Econometric Theory, 2011, Page 1 of 20.
doi:10.1017/S0266466611000181

DISCRETE TIME REPRESENTATION OF CONTINUOUS TIME ARMA PROCESSES

MARCUS J. CHAMBERS
University of Essex

MICHAEL A. THORNTON
University of Reading

This paper derives exact discrete time representations for data generated by a continuous time autoregressive moving average (ARMA) system with mixed stock and flow data. The representations for systems comprised entirely of stocks or of flows are also given. In each case the discrete time representations are shown to be of ARMA form, the orders depending on those of the continuous time system. Three examples and applications are also provided, two of which concern the stationary ARMA(2, 1) model with stock variables (with applications to sunspot data and a short-term interest rate) and one concerning the nonstationary ARMA(2, 1) model with a flow variable (with an application to U.S. nondurable consumers' expenditure). In all three examples the presence of an MA(1) component in the continuous time system has a dramatic impact on eradicating unaccounted-for serial correlation that is present in the discrete time version of the ARMA(2, 0) specification, even though the form of the discrete time model is ARMA(2, 1) for both models.

1. INTRODUCTION

A variety of methods can be utilized for the estimation of the parameters of linear continuous time systems, with approaches based on spectral representations, Kalman filtering of state space forms, and exact discrete time representations having generated the most interest in econometrics. A powerful method based on a discrete Fourier transform of the data was proposed by Robinson (1976) and is applicable to a wide class of stationary systems with exogenous variables that includes higher order continuous time autoregressive moving average (ARMA) models and models containing mixed differential-difference equations. An alternative spectral likelihood–based method was suggested by Robinson (1993) for

The second author's research was supported by an Economic and Social Research Council studentship (PTA-030-2006-00395). We thank the editor, Peter Phillips, and two anonymous referees for helpful comments. Address correspondence to Marcus J. Chambers, Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, United Kingdom; e-mail: mchamb@essex.ac.uk.

estimating continuous time ARMA systems that may include exogenous variables and mixtures of stocks and flows, and Phillips (1991) proposed spectral regession methods for the estimation of cointegrating vectors in cointegrated continuous time systems. Kalman filtering techniques based on state space representations have the advantage of being able to handle a wide range of models, including higher order systems, moving average (MA) disturbances, mixtures of stock and flow variables, exogenous variables, stochastic trends, cointegration, and irregularly sampled data, with important contributions being made by Harvey and Stock (1985, 1988) and Zadrozny (1988). Continuous time ARMA systems displaying fractional integration and driven by Lévy processes can also be handled by the state space/Kalman filtering approach, as in Brockwell (2001, 2004) and Brockwell and Marquardt (2005).

The advantages of using exact discrete time representations of stochastic differential equation systems were eloquently conveyed in Bergstrom (1990), and the algorithms currently available are able to deal with most of the features mentioned earlier; see, for example, Bergstrom (1983, 1986, 1997) and Chambers (1999, 2009). Bergstrom (1985) also argues that, in addition to being of interest in its own right, the exact discrete time model approach is computationally more efficient than the Kalman filter approach, once the setup costs of deriving the discrete time model have been borne. Features not currently handled by this approach include irregularly sampled data and MA disturbances, although Bergstrom (1984) does show how a model with MA disturbances can be transformed into a higher order model with a white noise disturbance, albeit at the cost of an identification problem. Although the algorithms could no doubt be modified to deal with both of these aspects, it is the latter one that is addressed here.

This paper aims to extend the range of continuous time models that can be estimated using an exact discrete time representation by including MA disturbances in the system of stochastic differential equations. Until now the exact discrete time representation approach has been restricted to models driven by some form of continuous time white noise process. However, MA disturbances can arise quite naturally in continuous time economic models, the consumption models of Christiano, Eichenbaum, and Marshall (1991) and Thornton (2009) being prime examples, and it is therefore pertinent to extend the range of models for which an exact discrete time representation has been derived. The variables in the continuous time model under consideration may be stationary, nonstationary, or cointegrated, and discrete time models are derived for the cases where the variables are stocks, flows, or mixtures of the two. The key to deriving suitable discrete time representations lies in writing the model in a particular state space form, based on Zadrozny (1988), and then utilizing the results in Chambers (1999). The approach therefore contrasts with the usual time domain approach to deriving discrete time models that is based on the state space form used by Bergstrom (1983) and that is standard in linear systems theory; see, for example, Kwakernaak and Sivan (1972). We are able to show that the discrete time representation of a continuous time ARMA(p,q) system (with $q < p$) is ARMA$(p, p-1)$ in the case of stock variables and ARMA(p, p) in the case of flows or mixtures of stocks and flows. Phadke and Wu (1974) proposed a coefficient-matching method to estimate a univariate continuous time ARMA(p, q) for a stock variable but did not derive the exact discrete model per se.

The paper is organized as follows. Section 2 defines the continuous time ARMA system and specifies the state space form that is used to derive the exact discrete time representation, pointing out why the form used in Chambers (1999) is not the most useful when MA disturbances are present. The discrete time representation in the cases of mixed stock and flow data is presented in Theorem 1 in Section 3, with the (nonnested) cases of pure stocks and pure flows given in Corollaries 1 and 2, respectively. Section 4 concentrates on three examples and applications, two of which concern the stationary ARMA(2, 1) model with stock variables (with applications to sunspot data and a short-term interest rate) and one concerning the nonstationary ARMA(2, 1) model with a flow variable (with an application to U.S. nondurable consumers' expenditure). In all three examples the presence of an MA(1) component in the continuous time system has a dramatic impact on eradicating unaccounted-for serial correlation that is present in the discrete time version of the ARMA(2, 0) specification, even though the form of the discrete time model is ARMA(2, 1) for both models. Section 5 concludes, and all proofs are contained in an Appendix.

2. THE MODEL AND ITS SOLUTION

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

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

where D denotes the mean square differential operator, $A_0, \ldots, A_{p-1}$ and $\Theta_1, \ldots, \Theta_q$ are $n \times n$ matrices of unknown coefficients, a_0 is an $n \times 1$ vector of unknown constants, and $u(t)$ is an $n \times 1$ continuous time white noise vector with variance matrix Σ. The matrices of unknown coefficients may, of course, depend on an underlying parameter vector β of more deeply embedded structural parameters, and the matrix Θ_0 is (implicitly) set equal to an identity matrix for purposes of identification. Although the process $u(t)$ and its derivatives are not physically realizable systems such as (1) are of widespread interest, and the condition $q < p$ is imposed so that $x(t)$ itself has an integrable spectral density matrix and, hence, finite variance. We therefore also assume that $p \geq 2$. Further details on the specification of continuous time random processes for the purposes of constructing econometric models can be found in Bergstrom (1984).

