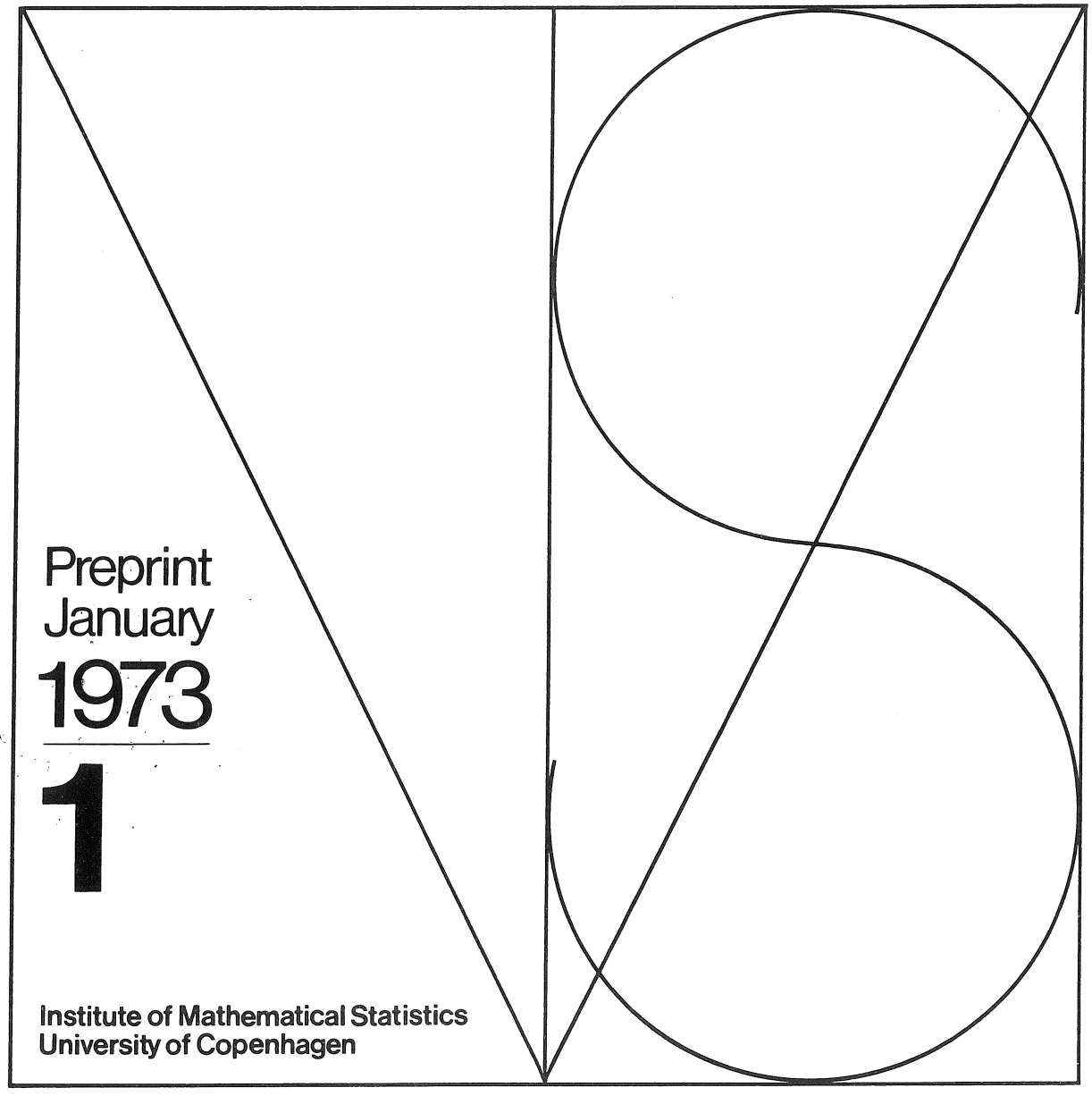
Fred L. Ramsey

On the Partial Autocorrelation Function in Time Series



Institute of Mathematical Statistics
University of Copenhagen

Fred L. Ramsey*

On the Partial Autocorrelation Function for Time Series.

Preprint 1972 No. 1

Institute of Mathematical Statistics
University of Copenhagen
January, 1973.

* Currently on leave from Dept. of Statistics, Oregon State University, Corvallis, Oregon.

