The Annals of Statistics
1975, Vol. 3, No. 3, 688–697

CONTINUOUS TIME REGRESSIONS WITH DISCRETE DATA

BY P. M. ROBINSON

Harvard University

A general continuous time distributed lag model is considered. The problem is that of estimating the parameters of the kernel when, as is often the case, the available data consist not of a continuous record but of discrete observations recorded at regular intervals of time. Fourier transformation of the model and insertion of the computable, discrete Fourier transforms of the variables produce an approximate model which is of non-linear regression type and is relatively easy to handle. Estimators are proposed and their asymptotic properties established, assuming principally that the variables are stationary and ergodic and that an "aliasing" condition on the independent variable is satisfied. The results of the paper imply a theory for the estimation of rather general continuous time systems, involving the operations of differentiation, integration and translation through time.

1. Introduction. The estimation of a relationship between stochastic processes that are of an underlying continuous nature will often proceed on the basis of a sample that is not continuous but discrete. Continuous observation is out of the question in the social sciences, and although in some natural sciences electrical or optical equipment can produce continuous measurements, the record will possess limited resolution and in many cases it may be feasible to obtain a discrete subsequence that embodies all the information. For convenience of both data collection and data analysis a sample of observations recorded at equal intervals of time (or whatever the dimension that is concerned) is often the most desirable. We shall be interested in circumstances in which we have available the sample

$$y(1), y(2), \cdots, y(N), z(1), z(2), \cdots, z(N)$$

from the continuous-time processes $y(t)$, $z(t)$, t indexing the processes over the entire real line, $\mathscr{R}$. (Our work is relevant also to circumstance in which $y(t)$ and $z(t)$ are not continuous but are observable more frequently than in the available record, the model relating the variables over only countably many time points.) For simplicity of exposition we have taken the sampling interval to be unity. We believe in the existence of the relationship

$$y(t) = \beta_0 \int_{\mathscr{R}} \gamma(\tau; \boldsymbol{\alpha}_0) z(t-\tau)\, d\tau + x(t)\,, \qquad t \in \mathscr{R}. \tag{1.1}$$

Here, $x(t)$ is a residual process and $y(t)$ and $z(t)$ are taken to represent, respectively, dependent and independent variables. By $\boldsymbol{\alpha}$ we mean a $1 \times p$ row vector

Received November 1972; revised June 1974.
AMS 1970 *subject classifications*. Primary 62E20, 62M10; Secondary 42A68, 62M15.
Key words and phrases. Stationary processes, non-linear regressions, Fourier methods, asymptotic theory.

of parameters (we use bold-face to denote vectors and matrices) and by β a scalar parameter, the zero subscript being appended in (1.1) to indicate that we are referring to the true parameter values rather than to any admissible values. The function or generalized function $\gamma(t; \boldsymbol{\alpha})$ is possibly non-linear in $(t, \boldsymbol{\alpha})$. In practice (1.1) will often be a solution of a system that arises more naturally from a priori considerations and involves $y(t)$ in a more complicated way. We note that $\mathscr{R}$ is taken as the range of integration, but when causation is implied, that is when $y(t)$ depends on $z(\tau)$ only for $\tau \leqq t$, we would have $\gamma(t; \boldsymbol{\alpha}) = 0$, $t > 0$. However, the direction of causation may be in doubt (and our methods may provide information on this question) so it seems worthwhile to allow for leads as well as lags.

In the past, (1.1) has usually been replaced by

$$y(n) = \sum_{j=-\infty}^{\infty} \beta_0(j) z(n-j) + x(n) , \qquad n = 0, \pm 1, \cdots . \tag{1.2}$$

The estimation of (1.2) is a relatively straightforward task, although of course the dimension of the parameter space must be finite, as is the case when the sum is truncated or when the $\beta_0(j)$ are functions of a finite number of parameters, for example when (1.2) is the solution of a linear difference equation. Sims [12] has shown that if a $y(t)$ generated by (1.1) is sampled at unit intervals, then, in a second-order sense, the discrete sequence is generated also by a model of the form (1.2). However, in general (1.1) and (1.2) are not isomorphic because of the impossibility of interpolating a continuum on the basis of merely a discrete equal-spaced sequence. It seems implausible to include only integral lags in a relationship between continuous processes, and undesirable to limit model specification in this way. We shall thus consider the estimation of (1.1) rather than (1.2). The method we propose and the asymptotic theory extend the treatment by Hannan and Robinson [6] of a lagged regression model that is a special case of (1.1), in which $\gamma(t; \boldsymbol{\alpha}) = \delta(t - \theta)$, where $\delta(t)$ is the Dirac delta function, so positive weight is given only to the lag θ. In the following section we suggest a procedure for estimating $\boldsymbol{\alpha}_0$ and β_0 and we establish theorems that confer desirable properties on our estimators under suitable circumstances.

