

Preliminary: Comments Welcome

Historical
Copy

Systematically Missing Data in Econometric Models:
Some Identification Considerations*

by Gary Skoog

Department of Economics
University of Minnesota

June 1975

u f s 6

*The Social Science Research Foundation and the Federal Reserve Bank of Minneapolis have given generous financial support for this project, which the author gratefully acknowledges.

Systematically Missing Data in Econometric Models--
Some Identification Considerations

I. Introduction

The variables which enter macroeconometric models are measured regularly, but at different time intervals. Thus, net national product data are available quarterly; money stock, biweekly; and interest rates, daily. This heterogeneity poses both theoretical and practical problems for the model builder. Despite attempts to construct "fine" models (in which the variables are, for example, measured monthly) it is fair to say that the method of resolution to date has been to construct "coarse" models (measuring all variables, say, quarterly), ignoring, or using in a nonsystematic or aggregated manner, the intraquarter observations.

The approach taken in this paper assumes the variables (which in practice will be residuals from fitted trends) are realizations of a continuous, second-order stationary stochastic process. Identification^{1/} is achieved by projection of the dependent process onto the independent process. Thus, if we had the continuous autocovariance and cross-covariance functions, or, equivalently, all of the own and cross-spectral densities, we would know the continuous model's parameters (lag distributions). It is with respect to this (idealized) data set that the continuous model is identified; with respect a set of data consisting only of regular point samplings, the continuous process is no longer identified. Nevertheless, by projecting the sampled dependent process onto the sampled independent process, a discrete model is identified with respect to the sampled data set. Relationships between the parameters of the discrete and continuous models were developed by Sims in [10], for scalar processes. Emphasis was placed on conditions under which the

parameters for the continuous and discrete models would be "close" to one another. In [3], Geweke extended Sims' results to the vector case. All data was gathered at the same time intervals in both of these papers.

We can now imagine that the discrete model just mentioned is a "fine" model in which, for example, all variables are observed, say, daily. With such frequent data this model is identified by projection and is thus analogous to the continuous model above. Further, with nonhomogeneously sampled time series or with uniformly less frequently sampled data, the "fine" model ceases to remain identified, but gives rise in turn to "coarser" models, which are identified with respect to the rate at which observations are being recorded. Thus we have several models, all interrelated, and possessing varying possibilities for identification.

The role of the continuous model may be likened to the role of preferences in the theory of consumer choice: even though they may not be our principal concern for inference, we are well advised to postulate them as our primitive notion. We will on occasion analyze the implications for the discrete models which flow from hypothesized restrictions of the underlying continuous model; and, reflecting our belief that economic processes are ideally best modeled as continuous phenomena, we should like to check the compatibility of any assumptions placed on a discrete model with a continuous model.

II. Notation, Framework, and the Continuous Model

1×1 $N \times 1$
Let $y(t)$, $x(t)$ be jointly covariance stationary, continuous, real, stochastic processes, where y will be interpreted as the dependent and x as an independent, linearly regular^{2/} process. We assume that $F_{x_i x_i}(d\lambda)$ $i=1, \dots, N$, and $F_{yy}(d\lambda)$ are absolutely continuous with respect

to Lebesgue measure, so in the $(N+1) \times (N+1)$ spectral distribution matrix of the $\begin{pmatrix} y \\ x \end{pmatrix}$ process, all elements have densities. Thus

$$R_{xy}^{Nx1}(t) \equiv \overline{Ex(s)y(s+t)'} = \int_{-\infty}^{\infty} e^{-i\lambda t} \overline{dF_{xy}(\lambda)} = \int_{-\infty}^{\infty} e^{i\lambda t} S_{xy}(\lambda) d\lambda$$

where

$$x(t) = \int_{-\infty}^{\infty} e^{i\omega t} dz_x(\omega), \quad y(t) = \int_{-\infty}^{\infty} e^{i\omega t} dz_y(\omega), \quad \frac{F_{xy}(d\lambda)}{d\lambda} = S_{xy}(\lambda),$$

and

$$Edz_x(\lambda) \overline{dz_y(\omega)} = \begin{cases} \overline{dF_{xy}(\lambda)}, & \lambda = \omega \\ 0 & \lambda \neq \omega. \end{cases}$$

Similar definitions and results hold for $R_{xx}(t)$ and $F_{xx}(d\lambda)$, in notation (e.g., Fishman [2]) which is close enough to standard^{3/} to be understood. All random variables are assumed to have mean 0, and we follow the usual econometric practice of limiting ourselves to second-order moments for identifying information.

We now consider the specification of model (A),

$$(1) \quad y(t) = x' * b(t) + u(t) = \left(\sum_{j=1}^N b_j * x_j \right) (t) + u(t) = \sum_{j=1}^N \int_{-\infty}^{\infty} b_j(s) x_j(t-s) ds + u(t)$$

in which, for $j=1, \dots, N$,

$$(2) \quad Ex_j(t)u(t+s) = 0,$$

all t, s , real. Substituting for $u(t+s)$ in the last expression, we have

$$Ex(t)y(t+s)' = E\left\{x(t) \left(\sum_{j=1}^N b_j * x_j \right)'(t+s)\right\},$$

or, again, $R_{xy}(s) = R_{xx} * \overline{b}(s)$ where $b' = (\overline{b_1} \dots \overline{b_N})$. But because x and y are real, $R_{xy}(s) = \overline{R_{xy}(s)}$ and $R_{xx}(s) = \overline{R_{xx}(s)}$, so we may conjugate, to obtain

$$(3) \quad R_{xy}(s) = R_{xx} * b(s).$$

We have arrived at (3), which is equivalent to (2), given (1), that is, given that such a $b(\cdot)$ exists.^{4/} But this heretofore tacit assumption must now be investigated, as well as the related question of convergence of the stochastic integrals in $x' * b$. Perhaps (1) and (2) are consistent--but with the implied $b(\cdot)$ a generalized function^{5/,6/} (see Proposition 4). More to the point, no such b may exist. It is $(y(t)|H_x)$, the projection^{7/} of $y(t)$ onto H_x (the space of values of the x process; see footnote 2) which always exists, and (1) would have been more accurately written

$$(1)' \quad y(t) = (y(t)|H_x) + u(t).$$

However, to restore (1) and to return to more familiar terrain, we have

Proposition 1 If $|S_{xx}(w)| \neq 0$ a.e., and $S_{xx}^{-1}(w)S_{xy}(w)$ is the (classical) Fourier transform of $b(\cdot)$ with absolutely integrable components, then (1)' becomes (1): specifically, $(y(t)|H_x)$ has the (kernel) representation

$$\int_{-\infty}^{\infty} b'(t-s)x(s)ds = x' * b(t).$$

Proof: The characterizing feature of orthogonal projection is that $\langle (y(t) - (y(t)|H_x)), x(s) \rangle = 0$, all t and s , where the notation $\langle y, z \rangle$ means Eyz , or Eyz if the random vectors y and z are real. Uniqueness follows essentially from the requirement that $(y(t)|H_x)$ be in H_x , the Hilbert space formed by completing, under the norm induced by $\langle \cdot, \cdot \rangle$ the set of all finite linear combinations of $x(t)$, t real. Now by the development in Rozanov, p. 28-35, there corresponds to $(y(t)|H_x)$ a S_x -unique spectral characteristic

$$\varphi(w), \varphi \in L^2[S_{xx}(w)dw],$$

the latter term meaning

$$\int_{-\infty}^{\infty} \varphi(w) S_x(w) \varphi'(w) dw < \infty,$$

such that

$$(y(t) | H_x) = \int_{-\infty}^{\infty} e^{it\lambda} \varphi(\lambda) z_x(d\lambda).$$

For future reference, let's call the general correspondence θ . Thus

$$\varphi \xrightarrow{\theta} x_\varphi, \text{ where } \varphi \in L^2[S_{xx}(w)dw] \text{ and } x_\varphi = \int_{-\infty}^{\infty} e^{it\lambda} \varphi(\lambda) z_x(d\lambda) \in H_x.$$

The integrability condition on φ is precisely that needed to ensure convergence of the stochastic integral. Rewriting the characterizing inner product which must be zero, substituting for $(y(t) | H_x)$, and conjugating yields

$$\int_{-\infty}^{\infty} e^{i\lambda(s-t)} E[\overline{z_y(d\lambda)} z_x(d\lambda) - z_x(d\lambda) z_x'(d\lambda) \varphi'(\lambda)] = 0$$

after exchanging the order of integration and using the orthogonality of the increments processes. We arrive at $\overline{S_{xy}} = \overline{S_{xx}} \varphi'$, or $S_{xy} = S_{xx} \overline{\varphi}'$, where $-'$ = ordinary transpose. But by hypothesis $\overline{\varphi}' \equiv S_{xx}^{-1} S_{xy} = \tilde{b}$, for " $b \in L^1$," i.e., some b all of whose components are in $L^1(-\infty, \infty)$. We may thus write

$$\begin{aligned} (y(t) | H_x) &= \int_{-\infty}^{\infty} e^{it\lambda} \overline{b}(\lambda) z_x(d\lambda) = \int_{-\infty}^{\infty} e^{it\lambda} \int_{-\infty}^{\infty} \overline{b}'(t-s) e^{-i\lambda(t-s)} ds z_x(d\lambda) \\ &= \int_{-\infty}^{\infty} \overline{b}'(t-s) \int_{-\infty}^{\infty} e^{i\lambda s} z_x(d\lambda) ds = \int_{-\infty}^{\infty} \overline{b}'(t-s) x(s) = b' * x(t) = \end{aligned}$$

$$x' * b(t)$$

The change of order of stochastic integration is possible by the matrix generalization of Rozanov's Fubini-like Theorem 2.4; $b \in \mathbb{L}_1$ is exactly what is needed to allow this application. Q.E.D.

We shall not pursue the question of describing the class $\tilde{\mathcal{B}}$ of $\tilde{b} \in L^2[S_{xx}(w)dw]$ which are Fourier transforms of $\mathbb{L}_1(-\infty, \infty)$ functions b except to say the elements of $\tilde{\mathcal{B}}$ must be uniformly continuous and bounded; and obviously $\tilde{\mathcal{B}}$ is nonempty, since it contains $e^{-ibw} e^{-iaw} = \tilde{\chi}_{[a, b]}$ (in the scalar case), where $\chi_{[a, b]}(t) = 1, t \in [a, b]$, and vanishes otherwise.

The continuous model may now be regarded as identified, at least where S_{xx} is nondegenerate, by $\tilde{b} = S_{xx}^{-1} S_{xy}$, where although we have rigorously worked out only a classical case, we shall not hesitate to interpret the symbols in the more generalized sense of footnote 6. To be clear about this point and to illustrate the force of Proposition 1 as well as the use of several definitions, consider the continuous time model

$$y(t) = x(t-1) + u(t), \quad E u(t)x(s) = 0 \text{ all } s \text{ and } t.$$

$R_{xy}(s) = R_{xx}(s-1)$ and $S_{xy}(w) = e^{-iw} S_{xx}(w)$ follow directly. Thus $\tilde{b} = S_{xx}^{-1} S_{xy} = e^{-iw}$. While \tilde{b} is bounded and uniformly continuous, it is not the Fourier transform of any $L^1(-\infty, \infty)$ function. It is, however, the Fourier transform of $\delta_1(\cdot)$, the Dirac delta "function" with unit mass at 1: $\delta_1(t) = \delta(t-1)$, where δ is the usual delta function. Thus $e^{-iw} = \int_{-\infty}^{\infty} e^{-iwt} \delta_1(t) dt$. Taking the point of view that R_{xx} and R_{xy} are the data, we form \tilde{b} and check whether or not we are lucky enough to be in the "smooth" case of Proposition 1, in which the effects of $x(t-s)$ on $y(t)$ are continuously spread over time. In this case, they are not-- they are concentrated at a lag of one unit. Nevertheless the systematic

effect, $(y(t)|H_x)$, is the (well-defined!) ordinary random variable $x(t-1)$ which corresponds under θ to the spectral characteristic $e^{-i\omega} = \varphi(\omega)$. It is only when we try to express $x(t-1)$ as a symbolic convolution, $x*b(t)$, that we run into the need of the generalized function $b(s) = \delta(s-1)$. In short, generalized functions (and especially their direct and inverse Fourier transforms) can be used as an aid in discovering images of the bijection θ , as we have done.