1. INTRODUCTION AND SUMMARY

Considerable attention has been given recently to the partial autocorrelation functions (PACF's) of stationary, discrete parameter, time series having particular model structure. See Box and Jenkins [3], for example. Barndorff-Nielsen and Schou [2] provided a fundamental theorem which argues strongly in favor of parameterizing autoregressive models by their partial autocorrelations.

This paper extends the Barndorff-Nielsen and Schou result by establishing a characterization of a stationary time series from its PACF, wherein the structural restrictions of stationarity are easily understood. Some well known structural features of stationary time series are discussed in terms of the PACF and the likelihood function is presented.

Parameterization of a time series by its PACF, though it provides a useful view of the structure, does not unfortunately facilitate the difficult inference problems.

2. PRELIMINARIES

Let $\acute{Z} = \{0, \pm 1, \pm 2, \ldots\}$ and $\acute{Z}_+ = \{1, 2, \ldots\}$. The discrete parameter time series $x = \{x_t, t \in \acute{Z}\}$ is called a second order time series if all second moments are finite. x is said to be wide-sense stationary if it is a second order time series whose first and second order moments are independent of cardinal time.

Let χ be the set of all Gaussian, wide-sense stationary time series with zero mean and unit variance. (It is convenient to consider only χ but to think of an element $x \in \chi$ as being a typical member of a broad equivalence class of w.s. stationary series obtainable by location and scale changes and, possibly, distributional changes leaving the first and second order moments fixed).

Let $\acute{R}$ be the set of all sequences $\rho = \{\rho_t, t \in \acute{Z}\}$ which satisfy

(i) $$\rho_0 = 1;$$

(ii) $$\rho_t = \rho_{-t}, \ \forall t, \ \text{(symmetry)};$$

and, for every $n \in \acute{Z}_+$ and all choices of indices $t_1 < \ldots < t_n$ from $\acute{Z}$ and of real numbers $\delta_1, \ldots, \delta_n$,

(iii) $$\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \delta_j \rho_{t_i - t_j} \geq 0 \ \text{(pos.semi-def.)}.$$

It is convenient to consider

(2.1) $$\acute{R} = B(\acute{R}) + I(\acute{R}),$$

where $I(\acute{R})$ - the "interior"-consists of all sequences which are strictly positive definite, while $B(\acute{R})$ - the "boundary"-consists of all sequences where equality in (iii) is achievable for some non-trivial set $\{\delta_i\}$.

It is well known that there is a 1 - 1 mapping Ψ: $\chi \to R$ for which $\Psi(x) = \rho$ if and only if $\rho_t = E\{x_t x_o\}$ for every $t \in \acute{Z}$. The sequence $\rho = \Psi(x)$ is called the autocorrelation function (ACF) for the time series x.

Definition: The partial autocorrelation function (PACF) for the second order process $x = \{x_t, t \in \acute{Z}\}$ is the doubly infinite sequence $\{\Phi_{s,t}; t > s \in \acute{Z}\}$ which has $\Phi_{s,s+1}$ the ordinary correlation between x_s and x_{s+1} and has, for $t > s+1$, $\Phi_{s,t}$ being the partial autocorrelation between x_s and x_t eliminating linear regressions on $x_{s+1}, \ldots, x_{t-1}$.

The PACF of an $x \in \chi$ is determined by a singly infinite sequence $\Phi = \{\Phi_t, t \in \acute{Z}_+\}$ where $\Phi_t = \Phi_{s,s+t}$ for all $s \in \acute{Z}$ and $t \in \acute{Z}_+$. Supposing that $\rho \in I(R)$, the sequence Φ can be determined by solving the sequence of matrix equations

$$\underline{R}_k \underline{\alpha}^{(k)} = \underline{\rho}_k, \text{ for } k \in \acute{Z}_+, \tag{2.2}$$

where

$$\underline{R}_k = \begin{Bmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & & \rho_{k-2} \\ \vdots & \vdots & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{Bmatrix}, \quad \underline{\rho}_k = \begin{Bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{Bmatrix}, \quad \underline{\alpha}^{(k)} = \begin{Bmatrix} \alpha_1^{(k)} \\ \alpha_2^{(k)} \\ \vdots \\ \alpha_k^{(k)} \end{Bmatrix}.$$