2. The estimation procedure. The presence of a convolution integral in (1.1) leads us to consider its Fourier representation. In the mean square sense, (1.1) is equivalent to

$$\int_{\mathscr{R}} e^{-it\lambda}\{d\chi_y(\lambda) - \beta_0 \tilde{\gamma}(\lambda; \boldsymbol{\alpha}_0)\, d\chi_z(\lambda) - d\chi_x(\lambda)\} = 0 , \qquad t \in \mathscr{R} .$$

Here it is implied that our processes are all mean square continuous (see Bartlett [1], page 138) and stationary with spectral representations typified by

$$x(t) = \int_{\mathscr{R}} e^{-it\lambda}\, d\chi_x(\lambda) , \qquad t \in \mathscr{R} ,$$

the stationary process $\chi_x(\lambda)$ having orthogonal increments. Also we assume that the transform

$$\tilde{\gamma}(\lambda; \boldsymbol{\alpha}) = \int_{\mathscr{R}} e^{it\lambda} \gamma(t; \boldsymbol{\alpha})\, dt , \qquad \lambda \in \mathscr{R} , \tag{2.1}$$

exists and is a known function of λ and $\boldsymbol{\alpha}$. Otherwise our methods cannot be used. Of course $\tilde{\gamma}$ is uniquely defined by γ, and numerous transform pairs are tabulated in the Fourier transform and control theory literature. (See e.g., Kaplan [8].) In fact when (1.1) is merely the solution of a more natural equation, $\tilde{\gamma}$ is the more basic function and it may not be possible to derive γ in an explicit, closed form from knowledge of $\tilde{\gamma}$. This is of no concern, however, since our computing formulas involve $\tilde{\gamma}$ rather than γ. We may describe γ and $\tilde{\gamma}$ as, respectively, the kernel and frequency response function of the filter. In (1.2) the parameter set could be reduced by taking the $\beta_0(j)$ to be ordinates of a polynomial or of a discrete frequency distribution. It may then be of interest to consider the "moments" of the "lag distribution". We can do the same here by considering, for $r = 1, 2, \cdots$,

$$\mathscr{E}\{t^r\} = [\textstyle\int_{\mathscr{R}} t^r \gamma(t; \boldsymbol{\alpha}_0)\, dt]/[\int_{\mathscr{R}} \gamma(t; \boldsymbol{\alpha}_0)\, dt] = \left\{ i^{-r} \frac{\partial^r}{\partial \lambda^r} \tilde{\gamma}(\lambda; \boldsymbol{\alpha}_0)\Big|_{\lambda=0} \right\} \Big/ \tilde{\gamma}(0, \boldsymbol{\alpha}) .$$

The increments $d\chi_x$, $d\chi_y$ and $d\chi_z$ are uniquely defined by knowledge of $x(t)$, $y(t)$ and $z(t)$, $t \in \mathscr{R}$, but in the absence of such knowledge we propose to replace them by the (normed) discrete Fourier transforms $w_x(s)$, $w_y(s)$ and $w_z(s)$, whose definitions are typified by

$$w_x(s) = (2\pi N)^{-\frac{1}{2}} \textstyle\sum_{n=1}^{N} x(n) \exp(in\lambda_s) , \qquad \lambda_s = \frac{2\pi s}{N} , \qquad -\frac{1}{2}N < s \leqq [\frac{1}{2}N] .$$

Then we consider the approximate model

$$w_y(s) = \beta_0 \tilde{\gamma}(\lambda_s; \boldsymbol{\alpha}_0) w_z(s) + w_x(s) . \tag{2.2}$$

Because $w_y(s)$ and $w_z(s)$ are computable and $\tilde{\gamma}$ is a given function, (2.2) is basically of regression type and is relatively easy to handle, although the estimation must generally rely on numerical methods. Now whereas the sample provides information on the discrete sequences over the frequency band $(-\pi, \pi]$ we may not believe (1.1) to be a valid model, or the aliasing condition (v) below to be reasonable, over the whole of this band. Thus, following [6], we consider a set $\mathscr{B} \subset (-\pi, \pi)$, composed of a finite number of disjoint open intervals that are symmetric about $\lambda = 0$, so that if $\lambda \in \mathscr{B}$, $-\lambda \in \mathscr{B}$ also. Then we estimate $\boldsymbol{\alpha}_0$ and β_0 by absolutely minimizing

$$Q_N(\boldsymbol{\alpha}, \beta) = N^{-1} \textstyle\sum_{\mathscr{B}} |w_y(s) - \beta \tilde{\gamma}(\lambda_s; \boldsymbol{\alpha}) w_z(s)|^2 \phi(\lambda_s) \tag{2.3}$$

over all admissible $\boldsymbol{\alpha}$, β, where the sum is over $\lambda_s \in \mathscr{B}$ and $\phi(\lambda)$ is a given real function that is positive, even and continuous in λ over $\mathscr{B}$. For convenience, we assume in our theorems that all our processes have zero means. In practice, mean correction can be accomplished by omitting from (2.3) the component for $\lambda_s = 0$, which involves the deviation

$$w_y(0) - \beta \tilde{\gamma}(0, \boldsymbol{\alpha}) w_z(0) = (N/2\pi)^{\frac{1}{2}} (\bar{y} - \beta \tilde{\gamma}(0; \boldsymbol{\alpha}) \bar{z}) .$$

3. Asymptotic theory. The justification for our estimation procedure is suggested by the following theorems.

THEOREM 1. *Let $\hat{\boldsymbol{\alpha}}$ and $\hat{b}$ minimize* (2.3) *over $\boldsymbol{\alpha} \in \mathscr{A}$. Let the following conditions hold.*

(i) (1.1) *is true.*

(ii) *$\mathscr{A}$ is a compact subset of $\mathscr{R}^p$ and $\boldsymbol{\alpha}_0 \in \mathscr{A}$.*

(iii) *Uniformly in $\lambda \in \mathscr{R}$, $\boldsymbol{\alpha} \in \mathscr{A}$, $\tilde{\gamma}(\lambda; \boldsymbol{\alpha})$ exists and is continuous.*

(iv) *$x(n)$ and $z(n)$ are mutually incoherent, strictly stationary and ergodic sequences with zero means, absolutely continuous spectral distribution functions and continuous spectral densities*

$$f_x^{\mathscr{A}}(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \mathscr{E}\{x(n)x(n+j)\}e^{-ij\lambda}\,, \qquad \lambda \in (-\pi, \pi]\,,$$

$$f_z^{\mathscr{A}}(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \mathscr{E}\{z(n)z(n+j)\}e^{-ij\lambda}\,, \qquad \lambda \in (-\pi, \pi]\,,$$

respectively.