The sad fact, however, is that the continuous data, (S_{xx}, S_{xy}) or (R_{xx}, R_{xy}) , with respect to which the continuous model is identified, are seldom at hand. Consequently, in general, many $b(\cdot)$ vectors will be observationally equivalent with respect to (even the doubly infinite idealizations of) the discrete observational patterns that we are likely to possess. Yet, as we will see, vestiges of identification may remain.

III. The Discrete Models, And Their Identifiability By Means of Projection

We consider three models which are all discrete samplings of the variables in model (A), $y(t) = x'*b(t)+u(t)$. Without loss of generality we may regard $Y(t) = y(t)$, t integer, as observed once per period, less often than any of the independent processes in models (B), (C), and (D) below.^{8/} In model (B), all of the variables X_1, \dots, X_N are observed n_1 times per period, where $n_1 \geq 1$. We call (B) the "fine" model, and write

$$(4) \quad Y(t) = \sum_{j=1}^N \sum_{s=-\infty}^{s=\infty} B_{j;f}(t-\frac{s}{n_1}) X_{j;f}(\frac{s}{n_1}) + U_f(t)$$

$$(5) \quad EX_{j;f}(\frac{s}{n_1}) U_f(t) = 0 \text{ all } s, t \text{ integers; } j=1, \dots, N.$$

The upper case variables will always denote point samplings of the lower case variables, so that

$$X_j\left(\frac{t}{n_1}\right) = x_j\left(\frac{t}{n_1}\right), \quad t \text{ integer}, \quad j=1, \dots, N.$$

In model (C), we assume that the independent variables are numbered in order of their observational frequency, so that x_i (or X_i), is observed n_i times per period, $n_1 \geq n_2 \geq \dots \geq n_N$. We also assume that $[n_i/n_{i+1}] = n_i/n_{i+1}$, where $[]$ is the greatest integer function. This assumption is likely to be met approximately in practice: for example, if n_1 is 12 (monthly data) and n_2 is 4 (quarterly data) then $[n_1/n_2] = n_1/n_2 = 3$. We call (C) the "systematically missing data" model--(discrete) data are "missing" with respect to (B)--and we write

$$(6) \quad Y(t) = \sum_{j=1}^N \sum_{s=-\infty}^{\infty} B_j \left(t - \frac{s}{n_j}\right) X_j\left(\frac{s}{n_j}\right) + U(t)$$

$$(7) \quad EU(t)X_j\left(\frac{s}{n_j}\right) = 0 \quad t, s, \text{ integer}, \quad j=1, \dots, N.$$

Finally, the "coarse" model (D) uses only the independent variables corresponding to the times at which the least frequently observed time series, X_N is measured. It is defined by

$$(8) \quad Y(t) = \sum_{j=1}^N \sum_{s=-\infty}^{\infty} B_j; c \left(t - \frac{s}{n_N}\right) X_j\left(\frac{s}{n_N}\right) + U_c(t)$$

$$(9) \quad EU_c(t)X_j\left(\frac{s}{n_N}\right) = 0 \quad t, s \text{ integer}; \quad j=1, \dots, N.$$

Before beginning a systematic analysis, several remarks are in order.

(i) Logically (C) may be regarded as the most general of the discrete models, with (B) and (D) constituting the special cases $n_i = n_1$ and $n_i = n_N$, $i=1, \dots, N$, respectively.

(ii) Models (B), (C), and (D) all tacitly define the distributed lag coefficients by the projection of sampled y onto the relevant spaces of values generated by the sampled $x = (x_1 \dots x_N)'$ process. Consequently, we should be prepared to put some effort into finding conditions under which the projections have the convolution representations indicated by (4), (6), and (8), respectively.

(iii) We could have rewritten (6) to appear closer to (4) as follows:

$$(6)' \quad Y(t) = \sum_{j=1}^N \sum_{s=-\infty}^{\infty} B_j(t - \frac{s}{n_1}) X_j(\frac{s}{n_j}) + U(t),$$

$$B_j(t - \frac{s}{n_1}) = 0, \quad \frac{s}{n_1} \neq \frac{r}{n_j}, \quad s, r \text{ integer; } j=1, \dots, N.$$

Here, (6) or (6)' appears as (4) subject to the constraint that when an independent variable is not available its coefficient is zero. Similarly, (8) could be written (8)'--(6) under the obvious constraints--or again as (8)''--(4) under even more constraints.

(iv) If our goal were to "predict" $Y(t)$ given the entire X process past, present, and future, sampled at the intervals indicated in the model (B), (C), and (D), the coefficients $\{B_{j;f}\}$, $\{B_j\}$, and $\{B_{j;c}\}$ would give the predictions which minimize mean-square error of predicted $Y(t)$, $\sigma_{Y(t)}^2(B) \leq \sigma_{Y(t)}^2(C) \leq \sigma_{Y(t)}^2(D)$. This follows because we are projecting onto successively smaller subspaces and consequently being left with successively longer normed residuals.

(v) To interpret (D) as an equation from the reduced form of a "current practice" econometric model would require that the X process be exogenous with respect to the Y process, or that there be no feedback from Y to X in the Granger-Sims sense.^{9/} Remark (iv) would then suggest

that "better practice" from the point of view of prediction, abstracting from sampling fluctuations, would involve use of (C). Thus, assuming that the data are heterogeneous, (C) uses all the data and optimizes; (D) ignores the more frequent observations, to its detriment; and (B) is not feasible.

The remainder of the paper will be devoted to analysis of the identifiability of and interrelationships among these models. We begin with more notation, adopting that employed in [3] and [10] wherever possible, and remaining consistent with that introduced already. Thus an upper case variable refers to a lower case variable, sampled. The same property holds between the covariance functions, since $R_{XY}(\frac{T}{n}) = EX(s-\frac{T}{n})Y(s) = Ex(s-\frac{T}{n})y(\frac{S}{n}) = R_{xy}(\frac{T}{n})$. Analogously, $R_X(\frac{T}{n}) = R_x(\frac{T}{n})$. Since the next few propositions concern the relation between models (A) and (B), to minimize subscripts we have set $n_1=n$ and have dropped the "f" on $B_{j;f}$. The standard manipulation yields the discrete spectral density matrix $S_X(w)$, defined on $[-n\pi, n\pi]$, in terms of

$$S_X(w): R_X(\frac{T}{n}) = \int_{-\infty}^{\infty} S_X(w) e^{i w \frac{T}{n}} dw = \int_{-n\pi}^{n\pi} \sum_{k=-\infty}^{k=\infty} S_X(w+2\pi nk) e^{i(w+2\pi nk)\frac{T}{n}} dw =$$

$$\int_{-n\pi}^{n\pi} F_n[S_X(\cdot)](w) e^{i w \frac{T}{n}} dw = \int_{-n\pi}^{n\pi} S_X(w) e^{i w \frac{T}{n}} dw,$$

the middle equality defining $F_n[]$.

By the assumption of finite variances of the X process,

$$R_X(0) = \int_{-\infty}^{\infty} S_X(w) dw = \int_{-n\pi}^{n\pi} S_X(w) dw,$$

so we have integrability of each term of this positive semidefinite Hermitian matrix. We already assumed $\det S_X(w)$ to be nonzero a.e. on $(-\infty, \infty)$ and it is natural to make the same assumption on $S_X(w)$ on $[-n\pi, n\pi]$.

Equivalently we order the eigenvalues of $S_X(w)$ as $\lambda_1(w) \geq \lambda_2(w) \geq \dots \lambda_N(w)$ and assume $\lambda_N(w) > 0$ for almost all such w . Perhaps obvious is the following.

Lemma 1 Finite variance of the X process-- $\text{Var } X_i < \infty, i=1, \dots, N$ --is equivalent to integrability of $\lambda_i(w), i=1, \dots, N$.

Proof: Since

$$\text{tr } S_X(w) = \sum_{i=1}^N S_{X_i}(w) = \sum_{i=1}^N \lambda_i(w),$$

$$\sum_{i=1}^N \text{Var } X_i = \sum_{i=1}^N \int_{-n\pi}^{n\pi} S_{X_i}(w) dw = \sum_{i=1}^N \int_{-n\pi}^{n\pi} \lambda_i(w) dw.$$

Finiteness of the left-hand side implies that each of $\int_{-n\pi}^{n\pi} \lambda_i(w) dw < \infty$.

The converse would also have content had we defined spectra for processes with infinite variance. Q.E.D.

As indicated in remark (ii), a development analogous to that employed in arriving at (3) would yield,

$$(10) \quad R_{XY}\left(\frac{t}{n}\right) = R_X * B\left(\frac{t}{n}\right)$$

where B is real, $B'\left(\frac{t}{n}\right) = (B_1\left(\frac{t}{n}\right), \dots, B_N\left(\frac{t}{n}\right))$, and now of course the convolution is in discrete time. The same objections apply: the existence of the representation $B' * X(t)$ for $(Y(t) | H_X)$ has been tacitly assumed. But as before we can make this assumption good by establishing

Proposition 2 Assume $\lambda_N(w) > 0$ a.e. on $[-n\pi, n\pi]$ and that $S_X^{-1}(w)S_{XY}(w)$ is the Fourier transform of $B\left(\frac{t}{n}\right)$ where

$$\sum_{j=1}^N \sum_{t=-\infty}^{t=\infty} |B_j\left(\frac{t}{n}\right)| < \infty.$$

Then

$$(Y(t) | H_X) = \sum_{j=1}^N \sum_{t=-\infty}^{t=\infty} B_j(t - \frac{s}{n}) X_j(\frac{s}{n}) \equiv B' * X(t),$$

so that model (B) is identified by projection.

Proof: Letting $\varphi(w)$ be the S_X -unique spectral characteristic of $(Y(t) | H_X)$, we have the latter equal to

$$\int_{-n\pi}^{n\pi} e^{itw} \varphi(w) z_X(dw).$$

Proceeding as in the proof of Proposition 1 to write out the meaning of orthogonality again yields $S_{XY}(w) = S_X(w) \bar{\varphi}'(w)$, or $\bar{\varphi}' \equiv S_X^{-1} S_{XY}$ a.e., using the invertibility of S_X ensured by $\lambda_N > 0$ a.e. Now by hypothesis

$$\bar{\varphi}'(w) = \sum_{t=-\infty}^{t=\infty} B(\frac{t}{n}) e^{-iw\frac{t}{n}},$$

where the absolute summability of $B(\cdot)$ guarantees absolute and uniform convergence (as well as the uniform continuity of $S_X^{-1}(w) S_{XY}(w)$). Let $\epsilon > 0$ be given, and choose M so large that, for all $w \in [-n\pi, n\pi]$ and all $j=1, \dots, N$,

$$|\bar{\varphi}'_j(w) - \sum_{t=-M}^{t=M} B_j(\frac{t}{n}) e^{-iw\frac{t}{n}}| \equiv |\delta_j(w)| < \sqrt{\frac{\epsilon}{N ||\lambda_1|| 2\pi n}}$$

where $||\lambda_1|| = \int_{-n\pi}^{n\pi} \lambda_1(w) dw$,

which is finite by the lemma. Now by the usual isometry between the time and frequency domains,

$$|| (Y(t) | H_X) - \sum_{j=1}^N \sum_{t=-M}^{t=M} B_j(t - \frac{s}{n}) X_j(\frac{s}{n}) ||^2 = \int_{-n\pi}^{n\pi} \delta'(w) S_X(w) \delta(w) dw =$$

$$\int_{-n\pi}^{n\pi} (\frac{\delta(w)}{||\delta(w)|| })' S_X(w) (\frac{\delta(w)}{||\delta(w)|| }) \cdot ||\delta(w)||^2 dw \leq \int_{-n\pi}^{n\pi} \lambda_1(w) \frac{N \cdot \epsilon}{N ||\lambda_1|| 2\pi n} dw = \epsilon,$$

where $||\delta(w)||^2 = \sum_{i=1}^N |\delta_i(w)|^2$

and the familiar inequality involving the Rayleigh quotient has been used.
Q.E.D.

