work directly from the continuous time system (2). The method of residues is used to evaluate the infinite summations that arise in moving from continuous time to discrete time in the frequency domain. Such methods could, in principle, be extended to the case of mixed frequency data, although such an investigation is beyond the scope of the present paper. Yet another alternative, and one that has already been mentioned earlier, is to employ the Kalman filter, as in Zadrozny (1988). However, it has been argued by Bergstrom (1990, pp.112–113) that the time domain approach outlined above, based on Cholesky factorisation, is less costly in terms of computational burden than the Kalman filter.

The properties of the Gaussian estimator, $\hat{\theta}$, obtained by maximising either $\ln L_\eta(\theta)$ or $\ln L_\xi(\theta)$ with respect to θ would depend on the time series properties of the variables. In stationary systems, subject to regularity conditions, the Gaussian estimator would typically have an asymptotically normal distribution and would converge to the true value (θ_0) at a rate equal to the square root of the sample size i.e. $\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V)$; see, for example, Bergstrom (1983). In cointegrated models, on the other hand, different rates of convergence are likely to apply to the parameters governing the short-run dynamics and the long run cointegrating vectors – the former typically converge to limiting normal distributions at the rate $\sqrt{T}$, while the latter converge at the rate T to a limiting mixed normal distribution. Asymptotic results relating to the estimation of continuous time models of cointegration can be found in Phillips (1991), Chambers and McCrorie (2007) and Chambers (2011).

5.2. Computation

Inspection of the discrete time representations in Theorems 1–3 reveals that the autoregressive matrices are related to the matrix exponential $F = e^{Ah}$ while the deterministic terms and autocovariance matrices depend on various integrals involving the function $F(r) = e^{Ar}$. A number of methods exist for the computation of matrix exponentials, the most straightforward involving truncation of the infinite series in (4) once a point is reached at which the change in successive values for each element is sufficiently small. Although this may not be the most efficient method a study of comparative techniques by Jewitt and McCrorie (2005) in the context of continuous time systems concluded that truncation is valid for the types of (typically) well conditioned problems to be found in models in economics, and is therefore the method adopted here.

The issues involved in computing the deterministic terms and autocovariance matrices is perhaps best illustrated with a concrete example, and for this purpose we focus on the discrete time representation in Theorem 1. The deterministic terms b_{1t} and b_{2t} cn be seen

to depend on $c(t)$ defined in (6). It is possible to show (see the Appendix) that

$$c(t) = (C_1\mu - C_2\gamma) + C_1\gamma t,$$

where

$$C_1 = \int_0^h e^{Ar}dr, \quad C_2 = \int_0^h re^{Ar}dr.$$

If A is nonsingular these matrices have exact representations in the form

$$C_1 = A^{-1}\left(e^{Ah} - I\right), \quad C_2 = hA^{-1}e^{Ah} - A^{-2}\left(e^{Ah} - I\right),$$

while if A is nonsingular the following representations follow from integration of the infinite series in (4) term by term:

$$C_1 = h\sum_{j=0}^{\infty}\frac{1}{(j+1)!}A^jh^j, \quad C_2 = h^2\sum_{j=0}^{\infty}\frac{1}{j!(j+2)}A^jh^j.$$

In this latter case the method of truncation can be used, as in the computation of e^A itself. The covariance matrix Ω_ϵ in Theorem 1 is of the form

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

Jewitt and McCrorie (2005), using results of van Loan (1978), show that this matrix (and e^{Ah} as well) can be obtained with the computation of a single matrix exponential. Let

$$Q = \begin{bmatrix} -A & \Sigma \\ 0 & A' \end{bmatrix}, \quad P = e^{Qh} = \begin{bmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{bmatrix}.$$

Then $\Omega_\epsilon = P_{22}'P_{12}$ and $e^{Ah} = P_{22}'$. The matrix exponential e^{Qh} can also be computed by truncating its infinite series representation. Similar techniques can be applied to the deterministic terms and covariance matrices arising in the more complicated representations in Theorems 2 and 3.

5.3. Simulation evidence

In order to assess the effects of explicitly incorporating mixed frequency data as opposed to aggregating to the lowest frequency, some simulations were carried out using a model containing two stock variables sampled at two different frequencies. Let $y(t) = [y_1(t), y_2(t)]'$, where $y_1(t)$ is sampled at the higher frequency with sampling interval $h = 1/3$ and $y_2(t)$ is the low frequency variable whose sampling frequency is normalised to unity. This scenario corresponds to y_2 being observed quarterly and y_1 being observed monthly. The model is given by

$$dy(t) = Ay(t)dt + \zeta(dt), \quad E[\zeta(dt)\zeta(dt)'] = \Sigma dt, \tag{25}$$

where A is a 2×2 matrix with coefficients a_{ij} $(i, j = 1, 2)$ and Σ is a 2×2 symmetric positive definite matrix with elements σ_{11}, σ_{21} and σ_{22}. This model nests a cointegrated system

where $A = \alpha\beta'$ with $\alpha = (\alpha_1, \alpha_2)'$ and $\beta = (\beta_c, 1)'$; in this case α is the vector of adjustment coefficients and β is the cointegrating vector, normalised on y_2. Three different parametric designs are considered for the matrix A, as follows:

$$\text{Design 1:} \quad A = \begin{bmatrix} -1 & 0.5 \\ 0.5 & -1 \end{bmatrix}; \quad \text{Design 2:} \quad A = \begin{bmatrix} -1 & -0.5 \\ -0.5 & -1 \end{bmatrix};$$

$$\text{Design 3:} \quad A = \alpha\beta', \quad \alpha = \begin{bmatrix} -1 \\ -2 \end{bmatrix}, \quad \beta = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Design 1 allows for positive feedback between y_1 and y_2, while in Design 2 this feedback is negative; the eigenvalues of A in both cases are -0.5 and -1.5 so the system is stationary.[11] In Design 3 the stationary cointegrating relationship is given by $y_2(t) - y_1(t)$; the eigenvalues of A are given by 0 and -1, the former indicating the zero root (corresponding to a unit root in discrete time) in the system. In all three experimental designs the covariance matrix was taken to be of the form

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{21} \\ \sigma_{21} & \sigma_{22} \end{bmatrix},$$

where $\sigma_{11} = \sigma_{22} = 1$ and $\sigma_{21} = \{0, 0.5\}$.[12] In order to ensure that Σ is positive definite during estimation (in particular, during optimisation of the likelihood function), estimates of the lower triangular matrix M, such that $\Sigma = MM'$, were computed; the elements of M are related to the elements of Σ as follows:

$$\sigma_{11} = m_{11}^2, \quad \sigma_{21} = m_{11}m_{21}, \quad \sigma_{22} = m_{21}^2 + m_{22}^2.$$

For all these designs the data are generated according to the exact discrete time representation in Theorem 1.

