10

Temporal Aggregation of Economic Time Series

by Albert MARCET

Introduction

We call temporal aggregation the situation in which a variable that evolves through time can not be observed at all dates. This phenomenon arises frequently in economics, where it is very expensive to collect data on certain variables, and there is no reason to believe that economic time series are collected at the frequency required to fully capture the movements of the economy. For example, we only have quarterly observations on GNP, but it is reasonable to believe that the behavior of GNP within a quarter carries relevant information about the structure of the economy.

In order to give a mathematical structure to this problem, we assume that there is an underlying stochastic process in continuous time that is observed only at discrete intervals. This structure has been used by Sims (1971) and Geweke (1978) to describe the effects of temporal aggregation in the distributed lag model.

We will be concerned with the issues that arise in the study of linear predictions of future values of the variables given all information up to the present. In other words, if we have a vector of n variables y, we try to predict $y_i(t+1)$ using a function of the form $\sum_{k=0}^{\infty} \mu_k' y(t-k)$.

Since the fundamental *moving average representation* (henceforth MAR) of a stochastic process (or Wold decomposition) is a good summary of the properties of those predictions, we will describe what is the relationship between the Wold decomposition of the unobserved continuous time process and the Wold decomposition of the discrete sampled process.

This approach to time series analysis has been used widely in economics. In rational expectations models, agents are often assumed to

form their expectations using the kind of predictions mentioned above and the MAR has proved to be a useful tool in solving these models. Also, Sims (1980) has developed a method for interpreting time series using the MAR; this approach has been used in Litterman and Weiss (1985) and Bernanke (1986).

We focus on the following type of questions. What features of the continuous MAR are captured by the discrete MAR? What type of systematic biases will exist in the discrete MAR? How are the one-step-ahead prediction errors in continuous and discrete time related? Can we approximate in any sense the continuous model arbitrarily well by sampling the continuous process more frequently?

Our study can be useful in deciding if the results obtained in a given application of time series analysis can be attributed to time aggregation bias. If the econometrician suspects that this is the case, he should look for data collected at a finer interval or, alternatively, he could estimate a structural model in continuous time with discrete sampled data using, for example the techniques developed by Hansen and Sargent (1980b) and (1981d). These methods have been applied in Hansen and Sargent (1983a), Christiano (1984) and Christiano and Eichenbaum (1985, 1986). Other techniques have also been developed by Bergstrom (1976, 1983) and by Lo (1985).

In order to analyze the relationship between the continuous and discrete MAR, we introduce an approach that relates these from the point of view of the space of functions L^2. As an application of this approach, we analyze the effects of using data that have been averaged over a certain period of time, and show that a systematic bias will be present when this type of data are mixed with sampled data.

Due to the nature of the subject, it is inevitable that the exposition will become technical. An effort has been made in order to clarify the meaning of the propositions, and to give them an intuitive interpretation. In the conclusion, the main results of the paper are restated in a simple way, and some direct implications of our work for econometric practice are briefly discussed.

In terms of the issues addressed, our work is in the same spirit as that of Sims (1971) and Geweke (1978). It is worth noting that the distributed lag model (studied by Sims and Geweke), and the study of predictions using all data up to the present (studied in this paper) are two basically different ways of analyzing time series, so that our work is not an extension or a particular case of their model. Interestingly enough, in the distributed lag model, one effect of temporal aggregation is that the first few coefficients of the discrete approximation will, in general, be small compared with the corresponding coefficients in continuous time, while the effect of temporal aggregation on the MAR is, in general, the oppposite (see Section 3).

We will be concerned with a very general class of n-dimensional stochastic processes to be specified in Section 1, where we set up our problem and our notation. In Section 2, we discuss the relationship between the discrete and the continuous one-step-ahead innovations. In Section 3, we display a formula relating the MAR of the continuous process, to the MAR of the discrete (sampled) process, and we compare it with the formulas obtained by Sims and Geweke in the distributed lag model. We find that there is an intimate relationship between how good the predictions with sampled data are, and how well the discrete MAR approximates the continuous MAR; also, we use these theoretical results to discuss how and when the shape of the MAR of the sampled process will be very different from the shape of the continuous MAR. Section 4 contains certain results and examples that illustrate how Granger causality relationships are affected by time aggregation. The effects of using unit-averaged data, are studied in Section 5. In Section 6, we obtain some preliminary results on the issue of convergence of the MAR as the sampling frequency increases; the results in this section analyze the approximation of the continuous projection with the discrete projection. Finally, we discuss in an appendix the autoregressive representation of the sampled process; the results in the appendix are useful in other parts of the paper.

1. Definition of the Problem

Let us consider an n-dimensional stochastic process in continuous time y; we will denote by $H_y(t)$ the set that contains all linear combinations of $y_i(s)$ for all $s \leq t$ and $i = 1, \ldots, n$, plus the limits of convergent sequences of these linear combinations. More formally, $H_y(t) = cl\{\text{linear subspace generated by } y_i(s), \text{ for all } s \leq t,\ i = 1, \ldots, n\}$, (where the closure is taken with respect to a metric to be specified below). Intuitively, we can think of this set as containing all random variables of the form

$$\sum_{k=0}^{\infty} \mu_k' y(t_k)$$

where $t_k \leq t$, and $\mu_k \in \mathbf{R}^n$ are constant vectors.

We will treat the space of random variables as a metric space with the distance given by the mean-square difference; i.e., for any two ran-

dom variables x, z,

$$d(x, z) = (E(x-z)^2)^{\frac{1}{2}} .$$

One of the implications of using this distance, is that two random variables x, z will be considered equal when they are equal with probability one.

Consider the problem of predicting $y(t+\alpha)$, $\alpha > 0$, using all the information available up to time t. We define the best linear predictor of $y(t+\alpha)$ with information up to t, as the element of $H_y(t)$ that minimizes the mean square error

$$E(y(t+\alpha) - x)^2 \qquad \text{for } x \in H_y(t) .$$

We will call this predictor the projection of $y(t+\alpha)$ on $H_y(t)$, and we will denote it by $\eta(y(t+\alpha)|H_y(t))$. More formally, we define $\eta(y(t+\alpha)|H_y(t))$ as the element of $H_y(t)$ such that

$$d\big(\eta(y(t+\alpha)|H_y(t)),\ y(t+\alpha)\big) \leq d\big(x,\ y(t+\alpha)\big) \quad \text{for any} \quad x \in H_y(t) .$$

Next, we are going to define the problem of temporal aggregation.

Let (Ω, S, P) be a probability space, and let y be a multivariate stochastic process in continuous time, $y : \mathbf{R} \times \Omega \to \mathbf{R}^n$. In order to guarantee that a moving average representation exists in continuous time, we make the following assumptions on this process:

1. $E(y_i(t)) = 0$, $\text{var}(y_i(t)) < \infty$ for all $i = 1, \ldots, n$, and all $t \in \mathbf{R}$
2. y is covariance stationary.
3. y is linearly regular.

Then, letting $m \leq n$ be the rank of the process y, this process has a fundamental (or Wold) moving average representation (MAR), which can be expressed by

$$(1) \qquad y(t) = \int_0^\infty a(u)\, \zeta\,(t - du)$$

where $a : \mathbf{R} \to \mathbf{R}^{n \times m}$, and ζ is a vector of m orthonormal random measures. The matrix function a is unique up to multiplication by an orthonormal matrix.[1]

In this paper, we will often use the following properties of random measures: for any two square integrable functions $f, g : \mathbf{R} \to \mathbf{R}^{n \times m}$, the integrals $\int f d\zeta$ and $\int g d\zeta$ are well defined random variables, and

they satisfy: $E(\int f d\zeta \cdot \int g' d\zeta) = \int f \cdot g'$. We can use this property to find the covariance function of y; since

$$E\,(y(t) \cdot y(t-s)') = \int_0^\infty a(u)a(u-s)'du \quad \text{all } s \in \mathbf{R} .$$

In the rest of the paper, the initials MAR will stand for the fundamental moving average representation.

An important property of the MAR is that

$$\eta(y(t)|H_y(t-\alpha)) = \int_\alpha^\infty a(u)\zeta(t-du) \qquad \text{for all } \alpha > 0 .$$

Using the process y, we define another stochastic process in discrete time $Y : I \times \Omega \to \mathbf{R}^n$ on the same probability space (Ω, S, P), by setting

$$(2) \qquad Y(t, \omega) = y(t, \omega) \qquad \text{for all } t \in I, \text{ all } \omega \in \Omega$$

where I is the set of all integers. We will call Y the sampled process. In our notation, we will suppress the dependence of Y on ω.

Most of this paper relates the MAR of y to the *MAR* of Y, so that we have to guarantee that the MAR of Y exists.

Indeed, we can see that Y inherits all the 'good' properties of y. By definition, Y satisfies assumptions 1 and 2. We show next that Y is also linearly regular.

Let $\{x_k\}$ be any sequence of random variables. By definition of mean square convergence, $x_k \to 0$ as $k \to \infty$ if and only if $\text{var}(x_k) \to 0$ as $k \to \infty$. Since for each integer s $H_Y(s) \subset H_y(s)$, we have

$$\begin{aligned} 0 \leq \text{var}\big(\eta(Y_i(t)|H_Y(s))\big) &= \text{var}\big(\eta(y_i(t)|H_Y(s))\big) \\ &\leq \text{var}\big(\eta(y_i(t)|H_y(s))\big), \qquad i = 1, \ldots, n . \end{aligned}$$

Since y is assumed to be linearly regular, the term to the right of this string of inequalities goes to zero as $s \to -\infty$, so that the variance of $\eta(Y_i(t)|H_Y(s))$ also goes to zero, and Y is linearly regular.

Hence, the MAR of Y exists, and we can write

$$(3) \qquad Y(t) = \sum_{k=0}^{\infty} A_k\, \epsilon(t-k)$$

where A_k are $n \times m'$ matrices, and ϵ is an m'-dimensional vector of white noises with

$$\begin{aligned} E(\epsilon(t) \cdot \epsilon'(s)) &= 0 \quad \text{if } t \neq s \\ &= \Sigma \quad \text{if } t = s . \end{aligned}$$

Note that, unlike in the continuous time case, we do not require that the elements of ϵ be uncorrelated contemporaneously.

The requirement that (3) is the fundamental moving average representation for Y guarantees that

$$\eta(Y(t)|H_Y(t-r)) = \sum_{k=r}^{\infty} A_k\, \epsilon(t-k) \quad \text{for any integer } r \geq 0\,.$$

We will say that ζ and ϵ are fundamental for y and Y, respectively; it can be shown that $H_y(t) = H_\zeta(t)$ and $H_Y(t) = H_\epsilon(t)$.

It is possible to normalize the discrete MAR so that the A_k's and ϵ are uniquely determined. In our study, different normalizations will be useful for different purposes, and we leave this question open for the moment.

We can extend the function a to the negative reals by setting $a(u) = 0$ if $u < 0$ and, analogously, we set $A_k = 0$ for any integer $k < 0$. After doing this, we can express (1) and (3) in convolution notation:

$$\begin{aligned} y(t) &= a * \zeta(t) \qquad t \in \mathbf{R} \\ Y(t) &= A * \epsilon(t) \qquad t \in \mathbf{I}\,. \end{aligned}$$

2. The Innovation of the Discrete Sampled Process

Throughout this and the next two sections, we normalize the discrete MAR by setting $\epsilon_i(t) = Y_i(t) - \eta(Y_i(t)|H_Y(t-1))$, $i = 1, \ldots, n$; that is, $\epsilon(t)$ is the vector of one-step-ahead innovations in Y.