The PACF is the sequence given by

$$\Phi_k = \alpha_k^{(k)}, \text{ for } k \in \acute{Z}_+. \tag{2.3}$$

Durbin [4] gave a method for sequentially solving (2.2). The relevant equations are:

$$\text{(E.1)} \qquad \Phi_1 = \alpha_1^{(1)} = \rho_1$$

$$\text{(E.2)} \qquad \sigma_1^2 = 1 - \Phi_1^2$$

$$\text{(E.3)} \qquad \Phi_{k+1} = \alpha_{k+1}^{(k+1)} = \{\rho_{k+1} - \sum_{j=1}^{k} \alpha_j^{(k)} \rho_{k+1-j}\} / \sigma_k^2$$

$$\text{(E.4)} \qquad \alpha_j^{(k+1)} = \alpha_j^{(k)} - \Phi_{k+1} \alpha_{k+1-j}^{(k)}, \quad (j=1,\ldots,k)$$

(E.5) $$\sigma_{k+1}^2 = \sigma_k^2(1-\Phi_{k+1}^2).$$

The first two equations of E = {(E.1), ... ,(E.5)} give starting values and the remaining three explain how to go from stage k to (k+1). Physically, $\alpha_j^{(k)}$ may be interpreted in one of two ways. It is the $j^{\underline{th}}$ coefficient in an autoregressive model of order k (AR(k)). It is also the coefficient of x_{k+1-j} in the linear regression of x_{k+1} on $\{x_1,\ldots,x_k\}$. The value of σ_k^2 is, physically, the variance of the residual from that regression. Equation (E.5) can be used to show further that

(2.4) $$|\underline{R}_{k+1}| = \prod_{j=1}^{k}(1-\Phi_j^2)^{k+1-j}.$$

What seems to have been overlooked is the simple observation that any sequence of constants Φ having $|\Phi_k| < 1$ also defines, via E, a unique sequence ρ which is positive definite because of (2.4). This is the essence of the proof for Theorem 1.

3. A REPRESENTATION OF χ

The intention here is to describe an $x \in \chi$ by its PACF; i.e. to describe what sequences are PACF's for wide sense stationary time series.

Definition: Let the set $\acute{S}$ consist of all real valued sequences $s = \{s_k, k \in \acute{Z}_+\}$ which satisfy

(a) $|s_k| \leq 1$, for all $k \in \acute{Z}_+$; and

(b) $|s_k| = 1$ implies $s_{k+1} = s_k$.

Definition: Let $\acute{S}$ be decomposed as

$$\acute{S} = B(\acute{S}) + I(\acute{S}), \tag{3.1}$$

where $I(\acute{S})$ consists of all $s \in \acute{S}$ for which (a) holds as a strict inequality for all k. Thus $B(\acute{S})$ consists of sequences which have $|s_k| = 1$ for some k. It is convenient to refer here, as before, to the components in (3.1) as, respectively, the "boundary" and "interior" of $\acute{S}$.

Theorem 1: The real, discrete parameter, second order time series x is wide sense stationary if and only if its PACF $\{\Phi_{s,t}, t > s \in \acute{Z}\}$ satisfies

(A) $\Phi_{s,s+k} = \Phi_k$ for all $s \in \acute{Z}$ and $k \in \acute{Z}_+$; and

(B) $\Phi = \{\Phi_k, k \in \acute{Z}_+\} \in \acute{S}$.

Furthermore, $\Phi \in I(\acute{S})$ if and only if $\Psi(x) = \rho \in I(\acute{R})$.

Equivalently

Theorem 1: There exists a one-to-one mapping $\xi: \acute{R} \to \acute{S}$ such that if $\rho = \Psi(x)$ for $x \in \chi$, then $\Phi = \xi(\rho)$ is the PACF of x. Furthermore, $\rho \in I(\acute{R})$ if and only if $\xi(\rho) \in I(\acute{S})$.

Proof: Case I - Necessity for $\rho \in I(\acute{R})$. E has a unique solution for Φ. Each Φ_k is the correlation between two well defined random variables and thus $|\Phi_k| \leq 1$. However (2.4) implies strict inequality must hold so that $\Phi \in I(\acute{S})$.

Case II - Sufficiency for $\Phi \in I(\acute{S})$. E has a unique solution for ρ. (2.4) implies $|\underline{R}_k|$ is strictly positive for all k. So for each N, all principal minorants of $\underline{R}_N$ have positive determinants. This implies $\underline{R}_N$ is positive definite for every $N \in \acute{Z}_+$ which implies that ρ is itself positive definite.

