Department of Economics

The Likelihood of a Continuous-time Vector Autoregressive Model

J. Roderick McCrorie

Working Paper No. 419 October 2000 ISSN 1473-0278

Queen Mary
University of London

THE LIKELIHOOD OF A CONTINUOUS-TIME VECTOR AUTOEGRESSIVE MODEL

J. RODERICK McCRORIE

Queen Mary and Westfield College

October 2000

ABSTRACT

This paper provides a method that weakens conditions under which the exact likelihood of a continuous-time vector autoregressive model can be derived. In particular, the method does not require the restrictions extant methods impose on discrete data that limit the applicability of continuous-time methods to real economic time series. The method applies generally to higher-order continuous-time systems involving mixed stock and flow data.

Key words: Continuous-time; vector autoregression; exact likelihood; time series
JEL classification: C32

Acknowledgements. The research was partially supported by the Economic and Social Research Council under grant R0042934211. I thank Lucien Foldes, David Fremlin, and Peter Robinson for comments and advice on a previous version. I remain solely responsible for all errors.

Address correspondence to:- Dr Roderick McCrorie, Department of Economics, Queen Mary and Westfield College, University of London, London, E1 4NS, U.K. Tel: +44 (20) 7882 5095; fax +44 (20) 8983 3580; e-mail: j.r.mccrorie@qmw.ac.uk

1. INTRODUCTION

Owing to developments in computer technology, it has become practical to estimate econometric models formulated in continuous-time, on the basis of discrete data. Common in modelling high-frequency financial data, this approach can also be useful in the analysis of macroeconomic time series, especially when *a priori* information is to be imposed on the distribution of the data and when the data themselves are generated in finer time intervals than the sampling interval. Studies of the temporal aggregation bias arising from equating the data generating interval with the sampling interval, when the former is in fact finer, invariably show that parameter estimates are distorted by the generation of spurious Granger causality relationships and serial persistence in the data.[1] This reflects the lack of time-invariance of discrete-time models. For example, if monthly observations of a certain variable satisfy a second-order autoregressive model, then quarterly observations *of the same variable* satisfy an autoregressive moving-average model. Although materially affecting statistical inference, this aspect is seldom appreciated in applied work.

The main reason in formulating the econometric model in continuous-time is that it allows us to tighten the link between theory and estimation by directing estimation towards the parameters of agents' objective functions rather than just towards the behaviour of the observations. Recognising that economic agents make decisions in finer time intervals than the sampling interval, we impose *a priori* restrictions on a continuous-time model as a means of accurately translating them to the distribution of the data. Although the approach offered below is more general, we focus on the continuous-time vector autoregressive (VAR) model which, like its discrete-time counterpart, has been popular in practice. This is largely because it generates discrete

[1] See Christiano and Eichenbaum (1985) and the references therein.

data that satisfy an *exact* discrete-time analogue: see Bergstrom (1996, 1997), Harvey and Stock (1993), Phillips (1991), and Robinson (1993). The purpose of this paper is to outline and formally justify the method by McCrorie (2000a), which allows the exact likelihood of a continuous-time VAR model to be computed without by itself entailing restrictions on the data that are capable of being rejected by a statistical test. The method involves deriving the covariance matrix required to compute the likelihood *via* a change in the order of three types of integration. The contribution is important because other methods *do* entail restrictions on the data that limit the applicability of continuous-time methods in practice. The time domain methods introduced by Bergstrom (1983) and Harvey and Stock (1985) require a steady-state assumption that does not rule out unit root processes *per se* but otherwise requires the variables to be transformed using *a priori* knowledge about the integration properties of the data and the dimension of the cointegration space. Phillips (1991) allows for observable stochastic trends but requires that certain time series are known to be co-integrated. The frequency domain approach by Robinson (1993) is motivated by the theory of stationary processes. Grossman, Melino, and Shiller (1987) do not require stationarity but their method relates to first-order models involving only flow variables and restricts the innovations to be Brownian motion. The last restriction, that the increments of the disturbance process are normally distributed, implies sample paths that are almost all continuous and is often not appropriate in econometric work. Our method relaxes all of these restrictions and, though pertaining to the usual stochastic differential equation system based on random measure, relies only on a technique from the ordinary differential equations literature. In this sense, the paper provides both a simplifying and unifying role that helps nullify the complexity of using continuous-time as compared with discrete-time models.

In the sequel, E will denote the expectation operator, and D the mean square differential operator with respect to continuous time. Stochastic integrals will be defined over intervals that are left-open and right-closed and the value of the integral at time zero will be 0.

2. THE MODEL

Consider a continuous-time vector autoregression in variables $y(t)$. The proposed method is general as it applies to the state-space form of the model: the (heuristic) system of first-order equations in the original variables and their derivatives

$$d\,\overline{y}(t) = \overline{A}(\theta)\overline{y}(t)\,dt + \overline{\zeta}(dt) \qquad (t \geq 0), \tag{2.1}$$

subject to the fixed initial conditions

$$\overline{y}(0) = y_0, \tag{2.2}$$

where $\{\overline{y}(t),\ t > 0\}$ is a real n-dimensional continuous time stochastic process of finite variance, $\overline{A}$ is an $n \times n$ matrix whose elements are known functions of a p-dimensional vector θ of unknown structural parameters, y_0 is an n-dimensional non-random vector, and $\overline{\zeta}(dt)$ is a white noise innovation vector defined precisely by the following assumption.[2]

Assumption 1. $\overline{\zeta}(\Delta)$ is an n-dimensional vector of random measures whose components are defined on $]0, \infty[$ such that every Borel subset Δ of $]0, \infty[$ is measurable and

(i) $E\,[\overline{\zeta}(\Delta)] = 0$;

(ii) for Borel subsets Δ_1 and Δ_2 of $]0, \infty[$,

[2] A formal definition of white noise is required owing to the fact that there is no wide-sense stationary process that is wide-sense integrable, whose integrals over every pair of disjoint integrals are uncorrelated. The approach now common in the literature was proposed by Bergstrom (1983) using the concept of random measure discussed by Rozanov (1967): see Bergstrom (1984) for a discussion.

$$E\,[\overline{\zeta}(\Delta_1)\ \overline{\zeta}'(\Delta_2)] = \lambda(\Delta_1 \cap \Delta_2)\ \overline{\Sigma}(\mu),$$

where λ is Lebesgue measure and $\overline{\Sigma}(\mu)$ is an unknown positive semi-definite matrix whose elements are known functions of a q-dimensional vector μ of unknown parameters ($q \leq n(n+1)/2$).