As Hansen and Sargent (1984, reprinted as Chapter 4 of this volume) point out, the innovation in discrete time can be written as

$$\epsilon(t) = \int_0^1 a(u)\zeta(t-du) + B_t \tag{4}$$

where $B_t = \eta(y(t)|H_y(t-1)) - \eta(Y(t)|H_Y(t-1))$. Clearly, since $H_Y(t-1) \subset H_y(t-1)$, we have $B_t \in H_y(t-1) = H_\zeta(t-1)$, so that $B_t \perp \int_0^1 a(u)\zeta(t-du)$. In words $\epsilon(t)$ is composed of two orthogonal parts: the one-step-ahead innovation in continuous time, and the difference between the one-step-ahead projections in continuous and in discrete time. By applying the law of iterated projections we have that

$$\eta(Y(t)|H_Y(t-1)) = \eta\left\{\eta(Y(t)|H_y(t-1))|H_Y(t-1)\right\},$$

so that B_t could be interpreted as the error made when we try to predict the *projection* in continuous time with sampled data.

From now on, we will assume that y has full rank. This is equivalent to assuming that the spectral density of y is positive definite at almost every frequency. This guarantees that Y has full rank as well, because the spectral density of Y is given by

$$f_Y(\omega) = \sum_{k=-\infty}^{\infty} f_y(\omega + 2\pi k) \tag{5}$$

(where f_χ denotes the spectral density of χ), so that f_Y is positive definite almost everywhere.[2] Hence, m' in the MAR of Y is equal to n.

Since $B_t \in H_\zeta(t-1)$, there exists a function $c\colon [1, \infty) \to \mathbf{R}^{n\times n}$ such that[3]

$$B_t = \int_1^\infty c(u)\zeta(t-du)\,.$$

If we let $c(u) = a(u)$ for $0 \leq u < 1$, equation (4) tells us that ϵ is related to ζ by:

$$\epsilon(t) = \int_0^\infty c(u)\zeta(t-du)\,. \tag{6}$$

Hansen and Sargent (1984) point out that one implication of (6) is that the one-step-ahead prediction error $\epsilon(t)$ will in general be correlated with innovations in continuous time that have happened before $t-1$. Also, they characterize c for the case that Y has an autoregressive representation.

Next, we give a general characterization of c in terms of a which will be useful later on in the paper. First of all, observe that using the same line of argument that led us to equation (6), we can conclude that there exists a function h such that

$$\eta(Y(t)|H_Y(t-1)) = \int_1^\infty h(u)\zeta(t-du) \tag{7}$$

where $h : \mathbf{R} \to \mathbf{R}^{n\times n}$. Clearly, by the definition of h and c, we conclude that $c = a - h$. We will now characterize h.

Define the space of functions L_n^2

$$L_n^2 = \left\{f : \mathbf{R} \to \mathbf{R}^n;\ \int_{-\infty}^{\infty} \|f(u)\|^2 du < \infty\right\}$$

where, the sign $\|\cdot\|$ inside the integral refers to the Eucledian norm in $\mathbf{R}^n$. We will adopt the convention that each $f(u)$ is a row vector. We endow the space L_n^2 with the inner product

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(u)\cdot g'(u)\, du = \int_{-\infty}^{\infty} \left[\sum_{i=1}^{n} f_i(u)\, g_i(u)\right] du\,.$$

It can be shown that this is a legitimate inner product, and that L_n^2 is a complete metric space in the metric induced by this inner product. Thus, L_n^2 is a Hilbert space.

Let us call a_i the i^{th} row of a, so that $a_i : \mathbf{R} \to \mathbf{R}^n$, and $a(u) = \begin{bmatrix} a_1(u) \\ \vdots \\ a_n(u) \end{bmatrix}$. Clearly, each a_i belongs to L_n^2, since:

$$||a_i||^2 = \int_{-\infty}^{\infty} ||a_i(u)||^2 du = \text{var}\,(y_i(t)) < \infty\,.$$

The following proposition displays a relationship between a and h in terms of the space L_n^2.

Define the set $A \subset L_n^2$ as follows:

$$A = \text{cl}\left\{ f \in L_n^2 : f(u) = \sum_{k=1}^{s} \sum_{j=1}^{n} \mu_k^j\, a_j\,(u-k) \text{ for some finite } s \right.$$
$$\left. \text{and some } \mu_k^j \in \mathbf{R} \right\}.$$

In words, A contains all finite linear combinations of the functions $a_j(u-k)$ for $j = 1, \ldots, n$ and $k = 1, 2, \ldots$, and the limits of convergent sequences of such linear combinations.

In the first section we defined the projection of a random variable on a certain set of random variables. We can now think of doing the same with functions in L_n^2 : for a given $f \in L_n^2$ and a set $S \subset L_n^2$, we define the projection of f on S denoted by $\eta\,(f|S)$ as the element of S such that

$$d\big(\eta\,(f|S),\, f\big) \le d(g,\, f) \qquad \text{for any} \quad g \in S$$

in the distance induced by the inner product of L_n^2.

Proposition 1: For each $i = 1, \ldots n$,[4]

$$h_i = \eta\,(a_i|A) \qquad \text{in the metric of } \; L_n^2\,.$$

Proof: Fix i. Since A is a closed linear subspace of a Hilbert space, the projection $\eta\,(a_i|A)$ exists and it is the only element of A for which the following orthogonality conditions hold:

$$\Big[a_i - \eta\,(a_i|A)\Big] \perp f \text{ for any } f \epsilon A. \quad \text{But}$$

$$\int_0^{\infty} [a_i\,(u) - h_i(u)] \cdot [a_j\,(u-k)]'du = \text{cov}\left[\int_0^{\infty} [a_i(u) - h_i(u)]\zeta(t-du)\,,\right.$$
$$\left.\int_0^{\infty} a_j\,(u-k)\zeta(t-du)\right] = \text{cov}\left[Y_i(t) - \eta(Y_i(t)|H_Y(t-1))\,,\, Y_i(t)\right] = 0$$

for all $k = 1, 2, \ldots$, $j = 1, \ldots, n$. Therefore, the orthogonality conditions in L_n^2 are satisfied with $\eta(a_i|A) = h_i$, and the only thing that is left to show is that $h_i \in A$.

By definition of $\eta\,(Y_i(t)|H_Y(t-1))$, there exists a sequence $\{z_\nu\}$ of random variables such that $z_\nu = \sum_{k=1}^{s_\nu} \sum_{j=1}^{n} \lambda_{k,j}^{\nu}\, Y_j(t-k)$ for some coefficients $\lambda_{k,j}^{\nu}$ and a finite s_ν, and such that $\{z_\nu\}$ satisfies

$$\text{(8)} \qquad \text{var}\left[\eta(Y_i(t)|H_Y(t-1)) - z_\nu\right] \to 0 \quad \text{as } \nu \to \infty\,.$$

Consider the functions f_ν given by $f_\nu(u) = \sum_{k=1}^{s_\nu} \sum_{j=1}^{n} \lambda_{k,j}^{\nu}\, a_j(u-k)$ for $\nu = 1, 2, \ldots$. Clearly $f_\nu \in A$, and it is easy to show that $h_i \in L_n^2$. From these observations, it is clear that we can write

$$||h_i - f_\nu||^2 = \text{var}\left[\int_0^{\infty} [h_i(u) - f_\nu(u)]\zeta(t-du)\right] =$$

$$\text{var}\left[\eta(Y_i(t)|H_Y(t-1)) - \sum_{k,j} \lambda_{k,j}^{\nu} \int_0^{\infty} a_j(u-k)\zeta(t-du)\right] =$$

$$\text{var}\left[\eta(Y_i(t)|H_Y(t-1)) - \sum_{k,j} \lambda_{k,j}^{\nu}\, Y_j(t-k)\right] \to 0 \quad \text{as } \nu \to \infty$$

where we have used (7), the fact that $\int_0^{\infty} a_j\,(u-k)\zeta(t-du) = \int_0^{\infty} a_j(u)\,\zeta(t-k-du) = Y_j(t-k)$, and (8). Therefore, $f_\nu \to h_i$ and, since A is a closed set, $h_i \in A$. ∎

This proposition tells us that h_i is (very close to) a function of the form $\sum_{k=1}^{\infty} \lambda_k\, a(u-k)$, $\lambda_k \epsilon R^n$, with the λ_k's chosen so as to make h_i "as close as possible" to a_i (where "close" is in terms of the distance $\int ||h_i - a_i||^2$).[5]

Once we have characterized h, we can find c in equation (6) by setting $c = a - h$.[6]

The characterization in Proposition 1 tells us the nature of h, and therefore of c, as functions in L_n^2, and allows us to handle examples quite easily. Later in this paper, we state some other properties of h.

The next proposition is very easy to prove, and is stated mainly for future reference.

Proposition 2: For any given $i = 1, \ldots n$, the following are equivalent:

i) $c_i(u) = 0$ for almost every $u \geq 1$

ii) $\eta(Y_i(t)|H_Y(t-1)) = \eta(y_i(t)|H_y(t-1))$

iii) $a_i(u) = \eta(a_i|A)(u)$ **for almost every** $u \geq 1$.

Proof: For *i)* $\Rightarrow$ *ii)*, use

$$y_i(t) = \int_0^1 a_i(u)\zeta(t-du) + \eta(y_i(t)|H_y(t-1))$$

$$Y_i(t) = \int_0^\infty c_i(u)\zeta(t-du) + \eta(Y_i(t)|H_Y(t-1)) \; .$$

Now, since $a = c$ in the interval $[0, 1)$ we get *ii)* by equating the right hand sides to these two equations.

To show *ii)* $\Rightarrow$ *iii)*, we note that *ii)* implies $\int_1^\infty h_i d\zeta = \int_1^\infty a_i d\zeta$, and use the comment in footnote (6).

That *iii)* $\Rightarrow$ *i)*, is obvious, since $c_i = a_i - \eta(a_i|A)$. ∎

3. MAR of the Sampled Process

In the last section, we found a characterization of the one-step-ahead innovation in discrete time in terms of the underlying continuous time process. We are now in a position to characterize the MAR coefficients of Y.

Proposition 3: The matrices A_k in the MAR of Y are given by:

$$(9) \qquad A_k = \left[\int_0^\infty a(u+k)c(u)'du\right] \cdot \left[\int_0^\infty c(u)\,c(u)'du\right]^{-1}$$

where $c = a - \eta(a \mid A) \equiv a - h$.

Proof: Let $V = E[\epsilon(t) \cdot \epsilon(t)']$. Now,

$$A_k V = E\left[Y(t+k)\cdot\epsilon(t)'\right] = E\left[\int_0^\infty a(u)\zeta(t+k-du)\cdot\int_0^\infty c(u)'\zeta(t-du)\right] =$$

$$E\left[\int_0^\infty a(u)\zeta(t+k-du)\cdot\int_0^\infty c(u-k)'\zeta(t+k-du)\right] = \int_0^\infty a(u)c(u-k)'du =$$

$$\int_0^\infty a(u+k)c(u)'du$$

where the first equality is easily derived by expressing Y in its MAR, and where we use (6). Finally, since

$$V = E\left[\epsilon(t)\cdot\epsilon(t)'\right] = \int_0^\infty c(u)c(u)'du$$

and the assumption that y is full rank guarantees that V is invertible, we have shown the proposition. ∎

In convolution notation, this result can be expressed as:

$$(10) \qquad A_k = \left[a^* c^\top(k)\right]\left[c^* c^\top(0)\right]^{-1} \quad \text{for all } k = 0, 1, \ldots$$

where $c(u)^\top = c(-u)'$ for all u.

Proposition 3 tells us that A_k is a weighted average of the function a over the interval $[k, \infty)$ with c as the weighting kernel.