Define $\widetilde{y}(t) = [x(t)', Dx(t)', \ldots, D^{p-1}x(t)']'$ and the following vectors and matrix:

$$a = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ a_0 \end{bmatrix}, \qquad \widetilde{A} = \begin{bmatrix} 0 & I & 0 & \dots & 0 \\ 0 & 0 & I & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \dots & I \\ A_0 & A_1 & A_2 & \dots & A_{p-1} \end{bmatrix}, \qquad \widetilde{u}(t) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \upsilon(t) \end{bmatrix},$$

where $\upsilon(t) = u(t) + \Theta_1 Du(t) + \cdots + \Theta_q D^q u(t)$. Then the state vector $\widetilde{y}(t)$ satisfies

$$D\widetilde{y}(t) = a + \widetilde{A}\widetilde{y}(t) + \widetilde{u}(t), \qquad t > 0. \tag{2}$$

The formulas in Chambers (1999) remain exactly valid for the model (1), but the derivation of the autocovariance properties of the resulting discrete time disturbance vector becomes more complicated because of the presence of the derivatives of $u(t)$ in $\widetilde{u}(t)$. We therefore work with an alternative state space representation, based on Zadrozny (1988), in which the $np \times 1$ state vector is defined as $y(t) = [y_1(t)', \dots, y_p(t)']'$ and with which we associate $y_1(t) = x(t)$. The state space form is based on the following set of p equations in the derivatives of the components of $y(t)$, given by

$$Dy_1(t) = A_{p-1} y_1(t) + y_2(t) + \Theta_{p-1} u(t), \tag{3}$$

$$Dy_2(t) = A_{p-2} y_1(t) + y_3(t) + \Theta_{p-2} u(t), \tag{4}$$

$$\vdots$$

$$Dy_{p-1}(t) = A_1 y_1(t) + y_p(t) + \Theta_1 u(t), \tag{5}$$

$$Dy_p(t) = a_0 + A_0 y_1(t) + u(t), \tag{6}$$

in which we define $\Theta_j = 0$ for $j > q$. To demonstrate that $y_1(t)$ satisfies (1), note that differentiation of (3) yields $D^2 y_1(t) = A_{p-1} Dy_1(t) + Dy_2(t) + \Theta_{p-1} Du(t)$ and substituting (4) for $Dy_2(t)$ gives

$$D^2 y_1(t) = A_{p-1} Dy_1(t) + A_{p-2} y_1(t) + y_3(t) + \Theta_{p-2} u(t) + \Theta_{p-1} Du(t).$$

Repeated further differentiation and substitution of the expressions for the $Dy_j(t)$ result in the equation for $y_1(t) = x(t)$ determined by the system (1). Combining the preceding expressions for $Dy_1(t), \dots, Dy_p(t)$, the state space form can therefore be written

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

where

$$A = \begin{pmatrix} A_{p-1} & I & 0 & \dots & 0 \\ A_{p-2} & 0 & I & \dots & 0 \\ \vdots & & & & \vdots \\ A_1 & 0 & 0 & \dots & I \\ A_0 & 0 & 0 & \dots & 0 \end{pmatrix}, \qquad \Theta = \begin{pmatrix} \Theta_{p-1} \\ \Theta_{p-2} \\ \vdots \\ \Theta_1 \\ I \end{pmatrix}.$$

Although it is the representation in (2) that is conventionally used in systems theory (e.g., Kwakernaak and Sivan, 1972) and also in time series analysis more generally (e.g., Lütkepohl, 2005), it is the form in (7) that is more amenable to dealing with the estimation of continuous time systems from discrete time data. In particular the representation (7) extracts the parameter matrices of interest in a way that facilitates the derivation of the covariance structure of the exact discrete time representation. Note, in particular, that the matrix A is different than the matrix $\widetilde{A}$ used in the purely autoregressive case by Chambers (1999). In view of the matrix exponential function playing an important role in deriving discrete time representations it is noteworthy that, in general, $e^A \neq e^{\widetilde{A}}$, where $e^{At} = I + \sum_{j=1}^{\infty}(At)^j/j!$ and t is a scalar. This issue is discussed further in Thornton and Chambers (2010), in which a detailed analysis of these matrix exponentials is used to characterize the discrete time representations obtained from the two different state space forms.

The solution to (7), conditional on $y(0)$, can be written

$$y(t) = e^{At}y(0) + \int_0^t e^{A(t-s)}[a + \Theta u(s)]ds, \qquad t > 0. \tag{8}$$

Furthermore, the solution can be shown to be unique in a mean square sense, the proof of Theorem 1(b) of Bergstrom (1983) remaining valid for the system (1). The solution (8) provides the basis for the derivation of the exact discrete time representations in Section 3.

3. DISCRETE TIME REPRESENTATION: STOCKS, FLOWS, AND MIXED SAMPLES

As in Chambers (1999) we assume that the system contains both stock and flow variables, and so the continuous time vector $x(t)$ satisfying (1) will be partitioned, without loss of generality, 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 $n^s + n^f = n$. Although we assume that the observation interval is of unit length we acknowledge that the results can be extended straightforwardly to an arbitrary interval length at the cost of increased notational complexity. By noting that the integrated state vector $Y_t = \int_{t-1}^t y(r)dr$ contains the subvector $\int_{t-1}^t x^f(r)dr$ and that

$$\int_{t-1}^t Dx^s(r)dr = x^s(t) - x^s(t-1),$$

the observed variables will be arranged in the form

$$x_t = \begin{bmatrix} x^s(t) - x^s(t-1) \\ \int_{t-1}^{t} x^f(r)dr \end{bmatrix}, \qquad t = 1, \ldots, T,$$

T denoting the sample size. The discrete time state vector Y_t, however, comprises not only the observable $n \times 1$ vector x_t but also the $n(p-1) \times 1$ vector of unobservable components, denoted w_t, that need to be eliminated from the system. It is convenient to use the expressions $x_t = S_1 Y_t$ and $w_t = S_2 Y_t$, respectively, where

$$S_1 = \begin{bmatrix} 0 & 0 & I_{n^s} & 0 \\ 0 & I_{n^f} & 0 & 0 \end{bmatrix}, \qquad S_2 = \begin{bmatrix} I_{n^s} & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{n^r} \end{bmatrix}, \tag{9}$$

and $n^r = n(p-1) - n^s$.

