CentER

Discussion Paper

No. 2004–39

GRANGER CAUSALITY AND THE SAMPLING OF ECONOMIC PROCESSES

By J.R. McCrorie, M.J. Chambers

April 2004

ISSN 0924-7815

TILBURG UNIVERSITY

Granger Causality and the Sampling of Economic Processes

J. Roderick McCrorie
University of Essex

and

Marcus J. Chambers
University of Essex

March 2004

Abstract: This paper provides a discussion of the developments in econometric modelling that are designed to deal with the problem of spurious Granger causality relationships that can arise from temporal aggregation. We outline the distortional effects of using discrete time models that explicitly depend on the unit of time and outline a remedy of constructing time-invariant discrete time models via a structural continuous time model. In an application to testing for money-income causality, we demonstrate the importance of incorporating exact temporal aggregation restrictions on the discrete time data. We do this by conducting causality tests in discrete time models that: (a) impose the temporal aggregation restrictions exactly; (b) impose the temporal aggregation restrictions approximately; and (c) do not impose these restrictions at all.

J.E.L. code: C32.

Keywords: Granger causality, temporal aggregation,

Acknowledgements: This paper is a revised version of a paper of the same title prepared for the $(EC)^2$ conference on *Causality and Exogeneity in Econometrics* held at CORE, Louvain-la-Neuve, December 2001. We thank an editor and two referees for helping to improve the focus, presentation and content of the original paper. McCrorie thanks the ESRC for partial support under award number R0042934211. His research was partially performed as a visiting fellow at CentER, Tilburg University. Chambers thanks the Leverhulme Trust for support provided by a Philip Leverhulme Prize.

Address for Correspondence: Dr. Roderick McCrorie, Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, U.K. Tel: +44 1206 872961; fax: +44 1206 872724; email: mccrorie@essex.ac.uk.

1. Introduction

Most models involving economic time series are fundamentally dependent on the unit of time and the observation interval. If agents' decision intervals do not coincide with the sampling interval, then inferences made about the behaviour of economic agents from observed time series can be distorted (see, e.g., Christiano and Eichenbaum, 1987; Marcellino, 1999). This distortion is known as 'temporal aggregation bias,' which can occur when observations are not collected frequently enough fully to capture the movements of economic variables. If our ultimate goal is to provide an *economic* interpretation of parameter estimates that relates to the behaviour of economic agents and not just to the behaviour of the observations, then taking account of the effects of temporal aggregation is important. This is because the time intervals between macroeconomic observations are typically much longer than the time intervals between the microeconomic decisions of economic agents that the observations reflect.

Even in a pure time series context, the lack of time invariance of discrete time models matters in general. For example, if monthly observations of certain variables satisfy (as common a model as) a vector autoregressive (VAR) model, then quarterly observations *of the same variables* can satisfy a vector autoregressive moving average (VARMA) model.[1] A coarsely sampled process, omitting information useful for predicting an economic time series, will exhibit bi-directional Granger causality with another sampled process in the coarser time interval provided they are correlated, even if there is only unidirectional causality in the finer time interval. This means that the observation of bi-directional Granger causality cannot constitute *prima facie* evidence that there is bi-directional causality in the data generating process relating to the behaviour of economic agents.

The purpose of this paper is to discuss the developments in formulating econometric models in continuous time as a means of dealing with the distortional effect of temporal aggregation bias in generating spurious Granger causality relationships among observed time series. Formulating an econometric model in continuous time offers a basis for importing causality restrictions to observed discrete data independently of the sampling interval, as a means of obtaining efficient estimates of the structural parameters of the model. Our main emphasis will be on how to deliver accurate restrictions to the data and on evaluating the importance of these restrictions for causality testing, although we shall also briefly consider the problem of defining what we mean by 'causality' in the context of continuous time models. To a large extent, we shall concentrate on continuous time VAR models because

[1] This can be true for stock variables, observed at points in time, such as the money stock and the rate of interest, and is certainly true for flow variables, measured as integrals over the observation interval, such as income and consumption.

discrete data generated by these models satisfy an *exact* discrete analogue.[2] In this context, restrictions on the underlying time series are imported exactly to the distribution of observed discrete data and, if the data are equispaced, independently of the rate of sampling. Here, we demonstrate in an application to testing for money-income causality that imposing restrictions that incorporate the effects of temporal aggregation is important for causality testing. We do this by conducting causality tests in discrete time models that incorporate the temporal aggregation restrictions exactly as well as approximately, and compare the results with discrete time models that do not impose any temporal aggregation restrictions. Throughout, in our discussions of continuous time models, we assume that the parameters of interest are identified from the discrete time data. We acknowledge, however, the results of McCrorie (2003) who showed that a sometimes stringent condition is required to identify the structural parameters on the basis of discrete data. A trade-off can potentially emerge between obtaining estimates that are robust to spurious Granger causality relationships against the imposition of identification restrictions on the structural continuous time model that would not ordinarily be provided by economic theory.

The paper is organized as follows. Section 2 contains a discussion to show the extent to which naïvely specified discrete time models can provide a distorted picture of Granger causality relationships and serial correlation properties in the data, and in a way that materially affects statistical inference and impacts upon applied work. In Section 3, with a view towards adopting a continuous time approach as a remedy to these problems, we briefly discuss and define what we mean by the concept of 'causality' or 'non-causality' in continuous time models. Our central focus, however, will not be on continuous time models *per se*, but on how these models form the basis of importing restrictions independently of the unit of time to discrete time analogues that have similar forms to the models discussed in Section 2. This approach is developed also in Section 3. In Section 4, we present an illustration of the issues in the context of testing for money-income causality. Section 5 concludes.

2. The distortional effects of naïvely using discrete time models

It is worth emphasizing that naïvely formulating discrete time models in the time unit that coincides with the collection of the data involves an approximation that carries a cost unless economic variables jump discretely at points-in-time coinciding with the endpoints of the observation interval. The practice is so common in applied work as to be taken for granted and so we begin by discussing the nature of the error involved in this type of approximation and the sense in which Granger causality relationships can get distorted.

[2] Examples of authors considering continuous time models and adopting the principle of using an exact discrete analogue include: Bergstrom (1997); Chambers (1999); Chambers and McCrorie (2003); Hansen and Sargent (1991b); Harvey and Stock (1989, 1993); McCrorie (2000); Phillips (1991); Renault, Sekkat and Szafarz (1998); Robinson (1993); and Yu and Phillips (2001).

Many results are available in the literature through examining the problem as one of fixed-interval time aggregation, where the analysis is predicated on the data being generated in some definite, finer time interval than the sampling interval. This approach is well known and continues to be an active avenue of research – see especially Marcellino (1999), and Breitung and Swanson (2002) and the references therein. The complementary approach of formulating models in continuous time has perhaps had less impact on applied work, although it can be advantageous in certain circumstances. Firstly, with a view towards testing for causality, it allows *a priori* information to be imported to the data independently of the rate of sampling, and in a way that does not affect Granger causality relationships. In the fixed-interval time aggregation approach, this is possible only if we know what the finer time interval is (and if we really *know* what the finer time interval is we could then correct for the effects of temporal aggregation along the lines followed by Marcellino, 1999, for VARIMA models.) Secondly, the continuous time approach makes arbitrary the time unit in the agents' decision rule, and in principle allows for decision intervals to vary across different economic agents.[3] Because results on the efficacy of discrete approximation to continuous time can be interpreted as limiting statements on replacing small-unit by large-unit discrete time, the framework enables a discussion of the effects of temporal aggregation without explicit reference to a time unit in which the data are generated. We discuss its effects using several models, as results can be model specific.

2.1. Distributed lag models

Sims (1971a,b) considered the effects of temporal aggregation by considering discrete approximation to continuous time using the univariate distributed lag model

$$y(t) = \int_{-\infty}^{\infty} b(s)x(t-s)ds + u(t), \quad (t \in \mathcal{R}), \tag{1}$$

where the endogenous variable $y(t)$, the exogenous variable $x(t)$, and the disturbance $u(t)$ are wide-sense stationary random processes. As $b(\cdot)$ may be a generalized function, (1) may involve discrete non-integer lags and derivatives of arbitrary order. The model is identified by the assumption that $u(t)$ and $x(t-s)$ are uncorrelated for all s. Estimating $b(\cdot)$ then requires two types of approximation: substituting discrete time for continuous time and substituting a finite-parameter model for (1). Sims isolates the effects of the former by considering the model that equispaced data generated by (1) exactly satisfy:

$$Y(t) = \sum_{s=-\infty}^{\infty} B(s)X(t-s) + U(t), \quad (t \in \mathcal{Z}), \tag{2}$$

[3]See Christiano and Eichenbaum (1987) for a discussion of temporal aggregation in a context where an economic time series is interpreted as the outcome of a well-specified dynamic equilibrium in which rational economic agents solve stochastic optimization problems.

where, for all $t \in \mathcal{Z}$, $Y(t) = y(t)$, $X(t) = x(t)$, and $U(t)$ is uncorrelated with $X(s)$ for all s. Sims showed that

$$B(t) = \int_{-\infty}^{\infty} b(s) r_x(t-s) ds, \tag{3}$$

where

$$r_x(t) = \sum_{s=-\infty}^{\infty} R_X^{-*}(s) R_x(t-s), \tag{4}$$

and R_x and R_X are the autocovariance functions of x and X, and R_X^{-*} is the inverse of R_X under convolution. Notice that the exact discrete analogue (2) does not depend on a particular choice of t, and so we can examine the effects of temporal aggregation including the generation of spurious causality relationships by considering the extent to which its lag structure resembles the lag structure of (1). The discrete lag distribution is a sampling, at unit time intervals, of the continuous lag distribution smoothed using the filter r_x. Sims deduces that there are two rough conditions that guarantee that $B(t)$ is close to a weighted average of $b(s)$ for s near t: that r_x must be small for $|t| > 1$, and must have integral near one. This can be the case if the independent variable x is locally smooth in the sense of not fluctuating too wildly but the situation is not always encouraging; for even if r_x has continuous derivative or the spectral mass of x is zero outside the Nyquist frequency, Sims showed that the 'side lobes' of r_x can be non-negligible, indicating that $B(\cdot)$ poorly represents $b(\cdot)$.