We saw in the last section that $c_i = a_i - \eta(a_i|A)$, so that the i^{th} row of c is the error made when we project a_i on the set A. It is easy to show that for any function f in A, $f(u) = [0, \ldots, 0]$ for all $u \in [0, 1)$,[7] so that $\eta(a_i|A)(u) = 0$ and $c_i(u) = a_i(u)$ for $u \in [0, 1)$. For $u \geq 1$, if $\eta(a_i|A)$ is any good in approximating a_i, we would expect $c_i(u)$ to be small, and the graph of the elements of c_i will look more or less like Figure 1:

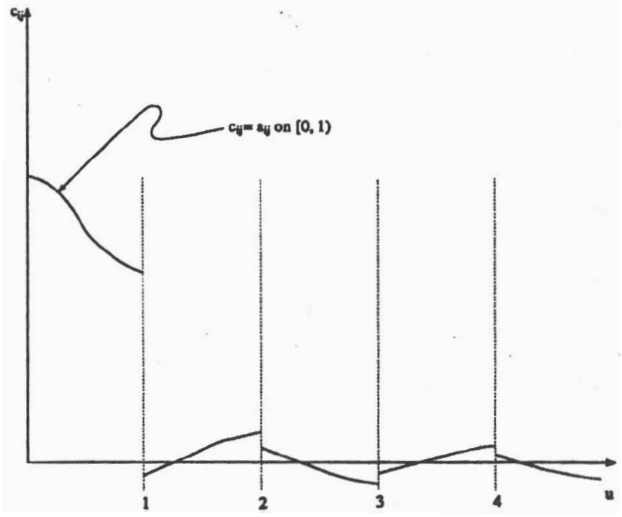

Figure 1.

This illustrates the fact that c will in general give most of the weight to values of a on the interval $[k, k+1)$. On the other hand, c will give more or less weight to values of a_{ij} on the interval $[k+1, \infty)$ depending

on how large c_{ij} is in the interval $[1, \infty)$; but since

$$\int_1^\infty ||c_i||^2 = \text{var}\left[\eta(Y_i(t)|H_Y(t-1)) - \eta(y_i(t)|H_y(t-1))\right] ;$$

we have that c_i is small on the interval $[1, \infty)$ when the projection in discrete time approximates well the projection in continuous time. In this case, A_k depends largely on values of a in the interval $[k, k+1)$. In the extreme case that those two projections are equal, Proposition 2 applies, and:

$$(11) \qquad A_k = \left[\int_0^1 a(u+k)a(u)'du\right] \cdot \left[\int_0^1 a(u)\,a(u)'du\right]^{-1}$$

so that A_k only depends on values of a on the interval $[k, k+1)$.

There are certain similarities between formula (10) and the formulas that Sims and Geweke obtain in their work on time aggregation of the distributed lag model. In particular, in the distributed lag model the discrete parameters are also obtained by applying a weighting kernel to the continuous time parameters. However, in our case, the kernel in (10) is one-sided, and in general discontinuous, while the opposite was true in the distributed lag model.

Next, we are going to substantiate our claim about the non-continuity of c. For this purpose, it is enough to think of the case of a univariate process. Assume that a is continuous everywhere except at zero (in other words, $a(0) \neq 0$). Since $c(u) = a(u) - \sum_{k=1}^{\infty} \lambda_k a(u-k)$, for any integer, ν, $c(\nu) = a(\nu) - \sum_{k=1}^{\nu} \lambda_k a(\nu-k)$, because $a(\nu-k) = 0$ for $k > \nu$. Therefore, at $u = \nu$ the function c is the sum of $\nu+1$ functions such that one of them is discontinuous at $u = \nu$ (because, for $k = \nu$ in the summation sign above, $\lambda_\nu a(\nu-\nu) = \lambda_\nu a(0)$, and we assumed that a was discontinuous at $u = 0$) and the remaining ν functions are continuous at $u = \nu$. Therefore, if $\lambda_k > 0$ for all k, c will be discontinuous at all integers.

Equation (10) shows how the coefficients in the i^{th} row of A_k (i.e., the coefficients of the i^{th} variable Y_i) are affected by all the rows in a, so that the i^{th} row of A_k will in general depend on the moving average coefficients in continuous time of all the elements of y. This is the phenomenon that Geweke called "contamination", and that also appeared in the model he studied. The coefficients A_k will be "contaminated" even when the projections in continuous and discrete time coincide, as formula (11) shows. The only general case in which contamination disappears, is when $a_{ij} \equiv 0$ for all $i = j$ (i.e., when the two matrices to

the right of (10) are both diagonal). This is a very special case, since it amounts to assuming that $E(y_i(t) \cdot y_j(t')) = 0$ for all $i \neq j$, and all $t, t' \in \mathbf{R}$.

The rest of this section discusses what distortions can be generated by temporal aggregation in the MAR in view of the above results.

We begin by displaying one type of distortion that will be present in most cases. For any integer ν, we can write

$$\text{var}\left(Y_i(t) - \eta(Y_i(t)|H_Y(t-\nu))\right) \geq \text{var}\left(y_i(t) - \eta(y_i(t)|H_y(t-\nu))\right)$$

and

$$\sum_{k=0}^{\nu-1} A_{k_i} \Sigma A_{k_i}' \geq \int_0^\nu a_i(u)a_i(u)'du ,$$

where A_{k_i} is the i^{th} row of A_k. In the one variable case, this can be written as

$$\sigma_\epsilon^2 \sum_{k=0}^{\nu-1} (A_k)^2 \geq \int_0^\nu a(u)^2 du .$$

One way to interpret this is that the first coefficients of the discrete MAR will be too large in absolute value so that, for a univariate model, if we plot the discrete and continuous MAR's in the same graph, we will obtain a version of Figure 2.

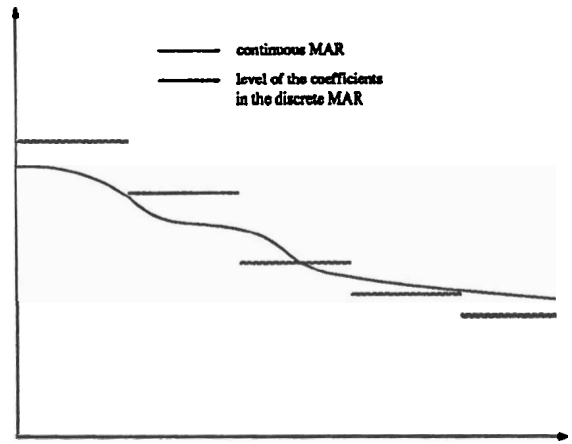

Figure 2.

We could distinguish two reasons why the discrete MAR can be a bad approximation of the continuous MAR; one is contamination, and the other A_k depending on values of a on the interval $[k+1, \infty)$. We now give examples that illustrate the type of distortions the second "reason" can generate, and in what cases this distortion will be severe. We will only consider one-variable examples, in this section, since their effect is present in these cases. These examples, illustrate how Propositions 1, 2 and 3 can be used to analyze time aggregation.

Examples 3.1 and 3.2 display two cases in which the projections in continuous and in discrete time coincide, so that $c(u) = 0$ for $u \in [1, \infty)$, and the shape of A_k is similar to the shape of a. In Example 3.3 we find a very simple MAR in continuous time that will have a distorted discrete MAR. We have said that when c is zero in the interval $[1, \infty)$, A_k is an average of a in the interval $[k, k+1)$; but even in this case, if c is positive *and* negative in the interval $[0, 1)$, A_k will not be a proper average of a in the interval $[0, 1)$; we illustrate this point in Example 3.4. Example 3.5 effectively shows one way to construct continuous MAR's that will generate distorted discrete MAR's. Example 3.6 discusses the effects of unit-averaging. Even though the main message of Example 3.6 is that when $a(0) = 0$ and a is continuous, we may expect distortions in the discrete MAR, Example 3.7 shows that this is not always true. Finally, Examples 3.1, 3.4 and 3.7 are concrete illustrations of the aliasing problem. They show three different continuous MAR's that generate exactly the same sampled process: in particular, Y is an autoregressive process of order one in all these examples.

Example 3.1: Probably the simplest case we can deal with is the AR(1) process in continuous time

$$y(t) = \int_0^\infty e^{-\lambda u}\zeta(t-du) \qquad t \in \mathbf{R}, \quad \lambda > 0\,.$$

In this case, $e^{-\lambda}y(t-1) = \int_1^\infty e^{-\lambda u}\zeta(t-du) = \eta(y(t)|H_y(t-1))$, and $e^{-\lambda u}\,y(t-1) \in H_Y(t-1)$; the orthogonality conditions in continuous time imply that, in particular, $[Y(t) - e^{-\lambda}Y(t-1)] \perp Y(t-k)$ for all $k = 1, 2, \ldots$, so that $e^{-\lambda}Y(t-1)$ *is* the projection in discrete time.

Therefore, Y has an autoregressive representation (henceforth ARR) of order 1, with parameter equal to $e^{-\lambda}$, so that the MAR of Y is given by:

$$A_k = e^{-\lambda k} = a(k) \qquad \text{for all integers } k\,.$$

For illustrative purposes, we next find A_k by using propositions 1, 2, 3. Since $e^{-\lambda}a(u-1) = a(u)$ for all $u \geq 1$, clearly $e^{-\lambda}a(u-1) = \eta(a|A)(u)$ for all $u \geq 1$, so that Proposition 2 applies, and we can use formula (11) to conclude:

$$A_k = \left[\int_0^1 e^{-\lambda(u+k)}\,e^{-\lambda u}\,du\right] \cdot \left[\int_0^1 e^{-\lambda 2u}\,du\right]^{-1} = e^{-\lambda k}$$

Example 3.2: If a is constant over intervals $[k, k+1)$ and y one-dimensional, letting $\epsilon(t) = \int_0^1 \zeta(t-du)$ and $A_k = a(k)$, we can write:

$$Y(t) = \sum_{k=0}^{\infty}\int_k^{k+1} a(u)\zeta(t-du) = \sum_{k=0}^{\infty} a(k)\int_k^{k+1}\zeta(t-du) = \sum_{k=0}^{\infty} A_k\epsilon(t-k)\,.$$

This expresses Y in moving average form. It has to be true that the Fourier transform of a has no zeroes on the right half of the complex plane. This guarantees that the Fourier transform of $\{A_k\}$ has no zeroes inside the unit disk, and that this is associated with the Wold representation of Y.

In this case, as in the previous example, the discrete MAR equals the continuous MAR sampled at integers, and from the way ϵ was chosen, it is clear that the one-step-ahead innovations in discrete and continuous time coincide, so that the one-step-ahead projections in discrete and continuous time are equal.

Example 3.3: Hansen and Sargent (1984) showed that for a MAR a that satisfies:

- a is continuous at all $u \geq 0$
- $a(0) \neq 0$
- Y has an ARR,

if $c = 0$ almost everywhere in $[1, \infty)$ then Y has an ARR of order 1.

In our framework, this is displayed in the following way: if Y had an autoregressive representation of order more than 1, some λ_k would be different from zero for $k > 1$. Then $\eta(a|A)$ would be discontinuous at $u = k$ (this can be deduced by using the line of argument on page 243 showing that c is in general discontinuous), so it is not possible that $\eta(a|A)(u) = a(u)$ for a.e. $u \in [1, \infty)$. Thus, $c \neq 0$ in a subset of $[1, \infty)$ of positive measure.

This implies that apparently well behaved a's will yield distorted A_k's. For example, the MAR given by : $a(u) = e^{-\lambda_1 u} + e^{-\lambda_2 u}$ has a function c with $\int_1^\infty ||c||^2 > 0$. Indeed, if this was not true, then $c \equiv 0$ almost everywhere in $[1, \infty)$, and by Proposition 2 there would exist a

constant μ such that $\mu\, a\,(u-1) = a(u)$ for all $u \geq 1$. In particular, for $u = 1$ and $u = 2$ this would imply

$$= e^{-\lambda_1} + e^{-\lambda_2} \qquad \text{and} \qquad \mu(e^{-\lambda_1} + e^{-\lambda_2}) = e^{-\lambda_1 2} + e^{-\lambda_2 2}$$

which in turn would imply $(e^{-\lambda_1} + e^{-\lambda_2})^2 = e^{-\lambda_1 2} + e^{-\lambda_2 2}$, which is not possible.