We observe that the construction guarantees that $\bar{\phi}' = \tilde{B}$ be in $L^2[S_X(w)dw]$, which is precisely the (so-called "matching") condition^{10/} that the convolution sum be convergent in mean square. Indeed, assume that we find R_X^{-1} and "solve" (10) for $B(\frac{t}{n}) = R_X^{-1} * R_{XY}(\frac{t}{n})$. Putting aside the obvious convergence question involving the right-hand side, when we ask about $X' * R_X^{-1} * R_{XY}$, we are led back to the matching condition, since the variance of this random variable will be, if finite,

$$\int_{-\pi\pi}^{\pi\pi} \tilde{R}_{XY} \tilde{R}_X^{-1} \cdot \tilde{S}_X \cdot \tilde{R}_X \cdot \tilde{R}_{XY} dw.$$

But now, the finiteness of this integral must be checked, or assumed. We will pursue this approach, to derive different conditions under which the desired representation $B' * X(t)$ holds.

The first step is to construct R_X^{-1} . Its defining property is

$$\sum_{s=-\infty}^{s=\infty} R_X(\frac{s}{n}) R_X^{-1}(\frac{t-s}{n}) = \delta_{\frac{t}{n}} \cdot I_N,$$

where $\delta_{\frac{t}{n}}$ is the Kronecker delta. This convolution requirement translates into $S_X(\omega) S_X^{-1}(\omega) = I_N$, $\omega \in [-\pi, \pi]$ if R_X and R_X^{-1} can both be Fourier transformed. Indeed, Wiener and Masani have provided the theory which extends to the Hilbert space \mathbb{L}_2 (of $N \times N$ matrices whose components are $L^2[-\pi, \pi]$, complex-valued functions) the results we need, particularly the Riesz-Fischer Theorem, Parseval's Identity, and the Convolution Rule. ([8], especially Theorem 3.9). Thus for $R_X(\frac{s}{n})$ to be in \mathbb{L}_2 , the corresponding (sequence) Hilbert space of square-summable matrices, we must have by definition that

$$(11) \quad \sum_{s=-\infty}^{s=\infty} \|R_X(\frac{s}{n})\|_E^2 \equiv \sum_{s=-\infty}^{s=\infty} \text{tr } R_X(\frac{s}{n})R_X'(\frac{s}{n}) = \sum_{s=-\infty}^{s=\infty} \left(\sum_{i=1}^N \sum_{j=1}^N |r_{ij}(\frac{s}{n})|^2 \right) < \infty.$$

In this case,

$$(12) \quad S_X(w) = \sum_{s=-\infty}^{s=\infty} R_X(\frac{s}{n})e^{iws/n} \text{ is in } \underline{L}_2, \text{ i.e.,}$$

$$\infty > \frac{1}{2\pi n} \int_{-n\pi}^{n\pi} \text{tr } \{S_X(w)S_X'(w)\}dw \equiv ((S_X(w), S_X(w))) = \|S_X(w)\|^2.$$

Consequently, we will be able to Fourier transform R_X to S_X , invert, and inverse Fourier transform provided that (12) is met by S_X and S_X^{-1} . 11/

Now S_X and S_X^{-1} are in \underline{L}_2 if and only if, respectively,

$$\sum_{i=1}^N \lambda_i^2 \text{ and } \sum_{i=1}^N \frac{1}{\lambda_i^2}$$

are integrable, that is, if and only if

$$\int_{-n\pi}^{n\pi} \lambda_1^2(w)dw < \infty \text{ and } \int_{-n\pi}^{n\pi} \frac{1}{\lambda_N^2(w)}dw < \infty.$$

We may summarize this discussion as

Lemma 2 Provided the largest and the reciprocal of the smallest eigenvalues of $S_X(w)$ are both in $L^2[-n\pi, n\pi]$, R_X^{-1} exists, uniquely, in \underline{L}_2 .

As noted, the matching condition must be dealt with explicitly.

Here, it is evidently

$$\int_{-n\pi}^{n\pi} \tilde{R}_{XY}' \tilde{R}_X^{-1} \tilde{R}_{XY} dw < \infty.$$

Now if \tilde{R}_X^{-1} is essentially bounded above by which we mean there exists $c > 0$ such that $cI_N - \tilde{R}_X^{-1}(w)$ is a.e. positive definite, then we would have the matching condition satisfied whenever $R_{XY}(\frac{s}{n})$ is a square summable vector sequence, i.e.,

$$\sum_{s=-\infty}^{s=\infty} R_{XY}' \left(\frac{s}{n} \right) R_{XY} \left(\frac{s}{n} \right) = \sum_{s=-\infty}^{s=\infty} \left\| R_{XY} \left(\frac{s}{n} \right) \right\|_E^2 =$$

$$\sum_{s=-\infty}^{s=\infty} \sum_{i=1}^N \left| R_{XY,i} \left(\frac{s}{n} \right) \right|^2 = \int_{-n\pi}^{n\pi} \tilde{R}_{XY}'(w) \tilde{R}_{XY}(w) dw < \infty,$$

where as before $\| \cdot \|_E$ stands for Euclidean norm, and we have tacitly filled in a gap by defining the concepts for $N \times 1$ vectors that had previously been defined for scalars and matrices. The assumption that $R_{XY}(\cdot)$ be square summable is a very natural one,^{12/} especially in conjunction with $R_X(\cdot) \in \underline{\ell}_2$. Finally, we observe that if \tilde{R}_X^{-1} is essentially bounded above, $\text{tr } S_X^{-1} S_X^{-1} < \text{tr } c^2 \cdot I_N = c^2 N$,^{13/} which is integrable over $[-n\pi, n\pi]$, so that the condition that $\frac{1}{\lambda_N(w)}$ be square integrable is satisfied. Likewise, if \tilde{R}_X were bounded above, $\lambda_1(w)$ would be square integrable. But $\tilde{R}_X < d \cdot I_N$ is equivalent to $\tilde{R}_X^{-1} > d^{-1} \cdot I_N$, i.e., \tilde{R}_X^{-1} being essentially bounded below. Taken together, the discussion above entails

Proposition 3 If $R_{XY} \left(\frac{t}{n} \right) \in \underline{\ell}_2$ and if also: (a) S_X is essentially bounded above and below (so that S_X^{-1} is as well) or, if (b) $\int_{-n\pi}^{n\pi} \lambda_1^2(w) dw < \infty$ and S_X is essentially bounded below (equivalently, that S_X^{-1} is essentially bounded above); then R_X has a unique $\underline{\ell}_2$ inverse such that $(R_X^{-1} * R_{XY})' * X(t) = (Y(t) | H_X)$, i.e., the matching condition is satisfied. Hence, the projection has a convolution representation, and model (B) is again identified.

At this point, several remarks are in order.

(i) The question raised earlier about convergence of $R_X^{-1} * R_{XY} \left(\frac{t}{n} \right)$ can be answered affirmatively, under the hypotheses of Proposition 3, since

$$\sum_{s=-\infty}^{s=\infty} \sum_{j=1}^N |R_X^{-1}(\frac{s}{n}; i, j) R_{XY}(\frac{t-s}{n}; j)| \leq \sum_{j=1}^N \{ \sum_{s=-\infty}^{s=\infty} |R_X^{-1}(\frac{s}{n}; i, j)|^2 \}^{1/2} x$$

$$\{ \sum_{s=-\infty}^{s=\infty} |R_{XY}(\frac{t-s}{n}; j)|^2 \}^{1/2} < \infty$$

by the Schwarz inequality.

(ii) The approach of Proposition 3 squares with our intuition in requiring a full rank condition on a structural characteristic of the x process to be able to uniquely assign coordinates to the coordinate-free concept of vector.

(iii) The two preceding propositions apply to establishing the identifiability of Model (D), by taking $n=1$.

(iv) While no direct reference to the spectral characteristic of $(Y(t) | H_X)$ was made, it clearly is $S_X^{-1} \cdot S_{XY}$. But now, as the sum of products of $L^2[-n\pi, n\pi]$ functions, this spectral characteristic is no longer subject to the harsh continuity requirements implied by the hypothesis of Proposition 2.

In summary, identification of models (A), (B), and (D) has been achieved in two ways: rigorously, Propositions 1-3 indicate precise conditions for the commonly written distributed lag relations to be valid; in the case of the continuous model, the more general methods of Fourier transformation of generalized functions were indicated to be capable of providing the desired identification. It is not evident (to the author!) how this latter technique would aid in identification of the discrete model.^{14/} Model (C) will be treated in Section V. We now address the relation between the continuous b and discrete B .

IV. On the Identification of the Continuous Model from Uniform Discrete Data

We begin by examining the relation of Model (A) to the model (B) (or (D) if $n=1$). It is useful to observe that S_{XY} is $S_{xy}(w)$ n -folded, ^{15/} just as S_X was S_x acted upon by F_n :

$$R_{XY}\left(\frac{\tau}{n}\right) = \int_{-\infty}^{\infty} S_{xy}(w) e^{i w \frac{\tau}{n}} dw = \int_{-n\pi}^{n\pi} F_n[S_{xy}(\cdot)](w) e^{i w \frac{\tau}{n}} dw =$$

$$\int_{-n\pi}^{n\pi} S_{XY}(w) e^{i w \frac{\tau}{n}} dw = R_{XY}\left(\frac{\tau}{n}\right)$$

where, as before,

$$F_n[S_{xy}(\cdot)](w) = \sum_{k=-\infty}^{k=\infty} S_{xy}(w+2\pi nk).$$

The force of the equality string is the third equality, which proves that

$$(11a) \quad F_n[S_{xy}(\cdot)](w) = S_{XY}(w) \text{ a.e.}$$

Taking $y=x$, or recalling the earlier result, yields

$$(11b) \quad F_n[S_x(\cdot)](w) = S_X(w) \text{ a.e.}$$

The left-hand side of both of these relations is $2\pi n$ -periodic, so that the right-hand side is also, despite the fact that when X and Y are regarded as discrete processes in their own right, S_X , S_{XY} , and S_{YY} are usually considered as defined only on $[-n\pi, n\pi]$. Denoting (perhaps overly suggestively!) the spectral characteristics of $(y(t)|H_X)$ and $(Y(t)|H_X)$ by \tilde{b} and \tilde{B} , no further assumptions were required to write

$$(12a) \quad S_{XY}(w) = S_X(w)\tilde{B}(w), \quad |w| \leq n\pi$$

and

$$(12b) \quad S_{xy}(w) = S_x(w)\tilde{b}(w).$$

Applying F_n to (12b) and equating with (12a) results in

$$(13a) \quad S_{XY}(w) = F_n[S_x(\cdot)\tilde{b}(\cdot)]$$

$$(13b) \quad S_X(w)B(w) = F_n[S_x(\cdot)\tilde{b}(\cdot)].$$

Assuming S_X^{-1} exists a.e. yields

$$(14) \quad \tilde{B}(w) = S_X^{-1}(w)F_n[S_x(\cdot)\tilde{b}(\cdot)] = F_n[S_X^{-1}(\cdot)S_x(\cdot)\tilde{b}(\cdot)](w) \text{ a.e.}$$

since a $2n\pi$ -periodic function can be passed through the F_n operator.

Equations (13) and (14) express the general underidentifiability of the continuous model from discrete data. We may observe S_X , S_{XY} and S_Y (which is irrelevant here), and, from (12a), compute $\tilde{B}(w)$. But since neither S_x nor S_{xy} is available (only their folded versions (11a) and (11b) are observable), we cannot compute $S_x^{-1}(w)S_{xy}(w) = \tilde{b}(w)$ from (12b). We interpret the right-hand sides of (13) and (14) as describing, without making any further assumptions about the $\binom{y}{x}$ process, precisely the set of $\tilde{b}(\cdot)$ which are consistent with the data-determined left-hand sides. The nonidentifiability of \tilde{b} results from the joint (and observationally inseparable) operations upon it of multiplication by an unknown $S_x(w)$ matrix, followed by a folding of the product.