The results of 10,000 simulations are contained in Table 3 for Designs 1 and 2 and in Table 4 for Design 3. The tables report the bias and root mean square error (RMSE) of three estimators of the model parameters. The column headed 'Low' contains the results when the data are aggregated to the lowest frequency, which would be the current default method for estimation. The estimated system in this case is given by (5) with $h = 1$. The next column, headed 'High', refers to the infeasible estimator obtained under the assumption that both variables can be observed at the highest frequency; the estimated system is again given by (5) but with $h = 1/3$.[13] The third column, headed 'Mixed', is the estimator using both frequencies of data i.e. the entire set of observations available to the researcher. In this case the model being estimated is given in Theorem 1 in which $k = h^{-1} = 3$; the vector of disturbances, η_t, is of dimension 4×1. In all the experimental designs the data were generated at the highest frequency according to (5) with $N = 300$; the $T = Nh = 100$ low frequency observations were then obtained from the high frequency data by selecting every

---

[11] The eigenvalues of A are required to have negative real parts in order for the system to be stationary.

[12] In view of $\sigma_{11} = \sigma_{22} = 1$ the parameter σ_{21} is, in effect, measuring the correlation between ζ_1 and ζ_2 in the continuous time system.

[13] Although y_1 is observed at the highest frequency, y_2 is not; this estimator is therefore infeasible because it is using observations that are not available to the researcher.

third observation. This is consistent with there being 25 years of quarterly data for y_2 and the same span of monthly data for y_1.

As can be seen from Table 3, the estimates obtained in the infeasible case (using high frequency observations on both variables) tend to produce the smallest bias and RMSE, the latter being approximately half the RMSE values for estimates based on the aggregated low frequency data. Estimates obtained using the mixed frequency data, however, show much lower bias and RMSE than the low frequency estimates even if they don't quite match the performance of the infeasible estimator (which is to be expected). A similar picture emerges from Table 4 in terms of the adjustment and covariance parameters, with all three methods producing very small bias and RMSE in estimating the cointegrating parameter β_c; this is presumably due to the superconsistency (faster convergence rate) of estimates of this parameter. The effect of the continuous time correlation parameter changing from $\sigma_{21} = 0$ to $\sigma_{21} = 0.5$ in Table 3 is to increase bias and RMSE in Design 1, particularly for the low frequency estimates, while the bias and RMSE remain broadly unaffected in Design 2. For Design 3 in Table 4 there is a tendency for both bias and RMSE to increase, though not uniformly.

In summary, the simulations suggest that neglecting high frequency data (when available) comes at a cost in terms of larger estimation bias and RMSE. Although the derivation of the exact discrete model is slightly more complicated when dealing with mixed frequencies than is the case when data are aggregated to the lowest frequency, the benefits of doing so would appear to be worthwhile. An empirical application is explored in the next section to ascertain the impact of using mixed frequency data in practice.

6. An empirical example: purchasing power parity

One of the most widely researched areas in the international macroeconomics/finance arena concerns the stationarity (or otherwise) of the real exchange rate between two currencies. The notion of purchasing power parity (PPP) – ignoring various nuances and subtleties – essentially suggests that, at least in the long run, the nominal exchange rate adjusts so that goods and services cost the same amount when prices in different currencies are expressed in a common currency; this, in turn, has implications for the real exchange rate. To focus ideas, let P denote the domestic price level, P^* denote the foreign price level, and S denote the nominal exchange rate expressed as units of domestic currency per unit of foreign currency. Then PPP implies that $S = P/P^*$ or $P = SP^*$, at least in the long run. Defining $s = \ln S$, $p = \ln P$ and $p^* = \ln P^*$, another way of writing this relationship is $s = p - p^*$, a form that readily suggests testing PPP by estimating a simple regression of s on p and p^* and then testing whether the coefficients on the two regressors are $+1$ and -1, respectively. However, given that exchange rates and price indices are often found to be characterised as containing unit roots (and possibly two in the case of price indices), cointegration techniques can be used to test the PPP restrictions on the cointegrating vector (if one is found), but the evidence is somewhat mixed. A comprehensive account of empirical research into PPP and the real exchange rate can be found in Sarno and Taylor (2002).

The price indices used in PPP research are typically observed on a monthly basis while

exchange rates can be observed at much higher frequencies.[14] The usual approach is to aggregate the exchange rate data to the monthly frequency by taking either a monthly average of daily closing prices or the value at a particular point in the month (such as a daily price in the middle or at the end of the month). This approach throws away a great deal of potentially useful information in the exchange rate data that may be pertinent to assessing the empirical validity of PPP. In what follows, the methods derived in this paper are utilised in an assessment of the implications for PPP of combining the high frequency exchange rate data with the lower frequency price index data.

In accordance with the notation used in previous sections, let $y_1(t) = s(t)$ denote the high frequency exchange rate variable, let $y_2(t) = [p(t), p^*(t)]'$ denote the vector of low frequency price index variables, let $y(t) = [y_1(t), y_2(t)']'$, and note that t is being treated as a continuous time parameter. The most general continuous time system under consideration has the representation

$$dy(t) = [\mu + \alpha\beta'y(t)]dt + \zeta(dt), \quad t > 0, \tag{26}$$

where μ is a 3×1 vector of intercepts, α is a 3×1 vector of adjustment parameters, $\beta = [1, \beta_p, \beta_{p^*}]'$ is a 3×1 vector of cointegrating parameters normalised on $s(t)$, and $\zeta(dt)$ is a 3×1 vector of random measures with mean vector zero and symmetric positive definite covariance matrix Σdt of dimension 3×3. The unrestricted intercept, μ, can be decomposed using the identity $I_3 = \alpha(\alpha'\alpha)^{-1}\alpha' + \alpha_\perp(\alpha_\perp'\alpha_\perp)^{-1}\alpha_\perp'$, where $\alpha_\perp$ is the 3×2 orthogonal complement of α satisfying $\alpha_\perp'\alpha = 0$. Post-multiplying this identity by μ yields $\mu = \alpha\kappa + \delta$ where the scalar $\kappa = (\alpha'\alpha)^{-1}\alpha'\mu$ and $\delta = \alpha_\perp(\alpha_\perp'\alpha_\perp)^{-1}\alpha_\perp'\mu$ is a 3×1 vector. This implies that

$$dy(t) = [\delta + \alpha(\kappa + \beta'y(t))]dt + \zeta(dt), \tag{27}$$

so that κ represents an intercept in the cointegrating relationship.[15] In terms of the system (2), $\gamma = 0$, $\mu = \delta + \alpha\kappa$ and $A = \alpha\beta'$; the system (27) therefore represents a continuous time cointegrated system with cointegrating relationship of the form $s(t) + \kappa + \beta_p p(t) + \beta_{p^*}p^*(t)$. The long-run PPP hypothesis consists of the two restrictions $\beta_p = -\beta_{p^*} = -1$, but we shall also consider a weaker version in which there is a single restriction of the form $\beta_p = -\beta_{p^*}$ that does not constrain these parameters to a particular value.

The data used in the empirical application are the monthly consumer price indices for the United States (P) and the United Kingdom (P^*) for the period January 1996 to March 2014 and the daily (closing) exchange rate (S), measured in US dollars per pound, from January 1, 1996 to March 28, 2014. The exchange rate data were also aggregated to weekly and monthly frequencies, there being four weekly observations and twenty daily observations corresponding to each month; further details concerning the data are provided in the Appendix.

