UNIVERSITY *of York*

Discussion Papers in Economics

No. 16/10

Continuous Time ARMA Processes: Discrete Time Representation and Likelihood Evaluation.

Michael A. Thornton and Marcus J. Chambers

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Continuous Time ARMA Processes: Discrete Time Representation and Likelihood Evaluation

Michael A. Thornton
University of York

and

Marcus J. Chambers
University of Essex

August 2016

Abstract

This paper explores the representation and estimation of mixed continuous time ARMA (autoregressive moving average) systems of orders p, q. Taking the general case of mixed stock and flow variables, we discuss new state space and exact discrete time representations and demonstrate that the discrete time ARMA representations widely used in empirical work, based on differencing stock variables, are members of a class of observationally equivalent discrete time ARMA$(p+1, p)$ representations, which includes a more natural ARMA(p, p) representation. We compare and contrast two approaches to likelihood evaluation and computation, namely one based on an exact discrete time representation and another utilising a state space representation and the Kalman-Bucy filter.

Keywords. Continuous time; ARMA process; state space; discrete time representation.

J.E.L. classification number. C32

Acknowledgements: We would like to thank Peter Phillips, Richard Smith, Karim Abadir, Peter Burridge and participants of the Peter C.B. Phillips honorary conference at the University of York, for helpful comments. All errors remain our own.

Address for Correspondence: Michael A. Thornton, Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, England.
Tel: +44 1904 324566; fax: +44 1904 ; e-mail: michael.thornton@york.ac.uk.

1. Introduction

It seems entirely natural for economic modelling to reflect the fact that economic activity is carried out more or less continuously through time. The advantages of so doing, summarised in Bergstrom (1990), include: the efficient use of information; the accurate imposition of *a priori* restrictions or hypotheses from economic theory on the parameters of the underlying model; and, the ability to forecast the real time economic variables.

Since a continuous record of data is seldom available (and any that is may well suffer from contamination by micro-structure noise) a variety of methods have been proposed enabling the analyst to estimate the parameters of linear continuous time systems using data available at regular, discrete intervals of time. An important feature of these methods is their ability to take account of whether the data are: *skip-sampled* stock variables; *time averaged* flow variables; or, a combination of the two.

The literature has concentrated on three broad methods for evaluating the pseudo-likelihood that data were generated by a given vector of parameters. Spectral (or frequency domain) representations, as in Robinson (1976, 1993) and Phillips (1991), match the periodogram of the observed data with the theoretical spectrum implied by the parameter vector, accounting for the phenomenon of aliasing by folding frequencies ranging across the real line into the range $(-\pi, \pi]$. Time domain techniques, based around translating a stochastic differential transition equation into a stochastic difference transition equation, making use of the matrix exponential, have, however, found more popularity. Provided some linear combination of the resulting discrete time state vector is observable, perhaps subject to observation noise, the pseudo-likelihood can be evaluated using the Kalman-Bucy filter. Important contributions in this area include Harvey and Stock (1985, 1988, 1989), who consider multivariate autoregressive (AR) models that can allow for stochastic trends, and Zadrozny (1988), whose approach considers multivariate autoregressive moving average (ARMA) models that can also allow for certain types of data irregularities such as mixed observation frequencies and irregular sampling intervals.

An alternative time domain approach proceeds by solving out the unobservable elements from the discrete time state vector using lags of the observable elements and disturbances and by so doing defines a mapping from the continuous time parameters to a discrete time model with identical first- and second-order moments. The resulting exact discrete time representations of the model can then be used to evaluate the likelihood; see, in particular, the contributions in Bergstrom (1990, 1997) and Bergstrom and Nowman (2007). As discussed in Bergstrom (1985), the exact discrete model can be a computationally more efficient way to evaluate the likelihood function at a given point in the parameter space, once the set-up costs of deriving the discrete time model have been borne, taking advantage of the sparse structure of the covariance matrix of the disturbances.

Our main contribution is to derive an exact discrete time representation for a continuous time ARMA process with mixed stock-flow data. Much of the original work on the exact discrete time representation of AR models, such as Bergstrom (1983, 1986, 1997) and Chambers (1999), used a state space representation in which the state vector contained a real-time vector of the variables and its derivatives, a property exploited for mixed stock-flow processes, where representations featured the first difference of the stock variables. Continuous time processes with moving average disturbances were not considered. Recent work by Chambers and Thornton (2012), however, was based on the state space form used by Zadrozny (1988), the structure of which makes it possible to incorporate a moving average disturbance, but did not offer the same route to handle mixed processes. This paper overcomes that hurdle by augmenting the form in Chambers and Thornton (2012). We then show how to recover an exact discrete representation from the resulting stochastic difference equation, taking account of the novel feature that the dimension of the system is not an integer multiple of the number of variables under consideration. In doing so we show that the representations delivered by existing methods are not unique and rely on an arbitrary identifying assumption (the differencing of the stocks). We propose a more natural identification method, establishing for the first time the existence and uniqueness of an exact discrete time ARMA(p, p) representation for continuous time mixed stock-flow data, bringing it into line with well known results for discrete time processes subject to high orders of temporal aggregation, see Brewer (1973). In addition, our framework unifies the treatment of pure stock data, which is identical to Chambers and Thornton (2012), and pure flow data, which previously relied on double integration of the stochastic difference equation, giving an attractive alternative calculation of the covariance matrix of discrete time flow data that requires the evaluation of only one matrix exponential.

The two time domain techniques depart from a common platform and our analysis also has potential uses in state-space estimation. Since an expression for unobservable components of the state vector in terms of lagged observables and disturbances is an intermediate step in the derivation of the exact discrete time distribution, a second contribution is to provide an expression for the expectation and variance of the state vector conditional on current and lagged data. Not only does this enable recovery of the state vector, in cases this is of interest, it also facilitates the calculation of the initial conditions for a Kalman iteration, given p observations, in cases where the unconditional mean is not appropriate, such as when the data are non-stationary; see, for example, Harvey and Stock (1985). Finally, we extend the range of state space models capable of modelling continuous time ARMA process with mixed stock-flow data beyond that in Zadrozny(1988), freeing the analyst to model unobserved components in the state vector, if desired.

The paper is organised as follows. Section 2 outlines the continuous time ARMA(p, q) model and discusses different state space forms. One of these forms, chosen for computational advantage, underpins the derivation of a class of exact discrete time ARMA representations for mixed stock-flow data observed at an arbitrary frequency in section 3. The discrete time ARMA(p, p) representation within this class forms the basis of discussion of time domain methods to evaluate the likelihood in section 4. Section 5 concludes, and an Appendix contains proofs of the results stated in the main body of the paper.

2. State space representations of linear continuous time mixed stock-flow processes

The continuous time ARMA(p, q) model for the $n \times 1$ vector $x(t)$ is given by

$$D^p x(t) = a_0 + A_{p-1} D^{p-1} x(t) + \ldots + A_0 x(t) + u(t) + \Theta_1 D u(t) + \ldots + \Theta_q D^q u(t), \ t > 0, \quad (1)$$

where D denotes the mean square differential operator, $u(t)$ is an $n \times 1$ continuous time white noise process, a_0 is an $n \times 1$ vector and $A_0, \ldots, A_{p-1}$ and $\Theta_1, \ldots, \Theta_q$ are $n \times n$ matrices of coefficients. More precisely, if $x(t)$ is mean square differentiable then there exists a process $\xi(t)$ satisfying

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

in which case $Dx(t) = \xi(t)$. In addition the white noise process $u(t)$ satisfies $E[u(t)] = 0$ and, for $t_2 > t_1$, has autocovariance properties

$$E\left[\int_{t_1}^{t_2} u(r)dr \int_{t_1}^{t_2} u(s)'ds\right] = \Sigma\left(t_2 - t_1\right),$$

$$E\left[\int_{t_1}^{t_2} u(r)dr \int_{t_1}^{t_2} u(\tau+s)'ds\right] = 0, \quad |\tau| > t_2 - t_1.$$

The task is to estimate the matrices $A_0, \ldots, A_{p-1}$ and $\Theta_1, \ldots, \Theta_q$ and the vector a_0 of unknown, but assumed finite, coefficients, plus the variance matrix Σ, not from a continuous record, but from a sequence of data observed at or over discrete intervals of time.