Geweke (1978) extended (3) and (4) to the multivariate setting and established a limiting result that helps explain the sense in which discrete time models specified with respect to the sampling interval can offer a good approximation. He essentially shows, subject to regularity conditions including that $b(\cdot)$ is an ordinary function, that

$$\sum_{s=-\infty}^{\infty} \|B(s) - \tau b(s\tau)\| \to 0 \quad \text{as} \quad \tau \to 0, \tag{5}$$

where $\|\cdot\|$ is the root-sum-of-squared-elements norm[4] and the continuous time process x and lag distribution $b(\cdot)$ are fixed as the time unit τ in the discrete model drops to zero. Intuitively, if the time unit is small enough compared with the rate of variation in the exogenous variables, the discrete time model should be adequate. Of course, there is no reason to believe that the relation between time series is truly specified by the model (1). Sims (1987) offers an example to show that if (1) were replaced by a model involving derivatives and x follows a second-order Markov process then, constraining the estimation equation to

[4]Norm convergence is required because the focus is on how well $B(s)$ approximates $\tau b(s\tau)$ *as a function* in the limit.

involve only positive lags of X, the limiting discrete time model has a form far different from an approximation based on the first difference operator. Other limitations of using the discrete time model arise because individual values of $B(t/\tau)$ are not precluded by (5) from not converging to corresponding $b(t)$ values. These include a monotone continuous time lag distribution not necessarily having a monotone discrete time analogue and, importantly for our purpose, a lack of a Granger causal relationship from y to x in (1) will not generally imply the same from Y to X in (2).[5] This is because $B(t)$ depends on $b(s)$ for all s, even if $b(s)$ is one-sided on the past and (3) implies $B(t) \neq 0$ for $t < 0$. Geweke demonstrates other distortional effects in the multivariate context, including what he calls 'contamination,' where each row of $B(t)$ confounds not just the corresponding row of $b(t)$ but *potentially all* its rows.

2.2. Vector autoregressive models

Hansen and Sargent (1991a) considered instead the circumstances under which the impulse-response function from the vector autoregression associated with a discretely sampled version of a continuous time process resembles the shape of the impulse-response function in continuous time. Marcet (1991) relaxed the assumption they made that the process of interest has a rational spectral density, and derived a useful characterization of the relation between the continuous time and discrete time Wold representations. Interest centres on a single vector y whose continuous time dynamics are summarized in the Wold representation

$$y(t) = \int_0^{\infty} a(u)\xi(t - du), \quad (t \in \mathcal{R}), \tag{6}$$

where the matrix function a is restricted to satisfy $a(u) = 0$ when $u < 0$ and ξ is a vector of orthonormal random measures.[6] The Wold representation of the sampled process is

$$Y(t) = \sum_{k=0}^{\infty} A_k \epsilon(t - k), \quad (t \in \mathcal{Z}), \tag{7}$$

where $A_k = 0$ for $k < 0$ and ϵ is a white noise vector. The substantive issues relate to how close the shape of A is to a and how ϵ is related to ξ. Marcet shows that the coefficients A_k in (7) – normalized so that $\epsilon(t)$ is the vector of one-step-ahead innovations in Y – are given by

$$A_k = \int_0^{\infty} a(u + k)c'(u)du \left(\int_0^{\infty} c(u)c'(u)du \right)^{-1}, \quad (k = 0, 1, \ldots), \tag{8}$$

[5] We shall formally discuss various concepts of causality in Section 3 but, for the moment, we can say in the context of two arbitrary time series Y_t and X_t that "Y_t causes X_t if we are able better to predict X_t using all available information than if the information apart from Y_t has been used" (Granger, 1969, p. 428).

[6] Random measures can be used to provide a formal definition of white noise in continuous time models that is analogous to the definition of uncorrelated errors in discrete time. See Bergstrom (1984) for a discussion.

where ϵ is related to ξ by

$$\epsilon(t) = \int_0^\infty c(u)\xi(t - du), \tag{9}$$

with $c(u) = a(u)$ on $0 \leq u < 1$. This means that when the function c is small on $[1, \infty)$, A_k is essentially an average of a on $[k, k+1]$ and in this sense A_k is a good approximation to a. Expression (9) is analogous to (3) in that the discrete parameters are obtained by applying a weighting function to the continuous-time parameters. And because the coefficients in the i-th row of A_k are affected by all the rows in a, there is also contamination in this model. The coefficients A_k will be contaminated even when the projections in continuous-time and discrete-time coincide, as can be seen by putting $c(u) = a(u)$ for all u in (8). Generally, only when $a_{ij} = 0$ for all i, j will the contamination disappear, which is the same as assuming that $E[y_i(t)y_j(t')] = 0$ for all $i \neq j$ and all $t, t' \in \mathcal{R}$. Marcet constructs various examples where the discrete Wold representation can be a poor approximation to the continuous one and demonstrates that in this context also the absence of Granger causality from one variable to another does not carry over to the sampled processes in general. Exceptions are processes in continuous time that are uncorrelated at all dates and the case where the first variable can be predicted with equal accuracy regardless of whether continuous or discrete data are used. The intuition of these results is that a sampled process omits information useful for predicting a continuous time process, namely past values between the integers. The other sampled process will be correlated with this information provided there is some correlation between the continuous time processes. Acting as a proxy for these past values, the latter sampled process will appear to cause the former in discrete time.

The effects of temporal aggregation on Granger causality testing have also been explored in empirical work. Christiano and Eichenbaum (1987) establish that time-averaging and sampling a continuous time process can increase the moving-average order of a time-series representation. The former result, which they attribute to Working (1960), has recently been characterized precisely in the multivariate setting by Breitung and Swanson (2002). Interestingly, Christiano and Eichenbaum suggest that temporal aggregation effects induced by shrinking the model timing interval can play a similar role in improving model fit as can adding costs of adjustment and serially correlated shocks because the qualitative effect on the reduced-form dynamics of the model for sampled data is the same. They find evidence for money Granger causing output with quarterly U.S. data that seems to be overturned when moving to a finer sampling interval. Harvey and Stock (1989) find evidence of money not Granger causing income in a continuous time model but a strong reversal of this finding when temporal aggregation is ignored in discrete time VARs. More recently, Renault, Sekkat and Szafarz (1998), using a continuous time model to distinguish between 'true' and 'spurious' causality, obtained evidence to suggest there is a 'discrete time illusion' of causality between the German mark and the Swiss franc. All of these results suggest that the common practice of naïvely formulating models in discrete time is not innocuous and that formulating econometric models in continuous time can help avoid misinterpreting the data in general, and Granger causality relationships in particular.

3. Defining causality in continuous time models

Our fundamental objective in this paper is to demonstrate that testing for causality between time series is not immune to the distortional effects of temporal aggregation and that continuous time models can offer the basis to correct for these distortions. Towards this end, we now discuss how to define causality in continuous time models, and how we can accurately import restrictions to economic data in a context where the data are generated in finer time intervals than the sampling interval.

3.1. Granger causality

Attempting to explain precisely what we mean by 'causality' is a forlorn task and is ultimately a philosophical problem. As Granger (1980, p. 330) notes, "Attitudes towards causality differ widely, from the defeatist one that it is impossible to define causality, let alone test for it, to the populist viewpoint that everyone has their own personal definition and so it is unlikely that a generally acceptable definition exists." The concept of 'Granger causality' is designed as an operational definition such that real statements can be made about causality on the basis of statistical data. The general principle can be set in a context where Ω_T represents all the information in the universe at time T. Let $F(A|B)$ be the conditional distribution of A given B and consider two series Y_t and X_t. Then if

$$F(X_{t+k}|\Omega_t) = F(X_{t+k}|\Omega_t - Y_t) \quad (\forall k > 0), \tag{10}$$

where $\Omega_t - Y_t$ is all the information in the universe apart from the values Y_t taken up to time t, then Y_t does not cause X_t. If condition (10) does not hold, then Y_t could be said to cause X_t on the grounds that there is special information contained in Y_t about X_t that is not available elsewhere. Granger and Thomson (1987) demonstrate that when using the causal variable Y_t to form forecasts of a function of X_t, then one is never worse off and usually better off using any cost function.