Under Assumption 1, $\overline{y}(t)$ is not mean square differentiable, and so (2.1) should be interpreted as representing the integral equation

$$\overline{y}(t) - \overline{y}(0) = \int_{]0,t]} \overline{A}\,\overline{y}(r)dr + \int_{]0,t]} \overline{\zeta}(dr) \qquad (t > 0), \tag{2.3}$$

where the first integral is defined in the wide sense (see Bergstrom (1984, p. 1152)) and $\int_{]0,t]} \overline{\zeta}(dr) = \overline{\zeta}(]0, t])$. Higher order systems reduce to the form of (2.3) in the original variables and their derivatives such that $\overline{A}$ is a block companion matrix and $\overline{\Sigma}$ is a block partitioned matrix all but one of whose blocks is zero. For example, the prototypical second-order model treated by Bergstrom (1986) follows by taking

$$\overline{y}'(t) = [y'(t) : Dy'(t)],\ \overline{A} = \begin{bmatrix} 0 & I \\ A_2 & A_1 \end{bmatrix},\ \overline{\zeta} = \begin{bmatrix} 0 \\ \zeta \end{bmatrix},\ \overline{\Sigma} = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma \end{bmatrix}, \tag{2.4}$$

where ζ is a vector and A_1, A_2 and Σ are matrices of appropriate dimensions. The technical requirement that all the variables possess mean-square derivatives of sufficient order with respect to a given order of system is overcome using a mixed-order model. McCrorie (2000c) has shown how to extend the method to continuous-time VARMA models, thereby including the closed models by Zadrozny (1988) and Robinson (1993). Unobservable stochastic trends can be treated by appending to the system another (usually first-order) differential equation. Deterministic trends including a constant vector, as polynomials in the time variable, can be treated exactly in an open system as elements of a vector of exogenous variables: see McCrorie

(2000b). In the interest of clarity, we shall outline the method using only (2.1) and (2.4): the unessential details are provided by McCrorie (2000a). The formal justification underlying the method is contained in the Appendix.

3. THE LIKELIHOOD

This section outlines the algorithm to compute the exact likelihood of the model.

3.1. *Defining observable vectors*

Consider the vector $y(t)$ that contains the levels of the variables ordered such that

$$y(t) = \begin{bmatrix} y^s(t) \\ y^f(t) \end{bmatrix}, \tag{3.1}$$

where $y^s(t)$ is a vector of stock variables observed at points in time ($t = 0, 1, 2, \ldots, T$) and $y^f(t)$ is a vector of flow variables observed as integrals over the intervals $]t-1, t]$ ($t = 1, 2, \ldots, T$). Here, we need to decide whether to compute the likelihood using the Kalman-Bucy filter, as introduced by Jones (1981) and Harvey and Stock (1985), or using the exact discrete analogue approach introduced by Bergstrom (1983).[3] This is because our method (applicable to both cases) relies on an integration that in the latter case requires the stock variables to be defined as first differences, or equivalently as the integral over $]t-1, t]$ of $Dx^s(t)$, in order that expressions are obtained for *both* stock and flow data. Observable vectors in the latter case are defined by the initial stock vector $y^s(0)$ and the vectors

$$y_t = \begin{bmatrix} y^s(t) - y^s(t-1) \\ \int_{]t-1,t]} y^f(r)dr \end{bmatrix} \qquad (t = 1, 2, \ldots, T). \tag{3.2}$$

[3] McCrorie (2000d) has compared the two approaches in a continuous-time model with unobservable stochastic trends. The former method treats unobservable variables without requiring an assumption to eliminate them and so the issue of defining observable stock variables as differences does not arise. Based on the Kalman-Bucy filter, it conveniently treats missing data, non-equispaced data, and errors

Taking the latter approach, if only because it is less well known, we equate the observation interval with the unit interval and derive the exact discrete analogue on the basis of the sequence $y_1, y_2, \dots, y_T$ having been generated by (2.1).

3.2. *The solution of the model*

The existence and uniqueness theorem established by Bergstrom (1983, Theorem 1) states that under Assumption 1 the solution of (2.1) subject to (2.2) is given by

$$\bar{y}(t) = e^{t\bar{A}}\,\bar{y}(0) + \int_{]0,t]} e^{(t-r)\bar{A}}\,\bar{\zeta}(dr) \qquad (t > 0), \tag{3.3}$$

where, for any square matrix A, $e^{A} = I + \sum_{r=1}^{\infty} (r!)^{-1} A^{r}$. The solution has the same shape as its ordinary-differential-equation analogue. Subtracting (3.3) lagged by one period yields

$$\bar{y}(t) = e^{\bar{A}}\,\bar{y}(t-1) + \int_{]t-1,t]} e^{(t-r)\bar{A}}\,\bar{\zeta}(dr). \tag{3.4}$$

3.3. *The problem and its solution*

The natural approach now would be to integrate (3.4) over the interval $]t{-}1, t]$ and solve the resulting system to obtain the exact discrete analogue of the continuous-time model in terms of the observable vectors $y_1, y_2, \dots, y_T$, and a disturbance vector that has the form of a double integral. (McCrorie, 2000c, has shown that essentially the same expression is integrated to derive the Kalman-Bucy filter.) The covariance matrix of the state innovation vector will have the form of a triple integral whose derivation, as explained in the Appendix, is non-standard. Once justified, however, it can be treated in principle by the method of Grossman, Melino, and Shiller (1987) for first-order models with flow data since we have *integrated* the solution of the

---

of measurement. On the other hand, it is computationally *less* efficient than the latter (provided the sample size is large enough to justify the fixed set up cost of deriving the exact discrete analogue).

continuous-time model in state-space form (namely, a *first*-order equation). The only essential difference is that our results pertain under Assumption 1.