Example 3.4: Let a be as depicted in Figure 3.

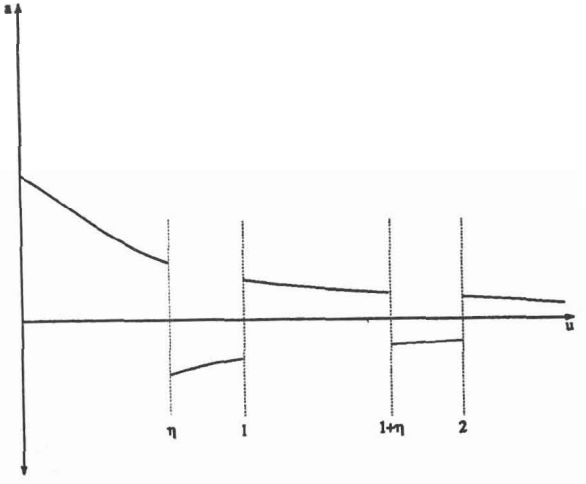

Figure 3.

In Figure 3, $|a(u)| = e^{-\lambda u}$ for all $u \geq 0$, but for each integer k, if $u \in [k+\eta,\ k+1)$, then $a(u) < 0$.

Again, $e^{-\lambda} a(u-1) = a(u)$ for all $u \geq 1$, so that, by Proposition 2, $\eta(Y_i(t)|H_Y(t-1)) = e^{-\lambda} Y(t)$. Note that $A_k = e^{-\lambda k}$, and the discrete MAR fails completely to capture the oscillations in a.

The interest of this example is to show that it is possible for the projection with discrete data to be as good as the projection in continuous time, and yet the discrete time MAR be very different from the continuous MAR. Indeed, in this case we can apply Proposition 2 to conclude that the two projections coincide, but the discrete MAR fails completely to capture the oscillations in the continuous MAR. We may expect this to happen when a is both positive and negative in large parts of the interval $[0, 1)$. Recall that, from formula (11), we know

that A_k is largely a weighted average of a on the interval $[k,\ k+1)$, where the weights are given precisely by a on the interval $[0, 1)$; therefore, if a takes both positive and negative values, A_k is not a proper average of a on the interval $[k,\ k+1)$.

Example 3.5: For $\{A_k\}$ not to be distorted, it is necessary that:

$$\int_1^2 c^2 = \int_1^2 (a(u) - h(u))^2 du = \int_1^2 (a(u) - \mu \cdot a(u-1))^2 du = 0$$

so that, unless there exits some constant μ such that $a(u) = \mu\, a(u-1)$ for a.e. $u \in [1, 2)$, we will have distortions of the discrete MAR. Keeping this in mind, it is quite easy to generate examples in which $\int_1^2 c^2$ is very large, by considering functions a that look very different in the interval $[0, 1)$ and in the interval $[1, 2)$. The graphs in Figure 4 illustrate this situation.

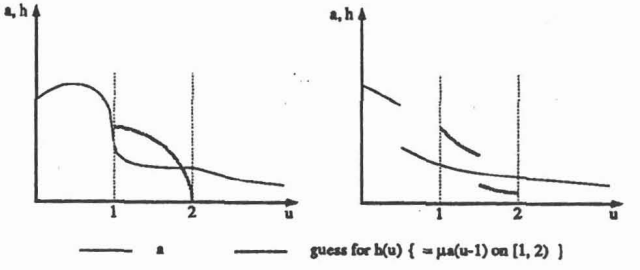

Figure 4.

Example 3.6: It has been suspected for a long time that when a is continuous as a function defined on the whole real line (i.e., when $a(0) = 0$), the discrete MAR does not approximate a well. Hansen and Sargent (1984), have an example with large distortions. We will now argue that this is in fact the case if a is continuous and positive at $u = 1$.

We have argued before that for $u \in [1, 2)$ $h(u) = \mu\, a(u-1)$ for some coefficient μ. Therefore, if a is continuous and strictly positive for all $u > 0$, but $a(0) = 0$, the graph of a and h will be as depicted in Figure 5.

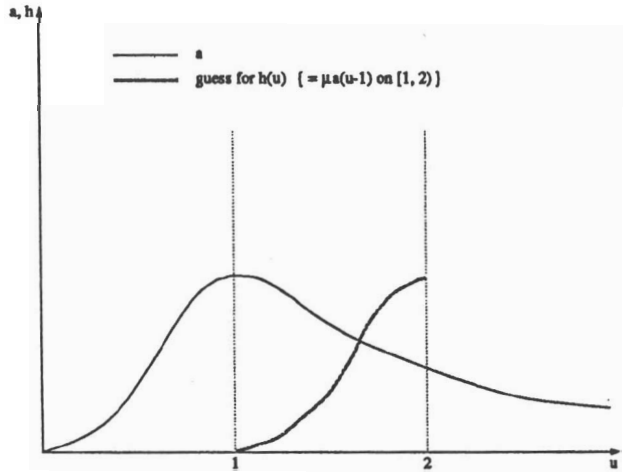

Figure 5.

As illustrated in Figure 5, $\int_1^2 c^2 = \int_1^2 (a-h)^2$ is large in relation to $\int_0^1 c^2 = \int_0^1 a^2$. In this case, A_k will depend more on values of a in the interval $[k+1,\ k+2)$ than on values of a in $[k,\ k+1)$.

Example 3.7: In view of the previous example, we should not conclude that any MAR in continuous time with $a(0) = 0$ will give a bad approximation in discrete time. For example, take a function a that is continuous in the interval $[0, 1]$, satisfies $a(0) = a(1) = 0$, and can be extended to $[1, \infty)$ by setting: $a(u) = \ell a(u-1)$ for some $0 < \ell < 1$, for all $u \in [1, \infty)$; the graph of this function will be as depicted in Figure 6.

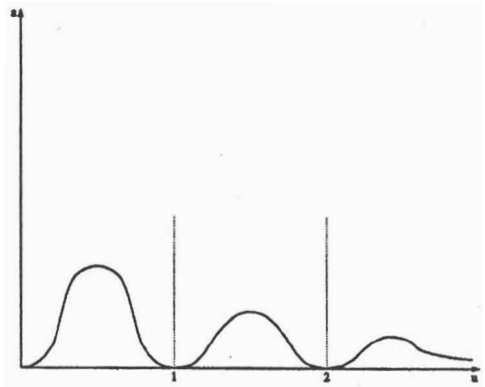

Figure 6.

Because of the way a was defined, clearly Proposition 2 (ii) applies, and

$$\eta(Y(t)|H_Y(t-1)) = \eta(y(t) \mid H_y(t-1))\ ,$$

so that Y has an autoregressive representation of order 1.

4. Granger Causality

Consider a bivariate process in continuous time y, such that y_2 does not Granger cause y_1. It is well known that, in general, Y_2 will Granger cause Y_1. It is possible to deduce this by combining results in Sims (1971, 1972b).[8]

There are cases, however, in which the absence of Granger causality from the second variable to the first, *does* carry over to the sampled process. One case in which this happens is when y_1 and y_2 are uncorrelated at all dates; then Y_1 and Y_2 will also be uncorrelated at all dates, so that past Y_2's will not help in predicting Y_1. Another case is given by the following proposition, which says that if the first variable can be predicted with equal accuracy whether we use continuous or discrete data and if y_2 does not Granger cause y_1, then Y_2 does not Granger cause Y_1.

Proposition 4: If $\eta(y_1(t) \mid H_y(t-1)) = \eta(Y_1(t) \mid H_Y(t-1))$ and y_2 fails to Granger cause y_1, then Y_2 will also fail to Granger cause Y_1.

Proof: The conditions in this proposition imply:

$$\begin{aligned}\eta(y_1(t) \mid H_y(t-1)) &= \eta(y_1(t) \mid H_{y_1}(t-1)) \\ &= \eta(Y_1(t) \mid H_Y(t-1)) ,\end{aligned}$$

therefore $\eta(Y_1(t) \mid H_y(t-1)) \in H_{y1}(t-1) \cap H_Y(t-1)$. By definition, $H_{y_1}(t-1) \cap H_Y(t-1) = H_{Y_1}(t-1)$, so that $\eta(Y_1(t) \mid H_y(t-1)) \in H_{Y_1}(t-1)$ and, since $H_{Y_1}(t-1) \subset H_y(t-1)$, we have that $\eta(Y_1(t) \mid H_y(t-1)) = \eta(Y_1(t) \mid H_{Y_1}(t-1))$. Combining this last equality with the first condition of the proposition tells us that $\eta(Y_1(t) \mid H_{Y_1}(t-1)) = \eta(Y_1(t) \mid H_Y(t-1))$. ∎

This type of result is not surprising. The reason that, in general, Y_2 will Granger cause Y_1 even when past y_2 do not help in predicting y_1, is that the values of y_1 between integers enter in $\eta(Y_1(t) \mid H_y(t-1))$. Then, in the discrete projection, past values of Y_2 could help predict Y_1 because they help predict values of y_1 between integers. However, in the case that the above proposition considers, there is no point in trying to predict the values of y_1 between integers, so that there is no room for past Y_2 to help predict Y_1.

We can construct cases in which the effect of temporal aggregation on Granger causality is the opposite of the effect described at the beginning of this section, that is, Y_2 does not Granger cause Y_1 even though y_2 Granger causes y_1. Consider the MAR in continuous time given by the functions a_{ij} depicted in Figure 7.

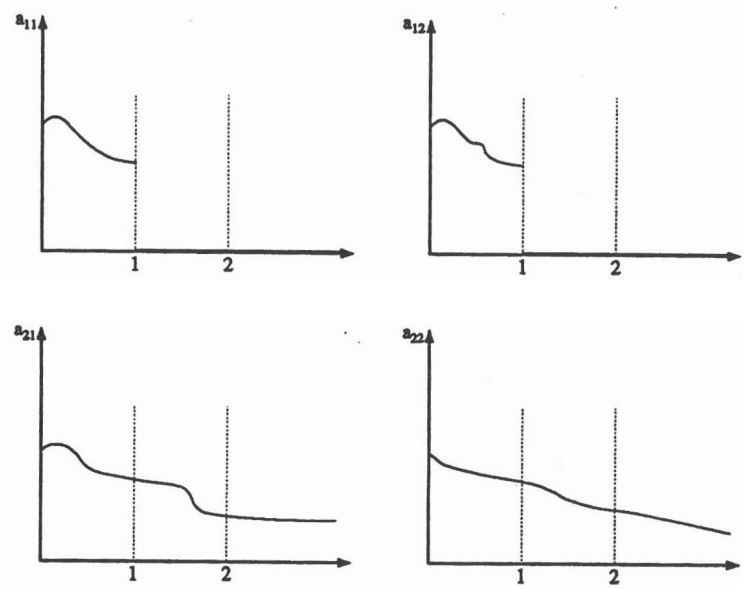

Figure 7.

In this case, clearly $\text{cov}(Y_1(t), Y_2(t-2)) = \int_0^\infty [a_{11}(u)a_{21}(u-s) + a_{12}(u)a_{22}(u-s)]\,du = 0$ for integers $s \geq 1$, and past Y_2 will not help in predicting current Y_1. But, since neither a_{11} nor a_{12} is identically zero, y_2 Granger causes y_1.

Another case in which this happens is shown in Figure 8.