The above definition of causality is, of course, too general to be testable. In practice, we have to replace Ω_t with a restricted information set containing present and past values of certain time series and we choose a criterion to decide on how one forecast is superior to another, often restricting attention to linear forecasts under the usual least-squares loss function.[7] If J_t is an information set available at time t that includes X_{t-j} and a vector Z_{t-j} of other series but excluding Y_{t-j} $(j \geq 0)$, and J_t' is the information set J_t expanded to include Y_{t-j} $(j \geq 0)$, then if there exists $k > 0$ such that

$$F(X_{t+k}|J_t') \neq F(X_{t+k}|J_t), \tag{11}$$

we could say that Y_t is a *prima facie* cause of X_{t+k} with respect to the information set J_t', on the grounds that Y_t is a possible cause of the future X's. We need to use the phrase *'prima facie'* because we might obtain a different result with respect to a different information set. The notion of Granger causality pertains to conditional expectations and as a condition is implied by (11). If there exists $k > 0$ such that

$$E(X_{t+k}|J_t') \neq E(X_{t+k}|J_t), \tag{12}$$

we could say that Y_t causes X_t in mean with respect to the information set J_t', or we say simply that Y_t 'Granger causes' X_t. Usually the definition is considered just for the case $k = 1$ and we focus on whether Y_t helps provide an improved least-squares forecast for X_{t+1} than if Y_t were not used. It is this basic definition that Florens and Fougère (1996) and Comte and Renault (1996) initially applied in the continuous time framework.

3.2. Non-causality in continuous time models

Formulating econometric models in continuous time offers several advantages in the context of causality testing: they can take account of the interaction among variables during the unit observation period; they can be represented as a causal chain where each of the variables responds directly to the stimulus of only a proper subset of the other variables while there is interaction between all the variables during the observation period; they allow a clear distinction to be made between stock and flow variables; and their form does not depend on the unit observation period.

Florens and Fougère (1996) and Comte and Renault (1996), denoted FF and CR, respectively, in what follows, offered a global definition of non-causality in continuous time. Following CR, suppose $X(t) = [X_1(t)', X_2(t)', X_3(t)]'$ is an n-dimensional continuous time stochastic process, where the X_i are n_i-dimensional processes and $n_1 + n_2 + n_3 = n$. We consider the non-causality of X_2 on X_1 with X_3 as the 'environment' variable. FF and CR both say that X_2 does not Granger cause X_1 if

$$\forall t, h \geq 0, \quad E[X_1(t+h)|I(t)] = E[X_1(t+h)|I(t) - \overline{X}_2(t)], \tag{13}$$

[7] See Ashley, Granger and Schmalensee (1980) for a celebrated, early application to advertising and aggregate consumption.

where the information sets in (13) are the σ-algebras $I(t) = \sigma\{X(\tau), \tau \leq t, \tau \in \mathcal{R}\}$ and $I(t) - \overline{X}_2(t) = \sigma\{X_1(\tau), X_3(\tau), \tau \leq t, \tau \in \mathcal{R}\}$.

What we are really interested in, of course, is a *local* notion of causality, when $h \downarrow 0$, but if $X_1(t)$ is a mean-square continuous process, we have

$$\text{L}^2\text{-}\lim_{h\downarrow 0} E[X_1(t+h)|I(t)] = \text{L}^2\text{-}\lim_{h\downarrow 0} E[X_1(t+h)|I(t) - \overline{X}_2(t)], \tag{14}$$

and so, as FF and CR pointed out, we cannot meaningfully define local non-causality in terms of the *levels* of processes. Instead, they define local Granger non-causality in terms of increments. FF do this implicitly by using the general notion of causality discussed above with the canonical decomposition of a semimartingale $X(t) = X(0) + M(t) + H(t)$, where $M(t)$ is a local martingale with respect to $I(t)$, $M(0) = H(0) = 0$, and $H(t)$ is a process of finite variation.[8] Here, it will be sufficient for our purpose to state CR's definition of local causality that is explicitly based in terms of the increments of a process, which is equivalent to FF's except that it applies to a narrower class of processes. Let $X_1(t) = X_1(0)+M_1(t)+H_1(t)$ be a càdlàg[9] semimartingale such that M_1 is a martingale with respect to $I(t)$ and H_1 is mean-square continuously differentiable. Then X_2 does not locally Granger cause X_1 if and only if

$$\text{L}^2\text{-}\lim_{h\downarrow 0} E\left[\left.\frac{X_1(t+h) - X_1(t)}{h}\right| I(t)\right] = \text{L}^2\text{-}\lim_{h\downarrow 0} E\left[\left.\frac{X_1(t+h) - X_1(t)}{h}\right| I(t) - \overline{X}_2(t)\right]. \tag{15}$$

The definition is analogous to (12) except that it is applied to the increments and not the levels, and under a different metric. It is implied by the global definition (13) above.

There exists a potential problem in applying the definitions to processes like continuous time VARs because the conditions could be written equivalently in terms of a process or its (mean square) derivatives. Suppose a process X has components X_1 and X_2 having the same order of differentiability, and let D denote the mean square differentiation operator. CR showed that X_2 does not locally Granger cause X_1 if $D^k X_2$ does not locally Granger cause $D^k X_1$ and, provided the non-causality is defined in terms of the derivative of maximal order k that effectively exists for the process, the problem is circumvented. We remark that in the case that there is no environment and the continuous time process admits a continuous time invertible moving average representation, then the local Granger non-causality from X_2 onto X_1 and of X_1 onto X_2 is not sufficient to ensure the independence between X_1 and X_2. CR offer a necessary and sufficient condition they call local instantaneous causality, whose global counterpart is "the natural generalization of discrete time instantaneous causality" (p. 221). What precisely 'instantaneous causality' is and whether it truly exists other than

[8] See Protter (1990, p. 107).

[9] *continue à droite, limite à gauche.*

as a facet of temporal aggregation or missing causal variables is a controversial issue (see Granger, 1988, pp. 204-208). Certainly, Granger felt that it would have as a concept to go beyond correlation, which measures the association of two variables, to indicate the *direction* of their relationship. In the absence of *a priori* information, this would rule out a definition that is symmetric in the variables, contrary to an earlier definition by Pierce and Haugh (1977), and indeed CR's definition. In the fixed-interval time aggregation literature, Breitung and Swanson (2002) have recently profitably revived a notion of (apparent) instantaneous causality in a multivariate context, giving sufficient conditions to rule it out as an artefact of temporal aggregation. However, for our purpose of using a structural continuous time model to help test for causal structure using observed discrete data, we shall not need such a concept, as we explain below.

3.3. Continuous time VAR models and their exact discrete analogues

One of the advantages of formulating econometric models in continuous time is that they can represent a causal chain model that can take account of *a priori* information concerning the ordering of the variables. As Bergstrom (1996) notes, we can impose the restrictions implied by our knowledge of the information available to agents on a particular day as a means of obtaining more efficient parameter estimates. This is not possible in a naïve discrete time framework because each variable will be a function of all the variables in the model during the observation period. What is more, in certain circumstances we can import exactly the restrictions on the continuous time model to a discrete time analogue that we can then use as the basis of estimating the structural parameters or for causality testing. For example, for the first order continuous time VAR model in stock variables, we have

$$dx(t) = A(\theta)x(t)dt + \zeta(dt), \tag{16}$$

where $\{x(t), -\infty < t < \infty\}$ is an n-dimensional continuous time random process, A is an $n \times n$ matrix whose eigenvalues have strictly negative real parts and whose elements are known functions of a p-dimensional vector θ of unknown parameters ($p < n^2$), and $\zeta(dt)$ is a vector of white noise innovations with covariance matrix Σdt. Bergstrom (1984) showed how to derive a system of stochastic difference equations satisfying the time-invariant linear stochastic differential equation system driven by white noise disturbances in (16). A sequence of equispaced observations $x(0), x(1), \ldots, x(T)$ generated by (16) satisfies the exact discrete model

$$x(t) = F(\theta)x(t-1) + \epsilon_t, \quad t = 1, \ldots, T, \tag{17}$$

where $F(\theta) = e^{A(\theta)} = I + \sum_{r=1}^{\infty} A(\theta)^r/r!$ and ϵ_t is a white noise disturbance vector with covariance matrix $\Omega(\theta) = \int_0^1 e^{rA(\theta)}\Sigma e^{rA(\theta)'}dr$. The exact discrete analogue (17) has the form of a VAR model in discrete time. In this context only, we note that if there is no environment and x has components x_1 and x_2, if the matrix A is lower triangular, implying local Granger non-causality of x_2 onto x_1, then F will also be lower triangular and so there will be discrete time Granger non-causality of x_2 onto x_1. In higher-order models, however, this property does not carry over. For example, in the second-order model discussed by Bergstrom (1985), we have

$$d[Dx(t)] = [A_1(\theta)Dx(t) + A_2(\theta)x(t)]dt + \zeta(dt), \quad t > 0, \tag{18}$$

where $x(0)$ and $Dx(0)$ are assumed to be non-random. As shown by Bergstrom (1985) the observations of $x(t)$ observed at integer points in time[10] satisfy the VARMA model

$$x(t) = F_1(\theta)x(t-1) + F_2(\theta)x(t-2) + \eta_t, \quad t = 3, \ldots, T, \tag{19}$$

where η_t is an MA(1) disturbance process and may be written $\eta_t = u_t + Gu_{t-1}$ for some particular white noise process u_t and matrix G. The non-causality of x_2 onto x_1 in continuous time is represented by the restrictions $[A_1]_{12} = 0$ and $[A_2]_{12} = 0$ in (18), where $[\cdot]_{12}$ denotes the (1,2)-block of a matrix. Following CR we can use the condition by Boudjellabah, Dufour and Roy (1992) to characterize the discrete time Granger non-causality of x_2 onto x_1 by the nullity of

$$[(I + Gz)^{-1}(I - F_1 z - F_2 z^2)]_{12} = 0 \quad \forall z. \tag{20}$$

While the Granger local non-causality of x_2 onto x_1 implies in discrete time that F_1 and F_2 are lower triangular, in general (20) will not be fulfilled, even when there is no environment, owing to the complication arising from the moving average term in the exact discrete model. In other words, the Granger local non-causality of x_2 onto x_1 does not imply Granger non-causality in the discretized process. It is worth remarking that this result is not dependent on the form of Σ. As CR show, in the absence of an environment variable, their 'local instantaneous causality' restrictions on Σ, necessary to establish the independence of x_1 and x_2, are independent of whether x_2 locally Granger causes x_1 or *vice versa*.