3.4. *Deriving the exact discrete analogue*

Integrating (3.4) over $]t-1, t]$, then, yields

$$\int_{]t-1,t]} \overline{y}(r)\,dr \;=\; \int_{]t-2,t-1]} e^{\overline{A}}\,\overline{y}(r)\,dr \;+\; \int_{]t-1,t]} \int_{]\tau-1,\tau]} e^{(\tau-r)\overline{A}}\,\overline{\zeta}(dr)\,d\tau\,, \tag{3.5}$$

the existence of the double integral guaranteed by Lemma A-4 in the Appendix. Note that we have introduced an order of autocorrelation to the system: the double integral is affected by $\overline{\zeta}(dt)$ over *two* observation periods $]t-2, t]$. The coefficient matrices of the exact discrete model can now be obtained by eliminating the unobservable vectors in the state-space representation under a rank condition: see McCrorie (2000a) for details. It is convenient to multiply (3.5) by the matrix P that permutes the state vector in such a way that its first elements are y. As P is orthogonal, we have

$$P\int_{]t-1,t]} \overline{y}(r)\,dr \;=\; (Pe^{\overline{A}}P')P\int_{]t-2,t-1]} \overline{y}(r)\,dr \;+\; P\int_{]t-1,t]} \int_{]\tau-1,\tau]} e^{(\tau-r)\overline{A}}\,\overline{\zeta}(dr)\,d\tau. \tag{3.6}$$

Suppose

$$P = \begin{bmatrix} S \\ W \end{bmatrix}, \qquad\qquad P\,e^{\overline{A}}P' \;=\; \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \tag{3.7}$$

where S is a matrix that selects y and C_{11}, C_{12}, C_{21}, and C_{22} are by construction matrices of the same name in Bergstrom (1986). Equation (3.6) is seen to involve a pair of equations in both observable and unobservable variables. If C_{12} is non-singular, we can then obtain the main part of the exact discrete analogue:

$$y_t = F_1\,y_{t-1} + F_2\,y_{t-2} + \eta_t \qquad (t = 3, \ldots, T), \tag{3.8}$$

where

$$F_1 = C_{11} + C_{12}C_{22}C_{12}^{-1}, \quad (3.9)$$

$$F_2 = C_{12}(C_{21} - C_{22}C_{12}^{-1}C_{11}), \quad (3.10)$$

$$\eta_t = u_{1t} + C_{12}[-C_{22}C_{12}^{-1} : I]\begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \end{bmatrix}, \quad (3.11)$$

$$u_{1t} = S\gamma_t, \quad u_{2t} = W\gamma_t, \quad (3.12)$$

and γ_t is the double integral in (3.5). To construct the likelihood, we need also to derive supplementary equations relating y_1 and y_2 to the initial state vector. These are given by

$$y_1 = G_1\, y(0) + \eta_1, \quad (3.13)$$

$$y_2 = C_{11}\, y_1 + G_2\, y(0) + \eta_2, \quad (3.14)$$

where

$$G_1 = S \int_{]0,1]} e^{r\bar{A}} dr, \quad (3.15)$$

$$G_2 = C_{12}\, W \int_{]0,1]} e^{r\bar{A}} dr, \quad (3.16)$$

$$\eta_1 = u_{11} = S\gamma_1, \quad (3.17)$$

$$\eta_2 = u_{12} + C_{12}\, u_{21} = S\gamma_2 + C_{12}W\gamma_1. \quad (3.18)$$

3.5. *The covariance matrix of the state innovation vector*

The exact discrete analogue is described in its most general form by (3.8), (3.11) and (3.12). In order to derive a compact form for the covariance matrix Ω of the $nT \times 1$ innovation vector $\eta = [\eta_1', \eta_2', \ldots, \eta_T']'$, we use the decomposition

$$\gamma_t = \int_{]t-1,t]} \int_{]\tau-1,\tau]} e^{(\tau-r)\bar{A}} \bar{\zeta}(dr)\, d\tau$$

$$= \gamma_{0,t} + \gamma_{1,t-1}, \quad (3.19)$$

where

$$\gamma_{0,t} = \int_{]t-1,t]} \int_{]r,\,t]} e^{(\tau-r)\bar{A}} \bar{\zeta}(dr)\, d\tau\,, \tag{3.20}$$

$$\gamma_{1,t-1} = \int_{]t-2,t-1]} \int_{]t-1,r+1]} e^{(\tau-r)\bar{A}} \bar{\zeta}(dr)\, d\tau\,. \tag{3.21}$$

The above decomposition is useful for deriving a moving average representation of the vectors $\eta_1, \ldots, \eta_T$ in terms of $[\gamma'_{0,t} \quad \gamma'_{1,t}]'$, and when applying Assumption 1 because $]t-2, t-1]$ and $]t-1, t]$ are disjoint. The results in the Appendix support the following theorem which contains the information (up to multiplication by a known matrix) to derive the autocovariance matrices of the disturbance vectors $\eta_1, \ldots, \eta_T$. This exploits the fact that the state-space form of the exact discrete analogue is a VARMA (1, 1) model.