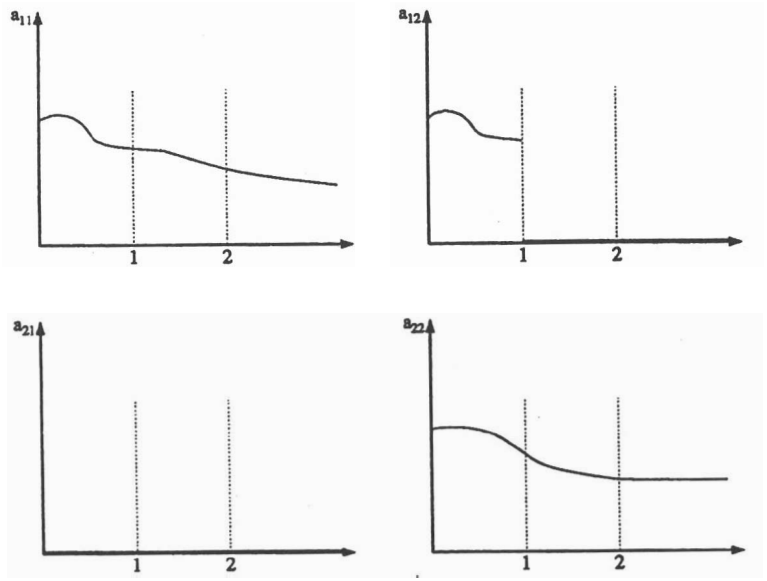

Figure 8.

Again, $Y_1(t)$ will be uncorrelated with past Y_2's and Y_2 will not Granger cause Y_1.

These two examples are very special cases; the characteristic they have in common is that much of the variance of $Y_1(t)$ is due to innovations in the interval $[t, t-1)$.

5. Unit-Averaged Data

Sometimes the data available to us on a given variable consist of averages of observations over a certain period of time. For example, our monthly data on a given series may consist of the average of the weekly data for that month. Say that we have unit-averaged observations on the first m-variables. We can model this situation as follows. We observe the process Y in discrete time given by

$$Y_i(t,\omega) = \int_0^1 y_i(t-s,\omega)ds \qquad i = 1,\ldots,m\ ;$$

$$Y_i(t,\omega) = y_i(t,\omega) \qquad i = m+1,\ldots,n, \text{ for all } t \in \mathbf{I}, \text{ and all } \omega \in \Omega$$

where the above integral sign refers to Lebesgue integral.

Clearly, any of the unit-averaged variables Y_i , $i \leq m$, can be obtained by sampling the continuous time process $\tilde{y}$ given by:

$$\tilde{y}_i(t) = \int_0^1 y_i(t-s)ds \qquad \text{all } t \in \mathbf{R}\ .$$

In order to apply the results of Section 3, we have to find the MAR of the continuous time process $\tilde{y}$. Let us define the function $\ell : \mathbf{R} \to \mathbf{R}$ as

$$\ell(u) = 1 \quad \text{if } u \in [0,1)$$
$$\ell(u) = 0 \quad \text{otherwise ;}$$

so that we can write

$$\tilde{y}_i(t) = \int_{-\infty}^{\infty} \ell(u)\, y_i(t-u)\, du$$

and we have that

$$(15)\quad f_{\tilde{y}_i}(\omega) = \hat{\ell}(\omega) f_{y_i}(\omega)\hat{\ell}'(\omega) = \hat{\ell}(\omega)\hat{a}_i(\omega)\hat{a}_i'(\omega)\hat{\ell}'(\omega), \qquad \omega \in [-\pi,\pi]$$

where f_χ denotes the spectral density of χ, and $\hat{b}$ denotes the Fourier transform of the function b; in the above equation we have used the formula for the spectral representation of linear combinations of random variables (see Rozanov (1967)). Since $\hat{\ell}$ has no zeroes in the left half plane, (15) indicates that $\ell^* a_i$ gives the MAR for $\tilde{y}_i$, and we can write

$$\tilde{y}_i(t) = \int_0^{\infty} \tilde{a}_i(u)\zeta(t-du)$$

where

$$(15')\qquad \tilde{a}_i(u) = \ell * a(u) = \int_0^1 a_i(u-s)ds\ ,\ t \in \mathbf{R},\ i = 1,\ldots,m\ .$$

The next proposition states that $\tilde{a}$ is always smoother than a. As usual, we denote the space of functions that can be differentiated s times by C^s.

Proposition 5: For any i, $j = 1, \ldots, n$

i) If $a_{ij} \in L^1$ [i.e., $\int |a_{ij}| < \infty$] then $\tilde{a}_{ij}$ is continuous.
ii) If, in addition, $a_{ij} \in C^s$ then $\tilde{a}_{ij} \in C^{s+1}$.

Proof:

i) Define $F(x) = \int_{-\infty}^{x} a_{ij}(s)ds$; then F is an indefinite integral; therefore, F is continuous.[9] Since $\tilde{a}_{ij}(u) = \int_{u-1}^{u} a_{ij}(s)ds = F(u) - F(u-1)$, $\tilde{a}_{ij}$ is continuous.

ii) If a_{ij} is continuous at x, then F has a derivative at x, and $F'(x) = a_{ij}(x)$.[10] ∎

Note that Proposition 5 refers to $\tilde{a}$ as a function defined on the whole real line, so that, since $\tilde{a}(u) = 0$ for $u < 0$, i) in Proposition 5 implies that $\tilde{a}(0) = 0$. Remember that Example 3.6 dealt with the case of a continuous function a. The comments made there apply to this section.

The rest of this section analyzes the effects of using unit-averaged and sampled data at the same time.

Consider a 2-dimensional process in continuous time; we have sampled observations on the second variable, but we have unit-averaged observations on the first. It is common practice in this case to estimate a 2-variable system in discrete time, consisting of point-in-time observations for one variable, and unit-averaged observations for the other. We are going to argue that this practice will systematically overstate the importance of the second variable (consisting of sampled observations) in predicting the first one.

We will discuss this by giving an informal, general argument, and by displaying several simulations.

The general argument goes as follows. Consider a two-dimensional process y with MAR represented by the graph in Figure 9.

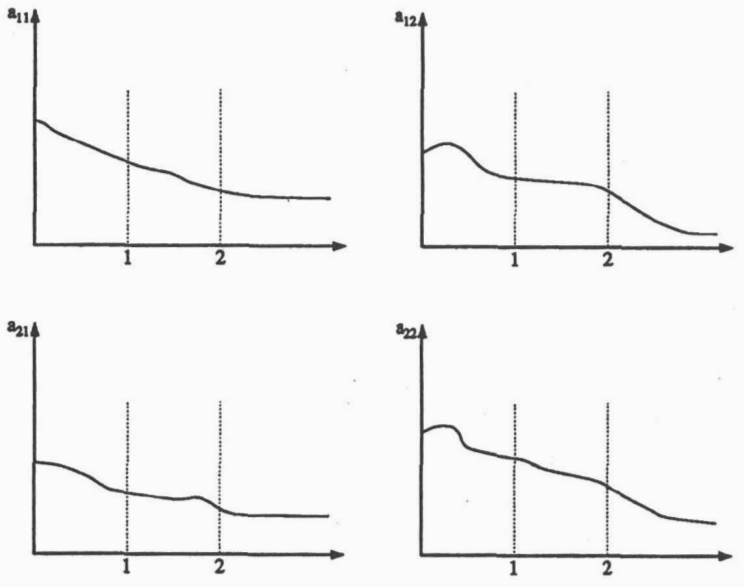

Figure 9.

The discrete system obtained by mixing unit-averaged and sampled data in the way described above, is equivalent to sampling the continuous time process $\hat{y}(t)' = (\tilde{y}_1(t),\ y_2(t))'$. The MAR of $\hat{y}$ is given by $\hat{a}(u) = \begin{bmatrix} \tilde{a}_1(u) \\ a_2(u) \end{bmatrix}$, where $\tilde{a}_1$ is given by equation (15′), so that the graph

of $\hat{a}$ will be as in Figure 10.

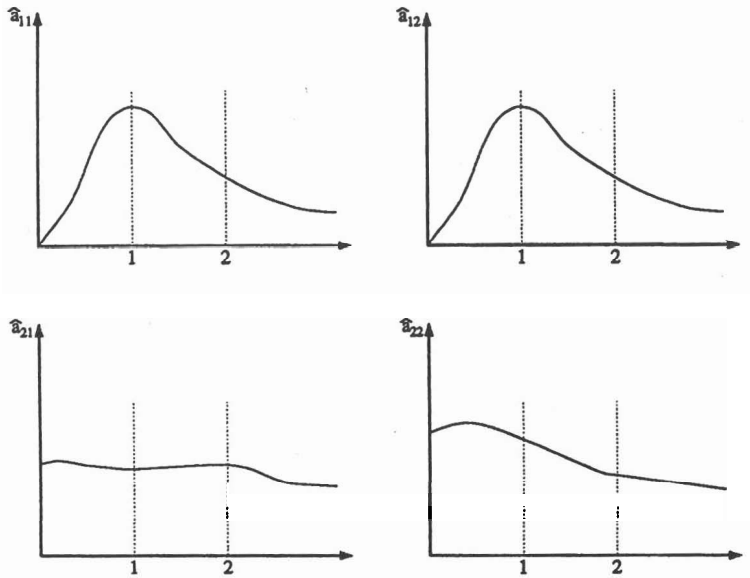

Figure 10.

From the discussion in Section 2 (and the appendix) we have learned that, letting $\mu_k \in \mathbf{R}^2$ be the vectors such that

$$\eta(Y^1(t) \mid H_Y(t-1)) = \sum_{k=1}^{\infty} \mu_k Y(t-k) ,$$

the μ's are those vectors that make the function $h_1(u) = \sum_{k=1}^{\infty} \mu_k\, a(u-k)$ as close as possible to the function a_1.

We will argue that, if Y_1 is unit-averaged, the portion of the variance of Y_1 explained by Y_2 will be larger than if both variables were sampled, due to the fact that the first elements of the vectors μ_k will, in general, be smaller.

We will look, first, at the sampled system. Let us consider how much the first elements of μ_k can contribute to make h_1 close to a_1. Assume we set the second element of μ_k equal to zero, and we try to approximate a_1 with a function of the type $g(u) = \sum \lambda_k a_1(u-k)$.

Then, g has to look like Figure 11.

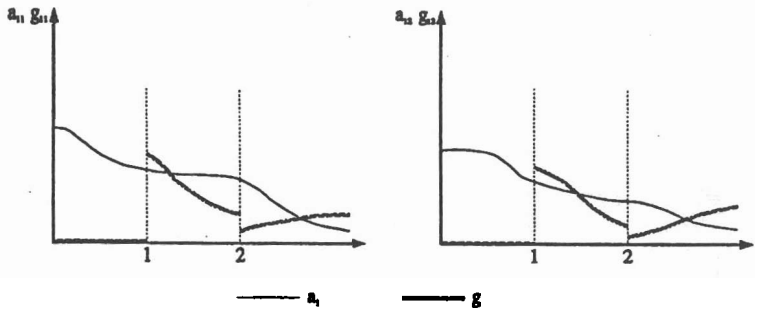

Figure 11.

However, if we did the same with $\hat{a}$ instead of a, and let $\hat{g}(u) = \sum_{k=1}^{\infty} \lambda_k \hat{a}_1(u-k)$ and then compare $\hat{g}$ with $\hat{a}_1$, we have a version of Figure 12 (see Example 3.6).

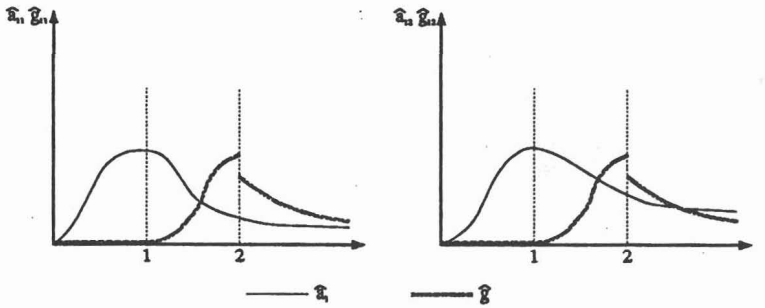

Figure 12.