The solution (8) can be used to derive the law of motion for Y_t. It is straightforward to show that $y(t)$ satisfies the stochastic difference equation

$$y(t) = c + Cy(t-1) + \epsilon(t), \qquad \epsilon(t) = \int_{t-1}^{t} C(t-s)\Theta u(s)ds, \qquad t = 1, \ldots, T, \tag{10}$$

where $C(r) = e^{rA}$, $c = \left[\int_0^1 C(r)dr\right] a$, and

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = C(1) = e^A.$$

Note that $C_{ij} = S_i C S_j'$ $(i, j = 1, 2)$. Integration of (10) over $(t-1, t]$ yields

$$Y_t = c + CY_{t-1} + v_t,$$
$$v_t = \int_{t-1}^{t} \epsilon(r)dr = \int_{t-1}^{t} \int_{r-1}^{r} C(r-s)\Theta u(s)dsdr, \qquad t = 2, \ldots, T, \tag{11}$$

from which the discrete time model can be derived. The objective is to eliminate the elements of w_t from (11) to derive a stochastic difference equation for x_t. To do this the following assumptions will be made.

Assumption 1. The $n(p-1) \times n(p-1)$ matrix C_{22} is nonsingular.

Assumption 2. The $n \times n(p-1)$ matrix C_{12} has rank n.

Assumption 3. The $n(p-1) \times n(p-1)^2$ matrix $\left[C_{22}^{-1}, \ldots, C_{22}^{-(p-1)}\right]$ has rank $n(p-1)$.

These assumptions correspond to Assumptions 2–4 in Chambers (1999) and are closely related to the concepts of reconstructibility and detectability employed in optimal control theory of linear systems; a discussion of their role in deriving the discrete time representation from the state space form (10) can be found in Chambers (1999, p. 626). Such assumptions on submatrices of the matrix exponential typically arise in the derivation of exact discrete time representations of continuous time systems; see, for example, Assumption 4 of Bergstrom (1983). They are not, however, required in Kalman filtering approaches for the purpose of estimation that use the transition equation directly; see, for example, Zadrozny (1988).

THEOREM 1. *Under Assumptions 1–3, the observed vector x_t of mixed stock and flow variables generated by the continuous time ARMA(p,q) system (1) satisfies the discrete time ARMA(p,p) system*

$$x_t = f + F_1 x_{t-1} + \cdots + F_p x_{t-p} + \zeta_t, \qquad t = p+1, \ldots, T,$$

where $f = c_1 + C_{12} M \bar{c}$, $F_1 = C_{11} + C_{12} M N_1$, $F_j = C_{12} M N_j$ $(j = 2, \ldots, p)$, and the vectors c_1 and $\bar{c}$ and matrices M and N_j are defined in the Appendix. Furthermore, the autocovariance matrices of ζ_t are given by

$$\mathrm{E}(\zeta_t \zeta_{t-j}') = \begin{cases} \sum_{i=j}^{p} K_i \Omega_\xi K_{i-j}', & j = 0, \ldots, p, \\ 0, & j > p, \end{cases}$$

where $K_0 = [S_1, 0]$, $K_1 = [C_{12}(M_{11} S_1 + M_2 S_2), S_1]$, $K_p = [0, C_{12}(M_{1,p-1} S_1 + M_p S_2)]$, $K_j = [C_{12}(M_{1j} S_1 + M_{j+1} S_2), C_{12}(M_{1,j-1} S_1 + M_j S_2)]$ $(j = 2, \ldots, p-1)$, and

$$\Omega_\xi = \begin{bmatrix} \Omega_{\xi,11} & \Omega_{\xi,12} \\ \Omega_{\xi,21} & \Omega_{\xi,22} \end{bmatrix} = \begin{pmatrix} \int_0^1 \Gamma_1(r) \Theta \Sigma \Theta' \Gamma_1(r)' dr & \int_0^1 \Gamma_1(r) \Theta \Sigma \Theta' \Gamma_2(r)' dr \\ \int_0^1 \Gamma_2(r) \Theta \Sigma \Theta' \Gamma_1(r)' dr & \int_0^1 \Gamma_2(r) \Theta \Sigma \Theta' \Gamma_2(r)' dr \end{pmatrix},$$

where $\Gamma_1(r) = \Phi(r)$, $\Gamma_2(r) = \Phi(1) - \Phi(r)$, and $\Phi(r) = \int_0^r C(s) ds = \sum_{j=0}^{\infty} r^{j+1} A^j/(j+1)!$.

Theorem 1 shows that the exact discrete time representation of a continuous time ARMA(p,q) process with mixed stock and flow data is of ARMA(p,p) form. Theorem 2.1 of Bergstrom (1986) established that a continuous time ARMA$(2,0)$ process with mixed sample has a discrete time ARMA$(2,2)$ representation, a result that was extended by Chambers (1999) to the continuous time ARMA$(p,0)$ case. Compared to the continuous time ARMA$(p,0)$ case, however, the presence of the continuous time MA component manifests itself in the covariance matrix Ω_ξ, via which the MA matrices contained in Θ have an impact. The discrete time ARMA representation in Theorem 1 can provide the basis for obtaining maximum likelihood estimates of the unknown parameters, conditional on $x_1, \ldots, x_p$. Unconditional estimates could be obtained by relating $x_1, \ldots, x_p$ to x_0 as in Theorem 2 of Chambers (1999), although for compactness we do not pursue this line of investigation here except to note that the methods used in that paper are equally valid for the continuous time ARMA system (1).

The cases where the variables are either all stocks or all flows can also be dealt with by similar methods although the exact discrete models are not special cases of Theorem 1 exactly. When $x(t)$ comprises stock variables the observations are now of the form $x_t = x(t)$ $(t = 1, \ldots, T)$, and the discrete time representation can be derived from (10) directly without recourse to (11). The relevant state vector is $y(t) = [x_t', w_t']'$, and the appropriate selection matrices are now

$$S_1 = [I_n, 0, \ldots, 0] \qquad (n \times np) \quad \text{and} \quad S_2 = \left[0, I_{n(p-1)}\right] \qquad (n(p-1) \times np). \tag{12}$$

The exact discrete model is defined in Corollary 1.