Table 5 contains summary results and test statistics relating to estimation of the unre-

---

[14]Although we focus on the temporal aggregation aspects of price indices a referee has highlighted that there are also cross-sectional issues involved but which are beyond the scope of this example.

[15]The specification of the deterministic component in (27) corresponds to model $H_1(1)$ in the terminology of Johansen (1995).

stricted model (26) and various nested models using the monthly price indices and monthly, weekly and daily exchange rate data. The Table contains the maximised log-likelihood for the unrestricted model as well as the Schwarz Information Criterion (SIC) for the unrestricted and restricted models; for the latter, the values of the likelihood ratio (LR) test statistics (and their associated p-values) are also reported.[16] For the monthly exchange rate data none of the restrictions is rejected at the 5% significance level and so the preferred model is one in which the PPP restriction is imposed. This restricted model is also chosen by the SIC, and a similar picture emerges with the weekly exchange rate data. However, moving to the daily exchange rate data, entirely different conclusions are drawn. All of the restrictions tested are comprehensively rejected, the preferred model, chosen on the basis of the likelihood ratio tests and the SIC, being the unrestricted model.

To gain further insight into these results, Table 6 reports full estimates of the unrestricted models for all three exchange rate frequencies, as well as the restricted (preferred) models for the monthly and weekly exchange rate data (the preferred model for the weekly data being the unrestricted model). In addition to α, β and μ Table 6 also contains estimates of the elements of the Cholesky matrix corresponding to Σ; this is a lower-triangular matrix M such that $\Sigma = MM'$ and is given by

$$M = \begin{bmatrix} m_{11} & 0 & 0 \\ m_{21} & m_{22} & 0 \\ m_{31} & m_{32} & m_{33} \end{bmatrix}.$$

The final row of Table 6 contains the implied values of the intercept in the cointegrating relationship, κ. As can be seen from Table 6, many of the estimated parameters in the models using monthly and weekly exchange rate data have large standard errors relative to the estimated parameter. It is, therefore, perhaps no great surprise that the restrictions are not rejected when imposed on the model, due to this level of parameter uncertainty. This is reflected by the small fall in the value of the log-likelihood reported in Table 5 when the PPP restrictions are imposed. Using daily exchange rate data, however, results in parameter estimates with much smaller standard errors, and hence there is much greater information available to reject the hypotheses being tested. For example, $\hat{\beta}_p = -3.0260$ with a standard error of 0.2601, while $\hat{\beta}_{p^*} = 3.2674$ with a standard error of 0.3223, so the rejection of the joint hypothesis $\beta_p = -\beta_{p^*} = -1$ should not be too surprising. The move from monthly and weekly exchange rate data to daily data has resulted in more precise estimates of the parameters of interest and a rather different outcome to the statistical tests carried out.

7. Conclusions

This paper has derived exact discrete time representations corresponding to a system of linear stochastic differential equations when the observed sample is available at different frequencies. The cases of common data sampling (all stock or all flow variables) and mixed data sampling (a combination of stocks and flows) have both been considered in this mixed frequency scenario. Simulations based on both stationary and cointegrated systems reveal

---

[16]In all cases the p-values are obtained assuming that the LR_q statistic has a chi-squared distribution with q degrees of freedom.

that there are substantial gains to be made in estimation (smaller bias and RMSE) by utilising all the higher frequency data in conjunction with the low frequency data instead of aggregating all data to the lowest frequency. An empirical application using monthly price indices for the UK and US and monthly, weekly and daily exchange rate data reveal that substantially different inferences can be drawn when the highest frequency exchange rate data are used compared with the lower frequency data. An advantage of considering mixed frequency data analysis based on an underlying continuous time model is that the model of interest is not tied to any particular sampling frequency and the discrete time representations hold exactly at whatever frequency the data are sampled.

There are a number of directions for further research that emerge from the results reported here. One obvious extension is to consider exact discrete time representations for mixed sample and mixed frequency data generated by an $ARMA(p, q)$ system in continuous time. This could be achieved by combining the techniques employed in the proofs of Theorems 1–3 with those used by Chambers and Thornton (2012) whose empirical examples demonstrated the benefits of the additional MA components in the continuous time model for improving model fit. In some situations it would also be beneficial to allow both high and low frequency variables to be mixtures of stocks and flows and also to allow for more than two sampling frequencies. A more realistic situation in some cases would be to allow for the sampling intervals to be of different lengths, which may be of some importance when combining monthly and daily data as the number of days in the months of the year are not equal. Another area in which continuous time models can be advantageous is in producing forecasts of variables at any future point in time as they are not tied to a particular sampling frequency. Again the techniques used in this paper form a basis for these, and other, extensions, and will be explored in future work.

Appendix.

Proofs of Theorems

Proof of Theorem 1. The aim is to eliminate the unobservable $y_2(t - h)$ from the system (8) and (9) and, effectively, replace it with $y_2(t - kh) = y_2(t - 1)$. Lagging (9) by h and repeatedly substituting backwards enables $y_2(t - h)$ to be expressed in terms of $y_2(t - 1)$ as follows:

$$y_2(t - h) = \sum_{j=1}^{k-1} F_{22}^{j-1} c_2(t - jh) + \sum_{j=1}^{k-1} F_{22}^{j-1} F_{21} y_1(t - jh - h)$$

$$+ F_{22}^{k-1} y_2(t - 1) + \sum_{j=1}^{k-1} F_{22}^{j-1} \epsilon_2(t - jh). \tag{28}$$

Substituting (28) into (9) yields the equation for y_2, (13), in which the disturbance term is

$$\eta_{2t} = \sum_{j=0}^{k-1} F_{22}^{j} \epsilon_2(t - jh).$$

More generally, lags of $y_2(t)$ can be expressed, for $1 \leq l \leq k - 1$, as

$$y_2(t - lh) = \sum_{j=0}^{k-l-1} F_{22}^{j} c_2(t - lh - jh) + \sum_{j=0}^{k-l-1} F_{22}^{j} F_{21} y_1(t - lh - jh - h)$$

$$+ F_{22}^{k-l} y_2(t - 1) + \sum_{j=0}^{k-l-1} F_{22}^{j} \epsilon_2(t - lh - jh) \tag{29}$$

so that, for $0 \leq l \leq k - 1$, lags of $y_1(t)$ can be expressed in terms of $y_2(t - 1)$ as

$$y_1(t - lh) = c_1(t - lh) + F_{11} y_1(t - lh - h) + F_{12} y_2(t - lh - h) + \epsilon_1(t - lh)$$

$$= b_1(t - lh) + \sum_{j=1}^{k-l} B_{11,j} y_1(t - lh - jh) + B_{12,l} y_2(t - 1) + \eta_{1,t-lh},$$

as required, where (29) has been substituted for $y_2(t - lh - h)$ and

$$\eta_{1,t-lh} = \epsilon_1(t - lh) + F_{12} \sum_{j=1}^{k-l-1} F_{22}^{j-1} \epsilon_2(t - lh - jh);$$

if $l = k - 1$ in this last summation then it is to be taken as zero. From these representations it is possible to write η_t in the form