In economic applications, the difficulty of this task is often compounded by an unavoidable heterogeneity in the way that the elements of $x(t)$ are observed: data on prices, interest rates or capital stocks are typically observed at a point in time; whereas measures of quantity, activity or accumulation are observed as aggregates over time. Many economic models feature both types of variable. Without loss of generality, we partition the vector of interest as

$$x(t) = \begin{bmatrix} x^s(t) \\ x^f(t) \end{bmatrix},$$

where $x^s(t)$ $(n^s \times 1)$ contains stock variables, $x^f(t)$ $(n^f \times 1)$ contains flow variables, and $n^s + n^f = n$. A pure stock (flow) process results when n^f (n^s) is zero. We suppose our data, observed at intervals of length $h > 0$, take the form

$$x_{th} = \begin{bmatrix} x^f_{th} \\ x^s_{th} \end{bmatrix} = \begin{bmatrix} h^{-1}\displaystyle\int_{th-h}^{th} x^f(r)dr \\ x^s(th) \end{bmatrix}, \quad t = 0, 1, 2, \ldots, T.$$

The flow data[1] are represented as time averages, corresponding to, say, the scaling of a quarterly figure (when $h = 1/4$) as an annualised equivalent. This scaling is of minimal consequence in stationary systems and a time aggregate is as easily constructed. The effect is to make the scale of the flow variables comparable in models estimated at different frequencies.[2] However, this scaling has been shown to be important in models containing nonstationary and cointegrated variables by Chambers (2009, 2011) and so we proceed using the scaled observations on flow variables.

Time domain methods involve translating (1) into a first order stochastic differential equation in a state vector, linear combinations of which correspond to the discretely observed data after integration. Estimation proceeds either using the Kalman-Bucy filter or

[1] Flows are brought to the top of the observed vector to ease exposition of the partitioned matrices to follow.

[2] The measurement of stock variables is often independent of sampling frequency although some stock variables, such as interest rates, may also need expressing as a particular rate of return.

by removal of unobservable (combinations of) components to reveal the discrete time dynamics. As with linear processes in discrete time, there is more than one viable state space representation of (1), with some particularly suited to moving average errors and others to modelling mixed stock-flow data.

Much of the work on linear continuous time systems in Econometrics, such as Bergstrom (1983, 1986), Harvey and Stock (1989) and Chambers (1999), used the representation

$$Dy^b(t) = \bar{a} + A^b y^b(t) + \Theta^b u^b(t), \quad t > 0. \tag{2}$$

where

$$\bar{a} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ a_0 \end{bmatrix}, \quad A^b = \begin{bmatrix} 0 & I & 0 & \dots & 0 \\ 0 & 0 & I & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \dots & I \\ A_0 & A_1 & A_2 & \dots & A_{p-1} \end{bmatrix}, \quad \Theta^b = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ I \end{bmatrix},$$

and $y^b(t) = [y_1^b(t)',\, Dy_1^b(t)',\, \dots,\, D^{p-1}y_1^b(t)']'$. For equation (1), the disturbance is defined by $u^b(t) = u(t) + \Theta_1 Du(t) + \ldots + \Theta_q D^q u(t)$. A discrete time representation is derived by manipulations of the integral

$$y^b(t) = e^{A^b t} y^b(0) + \int_0^t e^{(t-r)A^b} \left[\bar{a} + \Theta^b u^b(r)\right] dr, \quad t > 0,$$

which is unique in a mean square sense and where the matrix exponential is defined by $e^{At} = I + \sum_{j=1}^{\infty} (At)^j / j!$. Associating $x(t) = y_1^b(t)$ links (1) to (2) and places derivatives of $x(t)$ in the state vector. Bergstrom (1986) based his representation on the integral of the state vector over $(th - h, th]$, which contains the observable sub-vectors $\int_{th-h}^{th} x^f(r)dr$ and $\int_{th-h}^{th} Dx^s(r)dr = x^s(th) - x^s(th-h)$, enabling extraction of the observed variables in the form[3]

$$x_{th}^b = \begin{bmatrix} x^s(th) - x^s(th-h) \\ \displaystyle\int_{th-h}^{th} x^f(r)dr \end{bmatrix}, \quad t = 1, 2, \dots, T.$$

The evaluation of the second moments of the weighted integral of $u^b(t)$, which are key components in evaluating the likelihood, can be accomplished by methods in Chambers (1999), for the purely autoregressive case where $\Theta_1 = \ldots = \Theta_q = 0$. For CARMA processes, however, the weighted integral of $u^b(t)$ is not tractable. A modification of (2) has been used by Brockwell (2004, 2009) to model scalar CARMA processes, incorporating the moving average disturbance via the observation equation $x(th) = S^b y^b(th)$ where

$$S^b = [I,\, \Theta_1,\, \dots,\, \Theta_q],$$

and setting $u^b(t) = u(t)$. This avoids the problem of intractable derivatives in $u^b(t)$ at the expense of maintaining derivatives of the observables in the state vector. Because $x(t) \neq y_1^b(t)$ unless $\Theta_1 = \ldots = \Theta_q = 0$, however, the integral of the state vector no longer contains both stock and flow data.

[3]Note that Bergstrom normalised h to 1. Time averaging of the flow variables can accomplished by an appropriate scaling of the relevant rows and columns of selection matrices.

Chambers and Thornton (2012) utilised the state space representation in Zadrozny (1988) to derive the exact discrete time representation of a CARMA process, in which the $np \times 1$ state vector is defined as $y^c(t) = [y_1(t)', \dots, y_p(t)']'$ and with $y_1(t) = x(t)$. The state space form is based on the following set of p equations in the derivatives of the components of $y^c(t)$, given by

$$\begin{aligned} Dy_1(t) &= A_{p-1}y_1(t) + y_2(t) + \Theta_{p-1}u(t), && (3) \\ Dy_2(t) &= A_{p-2}y_1(t) + y_3(t) + \Theta_{p-2}u(t), && (4) \\ &\vdots \\ Dy_{p-1}(t) &= A_1y_1(t) + y_p(t) + \Theta_1u(t), && (5) \\ Dy_p(t) &= a_0 + A_0y_1(t) + u(t), && (6) \end{aligned}$$

in which we define $\Theta_j = 0$ for $j > q$. Combining the expressions for $Dy_1(t), \dots, Dy_p(t)$ above, the state space form can be written

$$Dy^c(t) = \bar{a} + Ay^c(t) + \Theta^c u(t), \tag{7}$$

where

$$A = \begin{bmatrix} A_{p-1} & I & 0 & \dots & 0 \\ A_{p-2} & 0 & I & \dots & 0 \\ \vdots & & & & \vdots \\ A_1 & 0 & 0 & \dots & I \\ A_0 & 0 & 0 & \dots & 0 \end{bmatrix}, \quad \Theta^c = \begin{bmatrix} \Theta_{p-1} \\ \Theta_{p-2} \\ \vdots \\ \Theta_1 \\ I \end{bmatrix},$$

and the vector $\bar{a}$ is defined following (2). Examination of equations (3)–(6) reveals that (7) incorporates a moving average error seamlessly, but that the state vector, $y^c(t)$, does not contain derivatives of $x(t)$.[4]

This paper extends the method of Chambers and Thornton (2012) to mixed sample processes. In doing so it also sheds new light on the exact discrete representation of purely autoregressive continuous time processes and offers some computational efficiencies for the modelling of pure flow data. It is well known that the effect of time aggregation on a discrete time mixed VAR(p) produces a VARMA(p, p) once the order of time aggregation becomes sufficiently large relative to the model order; see Brewer (1973). The discrete time representation for the continuous time mixed VAR(p), developed in Bergstrom (1983) and in Chambers (1999), produces an ARMA(p, p) in x^b_{th}, that is after the stock variables have been differenced. Once the stock variables are re-integrated, these representations correspond to an ARMA($p + 1, p$), which discrete time results would suggest is not the most parsimonious form. We show that the differencing of the stock variables identifies the representation among a wider class of ARMA($p + 1, p$) processes and that the more parsimonious ARMA(p, p) is also among this class.