Theorem 1. (*McCrorie*, *2000a*) *Let* $\xi'_t = [\gamma'_{0,t} \quad \gamma'_{1,t}]$ (t = 1, 2, . . . , T), *where* $\gamma_{0,t}$ *and* $\gamma_{1,t}$ *are defined by* (3.20) *and* (3.21). *Then, under Assumption 1,*

$$E(\xi_s \xi'_t) = \delta_{s,t}\,\Gamma, \tag{3.22}$$

where $\delta_{s,t} = 1$ *if* $s = t$, 0 *otherwise*,

$$\Gamma = \begin{bmatrix} \Gamma_{00} & \Gamma'_1 \\ \Gamma_1 & \Gamma_{01} \end{bmatrix}, \tag{3.23}$$

$$\Gamma_{00} = \int_{]0,1]} \int_{]0,s]} \int_{]0,s]} e^{u\bar{A}}\, \bar{\Sigma} e^{v\bar{A}'}\, du dv ds\,, \tag{3.24}$$

$$\Gamma_{01} = \int_{]0,1]} \int_{]s,1]} \int_{]s,1]} e^{u\bar{A}}\, \bar{\Sigma} e^{v\bar{A}'}\, du dv ds\,, \tag{3.25}$$

$$\Gamma_{1} = \int_{]0,1]} \int_{]s,1]} \int_{]0,s]} e^{u\bar{A}}\, \bar{\Sigma} e^{v\bar{A}'}\, du dv ds\,. \tag{3.26}$$

It follows that, for $t > 1$,

$$E(\gamma_t \gamma'_t) = E(\gamma_{0,t}\gamma'_{0,t}) + E(\gamma_{1,t-1}\gamma'_{1,t-1}) = \Gamma_{00} + \Gamma_{01}, \tag{3.27}$$

$$E(\gamma_t \gamma'_{t-1}) = E(\gamma_{1,t-1}\gamma'_{0,t-1}) = \Gamma_1. \tag{3.28}$$

McCrorie (2000a) has shown in addition that Γ_{00}, Γ_{01}, and Γ_1 can be expressed in terms of submatrices of the exponential of a certain block-triangular matrix, a result that considerably facilitates computing them.

3.6. *Computing the exact likelihood*

The proposed method of integrating the solution of the continuous time model yields an especially parsimonious form for the likelihood. We can immediately derive moving-average representations for the vectors $\eta_1, \ldots, \eta_T$ in terms of the ξ_t of Theorem 1 and then apply Assumption 1 via (3.22) to obtain the autocovariance matrices of $\eta_1, \ldots, \eta_T$. If we define

$$\Omega_{t,s} = E(\eta_t \eta_s'), \ \Omega_s = E(\eta_t \eta_{t-s}'), \qquad (3.29)$$

the non-zero submatrices of the covariance matrix of the state innovation vector are given by

$$\Omega_{11} = U_0 \Gamma U_0', \qquad (3.30)$$

$$\Omega_{21} = U_1 \Gamma U_0', \quad \Omega_{22} = \Omega_{11} + U_1 \Gamma U_1, \qquad (3.31)$$

$$\Omega_{31} = V_2 \Gamma U_0', \quad \Omega_{32} = \Omega_{31} + V_1 \Gamma U_0', \qquad (3.32)$$

$$\Omega_{42} = \Omega_{31}, \qquad (3.33)$$

$$\Omega_0 = U_0 \Gamma U_0' + V_1 \Gamma V_1' + V_2 \Gamma V_2', \qquad (3.34)$$

$$\Omega_1 = V_1 \Gamma U_0' + V_2 \Gamma V_1', \qquad (3.35)$$

$$\Omega_2 = V_2 \Gamma U_0', \qquad (3.36)$$

$$\Omega_j = 0 \quad (j > 2), \qquad (3.37)$$

where U_0, U_1, V_1, and V_2 are the coefficient matrices in the moving-average representations for η_1, η_2, and η_t ($t = 3, \ldots, T$) given explicitly in McCrorie (2000a). The autocovariance matrices are expressed in a closed form, and not in the usual integral form that arises when using the prototypical method by Bergstrom (1986).

For illustrative purposes, consider minus twice the logarithm of the Gaussian likelihood function[4] less a constant:

$$L(\theta, \mu, y') = \ln\left|\Omega(\theta, \mu)\right| + \eta'\Omega^{-1}(\theta, \mu)\eta, \tag{3.38}$$

where $"\left| \; \right|"$ is the determinant operator. A simpler form of this function, that exploits the sparseness of Ω, follows from the Cholesky factorization $\Omega = QQ'$, where Q is a lower triangular matrix with positive elements along the diagonal, whose submatrices can be obtained recursively. Define the vector ε by $Q\varepsilon = \eta$ so that $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = I$. This facilitates expressing the exact discrete model given by (3.8), (3.13) and (3.14) as a VARMA model whose moving-average coefficient matrices are time dependent but converge to constant matrices as $t \to \infty$: see Bergstrom (1990, Chapter 7, Theorem 1) and the comment thereafter. As Bergstrom (1997, p. 483) notes, unit roots in the MA process do not prevent this convergence. This leads to the simpler form

$$L(\theta, \mu, y') = \sum_{i=1}^{T}(\varepsilon_i^2 + 2\ln q_{ii}), \tag{3.39}$$

where ε_i is the i-th element of the nT-vector ε (whose elements can be evaluated recursively from $Q\varepsilon = \eta$) and q_{ii} is the i-th diagonal element of Q. Note that the inversion of the $nT \times nT$ matrix Ω has been circumvented. It is worth emphasising that (3.39) is indeed the (essential part of the) Gaussian likelihood function that *exactly* incorporates the restrictions of economic theory, in contrast to the frequency domain approximations proposed by Robinson (1993). In the usual way, we can use (3.39) to define the exact Gaussian estimator as

$$[\hat{\theta}, \hat{\mu}, \hat{y}'] = \arg\min_{[\theta, \mu, y']} L(\theta, \mu, y'). \tag{3.40}$$

---

[4] We explicitly use as an example the Gaussian likelihood function, i.e. the likelihood function that would be obtained if η were a $N(0, \Omega)$ random vector, although we do not assume η has this property. It is a function of the vector of parameters $[\theta, \mu]$ and the unobservable part of the initial state vector y.