Note E only specifies ρ_k for $k \in \acute{Z}_+$, but this uniquely determines ρ by extension using (i) and (ii) in the definition of $\acute{R}$.

Case III - Necessity for $\rho \in B(\acute{R})$. There exists a positive integer p for which $|\underline{R}_k| = 0$ for all $k \geq p$ and $|\underline{R}_k| > 0$ for all $k < p$. E has a unique solution for $\{\Phi_1, \ldots, \Phi_{p-1}\}$ where each is strictly less than one in magnitude. There exists, since $\underline{R}_p$ has rank (p-1), a vector $\underline{\lambda}' = (\lambda_1, \ldots, \lambda_p)$ whose direction is unique, which has $\lambda_1 \neq 0$ and $\lambda_p \neq 0$ and for which

$$\underline{\lambda}' \underline{R}_p \underline{\lambda} = \mathrm{Var}\{\sum_{j=1}^{p} \lambda_j x_j\} = 0,$$

(where $\rho = \Psi(x)$). Using stationarity of x, it follows that for every $t > s+p$, the residuals from regressions of x_s and x_t on $x_{s+1}, \ldots, x_{t-1}$ are zero with probability one. Hence, in the sense that zero predicts zero perfectly, $\Phi_{s,t} = \Phi_{t-s} = \pm 1$ for all $t-s \geq p$, where the sign is determined by whether λ_1 and λ_p have like or opposite signs. Clearly the resulting Φ is in $B(\acute{S})$.

Case IV - Sufficiency for $\Phi \in B(\acute{S})$. There exists a p for which $|\Phi_k| = 1$ for all $k \geq p$ and $|\Phi_k| < 1$ for all $k < p$. E has a unique solution for $\{\rho_0, \rho_1, \ldots, \rho_p\}$ where $\rho_0 = 1$, $|\underline{R}_k| > 0$ for $k \leq p$ and $|\underline{R}_{p+1}| = 0$. Let $\{x_1, \ldots, x_{p+1}\}$ be defined as having a multivariate Gaussian distribution with means zero, variances one and covariance matrix $\underline{R}_{p+1}$ (of rank p). The residuals from regressions of x_1 on $\{x_2, \ldots, x_{p+1}\}$ and of x_{p+1} on

$\{x_1,\ldots,x_p\}$ are zero with probability one. That is, with probability one,

$$x_t = \sum_{j=1}^{p} \beta_j x_{t+j} \qquad (t = 1) \tag{3.2}$$

and

$$x_t = \sum_{j=1}^{p} \alpha_j x_{t-j} \qquad (t = p+1). \tag{3.3}$$

Extend the sequence $\{x_1,\ldots,x_{p+1}\}$ according to the difference equation (3.2) for $t = 0, -1, -2, \ldots$ and according to (3.3) for $t = p+2, p+3, \ldots$. The result is a wide sense stationary time series for which Φ is the PACF. Clearly the corresponding ρ, derivable from the series, is in $B(\acute{R})$. q.e.d.

The advantage Theorem 1 enjoys over other ways of specifying the structure of wide sense stationary time series is that the stationarity region for the structural parameters is so simple. Whereas condition (iii) for $\acute{R}$ describes complicated regions for the individual correlations and whereas the stationarity conditions for autoregression models describe very complicated regions for the model parameters, the structure of $\acute{S}$ is such that each partial autocorrelation is allowed to vary over the open interval (-1, 1) independently of the others.

The time series x is called an autoregression of order p, AR(p), if it satisfies a $p^{\underline{th}}$ order stochastic difference equation with white noise "shocks"; i.e. if there exist real numbers $\alpha_1, \ldots, \alpha_p$ such that

$$x_t = \sum_{j=1}^{p} \alpha_j x_{t-j} + y_t, \text{ for all } t \in \acute{Z}, \tag{3.4}$$

where $y = \{y_t, t\in\acute{Z}\}$ is a sequence of independent $N(0,\sigma^2)$ random variables. It is well known that (3.4) has a stationary solution for x if and only if the roots of the characteristic polynomial equation

$$\alpha(x) = 1-\alpha_1 x- \ldots -\alpha_p x^p = 0 \tag{3.5}$$

all lie outside the unit circle in the complex plane. Refe-

rence to the proof of this statement (see, e.g. Anderson [1]) shows that the unit circle itself, the boundary, may be included provided $\sigma^2 = 0$.