$$\eta_t = \sum_{j=0}^{k-1} F_{\eta,j} \epsilon(t - jh), \tag{30}$$

where the $F_{\eta,j}$ $(j = 0, \ldots, k-1)$ are defined by

$$F_{\eta,0} = \begin{bmatrix} I & 0 \\ 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \\ 0 & I \end{bmatrix}, \quad F_{\eta,1} = \begin{bmatrix} 0 & F_{12} \\ I & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \\ 0 & F_{22} \end{bmatrix}, \quad F_{\eta,2} = \begin{bmatrix} 0 & F_{12}F_{22} \\ 0 & F_{12} \\ I & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \\ 0 & F_{22}^2 \end{bmatrix},$$

$$F_{\eta,j} = \begin{bmatrix} 0 & F_{12}F_{22}^{j-1} \\ 0 & F_{12}F_{22}^{j-2} \\ \vdots & \vdots \\ 0 & F_{12} \\ I & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & F_{22}^{j} \end{bmatrix}, \quad j = 3, \ldots, k-2, \quad F_{\eta,k-1} = \begin{bmatrix} 0 & F_{12}F_{22}^{k-2} \\ 0 & F_{12}F_{22}^{k-3} \\ \vdots & \vdots \\ 0 & F_{12}F_{22} \\ 0 & F_{12} \\ I & 0 \\ 0 & F_{22}^{k-1} \end{bmatrix}.$$

The form of the covariance matrix Ω_η then follows in view of $\epsilon(t)$ being vector white noise with covariance matrix Ω_ϵ. $\square$

Proof of Theorem 2. Integrating (10)–(13) over the interval $(t-h, t]$ and dividing the equations for y_1 by h yields, for $0 \leq l \leq k-1$,

$$Y_{1,t-lh} = h^{-1}b_{1,t-lh}^* + \sum_{j=1}^{k-l} B_{11,j}Y_{1,t-lh-jh} + h^{-1}B_{12,l}Y_{2,t-1}^u + h^{-1}e_{1,t-lh}, \tag{31}$$

$$Y_{2t}^u = b_{2t}^* + h\sum_{j=1}^{k} B_{21,j}Y_{1,t-jh} + B_{22}Y_{2,t-1}^u + e_{2t}, \tag{32}$$

where $Y_{2t}^u = \int_{t-h}^{t} y_2(r)dr$, the b_{jt}^* are defined in the Theorem, and $e_{jt} = \int_{t-h}^{t} \eta_{jr}dr$ $(j = 1, 2)$. The objective is to eliminate terms involving the unobervable Y_{2t}^u and replace them with the observable $Y_{2t} = \int_{t-1}^{t} y_2(r)dr$. Noting that $Y_{2t} = s(L^h)Y_{2t}^u$, the application of the filter $s(L^h)$ to (31) and (32) yields the equations in the Theorem, in which $u_{1t} = h^{-1}s(L^h)e_{1t}$ and $u_{2t} = s(L^h)e_{2t}$.

Turning to u_t it is possible to write

$$u_t = s(L^h)K_h e_t = s(L^h)K_h \int_{t-h}^{t} \eta_r dr = s(L^h)K_h F_\eta(L^h) \int_{t-h}^{t} \epsilon(r)dr,$$

where $F_\eta(z) = \sum_{j=0}^{k-1} F_{\eta,j}z^j$ from the representation for η_t in (30) and

$$K_h = \begin{bmatrix} h^{-1}I_{kn_1} & 0 \\ 0 & I_{n_2} \end{bmatrix}.$$

23

The role of the symmetric $(kn_1+n_2)\times(kn_1+n_2)$ matrix K_h is to account for the normalisation by h in the equation for the high frequency flows. From (7) the integral of $\epsilon(t)$ takes the form

$$\epsilon_t = \int_{t-h}^t \epsilon(r)dr = \int_{t-h}^t \int_{r-h}^r e^{A(r-s)}\zeta(ds)dr$$

$$= \int_{t-h}^t P_0(t-s)\zeta(ds) + \int_{t-2h}^{t-h} P_1(t-h-s)\zeta(ds),$$

where $P_0(r) = \int_0^r e^{As}ds$ and $P_1(r) = \int_r^h e^{As}ds$; the splitting of the double integral to two single integrals with respect to $\zeta(dt)$ follows along the same lines as in the proof of Lemma B2 in Chambers (2011). From this representation it follows that ϵ_t is an MA(1) process with autocovariances

$$\Omega_{\epsilon 0} = E(\epsilon_t \epsilon_t') = \int_0^h P_0(r)\Sigma P_0(r)'dr + \int_0^h P_1(r)\Sigma P_1(r)'dr,$$

$$\Omega_{\epsilon 1} = E(\epsilon_t \epsilon_{t-h}') = \int_0^h P_1(r)\Sigma P_0(r)'dr.$$

Using Lemma 1 the convolution $H(z) = s(z)F_\eta(z)$ has the form $H(z) = \sum_{j=0}^{2k-2} H_j z^j$ and hence $u_t = \sum_{j=0}^{2k-2} H_j \epsilon_{t-jh}$, where

$$H_j = \begin{cases} \displaystyle\sum_{l=0}^j F_{\eta,l}, & j = 1,\ldots,k-1, \\ \displaystyle\sum_{l=j-(k-1)}^{k-1} F_{\eta,l}, & j = k,\ldots,2k-2. \end{cases}$$

The autocovariance properties of u_t follow from this representation and the MA(1) properties of ϵ_t; see Remark 11. $\square$

Proof of Theorem 3. Integrating (2) over $(t-h,t]$ and partitioning conformably with y_1 and y_2 yields

$$\Delta_h y_1(t) = a_{1t} + A_{11}Y_{1t}^u + A_{12}Y_{2t}^u + z_{1t}, \tag{33}$$
$$\Delta_h y_2(t) = a_{2t} + A_{21}Y_{1t}^u + A_{22}Y_{2t}^u + z_{2t}, \tag{34}$$

where $Y_{jt}^u = \int_{t-h}^t y_j(r)dr$ $(j=1,2)$, $a_t = (a_{1t}', a_{2t}')'$ is defined in the Theorem, and

$$z_t = \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \int_{t-h}^t \zeta(dr).$$

Integrating (8) and (9) over the same interval gives

$$Y_{1t}^u = f_{1t} + F_{11}Y_{1,t-h}^u + F_{12}Y_{2,t-h}^u + \epsilon_{1t}, \tag{35}$$
$$Y_{2t}^u = f_{2t} + F_{21}Y_{1,t-h}^u + F_{22}Y_{2,t-h}^u + \epsilon_{2t}, \tag{36}$$

where $\epsilon_t = [\epsilon_{1t}', \epsilon_{2t}']' = \int_{t-h}^t \epsilon(r)dr$ and f_t is defined in the Theorem. Beginning with the

equation for flows, solving (33) for Y_{1t}^u under Assumption 1 results in

$$Y_{1t}^u = A_{11}^{-1}\left[\Delta_h y_1(t) - a_{1t} - A_{12} Y_{2t}^u - z_{1t}\right]. \tag{37}$$