COROLLARY 1. *Under Assumptions 1–3, the observed vector x_t of stock variables generated by the continuous time ARMA(p, q) system (1) satisfies the discrete time ARMA$(p, p-1)$ system*

$$x_t = f + F_1 x_{t-1} + \cdots + F_p x_{t-p} + \eta_t, \qquad t = p+1, \ldots, T,$$

where f and the F_j $(j = 1, \ldots, p)$ are defined in Theorem 1. Furthermore, the autocovariance matrices of η_t are given by

$$\mathrm{E}(\eta_t \eta_{t-j}') = \begin{cases} \sum_{i=j}^{p-1} C_i \Omega_\epsilon C_{i-j}', & j = 0, \ldots, p-1, \\ 0, & j \geq p, \end{cases}$$

where $C_0 = S_1$, $C_j = C_{12}(M_{1j} S_1 + M_{j+1} S_2)$ $(j = 1, \ldots, p-1)$, M_{1j} and M_j are defined in the Appendix, and $\Omega_\epsilon = \mathrm{E}(\epsilon(t)\epsilon(t)') = \int_0^1 C(r)\Theta\Sigma\Theta' C(r)' dr$.

The result that the discrete time representation of a continuous time ARMA(p, q) process in a stock variable is ARMA$(p, p-1)$ was established by Phadke and Wu (1974), although they did not derive the discrete time representation in the generality provided by Corollary 1. The exact discrete time ARMA$(p, p-1)$ representation of a continuous time ARMA$(p, 0)$ system was established in a form similar to that given in Corollary 1 by Bergstrom (1984).

In the case where $x(t)$ contains flow variables the observations are of the form

$$X_t = \int_{t-1}^{t} x(r) dr, \qquad t = 1, \ldots, T,$$

and (11) provides the basis for deriving the exact discrete time representation. Partitioning Y_t as $Y_t = [X_t', W_t']'$ the observable and unobservable components of Y_t are given by $X_t = S_1 Y_t$ and $W_t = S_2 Y_t$, respectively, where S_1 and S_2 are defined in (12). As before the objective is to eliminate the $n(p-1) \times 1$ vector W_t of unobservable variables from the system.

COROLLARY 2. *Under Assumptions 1–3, the observed vector X_t of flow variables generated by the continuous time ARMA(p,q) system (1) satisfies the discrete time ARMA(p, p) system*

$$X_t = f + F_1 X_{t-1} + \cdots + F_p X_{t-p} + \rho_t, \qquad t = p+1, \ldots, T,$$

where f and the F_j $(j = 1, \ldots, p)$ are defined in Theorem 1. Furthermore, the autocovariance matrices of ρ_t are given by

$$\mathrm{E}(\rho_t \rho_{t-j}') = \begin{cases} \sum_{i=j}^{p} B_i \Omega_\xi B_{i-j}', & j = 0, \ldots, p, \\ 0, & j \geq p, \end{cases}$$

where $B_0 = [S_1, 0]$, $B_1 = [C_{12}(M_{11} S_1 + M_2 S_2), S_1]$, $B_p = [0, C_{12}(M_{1,p-1} S_1 + M_p S_2)]$, $B_j = [C_{12}(M_{1j} S_1 + M_{j+1} S_2), C_{12}(M_{1,j-1} S_1 + M_j S_2)]$ $(j = 2, \ldots, p-1)$, and Ω_ξ is defined in Theorem 1.

When the variable of interest is a flow variable and is observed as an integral of the underlying flow the impact on the discrete time representation over the stock variable case is that the MA order increases from $p-1$ to p. This is well known in the discrete time representation of continuous time ARMA$(p, 0)$ processes (see Bergstrom, 1985) but appears not to have been established until now for continuous time ARMA(p, q) processes. Note that the form of the autocovariance structure is the same as for the mixed sample case in Theorem 1. We have chosen to use matrices B_j in Corollary 2 rather than K_j as in Theorem 1 in view of the selection matrices being different in the two cases.

4. SOME EXAMPLES AND APPLICATIONS

4.1. The Stationary Continuous Time ARMA(2, 1) Process with Stocks

In this section we provide two applications of our methodology with the stationary continuous time ARMA(2, 1) model, the first using sunspot data and the second using short-term interest rate data. The sunspot data were used by Phadke and Wu (1974), who obtained an estimated cycle length of 10.83 years using a continuous time ARMA(2, 1) model. Their method of estimating the continuous time parameters is essentially based on matching the autocovariance properties of the continuous time ARMA process with the autocovariances based on maximum likelihood estimates of the parameters of the discrete time ARMA process of appropriate order. It is of interest to compare their estimates with those that are obtained by computing maximum likelihood estimates based on the appropriate exact discrete time representation derived previously.

The univariate continuous time ARMA(2, 1) model under consideration is

$$D^2x(t) = a_1 Dx(t) + a_0 x(t) + u(t) + \theta_1 Du(t), \qquad -\infty < t < \infty, \tag{13}$$

where $x(t) = y(t) - \mu$, $y(t)$ denotes the sunspot number, $\mu = \mathrm{E}(y(t))$, a_0, a_1, and θ_1 are the unknown scalar parameters, and the variance of $u(t)$ will be denoted σ_u^2. The observations are assumed to take the form $y_t = y(t)$ $(t = 1, \ldots, T)$, and the variable $x_t = y_t - \hat{\mu}$ is used for estimation, where $\hat{\mu} = T^{-1}\sum_{t=1}^{T} y_t$. The model is stationary provided the roots of the characteristic equation $z^2 - a_1 z - a_0 = 0$ have negative real parts. The exact discrete time representation takes the ARMA(2, 1) form

$$x_t = f_1 x_{t-1} + f_2 x_{t-2} + \eta_t, \qquad t = 3, \ldots, T, \tag{14}$$

where η_t is MA(1). Let $T_0 = T - 2$ and let $\eta = (\eta_3, \eta_4, \ldots, \eta_T)'$ denote the $T_0 \times 1$ vector of discrete time disturbances with covariance matrix $\Omega_\eta = \mathrm{E}(\eta\eta')$, the latter being a sparse Toeplitz matrix in view of η being a vector of MA(1) components. Then, assuming that η is normally distributed, the logarithm of the likelihood function is

$$\log L(\beta) = -\frac{T_0}{2}\log 2\pi - \frac{1}{2}\log|\Omega_\eta| - \frac{1}{2}\eta'\Omega_\eta^{-1}\eta, \tag{15}$$

where $\beta = (a_0, a_1, \theta_1, \sigma_u^2)'$ denotes the unknown parameter vector and $|\Omega_\eta|$ denotes the determinant of Ω_η. Following Bergstrom (1985) it is possible to express $\log L(\beta)$ in terms of a vector of uncorrelated standard normal variates $\epsilon = P^{-1}\eta$, where P is a sparse lower triangular matrix satisfying $PP' = \Omega_\eta$, as follows:

$$\log L(\beta) = -\frac{T_0}{2}\log 2\pi - \sum_{t=3}^{T}\log p_{tt} - \frac{1}{2}\sum_{t=3}^{T}\epsilon_t^2, \tag{16}$$

where p_{tt} denotes the tth diagonal element of P. Bergstrom (1990) proposed that the vector of uncorrelated standard normal variates ϵ be used to compute a portmanteau-type test statistic given by

$$S_l = \frac{1}{T_0 - l}\sum_{r=1}^{l}\left(\sum_{t=l+1}^{T}\epsilon_t\epsilon_{t-r}\right)^2. \tag{17}$$