4. An empirical illustration

In this section we provide an empirical illustration of testing for Granger causality when the effects of temporal aggregation are explicitly taken into account.[11] We focus on the

[10] It is possible to allow for a sample of mixed stock and flow data but doing so adds nothing to the discussion here.

[11] This section was added to the original paper on the suggestion of the editor and a referee. The methods and results are meant to be illustrative rather than definitive and we have focused exclusively on the

widely-studied issue of money-income causality which, following Sims (1972), has been the subject of intense research activity. There has, however, been much conflicting evidence produced concerning the key issue of whether money Granger causes income. Recently, however, theoretical advances in the area of unit roots and cointegration have led to a re-examination of some of the earlier results in the literature. An important contribution in this vein is Stock and Watson (1989), who undertake a careful analysis of stochastic and deterministic trends in monthly U.S. industrial production and money stock (M1) data over the period January 1960 to December 1985. One of their key findings is that *innovations* in M1 i.e. in the appropriately detrended series, have statistically significant marginal predictive value for industrial production. This finding is robust to consideration of both a bivariate system and a multivariate system that also incorporates a price index and an interest rate.

Our empirical application also uses monthly U.S. data on the industrial production index (IP) and M1 but over a longer period than that considered by Stock and Watson (1989). Our sample period covers January 1960 to December 2001, with observations from 1959 used as intitial values in the dynamic models. The sample size used for all estimations is therefore $T = 504$. We define $y = \ln(\text{IP})$ and $m = \ln(\text{M1})$. Table 1 contains the results of testing for unit roots and cointegration in the two series y and m, as well as tests of the significance of deterministic time trends in these two series. The univariate Stock-Watson $q_c(1, 0)$ and augmented Dickey-Fuller τ_6 tests for unit roots provide strong evidence for the presence of unit roots in each of the series, although we note that the augmented Dickey-Fuller statistic with a linear time trend is significant for y. The two cointegration tests both fail to reject the null hypothesis of two unit roots in the bivariate system against the alternative of only one, suggesting that these two series are not cointegrated but that the system contains two unit roots. Although these findings are in accordance with those of Stock and Watson (1989), the t-statistics on the deterministic trend terms suggest the presence of a quadratic as opposed to a linear trend in m, although the growth of IP appears only to contain a drift component.[12] We therefore proceed under the maintained hypothesis that y and m are well described by the following univariate representations in discrete time:[13]

$$\Delta y_t = \alpha_{y0} + \Delta\eta_t, \tag{21}$$

$$\Delta m_t = \alpha_{m0} + \alpha_{m1} t + \alpha_{m2} t^2 + \Delta\mu_t, \tag{22}$$

importance for causality tests of incorporating temporal aggregation restrictions on discrete time data. We acknowledge that our results could be sensitive to changes in the sample period (Friedman and Kuttner, 1993) or to different orthogonalizations of the covariance matrices of residuals (Swanson and Granger, 1997). However, such issues are beyond the scope of the illustration provided here.

[12] Eliminating time^2 in the regression for y yields a t-ratio of -0.6620 on time; further eliminating time results in a t-ratio of 3.3524 on the drift term.

[13] We have borrowed the notation from Stock and Watson (1989).

where $\Delta\eta_t$ and $\Delta\mu_t$ denote mean zero stationary processes. Furthermore, defining the vector $x = (y, m)'$, we also maintain that x contains two unit roots and that its components are not cointegrated.

In view of the above properties of the series, the usual approach to testing for Granger causality from m to y would be to specify a VAR in first differences (with a quadratic trend) and to test the significance of the coefficients on the lagged Δm_t variables in the equation for Δy_t. The validity of such an approach is established by Toda and Phillips (1993),[14] who state (in the context of Wald tests) that "if it is known that the system is I(1) with no cointegration, causality tests based on difference VARs are also valid, and in these tests the usual chi-square critical values are employed" (pp. 1376–1377). Furthermore, "causality tests in difference VARs are likely to have higher power in finite samples" (p. 1377). Adopting this approach with a VAR specified in terms of the vector of detrended series $\widehat{x}_t = (\eta_t, \mu_t)'$ we consider testing causality in the system

$$\Delta\widehat{x}_t = b_0 + \sum_{j=1}^{p} B_j \Delta\widehat{x}_{t-j} + u_t, \tag{23}$$

where u_t is assumed to be vector white noise with covariance matrix Σ_u. Estimation of this system suggests that the null of no Granger causality from (detrended) money to (detrended) output is represented by the p restrictions $[B_j]_{12} = 0$ $(j = 1, \ldots, p)$, where $[B_j]_{12}$ denotes the second element in the first row of the matrix B_j i.e. the coefficient of $\Delta\mu_{t-j}$ in the equation for $\Delta\eta_t$. On the basis of the Akaike and Schwarz order selection criteria, as well as a likelihood ratio test for testing the null that the order is p against the alternative that it is $p+1$, the value of $p = 3$ is chosen. However, there is strong evidence of serial correlation in the residuals of both equations, and it therefore seems prudent to increase the order of the model in an attempt to eradicate the serial correlation. There is, however, a trade-off to be made. Incorporating too many lagged terms may adversely affect the power of the tests, but incorporating an insufficient number may lead to the tests being biased due to the presence of serial correlation. We therefore consider two values of p, namely $p = 6$ and $p = 12$, and note that Lagrange Multiplier tests of serial correlation up to order 12 do not reject the null of no serial correlation in either equation when $p = 12$, but do reject (at the 5% level) when $p = 6$.

Likelihood ratio tests of the hypothesis of no Granger causality for these two VARs are presented in Table 2. In neither case is the null hypothesis rejected at the 5% (or even the 10%) level of significance, although we note the marginal probability value is much higher (i.e. further away from rejecting the null) in the absence of serial correlation (when $p = 12$).

[14] See also Sims, Stock and Watson (1990) for related results.

These results, obtained without consideration of temporal aggregation issues, indicate that innovations in money *do not* Granger cause output growth, contrary to the findings of Stock and Watson (1989). We shall use these results as a benchmark against which to compare our findings when temporal aggregation is accounted for, and to which we now turn.

The continuous time model that we shall estimate is based around our earlier findings concerning the trend properties of the two variables y and m. The continuous time counterparts of (21) and (22) are, respectively,

$$d \ln \mathrm{IP}(t) = \gamma_{0y} dt + d\eta(t), \tag{24}$$

$$d \ln \mathrm{M1}(t) = \left(\gamma_{0m} + \gamma_{1m} t + \gamma_{2m} t^2\right) dt + d\mu(t), \tag{25}$$

where $\eta(t)$ and $\mu(t)$ denote mean zero stationary continuous time random processes. We shall treat both IP and M1 as being flow variables in view of the M1 data being monthly averages of daily values and the IP index being a measure of output produced during each month. The raw observations are therefore in the form of the integrals $Y_t = \int_{t-1}^{t} \mathrm{IP}(r)dr$ and $M_t = \int_{t-1}^{t} \mathrm{M1}(r)dr$. Ideally, because the model is specified in terms of logarithms, we would wish to observe the integrals of the logarithms themselves, but we shall proceed on the assumption that the logarithm of the observed integrals provides an accurate approximation.[15] Taking $y_t = \ln Y_t$ and $m_t = \ln M_t$, we are therefore assuming that

$$y_t = \ln \int_{t-1}^{t} \mathrm{IP}(r)dr \approx \int_{t-1}^{t} \ln \mathrm{IP}(r)dr, \quad m_t = \ln \int_{t-1}^{t} \mathrm{M1}(r)dr \approx \int_{t-1}^{t} \ln \mathrm{M1}(r)dr. \tag{26}$$

The least squares detrending that we applied in the discrete time approach remains equally valid here. To verify this, observe that integrating the left hand side of (24) twice over the unit interval yields

$$\int_{t-1}^{t} \int_{s-1}^{s} d \ln \mathrm{IP}(r) ds = \int_{s-1}^{s} \left[\ln \mathrm{IP}(s) - \ln \mathrm{IP}(s-1)\right] ds = \Delta y_t.$$

Integrating the right hand side twice yields $\gamma_{0y} + \Delta\eta_t$, where $\Delta\eta_t = \int_{t-1}^{t} \int_{s-1}^{s} d\eta(r)ds$ denotes the discrete time detrended series. Applying a similar procedure to (25) and dealing with the double integrals $\int_{t-1}^{t} \int_{s-1}^{s} r dr ds$ and $\int_{t-1}^{t} \int_{s-1}^{s} r^2 dr ds$ yields the equation

$$\Delta m_t = (\gamma_{0m} - \gamma_{1m} + \frac{7}{6}\gamma_{2m}) + (\gamma_{1m} - 2\gamma_{2m})t + \gamma_{2m} t^2 + \Delta\mu_t,$$

which is of the same form as (22).