Consider first the conceptual unraveling of (11b), or, equivalently the nature of the underidentification of S_x . For any real process, S_x must: (1) be integrable, (2) satisfy $S_x'(w) = S_x(w) = \overline{S_x(-w)}$, and (3) be positive semidefinite. The latter requirement is sometimes strengthened to (3') positive definite a.e. It is immediate from (11b) that the same properties will hold for S_x . Now define $N_{\lambda;n} = \bigcup_{k=-\infty}^{\infty} \{\lambda + 2\pi nk\}$. Choose any

"allowable" $S_X(\cdot)$ and pick any $\lambda \in N_{0; \frac{n}{2}}$, i.e., $\lambda \neq 0 \pmod{\pi n}$. Any "decomposition" of $S_X(\lambda)$, i.e., any sequence of positive definite Hermitian matrices $\{S_{X;k}(\lambda)\}$ such that $S_X(\lambda) = \sum_{k=-\infty}^{\infty} S_{X;k}(\lambda)$, can be spread over $N_{\lambda;n}$ to help form an observationally equivalent spectral density matrix function. To maintain the last equality in (2) we must form $N_{-\lambda;n}$ and a related sequence $S_{X;k}(-\lambda) = \overline{S_{X;k}(\lambda)}$. A perturbation matrix function (which is zero a.e.) is defined by:

$$(15) \quad P(w) = \begin{cases} 0 & w \in (N_{\lambda;n} \cup N_{-\lambda;n})^c \\ S_{X,k}(\lambda) & w = \lambda + 2\pi nk, k \neq 0 \\ - \sum_{k \neq 0} S_{X,k}(\lambda) & w = \lambda \\ S_{X,k}(-\lambda) & w = -\lambda - 2\pi nk, k \neq 0 \\ - \sum_{k \neq 0} S_{X,k}(-\lambda) & w = -\lambda \end{cases}$$

Consequently $\overline{P}(-w) = P(w) = P'(w)$

Finally, $S_{\underline{X}}(w) \equiv S_X(w) + P(w)$. Note that $F_n[P](w) = 0$. By construction $S_{\underline{X}}(w)$ is still an allowable spectral density matrix function, and $F_n[S_{\underline{X}}](w) = F_n[S_X](w) = S_X(w)$, $|w| \leq n\pi$, so that both $S_{\underline{X}}$ and S_X give rise to the same observed S_X . Admittedly, $S_{\underline{X}}$ and S_X don't constitute a very interesting observationally equivalent pair^{16/}; but this operation may be performed simultaneously for all λ , or all λ in a set, E , of positive measure.^{17/} This now describes exactly the set of all $S_{\underline{X}}(\cdot)$ which are observationally equivalent to $S_X(\cdot)$.^{18/} The easiest kind of nondegenerate special case would perhaps be to take $\alpha \cdot S_X(w)$

away from $S_x(w)$ for w in, say, $(\pi, 3\pi]$ and to add these values to $S_x(w)$ for w in $(3\pi, 5\pi]$, $0 < \alpha < 1$, doing the symmetric operation for negative w .

For short, we will describe this procedure as "selecting an allowable P_x " where P_x satisfies the restrictions just described for $P(w)$. Similarly, $S_{xy} = S_{xy} + Q$, where, because $S_{xy}(\lambda) = \overline{S_{xy}(-\lambda)}$, $Q(\lambda) = \overline{Q(-\lambda)}$; and, of course, Q must satisfy the "adding up" property $F_n[Q] = 0$. There is no requirement on Q analogous to the positive definite Hermitian restriction imposed on P_x . The formal definition of $Q(w)$ is identical to (15) except that the $N \times 1$ vector $S_{xy;k}$ replaces the matrix $S_{x;k}$. "Selecting an allowable Q_{xy} " means, as before, choosing a nontrivial decomposition satisfying the conjugate symmetry property, and is equivalent to selecting an observationally equivalent S_{xy} .

The observationally equivalent \tilde{b} may now be represented as $\tilde{b} + \Delta\tilde{b}$, with a formula for $\Delta\tilde{b}$ to be derived. The equation $\tilde{b} = S_x^{-1} S_{xy}$ has two meanings: for the "true" (S_x, S_{xy}) it gives the "true" \tilde{b} ; it also gives the form of identification--for any potential (S_x, S_{xy}) pair, the corresponding \tilde{b} is given by it. Consequently, for $\tilde{b} + \Delta\tilde{b}$ to be observationally equivalent to \tilde{b} , there must be at least one allowable P_x and at least one allowable Q_{xy} such that

$$(17) \quad \tilde{b} + \Delta\tilde{b} = (S_x + P_x)^{-1} (S_{xy} + Q_{xy}).$$

Rearranging yields $(S_x + P_x) \Delta\tilde{b} = -P_x \tilde{b} + Q_{xy}$ which, depending on whether it is folded or not, yields,

$$(18a) \quad \Delta\tilde{b}(w) = (S_x + P_x)^{-1} (-P_x \tilde{b} + Q_{xy})$$

or

$$(18b) \quad F_n [S_x \tilde{\Delta b}] (w) = -F_n [P_x \tilde{b}] (w)$$

(17) and (18a) are equivalent, and give the answer to the question: which \tilde{b} are observationally equivalent to a given (\tilde{b}, S_x, S_{xy}) when only (S_x, S_{xy}) ^{19/} may be observed? (18b), which follows from (13b) as well, suffers since it makes no mention of Q_{xy} : if a $\tilde{\Delta b}$ is found that satisfies it, the question of whether there exists an allowable Q_{xy} which is consistent with (18a) remains. Finally, while all the discussion has been in terms of the spectral characteristics, adoption of the assumptions of the previous propositions allow the interpretation of B and b as lag distributions.

Without making additional assumptions, it appears nothing more can be said. However, by placing specific restrictions on characteristics of the processes, it is possible to proceed: either as in Sims [10] and Geweke [3], to force $b(\cdot)$ and $B(\cdot)$ to be "close"; or, as here, to force identification. In fact, extending an assumption analyzed in each of these papers for different reasons^{20/} leads to our next result.

We need first to give two definitions, which might be understood from their symbols without explanation. For a sequence of vectors or matrices $(R_{XY}$ or $R_X)$ to be in \underline{L}_1 we require their components to be absolutely summable; and for a vector or matrix of (real- or complex-valued functions, S_{XY} or $S_X)$ to be in \underline{L}_1 we require that each component be absolutely integrable. As before, we have refrained from including in the notation the domains which will be clear from the context and may on occasion be noted explicitly.

Proposition 4 Assume: (i) $R_{xy}(\cdot)$ and $R_x(\cdot)$ are given between the lattice points $L_n \equiv \{\frac{t}{n}: t \text{ integer}\}$ by linear interpolation, (ii) $R_{xy}(\cdot)$

and $R_X(\cdot)$ are in \underline{L}_1 , and (iii) S_X is essentially bounded below. Then $B(\frac{t}{n})$ and $b(t)$ are both identified from the discrete data; moreover, $b(\frac{t}{n}) = \begin{cases} B(\frac{t}{n}) & t \text{ integer} \\ 0 & \text{otherwise} \end{cases}$, where the interpretation is that $b(\cdot)$ is a row of delta functions with weights given by $B(\frac{t}{n})$ on the lattice points L_n

Proof: (i) says that

$$R_X(\frac{t+s}{n}) = (1-|s|)R_X(\frac{t}{n}) + |s|R_X(\frac{t}{n} + \frac{1}{n} \cdot \frac{s}{|s|})$$

and

$$R_{XY}(\frac{t+s}{n}) = (1-|s|)R_{XY}(\frac{t}{n}) + |s|R_{XY}(\frac{t}{n} + \frac{1}{n} \cdot \frac{s}{|s|})$$

for t integer, and $|s| \leq 1$. ($\frac{s}{|s|}$ is taken to be zero at $s=0$.) Defining $r_0(w) = \begin{cases} 1-|w| & |w| \leq 1 \\ 0 & |w| > 1 \end{cases}$ and interpreting $R_X(\frac{t}{n})$ and $R_{XY}(\frac{t}{n})$ as rows of delta functions with weights given by their values on L_n , we may rewrite assumption (i) as

$$R_X(\frac{t+s}{n}) = \int_{-\infty}^{\infty} R_X(\frac{t+s}{n} - w) r_0(w) I_N = R_X * r_0 \cdot I_N(\frac{t+s}{n})$$

and

$$R_{XY}(\frac{t+s}{n}) = \int_{-\infty}^{\infty} R_{XY}(\frac{t+s}{n} - w) r_0(w) I_N = R_{XY} * r_0 \cdot I_N(\frac{t+s}{n})$$

where, because of the finite support of $r_0(\cdot)$ and the nature of $R_X(\cdot)$ and $R_{XY}(\cdot)$, the integrals are sums of at most two numbers. R_X and R_{XY} will be in \underline{L}_2 if they are in \underline{L}_1 ; that they are in \underline{L}_1 follows from (i), (ii), and their sampling relation to R_X and R_{XY} -- indeed, any corresponding components have equal \underline{L}_1 and L^1 norms. Thus, $R_X \in \underline{L}_2$ ($\underline{L}_2(-n\pi, n\pi)$ displaying the domain) so that

$$\int_{-n\pi}^{n\pi} \lambda_1^2(w) dw < \infty;$$

with (iii), we have Proposition 3 in force, so that

$$\tilde{B} = \tilde{R}_X^{-1} \tilde{R}_{XY} = S_X^{-1} S_{XY}$$

is also in \mathbb{L}_2 and the classical-inverse Fourier transform may be applied to \tilde{B} to yield the $B^{21/}$ of model (B). $R_X(\cdot)$ and $R_{XY}(\cdot)$ as generalized functions have Fourier transforms which are periodic. The Fourier transform of $r_0 \cdot I_N$ is $\frac{2}{w} (1 - \cos w) \cdot I_N$, obviously positive definite a.e. The convolution R_X thus has $\tilde{R}_X = \tilde{R}_X \cdot \frac{2}{w} (1 - \cos w)$ which is clearly in \mathbb{L}_2 from consideration of the right-hand side. The spectral characteristic, \tilde{b} , satisfies $\tilde{b} = S_X^{-1} S_{XY}^{21/}$ if we had $\tilde{R}_X = S_X$ and $\tilde{R}_{XY} = S_{XY}$, then

$$\tilde{b} = (\tilde{R}_X \cdot \frac{2}{w} (1 - \cos w))^{-1} (\frac{2}{w} (1 - \cos w) R_{XY}) = \tilde{R}_X^{-1} \tilde{R}_{XY} = \tilde{B}$$

and the conclusion follows. We have $R_X(t) = \int_{-\infty}^{\infty} S_X(w) e^{iwt} dt$, with $S_X(\cdot) \in \mathbb{L}_1$. But the relation $R_X \xleftrightarrow{U} \tilde{R}_X$ is the unitary Fourier isomorphism alluded to in footnote 3, extended to $\mathbb{L}_2(-\infty, \infty)$ matrices in the Wiener-Masani manner. (That the components of R_X (and so of \tilde{R}_X) are in $L^2(-\infty, \infty)$ can be seen directly since they are in $L^1(-\infty, \infty)$ and bounded.) With $S_X(\cdot)$ and $R_X(\cdot)$ both having absolutely integrable components, the classical inversion theorem of footnote 3 applies. But now $S_X(\cdot)$ is also seen to be bounded, hence in $\mathbb{L}_2(-\infty, \infty)$, and it follows that $S_X \xleftrightarrow{U} R_X$, since \xleftrightarrow{U} agrees with the ordinary Fourier integral when domain and range element are both in $L^1(-\infty, \infty) \cap L^2(-\infty, \infty)$. (See [5], especially p. 510-513 for \xleftrightarrow{U} .) But then $S_X = \tilde{R}_X$ by the one to one property of \xleftrightarrow{U} . Q.E.D.

A computation in the proof showed that $w^2 S_X(w) = \tilde{R}_X(1 - \cos w)$.