(v) *For $\lambda \in \mathscr{B}$ and $j = \pm 1, \pm 2, \cdots$,*

$$f_z^{c}(\lambda + 2\pi j) = 0$$

where $f_z^c(\lambda)$ is the spectral density of the continuous stationary process $z(t)$ given by

$$f_z^c(\lambda) = \frac{1}{2\pi} \int_{\mathscr{R}} \mathscr{E}\{z(t)z(t+\tau)\}e^{-i\lambda\tau}\, d\tau\,, \qquad \lambda \in \mathscr{R}\,.$$

(vi) *$(\boldsymbol{\alpha}_0, \beta_0)$ is the only admissible $(\boldsymbol{\alpha}, \beta)$ that is a zero of*

$$(3.1) \qquad (2\pi)^{-1} \int_{\mathscr{B}} |\beta\tilde{\gamma}(\lambda; \boldsymbol{\alpha}) - \beta_0\tilde{\gamma}(\lambda; \boldsymbol{\alpha}_0)|^2 f_z^{\mathscr{A}}(\lambda)\phi(\lambda)\, d\lambda\,.$$

Then $\lim_{N\to\infty} (\hat{\boldsymbol{\alpha}}, \hat{\beta}) = (\boldsymbol{\alpha}_0, \beta_0)$, *almost surely* (a.s.).

PROOF. The proof is similar to that of Theorem 6 of Jennrich [7], Theorem 3 of Hannan [5] and Theorem 2 of Hannan and Robinson [6], and is thus, like that of Theorem 2 below, somewhat abbreviated. We have

$$Q_N(\boldsymbol{\alpha}, \beta) = N^{-1} \sum_{\mathscr{B}} |w_y(s)|^2\phi(\lambda_s) - 2\beta\hat{a}(\boldsymbol{\alpha}) + \beta^2\hat{b}(\boldsymbol{\alpha})\,,$$

where

$$\hat{a}(\boldsymbol{\alpha}) = \frac{1}{N} \sum_{\mathscr{B}} \tilde{\gamma}(\lambda_s; \boldsymbol{\alpha})w_z(s)\overline{w_y}(s)\phi(\lambda_s)\,, \qquad \hat{b}(\boldsymbol{\alpha}) = \frac{1}{N} \sum_{\mathscr{B}} |\tilde{\gamma}(\lambda_s; \boldsymbol{\alpha})w_z(s)|^2\phi(\lambda_s)\,,$$

$\hat{a}(\boldsymbol{\alpha})$ being real because $\mathscr{B}$ is symmetric and ϕ is even. Now under (i)—(iv) we have

$$\lim_{N\to\infty} \hat{a}(\boldsymbol{\alpha}) = a(\boldsymbol{\alpha})\,, \qquad \lim_{N\to\infty} \hat{b}(\boldsymbol{\alpha}) = b(\boldsymbol{\alpha})\,, \quad \text{a.s.,}$$

uniformly in $\boldsymbol{\alpha} \in \mathscr{A}$, where

$$a(\boldsymbol{\alpha}) = \frac{\beta_0}{2\pi} \int_{\mathscr{B}} \tilde{\gamma}(\lambda; \boldsymbol{\alpha})\tilde{\gamma}(-\lambda; \boldsymbol{\alpha}_0)f_z^{\mathscr{A}}(\lambda)\phi(\lambda)\, d\lambda\,,$$

$$b(\boldsymbol{\alpha}) = \frac{1}{2\pi} \int_{\mathscr{B}} |\tilde{\gamma}(\lambda; \boldsymbol{\alpha})|^2 f_z^{\mathscr{A}}(\lambda)\phi(\lambda)\, d\lambda\,.$$

for pointwise convergence follows from the lemma in [6], because $\tilde{\gamma}$ and ϕ are continuous and, for $\lambda \in \mathscr{B}$,

$$f_z^{\mathscr{A}}(\lambda) = \sum_{j=-\infty}^{\infty} f_z^{c}(\lambda + 2\pi j) = f_z^{c}(\lambda) ,$$
$$f_{yz}^{\mathscr{A}}(\lambda) = \sum_{j=-\infty}^{\infty} \tilde{\gamma}(\lambda + 2\pi j; \boldsymbol{\alpha}_0) f_z^{c}(\lambda + 2\pi j) = \tilde{\gamma}(\lambda; \boldsymbol{\alpha}_0) f_z^{c}(\lambda)$$

under (v), $f_{yz}^{\mathscr{A}}$ being the cross-spectral density of $y(n)$, $z(n)$. Uniformity of convergence under (ii) follows from [7], Theorem 1. Now $\hat{\boldsymbol{\alpha}}$ as defined must maximize $\hat{a}(\boldsymbol{\alpha})^2/\hat{b}(\boldsymbol{\alpha})$ over $\mathscr{A}$, with $\hat{\beta} = \hat{a}(\boldsymbol{\alpha})/\hat{b}(\hat{\boldsymbol{\alpha}})$, (cf., Golub and Pereyra [2], Theorem 2.1). But (3.1) is

$$\beta^2 b(\boldsymbol{\alpha}) - 2\beta a(\boldsymbol{\alpha}) + a(\boldsymbol{\alpha}_0)^2/b(\boldsymbol{\alpha}_0) = a(\boldsymbol{\alpha}_0)^2/b(\boldsymbol{\alpha}_0) - a(\boldsymbol{\alpha})^2/b(\boldsymbol{\alpha}) ,$$

on putting $\beta = a(\boldsymbol{\alpha})/b(\boldsymbol{\alpha})$ and noting that (vi) with $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ implies $b(\boldsymbol{\alpha}_0) \neq 0$. Because this expression is positive under (vi), uniformly in admissible $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}_0$, the consistency of $\hat{\alpha}$, and thence of $\hat{\beta}$, follows from the type of argument used in [7], Theorem 6.