Lagging (37) by h and substituting into (36) for $Y_{1,t-h}^u$ yields

$$
\begin{aligned}
Y_{2t}^u &= f_{2t} + F_{21} A_{11}^{-1}\left[\Delta_h y_1(t-h) - a_{1,t-h} - A_{12} Y_{2,t-h}^u - z_{1,t-h}\right] + F_{22} Y_{2,t-h}^u + \epsilon_{2t} \\
&= g_{2t} + G_{21}\Delta_h y_1(t-h) + G_{22} Y_{2,t-h}^u + v_{2t},
\end{aligned} \tag{38}
$$

where g_{2t}, G_{21} and G_{22} are defined in the Theorem, and $v_{2t} = \epsilon_{2t} - F_{21}A_{11}^{-1} z_{1,t-h}$. It follows, by backward substitution, that

$$Y_{2,t-h}^u = \sum_{j=1}^{k-1} G_{22}^{j-1} g_{2,t-jh} + \sum_{j=2}^{k} G_{22}^{j-2} G_{21}\Delta_h y_1(t-jh) + G_{22}^{k-1} Y_{2,t-kh}^u + \sum_{j=1}^{k-1} G_{22}^{j-1} v_{2,t-jh}. \tag{39}$$

Substituting (39) into (38) yields

$$Y_{2t}^u = \bar{g}_{2t} + \sum_{j=1}^{k} \Phi_{21,j}\Delta_h y_1(t-jh) + \Phi_{22} Y_{2,t-1}^u + \bar{\xi}_{2t}, \tag{40}$$

where $\bar{g}_{2t} = \sum_{j=0}^{k-1} G_{22}^j g_{2,t-jh}$, $\Phi_{21,j}$ $(j=1,\ldots,k)$ and Φ_{22} are defined in the Theorem, and $\bar{\xi}_{2t} = \sum_{j=0}^{k-1} G_{22}^j v_{2,t-jh}$. In order to transform Y_{2t}^u into Y_{2t} the operator $s(L^h)$ is applied, resulting in (22), where $\xi_{2t} = s(L^h)\bar{\xi}_{2t}$ etc.

Turning to the equation for stocks, substituting (37) into (34) yields

$$
\begin{aligned}
\Delta_h y_2(t) &= a_{2t} + A_{21}A_{11}^{-1}\left[\Delta_h y_1(t) - a_{1t} - A_{12} Y_{2t}^u - z_{1t}\right] + A_{22} Y_{2t}^u + z_{2t} \\
&= \left(a_{2t} - A_{21}A_{11}^{-1} a_{1t}\right) + A_{21}A_{11}^{-1}\Delta_h y_1(t) + \left(A_{22} - A_{21}A_{11}^{-1} A_{12}\right) Y_{2t}^u \\
&\quad + z_{2t} - A_{21}A_{11}^{-1} z_{1t}.
\end{aligned} \tag{41}
$$

Applying the h-difference operator $\Delta_h = 1 - L^h$ to (8) results in (for $j = 0, \ldots, k-1$)

$$\Delta_h y_1(t-jh) = \Delta_h c_1(t-jh) + F_{11}\Delta_h y_1(t-jh-h) + F_{12}\Delta_h y_2(t-jh-h) + \Delta_h \epsilon_1(t-jh). \tag{42}$$

Now, lagging (41) by h and substituting for $\Delta_h y_2(t-jh-h)$ gives

$$\Delta_h y_1(t-jh) = g_{1,t-jh} + G_{11}\Delta_h y_1(t-jh-h) + G_{12} Y_{2,t-jh-h}^u + v_{1,t-jh}, \tag{43}$$

where $v_{1t} = \Delta_h \epsilon_1(t) + F_{12}\left(z_{2,t-h} - A_{21}A_{11}^{-1} z_{1,t-h}\right)$ and g_{1t}, G_{11} and G_{12} are defined in the Theorem. Generalising (39) yields

$$
\begin{aligned}
Y_{2,t-lh}^u &= \sum_{p=0}^{k-l-1} G_{22}^p g_{2,t-lh-ph} + \sum_{p=1}^{k-l} G_{22}^{p-1} G_{21}\Delta_h y_1(t-lh-ph) \\
&\quad + G_{22}^{k-l} Y_{2,t-kh}^u + \sum_{p=0}^{k-l-1} G_{22}^p v_{2,t-lh-ph}.
\end{aligned} \tag{44}
$$

Using (44) to substitute for $Y_{2,t-jh-h}^u$ in (43) provides

$$
\begin{aligned}
\Delta_h y_1(t-jh) &= g_{1,t-jh} + G_{11}\Delta_h y_1(t-jh-h) + G_{12}\left(\sum_{p=0}^{k-j-2} G_{22}^p g_{2,t-jh-h-ph}\right. \\
&\quad + \sum_{p=1}^{k-j-1} G_{22}^{p-1} G_{21}\Delta_h y_1(t-jh-h-ph) + G_{22}^{k-j-1} Y_{2,t-1}^u \\
&\quad \left. + \sum_{p=0}^{k-j-2} G_{22}^p v_{2,t-jh-h-ph}\right) + v_{1,t-jh} \quad (45) \\
&= \bar{g}_{1,t-jh} + \sum_{p=1}^{k-j}\Phi_{11,p}\Delta_h y_1(t-jh-ph) + \Phi_{12,j}Y_{2,t-1}^u + \bar{\xi}_{1,t-jh}, \quad (46)
\end{aligned}
$$

where $\bar{g}_{1,t-jh} = g_{1,t-jh} + G_{12}\sum_{p=1}^{k-j-1} G_{22}^{p-1} g_{2,t-jh-ph}$, the $\Phi_{11,j}$ $(j=2,\ldots,k)$ and $\Phi_{12,j}$ are defined in the Theorem, and $\bar{\xi}_{1,t-jh} = v_{1,t-jh} + G_{12}\sum_{p=1}^{kj-1} G_{22}^{p-1} v_{2,t-jh-ph}$. The final step is to apply the operator $s(L^h)$ to transform $Y_{2,t-1}^u$ into $Y_{2,t-1}$, resulting in (19) – (21) as required, where $\xi_{1,t-jh} = s(L^h)\bar{\xi}_{1,t-jh}$ etc.