[15]This approximation is not without precedent in the empirical continuous time literature. See, for example, Bergstrom, Nowman and Wymer (1992).

The specification of the continuous time model is completed by equations describing the dynamic evolution of $\eta(t)$ and $\mu(t)$. In discrete time, this is achieved by the specification of a VAR model containing a sufficient number of lags to render the disturbance vector approximately white noise. In continuous time, the specification is in the form of a system of stochastic differential equations whose order is sufficiently high to model the dynamics adequately. Following Harvey and Stock (1989) we assume the vector $\widehat{x}(t) = (\eta(t), \mu(t))'$ satisfies a third-order stochastic differential equation with zero roots, given by

$$d\left[D^2\widehat{x}(t)\right] = \left[A_2 D^2\widehat{x}(t) + A_1 D\widehat{x}(t) + A_0\widehat{x}(t)\right]dt + \zeta(dt), \quad t > 0, \tag{27}$$

where A_0, A_1 and A_2 are 2×2 matrices of unknown parameters and $\zeta(dt)$ is a 2×1 vector of random measures satisfying $E\zeta(dt) = 0$, $E\zeta(dt)\zeta(dt)' = dt\Sigma_\zeta$, and $E\zeta(\Delta_1)\zeta(\Delta_2)' = \emptyset$ for Δ_1 and Δ_2 any two disjoint subsets of $[0, T]$. The zero roots assumption is incorporated by setting $A_0 = 0$, resulting in a second-order stochastic differential equation in the stationary vector $w(t) = D\widehat{x}(t)$, given by

$$d\left[Dw(t)\right] = \left[A_2 Dw(t) + A_1 w(t)\right]dt + \zeta(dt), \quad t > 0, \tag{28}$$

Effectively, this is a third-order system in the underlying variables lnIP and lnM1 but which contains zero roots. The detrending equations (24) and (25) deal with the continuous time zero roots as well as the deterministic trends, the result being the second-order system in the vector of detrended variables $D\widehat{x}(t)$ in (28).

Our approach to estimation is based on the exact discrete representation of (28). Note that $\Delta\widehat{x}_t = \int_{t-1}^{t}\int_{s-1}^{s} d\widehat{x}(r)ds = \int_{t-1}^{t}\int_{s-1}^{s} w(r)drds$ and define θ to be the vector of unknown parameters comprising the elements of A_2, A_1 and Σ_ζ. The vector θ therefore has 11 free parameters. It is possible to show that $\Delta\widehat{x}_t$ satisfies a VARMA(2,3) system of the form

$$\Delta\widehat{x}_t = F_1(\theta)\Delta\widehat{x}_{t-1} + F_2(\theta)\Delta\widehat{x}_{t-2} + \xi_t, \quad t = 1, \ldots, T, \tag{29}$$

where ξ_t is a vector MA(3) disturbance satisfying $E\xi_t = 0$, $E\xi_t\xi'_{t-j} = \Omega_j(\theta)$ $(j = 0, \ldots, 3)$ and $E\xi_t\xi'_{t-j} = 0$ $(j > 3)$, and the elements of F_1, F_2 and the Ω_j are complicated functions of the elements of θ. Details of the precise formulae relating F_1, F_2 and the Ω_j to θ, along with their derivations, may be found in the Appendix. We take the pre-sample values $\Delta\widehat{x}_{-1}$ and $\Delta\widehat{x}_0$ to be fixed and condition the likelihood function accordingly. The null hypothesis that innovations in money do not Granger cause output is represented by the two restrictions $[A_2]_{12} = 0$ and $[A_1]_{12} = 0$, which we test using the likelihood ratio principle.

We shall also consider two approximate discrete time models derived from (28) in an attempt to assess the importance of imposing the exact restrictions on the discrete time data. The first approximation replaces the derivative $D^k\widehat{x}(t)$ in (27) with the difference $\Delta^k\widehat{x}_t$ to yield

$$\Delta^3\widehat{x}_t = A_2\Delta^2\widehat{x}_t + A_1\Delta\widehat{x}_t + v_t, \tag{30}$$

where v_t is assumed to be vector white noise with covariance matrix Σ_v. This equation can be rearranged into a VAR(2) in the variable $\Delta\widehat{x}_t$ of the form

$$\Delta\widehat{x}_t = C_1\Delta\widehat{x}_{t-1} + C_2\Delta\widehat{x}_{t-2} + e_t, \tag{31}$$

where $C_1 = K^{-1}(2I - A_2)$, $C_2 = -K^{-1}$, $K = (I - A_1 - A_2)$, and $e_t = K^{-1}v_t$ is vector white noise. The second approximation is more sophisticated and is derived by integrating (28) four times over the interval $(t-1, t)$ and using the fact that $\int_{t-1}^{t}\int_{s-1}^{s} w(r)drds = \Delta\widehat{x}_t$ as well as the approximation $\int_{t-1}^{t}\alpha(r)dr \approx [\alpha(t) + \alpha(t-1)]/2 = F(L)\alpha(t)$ for a continuous time integrable variable $\alpha(t)$, where $F(z) = (1+z)/2$ and L denotes the lag operator. The result is a VARMA(2,3) system in $\Delta\widehat{x}_t$ of the form

$$\Delta\widehat{x}_t = G_1\Delta\widehat{x}_{t-1} + G_2\Delta\widehat{x}_{t-2} + \epsilon_t, \tag{32}$$

where $G_1 = H^{-1}(2I + \frac{1}{2}A_1)$, $G_2 = H^{-1}(\frac{1}{4}A_1 - \frac{1}{2}A_2 - I)$, $H = (I - \frac{1}{4}A_1 - \frac{1}{2}A_2)$, and ϵ_t is a vector MA(3) process. The differences between the exact discrete time model (29) and the approximations (31) and (32) lie in the way in which the discrete time autoregressive matrices relate to the continuous time parameters and the nature of the disturbance vectors. In (31) the disturbance vector is assumed to be white noise, while in (32) it is MA(3) although, once again, the precise form of the autocovariance matrices in (29) and (32) are different. Each of the two discrete time approximations is estimated by maximising the (Gaussian) likelihood function, conditional on $\Delta\widehat{x}_{-1}$ and $\Delta\widehat{x}_0$ being fixed. Further details of the derivations leading to the approximations (31) and (32) may be found in the Appendix.

Table 3 presents estimates of the continuous time parameters obtained from the exact discrete model and the two approximations, both with and without the causality restrictions imposed. Rather than estimating the covariance matrix Σ_ζ directly, we estimated the elements of the lower triangular Cholesky factorisation M_ζ such that $M_\zeta M_\zeta' = \Sigma_\zeta$. This was done to ensure that the covariance matrix remained positive definite in the optimisation of the likelihood function, and Table 3 reports estimates of the elements of M_ζ (denoted $[M_\zeta]_{11}$, $[M_\zeta]_{21}$ and $[M_\zeta]_{22}$). Taking the results obtained from the exact discrete model first, it is clear to see that imposition of the causality restrictions has a dramatic impact on the estimates of the remaining free parameters and there is a corresponding sharp fall in the value of the maximised likelihood function. This is perhaps not surprising in view of the significance in the unrestricted model of the two parameters that are being constrained to be equal to zero in the restricted model. As a result the likelihood ratio statistic convincingly *rejects* the null of Granger non-causality when the temporal aggregation restrictions are accounted for exactly. In contrast, imposition of the restrictions via the two approximate discrete models results in much smaller changes in the remaining free parameters, and a much smaller drop in the maximised likelihood function, than when the exact discrete model is employed. As a result, neither likelihood ratio statistic rejects the null hypothesis in the two approximate discrete models, a finding that is in line with the discrete time VARs reported earlier.

Our empirical results suggest that correctly accounting for temporal aggregation restrictions can have an important bearing on inferences drawn when testing for Granger causality. It also appears that even the approximate discrete time models do not adequately reflect the temporal aggregation restrictions, in line with purely unrestricted discrete time VARs. It is also worth noting that the MA(3) disturbance in the exact discrete model appears to account for the serial correlation in the disturbance term, a feature not shared in the approximations (nor in the low order VARs). Our result, that innovations in money cause output (growth) in the continuous time system, is at variance with Harvey and Stock (1989), who found that, using data from January 1960 to December 1985, accounting for temporal aggregation resulted in a non-rejection (at the 5% level) of the restrictions. We note, however, that our method of detrending the data is different to theirs, as is our approach to estimation, which is based on the exact discrete model while Harvey and Stock used Kalman filtering techniques applied to the state space form of the model. Our sample period is also longer.

5. Conclusion

The paper has considered, from a continuous time perspective, the problem that spurious Granger causality relationships can arise due to temporal aggregation. We showed that formulating models in continuous time offers a basis for correcting for the effects of temporal aggregation in observed discrete data through a discrete time analogue, in a way that does not rely on our positing a definite time unit in which the data are generated. In an empirical application, we showed that imposing these restrictions, and precisely, matters in testing for Granger causality.