Note that we could prove consistency under conditions similar to those used by Grenander and Rosenblatt [3], pages 233–235, allowing $y(n)$ and $z(n)$ to be stochastic processes that are "slowly increasing", and thus normalizing Q_N differently and defining the spectral densities as transforms of autocorrelations. The condition deserving most comment is (v), however, which identifies the model to within equivalence classes agreeing over $\mathscr{B}$, the identification being completed by (vi). We define by $x_{\mathscr{B}}$, $y_{\mathscr{B}}$, $z_{\mathscr{B}}$ the outputs of ideal multiple bandpass filters which pass only the frequency components of $x(t)$, $y(t)$, $z(t)$ that are in $\mathscr{B}$. Thus, $z_{\mathscr{B}}$ is band-limited below Nyqvist frequency and a unique interpolation between data points is implied. Condition (v) weakens the condition

$$f_z^{c}(\lambda) = 0 , \qquad |\lambda| \geqq \pi + \delta , \quad 0 \leqq \delta \leqq \pi , \tag{3.2}$$

used in [6]. Spectral densities satisfying (v) but not (3.2) do not occur in practice, even after data transformation. However, both assumptions will nearly always be only approximations and there will be cases which make the weakening seem worthwhile, a fairly important one being the following. We assume

$$z(t) = \int_{t-1}^{t} v(u)\, du , \qquad t \in \mathscr{R} ,$$

where $v(t)$ is a strictly stationary process having spectral density f_v^{c} (which need not be integrable). Then

$$f_z^{c}(\lambda) = (2 \sin \tfrac{1}{2}\lambda/\lambda)^2 f_v^{c}(\lambda) , \qquad \lambda \in \mathscr{R} .$$

If f_v^{c} is differentiable within a neighborhood of $\lambda = 2\pi j$, uniformly in integral $j > 1$, f_z^{c} has a double zero and an analytic minimum at each of these points, whereas it may be substantial at the center of the first few "lobes". Thus, if $\mathscr{B} = \{\lambda \,|\, |\lambda| < \pi - \delta\}$, (v) seems reasonable for suitably large δ, while (3.2) may not be. In general, (v) seems reasonable if the bands where the density is "zero" contribute a relatively small amount to the total power, the amount possibly becoming negligible as sample size increases when data transformation is involved.

On the other hand, no deductions about the robustness of $\hat{\boldsymbol{\alpha}}$ and $\hat{\beta}$ with respect to even small departures from (v) can be made. We note finally that when $\gamma(t; \boldsymbol{\alpha}_0) = 0$ uniformly in irrational t, $\tilde{\gamma}$ is periodic and so (v) may be weakened (see Robinson [10]).

We consider now an asymptotic distribution theory for $\hat{\boldsymbol{\alpha}}$, $\hat{\beta}$. For a matrix function $\mathrm{A}(\boldsymbol{\theta})$ of a vector $\boldsymbol{\theta}$, we define $\boldsymbol{\theta}_0$ to be a regular point of $\mathrm{A}(\boldsymbol{\theta})$ if $\mathrm{A}(\boldsymbol{\theta})$ is continuous and has constant rank within a neighborhood of $\boldsymbol{\theta}_0$. (If $\mathrm{A}(\boldsymbol{\theta})$ is analytic the set of points that are not regular has measure zero.)

THEOREM 2. *Let the conditions of Theorem* 1 *and the following conditions hold.*

(vii) *There exists a positive function* $\psi(t)$, *such that* $\psi(t) \to 0$ *as* $t \to \infty$ *and, for* $t > \tau$,

$$\psi(t - \tau) > \sup_{A \in \mathscr{F}_{-\infty}^{\tau}, B \in \mathscr{F}_t^{\infty}} |\Pr(A \cap B) - \Pr(A)\Pr(B)| ,$$

where $\mathscr{F}_a^{\,b}$ *denotes the* σ*-field of events generated by* $x(t)$ *for* $a \leqq t \leqq b$ *only. Also*

$$\sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sum_{n_3=0}^{\infty} |\kappa(0, n_1, n_2, n_3)| < \infty ,$$

the summand being the absolute value of the fourth cumulant of $x(n)$, $x(n + n_1)$ $x(n + n_2)$ *and* $x(n + n_3)$, n, n_1, n_2 *and* n_3 *being integers.*

(viii) *Uniformly in* $t_j \in \mathscr{R}$, $j = 1, 2, 3, 4$, *the fourth cumulant of* $z(t_1)$, $z(t_2)$, $z(t_3)$ *and* $z(z_4)$ *is finite and given by*

$$\underset{\sum \lambda_j = 0}{\int\!\!\int\!\!\int\!\!\int} f(\lambda_1, \lambda_2 \lambda_3, \lambda_4) \exp(i \textstyle\sum t_j \lambda_j) \prod d\lambda_j ,$$

where f *is continuous over* $\sum \lambda_j = 0$.