4. CONCLUSION

This paper has outlined a method of computing the exact likelihood of a continuous-time vector autoregressive model without entailing the usual restrictions on discrete data that could be rejected by a statistical test. This is important because the restrictions used in rival methods have tended to limit the application of continuous-time methods to real economic time series. Once the technical results in the Appendix are given, we rely only on a technique used when considering *ordinary* differential equations, integrating the solution of the model, as a first step towards computing the likelihood. Observable vectors are defined such that the method yields expressions for both mixtures of stock and flow data in general. The exact discrete analogue, and in particular the elements of the covariance matrix, are expressed in an especially compact form and our expression for the Gaussian likelihood function exploits the sparseness of this covariance matrix. While we have focused on a prototypical model and have used an approach based on an exact discrete analogue, McCrorie (2000c) has shown that the same technique can be used for more general models and in constructing the Kalman-Bucy filter. It is hoped that this paper has made researchers aware of the possibilities of using econometric models formulated in continuous-time and, in particular, of their potential for imposing the restrictions of economic theory on the probability distribution of the data.

APPENDIX

The basic technical problem in justifying the method of this paper arises because the definition of white noise given in Assumption 1 leads to a stochastic integral whose properties are not immediately justified by an appeal to standard theorems. The definition based on random measure was constructed for econometric models by Bergstrom (1983, 1984) as the analogous definition to uncorrelated errors in discrete-time. This means in particular that theorems that require additional conditions than ours on the first and second moments, or that are based on processes with independent increments, are not applicable in general. Here, under Assumption 1, the covariance matrix relating to the disturbance term in (3.5) has the form of a triple integral whose derivation requires a change in the order of three types of integration:- the integration of a measurable function with respect to a random measure; the integration in the wide sense of a stochastic process of finite variance with respect to time; and an integration over the probability space to obtain expected values. (See Bergstrom, 1984, for a detailed construction of these integrals.) The non-standard change between the first and second types has been an outstanding problem in the continuous-time literature, and indeed the method by Bergstrom (1983), which relies on an additional assumption that can rule out unit root processes, was designed precisely to bypass this problem. Harvey and Stock (1988, p. 372), when deriving the Kalman-Bucy filter, also recognised such an interchange was non-standard but used it as "a heuristic device to obtain simplified expressions for evaluating the covariance matrix". It has also been used informally by other authors: see McCrorie (2000a).

Here, we establish the interchange as a multidimensional generalization of Rozanov (1967, Theorem 2.4, p. 12), setting up the problem so as to permit the standard application of Fubini's theorem. McCrorie (2000a) has offered an argument outside the paradigm of random measure, which is based on redefining and modifying the stochastic integral. The argument is in part heuristic, complicates the remainder of the proof of Theorem 1, and is less tractable for more complicated correlation structures. Otherwise, McCrorie's (2000a) approach to justifying Theorem 1 is entirely valid.

Let $(\Omega, \Im, P)$ be a probability space, let $X \subseteq\ ]0, T]$ be a half-open interval, and write L^2 for $L^2(\Omega, \Im, P)$, the space of (equivalence classes of) random variables of finite variance. A function $f : X \to L^2$ will be called *measurable* if the inverse image $f^{-1}[G]$ is Borel measurable for all open sets G, and *measurable in the wide sense* if the scalar function $t \mapsto \mathrm{E}\,(f(t) \times h)$: $X \to \mathrm{R}$ is Borel measurable for all random variables h of finite variance. The measurability concepts are extended to the multidimensional case on an element-by-element basis.

Let $(L^2)^n$ represent n copies of L^2 and let $L(\mathrm{R}^n, \mathrm{R}^n)$ be the space of (bounded) linear operators from R^n to R^n represented by the real $n \times n$ matrices. For $u \in (L^2)^n$, set

$$\|u\| = \sqrt{\sum_{i=1}^{n} \|u_i\|_{L^2}^2} \qquad (A.1)$$

where L^2 is given its usual norm:

$$\|u_i\|_{L^2} = \sqrt{\mathrm{E}|u_i|^2}\ . \qquad (A.2)$$

For $S \in L(\mathrm{R}^n, \mathrm{R}^n)$ and $v \in \mathrm{R}^n$, set

$$\|S\| = \sup_{\|v\|_2 \le 1} \|Sv\|_2\ , \qquad (A.3)$$

where R^n is given its Euclidean norm:

$$\|v\|_2 = \sqrt{\sum_{i=1}^{n} |v_i|^2}\ . \qquad (A.4)$$

Expression (A.3) defines the operator norm in $L(\mathrm{R}^n, \mathrm{R}^n)$ which, owing to the fact that any two norms on a finite-dimensional linear space are equivalent, is adopted for convenience. Matrices will be denoted by upper-case letters and their entries by corresponding lower-case letters.

The first lemma and corollary ensure that under Assumption 1 the set of values of random variables described by integrals of measurable functions with respect to the random measure is separable. Measurability is then sufficient for measurability in the wide sense, as a consequence of Rozanov (1967, Theorem 2.2, p. 9) which holds provided the elements of the covariance matrix are measurable. This condition holds if a (possibly vector-valued) function ϕ is measurable; for then $(t, s) \mapsto E\,(\phi(t)\phi'(s))$ is measurable as a function of two variables.

Lemma A-1. *Let* $X \subseteq \,]0, T]$. *Under Assumption* 1, $\{\bar{\zeta}(\Delta) : \Delta \in B(X)\}$ *is separable.*

Proof of Lemma A-1. Let H be the closed linear span of $\{\bar{\zeta}(]0, q]) : q \in \mathrm{Q}^{++}\}$, where Q^{++} is the set of strictly positive rational numbers. As Q^{++} is countable, H is separable.