Our results complement those in the fixed-interval time aggregation literature, especially those recently obtained by Marcellino (1999) and Breitung and Swanson (2002). Our application to money-income causality was designed to be illustrative of the effects of causality testing and no attempt was made to present a definitive study. One direction for future research would be to devise a data-determined method for continuous time models along the lines of Swanson and Granger (1997) for discrete time models, to examine the sensitivity of causality results to different residual orthogonalizations and under different *a priori* causal restrictions on the variables.

Appendix

Derivation of the exact discrete time representation

In this section we derive the formulae for the exact discrete model corresponding to a third-order continuous time system with flow variables and zero roots using the approach in Chambers (1999). The system in which we are interested may be written

$$d\left[D^2\widehat{x}(t)\right] = \left[A_2 D^2\widehat{x}(t) + A_1 D\widehat{x}(t) + A_0\widehat{x}(t)\right]dt + \zeta(dt), \quad t > 0, \quad A_0 = 0, \tag{33}$$

where $\zeta(dt)$ is a 2×1 vector of random measures satisfying $E\zeta(dt) = 0$, $E\zeta(dt)\zeta(dt)' = dt\Sigma_\zeta$, and $E\zeta(\Delta_1)\zeta(\Delta_2)' = \emptyset$ where Δ_1 and Δ_2 are any two disjoint subsets of $[0, T]$. It is convenient to rewrite (33) in terms of the stationary variable $w(t) = D\widehat{x}(t)$, which gives

$$d\left[Dw(t)\right] = \left[A_2 Dw(t) + A_1 w(t)\right]dt + \zeta(dt), \quad t > 0. \tag{34}$$

The underlying observations in our system are given by the $n \times 1$ vector

$$\Delta\widehat{x}_t = \int_{t-1}^{t}\int_{s-1}^{s} d\widehat{x}(r)ds = \int_{t-1}^{t}\int_{s-1}^{s} w(r)drds; \tag{35}$$

the first difference form reflects the zero roots in the continuous time system. The parameters to be estimated are the elements of A_1, A_2 and Σ_ζ; denote these by the vector θ.

Theorem. *Let $\widehat{x}(t)$ be generated by (33) and let the observations be given by (35). Then $\Delta\widehat{x}_t$ satisfies*

$$\Delta\widehat{x}_t = F_1(\theta)\Delta\widehat{x}_{t-1} + F_2(\theta)\Delta\widehat{x}_{t-2} + \xi_t, \quad t = 1, \ldots, T, \tag{36}$$

where $F_1 = F_{11} + F_{12}F_{22}F_{12}^{-1}$, $F_2 = F_{12}\left[F_{21} - F_{22}F_{12}^{-1}F_{11}\right]$,

$$\begin{aligned}
F &= \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} = e^A = \sum_{j=0}^{\infty}\frac{A^j}{j!}, \qquad A = \begin{bmatrix} 0 & I \\ A_1 & A_2 \end{bmatrix}, \\
E\xi_t\xi_{t-j}' &= \Omega_{\xi,j} \quad (j = 0, \ldots, 3), \\
\Omega_{\xi,0} &= S_0\Omega_0 S_0' + S_0\Omega_1 S_1' + S_1\Omega_1' S_0' + S_1\Omega_0 S_1', \\
\Omega_{\xi,1} &= S_0\Omega_1 S_0' + S_0\Omega_2 S_1' + S_1\Omega_0 S_0' + S_1\Omega_1 S_1', \\
\Omega_{\xi,2} &= S_0\Omega_2 S_0' + S_1\Omega_1 S_0' + S_1\Omega_2 S_1', \\
\Omega_{\xi,3} &= S_1\Omega_2 S_0',
\end{aligned}$$

$$
\begin{aligned}
S_0 &= [I, 0], \qquad S_1 = [-F_{12}F_{22}F_{12}^{-1}, F_{12}], \\
\Omega_0 &= \int_0^1 \Phi_0(r)\Sigma_\zeta \Phi_0(r)' dr + \int_0^1 \Phi_1(r)\Sigma_\zeta \Phi_1(r)' dr + \int_0^1 \Phi_2(r)\Sigma_\zeta \Phi_2(r)' dr, \\
\Omega_1 &= \int_0^1 \Phi_1(r)\Sigma_\zeta \Phi_0(r)' dr + \int_0^1 \Phi_2(r)\Sigma_\zeta \Phi_1(r)' dr, \\
\Omega_2 &= \int_0^1 \Phi_2(r)\Sigma_\zeta \Phi_0(r)' dr, \\
\Phi_0(r) &= -A^{-2} - A^{-1}r + A^{-2}e^{rA}, \\
\Phi_1(r) &= A^{-2}(F+I) - A^{-1} + A^{-1}(I+F)r - 2A^{-2}e^{rA}, \\
\Phi_2(r) &= A^{-1}(F - A^{-1}F) - A^{-1}Fr + A^{-2}e^{rA}.
\end{aligned}
$$

Proof. Let $z(t) = [w(t)', Dw(t)']'$. Then $z(t)$ satisfies

$$dz(t) = Az(t)dt + u(dt), \tag{37}$$

where $u(dt) = [0', \zeta(dt)']'$. Integrating (37) yields $\Delta z(t) = A\int_{t-1}^{t} z(r)dr + \int_{t-1}^{t} u(dr)$ so that

$$\int_{t-1}^{t} z(r)dr = A^{-1}\Delta z(t) - A^{-1}\int_{t-1}^{t} u(dr). \tag{38}$$

But, from Theorem 3 of Bergstrom (1984), $z(t) = Fz(t-1) + \int_{t-1}^{t} F(t-r)u(dr)$, where $F(r) = e^{rA}$. Hence

$$z(t) - z(t-1) = F\left[z(t-1) - z(t-2)\right] + \int_{t-1}^{t} F(t-r)u(dr) - \int_{t-2}^{t-1} F(t-1-r)u(dr). \tag{39}$$

Substituting (39) into (38) results in

$$
\begin{aligned}
\int_{t-1}^{t} z(r)dr &= F\int_{t-2}^{t-1} z(r)dr + A^{-1}\left\{\int_{t-1}^{t}\left[F(t-r) - I\right]u(dr)\right. \\
&\quad \left. - \int_{t-2}^{t-1}\left[F(t-1-r) - F\right]u(dr)\right\},
\end{aligned} \tag{40}
$$

which uses the properties $A^{-1}F = FA^{-1}$ and $A^{-1}FA = F$. Integrating (40) over $(t-1, t)$:

$$
\begin{aligned}
\int_{t-1}^{t}\int_{s-1}^{s} z(r)drds &= F\int_{t-1}^{t}\int_{s-2}^{s-1} z(r)drds + A^{-1}\left\{\int_{t-1}^{t}\int_{s-1}^{s}\left[F(s-r) - I\right]u(dr)ds\right. \\
&\quad \left. - \int_{t-1}^{t}\int_{s-2}^{s-1}\left[F(s-1-r) - F\right]u(dr)ds\right\}.
\end{aligned} \tag{41}
$$

Noting that

$$\int_{t-1}^{t}\int_{s-1}^{s} z(r)drds = \int_{t-1}^{t}\int_{s-1}^{s}\begin{bmatrix} w(r) \\ Dw(r)\end{bmatrix}drds = \begin{bmatrix}\Delta\widehat{x}_t \\ \Delta^2\widehat{x}(t)\end{bmatrix}$$

indicates that we need the first n equations of (41). Let $z_t = [z_{1t}', z_{2t}']' = \int_{t-1}^{t}\int_{s-1}^{s} z(r)drds$ and define $v_t = [v_{1t}', v_{2t}']'$ where

$$v_t = A^{-1}\left\{\int_{t-1}^{t}\int_{s-1}^{s}\left[F(s-r)-I\right]u(dr)ds - \int_{t-1}^{t}\int_{s-2}^{s-1}\left[F(s-1-r)-F\right]u(dr)ds\right\}.$$

Then (41) is $z_t = Fz_{t-1} + v_t$ and is comprised of

$$\begin{aligned} z_{1t} &= F_{11}z_{1t-1} + F_{12}z_{2t-1} + v_{1t}, &(42)\\ z_{2t} &= F_{21}z_{1t-1} + F_{22}z_{2t-1} + v_{2t}. &(43)\end{aligned}$$

From (42),

$$\begin{aligned} z_{2t-1} &= F_{12}^{-1}\left(z_{1t} - F_{11}z_{1t-1} - v_{1t}\right), &(44)\\ z_{2t-2} &= F_{12}^{-1}\left(z_{1t-1} - F_{11}z_{1t-2} - v_{1t-1}\right), &(45)\end{aligned}$$

while from (43),

$$z_{2t-1} = F_{21}z_{1t-2} + F_{22}z_{2t-2} + v_{2t-1}. \quad (46)$$

Substituting (44) and (45) into (46):

$$F_{12}^{-1}\left(z_{1t} - F_{11}z_{1t-1} - v_{1t}\right) = F_{21}z_{1t-2} + F_{22}F_{12}^{-1}\left(z_{1t-1} - F_{11}z_{1t-2} - v_{1t-1}\right) + v_{2t-1}. \quad (47)$$

Solving this equation for z_{1t} yields (29) as required, where

$$\xi_t = v_{1t} - F_{12}F_{22}F_{12}^{-1}v_{1t-1} + F_{12}v_{2t-1}. \quad (48)$$

