Locally Exact Discrete Time Representations of Non-Linear Continuous Time Models

Marcus J. Chambers[a], Theodore Simos[b] and Mike Tsionas[c]

[a] *Department of Economics, University of Essex*
[b] *Department of Economics, University of Ioannina*
[c] *Department of Economics, Lancaster University Management School*

September 2023

Abstract

This paper jointly addresses the issues of non-linearity and aggregation over time in continuous time dynamic systems. Our approach is based on a representation, called the locally exact discrete model (LEDM), which is derived for systems of non-linear stochastic differential equations (SDEs). The LEDM utilises a local approximation principle and shares many of the advantages of the well-known exact discrete model (EDM) which holds for linear continuous time systems. Not only does it nest the EDM when the underlying system of SDEs is linear, but it also provides a basis for the efficient Bayesian or Classical estimation of multivariate economic models formulated as systems of non-linear SDEs. We demonstrate the performance of the LEDM method in simulations with bivariate non-linear SDE systems containing stock variables as well as a mixture of a stock and a flow. The simulations show that our method has good accuracy in estimating the underlying structural parameters of the non-linear SDE system.

Keywords: non-linear stochastic differential equations; continuous time; locally exact discrete model.

J.E.L. classification number: C32.

Acknowledgements: We thank Michael Thornton for helpful comments on an earlier version of this paper.

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

1. Introduction

The estimation and use of continuous time dynamic models in macroeconomics and finance has matured into a well-established area of research and empirical practice. Early work on the estimation of continuous time models focused on systems of linear stochastic differential equations (SDEs) and a number of methods have been established for the estimation of such models using data observed at discrete time intervals. Prominent among these methods are Kalman filtering techniques based on appropriate state space forms (Harvey and Stock, 1985; Zadrozny, 1988), frequency domain approaches based on Fourier transforms (Robinson, 1976, 1993), and time domain methods based on an exact discrete time representation (Phillips, 1972; Bergstrom, 1983, 1985). In empirical applications Christiano, Eichenbaum and Marshall (1991), for example, used a frequency domain approximation to the Gaussian likelihood function in order to estimate their continuous time model of consumption behaviour based on a representative agent's first-order optimisation conditions, while Bergstrom and Nowman (2007) based their estimation of an eighteen-equation non-linear macroeconometric model of the United Kingdom on an exact discrete time representation corresponding to the underlying system of linearised SDEs.

A limitation of the above methods is that they are predominantly constructed for linear dynamic models.[1] Although the model of Bergstrom and Nowman (2007) consists of eighteen mixed first- and second-order non-linear SDEs the model was nevertheless linearised around its steady state prior to its estimation using an exact discrete model (EDM) corresponding to the linearised system. An EDM has the advantage that observations generated by the continuous time system satisfy the EDM without any approximation or interpolation errors. An alternative approach to the estimation of non-linear continuous time systems is to evaluate the likelihood function based on a numerical solution of the non-linear SDEs (Bailey, Hall and Phillips, 1987; Wymer, 1993). Non-linearities are, however, an important feature of many of the models of interest in macroeconomics and finance, and such non-linearities can influence dynamic features such as the rate of adjustment and, obviously, the steady state solution itself. The standard practice of linearising the system does not provide an adequate solution since non-linearities play a pivotal role in the system's reaction to shocks (Brunnermeier and Sannikov, 2014). Another drawback of linearisation is that the approximation errors that enter into the likelihood function can grow with the sample size (Fernandez-Villaverde, Rubio-Ramirez and Santos, 2006).

The present paper develops a method that tackles the twin problems of non-

[1] An exception is Robinson (1976) whose model is much more general than a linear system but has the drawbacks that it is not applicable to closed systems and also imposes a restrictive aliasing constraint on the spectral densities of the exogenous variables.

linearity and temporal aggregation that provide challenges to the estimation of continuous time dynamic stochastic models from discretely sampled data. We base our estimation method on what we call the *locally exact discrete model (LEDM)*, a non-linear discrete time dynamic model that results from the application of a local linearisation principle to the drift term originally proposed by Shoji and Ozaki (1997, 1998). Our data generating process (DGP) is formulated as a non-linear system of SDEs of the type that are frequently used in macroeconomics and finance, including dynamic stochastic general equilibrium (DSGE) models as well as more standard macro-dynamic models. We obtain the precise form of the LEDM, along with the covariance properties of the discrete time error terms, in three important cases of data sampling that are of empirical relevance: (i) stock data; (ii) flow data; and (iii) mixed stock and flow data. The last case is the most pertinent for macroeconomic models in which the variables are typically a mixture of stocks and flows. We include the other two cases, however, partly to help motivate ideas in the simplest setting of stock data, and also out of a desire for completeness. Although the LEDM is not the globally exact discrete time model, it can nevertheless be regarded in practice as a conditional Gaussian approximation of the DGP, possessing several advantages:

1. Alternative estimation methods – both classical and Bayesian – can be built on the basis of the LEDM. In our simulation study, for example, we implement a Bayesian estimation algorithm that is especially suited to the estimation of the deep structural parameters of the models.

2. The estimation algorithm is computationally efficient for medium size systems of SDEs of the type usually encountered in macroeconometric modelling. Alternative methods of solving, either numerically or analytically, the Kolmogorov partial differential equations that characterise the transition densities are rather challenging, especially for large systems.

3. An important advantage of the LEDM is that it nests the EDM when the data generating process is actually linear.

4. As in the exact discretisation of linear models, a useful by-product of our method is that it yields a non-linear model that has vector autoregressive (VAR) characteristics, which can also be useful as a basis for other purposes, such as hypothesis testing, forecasting, and Monte Carlo studies.

The LEDM method therefore provides a computationally efficient degree of generality and flexibility in non-linear continuous time modelling that extends the range of possibilities beyond purely linear models.

We derive the LEDM for each of the three data sampling schemes allowing for a general sampling interval of length h. This enables us to explore the order of the LEDM approximation error as the sampling frequency increases i.e. as h becomes smaller. We do this by considering a local strong measure of the approximation error and show that it is $O(h)$ for stock and mixed data and $O(h^2)$ with flow data. Use of a general sampling interval also enables further extensions to mixed frequency data of the type considered in continuous time systems by Chambers (2016) although such an extension lies considerably beyond the scope of the present paper.

The plan of the paper is as follows. In the next section the continuous time model is defined along with a set of assumptions appropriate for the case of stock variables. The precise form of the LEDM is presented in Theorem 1 which also contains the covariance properties of the discrete time disturbance vectors. The properties of the discrete time disturbances are then utilised in constructing the likelihood function based on the Gaussian properties that feed through from the disturbances in the continuous time model which is driven by a vector of Wiener processes. Under mild additional assumptions the local strong approximation error is shown to be $O(h)$ as $h \to 0$ (Proposition 1). Section 3 is devoted to deriving the LEDM in the cases of pure flow data (Theorem 2) and a mixture of stocks and flows (Theorem 3), the latter being the most relevant for the majority of empirical applications. The Gaussian likelihood functions are derived in each case based on the conditional first-order moving average properties of the discrete time disturbances and the local strong approximation error is shown to be $O(h^2)$ in the case of flow sampling (Proposition 2). The presence of stocks in the mixed sampling scenario suggests that the approximation error is $O(h)$ in that case.

Section 4 presents the results of a simulation study based on two different DGPs. Both DGPs are bivariate and non-linear in nature but one contains only stock variables while the other consists of a stock and a flow. The results demonstrate that our LEDM-based method performs well in finite samples. Some concluding comments are presented in section 5 and the proofs of the main results contained in the paper are provided in the Appendix.

In terms of notation, I_n denotes an $n \times n$ identity matrix, $\|A\| = \sqrt{\mathrm{tr}(AA')}$ denotes the Euclidean norm of the matrix A, $\mathrm{tr}\{A\}$ and $|A|$ denote the trace and determinant of a square matrix A, respectively, and the matrix exponential is defined by

$$e^A = \sum_{j=0}^{\infty} \frac{1}{j!} A^j,$$

also for a square matrix A. The notation $\lfloor x \rfloor$ denotes the integer part of x while $a \propto b$ denotes that a is proportional to b.

2. The model and the LEDM with stock sampling

Our continuous time model concerns an $n \times 1$ vector of observable variables, $y(t)$, that is related to an $n \times 1$ vector of state variables, $x(t)$, neither of which is observed as a continuous record. The model itself consists of the following two equations:

$$\textit{Measurement equation:} \quad y(t) = \gamma\big(x(t)\big), \quad t > 0; \tag{1}$$

$$\textit{Transition equation:} \quad dx(t) = \mu\big(x(t);\theta\big)dt + \Sigma\big(x(t);\theta\big)dW(t), \quad t > 0. \tag{2}$$

In (1), $\gamma(x)$ is a possibly non-linear function relating the observable variables to the state vector, while in (2) $\mu(x;\theta)$ and $\Sigma(x;\theta)$ represent a drift vector and a diffusion matrix, respectively, θ is a $p \times 1$ vector of unknown parameters, and $W(t)$ is an $n \times 1$ Wiener process (or standard Brownian motion) defined on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ with filtration $(\mathfrak{F}_t)_{t\geq 0}$ which satisfies $E\big(dW(t)\big) = 0$ and $E\big(dW(t)dW(t)'\big) = I_n dt$. The functions $\gamma(\cdot)$, $\mu(\cdot)$ and $\Sigma(\cdot)$ are all assumed to be known – we do not consider cases where these are unknown and need to be estimated non-parametrically. In addition $\mu(\cdot)$ and $\Sigma(\cdot)$ are assumed to be time invariant; the analysis can be extended to time varying diffusions, however, by incorporating time as an extra state variable, although we do not explicitly do so here.

In the standard case of a constant diffusion term the transformation in the measurement equation has the form $\gamma\big(x(t)\big) = \Sigma^{-\frac{1}{2}}x(t)$, where Σ is a positive definite symmetric matrix of constants independent of the state variables $x(t)$. For the rest of the paper, however, it is useful to work with the more general case where the diffusion term is state dependent as in (2). We therefore assume that the observable variables of interest satisfy a system of n non-linear transformations that act instantaneously on a state vector generated by a system of n first-order non-linear SDEs. Note that, in the case where $y(t) = x(t)$, the transition equation (2) represents a non-linear system of SDEs in the vector $y(t)$ directly. Our objective is to derive a representation – the LEDM – in terms of the discrete time observations on $y(t)$ that is consistent with the underlying non-linear continuous time system.

The specification of our proposed modelling framework is completed by a sampling equation that relates the discrete time observations to the continuous time process $y(t)$. We assume that the data are recorded at discrete time points $h, 2h, \ldots, Nh$ in the time interval $[0, T]$, where h is the observation (or sampling) interval, T denotes the data span and $N = T/h$ is an integer denoting the sample size. Often in the literature the convention $h = 1$ is adopted to imply an annual time interval, while $h = 1/12$ (1/52 or 1/250) a monthly (weekly or daily) time step, implying for most cases of practical importance that $h \leq 1$. In this section we consider the following sampling equation:

$$\textit{Sampling equation (stocks):}\ \ y_{th} = y(th), \quad t = 1, \ldots, N. \tag{3}$$

Stock variables therefore represent observations of the underlying continuous time process sampled at (equispaced) points in time at intervals of length h; this assumption will be relaxed in the next section to allow for flow and mixed stock and flow sampling.

Some additional assumptions are required on the drift and diffusion functions in (2), as well as the measurement equation function in (1), for the validity of our proposed method. In the sequel, to ease notation, it is sometimes convenient to suppress explicit dependence on the parameter vector θ and also, when there is no confusion, on the state variables.

Assumption 1. *The drift function $\mu(x;\theta)$ and diffusion matrix $\Sigma(x;\theta)$ are known, measurable functions defined on the domains of x and $\theta \in \Theta \subset \mathbb{R}^p$ and the matrix $V(x;\theta) = \Sigma(x;\theta)\Sigma(x;\theta)'$ is positive definite for all x and $\theta \in \Theta$.*

Assumption 2. *The drift and diffusion functions satisfy, for all $x, z \in \mathbb{R}^n$, the local Lipschitz condition*

$$\|\mu(x) - \mu(z)\| + \|\Sigma(x) - \Sigma(z)\| \leq K\|x - z\|$$

and the the linear growth condition[2]

$$\|\mu(x)\| + \|\Sigma(x)\| \leq K\left(1 + \|x\|\right),$$

for some constant $K > 0$.

Assumption 3. *The inverse of the measurement equation function, denoted $\gamma^{-1}(y)$, exists and is unique. Furthermore, the $n \times n$ matrix*

$$\Gamma(x) = \frac{\partial \gamma(x)}{\partial x'}$$

is of full rank for all x.

Assumption 1 is standard while Assumption 2 contains regularity conditions that imply the existence and uniqueness of a strong solution to the system of stochastic differential equations (2); see, *inter alia*, Arnold (1974, Theorem 6.2.2, p.105) and Karatzas and Shreve (1991, Theorem 5.2.9, p.289). Aït-Sahalia (2002) and Aït-Sahalia and Mykland (2004) discuss an analogous set of regularity conditions relevant for the existence of a weak solution of univariate diffusions. The first part of Assumption

[2]Some authors, for example Karatzas and Shreve (1991, equation 5.2.13), write the linear growth condition in the form $\|\mu(x)\|^2 + \|\Sigma(x)\|^2 \leq K^2\left(1 + \|x\|^2\right)$.