Under the null hypothesis that the model is correctly specified the statistic S_l will have an approximate chi-square distribution with l degrees of freedom for sufficiently large l and $T - l$. Alternative tests of serial correlation in ϵ can also be carried out, and the results that follow report the Box–Pierce Q statistic defined by

$$Q_l = T_0 \sum_{\tau=1}^{l} r_\tau^2, \tag{18}$$

where r_τ denotes the lag-τ sample autocorrelation of the ϵ_t. Under the null hypothesis that the ϵ_t are white noise the statistic Q_l will be approximately chi-square with l degrees of freedom in large samples.

Table 1 contains estimates of the model using sunspot data covering the period 1749–1924. The second column reports Phadke and Wu's results by way of comparison, and columns three and four contain the estimated parameters in the continuous time ARMA(2, 0) and ARMA(2, 1) models, respectively. The inadequacy of the continuous time ARMA(2, 0) model is clear from the low marginal probability values for the statistics S_{20} and Q_{20}, both of which are indicative of dynamic misspecification. Adding the continuous time MA(1) term to the model yields a statistically significant coefficient estimate, and neither portmanteau statistic is significant. The parameter estimates are very close to those obtained by Phadke and Wu although the resulting cycle length estimate is slightly higher and, furthermore, the standard errors of the estimated parameters (with the exception of θ_1) are smaller than those reported by Phadke and Wu. The fact that the continuous time MA(1) is able to remove the serial correlation from the discrete time MA component ϵ_t is of particular interest especially in view of the fact that the

TABLE 1. Estimates of continuous time ARMA models using sunspot data, 1749–1924

Parameter	PW (1974)	ARMA(2, 0)	ARMA(2, 1)
a_0	−0.3593	−0.4963	−0.3579
	(0.0751)	(0.0676)	(0.0444)
a_1	−0.3271	−0.7752	−0.3223
	(0.1251)	(0.1411)	(0.0925)
θ_1	0.6333	0.0000	0.6416
	(0.1249)		(0.1682)
σ_u	15.6990	30.4053	15.5068
	N.A.	(2.4382)	(2.6133)
$\log L$		−737.1520	−729.5901
S_{20}		0.0318	0.1835
Q_{20}		0.0045	0.3507
Roots	−0.1636	−0.3876	−0.1612
	$\pm 0.6767i$	$\pm 0.5883i$	$\pm 0.5761i$
Period	10.8329	10.6807	10.9058

Note: Figures in parentheses denote standard errors; N.A. denotes that the standard error is not available; the absence of a standard error indicates that the parameter was constrained to its reported value; and the entries for S_{20} and Q_{20} denote marginal probability values.

discrete time disturbance η_t is MA(1) for both exact discrete models. This empirical example highlights the importance of being able to allow for MA components in continuous time autoregressive models.

The second application employs our techniques in the modeling of short-term interest rates. Phillips and Yu (2009) provide an overview of maximum likelihood and Gaussian methods of estimating continuous time models used in finance, not all of which are amenable to analysis via exact discrete time representations because of their inherent nonlinearity. In the context of (essentially) linear models Nowman (1997, 1998) used the exact discrete representation of a continuous time AR(1) process to estimate a range of models of the term structure for interest rates, including the popular model of Vasicek (1977). Some of the models estimated by Nowman differ in their parameterizations of the volatility component, which here is taken to be constant. Nowman (1998) finds evidence of dynamic misspecification (as judged by the reported S statistics), which is of principal interest in this application. A further motivation for this application is the recent work by Benth, Koekebakker, and Zakamouline (2010), who propose generalizing the Vasicek model to include continuous time ARMA dynamics. By calibrating their model to minimize the mean square error between the actual yield curve and that consistent with their theoretical model, they conclude that a continuous time ARMA(2, 1) is better able to capture the key features of the data than the standard continuous time ARMA(1, 0). They do not, however, estimate the parameters directly. Our method enables the parameters of the model to be estimated directly from data on short-term interest rates, taking account of temporal aggregation as in Nowman (1997, 1998), while incorporating more sophisticated dynamics.

The model estimated for the short-term interest rate is given by

$$D^2x(t) = \kappa + a_1 Dx(t) + a_0 x(t) + u(t) + \theta_1 Du(t), \qquad -\infty < t < \infty, \tag{19}$$

where κ denotes the intercept. The only difference when comparing this model with (13) (the sunspot model) is that the interest rate data are not mean adjusted and the intercept is treated as a separate parameter. The discrete time model is ARMA(2, 1), and the same likelihood procedure as outlined before is used again. The data are the 1 month sterling interbank rate (the middle rate) from January 1978 to August 2008. Our estimates for a continuous time ARMA(1, 0), which is consistent with Vasicek's original model, a continuous time ARMA(2, 0), and a continuous time ARMA(2, 1) are in Table 2. Our results for the continuous time ARMA(1, 0) are broadly comparable with Nowman (1997). The continuous time ARMA(2, 1) model shows notable improvements in both Bergstrom's S and the Box–Pierce Q statistics relative to both the ARMA(1, 0) and ARMA(2, 0) specifications. The likelihood ratio statistic comparing the continuous time ARMA (2, 1) to the ARMA(1, 0) of 11.84 is highly significant, as is the value of 10.83 comparing it to the ARMA(2, 0). There is, therefore, some statistical evidence favoring the continuous time ARMA(2, 1) specification over purely autoregressive specifications in this example also.