The nonintegrability of the diagonal elements of the right-hand side shows

$$\int_{-\infty}^{\infty} w^2 S_{x_{ii}}(w) dw = \infty, \quad i=1, \dots, N, \quad \text{i.e.,}$$

that no component of the x vector can be mean-square differentiable, an observation also noted in [3], p. 7.

In the special case of $n=1$, the continuous model (A) is identified from only the coarse observational structure (D). It need hardly be said that the hypotheses are less likely to be fulfilled in this case: linear interpolability is required over longer segments. Indeed, by using X_i , i such that $n_i > n_N$, $R_{X_i}(\frac{t}{n_i})$ may be computed and part of the interpolability assumption thus checked when data of the model (C) pattern are available.

A second situation with identification possibilities requires the abandonment of the assumption that x be linearly regular,^{22/} since the latter implies that $S_x(\cdot)$ has the same rank a.e. If $S_x(w)$ vanishes for $w \in [-n\pi, n\pi]$, then $S_{xy}(w)$ vanishes outside this interval as well; consequently, $S_x \tilde{b} = S_{xy}$ is consistent with any values for $\tilde{b}(w)$ outside this interval. Otherwise put, model (A) is not identifiable with respect to continuous data, let alone discrete data. Processes of this type, whose spectral density matrix has for support a proper, compact set of the allowable support (here, $(-\infty, \infty)$ for real processes) are called band-limited. Their values $x(t)$, for all t , can be captured knowing only their values at $x(\frac{t}{n})$, provided the spectrum is bounded.^{22/} This result, bringing to mind interpolation and the previous result, suggests a closer examination.

The vanishing of S_x in an area says there is "no action" in x attributable to that part of the spectrum (here, the high-frequency components). A reasonable identifying assumption for $\tilde{b}(w)$ might therefore be to for it to have "no action" at these frequencies. We adopt

this for the purposes of the next proposition; now, with respect to the continuous data,

$$\tilde{b}(w) = \begin{cases} S_x^{-1}(w) S_{xy}(w) & |w| \leq n\pi \\ 0 & |w| > n\pi \end{cases}$$

Recalling (13b), $S_X \tilde{B} = F_n [S_x(\cdot) \tilde{b}(\cdot)]$, and noting that F_n effectively causes no loss of information in this situation because $S_x(w+2\pi nk)$ and $\tilde{b}(w+2\pi nk)$ both vanish for an integer k different from zero, we have

$$(19) \quad S_X(w) \tilde{B}(w) = \begin{cases} S_x(w) \tilde{b}(w) & |w| \leq n\pi \\ S_x(w^*) \tilde{b}(w^*) & |w| > n\pi \text{ and } w = w^* \pmod{2\pi n} \\ S_x(w^*) \tilde{b}(w^*) & |w^*| \leq n\pi \end{cases}$$

Since

$$S_X(w) = S_x(w), \quad |w| \leq n\pi$$

it follows that $\tilde{b}(w)$ is identified from the discrete data as

$$\tilde{b}(w) = \begin{cases} \tilde{B}(w) & |w| \leq n\pi \\ 0 & |w| > n\pi \end{cases}$$

Again we have equality of \tilde{b} and \tilde{B} in a sense. To translate this result into the time domain, we observe that

$$(20) \quad \tilde{b}(w) = \tilde{B}(w) \chi_{[-n\pi, n\pi]}(w) = \tilde{B}(w) \overbrace{\frac{\sin(\cdot)\pi n}{(\cdot)\pi}}(w), \frac{23}{}$$

where the indicator or characteristic function $\chi_{[a,b]}(\cdot)$ has been previously defined. Obviously $\tilde{b}(\cdot) \in \underline{L}_2(-\infty, \infty)$ if and only if $\tilde{B}(w) \in \underline{L}_2[-n\pi, n\pi]$, in which case we may take an inverse Fourier transform. By adopting the hypotheses of Proposition 3 this is ensured, and

$$(21) \quad \tilde{B}(w) = \sum_{t=-\infty}^{\infty} B\left(\frac{t}{n}\right) e^{-iw\left(\frac{t}{n}\right)}$$

holds for some $B\left(\frac{t}{n}\right) \in \underline{L}_2$. The inverse Fourier transform of $\tilde{B}(w)$ is thus a row of delta functions with weights on the lattice L_n given by $B\left(\frac{t}{n}\right)$.

Abusing the notation slightly by referring to this generalized function as $B(\cdot)$, we use the convolution rule on (20) to conclude

$$(22a) \quad b\left(\frac{r}{n}\right) = \int_{-\infty}^{\infty} B\left(\frac{r}{n} - s\right) \frac{\sin s\pi n}{s\pi} ds = n \cdot B\left(\frac{r}{n}\right) \text{ if } \frac{r}{n} \in L_n.$$

If $\frac{r}{n} \notin L_n$, i.e., if r is not an integer, the first equality holds and we may again derive $b\left(\frac{r}{n}\right)$ from observable $B(\cdot)$, but the relation is not so simple. We have $\frac{r}{n} = \frac{t}{n} + \delta$, $0 < \delta < \frac{1}{n}$, t integer. Then $B\left(\frac{r}{n} - s\right) = B\left(\frac{t}{n} + \delta - s\right)$ has its support located at s values in the displaced lattice $L_{\delta;n} \equiv \{\delta + \frac{t}{n}; t \text{ integer}\}$. Consequently

$$(22b) \quad b\left(\frac{r}{n}\right) = \sum_{j=-\infty}^{j=\infty} B\left(\frac{t-j}{n}\right) \frac{\sin(\delta + \frac{j}{n})\pi n}{(\delta + \frac{j}{n})\pi}.$$

It remains to check that our candidate for $(y(t) | H_X)$,

$$b' * x(t) \equiv \int_{-\infty}^{\infty} b(t-s)x(s)ds,$$

makes sense. We are given pause by the fact that Proposition 1 is not likely to be applicable: for $b(\cdot)$ to be in $L_1(-\infty, \infty)$, $\tilde{b}(w)$ must be continuous, and a glance at (20) shows this to be unlikely. However, the matching condition can be verified directly, since

$$\int_{-\infty}^{\infty} \tilde{b}'(w) S_X(w) \tilde{b}(w) dw = \int_{-n\pi}^{n\pi} \tilde{B}'(w) S_X(w) \tilde{B}(w) dw < \infty$$

by Proposition 3 and using the first equality in (20). This concludes the proof of

Proposition 5 If (a) the spectral characteristic $\tilde{b}(w)$ is chosen to vanish with S_X ; ^{24/} (b) the continuous process X is limited to the band $B \equiv [-n\pi, n\pi]$; (c) $S_X(w)$ ($=S_X(w)$ on B) is essentially bounded below and either $\lambda_1(w) \in L^2[-n\pi, n\pi]$ or S_X is essentially bounded above; and (d)

$R_{XY} \in \mathcal{L}_2$; then the continuous and discrete models (A) and (B) are both identified from the discrete data by projection. Moreover, equations (22) provide $b(\cdot)$ in terms of $B(\cdot)$.

In [3], Geweke's Proposition 3 shows, for the case $n=1$, an "inverse" result for band-limited processes:

$$(23) \quad B(t) = \int_{-\infty}^{\infty} \frac{\sin \pi(t-s)}{\pi(t-s)} \cdot b(s) ds, \quad t \text{ integer.}$$

His interest in (23) is that each component of B involves only the corresponding component of b --there is no contamination. Our result shows that, when $b(\cdot)$ is identified by (a), the integral equation (23), which holds only at integer t , inverts!

Even if our interest is not in the continuous model, the last two propositions provide some clues about when a "coarse" observational pattern will serve to identify a "less coarse" model. Before taking this up, we will respond to the challenge of remark (ii) as it pertains to model (C). It is still only tacitly defined by equations (6) and (7); Propositions 2 and 3 do not directly apply to it. We explore this question, and its relation to the other discrete models in the next section.

V. The Identification of Model (C) and the Relations Among the Discrete Models

One way to interpret (10), $R_{XY}(\frac{t}{n}) = R_X * B(\frac{t}{n})$, is as a necessary and sufficient condition on B for $B' * X(t)$ to be the projection $(Y(t) | H_X)$, provided that $B' * X(t)$ is well-defined. Put differently, given a B by whatever means, if $B' * X(t)$ can be shown to be in H_X , then checking (10) is equivalent to checking whether $B' * X(t)$ is the projection. While (10) referred to model (B), an analogous system of convolution equations

holds for model (C). Our plan is to develop them, solve them, and then apply the above considerations to justify our solution.

Let us recall the model (C) in the form using (6)':

$$(6)' \quad Y(t) = \sum_{j=1}^N \sum_{s=-\infty}^{\infty} B_j(t - \frac{s}{n_1}) X_j(\frac{s}{n_1}) + U(t)$$

$$B_j(t - \frac{s}{n_1}) = 0, \quad \frac{s}{n_1} \neq \frac{r}{n_j}; \quad s, r \text{ integer}; \quad j=1, \dots, N$$

$$(7) \quad EU(t) X_j(\frac{s}{n_j}) = 0, \quad t, s \text{ integer}, \quad j=1, \dots, N.$$

The following notation will be found useful. As before, denote by L_j the lattice $\{\frac{t}{n_j}: t \text{ integer}\}$ $j=1, \dots, N$. Recalling our assumption on the data structure, define the integers

$$m_{i,j} \equiv [\frac{n_i}{n_j}] = \frac{n_i}{n_j}; \quad N \geq j \geq i \geq 1; \quad i, j \text{ integer.}$$

By $H(S)$ will be meant the smallest Hilbert space containing, or spanned by, the set S . (This is the same concept mentioned in the proof of Proposition 1 and footnote 1.) In this terminology, we have: (i) the domain of $X_j(\cdot)$ is L_j ; (ii) the lattices are nested, $L_1 \supseteq L_2 \supseteq \dots \supseteq L_N$; (iii) $B_j(\cdot)$ is defined on L_1 but may be nonzero only on L_j, L_{j+1}, \dots, L_N . (iv)

$$H_D \equiv H(\bigcup_{j=1}^N (X_j(s), s \in L_N)) \subseteq H(\bigcup_{j=1}^N (X_j(s), s \in L_j)) \equiv H_C \subseteq H(\bigcup_{j=1}^N (X_j(s), s \in L_1)) \equiv H_B = H_X \subseteq H(\bigcup_{j=1}^N (x_j(s), s \text{ real})) \equiv H_A.$$

The models (A) to (D) are thus identified by projecting $Y(t)$ onto H_A to H_D , respectively, all of which may be regarded as closed subspaces of $H_{(x,y)} \equiv H(H_A \cup \{y(t), t \text{ real}\})$. This observation may clarify remarks (ii) and (iv) of Section III.

Multiplying both sides of (6)' by $X_i(t - \frac{r}{n_i})$, taking expectations, and making use of (7) yields

$$(24) \quad R_{X_i Y}(\frac{r}{n_i}) = \sum_{j=1}^N B_j * R_{X_i X_j}(\frac{r}{n_i}), \quad i=1, \dots, N,$$

where the j^{th} term in this sum is

$$(25) \quad B_j * R_{X_i X_j}(\frac{r}{n_i}) = \sum_{s=-\infty}^{s=\infty} B_j(t - \frac{s}{n_1}) R_{X_i Y_j}(\frac{r}{n_i} - (t - \frac{s}{n_1})),$$

and the constraints may be written

$$(26) \quad B_j(s) = 0, \quad s \in L_i \setminus L_j, \quad N \geq j > i \geq 1.$$

We will make use of the

Covariance Identification Lemma If $u(\cdot)$ and $v(\cdot)$ are two stationary processes observed on the lattices L_m and L_n , respectively, we may estimate consistently, and hence regard as identifiable, $R_{uv}(s)$, $s \in L_r$, r the least common multiple of m and n . ($L_m \equiv \{\frac{t}{m} : t \text{ integer}\}$)

Proof:

$$R_{uv}(\frac{kn-hm}{mn}) = E u(t - \frac{k}{m}) v(t - \frac{h}{n}),$$

by covariance stationarity, so that for any integers k and h ,

$$\frac{1}{|T|} \sum_{t \in T} u(t - \frac{k}{m}) v(t - \frac{h}{n}) \equiv \hat{R}_{uv}(\frac{kn-hm}{mn})$$

may be formed, where $|T|$ is the number of points in T , a finite set for which data on both terms in the product are available. This estimator is certainly consistent. Choosing $h=0$ (respectively $k=0$) shows $R_{uv}(s)$, $s \in L_m$ (respectively $R_{uv}(s)$, $s \in L_n$) are identified. Clearly the set of points so identified in arbitrarily large samples T is a lattice. The

precise description of the lattice gives its step, which is evidently $\frac{d}{mn}$, with d the minimum positive integer equal to $kn-hm$, where k and h vary over all integers. Since m and n have least common multiple r , $n\beta = \alpha m = r$ with α and β relatively prime integers, and $kn-hm = \frac{r}{\alpha\beta}(\alpha k - \beta h)$, we want the least positive value of $\alpha k - \beta h$ over all integers k and h . But $\alpha k - \beta h$ can always be made to equal 1, in which case $d = \frac{r}{\alpha\beta} = \frac{m}{\beta} = \frac{n}{\alpha}$, so that

$$\frac{d}{mn} = \frac{\frac{m}{\beta}}{mn} = \frac{1}{\beta n} = \frac{1}{r}.$$

This last observation must be evident to all number theorists, and its tedious induction proof is omitted. Q.E.D.