This completes the proof. □

Autocovariance properties of ξ_t

Note, first, that ξ_t has the representation $\xi_t = S_0v_t + S_1v_{t-1}$. The double integrals defining v_t can be reduced to single integrals using the methods of McCrorie (2000). This yields

$$\begin{aligned} &\int_{t-1}^{t}\int_{s-1}^{s}\left[F(s-r)-I\right]u(dr)ds\\ &\quad= \int_{t-2}^{t-1}\left[\int_{t-1}^{r+1}\left[e^{(s-r)A}-I\right]ds\right]u(dr) + \int_{t-1}^{t}\left[\int_{r}^{t}\left[e^{(s-r)A}-I\right]ds\right]u(dr)\\ &\quad= \int_{t-2}^{t-1}\left[A^{-1}F - I + (t-1-r)I - A^{-1}e^{(t-1-r)A}\right]u(dr)\end{aligned}$$

$$+\int_{t-1}^{t}\left[-A^{-1}-(t-r)I+A^{-1}e^{(t-r)A}\right]u(dr),$$

$$\begin{aligned}
&\int_{t-1}^{t}\int_{s-2}^{s-1}\left[F(s-1-r)-F\right]u(dr)ds \\
&= \int_{t-3}^{t-2}\left[\int_{t-1}^{r+2}\left[e^{(s-1-r)A}-F\right]ds\right]u(dr)+\int_{t-2}^{t-1}\left[\int_{r+1}^{t}\left[e^{(s-1-r)A}-F\right]ds\right]u(dr) \\
&= \int_{t-3}^{t-2}\left[A^{-1}F-F+(t-2-r)F-A^{-1}e^{(t-2-r)A}\right]u(dr) \\
&\quad +\int_{t-2}^{t-1}\left[-A^{-1}-(t-1-r)F+A^{-1}e^{(t-1-r)A}\right]u(dr).
\end{aligned}$$

Hence v_t has the representation

$$v_t=\int_{t-1}^{t}\Phi_0(t-r)u(dr)+\int_{t-2}^{t-1}\Phi_1(t-1-r)u(dr)+\int_{t-3}^{t-2}\Phi_2(t-2-r)u(dr). \tag{49}$$

The matrices Ω_j $(j=0,1,2)$ defined in the Theorem correspond to the autocovariances of v_t derived using (49). It is then a straightforward matter to derive the autocovariances of ξ_t using the relationship between ξ_t and v_t given in the first line of this section.

Discrete time approximation: method 1

Our simplest discrete time approximation replaces the derivatives $D^k\widehat{x}(t)$ in (27) with the differences $\Delta^k\widehat{x}_t$, yielding

$$\Delta^3\widehat{x}_t=A_2\Delta^2\widehat{x}_t+A_1\Delta\widehat{x}_t+v_t, \tag{50}$$

where v_t is assumed to be vector white noise. Expressing the higher-order differences in terms of $\Delta\widehat{x}_t$ gives

$$\Delta\widehat{x}_t-2\Delta\widehat{x}_{t-1}+\Delta\widehat{x}_{t-2}=A_2\left(\Delta\widehat{x}_t-\Delta\widehat{x}_{t-1}\right)+A_1\Delta\widehat{x}_t+v_t,$$

which, upon rearranging, yields

$$\Delta\widehat{x}_t=C_1\Delta\widehat{x}_{t-1}+C_2\Delta\widehat{x}_{t-2}+e_t, \tag{51}$$

where $C_1=(I-A_1-A_2)^{-1}(2I-A_2)$, $C_2=-(I-A_1-A_2)^{-1}$ and $e_t=(I-A_1-A_2)^{-1}v_t$ is also vector white noise.

Discrete time approximation: method 2

The second method is more sophisticated. Integrate (28) over $(t-1,t)$:

$$\Delta Dw(t)=A_2\Delta w(t)+A_1\int_{t-1}^{t}w(r)dr+\int_{t-1}^{t}\zeta(dr). \tag{52}$$

Integrating again yields

$$\Delta^2 w(t) = A_2 \Delta \int_{t-1}^{t} w(r)dr + A_1 \Delta \widehat{x}_t + \phi_t, \tag{53}$$

where $\phi_t = \int_{t-1}^{t} \int_{s-1}^{s} \zeta(dr)ds$. We shall use the approximation

$$\int_{t-1}^{t} \alpha(r)dr \approx [\alpha(t) + \alpha(t-1)]/2 = F(L)\alpha(t)$$

for a continuous time integrable variable $\alpha(t)$, where $F(z) = (1+z)/2$ and L denotes the lag operator. For future reference, note that $F(z)^2 = (1 + 2z + z^2)/4$. Integrating (53) again:

$$\Delta^2 \int_{t-1}^{t} w(r)dr = A_2 \Delta^2 \widehat{x}_t + A_1 F(L) \Delta \widehat{x}_t + F(L)\phi_t. \tag{54}$$

Integrating a fourth and final time yields

$$\Delta^3 \widehat{x}_t = A_2 F(L) \Delta^2 \widehat{x}_t + A_1 F(L)^2 \Delta \widehat{x}_t + F(L)^2 \phi_t. \tag{55}$$

Now, $F(L)\Delta^2 \widehat{x}_t = (\Delta \widehat{x}_t - \Delta \widehat{x}_{t-2})/2$, $F(L)^2 \Delta \widehat{x}_t = (\Delta \widehat{x}_t + 2\Delta \widehat{x}_{t-1} + \Delta \widehat{x}_{t-2})/4$, and define $\eta_t = F(L)^2 \phi_t = (\phi_t + 2\phi_{t-1} + \phi_{t-2})/4$, so that (55) can be written as

$$\Delta \widehat{x}_t - 2\Delta \widehat{x}_{t-1} + \Delta \widehat{x}_{t-2} = A_2(\Delta \widehat{x}_t - \Delta \widehat{x}_{t-2})/2 + A_1(\Delta \widehat{x}_t + 2\Delta \widehat{x}_{t-1} + \Delta \widehat{x}_{t-2})/4 + \eta_t.$$

Collecting terms and solving results in

$$\Delta \widehat{x}_t = G_1 \Delta \widehat{x}_{t-1} + G_2 \Delta \widehat{x}_{t-2} + \epsilon_t, \tag{56}$$

where $G_1 = H^{-1}(2I + \frac{1}{2}A_1)$, $G_2 = H^{-1}(\frac{1}{4}A_1 - \frac{1}{2}A_2 - I)$, $H = (I - \frac{1}{4}A_1 - \frac{1}{2}A_2)$, and $\epsilon_t = H^{-1}\eta_t$ is a vector MA(3) process.

The properties of ϵ_t may be established as follows. First, using the results of McCrorie (2000), it is possible to show that

$$\begin{aligned} \phi_t &= \int_{t-1}^{t} \int_{s-1}^{s} \zeta(dr)ds \\ &= \int_{t-1}^{t} (t-r)\zeta(dr) - \int_{t-2}^{t-1} (t-2-r)\zeta(dr). \end{aligned}$$

Hence ϕ_t is MA(1) with $E\phi_t\phi_t' = (2/3)\Sigma_\zeta$ and $E\phi_t\phi_{t-1}' = (1/6)\Sigma_\zeta$. The properties of η_t then follow from its definition in terms of ϕ_t, resulting in $E\eta_t\eta_t' = (1/3)\Sigma_\zeta$, $E\eta_t\eta_{t-1}' = (11/48)\Sigma_\zeta$, $E\eta_t\eta_{t-2}' = (1/12)\Sigma_\zeta$, and $E\eta_t\eta_{t-3}' = (1/96)\Sigma_\zeta$. The autocovariances of ϵ_t then follow directly.

References

Ashley, R., C.W.J. Granger and R. Schmalensee, 1980, Advertising and aggregate consumption: an analysis of causality, Econometrica 48, 1149-1167.

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

Bergstrom, A.R., 1996, Survey of continuous time econometrics, in: W.A. Barnett, G. Gandolfo and C. Hillinger, eds., Dynamic disequilibrium modelling (Oxford University Press, Oxford) 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.

Bergstrom, A.R., K.B. Nowman and C.R. Wymer, 1992, Gaussian estimation of a second order continuous time macroeconometric model of the UK, Economic Modelling 9, 313–351.

Boudjellabah, H., J.-M. Dufour and R. Roy, 1992, Testing causality between two vectors in a multivariate ARMA model, Journal of the American Statistical Association 87, 1082-1090.

Breitung, J. and N.R. Swanson, 2002, Temporal aggregation and spurious instantaneous causality in multiple time series models, Journal of Time Series Analysis 23, 651-665.

Chambers, M.J., 1999, Discrete time representation and Gaussian estimation of stationary and nonstationary continuous time systems, Journal of Economic Dynamics and Control 23, 619-639.

Chambers, M.J. and J.R. McCrorie, 2003, Identification and estimation of exchange rate models with unobservable fundamentals, preprint.

Christiano, L.J. and M.S. Eichenbaum, 1987, Temporal aggregation and structural inference in macroeconomics, Carnegie-Rochester Conference Series on Public Policy 26, 63-130.

Comte, F. and E. Renault, 1996, Noncausality in continuous time models, Econometric Theory 12, 215-256.

Florens, J.-P. and D. Fougère, 1996, Noncausality in continuous time, Econometrica 64, 1195-1212.