Table 2. Estimates of continuous time ARMA models using UK interest rate data, January 1978–August 2008

Parameter	ARMA(1, 0)	ARMA(2, 0)	ARMA(2, 1)
κ	0.0009	0.0116	0.0001
	(0.0008)	(0.0142)	(0.0001)
a_0	−0.0113	−0.1384	−0.0011
	(0.0081)	(0.1643)	(0.0008)
a_1	0.0000	−11.1581	−0.0658
		(0.0216)	(0.0247)
θ_1	0.0000	0.0000	19.0282
			(14.3238)
σ_u	0.0059	0.0687	0.0003
	(0.0002)	(0.0026)	(0.0002)
$\log L$	1,362.6305	1,363.1380	1,368.5527
S_{12}	0.0936	0.0896	0.2093
Q_{12}	0.0804	0.1045	0.2389
Roots	−0.0113	−11.1457	−0.0329
		−0.0124	$\pm 0.0080i$

Note: Figures in parentheses denote standard errors; the absence of a standard error indicates that the parameter was constrained to its reported value; and the entries for S_{12} and Q_{12} denote marginal probability values.

4.2. The Nonstationary Continuous Time ARMA(2, 1) Process with Flows

The focus now turns to the nonstationary univariate continuous time ARMA(2, 1) process with a flow variable. By nonstationary we are referring to the situation where one of the roots of the characteristic equation is identically equal to zero. The reason for considering this particular process is that models of consumption as continuous time processes with MA errors have received attention as generalizations of the martingale hypothesis. For example, Christiano et al. (1991) consider a structural model of the permanent income hypothesis in continuous time. They note that when technology shocks are included in their model, or when technology is deterministic but the representative household's discount rate does not equal the productivity of capital, then consumption follows an ARMA process. More recently Thornton (2009), building a continuous time analogue to the Pischke (1995) model of consumption under incomplete information, shows that household consumption would follow an integrated continuous time ARMA(1, 1) process and derives the exact discrete time representation. Continuous time ARMA models also provide scope to widen the range of models of consumer behavior beyond the convolution-based models of habit formation and exponential discounting considered by Heaton (1993).

Further support that a continuous time model of consumption might contain an MA disturbance is provided by a range of studies looking at income dynamics.

MaCurdy (1982), Abowd and Card (1989), and Pischke (1995) model changes in earnings using flow data gathered at various frequencies They each conclude that earnings are well described by an integrated MA(2) process, although Deaton (1992) notes that an integrated MA(1) process is also a relatively good approximation, and they each report a negative first-order autocorrelation in income changes. It is well known, following Working (1960), that such a process could not arise from the temporal aggregation of a continuous time martingale, which would appear as an integrated MA(1) process with first-order autocovariance of 0.25, suggesting that an equivalent underlying continuous time process should have an MA component. This being so, it would add an MA term into the consumption patterns of households who were liquidity constrained or otherwise displayed "Keynesian" behavior that linked current consumption to current income changes.

The model of interest is obtained from (1) by setting $p=2$, $q=1$, $n=1$, and $a_0 = A_0 = 0$:

$$D^2 x(t) = a_1 Dx(t) + u(t) + \theta_1 Du(t), \qquad t > 0, \tag{20}$$

where a_1 and θ_1 are the scalar unknown parameters and σ_u^2 will denote the variance of $u(t)$. The variable $x(t) = y(t) - \mu - \gamma t$ where $y(t)$ denotes the rate of flow of consumption expenditures on nondurables. Note that $\int_{t-1}^{t} x(r)dr = \int_{t-1}^{t} y(r)dr - \mu - \gamma(t-0.5)$. Based on the observed sequence $y_t = \int_{t-1}^{t} y(r)dr$ $(t = 1,\dots,T)$ we therefore use the variable $x_t = y_t - \hat{\mu}_y - \hat{\gamma}_y t$, where $\hat{\mu}_y$ and $\hat{\gamma}_y$ are obtained from an ordinary least squares regression of x_t on a constant and a time trend. The underlying state space matrix and its exponential for this model are given by

$$A = \begin{bmatrix} a_1 & 1 \\ 0 & 0 \end{bmatrix}, \qquad e^A = C = I + \sum_{k=1}^{\infty} \frac{1}{k!} \begin{bmatrix} a_1^k & a_1^{k-1} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} e^{a_1} & (e^{a_1}-1)/a_1 \\ 0 & 1 \end{bmatrix}.$$

Corollary 2 yields the exact discrete time representation

$$x_t = f_1 x_{t-1} + f_2 x_{t-2} + \rho_t, \qquad t = 3,\dots,T, \tag{21}$$

where it can be shown that $f_1 = e^{a_1} + 1$ and $f_2 = -e^{a_1}$, and so the first-differenced variable $\Delta x_t = x_t - x_{t-1}$ satisfies

$$\Delta x_t = e^{a_1} \Delta x_{t-1} + \rho_t, \qquad t = 3,\dots,T. \tag{22}$$

Furthermore, from the proof of Corollary 2, it follows that $\rho_t = B_0 \xi_t + B_1 \xi_{t-1} + B_2 \xi_{t-2}$, where ξ_t is the 4×1 vector

$$\xi_t = \begin{bmatrix} \displaystyle\int_{t-1}^{t} \Gamma_1(t-r)\Theta u(s)ds \\ \displaystyle\int_{t-1}^{t} \Gamma_2(t-r)\Theta u(s)ds \end{bmatrix},$$

where $\Theta = (\theta_1, 1)'$ and $\Gamma_1(r)$ and $\Gamma_2(r)$ are defined in Theorem 1, and it can also be shown that

$$B_0 = [S_1, 0] = [1, 0, 0, 0], \qquad B_1 = [-1, (e^{a_1} - 1)/a_1, 1, 0],$$

$$B_2 = [0, 0, -1, (e^{a_1} - 1)/a_1].$$

Hence x_t is an integrated ARMA$(2,2)$ process, and Δx_t is a stationary ARMA$(1,2)$ process, assuming that $a_1 < 0$.