The autocovariance structure of ξ_t can be obtained by noting that $\xi_t = s(L^h)\bar{\xi}_t$, and that, in turn, $\bar{\xi}_t = \sum_{j=0}^{k-1} R_j v_{t-jh}$, where $v_t = [v_{1t}', v_{2t}']'$ has the representation

$$
v_t = \int_{t-h}^t K_0(t-r)\zeta(dr) + \int_{t-2h}^{t-h} K_1(t-h-r)\zeta(dr),
$$

and the R_j $(j=0,\ldots,k-1)$ are defined by

$$
R_0 = \begin{bmatrix} I & 0 \\ 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \\ 0 & I \end{bmatrix}, \quad
R_1 = \begin{bmatrix} 0 & G_{12} \\ I & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \\ 0 & G_{22} \end{bmatrix}, \quad
R_2 = \begin{bmatrix} 0 & G_{12}G_{22} \\ 0 & G_{12} \\ I & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \\ 0 & G_{22}^2 \end{bmatrix},
$$

$$
R_j = \begin{bmatrix} 0 & G_{12}G_{22}^{j-1} \\ 0 & G_{12}G_{22}^{j-2} \\ \vdots & \vdots \\ 0 & G_{12} \\ I & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & G_{22}^j \end{bmatrix}, \quad j=3,\ldots,k-2, \quad
R_{k-1} = \begin{bmatrix} 0 & G_{12}G_{22}^{k-2} \\ 0 & G_{12}G_{22}^{k-3} \\ \vdots & \vdots \\ 0 & G_{12}G_{22} \\ 0 & G_{12} \\ I & 0 \\ 0 & G_{22}^{k-1} \end{bmatrix}.
$$

The precise formulae come about by noting that

$$
\begin{aligned}
v_{1t} &= \Delta_h \epsilon_1(t) + F_{12}\left(z_{2,t-h} - A_{21}A_{11}^{-1}z_{1,t-h}\right) \\
&= \int_{t-h}^{t} \left[F_{11}(t-r) : F_{12}(t-r)\right]\zeta(dr) - \int_{t-2h}^{t-h} \left[F_{11}(t-h-r) : F_{12}(t-h-r)\right]\zeta(dr) \\
&\quad + \left[-F_{12}A_{21}A_{11}^{-1} : F_{12}\right]\int_{t-h}^{t}\zeta(dr), \\
v_{2t} &= \epsilon_{2t} - F_{21}A_{11}^{-1}z_{1,t-h} \\
&= \int_{t-h}^{t}\left[P_{0,21}(t-r) : P_{0,22}(t-r)\right]\zeta(dr) \\
&\quad + \int_{t-2h}^{t-h}\left[P_{1,21}(t-h-r) : P_{1,22}(t-h-r)\right]\zeta(dr) + \left[-F_{21}A_{11}^{-1} : 0\right]\int_{t-h}^{t}\zeta(dr).
\end{aligned}
$$

The random vector v_t is thus an MA(1) process with variance matrix Ω_{v0} and autocovariance matrix Ω_{v1} as defined in Table 2. This enables the autocovariances of $\bar{\xi}_t$ and, hence, of ξ_t itself to be derived. $\square$

Supplementary results

Lemma 1. *Let $s(z) = \sum_{j=0}^{k-1} z^j$ and $F(z) = \sum_{j=0}^{k-1} F_j z^j$, where $s(z)$ is a scalar polynomial and $F_0, \ldots, F_k$ are matrices. Then the convolution $s(z)F(z)$ is of the form*

$$
H(z) = s(z)F(z) = \sum_{j=0}^{k-1}\sum_{l=0}^{k-1} F_l z^{j+l} = \sum_{j=0}^{2k-2} H_j z^j,
$$

where

$$
H_j = \sum_{l=\max\{0,j-(k-1)\}}^{\min\{j,k-1\}} F_l =
\begin{cases}
\sum_{l=0}^{j} F_l, & j = 0, \ldots, k-1, \\
\sum_{l=j-(k-1)}^{k-1} F_l, & j = k, \ldots, 2k-2.
\end{cases}
$$

Proof of Lemma 1. Collecting terms in common powers in the polynomial yields

$$
\begin{aligned}
H(z) &= F_0 + (F_0 + F_1)z + (F_0 + F_1 + F_2)z^2 + \ldots + (F_0 + F_1 + \ldots + F_{k-1})z^{k-1} \\
&\quad + (F_1 + \ldots + F_{k-1})z^k + (F_2 + \ldots + F_{k-1})z^{k+1} + \ldots + F_{k-1}z^{2k-2}.
\end{aligned}
$$

The matrices in this expression can be represented as stated above. $\square$

Derivation of expressions for $c(t)$ in section 5.2. From (6) we have

$$
c(t) = \int_{t-h}^{t} e^{A(t-s)}(\mu + \gamma s)ds = C_1\mu + C_1^*\gamma,
$$

where

$$
C_1 = \int_{t-h}^{t} e^{A(t-s)}ds = \int_0^h e^{Ar}dr,
$$

27

$$C_1^* = \int_{t-h}^{t} s e^{A(t-s)} ds = \int_0^h (t-r)e^{Ar} dr = C_1 t - C_2, \quad C_2 = \int_0^h r e^{Ar} dr.$$

Hence $c(t) = (C_1\mu - C_2\gamma) + C_1\gamma t$. When A is nonsingular the expressions for C_1 and C_2 follow from matrix generalisations of standard integral formulae. For nonsingular A the stated expressions follow from using the infinite series representation of e^{Ar} and integrating term by term, yielding

$$C_1 = \int_0^h \sum_{j=0}^{\infty} \frac{1}{j!} A^j r^j dr = \sum_{j=0}^{\infty} \frac{1}{j!} A^j \int_0^h r^j dr = \sum_{j=0}^{\infty} \frac{1}{j!} A^j \frac{h^{j+1}}{(j+1)} = h \sum_{j=0}^{\infty} \frac{1}{(j+1)!} A^j h^j,$$

$$C_2 = \int_0^h \sum_{j=0}^{\infty} \frac{1}{j!} A^j r^{j+1} dr = \sum_{j=0}^{\infty} \frac{1}{j!} A^j \int_0^h r^{j+1} dr = \sum_{j=0}^{\infty} \frac{1}{j!} A^j \frac{h^{j+2}}{(j+2)} = h^2 \sum_{j=0}^{\infty} \frac{1}{j!(j+2)} A^j h^j,$$

as required.

Data used in section 6. The monthly UK CPI data were obtained from the Office for National Statistics while the monthly US CPI data and daily exchange rate data were obtained from the FRED database provided by the Federal Reserve Bank of St. Louis. As mentioned in the text, the daily exchange rate observations were also aggregated to weekly and monthly frequencies. For the monthly series the last available daily observation was used. For the weekly series, the daily dates chosen depended on the number of days in the month. For months with 31 days, the four weekly values correspond to days 8, 16, 24 and 31; for months with 30 days, the observations correspond to days 8, 15, 23 and 30; while for February, days 7, 14, 21 and 28/29 were used. This ensures that the weekly data are in accordance with the monthly data i.e. the observation for the fourth week in the month corresponds to the last day of the month, as for the monthly data. In cases where the required day corresponds to a weekend or holiday, the value immediately prior to the required day was used.

References

Agbeyegbe, T.D., 1987. An exact discrete analog to a closed linear first-order continuous-time system with mixed sample. *Econometric Theory* 3, 143–149.

Agbeyegbe, T.D., 1988. An exact discrete analog of an open linear non-stationary first-order continuous-time system with mixed sample. *Journal of Econometrics* 39, 237–250.

Anderson, B.D.O., Deistler, M., Felsenstein, E., Funovits, B., Koelbl, L., Zamani, M., 2015. Multivariate AR systems and mixed frequency data: G-identifiability and estimation. *Econometric Theory* (forthcoming).

Andreou, E., Ghysels, E., Kourtellos, A., 2010. Regression models with mixed sampling frequencies. *Journal of Econometrics* 158, 246–261.