Applying this lemma to the convolution equation system (24)--(26) shows that all of the terms $R_{X_i Y}$ and $R_{X_i X_j}$ which effectively enter may be taken as known. For while the right-hand side of (25) contains all the terms $R_{X_i X_j}(s)$, $s \in L_1$, only when $t - \frac{s}{n_1} \in L_j$ will such a term have a nonzero coefficient, and it is in precisely that case that $R_{X_i X_j}$ is available, because the sum of the two lattice points $t - \frac{s}{n_1} \in L_j$ and $\frac{r}{n_i} \in L_i$ is a lattice point.

Solving this system--even "operationally"--appears to be much more formidable than solving (10), where we were able to justify multiplying by R_X^{-1} . If (24) were written in matrix notation as $R_{XY} = R_{XX} * B$, we evidently have two choices for the domain of the elements of R_{XX} : it may be uniform, $\frac{1}{n_1}$ units apart, in which case many components are unobservable; or, it may be nonuniform, defined only at the frequencies which effectively enter (25), in which case it is hard to imagine even the formal construction of R_X^{-1} . In any event, the constraints must also be dealt with. Finally, the frequency with which the equality in (24) holds varies with the component i .

These difficulties can again be dealt with in the frequency domain, with the aid of an entirely finite version of the folding principle used in the previous section. Stated in the form in which it will be used, with reference to the underlying continuous processes suppressed, we have the

Finite Folding Lemma If the discrete processes X and Y have autocorrelation function $R_{XY}(\cdot)$ which is in principle observable n_1 times per period but is actually observed n_j times per period, $n_1/n_j = m_{1,j}$, then the observed cross-spectral density is

$$F_{n_j;n_1}[S_{XY}(\cdot)](\omega) = \sum_{k=0}^{m_{1,j}-1} S_{XY}(\omega+2\pi n_j k), \text{ a.e., } \omega \in [0, 2\pi n_j],$$

where the densities are assumed to exist. ^{25/}

Proof: For any integer r,

$$\begin{aligned} \frac{r}{n_j} &= \frac{rm_{1,j}}{n_1}, \quad R_{XY}\left(\frac{r}{n_j}\right) = R_{XY}\left(\frac{rm_{1,j}}{n_1}\right) = \int_0^{2\pi n_1} e^{i\omega \frac{rm_{1,j}}{n_1}} S_{XY}(\omega) d\omega = \\ &= \sum_{k=0}^{m_{1,j}-1} \int_0^{2\pi n_j} e^{i(\omega+2\pi n_j k) \frac{rm_{1,j}}{n_1}} S_{XY}(\omega+2\pi n_j k) d\omega. \end{aligned}$$

The finite sum may be taken inside the integral and the exponential may be replaced by $e^{i\omega \frac{r}{n_j}}$. Q.E.D.

Observe that the notation $F_{n_j;n_1}[\](\cdot)$ indicates both domain and range of the operator. When the domain is clear from the context, as, say, in (30) where any fold is with respect to $[0, 2\pi n_1]$, the notation may be abbreviated to $F_{n_j}[\](\cdot)$. Valuable use will be made of the relation,

$$(27) \quad \sum_{s=-\infty}^{s=\infty} B_j \left(t - \frac{s}{n_1} \right) e^{i \left(t - \frac{s}{n_1} \right) w} \equiv \tilde{B}_j(w) = \tilde{B}_j(w + 2\pi n_j) = \sum_{s=-\infty}^{s=\infty} B_j \left(\frac{s}{n_j} \right) e^{i \left(\frac{s}{n_j} \right) w},$$

$$j=1, \dots, N,$$

which holds because of the constraint (26).

Finally, we Fourier transform (24) and consider the i^{th} equation.

The left-hand side is, by the Folding Lemma,

$$(28) \quad F_{n_i; n_1} [S_{X_i Y}] (w) = \sum_{k=0}^{m_{1,i} - 1} S_{X_i Y}(w + 2\pi n_i k).$$

The folding here is relative to the cross-spectrum we would observe if the data on X_i were on the lattice L_1 . There is "true" folding relative to the observational pattern of model (B): as long as we only observe X_i on L_1 , we cannot unfold. We will see that the right-hand side, however, contains a "pseudo-folding" for the terms j , $1 \leq j \leq i-1$. The i^{th} equation reads

$$(29) \quad F_{n_i; n_1} [S_{X_i Y}] (w) = \sum_{j=1}^{i-1} \sum_{k=0}^{m_{j,i} - 1} B_j(w + 2\pi n_i k) S_{X_i X_j}(w + 2\pi n_i k) + \sum_{j=i}^N \tilde{B}_j(w) F_{n_i; n_1} [S_{X_i X_j}] (w)$$

Consider, for example, $j=1$. We have seen that $R_{X_i X_1}(s)$ is known for $s \in L_1$; consequently, $S_{X_i X_1}(w)$ is known for all w , $0 \leq w \leq 2\pi n_1$. The $j=1$ term in (29) thus involves a (B_j -weighted) folding of what we know, as opposed to the left-hand side which folds part of a cross-spectrum we can't observe. This is the sense in which "pseudo-folding" is to be taken, since now we show how this latter can be undone.

Since (29) is to hold for all w (or, because of periodicity, for all w in $[0, 2\pi n_i]$) we may substitute $w + 2\pi h$, $h=0, 1, 2, \dots, n_i - 1$

successively for w . These n_i equations with w now regarded as belonging to $[0, 2\pi]$ are clearly equivalent to the original equation with w belonging to the entire $[0, 2\pi n_1]$ interval. By expanding in this way, we can express the desired \tilde{B} vector as the solution of the matrix equation system (30), displayed as p. 33 which will be on occasion abbreviated by $S_{XY}(w) = S_{X;C}(w)\tilde{B}(w)$.

We describe $S_{X;C}$ further, as its properties are of crucial importance for the identification of model (C). It is square, of dimension $(\sum_{i=1}^N n_i)$, and Hermitian. As indicated it is most easily visualized in terms of its N^2 blocks. The (j, k) block is: diagonal, of dimension n_j , if $j=k$; if $j < k$, it is $n_j \times n_k$ consisting of $m_{j;k}$ diagonal matrices stacked vertically. The elements on the diagonal are the cross-spectral density $F_{n_j, n_1} [S_{X_j X_k}] (w)$, as "truly" folded by the observational pattern. Although formally indicated to show the general pattern, no folding occurs in the first block-row or in the first block-column. Since all folding is with respect to $[0, 2\pi n_1]$, an n_1 has been omitted from $F_{n_i; n_1} [] (\cdot)$ in (30). In winding diagonally down a stack only one period is traversed--there is never any repetition. The width of the blocks diminishes with rightward movement. By inspecting the $n_1 + 1^{st}$ row, the "pseudo-folding" is seen, in that $S_{X_2 X_1}(w)$ and $S_{X_2 X_1}(w + 2\pi n_2)$, which are individually observable, occur together in the $n_1 + 1^{st}$ equation. Finally, all elements of $S_{XY}(\cdot)$ and $S_{X;C}$ are identified, or data-determined, so that we might hope that

$$(31) \quad \tilde{B}(w) = S_{X;C}^{-1}(w) S_{XY}(w), \quad w \in [0, 2\pi]$$

holds and our indicated programme may be carried out. Before continuing we make a slight digression in the next paragraph.

Since models (B) and (D) were seen as mathematically special cases of model (C), the equations

$$(32) \quad S_{XY}(w) = S_{X;B}(w) \tilde{B}_f, \quad w \in [0, 2\pi]; \text{ and}$$

$$(33) \quad S_{XY}(w) = S_{X;D}(w) \tilde{B}_c, \quad w \in [0, 2\pi],$$

are particularly instructive special cases of (30), in which the (j, k) block is always diagonal.^{26/} The technique of forming the super-matrix $S_{X;C}$ is necessitated by the fact that no orthogonal-increments representation of the subspace H_C we are projecting into exists, and creating one would be more difficult than our direct solution. That $S_{X;B}$ and $S_{X;D}$ actually yield the previous solutions derived on the basis of spectral representations is a gratifying check on the validity of the present technique.

The spirit of our inquiry involves making assumptions, insofar as possible, on the continuous process (y, x) and analyzing their implications for the discrete models. A natural assumption, suggested by the logical requirement that $|S_{x_i x_j}(\cdot)| \leq S_{x_i}(\cdot)^{1/2} S_{x_j}(\cdot)^{1/2}$, is that the (continuous) spectral density matrix have a dominant diagonal--the positive diagonal element exceeding the sum of the moduli of the off-diagonal elements of that row. But from the folding formulae (11a) and (11b) this property carries over to the discrete model (B)-spectra defined on $[-n\pi, n\pi]$ (or equivalently $[0, 2\pi n]$: in the sequel we may use these interchangeably); also, directly or by Finite Folding, to model (D). We have only to consider model (C) to establish

Proposition 6 $S_{X;B}$, $S_{X;C}$, $S_{X;D}$ are all positive definite, and hence invertible, under the assumption that the continuous spectral density matrix is pointwise dominant diagonal.

Proof: The matrices are all Hermitian. Having written out $S_{X;C}$ in (30), inspection shows that, upon unfolding the typical diagonal element $F_{n_i; n_1} [S_{X_i}] (w)$ into $S_{X_i} (w) + S_{X_i} (w + 2\pi n_1) + \dots + S_{X_i} (w + 2\pi n_1 (m_{1,i} - 1))$, each term bounds the sum of the off-diagonal terms in the same frequency range. Consequently, all the matrices are dominant diagonal, hence by the well-known theorem, positive definite. Q.E.D.

Two remarks are in order: (i) the immediacy of this result is illusory; rather, the power of (29) and the felicitous substitution which led to (30) are reflected; (ii) as satisfying as it is to derive a previous assumption, this result by itself, we emphasize, is inadequate: \tilde{B} , \tilde{B}_f , \tilde{B}_c may exist, but the nature of the implied inverse Fourier transform needs to be checked. This we do in the very important

Proposition 7 Any hypotheses on the continuous model which validate model (B) under Proposition 3 also validate model (C). Specifically, when $S_X^{N \times N}$ is essentially bounded below, and also

$$(i) \quad \int_{-n\pi}^{n\pi} \lambda_1^2(w) dw \quad ,$$

or (ii) S_X is essentially bounded above, then these same properties hold for $S_{X;C}$. Consequently with $R_{XY}(\frac{t}{n_1}) \in \underline{\ell}_2$, it follows that $\tilde{B} \in \underline{\ell}_2$ in (30)^{27/} so that: \tilde{B} has an inverse Fourier transform in $\underline{\ell}_2$, $B(\frac{t}{n_1})$, which satisfies the matching condition; hence, $B' * X(t)$ is well defined.