Friedman, B. and D. Kuttner, 1993, Another look at the evidence on money income causality, Journal of Econometrics 57, 189–203.

Geweke, J.B., 1978, Temporal aggregation in the multivariate regression model, Econometrica 46, 643-662.

Granger, C.W.J., 1969, Investigating causal relations by econometric methods and cross-spectral methods, Econometrica 37, 424-438.

Granger, C.W.J., 1980, Testing for causality: a personal viewpoint, Journal of Economic Dynamics and Control 2, 329-352.

Granger, C.W.J., 1988, Some recent developments in a concept of causality, Journal of Econometrics 39, 199-211.

Granger, C.W.J. and P.J. Thomson, 1987, Predictive consequences of using conditioning or causal variables, Econometric Theory 3, 150–152.

Hansen, L.P. and T.J. Sargent, 1991a, Two difficulties in interpreting vector autoregressions, in: L.P. Hansen and T.J. Sargent, eds., Rational expectations econometrics (Westview Press, Boulder) 77–119.

Hansen, L.P. and T.J. Sargent, 1991b, Identification of continuous time rational expectations from discrete data, in: L.P. Hansen and T.J. Sargent, eds., Rational expectations econometrics (Westview Press, Boulder) 219–235.

Harvey, A.C. and J.H. Stock, 1989, Estimating integrated higher-order continuous time autoregressions with an application to money-income causality, Journal of Econometrics 42, 319–336.

Harvey, A.C. and J.H. Stock, 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 (Blackwell, Cambridge MA) 55–70.

Marcellino, M., 1999, Some consequences of temporal aggregation in empirical analysis, Journal of Business and Economic Statistics 17, 129-136.

Marcet, A., 1991, Temporal aggregation of economic time series, in: L.P. Hansen and T.J. Sargent, eds., Rational expectations econometrics (Westview Press, Boulder) 237–282.

McCrorie, J.R., 2000, Deriving the exact discrete analog of a continuous time system, Econometric Theory 16, 998-1015.

McCrorie, J.R., 2003, The dimension of the aliasing problem for the Gaussian estimation of continuous time vector autoregressive models, preprint.

Ouliaris, S. and P.C.B. Phillips, 1994, COINT software for GAUSS.

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

Pierce, D.A. and L.D. Haugh, 1977, Causality in temporal systems: characterizations and a survey, Journal of Econometrics 5, 265-293.

Protter, P., 1990, Stochastic integration and differential equations (Springer, New York).

Renault, E., K. Sekkat and A. Szafarz, 1998, Testing for spurious causality in exchange rates, Journal of Empirical Finance 5, 47-66.

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 (Blackwell, Cambridge MA) 71–90.

Sims, C.A., 1971a, Discrete approximations to continuous time lag distributions in econometrics, Econometrica 39, 545-564.

Sims, C.A., 1971b, Approximate specification in distributed lag models, Bulletin of the International Statistical Institute 44, 285-294.

Sims, C.A., 1972, Money, income and causality, American Economic Review 62, 540–552.

Sims, C.A., 1987, Continuous and discrete time models, in: J. Eatwell, M. Milgate, and P. Newman, eds., The new Palgrave: A dictionary of economics (Macmillan, London) 626–628.

Sims, C.A., J.H. Stock and M.W. Watson, 1990, Inference in linear time series models with some unit roots, Econometrica 58, 113-144.

Stock, J.H. and M.W. Watson, 1988, Testing for common trends, Journal of the American Statistical Association 83, 1097–1107.

Stock, J.H. and M.W. Watson, 1989, Interpreting the evidence on money-income causality, Journal of Econometrics 40, 161–181.

Swanson, N.R., and C.W.J. Granger, 1997, Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions, Journal of the American Statistical Association 92, 357-367.

Toda, H.Y. and P.C.B. Phillips, 1993, Vector autoregressions and causality, Econometrica 61, 1367–1394.

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

Yu, J. and P.C.B. Phillips, 2001, A Gaussian approach for continuous time models of the short-term interest rate, Econometrics Journal 4, 221-225.

Table 1
Tests for integration, cointegration, and time trends

Univariate unit root tests

	Levels				First differences	
Series	$q_c^{t^2}(1,0)$	$q_c^t(1,0)$	$\tau_6^{t^2}$	τ_6^t	$q_c^t(1,0)$	τ_6^t
y	−11.7618	−10.3866	−3.5057	−3.7499	−421.1183	−5.8979
m	−2.0643	−1.5254	−1.4185	−1.3499	−563.1799	−3.8239
5% c.v.	−27.7289	−21.2162	−3.8607	−3.4318	− 21.2162	−3.4318

Significance of deterministic trends and multivariate cointegration tests

	t-statistics on:				Cointegration	
Series	constant	time	time^2		$q_c^t(2,1)$	$q_c^{t^2}(2,1)$
y	1.4416	−0.1332	−0.0255		−11.6123	−20.6597
m	0.4713	2.2801	−2.4459	5% c.v.	−30.3688	−36.9354

Note: $q_c^{t^2}(k,k-1)$ and $q_c^t(k,k-1)$ denote the Stock and Watson (1988) statistics for testing for k against $k-1$ unit roots using quadratic and linear time trends, respectively, based on a Parzen kernel estimate of the long run variance matrix with 12 lags; $\tau_p^{t^2}$ and τ_p^t denote the augmented Dickey-Fuller statistics for testing for a unit root using p lags of the dependent variable with quadratic and linear time trends, respectively; the t-statistics on the deterministic trend components are obtained in a regression of Δx on 6 of its lags plus a quadratic trend; 5% critical values are taken from the COINT package by Ouliaris and Phillips (1994).

Table 2
Causality tests in discrete time VARs

p	Unrestricted $\ln L$	Restricted $\ln L$	LR
6	3742.1	3737.3	9.6208 [0.1416]
12	3757.4	3752.0	10.7610 [0.5495]

Note: p denotes the order of the VAR; figures in square brackets denote marginal probability values; $\ln L$ denotes the maximised value of the likelihood function; LR denotes the likelihood ratio test statistic.

Table 3
Estimates of the continuous time model

Parameter	Exact discrete model		Approximation (31)		Approximation (32)	
Unrestricted model						
$[A_1]_{11}$	−8.6804	(1.0236)	−0.5532	(0.0510)	−15.9991	(0.3770)
$[A_1]_{12}$	−29.8579	(1.0303)	−0.0476	(0.0800)	0.2876	(2.1967)
$[A_1]_{21}$	0.9432	(0.3906)	0.0844	(0.0423)	0.5493	(0.4792)
$[A_1]_{22}$	−1.3574	(2.0038)	−0.6558	(0.0530)	−17.4982	(0.3775)
$[A_2]_{11}$	−3.4070	(0.4312)	1.6752	(0.0445)	−6.8921	(0.3376)
$[A_2]_{12}$	−9.2767	(0.7165)	−0.0834	(0.0767)	−1.2872	(0.3793)
$[A_2]_{21}$	1.7207	(0.2275)	−0.0238	(0.0445)	−0.5346	(0.3450)
$[A_2]_{22}$	−0.3231	(0.4892)	1.7651	(0.0442)	−5.3688	(0.3355)
$[M_\zeta]_{11}$	0.2072	(0.0095)	0.0078	(0.0002)	0.1759	(0.0070)
$[M_\zeta]_{21}$	−0.0146	(0.0025)	0.0050	(0.0002)	−0.1096	(0.0044)
$[M_\zeta]_{22}$	0.0005	(0.0131)	0.0003	(0.0002)	0.0010	(0.0065)
$\ln L$	3704.8497		3693.3063		3678.9276	
Restricted model						
$[A_1]_{11}$	−92.4106	(1.1647)	−0.5516	(0.0507)	−16.0899	(1.4624)
$[A_1]_{12}$	0.0000	na	0.0000	na	0.0000	na
$[A_1]_{21}$	−32.7328	(1.0850)	0.0843	(0.0327)	0.4378	(0.8883)
$[A_1]_{22}$	−35.3998	(1.1353)	−0.6578	(0.0529)	−17.4496	(1.4399)
$[A_2]_{11}$	−16.5473	(1.5857)	1.6814	(0.0444)	−6.7693	(0.7711)
$[A_2]_{12}$	0.0000	na	0.0000	na	0.0000	na
$[A_2]_{21}$	−5.9420	(1.3331)	−0.0241	(0.0272)	−0.5825	(0.5455)
$[A_2]_{22}$	−6.1268	(0.8638)	1.7616	(0.0442)	−5.4031	(0.7024)
$[M_\zeta]_{11}$	0.9158	(0.0314)	0.0078	(0.0002)	0.1758	(0.0152)
$[M_\zeta]_{21}$	−0.2192	(0.0100)	−0.0050	(0.0002)	−0.1096	(0.0092)
$[M_\zeta]_{22}$	0.3256	(0.0189)	0.0003	(0.0002)	0.0028	(0.0104)
$\ln L$	3687.4990		3691.2269		3677.7739	
Granger causality test						
LR	34.7014	[0.0000]	2.1588	[0.3398]	2.3074	[0.3155]

Note: Figures in parentheses denote asymptotic standard errors; figures in square brackets denote marginal probability values; $\ln L$ denotes the value of the maximised likelihood function; LR denotes the likelihood ratio tests statistic.