The reasoning behind our method is very simple. We construct our system in such a way that the integration transforming it from a stochastic differential equation to a stochastic

[4] The assertion underpinning the method in Chambers and Thornton (2012) for mixed processes was therefore in error. This error does not, however, affect Corollaries 1 and 2 of Chambers and Thornton (2012) that relate to pure stock and flow processes respectively which, along with the empirical applications in the paper, remain valid.

difference equation also produces the time-averaged flows. Let $S_0 = \left[S_0^{s\prime},\, S_0^{f\prime}\right]'$ denote the $n \times np$ matrix of rank n, such that

$$S_0 y^c(t) = x(t).$$

If $x(t)$ contains only stock process then this may be considered the observation equation, but not if any flow variables are present. To cover this case, define the $n^f \times 1$ vector, $y_0(t)$, such that

$$Dy_0(t) = h^{-1} S_0^f y^c(t) = h^{-1} x^f(t),$$

which usefully implies that its integral,

$$\int_{th-h}^{th} Dy_0(r)dr = y_0(th) - y_0(th-h) = h^{-1} \int_{th-h}^{th} x^f(r)dr = x_{th}^f,$$

is the vector of observed flow variables. Now consider the augmented state vector, $\tilde{y}(t) = [y_0(t)',\, y^c(t)']'$, which satisfies the system

$$D\tilde{y}(t) = a + H\tilde{y}(t) + \Theta u(t), \tag{8}$$

where

$$a = \begin{bmatrix} 0 \\ \bar{a} \end{bmatrix}, \quad H = \begin{bmatrix} 0 & h^{-1}S_0^f \\ 0 & A \end{bmatrix}, \quad \Theta = \begin{bmatrix} 0 \\ \Theta^c \end{bmatrix}.$$

The integral of (8), conditional on $\tilde{y}(0)$, can be written

$$\tilde{y}(t) = e^{Ht}\tilde{y}(0) + \int_0^t e^{H(t-s)} \left[a + \Theta u(s)\right] ds, \quad t > 0, \tag{9}$$

from which it follows that

$$\tilde{y}(th) = c + e^{Hh}\tilde{y}(th-h) + \epsilon(th), \quad t = 1, 2, \ldots, T, \tag{10}$$

where

$$c = \left[\int_{th-h}^{th} e^{H(th-s)}ds\right] a = \left[\int_0^1 e^{Hhr}dr\right] ah, \quad \epsilon(th) = \int_{th-h}^{th} e^{H(th-s)}\Theta u(s)ds.$$

It follows from the definition of the matrix exponential and the construction of H that the first n^f columns of e^{Hh} are $[I,\, 0]'$. Subtracting $y_0(th-h)$ from both sides of (10) leaves a state vector $y(th) = [x_{th}^{f\prime},\, y^c(th)']'$, with transition equation

$$y(th) = c + Cy(th-h) + \epsilon(th), \quad t = 1, 2, \ldots, T, \tag{11}$$

where $\epsilon(th)$ is defined following (10). It can be shown (see the Appendix) that

$$C = e^{Hh} - \begin{bmatrix} I_{n^f \times n^f} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & S_0^f \Phi \\ 0 & e^{Ah} \end{bmatrix},$$

in which

$$\Phi = \int_0^1 e^{Ahs} ds = \sum_{j=1}^{\infty} (Ah)^{j-1}/j!.$$

In cases where A is non-singular $\Phi = (A)^{-1}\left[e^{Ah} - I\right]$.

By construction, H has n^f zero roots. The first n^f columns of both H and C are null, reflecting the property that while the real time flow variables are influential their aggregates are not. In the case where A may be diagonalised, $A = Q\Lambda Q^{-1}$, where Q contains the eigenvectors of A and Λ is a diagonal matrix containing the corresponding eigenvalues, we can write $H = \tilde{Q}\tilde{\Lambda}\tilde{Q}^{-1}$, where

$$\tilde{\Lambda} = \begin{bmatrix} 0 & 0 \\ 0 & \Lambda \end{bmatrix}, \ \tilde{Q} = \begin{bmatrix} I & h^{-1}S_0^f Q\Lambda^{-1} \\ 0 & Q \end{bmatrix},$$

from which it follows that we can calculate

$$e^{Hh} = \tilde{Q}e^{\tilde{\Lambda}h}\tilde{Q}^{-1} = \begin{bmatrix} I & S_0^f Q[\Lambda h]^{-1}[e^{\Lambda h} - I]Q^{-1} \\ 0 & Qe^{\Lambda h}Q^{-1} \end{bmatrix}.$$

The associated observation equation is

$$x_{th} = \begin{bmatrix} x_{th}^f \\ x_{th}^s \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & S_0^{s\prime} \end{bmatrix} y(th) \equiv S_1 y(th), \ t = 1, 2, \ldots, T.$$

The advantage of the system in (7) is that the state vector, $y(th)$, contains the observed variables as a sub-vector; it is clear that the $n \times (np+n^f)$ observation matrix $S_1 = [I, 0]$ since our infeasible observation matrix $S_0 = [I, 0]$. In this case the system is a reordered version of that in Zadrozny (1988), with our derivation showing that the discrete time transition matrix, C, may be found from a single matrix exponential, avoiding calculation of the whole np square matrix Φ.

In fact the arguments above may be applied to any suitable state space translation of (1) or of any other linear continuous time model, that can be written in the form

$$Dy^a(t) = \bar{a} + Ay^a(t) + \Theta^a u(t), \quad t > 0, \tag{12}$$

$$x(t) = S_0 y^a(t), \tag{13}$$

where $y^a(t)$ is the state vector. When the form (2) is chosen our results translate arguments used in modelling mixed cointegrated processes by Harvey and Stock (1989) to processes with moving average errors through the choice of a different selection matrix. Other representations are available and may be preferred in different contexts, for example, to populate the state vector with unobserved but meaningful components. Our aim in the next section, however, is to derive a discrete time ARMA process that is satisfied exactly by equally spaced observations of x_{th}.

3. Exact discrete ARMA representations of mixed stock-flow processes

We return to the transition equation (11), which has the simplifying advantage that selection matrices

$$S_1 = [I_n,\, 0_{n\times r}]\,, \quad S_2 = [0_{r\times n},\, I_r]\,, \tag{14}$$

where $r = n(p-1) + n^f = np - n^s$ have the property that $S_1'S_1 + S_2'S_2 = I$. Further reasons for choosing this form will be discussed in the next section.[5] As in Chambers (1999) we define the $n \times 1$ vector $x_{th} \equiv S_1 y(th)$ and the $r \times 1$ vector $w_{th} \equiv S_2 y(th)$ $(t = 1, 2, \ldots, T)$ of unobservable variables and partition (11) as

$$\begin{aligned} x_{th} &= c_1 + C_{11}x_{th-h} + C_{12}w_{th-h} + \epsilon_{1,th}, & (15)\\ w_{th} &= c_2 + C_{21}x_{th-h} + C_{22}w_{th-h} + \epsilon_{2,th}, & (16) \end{aligned}$$

where $\epsilon_{th} \equiv (\epsilon_{1,th}', \epsilon_{2,th}')' = \epsilon(th)$, $c_i = S_i c$, $C_{ij} = S_i C S_j'$ $(i, j = 1, 2)$ and the first n^f columns of both C_{11} and C_{21} are null.

