7
The Formal Modelling of Stationary Time Series: Wold and the Russians

Moving averages and autoregressions

7.1 Slutzky's modelling of the 'summation of random causes', introduced in §§**5.11–5.16**, and the 'ordinary regression equations' (6.10) and (6.13) of Yule and Walker were to become the basic models of time series analysis. One of the reasons why they have been such enduring features, apart from their obvious usefulness, may be because of their renaming as *moving averages* and *linear autoregressions*, respectively, by Herman Wold (1938, page 2), as these are terms that convey their structure with great clarity and effectiveness.[1]

Wold regarded these two models as lying within the more general scheme of *linear regression*, having the common feature that a 'random element plays a fundamental, active role', a feature which 'constitutes a distinct contrast to the scheme of *hidden periodicities* – as we shall call the hypothesis of strict periods – and makes the schemes of linear regression *a priori* plausible in several instances where the scheme of hidden periodicities has been criticized' (*ibid.*, page 3; italics in original). Not content with simply introducing new terminology, however, Wold's real desire was to provide a formal theory, based on probability concepts, within which to set these models.

> While the schemes of linear regression thus form a type of hypothesis of the greatest importance, the development of the subject is still little advanced, both as to the theory and the application of the schemes. For instance, earlier definitions concerning the scheme of autoregression are incomplete. One of the chief purposes of the present volume is to give some contributions for completion in these respects. It also aims at bringing the schemes into place in the theory of probability, thereby uniting the rather isolated results hitherto reached.
>
> In the theory of probability, the schemes of linear regression fall under the heading of the discrete stationary random process as defined by [Khinchin (1932, 1933)]. (Wold, 1938, page 29)

T. C. Mills, *The Foundations of Modern Time Series Analysis*

Stationary random processes

7.2 As is made apparent in the second paragraph of the quote above, the models intuitively developed by Slutzky, Yule and Walker are all members of the class of discrete stationary random processes. Such a distinction had not been made explicitly before Wold and, in this context, it is worth quoting the opening two paragraphs of his introduction:

> Observational series which describe phenomena changing with time may be roughly classified in two broad categories, viz. *evolutive* and *stationary*. In the former case, different sections of the time series are dissimilar in one or more respects. For instance, the sectional averages may be distinctly different, or some other structural property of the series may present variation. In the analysis of evolutive time series, absolute time plays a fundamental role, e.g. as the independent variable in a trend function, or as a fixed scale in studying the development of a phenomenon from an initial state of rest.
>
> Stationary time series are unchanging in respect to their general structure. The fluctuations up and down in such a series may seem random or show tendencies to regularity – in any case, the character of the series is, on the whole, the same in different sections. Or otherwise expressed, in the analysis of stationary series time is allotted the secondary role of a passive medium. Even without preparation, observational time series are frequently stationary. On the other hand, the deviations from trend form a type of derived time series which is often stationary. (*ibid.*, page 1)

7.3 Wold's theoretical development of stationary random processes used the concepts and techniques introduced by the Russian mathematicians Khinchin (1932, 1933, 1934) and Kolmogorov (1931, 1933).[2] In this chapter we introduce the basic concepts employed by Wold and the fundamental theorems that he proved to obtain the representations that now form the formal basis of modern time series analysis. In order to keep the development manageable, however, no attempt is made to be inclusive and proofs are not provided.

In his formal development, Wold let $\{t\}$ stand for the set of values taken by a real parameter, assumed to represent time, and let one random variable $\xi(t)$ correspond to each time point t in $\{t\}$. The corresponding set of random variables, denoted $\{\xi(t)\}$, will then be a random process if the following conditions are satisfied.