3 ensures that we can express the unobservable state vector $x(t)$ in terms of the observable vector $y(t)$ at certain points in the proofs of the theorems that follow, while the second part ensures that the conditional covariance matrices of the disturbance vectors in the LEDMs are positive definite (in conjunction with Assumption 1). The derivation of the LEDM relies on a further assumption as follows:

Assumption 4. *Define, for $i = 1, \ldots, n$, the scalar functions $f_i(y) = g_i(\gamma^{-1}(y))$ and $n \times 1$ vector functions $\omega_i(y) = h_i(\gamma^{-1}(y))$, where*

$$g_i(x) = \left(\frac{\partial \gamma_i(x)}{\partial x}\right)' \mu(x) + \frac{1}{2}\mathrm{tr}\left\{\Sigma(x)' \frac{\partial^2 \gamma_i(x)}{\partial x \partial x'} \Sigma(x)\right\},$$

$$h_i(x)' = \left(\frac{\partial \gamma_i(x)}{\partial x}\right)' \Sigma(x).$$

Furthermore, let

$$A_i(y) = \frac{\partial f_i(y)}{\partial y}, \quad B_i(y) = \frac{\partial^2 f_i(y)}{\partial y \partial y'}, \quad i = 1, \ldots, n.$$

Then, for each $t = 1, \ldots, N-1$ and $i = 1, \ldots, n$, it is assumed that

$$\left.\begin{aligned} A_i\big(y(s)\big) &= A_i\big(y(th)\big) \equiv A_{i,th}, \\ B_i\big(y(s)\big) &= B_i\big(y(th)\big) \equiv B_{i,th}, \\ \omega_i\big(y(s)\big) &= \omega_i\big(y(th)\big) \equiv \omega_{i,th}, \end{aligned}\right\} \quad th \leq s < th + h.$$

Assumption 4 is used to ensure that the derivatives of certain functions arising in the derivation of the LEDM remain constant over each observation interval, their values being dependent on the observed value of $y(t)$ at the start of the relevant interval. A similar type of assumption was used by Nowman (1997) to capture the evolution of stochastic volatility in a single-factor model of interest rates and by Shoji and Ozaki (1997) in their local linearisation method for scalar SDEs. The LEDM pertaining to discrete time stock sampling defined in (3) is given in Theorem 1 below.

Theorem 1. *Let $y(t)$ $(0 < t \leq T)$ satisfy (1) and (2) and let the observations y_{th} $(t = 1, \ldots, N)$ satisfy (3). Then, under Assumptions 1–4, the LEDM is given by*

$$y_{th+h} = \Theta_{th} y_{th} + \Upsilon_{1,th+h} C_{th} + \Upsilon_{2,th+h} \Delta_{th} + \xi_{th+h}, \quad t = 1, \ldots, N-1, \tag{4}$$

where

$$\xi_{th+h} = \int_{th}^{th+h} e^{A_{th}(th+h-r)} \Omega_{th} dW(r)$$

and the matrices and vectors are defined in Table 1.

The LEDM in Theorem 1 is in the form of a vector autoregression that contains time-varying coefficients and conditional heteroskedasticity which arise mainly because of the influence of the local approximations that are utilised within each sampling interval. The disturbance vector ξ_{th+h} has zero mean in view of $e^{A_{th}(th+h-r)}\Omega_{th}$ and $dW(r)$ being uncorrelated[3] for $r \in [th, th+h)$. Defining $E_{th}(\cdot)$ to be the expectation operator conditional on the filtration $\mathfrak{F}_{th}$ i.e. $E_{th}(x_{th+h}) = E(x_{th+h}|\mathfrak{F}_{th})$ for a random variable x_{th} defined on the probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, the conditional covariance matrix is then given by

$$M_{th+h|th} = E_{th}\left(\xi_{th+h}\xi'_{th+h}\right) = \int_0^h e^{A_{th}r}\Omega_{th}\Omega'_{th}e^{A'_{th}r}dr.$$

The unconditional covariance matrix would be difficult to evaluate as it would require the determination of the unconditional expectation

$$E\left(e^{A_{th}r}\Omega_{th}\Omega'_{th}e^{A'_{th}r}\right)$$

in which each quantity is a complicated function of y_{th} itself. Following Shoji and Ozaki (1997) we proceed by treating y_{th} as given (non-random) in these expressions which enables the (quasi-)autocovariance structure to be derived; in this case, ξ_{th+h} is an uncorrelated normally distributed zero mean random disturbance vector with covariance matrix

$$M_{0,th+h} = \int_0^h e^{A_{th}r}\Omega_{th}\Omega'_{th}e^{A'_{th}r}dr, \tag{5}$$

the lack of serial correlation arising because the Wiener processes in ξ_{th+h} and $\xi_{th+h-jh}$ are uncorrelated for $j \neq 0$ when y_{th} is treated as fixed. It is straightforward to show that $M_{0,th+h}$ is positive definite because the matrix exponentials are nonsingular and

$$\Omega(x)\Omega(x)' = \Gamma(x)\Sigma(x)\Sigma(x)'\Gamma(x)'$$

is positive definite in view of the full rank of $\Gamma(x)$ in Assumption 3 and the positive definiteness of $\Sigma(x)\Sigma(x)'$ in Assumption 1.

The LEDM in Theorem 1, allied with the autocovariance properties of ξ_{th+h} described above, provides the basis for constructing the Gaussian likelihood function that can be used to obtain estimates of the parameters of interest. Let Y denote the $N \times n$ matrix of observations on the stock variables, with typical row y'_{th}, and, using Theorem 1, let

$$\xi_{th+h} = y_{th+h} - \Theta_{th}y_{th} - \Upsilon_{1,th+h}C_{th} - \Upsilon_{2,th+h}\Delta_{th}, \quad t = 1, \ldots, N-1$$

[3]In fact, due to the normality of the increments of $W(t)$, these quantities are also independent.

denote the vector of disturbances in the LEDM which we treat as being uncorrelated with covariance matrix $M_{0,th+h}$ defined in (5). Then the log-likelihood function is

$$\log L(\theta; Y) = -\frac{n(N-1)}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{N-1}\log|M_{0,th+h}| - \frac{1}{2}\sum_{t=1}^{N-1}\xi'_{th+h}M_{0,th+h}^{-1}\xi_{th+h}. \quad (6)$$

This is the most straightforward of the three sampling schemes to handle from a computational viewpoint as construction of the log-likelihood function only involves determinants and inverses of $n \times n$ matrices.

The log-likelihood function presented above can be utilised for estimation of the parameter vector θ using either classical or Bayesian methods. In the classical case the estimator, $\hat{\theta}$, is obtained as the argument that maximises the log-likelihood function itself, whereas in the Bayesian case prior information about θ is used to construct a posterior distribution which is maximised in order to determine $\hat{\theta}$. More precisely, let $p(\theta)$ denote the prior distribution reflecting beliefs about θ; the posterior distribution is then given by

$$p(\theta; Y) \propto L(\theta; Y)p(\theta). \quad (7)$$

The asymptotic properties of the resulting classical and Bayesian estimators can, in principle, be derived under appropriate regularity conditions.[4] In the case where $y(t)$ is stationary it can be expected that the resulting estimators will be asymptotically normally distributed with convergence to the limiting distribution taking place at a rate equal to the square root of the sample size, while when some nonstationarity and/or cointegration is present at least some parameter estimators can be expected to converge at the rate of the sample size to limiting distributions characterised by functionals of Brownian motion processes.

The accuracy of the LEDM approximation based on Assumption 4 can be expected to improve as the sampling frequency increases while in the case of linear models the LEDM corresponds to the exact discrete model in Bergstrom (1984, Theorem 3). This latter property can be demonstrated by observing that, when the drift term in (2) is linear, then C_{th} and Δ_{th} are zero vectors. It is also worth emphasising that, for non-linear models, the LEDM exploits information contained in the second-order derivatives through both C_{th} and Δ_{th}.

It is of interest to investigate more fully the accuracy of the approximation used in deriving the LEDM. With this in mind we compare the LEDM in Theorem 1 with the infeasible exact discrete time representation as follows. The vector $y(t)$ evolves in

[4]Establishing a set of general regularity conditions and providing a detailed analysis of the asymptotic properties of the likelihood-based estimators is beyond the scope of the present paper.

continuous time as[5]

$$dy(t) = f\big(y(t)\big)dt + \Omega\big(y(t)\big)dW(t), \quad t > 0, \tag{8}$$

where $f(y)$ is defined in Theorem 1 and

$$\Omega(y) = \Gamma(y)\Sigma(y) = \begin{pmatrix} \omega_1(y)' \\ \vdots \\ \omega_n(y)' \end{pmatrix},$$

the functions $\omega_i(y)$ $(i = 1, \ldots, n)$ being defined in Assumption 4 and $\Gamma(y)$ in Assumption 3. The discrete time observations then satisfy

$$y(th + h) = y(th) + \int_{th}^{th+h} f\big(y(r)\big)dr + \xi^a_{th+h}, \quad t = 1, \ldots, N-1, \tag{9}$$

where ξ^a_{th+h} denotes the actual discrete time disturbance given by

$$\xi^a_{th+h} = \int_{th}^{th+h} \Omega\big(y(r)\big)dW(r).$$

Our examination of the LEDM approximation error is then based on a comparison of ξ^a_{th+h} with ξ_{th+h}. The result is stated below.

Proposition 1. *Suppose that, in addition to Assumptions 1–4, $\|f\big(y(\tau)\big)\| < \infty$ for all $\tau \in (0, T]$ and $\|C_{th}\| < \infty$ for all $t = 1, \ldots, N-1$. Then, as $h \to 0$,*

$$E_{th}\left\|\xi^a_{th+h} - \xi_{th+h}\right\| = O(h), \quad t = 1, \ldots, N-1.$$

Proposition 1 provides the precise rate at which the LEDM approximation error vanishes, which is shown to be $O(h)$, and, hence, the LEDM has an approximation order of one. This rate is, in fact, the same as for the simple Euler approximation,[6] but the key feature of the LEDM is that it incorporates higher-order properties of the non-linear function in the SDE which can be important for a fixed sampling interval h. To see this, note that the disturbance in the simple Euler approximation is

$$\xi^E_{th+h} = y_{th+h} - y_{th} - hf(y_{th}),$$

whereas from (4) in Theorem 1 the LEDM disturbance is of the form

$$\xi_{th+h} = y_{th+h} - \Theta_{th}y_{th} - \Upsilon_{1,th+h}C_{th} - \Upsilon_{2,th+h}\Delta_{th},$$

[5]The SDE in (8) is obtained by stacking the equations in (A3) in the Appendix.

[6]This is relatively easy to establish given the results in the proof of Proposition 1.

in which the vector C_{th} depends on the second derivatives of the function $f(\cdot)$, the matrices Θ_{th}, $\Upsilon_{1,th+h}$ and $\Upsilon_{2,th+h}$ depend on exponential functions and integrals of first derivatives of $f(\cdot)$, and the vector Δ_{th} depends on both first and second derivatives of $f(\cdot)$. So, although both approximation errors are $O(h)$, the LEDM provides a more sophisticated approximation than the Euler scheme for any fixed h. Indeed, the simulation results of Shoji and Ozaki (1997) in a univariate setting suggest that a LEDM-type approximation outperforms the Euler scheme for a range of fixed values of h. We investigate the performance of our method in a subsequent section.

3. The LEDM with flow and mixed sampling

Some variables of interest to macroeconomists are not observable as stock variables in the manner defined in (3). Variables such as consumers' expenditure, gross domestic product and investment expenditures are all examples of flow variables whose observations consist of the accumulation of the underlying rate of flow over the observation interval. In this section we therefore investigate how the correct treatment of such variables affects the form of the LEDM. In the first instance we assume all of the variables in the vector $y(t)$ are flows before extending the analysis to the more challenging situation where $y(t)$ is comprised of a mixture of both stock and flow variables.

3.1. Flow sampling

In the case of flow variables the relevant sampling equation becomes:

$$\textit{Sampling equation (flows):}\ \bar{y}_{th} = \int_{th-h}^{th} y(r)dr, \quad t = 1, \ldots, N. \tag{10}$$

One immediate implication of this type of sampling is that it is not possible to construct the matrices in Assumption 4 owing to $y(th)$ not being observed. The following assumption therefore replaces Assumption 4 in the case of flow variables:

Assumption 5. *For $i = 1, \ldots, n$ let the functions $g_i(x)$, $c_i(x)$, $f_i(y)$, $A_i(y)$, $B_i(y)$ and $\omega_i(y)$ be defined as in Assumption 4. Then, for each $t = 1, \ldots, N-1$ and $i = 1, \ldots, n$, it is assumed that*

$$\left.\begin{aligned} A_i\big(y(s)\big) &= A_i(\bar{y}_{th}) \equiv \bar{A}_{i,th}, \\ B_i\big(y(s)\big) &= B_i(\bar{y}_{th}) \equiv \bar{B}_{i,th}, \\ \omega_i\big(y(s)\big) &= \omega_i(\bar{y}_{th}) \equiv \bar{\omega}_{i,th}, \end{aligned}\right\} \quad th \leq s < th + h,$$

where $\bar{y}_{th}$ is defined in (10).

Assumption 5 plays the same role as Assumption 4 in the case of stocks but the quantities are evaluated at the observed values of the flow variables rather than at the point-in-time values of stocks. The LEDM for flows is given in the following theorem.