The exact discrete time representation replaces the $r \times 1$ vector w_{th-h} in (15) with terms in $x_{(t-j)h}$ and $\epsilon_{(t-j)h}$. We introduce m as the number of times that equations (15) and (16) are lagged in order to perform this replacement and write the resulting system

$$\bar{M}\bar{w}_{th} = \bar{c} + N\bar{x}_{th} + e_{th}, \tag{17}$$

where $\bar{w}_{th} = [w_{th-h}', \ldots, w_{th-(m+1)h}']'$ is the vector containing the lagged unobservable variables, $\bar{x}_{th} = [x_{th-h}', \ldots, x_{th-(m+1)h}']'$ is the vector containing the lagged observable variables, $e_{th} = [\epsilon_{1,th-h}', \ldots, \epsilon_{1,th-mh}', \epsilon_{2,th-h}', \ldots, \epsilon_{2,th-mh}']'$ contains the lagged disturbances, the intercept term is $\bar{c} = \begin{bmatrix} i_m \otimes c_1 \\ i_m \otimes c_2 \end{bmatrix}$ and

$$\bar{M} = \left[\begin{array}{cccccc} 0 & -C_{12} & 0 & \ldots & 0 & 0 \\ 0 & 0 & -C_{12} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & -C_{12} \\ \hline I_r & -C_{22} & 0 & \ldots & 0 & 0 \\ 0 & I_r & -C_{22} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & I_r & -C_{22} \end{array}\right], \quad N = \left[\begin{array}{cccccc} -I_n & C_{11} & 0 & \ldots & 0 & 0 \\ 0 & -I_n & C_{11} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -I_n & C_{11} \\ \hline 0 & C_{21} & 0 & \ldots & 0 & 0 \\ 0 & 0 & C_{21} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & C_{21} \end{array}\right].$$

We further partition $N \equiv [N_1, \ldots, N_{m+1}]$, $N_j = \left[N_j^f, N_j^s\right]$ $(j = 1, \ldots, m+1)$, where N_j^s contains n^s columns, and $N_{m+1} \equiv \left[0, N_{m+1}^s\right]$.

The condition that $m \geq r/n$ ensures that (17) has at least as many equations as there are elements in $\bar{w}_{th}$. In order to solve the system we maintain the following assumptions, which correspond to the detectability and reconstructability of the state vector.

Assumption 1. The $r \times r$ matrix C_{22} is non-singular.

Assumption 2. The $n \times r$ matrix C_{12} has rank n.

Assumptions 1 and 2 are widely used in the exact discrete time representation literature

[5]To keep the stock variables at the top set $S_1 = \begin{bmatrix} 0_{n^s\times n^f} & I_{n^s} & 0_{n^s\times r} \\ I_{n^f} & 0_{n^f\times n^s} & 0_{n^f\times r} \end{bmatrix}$.

and guarantee that $\bar{M}$ has full column rank, see Lemmas A1 and A2, which is essential for Theorem 1 below. In the event that $\bar{M}$ does not have full column rank then a transformation of the system exists with the effect of reducing r, the dimension of the vector w_{th} and permitting construction of a smaller $\bar{M}$ matrix retaining full column rank, see Lemma A3. For convenience, we concentrate on the more commonly considered case that there are no redundant elements in w_{th}.

Theorem 1. *Under Assumptions 1 and 2, for $m \geq r/n$:*

(a) there exists at least one matrix R such that $R\bar{M}$ is non-singular; and,

(b) the observed vector x_{th} of mixed stock and flow variables generated by the continuous time ARMA(p, q) system (1) then satisfies the discrete time ARMA$(m+1, m)$ system

$$x_{th} = f + F_1 x_{th-h} + \ldots + F_{m+1} x_{th-(m+1)h} + \eta_{th}, \quad t = m+2, \ldots, T,$$

where $f = c_1 + C_{12}\hat{M}\bar{c}$, $F_1 = C_{11} + C_{12}\hat{M}N_1$, $F_j = C_{12}\hat{M}N_j$ $(j = 2, \ldots, m+1)$,

$$\begin{aligned}
\hat{M} &= [I_r, 0_{r\times mr}][R\bar{M}]^{-1}R = [\hat{M}_1, \hat{M}_2], \\
\hat{M}_1 &= [\hat{M}_{1,1}, \hat{M}_{1,2}, \ldots, \hat{M}_{1,m}], \\
\hat{M}_2 &= [\hat{M}_{2,1}, \hat{M}_{2,2}, \ldots, \hat{M}_{2,m}],
\end{aligned}$$

the matrices $\hat{M}_{1,i}$ and $\hat{M}_{2,i}$ $(i = 1, \ldots, m)$ being $r \times n$ and $r \times r$, respectively. Furthermore, the autocovariance matrices of η_{th} are given by

$$\Gamma_j^h = E(\eta_{th}\eta_{th-jh}') = \begin{cases} \sum_{i=j}^{m} C_i \Omega_\epsilon C_{i-j}', & j = 0, \ldots, m, \\ 0, & j > m, \end{cases}$$

where $C_0 = S_1$, $C_j = C_{12}(\hat{M}_{1,j}S_1 + \hat{M}_{2,j+1}S_2)$ $(j = 1, \ldots, m)$, and

$$\Omega_\epsilon = E(\epsilon(th)\epsilon(th)') = \int_0^1 e^{Hhr}\Theta\Sigma\Theta' e^{Hhr} h dr.$$

Theorem 1 presents a general set of ARMA processes consistent with (1). Once $m \geq r/n$ then $\bar{M}$ has full column rank (see Lemma A1) and $\bar{M}'$ is always a suitable choice of R. Alternative choices of m and of R, up to pre-multiplication by a non-singular $(m+1)r$ square matrix, would provide different exact discrete time ARMA$(m+1, m)$ representations. The form of N_{m+1} means that the coefficients on the first n^f elements of $x_{t-(m+1)}$ would always be zero. These representations are observationally equivalent in the sense of corresponding to the same causal rational transfer function, given the parameters of (1),

$$x_{th} = \sum_{j=0}^{\infty} S_1 C^j [c + \epsilon_{(t-j)h}],$$

as a result of sharing a common state space representation (11); see, for example, Hannan and Diestler (1988, p.17). Thus each representation would provide the same unconditional likelihood for a given data set.

In practice interest is likely to centre on the most parsimonious representations, when m is closest to r/n. When $x(t)$ is a pure stock process then the smallest $m \geq r/n$ is $p-1$. The matrix $\bar{M}$ is then square and invertible, and the results of Theorem 1 naturally become those of Corollary 1 of Chambers and Thornton (2012) for any non-singular R, producing an ARMA$(p, p-1)$ representation. When $x(t)$ is a pure flow process then the smallest $m \geq r/n$ is p, but again $\bar{M}$ is square and invertible. In this case the N_{p+1} matrix is null, producing an ARMA(p, p) representation equivalent to that in Corollary 2 of Chambers and Thornton (2012). It is worth noting that the evaluation of Γ_j^h simplifies existing methods that treat pure flow processes as integrals of pure stock processes. Such a treatment leads to a discrete time disturbance expressed using a double integral, requiring the evaluation of integrals of $\Phi(r) = \int_0^r e^{Ahs} ds$, something our augmented state equation avoids.

When $x(t)$ is a mixed process then p remains the smallest $m \geq r/n$. In this case, however, the matrix $\bar{M}$ has n^s fewer columns than rows. Multiplying (17) by any $(p+1)r \times p(r+n)$ linear transformation, R, such that $R\bar{M}$ is non-singular would give a possible ARMA$(p+1, p)$ representation with zero coefficients on $p+1$ lags of the flow variables, due to the n^f null columns of N_{p+1}. Different choices of R have the potential to produce different exact discrete time representations. The representations in Bergstrom (1983) and Chambers (1999) result from choosing R such that the last n^s columns of the matrix $[I - F_1 - \ldots - F_{p+1}]$ are null, which is equivalent to differencing the stock variables. This is not the most natural representation of the model, however. If R is chosen such that $RN_{p+1} = 0$ then the resulting discrete time process would have an ARMA(p, p) representation. Theorem 2 establishes the existence and uniqueness of such a discrete time representation for x_t in this case.