Definition: A second order time series x which is a stationary solution to the homogeneous ($y_t \equiv 0$) stochastic difference equation (3.4) is called a degenerate autoregression of order p, denoted DAR(p).

Definition: A stationary solution to (3.4) where $\sigma^2 > 0$ will be called non-degenerate, or NAR(p).

The following statements are clear. First, if according to the results above we decompose

$$\chi = B(\chi) + I(\chi), \tag{3.6}$$

then $x \in B(\chi)$ if and only if x is DAR(p) for some $p < \infty$. Secondly x is NAR(p) if and only if $\Phi_k = 0$ for all $k > p$. Finally,

Theorem 2: (Barndorff-Nielsen and Schou [2]). There exists a 1 - 1 mapping of the feasible $\{\alpha_1,\ldots,\alpha_p\}$ region for a NAR(p) onto the p-dimensional cube $(\Phi_1, \ldots,\Phi_p) \in (-1,+1)^p$.

4. A STRUCTURAL COMMENT

Rewrite equation (E.5) as

$$\sigma_N^2 = \prod_{k=1}^{N} (1-\Phi_k^2). \tag{4.1}$$

In the Wold decomposition for stationary time series, an x is purely deterministic if and only if $\sigma_N^2 \to 0$ as $N \to \infty$. Intuitively if we continue to gather information as more of the series is observed, perfect prediction will ultimately be possible; i.e.

Lemma: If

$$\overline{\lim_{k\to\infty}} |\Phi_k| = \omega > 0, \tag{4.2}$$

then x is purely deterministic.

Proof: Consider an ϵ in $(0,\omega^2)$ to be given. (We may consider $\Phi \in I(S)$ since the boundary is obvious.) If $\theta = 1-\omega^2+\epsilon$ choose M_ϵ so that

$$M_\epsilon > (\log\epsilon)/(\log\theta).$$

Let $\{k_i, i\in Z_+\}$ be a sequence of indices for which $|\Phi_{k_i}| \to \omega$ as $i \to \infty$ and let N_ϵ be large enough to insure that there are M_ϵ of the k_i less than N_ϵ which have

$$\Phi_{k_i}^2 > \omega^2 - \epsilon.$$

From (4.1),

$$\sigma_N^2 \leq (1-\omega^2+\epsilon)^{M_\epsilon} < \epsilon. \qquad \text{q.e.d.}$$

As a result of the Lemma, we may consider $\Phi_k \to 0$ in the following so that the right hand side of

$$\log \sigma_N^2 = \sum_{k=1}^{N} \log(1-\Phi_k^2)$$

will be dominated by

$$S_N = \sum_{k=1}^{N} \Phi_k^2.$$

Therefore, we have

Theorem 3: $x \in \chi$ is purely deterministic if and only if $\Phi = \xi(\Psi(x))$ is such that

$$\lim_{N\to\infty} \sum_{k=1}^{N} \Phi_k^2 = +\infty.$$

5. LIKELIHOOD FOR THE PACF

The results of Theorem 4 seem to be well known for the case where x is NAR(p). It is, in our statement of it, appropriate to consider the full parameterization for a stationary Gaussian time series as being $\{\mu, \gamma_0, \Phi\}$.

Theorem 4: Let $x = \{x_t,\ t\in\acute{Z}\}$ be such that, for every $t\in\acute{Z}$,

$$x_t = \mu + \gamma_0^{1/2} y_t,$$

where $y = \{y_t, t\in\acute{Z}\}$ is in $I(\chi)$. Then

(5.1) $$x_1 \sim N(\mu, \gamma_0)$$

and for $k = 1,2,\ldots,$ and the conditional distribution of x_{k+1} given $x_1,\ldots,x_k$ is Gaussian with mean and variance given respectively by

(5.2) $$E(x_{k+1}|x_1,\ldots,x_k) = \mu + \sum_{j=1}^{k} \alpha_j^{(k)} (x_{k+1-j} - \mu)$$

and

(5.3) $$\mathrm{Var}(x_{k+1}|x_1,\ldots,x_k) = \gamma_0(1-\Phi_1^2)\ldots(1-\Phi_k^2).$$

Here the $\alpha_j^{(k)}$ coefficients are defined in (2.2)