Theorem 2. *Let $y(t)$ $(0 < t \leq T)$ be generated according to* (1) *and* (2) *and let the observations $\bar{y}_{th}$ $(t = 1, \ldots, N)$ satisfy* (10)*. Then, under Assumptions 1–3 and 5, the LEDM is given by*

$$\bar{y}_{th+h} = \bar{\Theta}_{th}\bar{y}_{th} + \bar{\Upsilon}_{1,th+h}\bar{C}_{th} + \bar{\Upsilon}_{2,th+h}\bar{\Delta}_{th} + \bar{\xi}_{th+h}, \quad t = 1, \ldots, N-1, \tag{11}$$

where

$$\bar{\xi}_{th+h} = \int_{th}^{th+h} \bar{K}_{1,th}(th + h - r)\bar{\Omega}_{th}dW(r) + \int_{th-h}^{th} \bar{K}_{2,th}(th - r)\bar{\Omega}_{th}dW(r)$$

and the matrices and vectors are defined in Table 2.

The LEDM in Theorem 2 also possesses the time-varying autoregressive form evident in the LEDM in Theorem 1 although the autocovariance properties are more complicated with flow variables. Treating $\bar{y}_{th}$ as fixed in the components defining $\bar{\xi}_{th+h}$, as in Shoji and Ozaki (1997), it follows that $\bar{\xi}_{th+h}$ is a zero mean normally distributed MA(1) (first-order moving average) process with (conditional) covariance matrix

$$\bar{M}_{0,th+h} = \int_0^h \bar{K}_{1,th}(r)\bar{\Omega}_{th}\bar{\Omega}'_{th}\bar{K}_{1,th}(r)'dr + \int_0^h \bar{K}_{2,th}(r)\bar{\Omega}_{th}\bar{\Omega}'_{th}\bar{K}_{2,th}(r)'dr \tag{12}$$

and (conditional) first-order autocovariance matrix

$$\bar{M}_{1,th+h} = \int_0^h \bar{K}_{2,th}(r)\bar{\Omega}_{th}\bar{\Omega}'_{th}\bar{K}_{1,th}(r)'dr. \tag{13}$$

These time-varying second moments enable the Gaussian likelihood function to be constructed, although the process is more complicated than in the case of stock variables owing to the MA disturbance vector. While a number of methods exist to handle models with MA disturbances, we shall provide details of a method that is commonly associated with exact discrete time representations of continuous time systems that was first proposed in this context by Bergstrom (1985).

We begin by defining the disturbance vector in the LEDM as a function of the observable vectors and the parameter matrices which, from Theorem 2, takes the form

$$\bar{\xi}_{th+h} = \bar{y}_{th+h} - \bar{\Theta}_{th}\bar{y}_{th} - \bar{\Upsilon}_{1,th+h}\bar{C}_{th} - \bar{\Upsilon}_{2,th+h}\bar{\Delta}_{th}, \quad t = 1, \ldots, N-1. \tag{14}$$

Let $\bar{\xi}$ denote the $n(N-1) \times 1$ vector comprising the $N-1$ vectors $\bar{\xi}_{2h}, \ldots, \bar{\xi}_{Nh}$ stacked vertically on top of each other. Then the $n(N-1) \times n(N-1)$ covariance matrix of $\bar{\xi}$, which we denote by $\bar{\Omega} = E(\bar{\xi}\bar{\xi}')$, is a block-Toeplitz matrix with $\bar{M}_{0,th+h}$ constituting the blocks on the principle diagonal, $\bar{M}_{1,th+h}$ and $\bar{M}'_{1,th+h}$ occupying the bands below and above the principal diagonal, respectively, and zeros everywhere else. Denoting the $N \times n$ matrix of observations on the flow variables by $\bar{Y}$ we can write the log-likelihood function in the form

$$\log L(\theta; \bar{Y}) = -\frac{n(N-1)}{2} \log 2\pi - \frac{1}{2} \log |\bar{\Omega}| - \frac{1}{2} \bar{\xi}' \bar{\Omega}^{-1} \bar{\xi}. \tag{15}$$

From a computational point of view there is now an apparent requirement to compute the determinant and inverse of the sparse $n(N-1) \times n(N-1)$ matrix $\bar{\Omega}$. Bergstrom (1985), however, proposed a method that requires only the computation of the determinant and inverse of $n \times n$ matrices, as in the case of stock variables, by exploiting the sparsity of $\bar{\Omega}$. He noted that the Cholesky decomposition of $\bar{\Omega}$, denoted $\bar{C}$ and which satisfies $\bar{\Omega} = \bar{C}\bar{C}'$, is a lower triangular $n(N-1) \times n(N-1)$ matrix that is also sparse, having no more than $2n$ non-zero elements in any row in the case of a vector MA(1) disturbance structure.[7] The sparsity enables the non-zero $n \times n$ submatrices of $\bar{C}$ to be computed following a series of recursions which, as shown by Bergstrom (1990, pp.150–154), converge rapidly to constant matrices which endows the method with additional computational efficiencies. Let $\bar{c}_{ii}$ $(i = 1, \ldots, n(N-1))$ denote the diagonal elements of $\bar{C}$, and define $\bar{\epsilon}$ to be the $n(N-1)$ vector of random variables satisfying $\bar{C}\bar{\epsilon} = \bar{\xi}$. The $n \times 1$ subvectors of $\bar{\xi}$ can be calculated using recursive formulae that exploit the sparsity of $\bar{C}$ and it is straightforward to show that $E(\bar{\epsilon}\bar{\epsilon}') = I_{n(N-1)}$. By noting that

$$\log |\bar{\Omega}| = \log |\bar{C}\bar{C}'| = 2 \log \prod_{i=1}^{n(N-1)} \bar{c}_{ii} = 2 \sum_{i=1}^{n(N-1)} \log \bar{c}_{ii}$$

and that

$$\bar{\xi}' \bar{\Omega}^{-1} \bar{\xi} = \bar{\epsilon}' \bar{C}' (\bar{C}\bar{C}')^{-1} \bar{C} \bar{\epsilon} = \bar{\epsilon}' \bar{\epsilon} = \sum_{i=1}^{n(N-1)} \bar{\epsilon}_i^2$$

we find that (15) can be written

$$\log L(\theta; \bar{Y}) = -\frac{n(N-1)}{2} \log 2\pi - \sum_{i=1}^{n(N-1)} \log \bar{c}_{ii} - \frac{1}{2} \sum_{i=1}^{n(N-1)} \bar{\epsilon}_i^2. \tag{16}$$

The calculation of the elements in (16) requires nothing more demanding than the inversion of $n \times n$ matrices. Moreover these computations can also be incorporated in a Bayesian approach to estimation along the lines described in the case of stock variables which led to the posterior distribution in (7).

[7] In the case of a vector MA(q) process the Cholesky matrix has no more that $n(q+1)$ non-zero elements in any row.

It is also of interest to examine the order of the LEDM approximation for flow variables. In order to do so we compare the conditional expectation of the difference between $\bar{\xi}_{th+h}$ in (14) with the infeasible exact discrete time representation disturbance derived from (8). From (9) we have

$$y(s) = y(s-h) + \int_{s-h}^{s} f\big(y(r)\big)dr + \int_{s-h}^{s} \Omega\big(y(r)\big)dW(r).$$

Integrating over $s \in (th, th+h]$ results in

$$\bar{y}_{th+h} = \bar{y}_{th} + \bar{F}_{th+h} + \bar{\xi}^{a}_{th+h}, \quad t = 1, \ldots, N-1, \tag{17}$$

where

$$\bar{F}_{th+h} = \int_{th}^{th+h} \int_{s-h}^{s} f\big(y(r)\big)drds, \quad \bar{\xi}^{a}_{th+h} = \int_{th}^{th+h} \int_{s-h}^{s} \Omega\big(y(r)\big)dW(r)ds.$$

These double integrals can be simplified by writing them as the sum of two single integrals; details are in the proof of Proposition 2.

Proposition 2. *Suppose that, in addition to Assumptions 1–3 and 5,* $\|f\big(y(\tau)\big)\| < \infty$ *for* $\tau \in (0, T]$ *and* $\|\bar{C}_{th}\| < \infty$ *for all* $t = 1, \ldots, N-1$*. Then, as* $h \to 0$*,*

$$E_{th}\left\|\bar{\xi}^{a}_{th+h} - \bar{\xi}_{th+h}\right\| = O(h^2), \quad t = 1, \ldots, N-1.$$

The LEDM therefore has a higher order of approximation in the case of flow variables than in the case of stocks. This feature is a consequence of the observations on flow variables being in the form of an integral of the continuous time process over an interval of length h which has the effect of increasing the approximation order from one to two.[8] We now turn to the case of mixed stock and flow sampling.

3.2. Mixed stock and flow sampling

In the case where the vector of interest $y(t)$ consists of both stock and flow variables we assume that the vector $y(t)$ is comprised of n^s stock variables, $y^s(t)$, and n^f flow variables, $y^f(t)$, where $n^s + n^f = n$, so that $y(t) = \big(y^s(t)', y^f(t)'\big)'$; there is no loss of generality in the ordering of stocks before flows. Our discrete time sampling equation

[8]Note, however, that the approximation order would be $O(h)$ if the observations were normalised by dividing the integrals by h i.e. if the observations were of the form $h^{-1}\int_{th-h}^{th} y(r)dr$. Such a normalisation has been found to be important in models with integrated and cointegrated variables by Chambers (2011, p.160).

is then of the form:

Sampling equation (mixed):

$$\tilde{y}_{th} = \begin{pmatrix} \tilde{y}^s_{th} \\ \tilde{y}^f_{th} \end{pmatrix} = \begin{pmatrix} y^s(th) - y^s(th-h) \\ \displaystyle\int_{th-h}^{th} y^f(r)dr \end{pmatrix}, \quad t = 1, \ldots, N. \quad (18)$$

The observed vectors that comprise $\tilde{y}_{th}$ are denoted $\tilde{y}^s_{th}$ and $\tilde{y}^f_{th}$ for stocks and flows, respectively. Note that the stock variables appear as first-differences owing to the way in which unobservable variables, namely integrals of stocks and levels of flows, are eliminated in the derivation of the LEDM.

As in the case of flow variables it is no longer possible to construct the quantities in Assumption 4 owing to $y(t)$ not being completely observable. However, the vector $\tilde{y}_{th}$ does not seem to be entirely relevant for this purpose as it contains the first difference of the stock component whereas an estimate of the level of $y(t)$ is actually required. We therefore use the following observable vector in the relevant assumption:

$$\tilde{y}^\dagger_{th} = \begin{pmatrix} y^s(th) \\ \displaystyle\int_{th-h}^{th} y^f(r)dr \end{pmatrix}, \quad t = 1, \ldots, N. \quad (19)$$

This vector contains the observed levels of stocks and flows and is a more appropriate quantity to use than $\tilde{y}_{th}$ for the purposes of the approximation. The relevant assumption in the mixed sampling case is then:

Assumption 6. *For $i = 1, \ldots, n$ let the functions $g_i(x)$, $c_i(x)$, $f_i(y)$, $A_i(y)$, $B_i(y)$ and $\omega_i(y)$ be defined as in Assumption 4. Then, for each $t = 1, \ldots, N-1$ and $i = 1, \ldots, n$, it is assumed that*

$$\left.\begin{aligned} A_i\big(y(s)\big) &= A_i(\tilde{y}^\dagger_{th}) \equiv \tilde{A}_{i,th}, \\ B_i\big(y(s)\big) &= B_i(\tilde{y}^\dagger_{th}) \equiv \tilde{B}_{i,th}, \\ \omega_i\big(y(s)\big) &= \omega_i(\tilde{y}^\dagger_{th}) \equiv \tilde{\omega}_{i,th}, \end{aligned}\right\} \quad th \leq s < th + h,$$

where $\tilde{y}^\dagger_{th}$ is defined in (19)*.*

Assumption 6 plays the same role as Assumptions 4 and 5 in the cases of stocks and flows, respectively. It is also convenient to partition various matrices and vectors conformably with the stock and flow elements of $y(t)$ and $\tilde{y}_{th}$ so that, for a generic

$n \times n$ matrix A and $n \times 1$ vector a, we can write

$$A = \begin{pmatrix} A^{ss} & A^{sf} \\ A^{fs} & A^{ff} \end{pmatrix}, \quad a = \begin{pmatrix} a^s \\ a^f \end{pmatrix}.$$

In this notation A^{sf} is an $n^s \times n^f$ matrix, for example, while a^f is an $n^f \times 1$ vector. We also use this notation for partitioning the $n \times n$ identity and null matrices, I_n and 0_n, respectively. A further assumption is also utilised to derive the LEDM and relates to a sub-matrix of the $n \times n$ matrix

$$\tilde{A}_{th} = \begin{pmatrix} \tilde{A}'_{1,th} \\ \vdots \\ \tilde{A}'_{n,th} \end{pmatrix}, \tag{20}$$

where the $\tilde{A}_{i,th}$ $(i = 1, \ldots, n)$ are defined in Assumption 6.

Assumption 7. *The $n^s \times n^s$ sub-matrix $\tilde{A}^{ss}_{th}$ is nonsingular for all $t = 1, \ldots, N-1$.*

This assumption enables unobserved components[9] to be eliminated from the system when solving for the LEDM and has been made by a number of authors, including Agbeyegbe (1987, 1988), Simos (1996) and Chambers (2009).[10] It should be stressed that we do not require the invertibility of the entire matrix $\tilde{A}_{th}$ which would, for example, preclude stochastic trending behaviour in the system. The LEDM is presented below.