Theorem 2. *(a) There exists a $(p+1)r \times p(r+n)$ matrix R such that $RN_{p+1} = 0$ and $R\bar{M}$ is non-singular.*

(b) The observed vector x_{th} of mixed stock and flow variables generated by the continuous time ARMA(p, q) system (1) then satisfies a unique discrete time ARMA(p, p) with coefficients given in Theorem 1.

Since $(p-1)(r+n)$ rows of N_{p+1} are null, R may take $(p-1)(r+n)$ rows from the $p(r+n)$ identity matrix alongside rows of the form

$$\left[0_{r+n^f \times (p-1)n},\, T_1,\, 0_{r+n^f \times (p-1)n},\, T_2\right],$$

where T_1 and T_2 are chosen so that $T_1 C_{11} + T_2 C_{21} = 0$. For processes without moving averages the representation in Chambers (1999), which does not require the identification of such a space, has computational advantages in estimation. Ours may still be used, however, to translate the estimates of the parameters of a continuous time model into the most easily comparable discrete time form.

4. Evaluation of the pseudo-likelihood

We now consider the evaluation of the log-likelihood function for given values of the parameters of (1) $(A_0, \ldots, A_{p-1},\, \Theta_1, \ldots, \Theta_q,\, a_0,\, \Sigma)$ based on a sample of T observations, $x_h, x_{2h}, \ldots, x_{Th}$, using both the exact discrete representation, outlined in Theorems 1 and 2, and the Kalman-Bucy filter applied to the stochastic difference equation (11). We focus attention on estimation conditional on the first p of those observations, although both methods can be used to provide unconditional estimates, as demonstrated by Bergstrom (1990) and Harvey and Stock (1985). The techniques differ in the order of conditioning applied to the joint density of the sample, which can be decomposed as

$$f(x_{Th},\ldots,x_{(p+1)h}|x_{ph},\ldots,x_h)=\prod_{j=p+1}^{T} f(x_{jh}|x_{(j-1)h},\ldots,x_h).$$

Through the repeated conditioning stated in the expression on the right, the prediction error form of the Kalman-Bucy filter is based around uncorrelated errors resulting from optimal forecasts produced at the expense of calculating the full state vector for each observation. Gaussian estimation, on the other hand, reflects the joint distribution on the left using knowledge about the effects of time aggregation to characterise the the disturbance, η_{th}, with a sparse (auto-)covariance matrix. The two methods have much in common, however, and it is common practice in Gaussian estimation to calculate the likelihood via a normalised residual vector that is, in effect, the standardised prediction error vector.

4.1. Gaussian estimation using the exact discrete representation

Gaussian estimation uses the discrete time ARMA(p,q_d) representation, where $q_d=p-1$ for pure stock processes and $q_d=p$ if the data vector contains any flow variables, of equation (1) derived in Theorems 1 and 2. Let $\eta=(\eta'_{(p+1)h}\ldots,\eta'_{Th-h},\eta'_{Th})'$, denote the $n(T-p)$ vector of discrete time disturbances with

$$\eta_{th}=x_{th}-f-F_1x_{th-h}-\ldots-F_px_{th-ph},\quad t=p+1,\ldots,T.$$

Its covariance matrix, $E(\eta\eta')=\Omega_\eta$, has a block Toeplitz structure with ij'th block denoted by the n matrix

$$\Omega_{\eta,ij}=\begin{cases}\Gamma^h_{i-j}, & |i-j|\leq q_d,\\ 0, & |i-j|>q_d,\end{cases}$$

noting that $\Gamma^h_{-j}=\Gamma^{h\prime}_j$ with Γ^h_j defined in Theorem 1. Under the assumption that η has a multivariate normal distribution the likelihood may be evaluated as

$$\log L_G=-\frac{n(T-p)}{2}\log(2\pi)-\frac{1}{2}\log|\Omega_\eta|-\frac{1}{2}\eta'\Omega_\eta^{-1}\eta.$$

Two comments are worth making at this stage. The first is that the matrices $A_0,\ldots,A_{p-1}$ feature in both η, through the f vector and F_j matrices, and Ω_η, through the Γ^h_{i-j} matrices, a common result in temporally aggregated dynamic systems. The second is to highlight the computational advantages of the representation (7), which maintains the selection matrices S_1 and S_2 during the maximisation procedure and shows the impact of the Θ_j parameters to Ω_η via Ω_ϵ. In contrast, the form in (2) in which S_1 is a function Θ, requires repeated recalculation of the matrix S_2.

The sparse nature of Ω_η makes it possible to accelerate the calculation of this likelihood. Since Ω_η is positive definite and symmetric we can find a lower triangular matrix, U, with

ij'th block denoted (for consistency with the Kalman-Bucy filter) $U_{p+i,p+j}$, such that

$$UU' = \Omega_\eta,$$

with the sparse nature of Ω_η reflected in the sparse nature of U. The matrix U, which has a maximum of $q_d + 1$ non-zero blocks in each block row, can be calculated using the following recursions

$$\begin{aligned}
U_{p+1,p+1}U'_{p+1,p+1} &= \Gamma_0^h, \\
U_{ij} &= \left[\Gamma_{i-j}^h - \sum_{k=1}^{j-1} U_{ik}U'_{jk}\right] U_{jj}^{-1\prime},\ p+1 < i \leq p+q_d+1,\ p < j < i, \\
U_{ii}U'_{ii} &= \Gamma_0^h - \sum_{k=1}^{i-1} U_{ik}U'_{ik},\ p+1 < i \leq p+q_d+1, \\
U_{ij} &= 0,\ i > p+q_d+1, j < i-q_d, \\
U_{ij} &= \left[\Gamma_{i-j}^h - \sum_{k=i-q_d}^{j-1} U_{ik}U'_{jk}\right] U_{jj}^{-1\prime},\ i > p+q_d+1,\ q_d - i \leq p < j < i, \\
U_{ii}U'_{ii} &= \Gamma_0^h - \sum_{k=i-q_d}^{i-1} U_{ik}U'_{ik},\ i > p+q_d+1,
\end{aligned}$$

with U_{ii} chosen to be lower triangular. Only this lower triangular matrix need be inverted, firstly to find further block rows and secondly to calculate a vector of normalised residuals, $\zeta = (\zeta'_{(p+1)h} \ldots, \zeta'_{Th-h}, \zeta'_{Th})'$, satisfying $U\zeta = \eta$. It follows straightforwardly that $E(\zeta) = 0$ and $E(\zeta\zeta') = I_{n(T-p)}$, the subvectors of ζ being computed using

$$\begin{aligned}
\zeta_{(p+1)h} &= U_{p+1,p+1}^{-1}\eta_{(p+1)h}, \\
\zeta_{th} &= U_{tt}^{-1}\left(\eta_{th} - \sum_{k=1}^{\min\{t-p-1,q_d\}} U_{t,t-k}\eta_{(t-k)h}\right), \quad t = p+2, \ldots, T.
\end{aligned}$$

The log likelihood, conditional on $x_h, \ldots, x_{ph}$, can be evaluated as

$$\log L_G = -\frac{n(T-p)}{2}\log(2\pi) - \frac{1}{2}\sum_{t=p+1}^{T}\left(\zeta'_{th}\zeta_{th} + 2\log(|U_{tt}|)\right),$$

where $\log(|U_{tt}|)$ is easily calculated as the sum of the terms on the principle diagonal of U_{tt}. Calculation of U involves inverting a maximum of $T-p$ lower triangular matrices of dimension n, but in practice U_{ij} and $U_{i+1,j+1}$ often converge quickly, removing the need to calculate the rows of U further, see Bergstrom (1990, ch 7). The non-zero blocks in block row i contain coefficient matrices for a moving average representation of η_{ih}.