Bergstrom, A.R., 1983. Gaussian estimation of structural parameters in higher order continuous time dynamic models. *Econometrica* 51, 1117–1152.

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.

Chambers, M.J, 1998. The estimation of systems of joint differential-difference equations. *Journal of Econometrics* 85, 1–31.

Chambers, M.J, 2009. Discrete time representations of cointegrated continuous time models with mixed sample data. *Econometric Theory* 25, 1030–1049.

Chambers, M.J, 2011. Cointegration and sampling frequency. *Econometrics Journal* 14, 156–185.

Chambers, M.J, McCrorie, J.R., 2007. Frequency domain estimation of temporally aggregated Gaussian cointegrated systems. *Journal of Econometrics* 136, 1–29.

Chambers, M.J., Thornton, M.A., 2012. Discrete time representation of continuous time ARMA processes. *Econometric Theory* 28, 219–238.

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

Foroni, C., Marcellino, M., 2013. A survey of econometric methods for mixed frequency data. Norges Bank Working Paper 2013/06.

Foroni, C., Marcellino, M., 2014a. Mixed frequency structural VARs. Norges Bank Working Paper 2014/01.

Foroni, C., Marcellino, M., 2014b. Mixed frequency structural models: identification, estimation, and policy analysis. *Journal of Applied Econometrics* 29, 1118–1144.

Foroni, C., Marcellino, M., Schumacher, C., 2015. U-MIDAS: MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society, Series A* 178, 57–82.

Ghysels, E., 2014. Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics* (forthcoming).

Ghysels, E., Miller, J.I., 2014a. Testing for cointegration with temporally aggregated and mixed-frequency time series. *Journal of Time Series Analysis* (forthcoming).

Ghysels, E., Miller, J.I., 2014b. On the size distortion from linearly interpolating low-frequency series for cointegration tests. In: Fomby, T.B., Park, J.Y. and Chang, Y. (Eds.), *Advances in Econometrics: Esays in Honor of Peter C.B. Phillips*, vol. 33.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2002. The MIDAS touch: mixed data sampling regression models. Unpublished manuscript.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics* 131, 59–95.

Ghysels, E., Sinko, A., Valkanov, R., 2007. MIDAS regressions: further results and new directions. *Econometric Reviews* 26, 53–90.

Jewitt, G., McCrorie, J.R., 2005. Computing estimates of continuous time macroeconometric models on the basis of discrete data. *Computational Statistics and Data Analysis* 49, 397–416.

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

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

Phillips, P.C.B., 1991. Error correction and long run equilibrium in continuous time. *Econometrica* 59, 967–980.

Robinson, P.M., 1993. Continuous time models in econometrics: closed and open systems, stocks and flows. In: Phillips, P.C.B. (Ed.), *Models, Methods and Applications of Econometrics: Essays in Honor of A.R. Bergstrom.* Basil Blackwell, Oxford.

Sarno, L., Taylor, M.P., 2002. *The Economics of Exchange Rates.* Cambridge University Press, Cambridge.

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.

Simos, T., 1996. Gaussian estimation of a continuous time dynamic model with common stochastic trends. *Econometric Theory* 12, 361–373.

van Loan, C.F., 1978. Computing integrals involving the matrix exponential. *IEEE Transactions on Automatic Control* AC-23, 395–404.

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 1

Simulated bias and RMSE of maximum likelihood estimator of the parameter a in (1)

h	N	Bias	RMSE	h	N	Bias	RMSE
		$a = -5$, $T = 25$				$a = -5$, $T = 100$	
1	25	-0.8653	3.9986	1	100	-0.7135	3.2966
$\frac{1}{4}$	100	-0.4974	2.4389	$\frac{1}{4}$	400	-0.0873	0.7057
$\frac{1}{12}$	300	-0.1268	0.6545	$\frac{1}{12}$	1200	-0.0347	0.4029
		$a = -1$, $T = 25$				$a = -1$, $T = 100$	
1	25	-0.5697	1.9237	1	100	-0.0634	0.3334
$\frac{1}{4}$	100	-0.1000	0.3772	$\frac{1}{4}$	400	-0.0235	0.1655
$\frac{1}{12}$	300	-0.0908	0.3374	$\frac{1}{12}$	1200	-0.0217	0.1509

Table 2

Definitions of quantities in Theorem 3

Quantity	$i = 1$	$i = 2$
Autoregressive quantities		
g_{it}	$\Delta_h c_1(t) + F_{12}\left(a_{2,t-h} - A_{21}A_{11}^{-1}a_{1,t-h}\right)$	$f_{2t} - F_{21}A_{11}^{-1}a_{1,t-h}$
f_{it}	$\displaystyle\int_{t-h}^{t} c_1(r)dr$	$\displaystyle\int_{t-h}^{t} c_2(r)dr$
a_{it}	$\displaystyle\int_{t-h}^{t} (\mu_1 + \gamma_1 r)dr$	$\displaystyle\int_{t-h}^{t} (\mu_2 + \gamma_2 r)dr$
G_{1i}	$F_{11} + F_{12}A_{21}A_{11}^{-1}$	$F_{12}\left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)$
G_{2i}	$F_{21}A_{11}^{-1}$	$F_{22} - F_{21}A_{11}^{-1}A_{12}$

Autocovariance quantities

$$\Omega_{\bar{\xi}0} \qquad \sum_{j=0}^{k-1} R_j\Omega_{v0}R_j' + \sum_{j=1}^{k-1}\left(R_j\Omega_{v1}R_{j-1}' + R_{j-1}\Omega_{v1}'R_j'\right)$$

$$\Omega_{\bar{\xi}s} \qquad \sum_{j=s}^{k-1} R_j\Omega_{v0}R_{j-s}' + \sum_{j=s+1}^{k-1} R_j\Omega_{v1}'R_{j-s-1}' + \sum_{j=s-1}^{k-1} R_j\Omega_{v1}R_{j+1-s}' \quad (s = 1,\ldots,k)$$

$$\Omega_{v0} \qquad \int_0^h K_0(r)\Sigma K_0(r)'dr + \int_0^h K_1(r)\Sigma K_1(r)'dr$$

$$\Omega_{v1} \qquad \int_0^h K_1(r)\Sigma K_0(r)'dr$$

$$K_0(r) \qquad \begin{bmatrix} F_{11}(r) & F_{12}(r) \\ P_{0,21}(r) & P_{0,22}(r) \end{bmatrix}$$

$$K_1(r) \qquad \begin{bmatrix} -F_{11}(r) - F_{12}A_{21}A_{11}^{-1} & -F_{12}(r) + F_{12} \\ P_{1,21}(r) - F_{21}A_{11}^{-1} & P_{1,22}(r) \end{bmatrix}$$

NB: Additional quantities, e.g. the R_j, can be found in the proof of Theorem 3.