(A) If a subset of $\{t\}$, say $(t) = (t_1, \ldots, t_n)$, is arbitrarily chosen, then the variable $\xi(t_1, \ldots, t_n) = [\xi(t_1), \ldots, \xi(t_n)]$ will be well-defined. Let the distribution function of $\xi(t_1, \ldots, t_n)$ be denoted by

$$F(t_1, \ldots, t_n; u_1, \ldots, u_n) = P[\xi(t_1) \leq u_1, \ldots, \xi(t_n) \leq u_n] \tag{7.1}$$

Here the right-hand side of (7.1) denotes the joint probability that $\xi(t_1) \leq u_1, \ldots, \xi(t_n) \leq u_n$ and may be termed the probability function: let the sets of distribution and probability functions of $\xi(t_1, \ldots, t_n)$ be denoted by $\{F\}$ and $\{P\}$ respectively.

(B) Given $(t) = (t_1, \ldots, t_n)$ and with $(i_1, \ldots, i_n)$ being an arbitrary permutation of the sequence $(1, 2, \ldots, n)$, the functions $\{F\}$ will satisfy the following relations identically in $u_1, \ldots, u_n$:

$$F(t_{i_1}, \ldots, t_{i_n}; u_{i_1}, \ldots, u_{i_n}) = F(t_1, \ldots, t_n; u_1, \ldots u_n) \tag{7.2}$$

$$F(t_1, \ldots, t_m; u_1, \ldots, u_m) = F(t_1, \ldots, t_n; u_1, \ldots, u_m, +\infty, \ldots, +\infty) \tag{7.3}$$

where $m < n$. Equations (7.2) and (7.3) imply that the probability laws governing $\{\xi(t)\}$ must not contradict themselves and they are thus referred to as the *consistency relations*.

This random (or stochastic) process thus extends the notion of a random variable to an infinite number of dimensions. The sample elements of the process $\{\xi(t)\}$, also called the *realizations* of the process, are functions of t, say $\xi_i(t)$. Keeping t fixed at, say, $t = t_1$, the set of sample values $\xi_i(t_1)$ that constitute the random variable $\xi(t_1)$ are obtained. More generally, if we keep $t_1, \ldots, t_n$ fixed, the realizations will provide the 'universe' of sample elements $[\xi_i(t_1), \ldots, \xi_i(t_n)]$ that constitute the n-dimensional random variable $[\xi(t_1), \ldots, \xi(t_n)]$.

To be able to define stationarity, arbitrary translations within the set $\{t\}$ must be considered. Assume that $\{t\}$ consists either of all real values or is formed by an unbroken sequence of equidistant values, say $\ldots, -1, 0, 1, 2, \ldots$. A random process $\{\xi(t)\}$ as defined by a set $\{F\}$ is then termed *stationary* if, for an arbitrary subset $(t) = (t_1, \ldots, t_n)$, the relation

$$F(t_1 + t, \ldots t_n + t; u_1, \ldots, u_n) = F(t_1, \ldots t_n; u_1, \ldots, u_n)$$

is identically satisfied in $u_1, \ldots, u_n$ and in t. If t is restricted to be a sequence of equidistant values then $\{\xi(t)\}$ is called a *discrete* stationary random process: if t is arbitrary then the process will be *continuous*.

7.4 Expectations derived from the distribution functions $\{F\}$ determining a stationary process $\{\xi(t)\}$ are called the *characteristics* of the process and will be independent of t, as will be the distribution functions $F(t; u)$: the function of u so obtained is then the *principal* distribution function, $F(u)$. Restricting attention to a one-dimensional stationary process, the mean μ and variance σ^2 are then defined as[3]

$$\mu = E(\xi) = \int_{-\infty}^{\infty} u dF(u) \quad \sigma^2 = E[(\xi - \mu)^2] = \int_{-\infty}^{\infty} (u - \mu)^2 dF(u)$$

If the variance is finite, the *automoments* of second order, as defined by

$$\nu_2^{(k)} = E(\xi(t) \cdot \xi(t+k)) = \int_{R_2} uv \cdot d_{u,v}F(t, t+k; u, v) = \nu_2^{(-k)}$$

will also be finite. These characteristics determine the *autocorrelation coefficients* of the stationary process $\{\xi(t)\}$

$$r_k = r_k(\xi) = (\nu_2^{(k)} - \mu^2)/\sigma^2 = r_{-k}$$

If $r_k(\xi) = 0$ for all $k \neq 0$, the process $\{\xi(t)\}$ is termed *non-autocorrelated*.