(ix) $f_z^{\,c}(\lambda)$ *is continuous over* $\mathscr{R}$. *There exists a set* $\mathscr{C}$, $\mathscr{B} \subset \mathscr{C} \subset (-\pi, \pi)$, *such that the elements of* $\mathscr{B}$ *are bounded away from those of* $(-\pi, \pi) - \mathscr{C}$ *and, for* $\lambda \in \mathscr{C}$ *and* $j = \pm 1, \pm 2, \cdots$,

$$f_z^{\,c}(\lambda + 2\pi j) = 0 .$$

There exists a $K < \infty$ *such that* $f_z^{\,c}(\lambda) = 0$, $|\lambda| > K$.

(x) $\boldsymbol{\alpha}_0$ *is an interior point of* $\mathscr{A}$.

(xi) $\tilde{\gamma}(\lambda; \boldsymbol{\alpha}_0)$ *satisfies a Lipschitz condition of order greater than one half in* λ *over* $(-\pi, \pi)$.

(xii) *The first and second derivatives of* $\tilde{\gamma}(\lambda; \boldsymbol{\alpha})$ *exist and are continuous in* $\boldsymbol{\alpha}$ *and* λ *within a neighborhood of* $\boldsymbol{\alpha}_0$, *and* $(\boldsymbol{\alpha}_0; \beta_0)$ *is a regular point of*

$$\Psi\{\boldsymbol{\alpha}, \beta; \phi\} = \frac{1}{2\pi} \int_{\mathscr{B}} \begin{bmatrix} \dfrac{\beta\partial}{\partial \boldsymbol{\alpha}'} \tilde{\gamma}(\lambda; \boldsymbol{\alpha}) \\ \cdots\cdots \\ \tilde{\gamma}(\lambda; \boldsymbol{\alpha}) \end{bmatrix} \begin{bmatrix} \dfrac{\beta\partial}{\partial \boldsymbol{\alpha}'} \tilde{\gamma}(-\lambda; \boldsymbol{\alpha}) \\ \cdots\cdots \\ \tilde{\gamma}(-\lambda; \boldsymbol{\alpha}) \end{bmatrix}' f_z^{\mathscr{A}}(\lambda)\phi(\lambda)\, d\lambda .$$

Then as $N \to \infty$ $N^{\frac{1}{2}}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \,\vdots\, \hat{\beta} - \beta_0)$ *converges to a multivariate normal vector with null mean and covariance matrix*

$$\Psi\{\boldsymbol{\alpha}_0, \beta_0; \phi\}^{-1} \Psi\{\boldsymbol{\alpha}_0, \beta_0; \phi^2 f_x^{\mathscr{A}}\} \Psi\{\boldsymbol{\alpha}_0, \beta_0; \phi\}^{-1} . \tag{3.3}$$

PROOF. The basic method of proof is that of Theorem 7 of [7], Theorem 4 of [5] and Theorem 3 of [6], involving the construction of mean-value theorem relations, under (x) with N sufficiently large, for

$$(3.4)\qquad \frac{-N^{\frac{1}{2}}}{2}\left\{\frac{\partial}{\partial\hat{\boldsymbol{a}}} Q_N(\hat{\boldsymbol{a}}, \hat{\beta}) - \frac{\partial}{\partial\boldsymbol{\alpha}_0} Q_N(\boldsymbol{\alpha}_0, \beta_0), \right.$$

$$\left. \frac{\partial}{\partial\hat{b}} Q_N(\hat{\boldsymbol{a}}, \hat{\beta}) - \frac{\partial}{\partial\beta_0} Q_N(\boldsymbol{\alpha}_0\, \beta_0)\right\} .$$

Considering the final element, which is notationally the easiest to handle, we must prove in particular the asymptotic normality of

$$-\frac{N^{\frac{1}{2}}}{2}\frac{\partial}{\partial\beta_0} Q_N(\boldsymbol{\alpha}_0, \beta_0) = N^{-\frac{1}{2}} \textstyle\sum_{\mathscr{B}} w_y(s)\overline{w_z}(s)\tilde{\gamma}(-\lambda_s; \boldsymbol{\alpha}_0)\phi(\lambda_s)$$

$$(3.5)\qquad -N^{-\frac{1}{2}}\beta_0 \textstyle\sum_{\mathscr{B}} |\tilde{\gamma}(\lambda_s; \boldsymbol{\alpha}_0)w_z(s)|^2\phi(\lambda_s)$$

$$= N^{-\frac{1}{2}} \textstyle\sum_{\mathscr{B}} w_x(s)\overline{w_z}(s)\tilde{\gamma}(-\lambda_s; \boldsymbol{\alpha}_0)\phi(\lambda_s)$$

$$(3.6)\qquad +\beta_0 N^{-\frac{1}{2}} \textstyle\sum_{\mathscr{B}} [\int_{\mathscr{R}} \gamma(\tau; \boldsymbol{\alpha}_0)w_\tau(s)\, d\tau - \tilde{\gamma}(\lambda_s; \boldsymbol{\alpha}_0)w_z(s)]\overline{w_z}(s)\tilde{\gamma}(-\lambda_s; \boldsymbol{\alpha}_0)\phi(\lambda_s) ,$$

where

$$w_\tau(s) = (2\pi N)^{-\frac{1}{2}} \textstyle\sum_{n=1}^N z(n-\tau)\exp(in\lambda_s) , \qquad \tau \in \mathscr{R} .$$