Theorem 3. *Let $y(t)$ $(0 < t \leq T)$ be generated according to (1) and (2) and let the observations $\tilde{y}_{th}$ $(t = 1, \ldots, N)$ satisfy (18). Then, under Assumptions 1–3, 6 and 7, the LEDM is given by*

$$\tilde{y}_{th+h} = \tilde{\Phi}_{th}\tilde{y}_{th} + \tilde{\gamma}_{th+h} + \tilde{\lambda}_{th+h}, \quad t = 1, \ldots, N-1, \tag{21}$$

where

$$\tilde{\lambda}_{th+h} = \int_{th}^{th+h} \tilde{K}_{3,th}(th + h - r)\bar{\Omega}_{th} dW(r) + \int_{th-h}^{th} \tilde{K}_{4,th}(th - r)\tilde{\Omega}_{th} dW(r)$$

and the matrices and vectors are defined in Tables 3 and 4, respectively.

The LEDM in Theorem 3 displays similar time-varying parameters and heteroskedasticity as in the pure stock and flow cases although the formulae that underlie the parameter matrices and covariances are more complicated owing to the mixed nature

[9]These unobserved components refer to integrals of stock variables and to levels of flow variables.

[10]In fact, Agbeyegbe (1987, 1988) and Simos (1996) also assume that the entire matrix A is nonsingular, an assumption that is not required here.

of the data and the additional complexities involved in solving out unobservable components from the system. Conditional on $\tilde{y}_{th}$ the disturbance vector, $\tilde{\lambda}_{th+h}$, is a zero mean normally distributed process with covariance matrix

$$\tilde{M}_{0,th+h} = \int_0^h \tilde{K}_{3,th}(r)\tilde{\Omega}_{th}\tilde{\Omega}'_{th}\tilde{K}_{3,th}(r)'dr + \int_0^h \tilde{K}_{4,th}(r)\tilde{\Omega}_{th}\tilde{\Omega}'_{th}\tilde{K}_{4,th}(r)'dr \tag{22}$$

and first-order autocovariance matrix

$$\tilde{M}_{1,th+h} = \int_0^h \tilde{K}_{4,th}(r)\bar{\Omega}_{th}\tilde{\Omega}'_{th}\bar{K}_{3,th}(r)'dr. \tag{23}$$

These time-varying second moments can be used to construct the Gaussian likelihood function in the same way as with pure flow variables. From Theorem 3 we have the disturbance vectors in the form

$$\tilde{\lambda}_{th+h} = \tilde{y}_{th+h} - \tilde{\Phi}_{th}\tilde{y}_{th} - \tilde{\gamma}_{th+h}, \quad t = 1, \dots, N-1. \tag{24}$$

As in the case of flows these disturbance vectors are MA(1) processes and so the same procedure as in the previous subsection can therefore be followed, resulting in the log-likelihood function

$$\log L(\theta; \tilde{Y}) = -\frac{n(N-1)}{2}\log 2\pi - \sum_{i=1}^{n(N-1)} \log \tilde{c}_{ii} - \frac{1}{2}\sum_{i=1}^{n(N-1)} \tilde{\epsilon}_i^2, \tag{25}$$

where $\tilde{Y}$ denotes the $N \times n$ matrix of observations on the stocks and flows, $\tilde{c}_{ii}$ denotes the i'th diagonal element of the sparse lower triangular Cholesky matrix $\tilde{C}$, and $\tilde{\epsilon}_i$ denote the elements of the $n(N-1)$ random vector $\tilde{\epsilon}$ which satisfies $\tilde{C}\tilde{\epsilon} = \tilde{\lambda}$, $\tilde{\lambda}$ denoting the $n(N-1) \times 1$ vector with typical $n \times 1$ subvector $\tilde{\lambda}_{th+h}$. This log-likelihood can be combined with a prior on θ in a Bayesian approach, as in (7).

In terms of the order of approximation error associated with the LEDM in the case of a mixed sample, it can be expected that the rate $O(h)$ holds due to the presence of stock variables. We do not provide a formal result for this claim but inspection of the form of $\tilde{\lambda}_{th}$, allied with the proofs of Propositions 1 and 2, suggests that the same arguments apply and will lead to the stated result.

4. Simulation evidence

In this section we present some simulation results obtained from two different DGPs. The first – DGP1 – is based on a model used by Aït-Sahalia (2008) and assesses the LEDM approach using a non-linear SDE containing two stock variables. The second – DGP2 – is based on a simplified version of the model of Christensen, Posch and Wel (2016) and contains a stock and a flow variable. The relevant log-likelihood functions were maximised using Bayesian Markov chain Monte Carlo methods based on the fast Metropolis Adjusted Langevin Algorithm with flat priors; see Durmas et al. (2017) for details of this algorithm.

4.1. DGP1

The bivariate simulation model used by Aït-Sahalia (2008) is based on a pair of Ornstein-Uhlenbeck processes given by

$$\begin{aligned} dz_1(t) &= \big(\kappa_{11}(\eta_1 - z_1(t)) + \kappa_{12}(\eta_2 - z_2(t))\big)dt + dW_1(t), \\ dz_2(t) &= \big(\kappa_{21}(\eta_1 - z_1(t)) + \kappa_{22}(\eta_2 - z_2(t))\big)dt + dW_2(t), \end{aligned}$$

where W_1 and W_2 are independent Wiener processes. Define $x_1(t) = \exp(z_1(t))$ and $x_2(t) = \exp(z_2(t))$. Application of Ito's Lemma results in the pair of non-linear SDEs given by[11]

$$dx_1(t) = x_1(t)\left\{\frac{1}{2} + \kappa_{11}\big(\eta_1 - \log x_1(t)\big) + \kappa_{12}\big(\eta_2 - \log x_2(t)\big)\right\}dt + x_1(t)dW_1(t), \quad (26)$$

$$dx_2(t) = x_2(t)\left\{\frac{1}{2} + \kappa_{21}\big(\eta_1 - \log x_1(t)\big) + \kappa_{22}\big(\eta_2 - \log x_2(t)\big)\right\}dt + x_2(t)dW_2(t). \quad (27)$$

Equations (26) and (27) constitute the transition equations; in terms of the representation in (2) we have

$$\mu(x;\theta) = \begin{pmatrix} x_1\left\{\frac{1}{2} + \kappa_{11}\big(\eta_1 - \log x_1\big) + \kappa_{12}\big(\eta_2 - \log x_2\big)\right\} \\ x_2\left\{\frac{1}{2} + \kappa_{21}\big(\eta_1 - \log x_1\big) + \kappa_{22}\big(\eta_2 - \log x_2\big)\right\} \end{pmatrix}, \ \Sigma(x;\theta) = \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix},$$

where $x = (x_1, x_2)'$ and $\theta = (\eta_1, \eta_2, \kappa_{11}, \kappa_{12}, \kappa_{21}, \kappa_{22})'$. The continuous time measurement equation and discrete time sampling equation corresponding to (1) and (3) are then given by, respectively,

$$y(t) = x(t) \quad (t > 0) \quad \text{and} \quad y_{th} = y(th) \quad (t = 1, \ldots, N).$$

The variables are therefore of the stock variety and the LEDM in Theorem 1 is relevant.

The parameter values used in the simulations are taken from Aït-Sahalia (2008) and can be found in Table 5. One of the adjustment parameters, κ_{21}, is set to zero but this information is not used in the estimations, unlike Aït-Sahalia (2008) who imposes the zero restriction in estimation. There are, therefore, six unknown parameters (the elements of θ) to estimate. We consider three different combinations of data span (T)

[11] See equation (49) of Aït-Sahalia (2008).

and sampling interval (h) in order to examine, in particular, how the estimates behave as sampling becomes more frequent. Case I sets $T = 100$ and $h = 1$ which can be interpreted as 100 years of annual data. The sample size is $N = 100$ in this case. Case II sets $T = 35$ and $h = 1/4$ (so that $N = 140$) which would correspond to 35 years of quarterly data. Finally, Case III sets $T = 25$ and $h = 1/12$ which can be regarded as 25 years of monthly data; here, $N = 300$.

The results of 10,000 simulations of the non-linear bivariate model DGP1 are reported in Table 5. The biases associated with the estimation of η_1 and η_2 are small and negative and decrease as the sample size increases; the standard deviations also get smaller with increasing sample size. For the speed-of-adjustment parameters κ_{ij} it can be seen that there is a small, positive bias in the estimation of κ_{11}, slightly larger negative biases in respect of κ_{12} and κ_{21}, and a larger bias connected to the estimation of κ_{22}. This last feature is also evident in the simulation results reported in Table 1 of Aït-Sahalia (2008) against whose results ours compare favourably bearing in mind that he used a larger sample size ($N = 500$) and imposed the restriction that $\kappa_{21} = 0$.

4.2. DGP2

Our second DGP involves a flow variable in addition to a stock variable and is a simplified version of the model proposed in Christensen, Posch and Wel (2016) who derive the equilibrium dynamics for a small macroeconomic model that includes a financial sector. Their model involves three SDEs of which we focus on the following two:

$$dr(t) = \kappa\big(\gamma - r(t)\big)dt + \eta dB(t), \tag{28}$$

$$d\log C(t) = \Big(r(t) - \rho - \delta - \frac{1}{2}\sigma^2\Big)dt + \sigma dZ(t), \tag{29}$$

where $C(t)$ denotes consumption (a flow variable), $r(t)$ is the rental rate of capital (a stock variable), and $Z(t)$ and $B(t)$ are standard Brownian motion (or Wiener) processes. One difficulty with this formulation is that, although $C(t)$ is a flow variable, we don't observe the integrals of the logarithms which are the relevant quantities for the LEDM. We can clearly take the logarithms of the integrals, but these quantities are not the same i.e.

$$\log \int_{th-h}^{th} C(r)dr \neq \int_{th-h}^{th} \log C(r)dr.$$

This issue has been discussed by Bailey, Hall and Phillips (1987) in the context of a continuous time macroeconomic model and by Seong, Ahn and Zadrozny (2013) when considering discrete time temporal aggregation with mixed frequency data. However, we can overcome this problem by the application of Ito's Lemma on the transformation

$C(t) = \exp(\log C(t))$, in which case our bivariate SDE system becomes

$$dr(t) = \kappa\big(\gamma - r(t)\big)dt + \sigma_1 dW_1(t), \tag{30}$$

$$dC(t) = C(t)\big(r(t) - \rho\big)dt + \sigma_2 C(t)dW_2(t), \tag{31}$$

where $W_1(t)$ and $W_2(t)$ are standard Brownian motion processes. Although this transformation has made the consumption equation non-linear in $C(t)$ we are able to handle this non-linearity using our LEDM method, an outcome which is not an option with standard linear continuous time methods.

Equations (30) and (31) represent the continuous time transition equations corresponding to (2) in terms of which

$$\mu(x;\theta) = \begin{pmatrix} \kappa(\gamma - x_1) \\ x_2(x_1 - \rho) \end{pmatrix}, \quad \Sigma(x:\theta) = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 x_2 \end{pmatrix},$$

where $x = (x_1, x_2)' = (r, C)'$ and $\theta = (\kappa, \gamma, \rho, \sigma_1, \sigma_2)'$. The continuous time measurement equation corresponding to (1) and the discrete time sampling equation corresponding to (18) are then simply

$$y(t) = \begin{pmatrix} r(t) \\ C(t) \end{pmatrix} \quad (t > 0) \quad \text{and} \quad \tilde{y}_{th} = \begin{pmatrix} r_{th} - r_{th-h} \\ \int_{th-h}^{th} C(r)dr \end{pmatrix} \quad (t = 1, \ldots, N).$$

In view of the mixed sampling the LEDM in Theorem 3 is relevant.

In the simulations we use the same five parameter values as in Christensen, Posch and Wel (2016) which can be found in Table 6. We also consider three different combinations of data span (T) and sampling interval (h) as follows. Case I sets $T = 75$ and $h = 1$ which can be interpreted as 75 years of annual data in which case the sample size is $N = 75$. Case II sets $T = 25$ and $h = 1/4$ (so that $N = 100$) which would correspond to 25 years of quarterly data. Finally, Case III sets $T = 25$ and $h = 1/12$ which can be regarded as 25 years of monthly data, so that $N = 300$.

The results of 10,000 replications of the model can be found in Table 6. The estimation biases associated with each parameter are small and become smaller as the sample size increases; the same is true of the standard deviations. This suggests that the LEDM methods works well in finite samples even in the case of mixed data sampling.

5. Conclusions

In this paper we have developed a method for estimating the parameters of a system of non-linear SDEs under three sampling scenarios for the discrete time data generated by the underlying continuous time system. The LEDM method thereby enables us to handle the stock/flow distinction of economic variables while also taking into account the non-linear features of the economic model in a manageable way. The LEDM method also possesses two further advantages. First, it nests the well-known exact discrete model when the underlying system of SDEs is linear, and secondly it provides a basis for the efficient Bayesian or Classical estimation of multivariate economic models formulated as systems of non-linear SDEs (including DSGE models). Our simulation results explored the performance of the LEDM method in two bivariate non-linear SDE systems, one containing only stock variables, the other containing a stock and a flow. In both cases the estimates have small finite sample biases that decrease as the sample size increases, as do the standard deviations, suggesting that the method provides a feasible estimation procedure in practical situations.