The point of these graphs is to demonstrate that we can approximate a_1 better with a function like g than we can approximate $\hat{a}_1$ with a function like $\hat{g}$.

Now, since the vectors μ_k are chosen in order to make h_1 as close to a as possible, we would expect that after unit-averaging the first elements of μ_k are smaller. Therefore, past Y_1 will not be as helpful in predicting $Y_1(t)$ if Y_1 is unit-averaged.[11]

The same intuition can be used to justify the claim that if only Y_1 is unit-averaged, past Y_1 will not be as effective in predicting Y_2 as in the case that both variables are sampled, and that the same is true if

we compare the mixed system with the system in which both variables are averaged.

Next, we will present a few simulations that support our claims. We have simulated several AR(2), discrete time processes with two variables. Then we have temporarily aggregated them in three ways: the "sampled system" is obtained by sampling both variables every three periods; the "averaged system" is obtained by taking the average of each variable over three periods; finally, the "mixed system" is obtained by averaging the first and sampling the second variable every three periods. This does not correspond exactly with the situation discussed above, since the variables are not averaged in a continuous way, but it can be interpreted as an approximation; furthermore, it corresponds to a situation that arises often in practice, when the researcher has two variables recorded at different time-intervals (say monthly and quarterly) and has to aggregate over time one of the variables.

We have run a VAR for each of these temporally aggregated processes. In the tables below we report the decomposition of variance for each process. Since the variance decomposition can be affected by the order in which the Choleski decomposition is obtained, we have computed the decomposition for both orderings in all the cases.[12]

If our claims are true, the percentage variance of Y_1 and Y_2 explained by Y_1 would be smaller in the "mixed" than in the other two systems. In these simulations, the effect of mixing averaged and sampled data is almost always the one predicted on the last page, and in some cases the changes are very large (particularly in Table 1).

Only in one out of the six simulations we report did the change in the decomposition of variance of one of the variables go clearly in the other direction (Table 2, first ordering, decomposition of Y_1). Far from proving a theorem, we have just presented a heuristic argument, so that, probably, our intuition is not true for all processes. On the other hand, the result in Table 2 may be due to sampling error or to the fact that, as we pointed before, the type of unit-averaging that we have performed in the simulations is not quite the continuous time averaging discussed in the rest of this section.

In any event, these simulations do support the claim that, in general, by unit-averaging the first variable, Y_1 becomes less important in determining both Y_1 and Y_2.

Table 1

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = y(t) = \begin{bmatrix} 1.04, & .65 \\ .59, & .98 \end{bmatrix} y(t-1) + \begin{bmatrix} -.36, & -.33 \\ -.29, & -.33 \end{bmatrix} y(t-2) + \varepsilon(t)$$

Ordering 1: Y_1, Y_2

Variance of ...	Percentage explained by Y_1 in each system		
	Y_1, Y_2 sampled	Y_1, Y_2 averaged	Y_1 aver. Y_2 sampled
Y_1	56	35	15
Y_2	23	10	4

Ordering 2: Y_2, Y_1

Variance of ...	Percentage explained by Y_1 in each system		
	Y_1, Y_2 sampled	Y_1, Y_2 averaged	Y_1 aver. Y_2 sampled
Y_1	85	93	.5
Y_2	85	92	.2

Table 2

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = y(t) = \begin{bmatrix} 1.37, & -.03 \\ .41, & 1.06 \end{bmatrix} y(t-1) + \begin{bmatrix} -.44, & .02 \\ -.28, & -.23 \end{bmatrix} y(t-2) + \varepsilon(t)$$

Ordering 1: Y_1, Y_2

Variance of ...	Percentage explained by Y_1 in each system		
	Y_1, Y_2 sampled	Y_1, Y_2 averaged	Y_1 aver. Y_2 sampled
Y_1	87	92	95
Y_2	49	55	40

Ordering 2: Y_2, Y_1

Variance of ...	Percentage explained by Y_1 in each system		
	Y_1, Y_2 sampled	Y_1, Y_2 averaged	Y_1 aver. Y_2 sampled
Y_1	97	97	97
Y_2	37	40	21

Table 3

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = y(t) = \begin{bmatrix} 1.2, & .5 \\ .2, & .9 \end{bmatrix} y(t-1) + \begin{bmatrix} -.4, & -.2 \\ -.1, & -.2 \end{bmatrix} y(t-2) + \varepsilon(t)$$

Ordering 1: Y_1, Y_2

Variance of ...	Percentage explained by Y_1 in each system		
	Y_1, Y_2 sampled	Y_1, Y_2 averaged	Y_1 aver. Y_2 sampled
Y_1	73	66	48
Y_2	8	2	2

Ordering 2: Y_2, Y_1

Variance of ...	Percentage explained by Y_1 in each system		
	Y_1, Y_2 sampled	Y_1, Y_2 averaged	Y_1 aver. Y_2 sampled
Y_1	55	57	36
Y_2	10	10	4

Table 4

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = y(t) = \begin{bmatrix} 1.09, & .33 \\ .29, & 1.06 \end{bmatrix} y(t-1) + \begin{bmatrix} -.32, & -.18 \\ -.16, & -.3 \end{bmatrix} y(t-2) + \varepsilon(t)$$

Ordering 1: Y_1, Y_2

Variance of ...	Percentage explained by Y_1 in each system		
	Y_1, Y_2 sampled	Y_1, Y_2 averaged	Y_1 aver. Y_2 sampled
Y_1	81	87	67
Y_2	21	26	9

Ordering 2: Y_2, Y_1

Variance of ...	Percentage explained by Y_1 in each system		
	Y_1, Y_2 sampled	Y_1, Y_2 averaged	Y_1 aver. Y_2 sampled
Y_1	80	83	61
Y_2	23	21	8

Table 5

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = y(t) = \begin{bmatrix} .43, & .21 \\ .34, & .29 \end{bmatrix} y(t-1) + \begin{bmatrix} .14, & .03 \\ .05, & .12 \end{bmatrix} y(t-2) + \varepsilon(t)$$

Ordering 1: Y_1, Y_2

Variance of ...	Percentage explained by Y_1 in each system		
	Y_1, Y_2 sampled	Y_1, Y_2 averaged	Y_1 aver. Y_2 sampled
Y_1	89	84	87
Y_2	20	23	23

Ordering 2: Y_2, Y_1

Variance of ...	Percentage explained by Y_1 in each system		
	Y_1, Y_2 sampled	Y_1, Y_2 averaged	Y_1 aver. Y_2 sampled
Y_1	95	88	86
Y_2	15	20	17

Table 6

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = y(t) = \begin{bmatrix} .37, & .26 \\ .28, & .34 \end{bmatrix} y(t-1) + \begin{bmatrix} .13, & .04 \\ .04, & .12 \end{bmatrix} y(t-2) + \varepsilon(t)$$

Ordering 1: Y_1, Y_2

Variance of ...	Percentage explained by Y_1 in each system		
	Y_1, Y_2 sampled	Y_1, Y_2 averaged	Y_1 aver. Y_2 sampled
Y_1	90	93	78
Y_2	20	36	15

Ordering 2: Y_2, Y_1

Variance of ...	Percentage explained by Y_1 in each system		
	Y_1, Y_2 sampled	Y_1, Y_2 averaged	Y_1 aver. Y_2 sampled
Y_1	88	86	81
Y_2	18	31	18

6. Convergence of the Discrete MAR as the Sampling Interval Goes to Zero

For each $\delta > 0$, we define a discrete stochastic process Y^δ on our original probability space, by setting

$$Y^\delta(t,\omega) = y(t\delta,\omega) \quad \text{for all } t \in \mathrm{I}, \quad \text{all } \omega \in \Omega$$

and let the MAR of Y^δ be:

$$Y^\delta(t) = \sum_{k=0}^{\infty} A_k^\delta \varepsilon^\delta(t-k)\,. \tag{17}$$

We are interested in determining under what conditions and in what sense the δ-discrete MAR converges to the continuous MAR, and the discrete innovation converges to the continuous one, as the sampling interval δ goes to zero.

Note that the way we have defined $Y^\delta(t)$, this 't' does not correspond to the point 't' in real time (instead, it corresponds to the date '$t\delta$').

We will change our notation slightly in this section, by setting:

$$H_{Y^\delta}(q) = \mathrm{cl}\left\{ \text{linear subspace spanned by } Y^\delta(t) \text{ for } \delta t \leq q \right\} \text{ for } q \in R$$

so that 'q' now refers to the corresponding date *in real time*. Actually, $Y^\delta(q)$ itself may not be defined (when q/δ is not an integer). We will write $Y^\delta(t/\delta)$ regularly, assuming implicitly that this is well defined, that is, assuming that t/δ is an integer.

One type of convergence that will prove useful for our purposes and that has interest in itself, is whether, for a given $\alpha > 0$, the projection of $y(t+\alpha)$ on information up to t in continuous time, can be approximated arbitrarily well by observations of Y^δ up to 't' ('t' a date in real time), as δ goes to zero. The next Proposition deals with this question.

In this section we will use repeatedly the following property of Hilbert spaces: let S be such a space; for any sequence $\{x_\nu\}$ in S such that x_ν converges to x (in the distance induced by the inner product on S) then $\langle x_\nu, y\rangle \to \langle x, y\rangle$ as $\nu \to \infty$ for all $y \in S$.

Let $B : \mathbf{R} \to \mathbf{R}^{n\times n}$ be the autocovariance function of y, defined by:

$$B(s) = E(y(t+s)\cdot y(t)'), \quad s \in \mathbf{R}\,.$$

Lemma 3: Let y satisfy assumptions 1, 2, 3 in the introduction, then $y(t+s) \to y(t)$ in mean square as $s \to 0$. This in turn implies that B is continuous at 0.

Proof: For any $i = 1, \ldots, n$, using the MAR of y, we have

$$E(y_i(t) - y_i(t+s))^2 = \int_0^\infty ||a_i(u) - a_i(u+s)||^2 du =$$
$$= \sum_{j=1}^{n} \int_0^\infty |a_{ij}(u) - a_{ij}(u+s)|^2 \, du \, .$$

By Theorem 8.19 in Wheeden and Zygmund (1977), we know that any square integrable function $f : R \to R$, has the property that $\int |f(u) - f(u+s)|^2$ goes to zero as s goes to zero. Since each a_{ij} is square integrable, each element in the above sum goes to zero as $s \to 0$. Thus $y_i(t+s) \to y_i(t)$ in mean square.

Therefore, by the property of Hilbert spaces stated above, we have:

$$B_{ij}(s) = E(y_i(t) \cdot y_j(t+s)') \to E(y_i(t) \cdot y_j(t)') = B_{ij}(0) \text{ as } s \to 0$$

and B is continuous at 0. ∎

Proposition 6: If y satisfies assumptions 1, 2, 3 in the introduction, then

$$\eta\left[Y_i^\delta(t/\delta) \mid H_{Y^\delta}(t-\alpha)\right] \to \eta\left[y_i(t) \mid H_y(t-\alpha)\right] \quad \text{as } \delta \downarrow 0 \, .$$

for any $i = 1, \ldots, n$, $\alpha > 0$, and $t \in \mathbf{R}$.

Proof: Because of stationarity, it is enough to show the theorem for $t = 0$.