Table 3

Simulation results for Designs 1 and 2

	$\sigma_{21} = 0$			$\sigma_{21} = 0.5$		
	Low	High	Mixed	Low	High	Mixed
Design 1: Bias						
a_{11}	−0.0472	−0.0239	−0.0264	−0.0670	−0.0251	−0.0288
a_{12}	0.0141	−0.0033	−0.0001	0.0370	0.0012	0.0088
a_{21}	0.0178	0.0018	0.0025	0.0533	0.0054	0.0143
a_{22}	−0.0488	−0.0246	−0.0273	−0.0829	−0.0273	−0.0363
σ_{11}	0.0186	0.0028	0.0034	0.0210	0.0037	0.0035
σ_{21}	−0.0031	0.0008	0.0019	−0.0039	0.0020	−0.0002
σ_{22}	0.0216	0.0022	0.0108	0.0276	0.0037	0.0115
Design 1: RMSE						
a_{11}	0.3123	0.1550	0.1836	0.4249	0.1687	0.2333
a_{12}	0.3063	0.1558	0.2062	0.4183	0.1674	0.2508
a_{21}	0.3072	0.1539	0.2093	0.4366	0.1687	0.2611
a_{22}	0.3124	0.1578	0.2366	0.4478	0.1705	0.2833
σ_{11}	0.2358	0.0919	0.0968	0.2155	0.0917	0.0927
σ_{21}	0.1665	0.0683	0.1358	0.1729	0.0741	0.1249
σ_{22}	0.2387	0.0914	0.2050	0.2242	0.0908	0.1869
Design 2: Bias						
a_{11}	−0.0468	−0.0251	−0.0281	−0.0462	−0.0233	−0.0279
a_{12}	−0.0118	0.0001	−0.0027	−0.0085	0.0035	0.0023
a_{21}	−0.0067	0.0063	0.0039	−0.0037	0.0093	0.0101
a_{22}	−0.0422	−0.0224	−0.0257	−0.0423	−0.0211	−0.0227
σ_{11}	0.0200	0.0031	0.0041	0.0200	0.0016	0.0031
σ_{21}	0.0007	−0.0008	−0.0013	0.0081	−0.0008	−0.0039
σ_{22}	0.0178	0.0018	0.0110	0.0171	−0.0002	0.0045
Design 2: RMSE						
a_{11}	0.3049	0.1539	0.1837	0.3032	0.1540	0.1755
a_{12}	0.2996	0.1539	0.2085	0.3026	0.1537	0.1869
a_{21}	0.3039	0.1534	0.2170	0.2972	0.1541	0.1898
a_{22}	0.3104	0.1535	0.2450	0.3010	0.1544	0.2041
σ_{11}	0.2354	0.0919	0.0964	0.2676	0.0948	0.0983
σ_{21}	0.1640	0.0680	0.1398	0.2237	0.0743	0.1300
σ_{22}	0.2376	0.0916	0.2253	0.2637	0.0942	0.1982

Table 4

Simulation results for Design 3

	$\sigma_{21} = 0$			$\sigma_{21} = 0.5$		
	Low	High	Mixed	Low	High	Mixed
Design 3: Bias						
β_c	0.00014	0.00006	0.00003	0.00013	0.00009	0.00006
α_1	0.0005	0.0052	0.0073	0.0308	0.0052	0.0078
α_2	-0.0359	-0.0183	-0.0206	-0.0601	-0.0262	-0.0366
σ_{11}	-0.0051	-0.0012	-0.0003	-0.0061	-0.0019	-0.0008
σ_{21}	-0.0193	-0.0052	-0.0115	-0.0268	-0.0050	-0.0132
σ_{22}	-0.0076	-0.0010	-0.0024	-0.0111	-0.0026	-0.0022
Design 3: RMSE						
β_c	0.0067	0.0051	0.0060	0.0082	0.0071	0.0079
α_1	0.1736	0.1151	0.1369	0.2336	0.1561	0.1921
α_2	0.2231	0.1222	0.1698	0.2984	0.1610	0.2330
σ_{11}	0.2199	0.0892	0.0977	0.1891	0.0875	0.0926
σ_{21}	0.2099	0.0721	0.1310	0.1821	0.0710	0.1265
σ_{22}	0.2213	0.0892	0.1745	0.1913	0.0870	0.1616

Table 5

Summary results of estimating (26) and restricted models using monthly price indices and monthly, weekly and daily exchange rates

	Exchange rate frequency		
	Monthly	Weekly	Daily
Unrestricted model (26)			
$\ln L$	2543.0592	4364.8994	16,154.1533
SIC	-5010.7355	-8654.4159	$-32{,}232.9237^*$
Restriction: $\beta_p = -\beta_{p^*}$			
$\ln L$	2543.0575	4364.8834	16,133.7269
SIC	-5016.1166	-8659.7684	$-32{,}197.4554$
LR_1	0.0034	0.0320	40.8528
p-value	[0.9535]	[0.8580]	[0.0000]
Restriction: $\beta_p = -\beta_{p^*} = -1$			
$\ln L$	2543.0439	4364.8833	15,948.8692
SIC	-5021.4739^*	-8665.1528^*	$-31{,}833.1245$
LR_2	0.0306	0.0322	410.5682
p-value	[0.9848]	[0.9840]	[0.0000]

A * denotes the minimum SIC in the column.

Table 6
Estimates of unrestricted (U) and 'preferred' (P) models

	Monthly		Weekly		Daily
	U	P	U	P	U/P
α_s	−0.0391	−0.0403	−0.0473	−0.0491	−0.9267
	(0.0210)	(0.0213)	(0.0178)	(0.0231)	(0.0852)
α_p	0.0056	0.0059	0.0063	0.0063	0.3317
	(0.0035)	(0.0031)	(0.0029)	(0.0031)	(0.1236)
α_{p*}	0.0011	0.0004	−0.0001	0.0000	−0.3827
	(0.0052)	(0.0031)	(0.0015)	(0.0031)	(0.0936)
β_p	−0.6313	−1.0000	−0.8815	−1.0000	−3.0260
	(1.0707)		(1.4411)		(0.2601)
β_{p*}	0.5914	1.0000	0.8190	1.0000	3.2674
	(0.9253)		(1.6890)		(0.3223)
μ_s	0.0137	0.0221	0.0121	0.0267	1.6213
	(0.0641)	(0.0116)	(0.0414)	(0.0125)	(0.3253)
μ_p	0.0001	−0.0012	0.0004	−0.0014	−0.5634
	(0.0082)	(0.0017)	(0.0049)	(0.0017)	(0.1353)
μ_{p*}	0.0014	0.0015	0.0018	0.0018	0.6541
	(0.0032)	(0.0017)	(0.0001)	(0.0017)	(0.2379)
m_{11}	0.0246	0.0246	0.0269	0.0269	0.0162
	(0.0012)	(0.0012)	(0.0006)	(0.0006)	(0.0005)
m_{21}	0.0007	0.0007	0.0008	0.0008	0.0934
	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0302)
m_{22}	0.0035	0.0035	0.0035	0.0035	0.0839
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0268)
m_{31}	0.0002	0.0002	0.0002	0.0002	−0.1066
	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0254)
m_{32}	0.0006	0.0006	0.0006	0.0006	−0.0971
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0228)
m_{33}	0.0036	0.0036	0.0036	0.0036	0.0059
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0018)
κ	−0.3424	−0.5409	−0.2511	−0.5398	−1.7392

Numbers in parentheses denote standard errors.