As an application of this model its parameters are estimated using quarterly U.S. nondurable consumption over the period 1986(1)–2005(4), a total of 80 seasonally adjusted observations. Estimates of two models are reported in Table 3, the difference between them being that one constrains the MA parameter $\theta_1 = 0$. In the constrained model there is evidence of residual serial correlation in the discrete time representation, but allowing for θ_1 to be estimated freely appears to alleviate the serial correlation with a statistically significant estimate of θ_1 (the likelihood ratio statistic for testing $\theta_1 = 0$ is 13.84), which also results in a small reduction of the estimated variance parameter σ_u^2. By way of comparison the discrete time random walk, $\Delta x_t = \epsilon_t$, which would correspond to the underlying continuous time model $Dx(t) = u(t)$ with $x(t)$ (incorrectly) treated as a stock variable, yielded a log-likelihood value of 269.5929, whereas introducing a (statistically significant) MA component (corresponding to the same underlying continuous time model but treating $x(t)$ as a flow) resulted in a log-likelihood value of 278.0518. A comparison of these log-likelihood values suggests that

TABLE 3. Estimates of continuous time ARMA models using U.S. consumption data, 1986(1)–2005(4)

Parameter	ARMA$(2,0)$	ARMA$(2,1)$
a_0	0.0000	0.0000
a_1	-11.6969	-16.8639
	(0.3492)	(0.3492)
θ_1	0.0000	0.0596
		(0.0014)
σ_u	0.0872	0.0854
	(0.0076)	(0.0102)
$\log L$	298.5008	305.4183
S_{20}	0.0712	0.3949
Q_{20}	0.0196	0.5003
Roots	0.0000	0.0000
	-11.6969	-16.8639

Note: Figures in parentheses denote standard errors; the absence of a standard error indicates that the parameter was constrained to its reported value; and the entries for S_{20} and Q_{20} denote marginal probability values.

the continuous time specification has empirical content over the discrete time specification. Furthermore, as with the sunspot and interest rate data, the addition of an MA component in the continuous time model appears to have a substantial and important impact on the properties of the estimated model. Moreover, in the case of consumption, there are underlying theoretical reasons to support the presence of the MA component.

5. CONCLUDING COMMENTS

This paper has provided exact discrete time representations for data generated by a continuous time ARMA(p, q) system when the variables may be stocks, flows, or a combination of the two. In the case of stocks the exact discrete time model is ARMA$(p, p-1)$, whereas for flows or a mixture of stocks and flows it is shown to be ARMA(p, p). Three univariate applications of the methodology have also been reported, two using the stationary ARMA$(2, 1)$ model with stock variables (sunspot data and a short-term interest rate) and one using the nonstationary ARMA$(2, 1)$ model with a flow variable (U.S. nondurable consumers' expenditure). In all cases the presence of the continuous time MA component has empirical impact, eliminating the presence of serial correlation in the residuals of the nested (and more common) ARMA$(2, 0)$ model. In view of the potential importance of MA disturbances in continuous time systems, our results provide exact discrete time representations that can be utilized in the estimation of such models and that provide alternatives to existing frequency domain and Kalman filter/state space approaches that could equally well be used. Further work, comparing the relative properties of these different approaches to the estimation of continuous time ARMA systems, would appear to be apposite.

Two further points are worthy of mention. The first is that it is possible, in principle, to derive the discrete time representations using an alternative approach, based on Bergstrom (1983), allied to an integration-by-parts formula to deal with the presence of the continuous time MA component. For reasons mentioned in Section 2 this approach is less appealing than the one adopted here and becomes increasingly complicated to implement beyond the ARMA$(2, 1)$ case with stock variables. Nevertheless its validity is demonstrated in Thornton and Chambers (2010). The second point concerns inference in the presence of continuous time zero roots. Phillips (1991) demonstrated how zero roots and cointegration in continuous time systems feed through into unit roots and cointegration in the observed discrete time data, and standard methods of testing for unit roots and cointegration rank can be applied. However, if the continuous time model is taken seriously, it will impose complicated restrictions on the discrete time data as a result of the process of temporal aggregation, and the testing of zero roots and cointegration restrictions imposed directly on the continuous time system may yield efficiency gains as opposed to tests that ignore such restrictions. Park and Jeong (2010) have developed an asymptotic theory for maximum likelihood estimators of the parameters of diffusion models that contain a zero root, and these results will, of course, be relevant to this type of analysis. The investigation of such issues is currently being undertaken by the authors.

REFERENCES

Abowd, J.M. & D. Card (1989) On the covariance structure of earnings and hours changes. *Econometrica* 57, 411–445.

Benth, F.E., S. Koekebakker, & E.V. Zakamouline (2010) The CARMA Interest Rate Model. Manuscript, University of Agder.

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 Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, pp. 1145–1212. North-Holland.

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.

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

Brockwell, P.A. (2001) Lévy-driven CARMA processes. *Annals of the Institute of Statistical Mathematics* 53, 113–124.

Brockwell, P.A. (2004) Representations of continuous-time ARMA processes. *Journal of Applied Probability* 41A, 375–382.

Brockwell, P.A. & T. Marquardt (2005) Lévy-driven and fractionally integrated ARMA processes with continuous time parameter. *Statistica Sinica* 15, 477–494.

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.

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

Deaton, A. (1992) *Understanding Consumption*. Oxford University Press.

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

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

Heaton, J. (1993) The interaction between time non-separable preferences and time aggregation. *Econometrica* 61, 353–385.

Kwakernaak, H. & R. Sivan (1972) *Linear Optimal Control Systems.* Wiley.

Lütkepohl, H. (2005) *New Introduction to Multiple Time Series Analysis.* Springer-Verlag.

MaCurdy, T.E. (1982) The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics* 18, 83–114.

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.

Nowman, K.B. (1998) Continuous-time short term interest rate models. *Applied Financial Economics* 8, 401–407.

Park, J.Y. & M. Jeong (2010) Asymptotic Theory of Maximum Likelihood Estimator for Diffusion Model. Manuscript, Indiana University.

Phadke, M.S. & S.M. Wu (1974) Modeling of continuous stochastic processes from discrete observations with application to sunspots data. *Journal of the American Statistical Association* 69, 325–329.

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

Phillips, P.C.B. & J. Yu (2009) Maximum likelihood and Gaussian estimation of continuous time models in finance. In T.G. Anderson, R.A. Davis, J.P. Kreiss, & T. Mikosch (eds.), *Handbook of Financial Time Series*, pp. 497–530. Springer-Verlag.

Pischke, J.-S. (1995) Individual income, incomplete information, and aggregate consumption. *Econometrica* 63, 805–840.

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

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.

Thornton, M.A. (2009) Information and aggregation: The econometrics of dynamic models of consumption under cross-sectional and temporal aggregation. Ph.D. Dissertation, University of Essex.

Thornton, M.A. & M.J. Chambers (2010) Continuous Time ARMA Processes in Rival State Space Forms. Manuscript, University of Reading.

Vasicek, O. (1977) An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.