By definition of $H_y(-\alpha)$, there exists a sequence of random variables $\{a_\nu\}$ such that each a_ν can be written as

$$a_\nu = \sum_{k=1}^{s_\nu} \left[\lambda_k^\nu\right]' \cdot y(t_k^\nu) \quad \text{for some } \lambda_k^\nu \in \mathbf{R}^n, \, s_\nu \text{ finite, and } t_k^\nu < -\alpha \, ,$$

with the property that $a_\nu \to \eta(y_i(0)|H_y(-\alpha))$ as $\nu \to \infty$.

Now choose any $\theta > 0$, and choose ν' such that

$$\left[E(a_{\nu'} - \eta(y_i(0) \mid H_y(-\alpha)))^2\right]^{\frac{1}{2}} < \theta/2 \, .$$

First, we want to show that for each δ, we can select an element b_δ of $H_{Y^\delta}(-\alpha)$ such that $b_\delta \to a_\nu$, as $\delta \downarrow 0$ (remember that we already fixed ν'). Set:

$$b_\delta = \sum_{k=1}^{s_{\nu'}} \left[\lambda_k^{\nu'}\right]' y(t_k^\delta)$$

where t_k^δ is the closest number to $t_k^{\nu'}$ that is divisible by δ; here $t_k^{\nu'}$, $\lambda_k^{\nu'}$ and $s_{\nu'}$ are those in the above expression for $a_{\nu'}$.

Using the previous lemma, since $t_k^\delta \to t_k^{\nu'}$, clearly $y(t_k^\delta) \to (t_k^{\nu'})$ in mean square as $\delta \downarrow 0$, so that $b_\delta \to a_{\nu'}$. Hence, for the θ chosen above, there exists some δ' such that if $\delta < \delta'$ we can apply the triangle inequality to conclude that

$$\left[E\left(b_\delta - \eta(y_i(0)|H_y(-\alpha))\right)^2\right]^{\frac{1}{2}} \equiv d\left[b_\delta, \, \eta(y_i(0)|H_y(-\alpha))\right] \leq$$

$$\leq d\left[b_\delta, \, a_{\nu'},\right] + d\left[a_{\nu'}, \, \eta(y_i(0)|H_y(-\alpha))\right] < \theta/2 + \theta/2 = \theta \, .$$

Finally, the law of iterated projections implies that

$$\eta\left(Y_i^\delta(0)|H_{Y^\delta}(-\alpha)\right) = \eta\left[\eta(y_i(0)|H_y(-\alpha)) \;\middle|\; H_{Y^\delta}(-\alpha)\right] \, ,$$

so that, given θ, there exists δ' such that for any $\delta < \delta'$

$$d\left[\eta(Y_i^\delta(0)|H_{Y^\delta}(-\alpha)), \, \eta(y_i(0)|H_y(-\alpha))\right] \leq$$

$$\leq d\left[b_\delta \, , \eta(y_i(0)|H_y(-\alpha))\right] < \theta$$

and we have shown the proposition. ∎

This proposition assures us that the convergence of the α-step-ahead projections obtains under very general conditions. Next we turn to the discussion of the convergence of the MAR.

The first problem we have to deal with when comparing different δ-MAR's is how to normalize them. There are two standard ways of normalizing a discrete MAR: the first requires that $\epsilon(t)$ equals the one-step-ahead innovation, the other is to multiply this innovation by one of the 'square roots' of its variance-covariance matrix, in order to get a white noise in the MAR that has a covariance matrix equal to the identity matrix.

Clearly, if we insist that ϵ^δ be equal to the one-step-ahead innovation (i.e. that $\epsilon^\delta(t) = Y^\delta(t) - \eta(Y^\delta(t)|H_{Y^\delta}(t\delta - \delta))$) we will not get convergence to the continuous MAR, since $A_0^\delta = I$ for all δ (where I is the identity matrix), and $a(0)$ need not be close to I. Later on we will discuss another important problem that this normalization presents.

On the other hand, we should not insist on normalizing by setting $E(\epsilon^\delta(t) \cdot \epsilon^\delta(t)') = I$ either. In this case, it can be shown that as $\delta \downarrow 0$ the δ-MAR goes to zero for any process.

We propose the following normalization. Let ξ^δ be the one-step-ahead innovation of the δ-process, i.e., $\xi^\delta(t) = Y^\delta(t) - \eta(Y^\delta(t)|H_{Y^\delta}(t\delta - \delta))$. We define the white noise ϵ^δ in (17) as

$$(18) \quad \epsilon^\delta(t) = (\delta)^{\frac{1}{2}} W^{-1} \xi^\delta(t) \quad \text{where} \quad WW' = E\left[(\xi^\delta(t) \cdot \xi^\delta(t)')\right]^{-1} .$$

There are many matrices W that satisfy the above condition. By placing some additional requirements on W, we can resolve this uniqueness problem.[13]

This is a natural normalization in the sense that $\epsilon^\delta(t)$ mimics the properties of a random measure in continuous time (see section 1). Recall that the crucial property of a random measure ζ is that for any interval $\Delta \subset \mathbf{R}$ of length δ, $E(\zeta(\Delta) \cdot \zeta(\Delta)') = \delta \cdot I$. We could interpret ϵ^δ as a random measure defined only on intervals of the form $[k\delta, (k+1)\delta)$, and assigning measure $\epsilon^\delta(k)$ to each of these intervals. Then, the analogue to the above property of ζ is to require that $E(\epsilon^\delta(k) \cdot \epsilon^\delta(k)') = \delta \cdot I$, which is attained with the normalization given by (18).

This is relevant for econometric practice. If one were to compare the MAR's of two series collected at different intervals, the normalization that should be used for each series is the one given by (18).

Define the *function* $A^\delta : \mathbf{R} \to \mathbf{R}^{n \times n}$,

$$\begin{aligned} A^\delta(u) &= A_k^\delta \quad \text{for } u \geq 0 , \quad k \text{ such that } u \in [k\delta, (k+1)\delta) \\ &= 0 \quad \text{for } u < 0 \end{aligned}$$

where A_k^δ corresponds to a given normalization that agrees with (18). In words, A^δ is a step function, and the value that A^δ takes at each step is the value of the MAR coefficient that corresponds to that date.

We will discuss one sense in which A^δ approximates a as $\delta \downarrow 0$. It can be easily checked that:

$$(19) \qquad \int_0^\infty ||A_i^\delta||^2 = \int_0^\infty ||a_i||^2 .$$

This tells us that each row of A^δ belongs to L_n^2, and it has the same norm as the corresponding row of a.

The next proposition states one sense in which A^δ approximates a.

Proposition 7: For any $\alpha > \beta \geq 0$, and any $q \geq 0$:

$$(20) \quad \int_\beta^\alpha A^\delta(u+q) \cdot A^\delta(u)' \, du \to \int_\beta^\alpha a(u+q) \cdot a(u)' \, du \quad \text{as } \delta \downarrow 0 .$$

Proof: Fix α and q; from Proposition 6 and the property of Hilbert spaces mentioned above, the following is true:

$$(21) \qquad \begin{aligned} \rho_q^\delta &\equiv E\left[y(q) \cdot [Y^\delta(0) - \eta(Y^\delta(0)|H_{Y^\delta}(-\alpha))]'\right] \\ &\to E\left[y(q) \cdot [y(0) - \eta(y(0)|H_y(-\alpha))]'\right] \equiv \rho \quad \text{as } \delta \downarrow 0 . \end{aligned}$$

Using the MAR of y (in continuous time) we can write ρ in (21) as

$$(22) \qquad \rho = E\left[\int_0^\infty a(u)\zeta(q-du) \cdot \left[\int_0^\alpha a(u)\zeta(0-du)\right]'\right] = \int_0^\alpha a(u+q)a(u)' \, du .$$

It is enough to show that (20) holds for any sequence $\{\delta_\nu\}$ such that $\delta_\nu \downarrow 0$ as $\nu \to \infty$. Let q_ν be the closest real number to q that is divisible by δ_ν. Let us define ρ^ν by

$$\rho^\nu \equiv E\left[y(q_\nu) \cdot [Y^{\delta_\nu}(0) - \eta(Y^{\delta_\nu}(0)|H_{Y^{\delta_\nu}}(-\alpha))]'\right]$$

Since $q_\nu \to q$, $y(q_\nu) \to y(q)$ as $\nu \to \infty$, and it is true that $|\rho_q^{\delta_\nu} - \rho^\nu| \to 0$ as $\nu \to \infty$. Therefore, using (21), $\rho^\nu \to \rho$.

Now, using the δ_ν-discrete MAR of Y^{δ_ν} we can show that

$$\rho^\nu = \sum_{k=0}^{\alpha/\delta_\nu} A_k^{\delta_\nu} \cdot \left[A_k^{\delta_\nu} + (q_\nu/\delta_\nu)\right]' \cdot \delta_\nu = \int_0^\alpha A^\delta(u+q) \cdot A^\delta(u) \, du .$$

This, (22), and the fact that $\rho^\nu \to \rho$, finishes the proof for the case $\beta = 0$.

The case $\beta > 0$ is easily shown by noting that for any function f,

$$\int_\beta^\alpha f = \int_0^\alpha f - \int_0^\beta f . \qquad \blacksquare$$

The interpretation of this proposition is that the impulse-response function in continuous time of innovations that happen during a fixed period in real time can be approximated by the impulse-response function of innovations over the same period of real time derived from the δ-sampled model.

The results reported in this section so far seem to indicate that with data collected at fine enough intervals, the discrete time model will be a good approximation to the continuous time model. Nonetheless, we have to interpret these results with caution. Propositions 6 and 7

depend crucially on the fact that we fix a period in continuous time first, and then let δ vary. Also, the result in Proposition 7 depends on the normalization given by (18). In the rest of this section, we intend to show how important these points are.

Assume that instead of normalizing ϵ^δ by (18), we set it equal to the one-step-ahead innovation of the δ-sampled process, so that the MAR of Y^δ is now:

$$Y^\delta(t) = \sum_{k=0}^{\infty} A_k \xi^\delta(t-k) \quad \text{for} \quad \xi^\delta(t) = Y^\delta(t) - \eta(Y^\delta(t)|H_{Y^\delta}(t-1)) .$$

It makes no sense to ask whether $\xi^\delta(t)$ converges to some random variable, since $\text{var}(\xi^\delta(t)) \to 0$ as $\delta \downarrow 0$, but we may be interested in determining if the correlation coefficient between ξ^δ and the δ-step-ahead innovation in continuous time ρ_i^δ

$$\rho_i^\delta \equiv \frac{E\left[\xi_i^\delta(t) \cdot \int_0^\delta a_i(u)\zeta(t-du)\right]}{\left[\text{ var }(\xi_i^\delta(t)) \cdot \text{ var }\left(\int_0^\delta a_i(u)\zeta(t-du)\right)\right]^{\frac{1}{2}}}$$

goes to one as $\delta \downarrow 0$. If this convergence does not obtain, it will cast some doubt on the usual practice of interpreting the one-step-ahead innovation as an approximation to the innovation in continuous time over a period of length δ.

We have not yet explored this type of convergence under general conditions, but the next proposition shows that for a certain class of one-dimensional processes in continuous time, ρ^δ does not go to one. We denote the first and second derivatives from the right by $a'(u+)$ and $a''(u+)$.

Proposition 8: Let a be a one-dimensional, continuous time MAR that satisfies:

i) $a(0) = 0$
ii) $a(u) > 0$ for $u > 0$ and u close to zero
iii) $a''(\cdot+)$ exists at $u = 0$
iv) $a'(\cdot-)$ is bounded near zero.

Then ρ^δ does not go to one as $\delta \downarrow 0$.

Proof: (in appendix available from the author).

The set of MAR's in continuous time that satisfy the conditions of this proposition is an important class of models. Imagine that we have a continuous time process with a function a in the MAR that has a first derivative from the right at zero. Then, by unit-averaging this process for all $t \in \mathbf{R}$, we obtain another process that satisfies conditions i) to iv).