4.2. The Kalman prediction error

Letting $\hat{x}_{th|th-h}$ denote the predicted value of the vector x_{th} given information at time $th-h$ and $\Omega_{th|th-h} = E[\hat{x}_{th|th-h} - x_{th}][\hat{x}_{th|th-h} - x_{th}]'$ denote the variance of the prediction error, then the likelihood can be written in the form

$$
\begin{aligned}
\log L_K &= -\frac{n(T-p)}{2}\log(2\pi) - \frac{1}{2}\sum_{t=p+1}^{T}\log\left|\Omega_{th|th-h}\right| \\
&\quad -\frac{1}{2}\sum_{t=p+1}^{T}\left(x_{th}-\hat{x}_{th|th-h}\right)'\Omega_{th|th-h}^{-1}\left(x_{th}-\hat{x}_{th|th-h}\right).
\end{aligned}
$$

The one step ahead predictions and forecast variances can be calculated recursively using the Kalman-Bucy filter, with

$$
\begin{aligned}
\hat{x}_{th|th-h} &= S_1\hat{y}(th|th-h), && (18)\\
\Omega_{th|th-h} &= S_1P_{th|th-h}S_1', && (19)\\
\hat{y}(th+h|th) &= c + C\hat{y}(th|th-h) + K_{th}\left[x_{th}-\hat{x}_{th|th-h}\right], && (20)\\
K_{th} &= CP_{th|th-h}S_1'\left[\Omega_{th|th-h}\right]^{-1}, && (21)\\
P_{th+h|th} &= \left[C-K_{th}S_1\right]P_{th|th-h}\left[C-K_{th}S_1\right]' + \Omega_\epsilon, && (22)
\end{aligned}
$$

$\hat{y}(th|th-h)$ denoting the one step ahead prediction of the full state vector $y(th)$, $P_{th|th-h}$ its covariance matrix and K_{th} the Kalman gain. Computation of the likelihood eases considerably once changes in $P_{th+h|th}$ become relatively small and it no longer becomes necessary to compute equations (19), (21) or (22).

In the case that the time series are stationary, it is common to commence the Kalman iterations with the unconditional mean $\hat{y}(h|0) = [I-C]^{-1}c$ and with $P_{h|0}$ solving the discrete time Lyapunov equation $P_{h|0} = CP_{h|0}C' + \Omega_\epsilon$; see, for example, Hamilton (1994, p.378). If the analyst does not wish to evaluate the initial conditions in this way, or if the data are non-stationary, Harvey and Stock (1985) suggest estimating the initial condition of the state vector as a parameter in the model, which can be concentrated out of the likelihood function. The exact discrete representation, which concentrates w_{th} out of the dynamics of x_{th}, provides an alternative method. The expression in (A8) provides an expression for the unobserved elements of the state vector that can be used to appraise $w_{ph|ph}$ and its variance. We might then chose to commence the recursions with

$$
\begin{aligned}
\hat{y}_{ph+h|ph} &= c + C\hat{y}(ph|ph),\\
P_{ph+h|ph} &= CP_{ph|ph}C' + \Omega_\epsilon,\\
\hat{y}(ph|ph) &= (x_{ph}', \hat{w}_{ph|ph}')',\\
\hat{w}_{ph|ph} &= \hat{M}[\bar{c} + N\bar{x}_{ph+h}],
\end{aligned}
$$

with

$$
P_{ph|ph} = \left[\begin{array}{cc} 0 & 0 \\ 0 & \tilde{P}_{ph|ph}\end{array}\right], \quad \tilde{P}_{ph|ph} = \hat{M}\left[\begin{array}{cc} I\otimes\Omega_{11} & I\otimes\Omega_{12} \\ I\otimes\Omega_{21} & I\otimes\Omega_{22}\end{array}\right]\hat{M}',
$$

where $\Omega_{ij} = S_i\Omega_\epsilon S_j'$.

Comparing the two methods, it is clear that in Gaussian estimation the vector $U_{tt}\zeta_{th}$

is the part of x_{th} that is not explained by lagged values of x_{th} or of η_{th}. This naturally corresponds to the Kalman prediction error $x_{th} - \hat{x}_{th|th-h}$. If the model is correctly specified then ζ_{th} is a standard normal n vector and the variance of this error, $U_{tt}U_{tt}'$ corresponds to $\Omega_{th|th-h}$.

5. Concluding comments

This paper has explored the structure of state space representations of continuous time ARMA processes. Furthermore, it has provided exact discrete time representations for data generated by a continuous time ARMA(p, q) system in the general case where the data may be stocks, flows, or a combination of the two. We have demonstrated that, once flow variables are used, an ARMA(p, p) without differencing the stock variables is an appropriate discrete time representation, in line with the results for higher orders of discrete time aggregation. Intermediate results can be used to estimate the expectation and variance of the state vector conditional on p observations.

References

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

Bergstrom, A.R., 1985. The estimation of parameters in nonstationary higher-order continuous-time dynamic models. *Econometric Theory* 1, 369–385.

Bergstrom, A.R., 1986. The estimation of open higher-order continuous time dynamic models with mixed stock and flow data. *Econometric Theory* 2, 350–373.

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

Bergstrom, A.R., 1997. Gaussian estimation of mixed-order continuous-time dynamic models with unobservable stochastic trends from mixed stock and flow data. *Econometric Theory* 13, 467–505.

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

Brewer, K.R.W., 1973. Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models. *Journal of Econometrics* 1, 133–154.

Brockwell, P.J., 2004. Representations of continuous-time ARMA processes. *Journal of Applied Probability* 41, 375-38.

Brockwell, P.J., 2009. Lévy-driven continuous time ARMA processes. In: *Handbook of Financial Time Series*, Andersen, T.G., Davis, R.A., Kreiss, J.-P., Mikosch, Th.V., (Eds.). Springer, New York.

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., Thornton, M.A., 2012. Discrete time representations of continuous time ARMA processs. *Econometric Theory* 28, 219–238.

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

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

Harvey, A.C., Stock, J.H., 1988. Continuous time autoregressive models with common stochastic trends. *Journal of Economic Dynamics and Control* 12, 365–384.

Harvey, A.C., Stock, J.H., 1989. Estimating integrated higher-order continuous time autoregressions with an application to money-income causality. *Journal of Econometrics* 42, 313–336.

Hannan, E.J., Deistler, M., 1988. *The Statistical Theory of Linear Systems.* John Wiley & Sons, Chichester.

Kailath, T., 1980. *Linear Systems.* Prentice Hall, Englewood Cliffs, N.J.

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

Robinson, P.M., 1976. Fourier estimation of continuous time models. In A.R. Bergs-trom (ed.), *Statistical Inference in Continuous Time Economic Models*, pp. 215–266. North-Holland, Amsterdam.

Robinson, P.M., 1993. Continuous-time models in econometrics: closed and open systems, stocks and flows. In P.C.B. Phillips (ed.), *Models, Methods, and Applications of Econometrics: Essays in Honor of A.R. Bergstrom*, pp. 71–90. Blackwell, Oxford.

Szidarovsky, F., Bahill, A.T., 1991. *Linear Systems Theory.* CRC Press, Boca Raton.

Zadrozny, P., 1988. Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies. *Econometric Theory* 4, 108–124.

Appendix

The dimensions and forms of the vectors and matrices used in (17) are:

$\bar{M}$:	$m(r+n) \times (m+1)r$;	R :	$(m+1)r \times m(r+n)$;
$\hat{M}$:	$r \times m(r+n)$;	N :	$m(r+n) \times (m+1)n$;
N_j :	$m(r+n) \times n \quad (j = 1, \ldots, m+1)$;	N^s_{m+1} :	$m(r+n) \times n^s$;
$\bar{w}_t$:	$(m+1)r \times 1$;	$\bar{x}_t$:	$(m+1)n \times 1$;
e_t :	$m(r+n) \times 1$;	$\bar{c}$:	$m(r+n) \times 1$.