The asymptotic normality of (3.5) under (iii), (iv), (vii) may be established in a relatively straightforward way by modifying a central limit theorem in Hannan [4], Theorem 10′, page 227, for linear regression coefficients (see also [5], pages 776–777). The cited theorem in [4] relates to a discrete time model, but it applies here because only discrete $x(n)$ are involved in (3.5). The first part of (vii) is the strong mixing condition described in [4], pages 207–208 (and referred to there as the uniform mixing condition). No such condition is required on $z(t)$. Now, unless $\gamma(t; \boldsymbol{\alpha}_0) \propto \delta(t)$, (2.2) is not the exact discrete Fourier transform of (1.1) and (3.6) is not identically zero. But we can prove the convergence to zero of its mean square, which is $\beta_0{}^2$ times

$$(4\pi^2N^3)^{-1} \textstyle\sum' \mathscr{E}\{[[\int_{\mathscr{R}}\int_{\mathscr{R}} \gamma(\tau_1; \boldsymbol{\alpha}_0)\gamma(\tau_2; \boldsymbol{\alpha}_0)z(k-\tau_1)z(l-\tau_2)\, d\tau_1\, d\tau_2$$

$$- \tilde{\gamma}(-\lambda_t; \boldsymbol{\alpha}_0) \textstyle\int_{\mathscr{R}} \gamma(\tau; \boldsymbol{\alpha}_0)z(k-\tau)z(l)\, d\tau - \tilde{\gamma}(\lambda_s; \boldsymbol{\alpha}_0) \int_{\mathscr{R}} \gamma(\tau; \boldsymbol{\alpha}_0)z(k)z(l-\tau)\, d\tau$$

$$+ \tilde{\gamma}(\lambda_s; \boldsymbol{\alpha}_0)\tilde{\gamma}(-\lambda_t; \boldsymbol{\alpha}_0)z(k)z(l)]z(m)z(n)\}\tilde{\gamma}(-\lambda_s; \boldsymbol{\alpha}_0)\tilde{\gamma}(\lambda_t; \boldsymbol{\alpha}_0)\phi(\lambda_s)\phi(\lambda_t)$$

$$\times \exp\{i[(k-m)\lambda_s - (l-n)\lambda_t]\} ,$$

where we use the abbreviation

$$\textstyle\sum' = \sum_{\mathscr{B}} \sum_{\mathscr{B}} \sum_{k=1}^N \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N .$$

This expression, under (viii), is

$$(4\pi^2N^3)^{-1} \textstyle\sum' \iiiint_{\sum\lambda_j=0} [\tilde{\gamma}(-\lambda_1; \boldsymbol{\alpha}_0) - \tilde{\gamma}(\lambda_s; \boldsymbol{\alpha}_0)][\tilde{\gamma}(-\lambda_2; \boldsymbol{\alpha}_0) - \tilde{\gamma}(-\lambda_t; \boldsymbol{\alpha}_0)]$$

$$\times \tilde{\gamma}(-\lambda_s; \boldsymbol{\alpha}_0)\tilde{\gamma}(\lambda_t; \boldsymbol{\alpha}_0)\phi(\lambda_s)\phi(\lambda_t)[f_z{}^c(\lambda_1)f_z{}^c(\lambda_2)\delta(\lambda_1+\lambda_3)\delta(\lambda_2+\lambda_4)$$

$$+ f_z{}^c(\lambda_1)f_z{}^c(\lambda_2)\delta(\lambda_1+\lambda_4)\delta(\lambda_2+\lambda_3) + f_z{}^c(\lambda_1)f_z{}^c(\lambda_3)\delta(\lambda_1+\lambda_2)\delta(\lambda_3+\lambda_4)$$

$$+ f(\lambda_1, \lambda_2, \lambda_3, \lambda_4)] \exp\{i[k(\lambda_1+\lambda_s) + l(\lambda_2-\lambda_t) + m(\lambda_3-\lambda_s)$$

$$+ n(\lambda_4+\lambda_t)]\} \textstyle\prod d\lambda_j .$$

We consider first the term in $f_z^c(\lambda_1)f_z^c(\lambda_2)\delta(\lambda_1+\lambda_3)\delta(\lambda_2+\lambda_4)$, which corresponds to the squared mean of (3.6), and is

$$(4\pi^2N^3)^{-1}\sum{}' \int_{\mathscr{B}}\int_{\mathscr{B}}[\tilde{\gamma}(-\lambda_1;\boldsymbol{\alpha}_0)-\tilde{\gamma}(\lambda_s;\boldsymbol{\alpha}_0)][\tilde{\gamma}(-\lambda_2;\boldsymbol{\alpha}_0)-\tilde{\gamma}(-\lambda_t;\boldsymbol{\alpha}_0)]$$
$$\times\,\tilde{\gamma}(-\lambda_s;\boldsymbol{\alpha}_0)\tilde{\gamma}(\lambda_t;\boldsymbol{\alpha}_0)\phi(\lambda_s)\phi(\lambda_t)f_z^c(\lambda_1)f_z^c(\lambda_2)$$
$$\times\,\exp\{i[(k-m)(\lambda_1+\lambda_s)+(l-n)(\lambda_2-\lambda_t)]\}\,d\lambda_1\,d\lambda_2\,.$$