Consider $\mathcal{A} = \{\Delta \in B(X) : \bar{\zeta}(\Delta) \in H\}$. Then $\mathcal{I} = \{]t, u] : t, u \in \mathrm{Q}^{++}, t \le u \le T\} \subseteq \mathcal{A}$

because $\bar{\zeta}(]t, u]) = \bar{\zeta}(]0, u]) - \bar{\zeta}(]0, t]) \in H$.

For $]t', u'],]t'', u''] \subseteq \mathcal{I}$,

$$]t', u'] \cap \,]t'', u''] = \,]\max(t', t''), \min(u', u'')] \in \mathcal{I}.$$

As $]t', u']$ and $]t'', u'']$ are arbitrary, $I \cap J \in \mathcal{I} \quad \forall\, I, J \in \mathcal{I}$.

The class $\mathcal{A}$ has the properties:-

(i) $X \in \mathcal{A}$;

(ii) if $\Delta, \Delta' \in \mathcal{A}$ and $\Delta \subseteq \Delta'$, then $\Delta' \setminus \Delta \subseteq B(X)$ and

$$\bar{\zeta}(\Delta' \setminus \Delta) = \bar{\zeta}(\Delta') - \bar{\zeta}(\Delta) \in H; \text{ hence } \Delta' \setminus \Delta \in \mathcal{A};$$

(iii) if $<\Delta_m>$ is a non-decreasing sequence in $\mathcal{A}$, $\Delta_m \uparrow \Delta$, and $c = \mathrm{tr}\, \bar{\Sigma}$,

$$\left\|\bar{\zeta}(\Delta) - \bar{\zeta}(\Delta_m)\right\|^2 = \sum_i E\left|\bar{\zeta}_i(\Delta) - \bar{\zeta}_i(\Delta_m)\right|^2 = \sum_i E\left|\bar{\zeta}_i(\Delta \setminus \Delta_m)\right|^2 = \sum_i \bar{\sigma}_{ii} \lambda(\Delta \setminus \Delta_m)$$

$= c\lambda(\Delta \setminus \Delta_m) \to 0$ as $m \to \infty$; on taking square roots, $\bar{\zeta}(\Delta) = \lim_{m\to\infty} \bar{\zeta}(\Delta_m) \in H$;

hence $\Delta \in \mathcal{A}$. ($\mathcal{A}$ is a monotone class.)

By the Monotone Class Theorem (see, e.g., Billingsley, 1995, p.43), $\mathcal{A}$ includes the σ-algebra of subsets of X generated by $\mathcal{I}$, namely the Borel σ-algebra. Hence, $\bar{\zeta}(\Delta) \in H \;\; \forall \Delta \in B(X)$. Lemma 1 holds because H is separable. $\|$

Corollary A-2. $\phi(\tau) = \int_{]\tau-1, \tau]} e^{(\tau - r)\bar{A}} \bar{\zeta}(dr) \in H$.

Proof of Corollary A-2. The existence of the integral is established by Bergstrom (1983, p. 123). Since $e^{(\tau-r)\bar{A}}$ is a bounded linear operator, the result is immediate. $\|$

The next lemma bounds the norm of the type of process considered above on the assumption that the integrand, viewed as a matrix-valued function of a scalar, is continuous (where continuity is interpreted entrywise).

Lemma A-3. *Suppose that* $a < b$ *in* $\mathbb{R}$ *and that* $t \mapsto B(t)$: $]a, b] \to L(\mathbb{R}^n, \mathbb{R}^n)$ *is continuous at every point in* $]a, b]$. *Then*, *under Assumption* 1,

$$\left\|\int_{]a,b]} B(r)\bar{\zeta}(dr)\right\| \leq \sqrt{c(b-a)}\,\sup_{r\in]a,b]}\|B(r)\|, \tag{A.5}$$

where $c = \mathrm{tr}\,\bar{\Sigma}$.

Proof of Lemma A-3. Since $B(t)$ is continuous, there exists a dissection of $]a, b]$ such that every element $b_{ij}(t)$ can be uniformly approximated to arbitrary degree by a sequence of simple functions. It follows, by a straightforward argument considering its norm, that under Assumption 1, $\int_{]a,b]} B(r)\bar{\zeta}(dr)$ can be approximated in mean-square by a sequence of simple functions to a degree that depends only on the fineness of the chosen dissection, and so exists by the definition of Rozanov (1967, p. 7).

Let $D = (r_0, r_1, \ldots, r_d)$, with $a = r_0 \leq r_1 \leq \ldots \leq r_d = b$, be a dissection of $]a, b]$, set $\Delta_m =]r_{m-1}, r_m]$, and define, using the indicator function,

$$B^D(r) = \sum_{m=1}^{d} B(r_m) 1_{\Delta_m}(r). \tag{A.6}$$

We have

$$\left\|\int_{]a,b]} B^D(r)\bar{\zeta}(dr)\right\|^2 = \sum_{i=1}^{n}\left\|\int_{]a,b]}\sum_{j=1}^{n} b_{ij}{}^D(r)\bar{\zeta}_j(dr)\right\|_{L^2}^2 = \sum_i \left\|\sum_m \sum_j b_{ij}^D(r)\zeta_j(\Delta_m)\right\|_{L^2}^2$$

$$= \sum_i \sum_m \left\|\sum_j b_{ij}^D(r_m)\bar{\zeta}_j(\Delta_m)\right\|^2 \text{ (under Assumption 1, } E[\bar{\zeta}_j(\Delta_m)\bar{\zeta}_{j'}(\Delta_{m'})] = 0 \text{ for } m \neq m')$$

$$=\sum_m \sum_i \left\| \sum_j b_{ij}{}^D(r_m)\overline{\zeta}_j(\Delta_m) \right\|_{L^2}^2 = \sum_m \left\| B^D(r_m)\overline{\zeta}(\Delta_m) \right\|^2 \le \sum_m \left\| B^D(r_m) \right\|^2 \left\| \overline{\zeta}(\Delta_m) \right\|^2 . \qquad (A.7)$$