The LEDM method in this paper has been applied to a system of diffusion equations in which the drift and volatility terms can be non-linear functions of the parameters and variables, features which are of empirical relevance when dealing with macroeconomic and financial data. In the same way that exact discrete models can be extended from first- to higher-order systems in the linear case (see, for example, Bergstrom, 1983, Chambers, 1999, Chambers and Thornton, 2012) it is, in principle, possible to extend the LEDM method to higher-order non-linear SDEs, albeit with a rise in notational complexity. Moreover, in view of the relevance of our model to macroeconomics and finance and the fact that financial variables are typically observed more frequently than macroeconomics variables, the LEDM could realistically form the basis for handling systems with mixed frequency data, thereby extending the results for linear systems developed by Chambers (2016). Such extensions, as well as additional empirical applications, are beyond the scope of the present paper but represent potentially fruitful avenues for future research.

Appendix: Proofs

Proof of Theorem 1. Under Assumptions 1 and 2 there exists a unique strong solution to the stochastic differential equation system (2). Applying Ito's lemma to the i'th element of (1), and making use of the diffusion (2), results in

$$\begin{aligned} dy_i(t) &= \left[\left(\frac{\partial \gamma_i\big(x(t)\big)}{\partial x(t)} \right)' \mu\big(x(t)\big) + \frac{1}{2}\mathrm{tr}\left\{ \Sigma\big(x(t)\big)' \frac{\partial^2 \gamma_i\big(x(t)\big)}{\partial x(t) \partial x(t)'} \Sigma\big(x(t)\big) \right\} \right] dt \\ &\quad + \left(\frac{\partial \gamma_i\big(x(t)\big)}{\partial x(t)} \right)' \Sigma\big(x(t)\big) dW(t) \\ &= g_i\big(x(t)\big)dt + h_i\big(x(t)\big)' dW(t), \quad i = 1, \ldots, n, \end{aligned} \tag{A1}$$

where the functions $g_i(x)$ and $h_i(x)$ are defined in Assumption 4. We now use the inverse transformation in Assumption 3 allied with the functions $f_i(y)$ and $\omega_i(y)$ defined in Assumption 4 to write (A1) in terms of $y(t)$ in the form

$$dy_i(t) = f_i\big(y(t)\big)dt + \omega_i\big(y(t)\big)' dW(t), \quad i = 1, \ldots, n. \tag{A2}$$

Now, applying Ito's Lemma to the function $f_i(y)$ we obtain

$$df_i\big(y(t)\big) = \left(\frac{\partial f_i\big(y(t)\big)}{\partial y(t)} \right)' dy(t) + \frac{1}{2}\mathrm{tr}\left\{ \Omega\big(y(t)\big)' \frac{\partial^2 f_i\big(y(t)\big)}{\partial y(t) \partial y(t)'} \Omega\big(y(t)\big) \right\} dt, \quad i = 1, \ldots, n,$$

which is of the form

$$df_i\big(y(t)\big) = A_i\big(y(t)\big)' dy(t) + \frac{1}{2}\mathrm{tr}\left\{ \Omega\big(y(t)\big)' B_i\big(y(t)\big) \Omega\big(y(t)\big) \right\} dt, \quad i = 1, \ldots, n, \tag{A3}$$

where $A_i(y)$ and $B_i(y)$ are defined in Assumption 4 and $\Omega(y)$ is defined following (8). However, under Assumption 4, for $t \in [th, s)$ where $th \leq s < th + h$,

$$df_i\big(y(t)\big) = A'_{i,th} dy(t) + \frac{1}{2}\mathrm{tr}\left\{ \Omega'_{th} B_{i,th} \Omega_{th} \right\} dt, \quad i = 1, \ldots, n. \tag{A4}$$

Integrating (A4) from th to s yields

$$f_i\big(y(s)\big) - f_i\big(y(th)\big) = A'_{i,th}\big(y(s) - y(th)\big) + \frac{1}{2}\mathrm{tr}\left\{ \Omega'_{th} B_{i,th} \Omega_{th} \right\}(s - th), \quad i = 1, \ldots, n. \tag{A5}$$

Now evaluate (A2) at $t = s$:

$$dy_i(s) = f_i\big(y(s)\big)ds + \omega_i\big(y(s)\big)' dW(s), \quad i = 1, \ldots, n. \tag{A6}$$

Solving (A5) for $f_i\big(y(s)\big)$ and substituting the resulting expression into (A6) yields, for $i = 1, \ldots, n$,

$$\begin{aligned} dy_i(s) &= \Big(f_i\big(y(th)\big) + A'_{i,th}\big(y(s) - y(th)\big) + \frac{1}{2}\text{tr}\left\{\Omega'_{th} B_{i,th} \Omega_{th}\right\}(s - th) \Big) ds \\ &\quad + \omega_i\big(y(s)\big)' dW(s) \\ &= A'_{i,th} y(s) ds + \Big(f_i\big(y(th)\big) - A'_{i,th} y(th) - \frac{1}{2}\text{tr}\left\{\Omega'_{th} B_{i,th} \Omega_{th}\right\} th \Big) ds \\ &\quad + \frac{1}{2}\text{tr}\left\{\Omega'_{th} B_{i,th} \Omega_{th}\right\} s ds + \omega_i\big(y(s)\big)' dW(s). \end{aligned} \tag{A7}$$

Stacking over $i = 1, \ldots, n$ we then obtain the system

$$dy(s) = \big(A_{th} y(s) + C_{th} s + \Delta_{th}\big) ds + \Omega_{th} dW(s), \tag{A8}$$

where

$$A_{th} = \begin{pmatrix} A'_{1,th} \\ \vdots \\ A'_{n,th} \end{pmatrix}, \quad C_{th} = \begin{pmatrix} \frac{1}{2}\text{tr}\left\{\Omega' B_{1,th} \Omega\right\} \\ \vdots \\ \frac{1}{2}\text{tr}\left\{\Omega' B_{n,th} \Omega\right\} \end{pmatrix}, \quad \Delta_{th} = f\big(y(th)\big) - A_{th} y(th) - C_{th} th$$

and we have also used Assumption 4 to set $\omega_i\big(y(s)\big) = \omega_{i,th}$. Now define the indicator function

$$1_{[th,th+h)}(s) = \begin{cases} 1 & \text{if } \ th \leq s < th + h, \\ 0 & \text{otherwise.} \end{cases}$$

as well as the functions $A(s) = A_{th} 1_{[th,th+h)}(s)$, $\Psi(s) = (C_{th} s + \Delta_{th}) 1_{[th,th+h)}(s)$ and $\Omega(s) = \Omega_{th} 1_{[th,th+h)}(s)$ for $th \leq s < th + h$. Then (A8) can be written in the form

$$dy(s) = \big(A(s) y(s) + \Psi(s)\big) ds + \Omega(s) dW(s). \tag{A9}$$

Note that $A(s)$ and $\Omega(s)$ are matrices, and $\Psi(s)$ is a vector, of non-anticipating step functions on $[0, T]$ that remain constant over each interval $[th, th+h)$ $(t = 1, \ldots, N-1)$; for a definition of integrals of non-anticipating functions with respect to a Wiener process see, for example, Arnold (1974, p.65). Thus, under Assumptions 1–4, (A9) is a well-defined linear stochastic differential equation system. Moreover, given that the elements of the coefficient matrix and vector are time-varying step functions enables us to derive an exact discrete model along the lines of Robinson (2009); see also Bergstrom (1983) and Chambers (1999) for the constant coefficient case. The resulting exact discrete model is given by

$$y_{th+h} = \Theta_{th} y_{th} + \Upsilon_{1,th+h} C_{th} + \Upsilon_{2,th+h} \Delta_{th} + \xi_{th+h}, \quad t = 1, \ldots, N-1, \tag{A10}$$

where Θ_{th}, $\Upsilon_{1,th+h}$, $\Upsilon_{2,th+h}$ and ξ_{th+h} are as defined in the Theorem. The covariance properties of ξ_{th+h} discussed following Theorem 1 follow from the expression for ξ_{th+h} given there. □

Proof of Proposition 1. We begin by examining the difference

$$\begin{aligned}
\xi^a_{th+h} - \xi_{th+h} &= \left(y(th+h) - y(th) - \int_{th}^{th+h} f\big(y(r)\big)dr \right) \\
&\quad -\big(y(th+h) - \Theta_{th} y(th) - \Upsilon_{1,th+h} C_{th} - \Upsilon_{2,th+h}\Delta_{th}\big) \\
&= \left(\Theta_{th} - I_n\right) y(th) + \Upsilon_{1,th+h} C_{th} + \Upsilon_{2,th+h}\Delta_{th} - \int_0^h f\big(y(th+h-s)\big)ds
\end{aligned}$$

where a change of variable to $s = th + h - r$ has resulted in the final expression for the integral involving $f(\cdot)$. Define the following matrices:

$$P_{1,th} = \int_0^h s e^{A_{th}s} ds, \quad P_{2,th} = \int_0^h e^{A_{th}s} ds.$$

Then, by the same change of variable as above, we obtain

$$\begin{aligned}
\Upsilon_{1,th+h} &= \int_{th}^{th+h} e^{A_{th}(th+h-r)} r dr = \int_0^h e^{A_{th}s}(th+h-s)ds = (th+h)P_{2,th} - P_{1,th}, \\
\Upsilon_{2,th+h} &= \int_{th}^{th+h} e^{A_{th}(th+h-r)} dr = \int_0^h e^{A_{th}s} ds = P_{2,th}.
\end{aligned}$$

Combining the above and using the definition of Δ_{th} we then find that

$$\begin{aligned}
&\Upsilon_{1,th+h} C_{th} + \Upsilon_{2,th+h}\Delta_{th} \\
&\quad = \big((th+h)P_{2,th} - P_{1,th}\big) C_{th} + P_{2,th}\Big(f\big(y(th)\big) - A_{th} y(th) - C_{th} th \Big) \\
&\quad = \left(hP_{2,th} - P_{1,th}\right) C_{th} + P_{2,th}\Big(f\big(y(th)\big) - A_{th} y(th) \Big).
\end{aligned}$$

Hence, recalling that $\Theta_{th} = e^{A_{th}h}$, we find that

$$\begin{aligned}
\xi^a_{th+h} - \xi_{th+h} &= \big(e^{A_{th}h} - I_n\big) y(th) + \left(hP_{2,th} - P_{1,th}\right) C_{th} \\
&\quad + P_{2,th}\Big(f\big(y(th)\big) - A_{th} y(th) \Big) - \int_0^h f\big(y(th+h-s)\big)ds \\
&= \big(e^{A_{th}h} - I_n - P_{2,th}A_{th}\big) y(th) + \left(hP_{2,th} - P_{1,th}\right) C_{th} \\
&\quad + P_{2,th} f\big(y(th)\big) - \int_0^h f\big(y(th+h-s)\big)ds. \qquad \text{(A11)}
\end{aligned}$$

In what follows we shall make use of the following expansions:

$$
\begin{aligned}
P_{1,th} = \int_0^h s e^{A_{th} s} ds = \int_0^h s \left(\sum_{j=0}^{\infty} \frac{(A_{th} s)^j}{j!} \right) ds &= \sum_{j=0}^{\infty} \left(\int_0^h s^{j+1} ds \right) \frac{A_{th}^j}{j!} \\
&= \sum_{j=0}^{\infty} \frac{h^{j+2} A_{th}^j}{(j+2) j!} \\
&= \sum_{j=2}^{\infty} \frac{h^j A_{th}^{j-2}}{j(j-2)!}, \\
P_{2,th} = \int_0^h e^{A_{th} s} ds = \int_0^h \left(\sum_{j=0}^{\infty} \frac{(A_{th} s)^j}{j!} \right) ds &= \sum_{j=0}^{\infty} \left(\int_0^h s^j ds \right) \frac{A_{th}^j}{j!} \\
&= \sum_{j=0}^{\infty} \frac{h^{j+1} A_{th}^j}{(j+1) j!} \\
&= \sum_{j=1}^{\infty} \frac{h^j A_{th}^{j-1}}{j!}.
\end{aligned}
$$

Using the second result we find that the first term in (A11) is zero because

$$
e^{A_{th} h} - I_n - P_{2,th} A_{th} = I_n + \sum_{j=1}^{\infty} \frac{h^j A_{th}^j}{j!} - I_n - \left(\sum_{j=1}^{\infty} \frac{h^j A_{th}^{j-1}}{j!} \right) A_{th} = 0;
$$

this is a generalisation of the result that $\int_0^h e^{As} ds = A^{-1}(e^{hA} - I_n)$ (provided A is nonsingular), which can be written in the form $e^{hA} - I_n - \int_0^h e^{As} ds A$ using the commutability of e^A and A. Hence

$$
\begin{aligned}
E_{th} \left\| \xi_{th+h}^a - \xi_{th+h} \right\| &\leq \left\| (h P_{2,th} - P_{1,th}) C_{th} \right\| + \left\| P_{2,th} f\big(y(th)\big) \right\| \\
&\quad + E_{th} \left\| \int_0^h f\big(y(th+h-s)\big) ds \right\| \\
&\leq \left\| h P_{2,th} - P_{1,th} \right\| \left\| C_{th} \right\| + \left\| P_{2,th} \right\| \left\| f\big(y(th)\big) \right\| \\
&\quad + \int_0^h E_{th} \left\| f\big(y(th+h-s)\big) \right\| ds. \qquad \text{(A12)}
\end{aligned}
$$