7. Conclusion

In what follows, we make an attempt to summarize our results in a non-technical way..

The coefficients of the fundamental moving average representation (MAR) of the discrete, sampled process, are given by

$$A_k = \left[\int_0^\infty a(u+k)\, c(u)' du\right] \left[\int_0^\infty c(u)c(u)' du\right]^{-1}$$

where a is the MAR in continuous time, and c operates as a weighting function of the continuous MAR. The above equation tells us that when the function c is small on the interval $[1, \infty)$ A_k is mostly an average of a on the interval $[k, k+1)$, and we can consider A_k as a good approximation to a in this sense.

The function c has been thoroughly analyzed in Section 2. There, we have seen that c is large in the interval $[1, \infty)$ when the projection with continuous data is much more accurate than the projection with discrete data.

This restricts largely the type of distortions that can be attributed to temporal aggregation: if an econometrician estimates a MAR that does not agree with what he expects for the particular problem that he is studying, he can ask himself whether the coefficients he observes may have been generated by the above formula, given a function a that *is* of the right form. Also, he may have an *a priori* idea of how good predictions with discrete data are, compared to predictions with continuous data; from this, he can form an *a priori* idea about the function c, and guess how distorted the discrete MAR may be. For example, we were able to show that a systematic effect of time aggregation is to increase the absolute size of the first few coefficients of the MAR.

In Section 5, we have analyzed the problem of unit-averaging data. The formula for the MAR in continuous time that corresponds to unit-averaged data is given by (15'), and it basically implies that this MAR will be smoother, and its mass is shifted one unit to the right, in relation to the MAR and the original process.

Another result of that section is that by mixing variables that are unit-averaged and variables that are sampled in the same discrete time series, we will systematically overstate the importance of the variables that are sampled in determining all the variables in the system. This

result is obtained by using the projection approach of section 2, and it illustrates how this approach can be used to derive conclusions that are relevant in econometric practice.

Finally, we study some issues related to the approximation of the model in continuous time by collecting data at finer and finer intervals. We show convergence of the α-step ahead forecast, and we propose a normalization that is convenient for the study of these approximations. Convergence of the discrete time Wold decompositions to the continuous time is studied in Marcet (1987).

Appendix A: Autoregressive Representation of Y

We begin this section by noting that the function h in (7), can not only be approximated by a function of the form $f(u) = \sum \mu_k\, a(u-k)$, for $k\mu_k \in \mathbf{R}^n$, but that it actually is such a function. This is established in the following lemma.

Lemma 1: Given that $i = 1, \ldots, n$, there exist $\lambda_k \in \mathbf{R}^n$ $k = 1, 2, \ldots$ such that:

$$h_i(u) = \sum_{k=1}^{\infty} \lambda_k'\, a(u-k) \quad \text{for almost every } u \in \mathbf{R}$$

where the λ_k's depend on i.

Proof:

By Proposition 1, $h \in A$, so that there exist functions $f_\nu \in L_n^2$, $\nu = 1, 2, \ldots$, such that $f_\nu(u) = \sum_{k=1}^{s_\nu} \mu_{k,\nu}'\, a(u-k)$ for some $\mu_{k,\nu} \in \mathbf{R}^n$, and s_ν finite, and such that $f_\nu \to h_i$ in the metric of L_n^2 as $\nu \to \infty$. In particular, this implies that

$$\int_1^2 ||f_\nu - f_m||^2 \to 0 \text{ as } \nu, m \to \infty\,. \tag{12}$$

Since $a(s) = 0$ for $s < 0$, each f_ν has the property:

$$f_\nu(u) = \mu_{1,\nu}'\, a(u-1) \quad \text{for } u \in [1, 2)\,,$$

which together with (12) implies that

$$\begin{aligned} &\int_1^2 ||\mu_{1,\nu}' \cdot a(u-1) - \mu_{1,m}' \cdot a(u-1)||^2 = \\ &\int_0^1 ||(\mu_{1,\nu}' - \mu_{1,m}') \cdot a(u)||^2 \to 0 \text{ as } \nu, m \to \infty\,. \end{aligned} \tag{13}$$

Next, we will argue that $\{\mu_1, \nu\}$ converges in $\mathbf{R}^n$ as $\nu \to \infty$. We can rewrite (13) as:

$$(\mu_{1,\nu} - \mu_{1,m})' \cdot \left[\int_0^1 a(u)a(u)'du\right] \cdot (\mu_{1,\nu} - \mu_{1,m}) \to 0\,.$$

We claim that for any positive definite matrix C, and any sequence $\{x_n\}$ in $\mathbf{R}^n$, $x_n C x_n \to 0$ if and only if $x_n \to 0$ in $\mathbf{R}^n$.[14] But the assumption that y is a full rank process in continuous time implies that $\int_0^1 aa'$ is positive definite, because this is the covariance matrix of the one-step-ahead innovation of y in continuous time. Therefore, (13) implies that $\mu_{1,\nu} \to \lambda_1$ as $\nu \to \infty$ for some $\lambda_1 \in \mathbf{R}^n$.

It is easy to show that, if $\mu_{k,\nu} \to \lambda_k$ for all $k = 1, \ldots, r$ for a finite integer r, then $\mu_{r+1,\nu} \to \lambda_{r+1}$ as $\nu \to \infty$. To see this, note that

$$\int_0^{r+1} ||f_\nu - f_m||^2 \to 0 \text{ as } \nu, m \to \infty\,;$$

by the inductive assumption, this implies that $\int_0^1 ||(\mu_{r+1,\nu} - \mu_{r+1,m}) \cdot a(u)||^2 \to 0$, and $\mu_{r+1,\nu} \to \lambda_{r+1} \in \mathbf{R}^n$ as $\nu \to \infty$.

Now, choose any integer K. Clearly

$$\int_0^K ||f_\nu(u) - \sum_{k=1}^{\infty} \lambda_k a(u-k)||^2 du \to 0 \text{ and } \int_0^K ||f_\nu - h_i||^2 \to 0 \text{ as } \nu \to \infty\,.$$

Since the limits are unique, $h_i(u) = \sum_{k=1}^{\infty} \lambda_k' a(u-k)$ for almost every $u < K$. Therefore, this equality holds for almost every $u \in \mathbf{R}$. ∎

If we apply this lemma to equation (7), we can write Y as

$$Y_i(t) = \varepsilon(t) + \int_1^{\infty} \sum_{k=1}^{\infty} \lambda_k' a(u-k)\zeta(t-du)$$

so that if we can pull the summation outside the integral sign, we can write:

$$Y_i(t) = \varepsilon_i(t) + \sum_{k=1}^{\infty} \lambda_k' Y(t-k) \tag{14}$$

and in this case Y has an autoregressive representation, with the coefficients of this representation being the λ's in the above lemma.

To "pull the summation sign outside the integral" in fact means that the following is true:

$$\int_1^{\infty} \sum_{k=1}^{\infty} \lambda_k' a(u-k)\zeta(t-du) = \lim_{m\to\infty} \sum_{k=1}^{m} \lambda_k' \int_1^{\infty} a(u-k)\zeta(t-du)$$

which, by definition of convergence in mean square, means:

$$\lim_{m\to\infty} \int_1^\infty \| \sum_{k=1}^{\infty} \lambda_k' a(u-k) - \sum_{k=1}^{m} \lambda_k' a(u-k)\|^2 du = 0, \text{ or}$$

$$\lim_{m\to\infty} \int_1^\infty \| \sum_{k=m}^{\infty} \lambda_k' a(u-k)\|^2 du = 0 .$$

The next lemma gives a sufficient condition for the interchange of summation and integration.

Lemma 2: If the sequence $\{\lambda_k\}$ is absolutely summable (element by element), then (14) holds.

Proof:

By applying twice the triangle inequality and pulling the constants λ out of the norm in L_n^2, we can justify the following steps:

$$\left[\int_1^\infty \| \sum_{k=m}^{\infty} \lambda_k' a(u-k)\|^2\right]^{-\frac{1}{2}} \le \sum_{j=1}^{n} \left[\int_1^\infty \| \sum_{k=m}^{\infty} \lambda_k^j a_j(u-k)\|^2\right]^{-\frac{1}{2}} \le$$

$$\sum_{j=1}^{n} \sum_{k=m}^{\infty} |\lambda_k^j| \left[\int_1^\infty \|a_j(u-k)\|^2\right]^{-\frac{1}{2}} \le K \sum_{j=1}^{n} \sum_{k=m}^{\infty} |\lambda_k^j| ,$$

where K is constant. If λ_k is absolutely summable the right hand side goes to zero, and by our earlier comment, (14) will hold. ∎

Notes

1. Rozanov (1967) contains the definitions of linear regularity, stationarity, random measures, integration with respect to random measures, and a proof of the existence of the MAR in continuous and discrete time. Throughout the paper, we will only use orthonormal random measures, so that for any pair Δ, δ' of disjoint Borel subsets of R, $E(\zeta_i(\Delta))^2 = |\Delta|$ for all i, $E(\zeta_j(\Delta)\cdot\zeta_i(\Delta)) = 0$ if $i \neq j$, and for any i, j, $E(\zeta_j(\Delta)\cdot\zeta_j(\Delta')) = 0$.

2. See Rozanov (1967) for a definition of spectral density and its properties. Equation (5) is known as the folding formula. In most of our discussion, it is quite clear how our results could be generalized to the case of a non-full rank process y, but allowing for this case would complicate our notation without adding much substance to the problem.

3. See Rozanov (1967), page 3.

4. Rozanov (1967) observes that the spaces $H_\varepsilon(\infty)$ (for ε an uncorrelated random measure) and L_n^2 are unitarily isometric, because:

$$\| \int f d\varepsilon - \int g d\varepsilon \| = \|f - g\|$$

where the first norm is for random variables, and the second represents the norm in L_n^2. This tells us that $H_\varepsilon(\infty)$ and L_n^2 "are arranged in the same way" (page 3, Rozanov (1967)); therefore, in view of equations (1) and (7), we would expect that there was a relationship between h and a in terms of the topology of L_n^2. This relationship is precisely what Proposition 1 uncovers.

5. See Rudin, theorems 12.3 and 12.4.

6. Here, we have used: $\varepsilon(t) = \int_0^\infty c\, d\zeta = Y(t) - \eta(Y - i(t) | H_Y(t-1)) = \int_0^\infty (a-h) d\zeta$. We also have used the fact that for any $f, g \in L_n^2$, if $\int f d\varepsilon = \int g d\varepsilon$, then $\text{var}[\int f d\varepsilon - \int g d\varepsilon] = 0$, so that $\int \| f - g \|^2 = 0$ and $f = g$ almost everywhere. Therefore, $c = a - h$.

7. Since $f(u) = \sum_{k=1}^{\infty} \sum_{j=1}^{n} \lambda_{k,j}\, a_j(u-k)$, if $u < 1$ then $u - k < 0$ for any of the k's in the summation; $a_j(u-k) = 0$ and $f(u) = 0$.

8. Sims (1972b) shows that Y_2 does not Granger cause Y_1 iff $\eta(Y_2(t) \mid H_{Y_1}(t)) = \eta(Y_2(t) \mid H_{Y_1}(\infty))$. But from Sims (1971), it is clear that, in general, this equality will not hold for the sampled process, even if it holds for the underlying continuous time process.

9. See Wheeden and Zygmund (1977) for the definition of indefinite integral and its properties.

10. See Wheeden and Zygmund, (1977) page 101.

11. To consider a correct example, take $a_{11}(u) = e^{-\lambda u}$, $a_{21}(u) = Ke^{-\lambda u}$ for all $u \geq 0$, and $a_{12} \equiv 0$. In continuous time, the second variable does not Granger cause the first one and, by Proposition 4, the same is true in the sampled process. However it can be shown that if we unit-average the first variable, the second variable will Granger cause the first one.