Now

$$\left\| \overline{\zeta}(\Delta_m) \right\|^2 = \sum_i \left\| \overline{\zeta}_i(\Delta_m) \right\|_{L^2}^2 = \sum_i E\left| \overline{\zeta}_i(\Delta_m) \right|^2 = \sum_i \overline{\sigma}_{ii}\lambda(\Delta_m) = c\,\lambda(\Delta_m) . \qquad (A.8)$$

Hence

$$\sum_m \left\| \overline{\zeta}(\Delta_m) \right\|^2 \le c\ \lambda(]a,\, b]) = c\,(b-a). \qquad (A.9)$$

Letting $\|D\| = \max_{1\le m\le d} \left| r_m - r_{m-1} \right| \to 0$, so that $B^D(t) \to B(t)$ by continuity, gives

$$\left\| \int_{]a,b]} B(r)\overline{\zeta}(dr) \right\|^2 \le c\,(b-a)\sup_{r\in]a,b]} \left\| B(r) \right\|^2 . \qquad (A.10)$$

Expression (A.5) follows on taking square roots. ‖

The following lemma establishes the existence of the double integral in (3.5).

Lemma A-4. $\gamma_t = \int_{]t-1,t]} [\int_{]\tau-1,\tau]} e^{(\tau-r)\overline{A}}\overline{\zeta}(dr)]\, d\tau$ exists.

Proof of Lemma A-4. Since $e^{(\tau-r)\overline{A}}$ is continuous in r, $\phi(\tau) = \int_{]\tau-1,\tau]} e^{(\tau-r)\overline{A}}\overline{\zeta}(dr)$ exists as a random vector of finite variance. For $\rho \in\ ]\tau-1,\ \tau]$,

$$\phi(\tau) - \phi(\rho) = \int_{]\tau-1,\rho]} (e^{(\tau-r)\overline{A}} - e^{(\rho-r)\overline{A}})\overline{\zeta}(dr) - \int_{]\rho-1,\tau-1]} e^{(\rho-r)\overline{A}}\overline{\zeta}(dr)$$

$$+ \int_{]\rho,\tau]} e^{(\tau-r)\overline{A}}\overline{\zeta}(dr), \qquad (A.11)$$

and by Lemma A-3,

$$\left\| \int_{]\tau-1,\rho]} (e^{(\tau-r)\overline{A}} - e^{(\rho-r)\overline{A}})\,\overline{\zeta}(dr) \right\| \le \sqrt{c(\rho-\tau+1)}\ \sup_{r\in]\tau-1,\rho]} \left\| e^{(\tau-r)\overline{A}} - e^{(\rho-r)\overline{A}} \right\|$$

$$\le \sqrt{c}\,\sup_{r\in]\tau-1,\rho]} \left\| e^{(\tau-r)\overline{A}} \right\| \left\| I - e^{(\rho-\tau)\overline{A}} \right\| \to 0 \ \text{ as } \rho \uparrow \tau ;$$

$$\left\| \int_{]\rho-1,\tau-1]} e^{(\rho-r)\overline{A}} \overline{\zeta}(dr) \right\| \le \sqrt{c(\tau-\rho)} \sup_{r\in]\rho-1,\tau-1]} \left\| e^{(\rho-r)\overline{A}} \right\| \to 0 \text{ as } \rho \uparrow \tau,$$

and in a similar way the third integral in (A.11) converges to zero in norm as $\rho \uparrow \tau$.

By the triangle inequality,

$$\|\phi(\tau)-\phi(\rho)\| \le \left\| \int_{]\rho-1,\tau-1]} e^{(\rho-r)\overline{A}} \overline{\zeta}(dr) \right\| + \left\| \int_{]\tau-1,\rho]} (e^{(\tau-r)\overline{A}} - e^{(\rho-r)\overline{A}}) \overline{\zeta}(dr) \right\|$$

$$+ \left\| \int_{]\rho,\tau]} e^{(\tau-r)\overline{A}} \overline{\zeta}(dr) \right\| \to 0 \text{ as } \rho \uparrow \tau. \qquad \text{(A.12)}$$

Hence,

$$\lim_{\rho\uparrow\tau} \phi(\rho) = \phi(\tau) \;\; \forall\, \tau. \qquad \text{(A.13)}$$

As ρ is arbitrary, ϕ is left continuous in τ and so is Borel measurable. By Corollary A-2 and Rozanov (1967, Theorem 2.2, p.9), ϕ is measurable in the wide sense.

Set $M = \sqrt{c}\, \sup_r \left\| e^{(\tau-r)\overline{A}} \right\|$ in R. By Lemma A.3,

$$\|\phi(\tau)\| = \left\| \int_{]\tau-1,\tau]} e^{(\tau-r)\overline{A}} \overline{\zeta}(dr) \right\| \le \sqrt{c}\, \sup_{r\in]\tau-1,\tau]} \left\| e^{(\tau-r)\overline{A}} \right\| = M < \infty. \qquad \text{(A.14)}$$

As ϕ is bounded and measurable in the wide sense, by an obvious multidimensional generalisation of Rozanov (1967, Theorem 2.3, p.11), it is integrable in the wide sense on $]t-1, t]$ with respect to Lebesgue measure. ‖

Lemma A-5. $\overset{*}{\gamma}_t = \int_{]t-2,t]} \left(\int_{]\max(t-1,r),\min(t,r+1)]} e^{(\tau-r)\overline{A}} d\tau \right) \overline{\zeta}(dr)$ exists.