From the series expansions of $P_{1,th}$ and $P_{2,th}$ it is clear that $P_{2,th} = O(h)$ while

$$\begin{aligned} hP_{2,th} - P_{1,th} &= h\sum_{j=1}^{\infty}\frac{h^j A_{th}^{j-1}}{j!} - \sum_{j=2}^{\infty}\frac{h^j A_{th}^{j-2}}{j(j-2)!} \\ &= h\left(hI_n + \frac{h^2 A_{th}}{2} + \dots\right) - \left(\frac{h^2}{2}I_n + \frac{h^3 A_{th}}{3} + \dots\right) = \frac{h^2}{2}I_n + O(h^3) \end{aligned}$$

and so this matrix is $O(h^2)$. It then follows that

$$E_{th}\left\|\xi^a_{th+h} - \xi_{th+h}\right\| \leq O(h) + O(h^2) + O(h) = O(h)$$

under the conditions stated in the Proposition. $\square$

Proof of Theorem 2. The proof starts with (A10) in which we choose a value $s \in (th, th+h]$ so that, using Assumption 5,

$$\begin{aligned} y(s) &= e^{h\bar{A}_{th}}y(s-h) + \int_{s-h}^{s} e^{\bar{A}_{th}(s-r)}rdr\bar{C}_{th} \\ &\quad + \int_{s-h}^{s} e^{\bar{A}_{th}(s-r)}dr\bar{\Delta}_{th} + \int_{s-h}^{s} e^{\bar{A}_{th}(s-r)}\bar{\Omega}_{th}dW(r). \end{aligned} \tag{A13}$$

Integrating (A13) over $s \in [th, th+h)$ yields

$$\begin{aligned} \int_{th}^{th+h} y(s)ds &= e^{h\bar{A}_{th}}\int_{th}^{th+h} y(s-h)ds + \int_{th}^{th+h}\left(\int_{s-h}^{s} e^{\bar{A}_{th}(s-r)}rdr\right)ds\bar{C}_{th} \\ &+ \int_{th}^{th+h}\left(\int_{s-h}^{s} e^{\bar{A}_{th}(s-r)}dr\right)ds\bar{\Delta}_{th} + \int_{th}^{th+h}\left(\int_{s-h}^{s} e^{\bar{A}_{th}(s-r)}ds\right)\bar{\Omega}_{th}dW(r), \end{aligned}$$

which is of the form

$$\bar{y}_{th+h} = \bar{\Theta}_{th}\bar{y}_{th} + \bar{\Upsilon}_{1,th+h}\bar{C}_{th} + \bar{\Upsilon}_{2,th+h}\bar{\Delta}_{th} + \bar{\xi}_{th+h},$$

where $\bar{\Theta}_{th} = e^{h\bar{A}_{th}}$,

$$\begin{aligned} \bar{\Upsilon}_{1,th+h} &= \int_{th}^{th+h}\left(\int_{s-h}^{s} e^{\bar{A}_{th}(s-r)}rdr\right)ds, \\ \bar{\Upsilon}_{2,th+h} &= \int_{th}^{th+h}\left(\int_{s-h}^{s} e^{\bar{A}_{th}(s-r)}dr\right)ds, \\ \bar{\xi}_{th+h} &= \int_{th}^{th+h}\left(\int_{s-h}^{s} e^{\bar{A}_{th}(s-r)}ds\right)\bar{\Omega}_{th}dW(r). \end{aligned}$$

The double integrals can be simplified to yield the expressions in the Theorem. For example, we can re-write $\bar{\Upsilon}_{1,th+h}$ in the form

$$\bar{\Upsilon}_{1,th+h} = \int_{th}^{th+h} \left(\int_{r}^{th+h} e^{\bar{A}_{th}(s-r)} ds \right) rdr + \int_{th-h}^{th} \left(\int_{th}^{r+h} e^{\bar{A}_{th}(s-r)} ds \right) rdr.$$

But, by a simple change of variable,

$$\begin{aligned} \int_{r}^{th+h} e^{\bar{A}_{th}(s-r)} ds &= \int_{0}^{th+h-r} e^{\bar{A}_{th}w} dw = \bar{K}_{1,th}(th+h-r), \\ \int_{th}^{r+h} e^{\bar{A}_{th}(s-r)} ds &= \int_{th-r}^{h} e^{\bar{A}_{th}w} dw = \bar{K}_{2,th}(th-r), \end{aligned}$$

where $\bar{K}_{1,th}(r)$ and $\bar{K}_{2,th}(r)$ are defined in the Theorem; this yields the stated expression for $\bar{\Upsilon}_{1,th+h}$. Similar arguments apply to $\bar{\Upsilon}_{2,th+h}$ and $\bar{\xi}_{th+h}$; in the latter case we obtain

$$\bar{\xi}_{th+h} = \int_{th}^{th+h} \bar{K}_{1,th}(th+h-r)\bar{\Omega}_{th} dW(r) + \int_{th-h}^{th} \bar{K}_{2,th}(th-r)\bar{\Omega}_{th} dW(r).$$

The mean and autocovariance properties follow directly from this expression. □

Proof of Proposition 2. The proof of Proposition 2 follows in a similar way to that of Proposition 1. Using Theorem 2 and (17) we find that

$$\begin{aligned} \bar{\xi}^{a}_{th+h} - \bar{\xi}_{th+h} &= \left(\bar{y}_{th+h} - \bar{y}_{th} - \bar{F}_{th+h}\right) - \left(\bar{y}_{th+h} - \bar{\Theta}_{th}\bar{y}_{th} - \bar{\Upsilon}_{1,th+h}\bar{C}_{th} - \bar{\Upsilon}_{2,th+h}\bar{\Delta}_{th}\right) \\ &= \left(\bar{\Theta}_{th} - I_{n}\right)\bar{y}_{th} + \bar{\Upsilon}_{1,th+h}\bar{C}_{th} + \bar{\Upsilon}_{2,th+h}\bar{\Delta}_{th} - \bar{F}_{th+h}. \end{aligned}$$

Using the definitions of $\bar{K}_{1,th}(r)$ and $\bar{K}_{2,th}(r)$ in Theorem 2 we obtain

$$\bar{\Upsilon}_{2,th+h} = \int_{th}^{th+h} \int_{0}^{th+h-r} e^{\bar{A}_{th}s} ds dr + \int_{th-h}^{th} \int_{th-r}^{h} e^{\bar{A}_{th}s} ds dr.$$

Using the change of variable $w = th + h - r$ in the first integral and $w = th - r$ in the second it follows that

$$\begin{aligned} \bar{\Upsilon}_{2,th+h} &= \int_{0}^{h} \int_{0}^{w} e^{\bar{A}_{th}s} ds dw + \int_{0}^{h} \int_{w}^{h} e^{\bar{A}_{th}s} ds dw \\ &= \int_{0}^{h} \int_{0}^{h} e^{\bar{A}_{th}s} ds dw \\ &= h\bar{P}_{2,th} \end{aligned}$$

where

$$\bar{P}_{2,th} = \int_{0}^{h} e^{\bar{A}_{th}s} ds.$$

Applying similar reasoning to $\bar{\Upsilon}_{1,th+h}$ yields

$$\begin{aligned}
\bar{\Upsilon}_{1,th+h} &= \int_0^h \int_0^w e^{\bar{A}_{th}s} ds(th+h-w)dw + \int_0^h \int_w^h e^{\bar{A}_{th}s} ds(th-w)dw \\
&= (th+h)\int_0^h \int_0^w e^{\bar{A}_{th}s} dsdw - \int_0^h \int_0^w e^{\bar{A}_{th}s} dswdw \\
&\quad +th\int_0^h \int_w^h e^{\bar{A}_{th}s} dsdw - \int_0^h \int_w^h e^{\bar{A}_{th}s} dswdw \\
&= th\bar{\Upsilon}_{2,th+h} + h\int_0^h \int_0^w e^{\bar{A}_{th}s} dsdw - \int_0^h \int_0^h e^{\bar{A}_{th}s} dswdw.
\end{aligned}$$

Combining these expressions for $\bar{\Upsilon}_{1,th+h}$ and $\bar{\Upsilon}_{2,th+h}$ with $\bar{C}_{th}$ and the definition of $\bar{\Delta}_{th}$ in Theorem 2 gives

$$\begin{aligned}
&\bar{\Upsilon}_{1,th+h}\bar{C}_{th} + \bar{\Upsilon}_{2,th+h}\bar{\Delta}_{th} \\
&= \left(th\bar{\Upsilon}_{2,th+h} + h\int_0^h \int_0^w e^{\bar{A}_{th}s} dsdw - \int_0^h \int_0^h e^{\bar{A}_{th}s} dswdw \right) \bar{C}_{th} \\
&\quad +\bar{\Upsilon}_{2,th+h}\left(f(\bar{y}_{th}) - \bar{A}_{th}\bar{y}_{th} - \bar{C}_{th}th \right) \\
&= \left(h\int_0^h \int_0^w e^{\bar{A}_{th}s} dsdw - \int_0^h \int_0^h e^{\bar{A}_{th}s} dswdw \right) \bar{C}_{th} \\
&\quad +h\bar{P}_{2,th}\left(f(\bar{y}_{th}) - \bar{A}_{th}\bar{y}_{th} \right).
\end{aligned}$$

Turning to $\bar{F}_{th+h}$ it is possible to simplify the double integral as follows:

$$\begin{aligned}
\bar{F}_{th+h} &= \int_{th-h}^{th} \int_{th}^{r+h} f\big(y(r)\big) dsdr + \int_{th}^{th+h} \int_{r}^{th+h} f\big(y(r)\big) dsdr \\
&= \int_{th-h}^{th} (r+h-th) f\big(y(r)\big) dr + \int_{th}^{th+h} (th+h-r) f\big(y(r)\big) dr \\
&= \int_0^h (h-u) f\big(y(th-u)\big) du + \int_0^h u f\big(y(th+h-u)\big) du
\end{aligned}$$

where the final integrals are obtained by the changes of variable to $u = th - r$ and

$u = th + h - r$, respectively. Using these results we obtain

$$\begin{aligned}
\bar{\xi}^a_{th+h} - \bar{\xi}_{th+h} &= \left(\bar{\Theta}_{th} - I_n - h\bar{P}_{2,th}\bar{A}_{th}\right)\bar{y}_{th} + h\bar{P}_{2,th} f(\bar{y}_{th}) \\
&\quad + \int_0^h \left(h \int_0^w e^{\bar{A}_{th}s} ds - w \int_0^h e^{\bar{A}_{th}s} ds \right) dw\, \bar{C}_{th} \\
&\quad + \int_0^h (h-u) f\big(y(th-u)\big) du + \int_0^h u f\big(y(th+h-u)\big) du
\end{aligned}$$

and so the quantity of interest satisfies

$$\begin{aligned}
E_{th}\left\|\bar{\xi}^a_{th+h} - \bar{\xi}_{th+h}\right\| &\leq \left\|\bar{\Theta}_{th} - I_n - h\bar{P}_{2,th}\bar{A}_{th}\right\| \|\bar{y}_{th}\| + h\left\|\bar{P}_{2,th}\right\| \|f(\bar{y}_{th})\| \\
&\quad + \int_0^h \left\| h \int_0^w e^{\bar{A}_{th}s} ds - w \int_0^h e^{\bar{A}_{th}s} ds \right\| dw \left\|\bar{C}_{th}\right\| \\
&\quad + h \int_0^h \left\| f\big(y(th-u)\big) \right\| du + h \int_0^h \left\| f\big(y(th+h-u)\big) \right\| du.
\end{aligned}$$

The first term involves the matrix

$$\begin{aligned}
\bar{\Theta}_{th} - I_n - h\bar{P}_{2,th}\bar{A}_{th} &= e^{\bar{A}_{th}h} - I_n - h\sum_{j=1}^{\infty} \frac{h^j \bar{A}_{th}^{j-1}}{j!} \bar{A}_{th} \\
&= \sum_{j=1}^{\infty} \frac{h^j \bar{A}_{th}^{j}}{j!} - h\sum_{j=1}^{\infty} \frac{h^j \bar{A}_{th}^{j}}{j!} \\
&= (1-h)\sum_{j=1}^{\infty} \frac{h^j \bar{A}_{th}^{j}}{j!} = O(h)
\end{aligned}$$

while the second term involves

$$h\bar{P}_{2,th} = h\sum_{j=1}^{\infty} \frac{h^j \bar{A}_{th}^{j-1}}{j!} = O(h^2).$$

In the third term, consider

$$\begin{aligned}
\left\| h \int_0^w e^{\bar{A}_{th}s} ds - w \int_0^h e^{\bar{A}_{th}s} ds \right\| &\leq h \left\| \int_0^w e^{\bar{A}_{th}s} ds \right\| + w \left\| \int_0^h e^{\bar{A}_{th}s} ds \right\| \\
&\leq h \int_0^w \left\| e^{\bar{A}_{th}s} \right\| ds + w \int_0^h \left\| e^{\bar{A}_{th}s} \right\| ds \\
&\leq (h+w) \int_0^h \left\| e^{\bar{A}_{th}s} \right\| ds.
\end{aligned}$$