Partitions of matrix exponentials

The following partitions are needed to establish Theorems 1 and 2. Given the matrix A we can write, for integer j,

$$[hH]^j = \begin{bmatrix} 0 & S^f_0 \\ 0 & hA \end{bmatrix}^j = \begin{bmatrix} 0 & S^f_0(hA)^{j-1} \\ 0 & (hA)^j \end{bmatrix}.$$

Applying the formula for the matrix exponential we see that

$$e^{hH} = \begin{bmatrix} I & S^f_0\left(I + \sum_{j=2}^{\infty}(A^h)^{j-1}/j!\right) \\ 0 & I + \sum_{j=1}^{\infty}(A^h)^j/j! \end{bmatrix} = \begin{bmatrix} I & S^f_0\Phi \\ 0 & e^{Ah} \end{bmatrix},$$

where $\Phi = I + \sum_{j=2}^{\infty}(Ah)^{j-1}/(j)! = \int_0^1 e^{Ahs}ds$. Since the first n^f columns of H^j (for $j > 0$) contain only zeros, the top left n^f square of C is equal to the identity.

The following sub-matrices constitute the partitions of C used in Theorems 1 and 2.

$$C_{11} = \begin{bmatrix} 0_{n^f \times n^f} & \Phi[n^s+1\!:\!n,\, 1\!:\!n^s] \\ 0_{n^s \times n^f} & e^{Ah}[1\!:\!n^s,\, 1\!:\!n^s] \end{bmatrix} \equiv \begin{bmatrix} 0 & C^f_{11} \\ 0 & C^s_{11} \end{bmatrix},$$

$$C_{12} = \begin{bmatrix} \Phi[n^s+1\!:\!n,\, n^s+1\!:\!np] \\ e^{Ah}[1\!:\!n^s,\, n^s+1\!:\!np] \end{bmatrix} \equiv \begin{bmatrix} C^f_{12} \\ C^s_{12} \end{bmatrix},$$

$$C_{21} = \left[0_{np \times n_f},\, e^{Ah}[n^s+1\!:\!np,\, 1\!:\!n^s]\right] \equiv [0,\, C^s_{21}],$$

$$C_{22} = e^{Ah}[n^s+1\!:\!np,\, n^s+1\!:\!np],$$

where $B[a\!:\!b,\, c\!:\!d]$ denotes the sub-matrix containing rows a to b and columns c to d of the matrix B. Defining $\tilde{C}_{11} = S_1 e^{Hh} S_1'$, the matrix

$$W = \tilde{C}_{11} - C_{12}C_{22}^{-1}C_{21} = \begin{bmatrix} I_{n^f} & C^f_{11} - C^f_{12}C_{22}^{-1}C^s_{21} \\ 0_{n^s \times n^f} & C^s_{11} - C^s_{12}C_{22}^{-1}C^s_{21} \end{bmatrix} \equiv \begin{bmatrix} I_{n^f} & W^f \\ 0_{n^s \times n^f} & W^s \end{bmatrix} \tag{A1}$$

is the Schur complement of C_{22} in the matrix e^{Hh}. Since e^{Hh} is non-singular, Assumption 1 implies that both W and W^s are non-singular.

The following lemmas establish the ranks of the large partitioned matrices needed in Theorems 1 and 2. The proofs are well-known in the linear systems literature but are included for completeness. Lemma A1 establishes the necessary and sufficient conditions for the Popov-Belevitch-Hautus eigenvector tests (see Kailath, 1980, p.135), which are then deployed in Lemmas A2 and A4. Lemma A3 considers the elimination of unobservable variables in the event that assumptions 1 and 2 do not hold.

Lemma A1. *Under Assumptions 1 and 2, the nullspace of C_{12} does not contain any (non-trivial) eigenvectors of C_{22}.*

Proof. Let d be an eigenvector of C_{22} such that $C_{12}d = 0$. Then the $n+r$ vector $q = (0', d')'$ must be an eigenvector of e^H. Since this is an exponential, the associated eigenvalue, $\lambda > 0$, and $[H - \log\lambda I_{r+n}]q = 0$. Using (8), this gives

$$\begin{array}{rcll}
 & & q[n+1:n+n^f] & = q[1:n^f]\log\lambda, \\
A_{p-1}q[n^f+1:n^f+n] & + & q[n^f+n+1:n^f+2n] & = q[n^f+1:n^f+n]\log\lambda, \\
A_{p-2}q[n^f+1:n^f+n] & + & q[n^f+2n+1:n^f+3n] & = q[n^f+n+1:n^f+2n]\log\lambda, \\
\vdots & & \vdots & \quad\vdots \\
A_1q[n^f+1:n^f+n] & + & q[n^f(p-1)n+1:n^f+pn] & = q[n^f(p-2)n+1:n^f+(p-1)n+1]\log\lambda, \\
A_0q[n^f+1:n^f+n] & & & = q[n^f(p-1)n+1:n^f+pn]\log\lambda.
\end{array}$$

Since $q[1:n] = 0$ the top line implies that $q[n^f+1:n^f+n] = 0$. In the second line this implies that $q[n^f+n+1:n^f+2n] = 0$ and a recursive argument establishes that $q = 0$ and hence $d = 0$. $\square$

Lemma A2. *Under Assumptions 1 and 2, the matrix $\bar{M}$ has full column rank for $m \geq r/n$.*

Proof. Suppose there were an $(m+1)r$ non-null vector $d = [d_1', d_2', \ldots, d_{m+1}']'$ such that $\bar{M}d = 0$. Given Assumption 1, it would have to be the case that $C_{22}d_i = d_{i-1}$ and $C_{12}d_i = 0$, for $i = 2, 3, \ldots, m+1$, or equivalently, a non-null r vector d_{m+1} such that

$$\begin{bmatrix} C_{12} \\ C_{12}C_{22} \\ C_{12}C_{22}^2 \\ \vdots \\ C_{12}C_{22}^m \end{bmatrix} d_{m+1} \equiv M^* d_{m+1} = 0.$$

Hence $\bar{M}$ has full column rank provided the matrix M^* does. The matrix M^* has the form of the observability matrix of a linear system with transition matrix C_{22} and observation matrix C_{12}. It is well known from the linear systems literature – see, for example, Theorems 6.4 and 6.5 of Szidarovsky and Bahill (1991) – that Lemma A1 is necessary and sufficient for M^* to have full column rank (and for such a system to be observable), but the following is included for completeness.

Suppose that M^* has rank $k < r$. Let the $k \times r$ matrix G_1 consist of the k linearly independent rows of M^*, spanning its rowspace, and let G_2 be an $r - k \times r$ matrix spanning the rest of r space. Then the $r \times r$ matrix

$$G = \left[\begin{array}{c} G_1 \\ G_2 \end{array} \right],$$

is non-singular. Partition its inverse $G^{-1} = [\ B_1 \quad B_2\ ]$, where B_2 is $r \times r - k$, then it follows that $G_1 B_2 = 0$. Then $M^* B_2 = 0$ as G_1 spans the rowspace of M^* and B_2 spans the $r - k$ dimensional nullspace of M^*. It then follows that $C_{12} B_2 = 0$ and that $G_1 C_{22} B_2 = 0$ as space spanned by the columns of B_2 is invariant under C_{22}. We can therefore write Given that

$$V = G C_{22}^{-1} G^{-1} = \left[\begin{array}{cc} V_{11} & 0 \\ V_{12} & V_{22} \end{array} \right],\ Z = C_{12} G^{-1} = [Z_1,\, 0]\,, \tag{A2}$$

where V_{11} is $k \times k$ and Z_1 has k columns. Let $q \neq 0$ be an arbitrary $r - k$ eigenvector of V_{22} with associated eigenvalue, λ, so that the r vector $v = [0',\, q']'$ is an eigenvector of V in the nullspace of Z. Let $x = G^{-1} v \neq 0$. Then

$$C_{22}^{-1} x = G^{-1} G C_{22}^{-1} G^{-1} G x = G^{-1} V v = G^{-1} v \lambda = x \lambda, \text{ and } C_{12} x = Z v = 0.$$

As Lemma A1 shows that such no eigenvectors of C_{22}^{-1} lie in the nullspace of C_{12} then $\bar{M}$ has full column rank. $\square$

Lemma A3. *If the matrix $\bar{M}$ is of rank $k < m$ then the dimension of the unobservable state variables in* (15) *and* (16) *can be reduced to k.*