Proof: The covariance matrix of $\underline{x}' = (x_1,\ldots,x_{k+1}) = (\tilde{\underline{x}}', x_{k+1})$ is simply $\gamma_0 \underline{R}_{k+1}$. But if we define $\underline{\rho}_k^*$ as the $(k\times 1)$ vector whose elements are the same as in $\underline{\rho}_k$ of (2.2) but in reverse order, then in the exponent of the joint density for $\underline{x}$ we find the quadratic form

$$Q_{k+1} = (\underline{x}-\underline{\mu})'\underline{R}_{k+1}^{-1}(\underline{x}-\underline{\mu}) = (\underline{x}-\underline{\mu})'\left\{\begin{array}{c|c} \underline{R}_k & \underline{\rho}_k^* \\ \hline \underline{\rho}_k^{*\prime} & 1 \end{array}\right\}^{-1}(\underline{x}-\underline{\mu})$$

$$= (\underline{x}-\underline{\mu})'\left\{\begin{array}{c|c} \underline{R}_k^{-1} + (\underline{R}_k^{-1}\underline{\rho}_k^*\underline{\rho}_k^{*\prime}\underline{R}_k^{-1})/\sigma_k^2 & -(\underline{R}_k^{-1}\underline{\rho}_k^*)/\sigma_k^2 \\ \hline \text{(symmetry)} & 1/\sigma_k^2 \end{array}\right\}(\underline{x}-\underline{\mu})$$

$$= Q_k + \sigma_k^{-2}\{x_{k+1} - E(x_{k+1}|x_1,\ldots,x_k)\}^2,$$

since $\underline{R}_k^{-1}\underline{\rho}_k^*$ is the vector $\underline{\alpha}^{(k)}$ with its elements reversed in order. The rest is straightforward.

Two comments are in order at this point. First, it is even fairly clear how to determine the likelihood function when $\Phi \in B(S)$. Secondly, (E.4) shows that though $E(x_{k+1}|x_1,\ldots,x_k)$ is a linear function in each of the parameters $\Phi_1,\ldots,\Phi_k$, it is a complicated function of all of them. In fact the entire joint distribution of $x_1,\ldots,x_{k+1}$ depends only upon $\Phi_1,\ldots,\Phi_k$ for each k, so that conditional inferences given $x_1,\ldots,x_p$ about $\Phi_p,\ldots,\Phi_{p+q}$ using the Theorem 4 likelihood are unconditionally valid.

In the usual parameterization for NAR(p) it is customary to follow Mann and Wald [5] and to condition on $x_1,\ldots,x_p$ when constructing inference procedures. In that situation, where such conditioning is not technically justified for inferences on $\Phi_{p+1},\ldots,\Phi_{p+q}$, the effect of the conditioning is to greatly simplify the likelihood. When parameterizing a NAR(p) by its PACF, where conditioning <u>is</u> appropriate, the resulting likelihood is, alas, a nightmare.

One may nevertheless draw some (well known) conclusions from it. With the sample $x_1,\ldots,x_N$ write L as the total likelihood and

$$\ell = -\log L = \sum_{n=1}^{N} \ell_n, \tag{5.4}$$

where ℓ_n comes from the conditional distribution of x_n given its predecessors. That is, ignoring irrelevant constants,

$$\ell_{n+1} = \frac{1}{2}\log\gamma_0 + \frac{1}{2}\sum_{j=1}^{n}\log(1-\Phi_j^2)$$

$$+\{2\gamma_0\prod_{j=1}^{n}(1-\Phi_j^2)\}^{-1}\{(x_{n+1}-\mu) - \sum_{j=1}^{n}\alpha_j^{(n)}(x_{n+1-j}-\mu)\}^2.$$

Straightforward calculation gives

$$E\left(\frac{\partial^2 \ell_{n+1}}{\partial \mu^2}\right) = \gamma_0^{-1} \prod_{j=1}^{n} \frac{(1-\Phi_j)}{(1+\Phi_j)}$$

$$E\left(\frac{\partial^2 \ell_{n+1}}{\partial \gamma_0 \partial \mu}\right) = E\left(\frac{\partial^2 \ell_{n+1}}{\partial \Phi_k \partial \mu}\right) = 0, \quad \forall k,$$