Then

$$\begin{aligned}
\int_0^h \left\| h\int_0^w e^{\bar{A}_{th}s}ds - w\int_0^h e^{\bar{A}_{th}s}ds \right\| dw &\leq \int_0^h (h+w)\int_0^h \left\| e^{\bar{A}_{th}s} \right\| dsdw \\
&= h^2 \int_0^h \left\| e^{\bar{A}_{th}s} \right\| ds + \frac{h^2}{2}\int_0^h \left\| e^{\bar{A}_{th}s} \right\| ds \\
&= \frac{3h^2}{2}\int_0^h \left\| e^{\bar{A}_{th}s} \right\| ds = O(h^3).
\end{aligned}$$

Hence, given the stated assumptions and noting that $\|\bar{y}_{th}\| = O(h)$, we have

$$E_{th}\left\|\bar{\xi}^a_{th+h} - \bar{\xi}_{th+h}\right\| \leq O(h^2) + O(h^2) + O(h^3) + O(h^2) + O(h^2) = O(h^2)$$

as claimed. □

Proof of Theorem 3. Consider, first, the integral of $y(t) = [y^s(t)', y^f(t)']'$, given by

$$\int_{th}^{th+h} y(r)dr = \begin{pmatrix} \tilde{w}^s_{th+h} \\ \tilde{y}^f_{th+h} \end{pmatrix},$$

where $\tilde{y}^f_{th+h}$ is the observable flow component and $\tilde{w}^s_{th+h}$ denotes the unobservable integral of the stock variables. From the form of the LEDM for flows in Theorem 2 we can evaluate its components at the observed vector $\tilde{y}^\dagger_{th}$ using Assumption 6 to obtain

$$\begin{pmatrix} \tilde{w}^s_{th+h} \\ \tilde{y}^f_{th+h} \end{pmatrix} = \begin{pmatrix} \tilde{\Theta}^{ss}_{th} & \tilde{\Theta}^{sf}_{th} \\ \tilde{\Theta}^{fs}_{th} & \tilde{\Theta}^{ff}_{th} \end{pmatrix} \begin{pmatrix} \tilde{w}^s_{th} \\ \tilde{y}^f_{th} \end{pmatrix} + \begin{pmatrix} \tilde{c}^s_{th+h} \\ \tilde{c}^f_{th+h} \end{pmatrix} + \begin{pmatrix} \tilde{\xi}^s_{th+h} \\ \tilde{\xi}^f_{th+h} \end{pmatrix}, \tag{A14}$$

where

$$\tilde{\xi}_{th+h} = \int_{th}^{th+h} \tilde{K}_{1,th}(th+h-r)\tilde{\Omega}_{th}dW(r) + \int_{th-h}^{th} \tilde{K}_{2,th}(th-r)\tilde{\Omega}_{th}dW(r)$$

and the other components are defined in the Theorem. Partitioning (A9), integrating over $s \in [th, th+h)$ and using Assumption 6 yields

$$\begin{pmatrix} \tilde{y}^s_{th+h} \\ \tilde{w}^f_{th+h} \end{pmatrix} = \begin{pmatrix} \tilde{A}^{ss}_{th} & \tilde{A}^{sf}_{th} \\ \tilde{A}^{fs}_{th} & \tilde{A}^{ff}_{th} \end{pmatrix} \begin{pmatrix} \tilde{w}^s_{th+h} \\ \tilde{y}^f_{th+h} \end{pmatrix} + \begin{pmatrix} \tilde{g}^s_{th+h} \\ \tilde{g}^f_{th+h} \end{pmatrix} + \begin{pmatrix} \tilde{\eta}^s_{th+h} \\ \tilde{\eta}^f_{th+h} \end{pmatrix}, \tag{A15}$$

where $\tilde{w}^f_{th+h} = y^f(th+h) - y^f(th)$ is unobservable, $\tilde{g}_{th+h}$ is defined in the Theorem and

$$\tilde{\eta}_{th+h} = \tilde{\Omega}_{th}\int_{th}^{th+h} dW(r).$$

The objective is to eliminate the unobservable components from these representations that link observables to unobservables. From the first n^s equations of (A15) we obtain

$$\tilde{y}^s_{th+h} = \tilde{A}^{ss}_{th}\tilde{w}^s_{th+h} + \tilde{A}^{sf}_{th}\tilde{y}^f_{th+h} + \tilde{g}^s_{th+h} + \tilde{\eta}^s_{th+h}$$

which can be solved, under Assumption 7, for $\tilde{w}^s_{th+h}$:

$$\tilde{w}^s_{th+h} = \left(\tilde{A}^{ss}_{th}\right)^{-1}\left(\tilde{y}^s_{th+h} - \tilde{A}^{sf}_{th}\tilde{y}^f_{th+h} - \tilde{g}^s_{th+h} - \tilde{\eta}^s_{th+h}\right). \tag{A16}$$

This equation can be used to substitute for $\tilde{w}^s_{th+h}$ and $\tilde{w}^s_{th}$ in the first n^s equations of (A14) to give

$$\begin{aligned}
&\left(\tilde{A}^{ss}_{th}\right)^{-1}\left(\tilde{y}^s_{th+h} - \tilde{A}^{sf}_{th}\tilde{y}^f_{th+h} - \tilde{g}^s_{th+h} - \tilde{\eta}^s_{th+h}\right) \\
&= \tilde{\Theta}^{ss}_{th}\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\left(\tilde{y}^s_{th} - \tilde{A}^{sf}_{th-h}\tilde{y}^f_{th} - \tilde{g}^s_{th} - \tilde{\eta}^s_{th}\right) + \tilde{\Theta}^{sf}_{th}\tilde{y}^f_{th} + \tilde{c}^s_{th+h} + \tilde{\xi}^s_{th+h}.
\end{aligned}$$

This equation can be rearranged to give

$$\begin{aligned}
\tilde{y}^s_{th+h} &= \tilde{A}^{sf}_{th}\tilde{y}^f_{th+h} + \tilde{A}^{ss}_{th}\tilde{\Theta}^{ss}_{th}\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{y}^s_{th} + \tilde{A}^{ss}_{th}\left(\tilde{\Theta}^{sf}_{th} - \tilde{\Theta}^{ss}_{th}\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{A}^{sf}_{th-h}\right)\tilde{y}^f_{th} \\
&\quad + \tilde{g}^s_{th+h} + \tilde{A}^{ss}_{th}\tilde{c}^s_{th+h} - \tilde{A}^{ss}_{th}\tilde{\Theta}^{ss}_{th}\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{g}^s_{th} \\
&\quad + \tilde{\eta}^s_{th+h} + \tilde{A}^{ss}_{th}\tilde{\xi}^s_{th+h} - \tilde{A}^{ss}_{th}\tilde{\Theta}^{ss}_{th}\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{\eta}^s_{th}.
\end{aligned} \tag{A17}$$

Now substitute the lag of (A16) into the last n^f equations of (A14) to give

$$\begin{aligned}
\tilde{y}^f_{th+h} &= \tilde{\Theta}^{fs}_{th}\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{y}^s_{th} + \left(\tilde{\Theta}^{ff}_{th} - \tilde{\Theta}^{fs}_{th}\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{A}^{sf}_{th-h}\right)\tilde{y}^f_{th} \\
&\quad + \tilde{c}^f_{th+h} - \tilde{\Theta}^{fs}_{th}\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{g}^s_{th} + \tilde{\xi}^f_{th+h} - \tilde{\Theta}^{fs}_{th}\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{\eta}^s_{th} \\
&= \tilde{\Phi}^{fs}_{th}\tilde{y}^s_{th} + \tilde{\Phi}^{ff}_{th}\tilde{y}^f_{th} + \tilde{\gamma}^f_{th+h} + \tilde{\lambda}^f_{th+h},
\end{aligned} \tag{A18}$$

where

$$\tilde{\lambda}^f_{th+h} = \tilde{\xi}^f_{th+h} - \tilde{\Theta}^{fs}_{th}\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{\eta}^s_{th}$$

and all other terms are defined in the Theorem. Then, substituting (A18) for $\tilde{y}^f_{th+h}$

on the right-hand-side of (A17) yields

$$\begin{aligned}
\tilde{y}^s_{th+h} &= \left(\tilde{A}^{ss}_{th}\tilde{\Theta}^{ss}_{th} + \tilde{A}^{sf}_{th}\tilde{\Theta}^{fs}_{th}\right)\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{y}^s_{th} + \Bigg[\tilde{A}^{sf}_{th}\left(\tilde{\Theta}^{ff}_{th} - \tilde{\Theta}^{fs}_{th}\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{A}^{sf}_{th-h}\right) \\
&\quad + \tilde{A}^{ss}_{th}\left(\tilde{\Theta}^{sf}_{th} - \tilde{\Theta}^{ss}_{th}\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{A}^{sf}_{th-h}\right)\Bigg]\tilde{y}^f_{th} \\
&\quad + \tilde{A}^{ss}_{th}\tilde{c}^s_{th+h} + \tilde{A}^{sf}_{th}\tilde{c}^f_{th+h} + \tilde{g}^s_{th+h} - \left(\tilde{A}^{ss}_{th}\tilde{\Theta}^{ss}_{th} + \tilde{A}^{sf}_{th}\tilde{\Theta}^{fs}_{th}\right)\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{g}^s_{th} \\
&\quad + \tilde{A}^{ss}_{th}\tilde{\xi}^s_{th+h} + \tilde{A}^{sf}_{th}\tilde{\xi}^f_{th+h} + \tilde{\eta}^s_{th+h} - \left(\tilde{A}^{ss}_{th}\tilde{\Theta}^{ss}_{th} + \tilde{A}^{sf}_{th}\tilde{\Theta}^{fs}_{th}\right)\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{\eta}^s_{th} \\
&= \tilde{\Phi}^{ss}_{th}\tilde{y}^s_{th} + \tilde{\Phi}^{sf}_{th}\tilde{y}^f_{th} + \tilde{\gamma}^s_{th+h} + \tilde{\lambda}^s_{th+h}, \qquad \text{(A19)}
\end{aligned}$$

where

$$\tilde{\lambda}^s_{th+h} = \tilde{A}^{ss}_{th}\tilde{\xi}^s_{th+h} + \tilde{A}^{sf}_{th}\tilde{\xi}^f_{th+h} + \tilde{\eta}^s_{th+h} - \left(\tilde{A}^{ss}_{th}\tilde{\Theta}^{ss}_{th} + \tilde{A}^{sf}_{th}\tilde{\Theta}^{fs}_{th}\right)\left(\tilde{A}^{ss}_{th-h}\right)^{-1}\tilde{\eta}^s_{th}$$

and all other terms are defined in the Theorem. The representation for $\tilde{\lambda}_{th}$ is obtained by first stacking the expressions for $\tilde{\lambda}^s_{th+h}$ and $\tilde{\lambda}^f_{th+h}$ in a vector, so that

$$\tilde{\lambda}_{th} = \tilde{H}_{1,th}\tilde{\xi}_{th+h} + \tilde{H}_{2,th}\tilde{\eta}_{th+h} - \tilde{H}_{3,th}\tilde{\eta}_{th},$$

where $\tilde{H}_{1,th}$, $\tilde{H}_{2,th}$ and $\tilde{H}_{3,th}$ are defined in the Theorem. The expression in terms of the integrals with respect to the Wiener process is obtained by substituting the definitions of $\tilde{\xi}_{th+h}$, $\tilde{\eta}_{th+h}$ and $\tilde{\eta}_{th}$ following (A14) and (A15), respectively. □

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.

Aït-Sahalia, Y., 2002. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica* 70, 223–262.

Aït-Sahalia, Y., 2008. Closed-form likelihood expansions for multivariate diffusions. *Annals of Statistics* 36, 906–937.

Aït-Sahalia, Y., Mykland, P.A., 2004. Estimating diffusions with discretely and possibly randomly spaced data: a general theory. *Annals of Statistics* 32, 2186–2222.

Arnold, L., 1974. *Stochastic Differential Equations: Theory and Applications.* Wiley.

Bailey, R.W., Hall, V.B., Phillips, P.C.B., 1987. A model of output, employment, capital formation and inflation. In: Gandolfo, G. and Marzano, F. (eds.), *Saggi in memoria di Victorio Marrama.* Giuffrè.

Bergstrom, A.R., 1983. Gaussian estimation of structural parameters in higher-order continuous time dynamic models. *Econometrica* 51, 117–152.

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.

Bergstrom, A.R., 1990. *Continuous Time Econometric Modelling.* Oxford University Press.

Bergstrom, A.R., Nowman, K.B., 2007. A Continuous Time Econometric Model of the United Kingdom with Stochastic Trends. Cambridge.

Brunnermeier, M.K., Sannikov, Y., 2014. A macroeconomic model with a financial sector. *American Economic Review* 103, 379–421.

Chambers, M.J., 1999. Discrete time representation of stationary and non-stationary continuous time systems. *Journal of Economic Dynamics and Control* 23, 619–639.

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., 2016. The estimation of continuous time models with mixed frequency data. *Journal of Econometrics* 193, 390–404.

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

Christensen, B.J., Neri, L., Parra-Alvarez, J.C., 2022. Estimation of continuous-time linear DSGE models from discrete-time measurements. Manuscript.

Christensen, B.J., Posch, O., Wel, M., 2016. Estimating dynamic equilibrium models using mixed frequency macro and financial data. *Journal of Econometrics* 194, 116–137.

Christiano, L.J, Eichenbaum, M., Marshall, D., 1991. The permanent income hypothesis revisited. *Econometrica 59*, 397–423.