This is the square of

$$N^{-\frac{1}{2}}\sum_{\mathscr{B}}\tilde{\gamma}(-\lambda_s;\boldsymbol{\alpha}_0)\phi(\lambda_s)\int_{\mathscr{B}}[\tilde{\gamma}(\lambda;\boldsymbol{\alpha}_0)-\tilde{\gamma}(\lambda_s;\boldsymbol{\alpha}_0)]f_z^c(\lambda)F_N(\lambda-\lambda_s)\,d\lambda\,, \tag{3.7}$$

where

$$F_N(\lambda)=\frac{1}{2\pi N}\,|\textstyle\sum_{n=1}^N\exp(in\lambda)|^2$$

is Fejér's kernel. However,

$$N^{\frac{1}{2}}\int_{\mathscr{B}}[\tilde{\gamma}(\lambda;\boldsymbol{\alpha}_0)-\tilde{\gamma}(\lambda_s;\boldsymbol{\alpha}_0)]f_z^c(\lambda)F_N(\lambda-\lambda_s)\,d\lambda$$
$$=N^{\frac{1}{2}}\int_{-\pi}^{\pi}[\tilde{\gamma}(\lambda;\boldsymbol{\alpha}_0)-\tilde{\gamma}(\lambda_s;\boldsymbol{\alpha}_0)]f_z^c(\lambda)F_N(\lambda-\lambda_s)\,d\lambda$$
$$+\,N^{\frac{1}{2}}\sum_{1\leqq|j|\leqq J}\int_{(-\pi,\pi)-\mathscr{C}}[\tilde{\gamma}(\lambda+2\pi j;\boldsymbol{\alpha}_0)-\tilde{\gamma}(\lambda_s;\boldsymbol{\alpha}_0)]$$
$$\times\,f_z^c(\lambda+2\pi j)F_N(\lambda+2\pi j-\lambda_s)\,d\lambda\,,$$

$J<\infty$, because of the final part of (ix) (which will hold to a close uniform approximation for all integrable f_z^c). Under the first part of (ix), the first term on the right is not greater than

$$N^{\frac{1}{2}}\sup_{|\lambda|<\pi}[f_z^c(\lambda)]\int_{-\pi}^{\pi}|\tilde{\gamma}(\lambda;\boldsymbol{\alpha}_0)-\tilde{\gamma}(\lambda_s;\boldsymbol{\alpha}_0)|F_N(\lambda-\lambda_s)\,d\lambda=O(N^{\frac{1}{2}-\theta}) \tag{3.8}$$

if $\tilde{\gamma}(\lambda;\boldsymbol{\alpha}_0)$ satisfies a Lipschitz condition of order θ (Zygmund [13], page 91); since $\theta>\frac{1}{2}$ under (xi), (3.8) $\to 0$. (In practice, (xi) is certain to hold.) As far as the second term is concerned, we note from the definition of $\mathscr{C}$ that for each j, $1\leqq|j|\leqq J$, there exists an integer k such that, for all $\lambda_s\in\mathscr{B}$, $\lambda\in(-\pi,\pi)-\mathscr{C}$,

$$0<\Delta\leqq|\lambda+2\pi(j+k)-\lambda_s|\leqq\pi\,.$$

(The replacing of (v) by (ix) is not of practical concern in that $\mathscr{C}$ can be chosen arbitrarily close to $\mathscr{B}$.) Thus, under (ix), the second term converges to zero since $F_N(\lambda)=O(N^{-1}\operatorname{cosec}^2\Delta)$, $0<\Delta\leqq|\lambda|\leqq\pi$, F_N has period 2π, f_z^c is continuous and the number of summands is finite. Then, because $\tilde{\gamma}$, ϕ are continuous, (3.7) converges to zero. The remaining terms in the mean square of (3.6) are handled in a very different way but the proof that they converge to zero is basically the same as that of a result given in [6], although it is somewhat lengthier, so we shall not describe it in detail. However, it may be shown that each of the terms in $f_z^c(\lambda_1)f_z^c(\lambda_2)\delta(\lambda_1+\lambda_4)\delta(\lambda_2+\lambda_3)$ and in $f_z^c(\lambda_1)f_z^c(\lambda_3)\delta(\lambda_1+\lambda_2)\delta(\lambda_3+\lambda_4)$ converges to

$$(2\pi)^{-1}\int_{\mathscr{B}}|\tilde{\gamma}(\lambda;\boldsymbol{\alpha}_0)|^4[\phi(\lambda)f_z^c(\lambda)]^2\,d\lambda$$

and so the sum of all these terms converges to zero. Moreover, each of the terms in $f(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ converges to

$$(2\pi)^{-1} \int_{\mathscr{B}} \int_{\mathscr{B}} |\tilde{\gamma}(\mu; \boldsymbol{\alpha}_0)\tilde{\gamma}(\lambda; \boldsymbol{\alpha}_0)|^2 \phi(\mu)\phi(\lambda) f(-\mu, \lambda, \mu, -\lambda)\, d\mu\, d\lambda ,$$

so their sum also converges to zero. It follows that (3.6) converges to zero in mean square and so this term has no effect on the asymptotic distributional properties of our estimates. A similar term arising from the first expression in (3.4) may be dealt with in an analogous way, and (3.4) is asymptotically normal $(0, \Psi\{\boldsymbol{\alpha}_0, \beta_0; \phi^2 f_x^{\mathscr{A}}\})$. Now from the mean-value theorem, (3.4) equals $N^{\frac{1}{2}}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0, \hat{\beta} - \beta_0)$ times a matrix which, because of the consistency of $\hat{\alpha}$, $\hat{\beta}$, converges a.s. to $\Psi(\boldsymbol{\alpha}_0, \beta_0; \phi)$. Because $(\boldsymbol{\alpha}_0, \beta_0)$ is identified, it follows from (xii) and Rothenberg [11], page 581 that $\Psi\{\boldsymbol{\alpha}_0, \beta_0; \phi\}$ has full rank, and the theorem is established.