(5.5)
$$E\left(\frac{\partial^2 \ell_{n+1}}{\partial \gamma_0^2}\right) = (2\gamma_0^2)^{-1}$$

$$E\left(\frac{\partial^2 \ell_{n+1}}{\partial \Phi_k \partial \gamma_0}\right) = -\Phi_k / \{\gamma_0 (1-\Phi_k^2)\}$$

and

$$E\left(\frac{\partial^2 \ell_{n+1}}{\partial \Phi_h \partial \Phi_k}\right) = \frac{2\Phi_k \Phi_h}{(1-\Phi_k^2)(1-\Phi_h^2)} + \frac{Q_n(h,k)}{\prod_{j=1}^{n} (1-\Phi_j^2)}$$

In the last expression

$$Q_n(h,k) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{\partial \alpha_i^{(n)}}{\partial \Phi_h}\right)\left(\frac{\partial \alpha_j^{(n)}}{\partial \Phi_k}\right)\rho_{i-j}.$$

Though these are difficult to evaluate, the equations in E allow certain reductions.

(5.6)
$$Q_n(n,k) = 0, \text{ for } k < n.$$

(5.7)
$$Q_n(n,n) = \prod_{j=1}^{n-1} (1-\Phi_j^2)$$

and if both k and h are less than n,

(5.8)
$$Q_n(h,k) = \{1+\Phi_n^2\}Q_{n-1}(h,k) - 2\Phi_n P_{n-1}(h,k),$$

where

(5.9)
$$P_n(h,k) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{\partial \alpha_i^{(n)}}{\partial \Phi_h}\right)\left(\frac{\partial \alpha_{n+1-j}^{(n)}}{\partial \Phi_k}\right)\rho_{i-j}.$$

Now

$$P_n(n,k) = 0, \text{ for } k \leq n$$

and

$$P_n(n,n) = -\Phi_{n-1}Q_n(n,n),$$

but for k and h less than n, (5.9) is a mess. However, (5.6 ⇔ 8) imply that when $\Phi_n = 0$ for $n > p$, then

$$E\left(\frac{\partial^2 \ell_{n+1}}{\partial\Phi_h \partial\Phi_k}\right) = 0, \text{ when } k \neq h,$$

if either h or k exceeds p. And if $k > p$, then

$$E\left(\frac{\partial^2 \ell_{n+1}}{\partial\Phi_k^2}\right) = 1$$

In summary, the total information matrix for the parameter vector $\underline{\theta}' = (\mu,\gamma_0,\Phi_1,\ldots,\Phi_p,\Phi_{p+1},\ldots,\Phi_{p+q})$ under the hypothesis of a NAR(p) is given by

$$\left\{\begin{array}{c|c|c|c} i_N(\mu) & & & \\ \hline & i_N(\gamma_0) & \underline{i}_N(\gamma_0,\cdot)' & \\ \hline & \underline{i}_N(\gamma_0,\cdot) & \underline{G} & \\ \hline & & & \underline{D} \end{array}\right\},$$

where

$$i_N(\mu) = \gamma_0^{-1}\sum_{n=1}^{p-1}\prod_{j=1}^{n}\left[\frac{1-\Phi_j}{1+\Phi_j}\right] + (N-p)\gamma_0^{-1}\prod_{j=1}^{p}\left[\frac{1-\Phi_j}{1+\Phi_j}\right]$$

$$i_N(\gamma_0) = N/(2\gamma_0^2)$$

$$i_N(\gamma_0,k) = -(N-k-1)\Phi_k/\{\gamma_0(1-\Phi_k^2)\}, \text{ for } k = 1,\ldots,p.$$

$\underline{G}$ is a horrible $p \times p$ matrix, but D is a diagonal $q \times q$ matrix with $k^{\underline{th}}$ element on the diagonal

$$i_N(\Phi_{p+k}) = (N-p-k).$$

REFERENCES

[1] Anderson, T.W. The Statistical Analysis of Time Series. John Wiley & Sons, Inc. New York, 1972.

[2] Barndorff-Nielsen, O. and G. Schou. "On the Parametization of Autoregressive Models by Partial Autocorrellations." RESEARCH REPORT No. 2, Dept. of Theoretical Stat., Aarhus University, Aarhus, Denmark. September, 1972.

[3] Box, G.E.P. and G.M. Jenkins. Time Series Analysis: Forecasting and Control. Holden Day. San Francisco, 1970.

[4] Durbin, J. "The Fitting of Time Series Models". Rev. Int. Stat. Inst. 28, 1960, pp. 233-244.

[5] Mann H.B. and A. Wald. "On the Statistical Treatment of Linear Stochastic Difference Equations". Econometrica, 11, 1943, pp. 173-220.