Durmas, A., Roberts, G.O., Vilmart, G., Zygalakis, K.C., 2017. Fast Langevin based algorithm for MCMC in high dimensions. *Annals of Applied Probability* 27, 2195–2237.

Fernandez-Villaverde, J., Rubio-Ramirez, J., Santos, M.S., 2006. Convergence properties of the likelihood of computed dynamic models. *Econometrica* 74, 93–119.

Harvey, A.C., Stock, J.H., 1985. The estimation of higher-order continuous time autoregressive models. *Econometric Theory* 1, 97–117.

Karatzas, I., Shreve, S.E., 1991. *Brownian Motion and Stochastic Calculus* (Second Edition). Springer-Verlag.

Nowman, K.B., 1997. Gaussian estimation of single-factor continuous time models of the term structure of interest rates. *Journal of Finance* 52, 1695–1706.

Phillips, P.C.B., 1972. The structural estimation of a stochastic differential equation system. *Econometrica* 40, 1021–1041.

Robinson, P.M, 1976. The estimation of linear differential equations with constant coefficients. *Econometrica* 44, 751–764.

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.* Blackwell.

Robinson, P.M., 2009. On discrete sampling of time-varying continuous-time systems. *Econometric Theory* 25, 985–994.

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.

Shoji, I., Ozaki, T., 1997. Comparative study of estimation methods for continuous time stochastic processes. *Journal of Time Series Analysis* 18, 485–506.

Shoji, I., Ozaki, T., 1998. Estimation for nonlinear stochastic differential equations by a local linearization method. *Stochastic Analysis and Applications* 16, 733–752.

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

Wymer, C.R., 1993. Estimation of nonlinear continuous-time models from discrete data. In: Phillips, P.C.B. (ed.), *Models, Methods, and Applications of Econometrics: Essays in Honor of A.R. Bergstrom.* Blackwell.

Zadrozny, P., 1988. Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies. *Econometric Theory* 4, 108–24.

Table 1: Matrices and Vectors in Theorem 1

Matrix	Definition	Matrix	Definition
A_{th}	$\begin{pmatrix} A'_{1,th} \\ \vdots \\ A'_{n,th} \end{pmatrix}$	Ω_{th}	$\begin{pmatrix} \omega'_{1,th} \\ \vdots \\ \omega'_{n,th} \end{pmatrix}$
$\Upsilon_{1,th+h}$	$\int_{th}^{th+h} e^{A_{th}(th+h-r)} r dr$	$\Upsilon_{2,th+h}$	$\int_{th}^{th+h} e^{A_{th}(th+h-r)} dr$
Θ_{th}	$e^{hA_{th}}$		
Vector	**Definition**	**Vector**	**Definition**
C_{th}	$\frac{1}{2}\begin{pmatrix} \mathrm{tr}\left\{\Omega'_{th} B_{1,th} \Omega_{th}\right\} \\ \vdots \\ \mathrm{tr}\left\{\Omega'_{th} B_{n,th} \Omega_{th}\right\} \end{pmatrix}$	Δ_{th}	$f(y_{th}) - A_{th} y_{th} - C_{th} th$
$f(y)$	$\big(f_1(y), \ldots, f_n(y)\big)$		

Note: $A_{i,th}$, $B_{i,th}$, $\omega_{i,th}$ and $f_i(y)$ $(i = 1, \ldots, n)$ are defined in Assumption 4.

Table 2: Matrices and Vectors in Theorem 2

Matrix	Definition	Matrix	Definition
$\bar{A}_{th}$	$\begin{pmatrix} \bar{A}'_{1,th} \\ \vdots \\ \bar{A}'_{n,th} \end{pmatrix}$	$\bar{\Omega}_{th}$	$\begin{pmatrix} \bar{\omega}'_{1,th} \\ \vdots \\ \bar{\omega}'_{n,th} \end{pmatrix}$
$\bar{\Theta}_{th}$	$e^{h\bar{A}_{th}}$		

Matrix	Definition
$\bar{\Upsilon}_{1,th+h}$	$\int_{th}^{th+h} \bar{K}_{1,th}(th+h-r)rdr + \int_{th-h}^{th} \bar{K}_{2,th}(th-r)rdr$
$\bar{\Upsilon}_{2,th+h}$	$\int_{th}^{th+h} \bar{K}_{1,th}(th+h-r)dr + \int_{th-h}^{th} \bar{K}_{2,th}(th-r)dr$

Matrix	Definition	Matrix	Definition
$\bar{K}_{1,th}(r)$	$\int_0^r e^{s\bar{A}_{th}}ds$	$\bar{K}_{2,th}(r)$	$\int_r^h e^{s\bar{A}_{th}}ds$

Vector	Definition	Vector	Definition
$\bar{C}_{th}$	$\frac{1}{2}\begin{pmatrix} \mathrm{tr}\left\{\bar{\Omega}'_{th}\bar{B}_{1,th}\bar{\Omega}_{th}\right\} \\ \vdots \\ \mathrm{tr}\left\{\bar{\Omega}'_{th}\bar{B}_{n,th}\bar{\Omega}_{th}\right\} \end{pmatrix}$	$\bar{\Delta}_{th}$	$f(\bar{y}_{th}) - \bar{A}_{th}\bar{y}_{th} - \bar{C}_{th}th$

Note: $\bar{A}_{i,th}$, $\bar{B}_{i,th}$ and $\bar{\omega}_{i,th}$ are defined in Assumption 5 and $f(y)$ is defined in Table 1.

Table 3: Matrices in Theorem 3

Matrix	Definition	Matrix	Definition
$\tilde{A}_{th}$	$\begin{pmatrix} \tilde{A}'_{1,th} \\ \vdots \\ \tilde{A}'_{n,th} \end{pmatrix}$	$\tilde{\Omega}_{th}$	$\begin{pmatrix} \tilde{\omega}'_{1,th} \\ \vdots \\ \tilde{\omega}'_{n,th} \end{pmatrix}$
$\tilde{\Phi}^{ss}_{th}$	$\left(\tilde{A}^{sf}_{th}\tilde{\Theta}^{fs}_{th} + \tilde{A}^{ss}_{th}\tilde{\Theta}^{ss}_{th}\right)\left(\tilde{A}^{ss}_{th-h}\right)^{-1}$	$\tilde{\Phi}^{sf}_{th}$	$\tilde{A}^{sf}_{th}\tilde{\Theta}^{ff}_{th} + \tilde{A}^{ss}_{th}\tilde{\Theta}^{sf}_{th} - \tilde{\Phi}^{ss}_{th}\tilde{A}^{sf}_{th-h}$
$\tilde{\Phi}^{fs}_{th}$	$\tilde{\Theta}^{fs}_{th}\left(\tilde{A}^{ss}_{th-h}\right)^{-1}$	$\tilde{\Phi}^{ff}_{th}$	$\tilde{\Theta}^{ff}_{th} - \tilde{\Phi}^{fs}_{th}\tilde{A}^{sf}_{th-h}$
$\tilde{\Theta}_{th}$	$e^{h\tilde{A}_{th}}$		

Matrix	Definition
$\tilde{\Upsilon}_{1,th+h}$	$\int_{th}^{th+h}\tilde{K}_{1,th}(th+h-r)rdr + \int_{th-h}^{th}\tilde{K}_{2,th}(th-r)rdr$
$\tilde{\Upsilon}_{2,th+h}$	$\int_{th}^{th+h}\tilde{K}_{1,th}(th+h-r)dr + \int_{th-h}^{th}\tilde{K}_{2,th}(th-r)dr$

Matrix	Definition	Matrix	Definition
$\tilde{K}_{1,th}(r)$	$\int_0^r e^{s\tilde{A}_{th}}ds$	$\tilde{K}_{2,th}(r)$	$\int_r^h e^{s\tilde{A}_{th}}ds$
$\tilde{K}_{3,th}(r)$	$\tilde{H}_{1,th}\tilde{K}_{1,th}(r) + \tilde{H}_{2,th}$	$\tilde{K}_{4,th}(r)$	$\tilde{H}_{1,th}\tilde{K}_{2,th}(r) - \tilde{H}_{3,th}\tilde{\Omega}_{th-h}\tilde{\Omega}^{-1}_{th}$
$\tilde{H}_{1,th}$	$\begin{pmatrix} \tilde{A}^{ss}_{th} & \tilde{A}^{sf}_{th} \\ 0^{fs}_n & I^{ff}_n \end{pmatrix}$	$\tilde{H}_{2,th}$	$\begin{pmatrix} I^{ss}_n & 0^{sf}_n \\ 0^{fs}_n & 0^{ff}_n \end{pmatrix}$
$\tilde{H}_{3,th}$	$\begin{pmatrix} \tilde{\Phi}^{ss}_{th} & 0^{sf}_n \\ \tilde{\Theta}^{fs}_{th}\left(\tilde{A}^{ss}_{th-h}\right)^{-1} & 0^{ff}_n \end{pmatrix}$		

Note: $\bar{A}_{i,th}$, $\bar{B}_{i,th}$ and $\bar{\omega}_{i,th}$ are defined in Assumption 6.

Table 4: Vectors in Theorem 3

Vector	Definition	Vector	Definition
$\tilde{C}_{th}$	$\frac{1}{2}\begin{pmatrix} \mathrm{tr}\left\{\tilde{\Omega}'_{th}\tilde{B}_{1,th}\tilde{\Omega}_{th}\right\} \\ \vdots \\ \mathrm{tr}\left\{\tilde{\Omega}'_{th}\tilde{B}_{n,th}\tilde{\Omega}_{th}\right\} \end{pmatrix}$	$\tilde{\Delta}_{th}$	$f(\tilde{y}_{th}) - \tilde{A}_{th}\tilde{y}_{th} - \tilde{C}_{th}th$
$\tilde{\gamma}^s_{th+h}$	$\tilde{A}^{ss}_{th}\tilde{c}^s_{th+h} + \tilde{A}^{sf}_{th}\tilde{c}^f_{th+h} + \tilde{g}^s_{th+h} - \tilde{\Phi}^{ss}_{th}\tilde{g}^s_{th}$	$\tilde{\gamma}^f_{th+h}$	$\tilde{c}^f_{th+h} - \tilde{\Phi}^{fs}_{th}\tilde{g}^s_{th}$
$\tilde{c}_{th+h}$	$\tilde{\Upsilon}_{1,th+h}\tilde{C}_{th} + \tilde{\Upsilon}_{2,th+h}\tilde{\Delta}_{th}$	$\tilde{g}_{th}$	$\frac{h}{2}(2th+h)\tilde{C}_{th} + h\tilde{\Delta}_{th}$

Note: $A_{i,th}$, $B_{i,th}$ and $\omega_{i,th}$ $(i = 1, \ldots, n)$ are defined in Assumption 6 and $f(y)$ is defined in Table 1.

Table 5: Monte Carlo Results for DGP 1

		Case I ($T = 100,\ h = 1,\ N = 100$)		Case II ($T = 35,\ h = 1/4,\ N = 140$)		Case III ($T = 25,\ h = 1/12,\ N = 300$)	
Parameter	Value	Bias	Std.Dev.	Bias	Std.Dev.	Bias	Std.Dev.
η_1	0	-0.004	0.089	-0.003	0.071	-0.002	0.071
η_2	0	-0.007	0.075	-0.006	0.042	-0.004	0.035
κ_{11}	5	1.721×10^{-3}	1.651×10^{-3}	1.651×10^{-3}	1.351×10^{-3}	1.251×10^{-3}	1.161×10^{-3}
κ_{12}	1	-0.083	1.801×10^{-3}	-0.075	1.601×10^{-3}	-0.060	1.521×10^{-3}
κ_{21}	0	-0.079	1.541×10^{-3}	-0.067	1.101×10^{-3}	-0.032	0.871×10^{-3}
κ_{22}	10	0.560	1.971×10^{-3}	0.443	1.741×10^{-3}	0.380	1.620×10^{-3}

Note: "Bias" refers to the mean value of $\hat{\theta} - \theta_0$ across all replications, where $\hat{\theta}$ is the posterior mean estimate using a flat prior for θ and θ_0 is the true parameter vector, and "Std.Dev." denotes the standard deviation of the same quantity. The reported results are based on 10,000 Monte Carlo replications.

Table 6: Monte Carlo Results for DGP 2

		Case I ($T = 75,\ h = 1,\ N = 75$)		Case II ($T = 25,\ h = 1/4,\ N = 100$)		Case III ($T = 25,\ h = 1/12,\ N = 300$)	
Parameter	Value	Bias	Std.Dev.	Bias	Std.Dev.	Bias	Std.Dev.
κ	0.20	0.021	0.039	1.861×10^{-3}	0.034	1.131×10^{-3}	0.027
γ	0.10	0.055	0.032	1.671×10^{-3}	0.025	1.171×10^{-3}	0.019
ρ	0.03	0.011	0.047	1.871×10^{-3}	0.039	1.441×10^{-3}	0.032
σ_1	0.02	1.921×10^{-3}	0.035	1.881×10^{-3}	0.024	1.791×10^{-3}	0.019
σ_2	0.01	2.111×10^{-3}	0.047	1.901×10^{-3}	0.032	1.851×10^{-3}	0.023

Note: "Bias" refers to the mean value of $\hat{\theta} - \theta_0$ across all replications, where $\hat{\theta}$ is the posterior mean estimate using a flat prior for θ and θ_0 is the true parameter vector, and "Std.Dev." denotes the standard deviation of the same quantity. The reported results are based on 10,000 Monte Carlo replications.