Proof of Lemma A-5. Since $e^{(\tau-r)\overline{A}}$ is continuous in τ in any bounded interval,

$$\varphi(r) = \int_{]\max(t-1,r),\, \min(t,r+1)]} e^{(\tau-r)\overline{A}} d\tau \text{ exists in } L(\mathrm{R}^n,\ \mathrm{R}^n).$$

The result follows on establishing that φ is a right-continuous function of r, and so is integrable over $]t-2, t]$ with respect to $\overline{\zeta}$. ‖

In order to show that $\gamma_t = \gamma^*_t$, a.s., we need only show that $E(\gamma_t \times \overline{h}) = E(\gamma^*_t \times \overline{h})$ for all test random variables $\overline{h}$ of finite variance. This allows us to apply Fubini's theorem in a conventional setting and to follow the argument by Rozanov (1967, pp. 12-13). Although Rozanov did not invoke Fubini's theorem *per se*, its use can be justified by a product measure argument.

Lemma A-6. $\gamma_t = \gamma_t^*$.

Proof of Lemma A-6. Let $\overline{h}$ be a random variable of finite variance and define the measure $F_h(dr) = E(\overline{\zeta}(dr) \times \overline{h})$. Then

$$E(\gamma_t \times \overline{h}) = E\int_{]t-1,\, t]} [\int_{]\tau-1,\, \tau]} e^{(\tau-r)\overline{A}}\,\overline{\zeta}(dr)]\, d\tau\; \overline{h}$$

$$= \int_{]t-1,\, t]} \int_{]\tau-1,\, \tau]} e^{(\tau-r)\overline{A}} F_h(dr)\, d\tau$$

$$= \int_{]t-2,\, t]} [\int_{]\max(t-1, r), \min(t,\, r+1)]} e^{(\tau-r)\overline{A}} d\tau]\, F_h(dr)$$

(by Fubini's theorem)

$$= E\int_{]t-2,\, t]} [\int_{]\max(t-1,\, r), \min(t,\, r+1)]} e^{(\tau-r)\overline{A}} d\tau]\, \overline{\zeta}(dr)\, \overline{h}$$

$$= E(\gamma_t^* \times \overline{h}). \qquad \text{(A.17)}$$

As $\overline{h}$ is arbitrary, $\gamma_t = \gamma_t^*$. ‖

REFERENCES

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 and M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 2 (North-Holland: Amsterdam), 1145-1212.

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: Oxford University Press).

BERGSTROM, A.R. (1996), “Survey of Continuous Time Econometrics”, in W.A. Barnett, G. Gandolfo and C. Hillinger (eds.), *Dynamic Disequilibrium Modeling* (Cambridge: Cambridge University Press), 3-25.

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.

BILLINGSLEY, P. (1995), *Probability and Measure*, 3rd edition (New York: Wiley).

CHAMBERS, M.J. (1999), “Discrete Time Representation of Stationary and Nonstationary Continuous Time Systems”, *Journal of Economic Dynamics and Control*, **23**, 619-639.

CHRISTIANO, L.J. and EICHENBAUM, M.S. (1987), “Temporal Aggregation and Structural Inference in Macroeconomics”, *Carnegie-Rochester Conference Series on Public Policy*, **26**, 63-130.

GROSSMAN, S.J., MELINO, A. and SHILLER, R.J. (1987), “Estimating the Continuous-Time Consumption-Based Asset-Pricing Model”, *Journal of Business and Economic Statistics*, **5**, 315-327.

HARVEY, A.C. and STOCK, J.H. (1985), “The Estimation of Higher-Order Continuous Time Autoregressive Models”, *Econometric Theory*, 1, 97-117.

HARVEY, A.C. and STOCK, J.H. (1993), “Estimation, Smoothing, Interpolation, and Distribution for Structural Time-Series Models in Continuous Time”, in P.C.B. Phillips (ed.), *Models, Methods and Applications of Econometrics* (Cambridge, MA: Blackwell), 55-70.

JONES, R.H. (1981), “Fitting a Continuous Time Autoregression to Discrete Data”, in D.F. Findley (ed.), *Applied Time Series Analysis II* (New York:Academic Press), 651-682.

McCRORIE, J.R. (2000a), “Deriving the Exact Discrete Analog of a Continuous Time System”, *Econometric Theory*, **16**, in press.

McCRORIE, J.R. (2000b), “Interpolating Exogenous Variables in Open Continuous Time Macroeconometric Models”, *Journal of Economic Dynamics and Control*, **24**, in press.

McCRORIE J.R. (2000c), “The Exact Gaussian Likelihood of a Continuous Time ARMA model”, preprint.

McCRORIE, J.R. (2000d), “Estimating a Continuous Time Macroeconometric Model”, preprint.

PHILLIPS, P.C.B. (1991), “Error Correction and Long-Run Equilibrium in Continuous Time”, *Econometrica*, **59**, 967-980.

ROBINSON, P.M. (1993), “Continuous Time Models in Econometrics: Closed and Open Systems, Stock and Flows”, in P.C.B. Phillips (ed.), *Models, Methods and Applications of Econometrics* (Cambridge, MA: Blackwell), 71-90.

ROZANOV, Y.A. (1967), *Stationary Random Processes* (Holden Day, San Francisco).

Van LOAN, C.F. (1978), “Computing Integrals Involving the Matrix Exponential”, *IEEE Transactions on Automatic Control*, **AC-23**, 395-404.

ZADROZNY, P. (1988), “Gaussian Likelihood of Continuous-Time ARMAX Models When Data are Stocks and Flows at Different Frequencies”, *Econometric Theory*, **4**, 108-124.

Queen Mary
University of London

This working paper has been produced by
the Department of Economics at
Queen Mary, University of London

Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096 or Fax: +44 (0)20 8983 3580
Email: j.conner@qmw.ac.uk
Website: www.econ.qmw.ac.uk/papers/wp.htm