Proof: Assume $c_1 I_N < S_X(w) < c_2 I_N$, $w \in [0, 2\pi n]$, where of course $n = n_1$. We now seek to find similar bounds for $S_{X;C}$, $w \in [0, 2\pi]$. To this end, the notation for a typical test vector x_c for the Rayleigh quotient is needed. Consider the layout:

$\| \cdot \|_E$ the Euclidean norm, and

$$\|x_c\|_E^2 < \sum_{k=1}^{n_1} x_k' x_k. \quad .$$

Consequently,

$$(c_1) x_c' I_N x_c \leq x_c' S_{X;C}(w) x_c \leq (c_2^{m_{1,N}}) x_c' I_N x_c,$$

so that $S_{X;C}(w)$ is also essentially bounded from above and below. In case $S_X(w)$ meets condition (i) instead, we note that the same expansion for the quadratic form (or Rayleigh quotient) yields the inequality

$$\sup_{\|x_c\|_E = 1} (x_c' S_{X;C}(w) x_c) = \lambda_{1;C}(w) \leq \sum_{k=0}^{n_1} \lambda_1(w+2\pi(k-1)),$$

where $\lambda_{1;C}(w)$ is the largest eigenvalue of $S_{X;C}$ and $\lambda_1(\cdot)$ has the same meaning as before. Both sides of this last inequality are positive a.e.,

since $S_{X;C}$ is bounded from below by hypothesis. By squaring both sides and integrating over $[0, 2\pi]$, the fact that $\int_0^{2\pi} \lambda_{1;C}^2(w) dw < \infty$ follows from hypothesis (i) (with the trivial change of variable) and the Hölder inequality applied to the cross-product terms in the sum. (The equality involving $\lambda_{1;C}(w)$ was also used in the proof of Proposition 2.) Consequently, $S_{X;C}$ is in $\underline{\mathbb{L}}_2$, and since it is bounded from below, its inverse exists, is bounded from above, and consequently is in $\underline{\mathbb{L}}_2$ as well. Finally, we indicate more explicitly than on p. 14-15 the finiteness of the matching condition:

$$\int_0^{2\pi} R_{XY}(w)' R_X^{-1}(w) R_{XY}(w) dw \leq \int_0^{2\pi} R_{XY}'(w) c \cdot I R_{XY}(w) dw \quad ,$$

where I is of dimension $\sum_{i=1}^N n_i$ and c is the constant in the upper bound for $S_{X;C}^{-1} = \tilde{R}_X^{-1}$. We used the fact that $R_{XY}(\cdot) \in \underline{\mathbb{L}}_2$ implies $\tilde{R}_{XY}(\cdot) \in \underline{\mathbb{L}}_2$. Q.E.D.

Model (D)

Note: As above, we take $n_N=n$, and, to avoid repetition, "factor" out F_n which is common to all terms.

$$(33) \quad F_n \begin{pmatrix} S_{X_1 Y}(w) \\ S_{X_2 Y}(w) \\ \vdots \\ S_{X_N Y}(w) \end{pmatrix} = F_n \begin{pmatrix} S_{X_1}(w) & S_{X_1 X_2}(w) & \dots & S_{X_1 X_N}(w) \\ S_{X_2 X_1}(w) & S_{X_2}(w) & & S_{X_2 X_N}(w) \\ \vdots & \vdots & & \vdots \\ S_{X_N X_1}(w) & S_{X_N X_2}(w) & \dots & S_{X_N}(w) \end{pmatrix} \begin{pmatrix} \tilde{B}_{1;c}(w) \\ \tilde{B}_{2;c}(w) \\ \vdots \\ \tilde{B}_{N;c}(w) \end{pmatrix}$$

$$\text{or, } S_{XY}(w) = S_{X;D}(w) \tilde{B}_c(w)$$

With this notation, and adding a subscript to S_{XY} to identify the model to which it pertains, we have

$$I_{C;B} S_{XY;B} = I_{C;B} S_{X;B} \tilde{B}_f = S_{XY;C} = S_{X;C} \tilde{B}$$

and

$$I_{D;C} S_{XY;C} = I_{D;C} S_{X;C} \tilde{B} = S_{XY;D} = S_{X;D} \tilde{B}_c$$

Just as Proposition 7 proved properties about $S_{X;C}$ from assumptions about $S_{X;B}$, so could we prove that the same properties follow for $S_{X;D}$ under the same assumptions. In particular, one consequence of "essential boundedness above and below" was invertibility. The above equations may thus be rewritten, for $w \in [0, 2\pi]$, and suppressing w which occurs as the argument in all of these matrix functions, as

$$(34) \quad \tilde{B} = S_{X;C}^{-1} I_{C;B} S_{X;B} \tilde{B}_f \equiv \tilde{r}_{X;CB} \cdot \tilde{B}_f$$

and

$$(35) \quad \tilde{B}_c = S_{X;D}^{-1} I_{D;C} S_{X;C} \tilde{B} \equiv \tilde{r}_{X;DC} \cdot \tilde{B}.$$

Combining these equations suggests the definition for

$$I_{D;B} \equiv I_{D;C} I_{C;B} \quad \text{and} \quad \tilde{r}_{X;DB} \equiv \tilde{r}_{X;DC} \tilde{r}_{X;CB}.$$

These entail

$$(36) \quad \tilde{B}_c = \tilde{r}_{X;DC} \tilde{B} = \tilde{r}_{X;DC} \tilde{r}_{X;CB} \tilde{B}_f = \tilde{r}_{X;DB} \tilde{B}_f,$$

where of course,

$$\tilde{r}_{X;DB} = S_{X;D}^{-1} I_{D;C} I_{C;B} S_{X;B} = S_{X;D}^{-1} I_{D;B} S_{X;B}.$$

The equations (34)-(36) express the relations among the discrete models (B)-(D). They are, as perhaps might have been expected, similar to the result obtained in [3] and [10], $\tilde{B}_c = F_{n_N} [\tilde{r}_x \cdot \tilde{b}]$, which relates, in our terminology, models (A) and (D), where $\tilde{r}_x(w) = S_{X;D}^{-1}(w) S_X(w)$. Because multiplication in the frequency domain corresponds to convolution in the time domain, these equations are readily interpreted.

The general lack of identification of model (B) (respectively, model (C)) from the observational pattern of model (C) (respectively, model (D)) appears in the rectangular dimensionality of

$$\begin{array}{cc} \left(\sum_{i=1}^N n_i \right) \times N n_1 & N \times \left(\sum_{i=1}^N n_i \right) \\ \tilde{r}_{X;CB} & \text{and} \quad \tilde{r}_{X;DC} \end{array}$$

Consequently, neither (34) nor (35) may be inverted if any independent variable is observed more often than another. A discussion of the

precise nature of the underidentifiability would follow the lines of p. 18-21. Results analogous to those identifying \tilde{b} from \tilde{B} given above could now be given. For the reader who has followed the argument to this stage, however, to mention such results is almost to prove them.

We do call attention, nevertheless, to a special case when the data are in the observation pattern of model (C) and when the coefficients in model (B) are to be estimated. In general, owing to nonzero nondiagonal elements in $\tilde{r}_{X;CB}$ (the discrete analogue of "contamination") (34) shows that this cannot be done. However, for X_i , the desired coefficients $\tilde{B}_{i,f}^{(w+2\pi(k-1)n_i)}$ are identified, and consequently estimable, in the special case in which $n_i = n_1$ and $R_{X_i X_j}(\frac{t}{n_1}) = 0$ holds for all integer t , and for all j different from the fixed i . This can be seen directly from (30) in its expanded form. Of course, it is the strength of these assumptions which allows the achievement of more identification than we are otherwise entitled to. As with the other such special cases, we caution that a good deal of robustness is required for their practical application.

Finally, reflecting on the very plausible assumption of the nested data sets of model (C), we remark on the effect of its relaxation. If observations were available on X_i at the rate of n_i times per period, ($\frac{n_i}{n_{i+1}}$ now not necessarily an integer) the least common multiple, r , say, would have been formed. The appropriate "fine" model would then have had all variables defined on the lattice $L_r = \{\frac{t}{r}, t \text{ integer}\}$. We would then have proceeded in the same manner, with the only difference being that the pattern in (30) would not be quite so simple, although it would remain "quasi-block-diagonal." The full generality of the Covariance Identification Lemma would then have been required.

VI. Conclusion

Within the framework of continuous time, jointly covariance stationary stochastic processes, conditions on a continuous process were given under which the usual distributed lag regressions were proved to be identified by projection for very general classes of observation patterns. For different observation patterns, different but related distributed lags were identified in this manner.

The frequency domain was seen to be the natural habitat for the study of the relationships among these projections, despite the lack at times of a Cramer-type representation of the sampled process. Propositions, meant to be interpreted as limiting cases, were advanced in which conditions that have been studied elsewhere allowed fine distributed lags to be identified from relatively coarse data structures. Emphasis was placed on the development of an apparatus which is in principle capable of analyzing the effects of temporal aggregation on a case by case basis.

Footnotes

1/In [10], Sims uses the same method of identification. As he notes, it is not the only route that can be taken: projection on the past and present is also possible, but not so tractable analytically. More accurately, the projection is onto the space of values, H_x , of the independent process. The latter concept is rigorously defined in Rozanov [9], p. 3. Briefly, it is the completion under the quadratic mean norm of the linear space of random variables spanned by $x(s)$, s real. As a general rule, the meaning that is understood by any undefined symbols such as

$$\int_{-\infty}^{\infty} e^{itw} dz_x(w),$$

the integral of a complex valued function with respect to a random measure, may be found in this source (with $\phi(d\lambda)$ replacing $dz_x(w)$) or Koopmans [6]. Since all future reference to Rozanov will be to his book, [9], we will omit this reference number in the sequel.

2/For a rigorous definition of linear regularity, consult Rozanov, p. 53 for the discrete case, p. 110 for the continuous case. In both cases the intuitive meaning is the same: the best forecast of the "infinitely removed" future is the mean; or, there are no deterministic components remaining in the processes being studied. Linear regularity of x implies S_x has constant rank a.e. We assume $|S_x| \neq 0$ a.e., unless noted in the sequel, which is consistent with linear regularity. (Theorem 2.4, p. 115, Rozanov).

3/Unfortunately, there are several "standards," as casual perusal of the sources cited in this paper shows. Differences can often be resolved by checking whether $R_{xy}(s)$ means $Ex(t)y(t+s)'$ or $Ex(t+s)y(t)'$, where y as well as x may be a vector. (The symbol $'$ will always denote (complex) conjugate transpose; \sim will denote Fourier transform.) In the former case, the usual equality

$$R_{xy}(t) = \int_{-\infty}^{\infty} e^{i\lambda t} S_{xy}(\lambda) d\lambda$$

is maintained (when working with x and y real) by defining

$$E z_x(\lambda) \overline{dz_y(w)} = \begin{cases} \overline{dF_{xy}(\lambda)} & \lambda = w \\ 0 & \lambda \neq w; \end{cases}$$

which contrasts with the latter (more common) definition of $R_{xy}(s)$, in which the right-hand side of the last expression is not conjugated. Whichever definition we choose, we also have

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} R_{xy}(t) e^{-i\lambda t} dt = S_{xy}(\lambda)$$

when we assume that

7/ See any text on functional analysis for the technical definition. One the author has found eminently readable is Bachman and Narici [1], in which Theorem 10.8 assures that projections always exist.

8/ From an identification viewpoint the observational frequency of $y(t)$ is unimportant. Say $Y(\frac{t}{n}) = y(\frac{t}{n})$, t integer, so that $Y(\cdot)$ is $y(\cdot)$ sampled n times per period, instead of once. Two cases now arise: i) $[\frac{t}{n}] = \frac{t}{n}$, some t , in which case the previous analysis applies, because we are relating Y at an integer to given set of independent variables, and, because of stationarity, the relation is the same regardless of the integer; and, ii) $\frac{t}{n} = [\frac{t}{n}] + \frac{j}{n}$, $1 \leq j \leq n-1$: now define $Y_j(t) = y(t + \frac{j}{n})$ and apply the previous analysis to Y_j and the X process.