Proof Suppose M^* has rank $k < r$. Define a new unobservable variable $\tilde{w}_{th} = G w_{th}$, and rewrite (15) and (16) as

$$\begin{aligned} x_{th} &= c_1 + C_{11} x_{th-h} + C_{12} G^{-1} \tilde{w}_{th-h} + \epsilon_{1,th}, && \text{(A3)} \\ \tilde{w}_{th} &= G c_2 + G C_{21} x_{th-h} + G C_{22} G^{-1} \tilde{w}_{th-h} + G \epsilon_{2,th}, && \text{(A4)} \end{aligned}$$

Partitioning $\tilde{w}_{th} = [\tilde{w}'_{1,th} \tilde{w}'_{2,th}]'$, the expressions in (A2) mean that the columns of $C_{12} G^{-1}$ relating the $r - k$ vector $\tilde{w}'_{2,th-h}$ to x_{th} in in (A3) and the block of $G C_{22} G^{-1}$ relating it to the k vector $\tilde{w}_{1,th}$, in (A4) are zero. We may therefore write the system as

$$\begin{aligned} x_{th} &= c_1 + C_{11} x_{th-h} + Z_1 \tilde{w}_{1,th-h} + \epsilon_{1,th}, && \text{(A5)} \\ \tilde{w}_{1,th} &= G_1 c_2 + G_1 C_{21} x_{th-h} + V_{11} T_1 \tilde{w}_{1,th-h} + G_1 \epsilon_{2,th}, && \text{(A6)} \\ \tilde{w}_{2,th} &= G_2 c_2 + G_2 C_{21} x_{th-h} + V_{12} \tilde{w}_{1,th-h} + V_{22} \tilde{w}_{2,th-h} + G_2 \epsilon_{2,th}, && \text{(A7)} \end{aligned}$$

Since $\tilde{w}_{2,th}$ neither appears in, nor is not needed to remove $\tilde{w}_{1,th-h}$ from, (A5), we are free to ignore (A7)in deriving the exact discrete time representation, reducing the dimension of the unobservable vector from r, to k.

Lemma A4. *Under Assumptions 1 and 2, with $m = p$ and $0 < n^s < n$ the matrix* $\left[\bar{M},\, N^s_{p+1}\right]$ *is non-singular.*

Proof. Suppose there were an $(p+1)r$ non-null vector $d = [d_1',\, d_2',\, \ldots,\, d_{p+1}']'$ and a n^s non-null vector z such that $\bar{M}d + N^s_{p+1}z = 0$. Following the argument used in Lemma A2, it must be that $d_i = C_{22}^{-1}d_{i-1}$, $C_{12}d_i = 0$ $(i = 2, 3, \ldots, p)$, while $d_p = C_{22}d_{p+1} - C^s_{21}z$ and $C_{12}d_{p+1} - C^s_{11}z = 0$. Pre-multiplying the penultimate expression by C_{22}^{-1} and substituting in this is equivalent to a non-null pn vector, $(d_1',\, z')'$, such that

$$\begin{bmatrix} C_{12}C_{22}^{-1} & 0 \\ C_{12}C_{22}^{-2} & 0 \\ \vdots & \vdots \\ C_{12}C_{22}^{-p} & C_{12}C_{22}^{-1}C^s_{21} - C^s_{11} \end{bmatrix} \begin{bmatrix} d_1 \\ z \end{bmatrix} = 0.$$

Note that $C_{12}C_{22}^{-1}C^s_{21} - C^s_{11}$ is (-1) times the final n^s columns of W. Using (A1) the bottom block row produces

$$\begin{aligned} -W^s z = C^s_{12}C_{22}^{-p}d_1 \quad &\Rightarrow \quad z = -(W^s)^{-1}C^s_{12}C_{22}^{-p}d_1, \\ -W^f z = C^f_{12}C_{22}^{-p}d_1 \quad &\Rightarrow \quad W^f(W^s)^{-1}C^s_{12}C_{22}^{-p}d_1 = C^f_{12}C^p_{22}d_1, \end{aligned}$$

and hence an equivalent condition is the non-singularity of the $r \times r$ matrix,

$$\bar{W} = \begin{bmatrix} C_{12}C_{22}^{-1} \\ C_{12}C_{22}^{-2} \\ \vdots \\ C_{12}C_{22}^{-(p-1)} \\ W^*C_{22}^{-(p-1)} \end{bmatrix},$$

where $W^* = \left[I_{n^f},\, -W^f(W^s)^{-1}\right]C_{12}C_{22}^{-1}$ has full row rank and can be shown to contain the top n^f rows of $S_1e^{-Hh}S_2'$. The matrix $[C_{12}',\, W^{*\prime}]'$ must also have full row rank, otherwise there would exist n vectors $z_1 \neq 0$ and $z_2 \neq 0$ such that $z_1'C_{12}C_{22}^{-i} + z_2'C_{12}C_{22}^{-i-1} = 0$ and M^*_p would have rank less than r, in contradiction to Lemma A2. Identical arguments to Lemma A2 establish that the matrix

$$\begin{bmatrix} \begin{pmatrix} C_{12} \\ W^* \end{pmatrix} C_{22}^{-1} \\ \begin{pmatrix} C_{12} \\ W^* \end{pmatrix} C_{22}^{-2} \\ \vdots \\ \begin{pmatrix} C_{12} \\ W^* \end{pmatrix} C_{22}^{-(p-1)} \end{bmatrix},$$

has full column rank r. The rows of $W^*C_{22}^{-i}$ are not linearly independent of the rows of

$C_{12}C_{22}^{-(i+1)}$ $(i = 1, \ldots, p-2)$ and can be withdrawn from the matrix without reducing its rank, leaving $\bar{W}$. $\square$

Proof of Theorem 1. (a) The existence of an $(m+1)r \times m(r+n)$ matrix R such that $R\bar{M}$ that is non-singular follows naturally from Lemma A2. (b) Pre-multiplying (17) by R and solving leads to

$$w_{th-h} = \hat{M}\left[\bar{c} + N\bar{x}_{th} + e_{th}\right], \tag{A8}$$

where $\hat{M} = [I_r, 0_{r\times mr}][R\bar{M}]^{-1}R$. The matrices F_j $(j = 1, \ldots, m+1)$ and vector f follow directly from substituting back into (15). The error process, η_{th}, can be written

$$\begin{aligned}
\eta_{th} &= \epsilon_{1,th} + C_{12}\hat{M}e_{th} = \epsilon_{1,th} + C_{12}\sum_{i=1}^{m}\left[\hat{M}_{1,i}\epsilon_{1,th-ih} + \hat{M}_i\epsilon_{2,th-ih}\right] \\
&= \int_0^1 \left\{ S_1\tilde{C}(r)\tilde{\Theta}hu(r) + C_{12}\sum_{i=1}^{m}\left[\hat{M}_{1,i}S_1 + \hat{M}_iS_2\right]\tilde{C}(r)\tilde{\Theta}hu(r-i) \right\} dr.
\end{aligned}$$

The covariance structure of η_{th} then follows from the properties of $u(t)$. $\square$

Proof of Theorem 2. (a) The $p(r+n) \times n^s$ matrix N^s_{p+1}, which contains n^s columns of the non-singular matrix e^{Hh}, appended with additional rows of zeros, has rank n^s and the dimension of its left nullspace is $p(r+n) - n^s = (p+1)r$. Let R denote any matrix whose rows span this space. $R\bar{M}$ is non-singular provided the column space of $\bar{M}$ and the nullspace of R are disjoint. Since, by construction, the nullspace of R is spanned by the columns of N^s_{p+1}, this follows from Lemma A4. (b) Given this choice of R, theorem 1 then gives the coefficients of an ARMA$(p+1, p)$ process but with $F_{p+1} = 0$. Since any matrix, $\tilde{R}$, whose rows span the left nullspace of N^s_{p+1} can be written ΥR, where Υ is a $(p+1)r$ invertible matrix, the representation is unique. $\square$