Working, H. (1960) Note on the correlation of first differences of averages in a random chain. *Econometrica* 28, 916–918.

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 following matrices are used to define the autoregressive and MA matrices and vector of intercepts in Theorem 1 and the two corollaries:

$$M = \left[M_1, M_2, \ldots, M_p\right] = \hat{M}^{-1}\left[-I_{n(p-1)}, \widetilde{M}\right], \qquad M_1 = \left[M_{11}, M_{12}, \ldots, M_{1,p-1}\right],$$

$$\hat{M} = \begin{bmatrix} C_{12}C_{22}^{-1} \\ C_{12}C_{22}^{-2} \\ \vdots \\ C_{12}C_{22}^{-(p-1)} \end{bmatrix}, \qquad \widetilde{M} = \begin{bmatrix} C_{12}C_{22}^{-1} & 0 & \ldots & 0 & 0 \\ C_{12}C_{22}^{-2} & C_{12}C_{22}^{-1} & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \\ C_{12}C_{22}^{-(p-2)} & C_{12}C_{22}^{-(p-1)} & \ldots & C_{12}C_{22}^{-1} & 0 \\ C_{12}C_{22}^{-(p-1)} & C_{12}C_{22}^{-(p-2)} & \ldots & C_{12}C_{22}^{-2} & C_{12}C_{22}^{-1} \end{bmatrix},$$

$$N = \left[N_1, N_2, \ldots, N_p\right] = \left[\begin{array}{cccccc} -I_n & C_{11} & 0 & \ldots & 0 & 0 \\ 0 & -I_n & C_{11} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -I_n & C_{11} \\ \hline 0 & C_{21} & 0 & \ldots & 0 & 0 \\ 0 & 0 & C_{21} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & C_{21} \end{array}\right], \qquad \overline{c} = \begin{bmatrix} i_{p-1} \otimes c_1 \\ i_{p-1} \otimes c_2 \end{bmatrix},$$

where $C_{21} = S_2 F S_1'$, $c_1 = S_1 c$, $c_2 = S_2 c$, and i_{n-1} denotes a unit vector of dimension $n-1 \times 1$. The dimensions of these matrices and vectors are as follows:

$$M : n(p-1) \times np(p-1); \qquad M_j : n(p-1) \times n(p-1) \qquad (j = 1, \ldots, p);$$

$$\hat{M} : n(p-1) \times n(p-1); \qquad M_{1j} : n(p-1) \times n \qquad (j = 1, \ldots, p-1);$$

$$\widetilde{M} : n(p-1) \times n(p-1)^2;$$

$$N : np(p-1) \times np; \qquad N_j : np(p-1) \times n \qquad (j = 1, \ldots, p);$$

$$\bar{c} : np(p-1) \times 1; \qquad c_j : n \times 1 \qquad (j = 1), \qquad n(p-1) \times 1 \qquad (j = 2).$$

The matrix $\hat{M}$ is nonsingular under Assumptions 1–3; see Lemma A.1 of Chambers (1999).

Proof of Theorem 1. Premultiplying (11) by S_1 and noting that $S_1' S_1 + S_2' S_2 = I_{np}$ yields $x_t = S_1 c + S_1 C(S_1' S_1 + S_2' S_2) Y_{t-1} + S_1 v_t$, which can be written

$$x_t = c_1 + C_{11} x_{t-1} + C_{12} w_{t-1} + S_1 v_t, \qquad \textbf{(A.1)}$$

whereas premultiplication of (10) by S_2 yields, in a similar fashion,

$$w_t = c_2 + C_{21} x_{t-1} + C_{22} w_{t-1} + S_2 v_t. \qquad \textbf{(A.2)}$$

Elimination of w_{t-1} from (A.1) follows the same steps as in the proof of Theorem 1 of Chambers (1999) and hence results in the stated expressions for the autoregressive matrices in the discrete time representation. In dealing with the disturbance term note that we can write

$$\begin{aligned}
v_t &= \int_{t-1}^{t} \int_{r-1}^{r} C(r-s) \Theta u(s) ds dr \\
&= \int_{t-1}^{t} \left[\int_{s}^{t} C(r-s) dr \right] \Theta u(s) ds + \int_{t-2}^{t-1} \left[\int_{t-1}^{s+1} C(r-s) dr \right] \Theta u(s) ds \\
&= \int_{t-1}^{t} \Gamma_1(t-s) \Theta u(s) ds + \int_{t-2}^{t-1} \Gamma_2(t-1-s) \Theta u(s) ds \\
&= \xi_{1t} + \xi_{2,t-1},
\end{aligned}$$

where we have used (14), (16), and Lemma 1 of Chambers (1999). Theorem 3 of Chambers (1999) then provides the MA(p) representation

$$\zeta_t = K_0 \xi_t + \cdots + K_p \xi_{t-p}, \qquad t = p+1, \ldots, T,$$

where $\xi_t = [\xi_{1t}', \xi_{2t}']'$. ■

Proof of Corollary 1. Applying the same procedure as in Theorem 1 to (10) we obtain

$$x_t = c_1 + C_{11} x_{t-1} + C_{12} w_{t-1} + S_1 \epsilon(t), \qquad \textbf{(A.3)}$$

$$w_t = c_2 + C_{21} x_{t-1} + C_{22} w_{t-1} + S_2 \epsilon(t). \qquad \textbf{(A.4)}$$

The elimination of w_t proceeds in the same way, and hence the autoregressive matrices in the discrete time representation are identical to those in Theorem 1. Theorem 3 of Chambers (1999) enables the disturbance vector to be written

$$\eta_t = C_0\epsilon(t) + \cdots + C_{p-1}\epsilon(t-p+1), \qquad t = p+1, \ldots, T, \tag{A.5}$$

which follows by replacing his ξ_t with the vector $[\epsilon(t)', 0']'$. ■

Proof of Corollary 2. Proceeding as in Theorem 1 we obtain, from (11) and with S_1 and S_2 defined in (12) rather than in (9),

$$X_t = c_1 + C_{11}X_{t-1} + C_{12}W_{t-1} + S_1 v_t, \tag{A.6}$$

$$W_t = c_2 + C_{21}X_{t-1} + C_{22}W_{t-1} + S_2 v_t. \tag{A.7}$$

The elimination of W_t follows the same steps as in Theorem 1, and the MA representation for the disturbance is obtained as in Theorem 1 but with the different definition of S_1 and S_2 being reflected in the use of B_j instead of K_j $(j = 1, \ldots, p)$ in the MA representation. ■