9/ See Sims [11] for a proof of the scalar version of this assertion.

10/ As employed in [6], p. 162-164. The term is from filter theory.

11/ To be explicit about the identification tacitly made here, we have two relations involving R_X :

$$R_X(\frac{t}{n}) = \int_{-n\pi}^{n\pi} S_X(w) e^{iw(\frac{t}{n})} dw \text{ and } R_X(\frac{t}{n}) = \int_{-n\pi}^{n\pi} \tilde{R}_X(w) e^{iw(\frac{t}{n})} dw,$$

The first is the usual spectral decomposition, with elements of S_X in $L^1[-n\pi, n\pi]$. The second is the classical Fourier transform L^1 - L^2 pair relation, with elements of \tilde{R}_X in $L^2[-n\pi, n\pi]$. Since these elements will also be in $L^1[-n\pi, n\pi]$, we have two $L^1[-n\pi, n\pi]$ functions with the same Fourier coefficients. Hence, $\tilde{R}_X(w) = S_X(w)$ a.e., by the classical uniqueness theorem for $L^1[-n\pi, n\pi]$ functions.

12/ This is especially so if an abstract point of view is taken toward (10), in which R_X is viewed as an \mathcal{L}_2 -operator which is to take \mathcal{L}_2 vectors to \mathcal{L}_2 vectors. Thus stated, we might seek conditions on R_X that it and its inverse have this property. My conjecture is that essential boundedness is the desired condition.

13/ We have had in mind the usual diagonalization. Choose U , unitary, such that $U'S_X U = \Lambda$. Then $U'S_X^{-1} U = \Lambda^{-1}$, and $U'S_X^{-1} S_X^{-1} U = \Lambda^{-1} \Lambda^{-1}$. But Λ is real because S_X is Hermitian, so that taking traces, permuting u' , and using $uu' = I$ yields $\text{tr } S_X^{-1} S_X^{-1} = \sum_{i=1}^N \lambda_i^{-2}$. Also, using the positive definite property of S_X ($S_X > 0$ for short) and of S_X^{-1} ,

$$R_{xy}(\cdot) \in L^1(-\infty, \infty)$$

as well as that $S_{xy}(\lambda)$ exists. This follows from

$$F_{xy}(\lambda+\epsilon) - F_{xy}(\lambda) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T R_{xy}(t) \frac{e^{-i(\lambda+\epsilon)t} - e^{-i\lambda t}}{-it} dt. \quad ([5], p. 41)$$

Dividing by ϵ and passing to the limit, we get the desired result by pulling the limit on ϵ inside the integral sign, which is permissible by the Dominated Convergence Theorem and uses the integrability of $R_{xy}(\cdot)$.

Of course, all integrals involving matrix-valued integrands are to be interpreted component-wise. Finally, we note that that a bounded function ($|R_{xy}(\cdot)|$ is bounded by $|R_{xy}(0)|$) in $L^1(-\infty, \infty)$ is also in $L^2(-\infty, \infty)$ and that when $R_{xy}(\cdot) \in L^1(-\infty, \infty)$, by the preceding inversion formula, $|S_{xy}(\cdot)|$ is bounded. Hence both $R_{xy}(\cdot)$ and $S_{xy}(\cdot)$ are in $L^2(-\infty, \infty)$, and may thus be regarded as Fourier transform pairs under the classical unitary mapping of this Hilbert space onto itself ([5], p. 513.) Finally, we will use the symbols $R_x(\cdot)$ and $R_{xx}(\cdot)$ as well as $S_x(\cdot)$ and $S_{xx}(\cdot)$ interchangeably.

4/ This tact was taken because of precedent ([3], [10]) and because it is a natural way to proceed.

5/ While Lighthill [7] is a standard reference which gives a good "physical" motivation for this subject, he never mentions the convolution of distributions. This topic is crucial for our purposes, however, as we wish to be able to Fourier transform (3), say, to $S_{xy} = S_{xx} \cdot b$ for a wide variety of ordinary and generalized $R_{xx}(\cdot)$ and $b(\cdot)$ functions. We refer the interested reader to [13] or [14] for justification of any underqualified use of such procedures. Briefly, ordinary functions can be embedded in a suitable space of generalized functions. The latter are defined so as always to permit Fourier and inverse Fourier transformation. In this way, we can operate with ordinary functions which don't meet the "classical" conditions for Fourier representations. For example,

$$\sum_{t=-\infty}^{t=\infty} e^{iwt} = \delta(w),$$

the Dirac delta function, where convergence of the infinite series is interpreted in the distribution sense. Indeed, the series doesn't converge in the "classical" modes (pointwise, $L^2(-\infty, \infty)$, Cesaro, etc.).

6/ A related way that generalized functions or distributions might enter is if we extended our considerations to generalized random processes. Continuous time white noise is an example; its correlation function is the Dirac delta (generalized) function of the previous footnote.

if $S_X < c \cdot I_N$, then (i) $S_X^{-1} \equiv S_X^{-1} > c^{-1} \cdot I_N$, and (ii) $S_X^{-1} S_X^{-1} > S_X^{-1} \cdot c^{-1} \cdot I_N > c^{-2} I_N$. (ii) follows from (i) and the result that "A > B > 0, C > 0; C commutes with A and B implies AC > BC." Facts (i) and (ii) may be found in Halmos [4], p. 167-8.

14/ Recall that the artifice of the generalized function $b(t) = \delta_\alpha(t)$ allowed us to "sift" $x(t-\alpha)$ from $\int_{-\infty}^{\infty} b(t-s)x(s)ds$. Had $b(t)$ been the derivative of a delta function, $x'(t-\alpha)$ could have been realized, if the x process were mean square differentiable. These $b(\cdot)$ "functions" may be thought of as the inclusion of "ideal elements": limits of sequences $\{b_n\}$ of ordinary functions $b_n(\cdot)$ which become more concentrated around α as n increases [7]. In the discrete model there again may be elements of H_X which are "ideal" in the sense that they are not representable as convolutions, but only as limits of convolutions. It is unclear how the concept of generalized function would be useful in describing these limiting elements.

15/ Most texts on time series treat this topic under the heading "aliasing." To describe the use of the term "folding operator," imagine the graph of a (symmetric) spectral-density function, drawn on both sides of a piece of an (infinitely long) piece of paper. Fold the paper, so as to make an accordion, with pleats at $n\pi k$, k integer. Compress the positive half of accordion; and superimpose (add vertically) to make the spectrum on $[0, n\pi]$.

16/ Strictly speaking, since S_X and S_{-X} differ on a countable set, they are equal almost everywhere, and by the usual identification of such matrix functions, were considered to be the same matrix function all along.

17/ Technically, the elements of the perturbation matrix, here, $P(w)$, must be measurable.

18/ Again by footnote 16, there is no loss of generality by forcing our perturbation matrices to be zero on any set of measure zero, which is the effect of not allowing λ to be in $N_{0; \frac{n}{2}}$.

19/ Of course S_Y may be observed, but since $S_Y = S_X |B|^2 + S_U$ and nothing is known about S_U , S_Y provides no additional information.

20/ In [10] and [3], $r_x(t) \equiv R_X^{-1} * R_x(t)$ and $\tilde{r}_x(w) \equiv S_X^{-1} S_x(w)$ are defined. Under the "linear interpolation" assumption on R_x , in [10] it is shown that $r_x(t) = r_0(t) \equiv \begin{cases} 1-|u| & |u| \leq 1 \\ 0 & |u| > 1 \end{cases}$ and it is emphasized that this filter has the desirable properties of having its integral equal unity and vanishing off an interval $([-1, 1])$ around the origin. In [3], the further desirable property of "no contamination," or, no confounding of different components of the b vector in the B vector of (14), is proved under this assumption.

21/ Despite the notation we emphasize that \tilde{B} and \tilde{b} are the primitive concepts here.

22/ We refer the reader to Rozanov, Theorem 2.4, p. 115 and Theorem 6.4, p. 27 for precise statements and proofs of these results.

$$23/ \int_{-\infty}^{\infty} \frac{\sin t \pi}{t \pi} e^{-itw} dt = \chi_{[-n\pi, n\pi]}^{(w)}$$
 is not straight-forward

without a handbook of definite integrals, since the computation

$$\int_0^{\infty} \sin t \pi \cos tw dt$$
 is needed. Actually, the integral equals $\frac{1}{2} \begin{cases} |w| > n\pi \\ |w| < n\pi \end{cases} = n\pi$

which is $\chi_{[-n\pi, n\pi]}^{(\cdot)}$ except at the endpoints. Of course, at $t=0$, $\frac{\sin t \pi}{t \pi}$ is defined (by continuity) as n .

24/ Spectral characteristics being S_X -unique means that two such elements, $\tilde{b}_1(w)$ and $\tilde{b}_2(w)$, are identified whenever

$$\int_{-\infty}^{\infty} (\tilde{b}_1 - \tilde{b}_2)' S_X (\tilde{b}_1 - \tilde{b}_2) dw = 0.$$

Where S_X vanishes, the equivalence class of such identified elements is large. Hypothesis (a) tells us which element to select.

25/ We depart slightly from tradition here by regarding cross-spectra as defined on $[0, 2\pi_1]$ and $[0, 2\pi_i]$ rather than on $[-\pi_1, \pi_1]$ and $[-\pi_i, \pi_i]$ because the limits in the finite summations are more tractable. Of course since our functions are of the required periodicity, they may be extended back to the interval which is symmetric about the origin, if desired.

26/ The left-hand side vector $S_{XY}(w)$ is, respectively, of dimension $n \times 1$, $(\sum_{i=1}^N n_i) \times 1$, and $N \times 1$ in equations (32), (30), and (33), respectively. The meaning of the terms $S_{X;B}$, $S_{X;D}$, \tilde{B}_f , and \tilde{B}_c should be clear, but in any event is indicated explicitly in the matrix displays.

27/ The pedantic reader will note that the same symbol, \tilde{B} , is used in (30), where its domain is $[0, 2\pi]$ and it is $(\sum_{i=1}^N n_i) \times 1$, and in (27), where its domain is $[0, 2\pi_1]$ and it is $N \times 1$. Not making this distinction, which causes no real difficulty, reflects itself in the minor ambiguity regarding the domain of $\| \cdot \|_2$.

References

- [1] Bachman, G. and Narici, L., Functional Analysis, Academic Press, New York, 1966.
- [2] Fishman, G. S. Spectral Methods in Econometrics, Harvard Univ. Press, Cambridge, Massachusetts, 1969.
- [3] Geweke, J. G., "Temporal Aggregation in the Multivariate Regression Model," Discussion Paper No. 74-50, University of Minnesota, Minneapolis, Minnesota, November 1974.
- [4] Halmos, Finite Dimensional Vector Spaces (2nd Ed.), Princeton University Press, Princeton, New Jersey, 1947.
- [5] Hamman, E. J., Multiple Time Series, Wiley, New York, 1970.
- [6] Koopmans, L. H., The Spectral Analysis of Time Series, Academic Press, New York, 1974.
- [7] Lighthill, M. J., Introduction to Fourier Analysis and Generalized Functions, The Cambridge University Press, London, 1964.
- [8] Masani, P. and Wiener, W., "The Prediction Theory of Multivariate Stochastic Processes I and II," Acta. Math. 98 and 99; 111-150 and 93-137; 1958 and 1959.
- [9] Rozanov, Yu. A., Stationary Random Processes, Holden-Day, San Francisco, California, 1967.
- [10] Sims, C. A., "Discrete Approximations to Continuous Time Distributed Lags in Econometrics," Econometrica 39, (1971), pp. 545-563.
- [11] Sims, C. A. "Money, Income, and Causality," The American Economic Review, Vol. 62, No. 4, September 1972, pp. 540-552.
- [12] Whittle, P. Prediction and Regulation, Van Nostrand-Reinhold, Princeton, New Jersey, 1963.
- [13] Yosida, K., Functional Analysis, Berlin, Springer-Verlag, 1965.
- [14] Zemanian, A. H., Distribution Theory and Transform Analysis: An Introduction to Generalized Functions, with Applications, McGraw-Hill, New York, 1965.