4. Final comments. It follows from [4], Theorem 5 that the $\phi(\lambda)$ minimizing (3.3) is $f_x^{\mathscr{A}}(\lambda)^{-1}$, assuming it exists over $\mathscr{B}$. There are many ways in which $f_x^{\mathscr{A}}$, might be estimated, some of which fall into the following scheme. After using an arbitrary ϕ to find initial estimates $\bar{\boldsymbol{\alpha}}$, $\bar{\beta}$ we then compute

$$\hat{f}_x^{\mathscr{A}}(\omega_m) = N^{-1} \sum_{\mathscr{B}} |w_y(s) - \bar{\beta}\tilde{\gamma}(\lambda_s; \bar{\boldsymbol{\alpha}}) w_z(s)|^2 K_M(\lambda_s - \omega_m) ,$$

$\omega_m = \pi m/M$, for integral m such that $\omega_m \in \mathscr{B}$ and for some suitable integral $M \ll N$. The kernel K_M is chosen so that

$$K_M(\lambda) \geqq 0 , \quad \int_{\mathscr{B}} K_M(\lambda)\, d\lambda = 2\pi , \qquad \lim_{M\to\infty} K_M(\lambda) = 2\pi\delta(\lambda) .$$

For example, we might have the rectangular spectral window

$$(4.1) \qquad K_M(\lambda) = 2M , \quad -\pi/2M < \lambda \leqq \pi/2M ; \qquad = 0 , \quad \text{otherwise,}$$

or alternatively, when $\mathscr{B} = (-\pi, \pi)$,

$$K_M(\lambda) = M , \quad \lambda = 0 ; \qquad = 2\pi F_M(\lambda) , \quad \text{otherwise.}$$

Then if we replace $\phi(\lambda_s)$ in (2.3) by $\hat{f}_x^{\mathscr{A}}(\omega_m)^{-1}$, $\omega_m - \pi/2M < \lambda_s \leqq \omega_m + \pi/2M$, it may be shown, as in [6], Section 3 (where the kernel (4.1) is used), that our theorems hold for fixed M with $\phi(\lambda)$ replaced in (3.3) by the reciprocal of

$$\lim_{N\to\infty} \hat{f}_x^{\mathscr{A}}(\omega_m) = \frac{1}{2\pi} \int_{\mathscr{B}} f_{xx}^{\mathscr{A}}(\mu) K_M(\mu - \omega_m)\, d\mu ,$$

$\omega_m - \pi/2M < \lambda \leqq \omega_m + \pi/2M$. Moreover, since $f_x^{\mathscr{A}}$ is continuous the last expression converges as $M \to \infty$ to $f_x^{\mathscr{A}}(\omega_m)$ and the argument in [6] Section 3 enables us to assert that there exists some sequence M increasing with N such that our theorems hold and (3.3) is $\Psi\{\boldsymbol{\alpha}_0, \beta_0; f_x^{\mathscr{A}}(\lambda)^{-1}\}^{-1}$. To obtain an estimator of maximum likelihood type one would iterate with respect to the estimation of $f_x^{\mathscr{A}}$.

Of the special cases of (1.1) that are of greatest practical importance, two have been dealt with in [6], [10] and the others are solutions of various functional equations, such as linear differential equations with constant coefficients, difference-differential equations, delay-differential equations and integral equations. (These, and a multivariate analogue of (1.1), are discussed in [9].) While we have given precise results for the general case, we note that the most suitable method of estimation may well, depending on γ, vary from that here proposed; some choices of γ allow a weakening of our conditions; robustness to departures from (v) and relevance of asymptotic theory will depend on γ; estimators may be constructed that are based on alternative solutions to the aliasing problem; assessing whether the identification condition (vi) holds is for some γ a difficult problem in itself.

Acknowledgments. I am very grateful to Professor E. J. Hannan for advice on the original version, which was written while I was at the Australian National University and was adapted from my Ph. D. thesis. I am also very grateful for the constructive criticism of the referees, one of whom offered particularly detailed comments. The paper was revised with partial support from NSF Grant GS-32327XI.

REFERENCES

[1] BARTLETT, M. S. (1955). *An Introduction to Stochastic Processes.* Cambridge Univ. Press.
[2] GOLUB, G. H. and PEREYRA, V. (1973). The differentiation of pseudo-inverses and non-linear least squares problems whose variables separate. *SIAM J. Numer. Anal.* **10** 412–432.
[3] GRENANDER, U. and ROSENBLATT, M. (1957). *The Statistical Analysis of Stationary Time Series.* Wiley, New York.
[4] HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.
[5] HANNAN, E. J. (1971). Non-linear time series regression. *J. Appl. Probability* **8** 767–780.
[6] HANNAN, E. J. and ROBINSON, P. M. (1973). Lagged regression with unknown lags. *J. Roy. Statist. Soc. Ser. B.* **35** 252–267.
[7] JENNRICH, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40** 633–643.
[8] KAPLAN, W. (1662). *Operational Methods for Linear Systems.* Addison-Wesley, Reading, Mass.
[9] ROBINSON, P. M. (1972). The estimation of continuous-time systems using discrete data. Ph. D. thesis, Australian National Univ.
[10] ROBINSON, P. M. (1974). Stochastic difference equations with non-integral differences. *Adv. Appl. Probability* **6** 524–545.
[11] ROTHENBERG, T. J. (1971). Identification in parametric models. *Econometrica* **39** 577–591.
[12] SIMS, C. A. (1971). Discrete approximations to continuous time distributed lags in econometrics. *Econometrica* **39** 545–563.
[13] ZYGMUND, A. (1959). *Trigonometric Series*, **1**. Cambridge Univ. Press.

DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY
ONE OXFORD STREET
CAMBRIDGE, MASSACHUSETTS 02138