Now consider a set of random processes $\{\xi^{(1)}(t)\}, \ldots, \{\xi^{(k)}(t)\}$ with an arbitrarily chosen set of time points $(t) = (t_1, \ldots, t_n)$ and k sets of real numbers $(u^{(s)}) = (u_1^{(s)}, \ldots, u_n^{(s)}), s = 1, \ldots, k$. The processes $\{\xi^{(i)}(t)\}$ will be called *independent* if the following relation is satisfied

$$\begin{aligned} &P[\xi^{(1)}(t_1) \leq u_1^{(1)}, \ldots, \xi^{(1)}(t_n) \leq u_n^{(1)}; \ldots; \xi^{(k)}(t_1) \leq u_1^{(k)}, \ldots, \xi^{(k)}(t_n) \leq u_n^{(k)}] \\ &\quad = P[\xi^{(1)}(t_1) \leq u_1^{(1)}, \ldots, \xi^{(1)}(t_n) \leq u_n^{(1)}] \ldots P[\xi^{(k)}(t_1) \leq u_1^{(k)}, \ldots, \xi^{(k)}(t_n) \leq u_n^{(k)}] \end{aligned}$$

If it is assumed that the independent processes $\{\xi^{(i)}(t)\}$ are stationary and have finite variances $\sigma^2(\xi^{(i)})$ then the sum process

$$\{\zeta_k(t)\} = a_1\{\xi^{(1)}(t)\} + \cdots + a_k\{\xi^{(k)}(t)\}$$

is stationary with expectation, variance and autocorrelation coefficients given by

$$E\{\zeta_k\} = a_1 E\{\xi^{(1)}\} + \cdots + a_k E\{\xi^{(k)}\}$$

$$\sigma^2(\zeta_k) = a_1^2 \sigma^2(\xi^{(1)}) + \cdots + a_k^2 \sigma^2(\xi^{(k)}) \tag{7.4}$$

$$r_p(\zeta_k) = a_1^2 \frac{\sigma^2(\xi^{(1)})}{\sigma^2(\zeta_k)} r_p(\xi^{(1)}) + \cdots + a_k^2 \frac{\sigma^2(\xi^{(k)})}{\sigma^2(\zeta_k)} r_p(\xi^{(k)}) \tag{7.5}$$

The expressions (7.4) and (7.5) depend on the identities

$$r(\xi^{(r)}(t \pm p); \xi^{(s)}(t \pm q)) = 0 \quad p \geq 0, \ \ q \geq 0 \tag{7.6}$$

where r and s are arbitrary. If (7.6) is satisfied then $\{\xi^{(r)}\}$ and $\{\xi^{(s)}\}$ are said to be *uncorrelated*. In fact, (7.4) and (7.5) will hold if $\{\xi^{(r)}\}$ and $\{\xi^{(s)}\}$ are simply uncorrelated processes, stationary or otherwise.

Similarly, the moving average process defined by

$$\zeta(t) = a_0\xi(t) + a_1\xi(t-1) + \cdots + a_h\xi(t-h) \tag{7.7}$$

will also be stationary if $\xi(t)$ is stationary (*ibid.*, page 38).

These operations may also be applied to observed time series, so that, if $\ldots, \bar{\xi}_{t-1}, \bar{\xi}_t, \bar{\xi}_{t+1}, \ldots$ represents such a series, the counterpart to (7.7), for example, is

$$\bar{\zeta}_t = a_0\bar{\xi}_t + a_1\bar{\xi}_{t-1} + \cdots + a_h\bar{\xi}_{t-h}$$

7.5 Wold then utilized the concept of convergence in probability to state his first theorem. A sequence $\xi^{(1)}, \xi^{(2)}, \ldots$ of random variables is said to *converge in probability* to a random variable ξ if, for every $\varepsilon > 0$

$$P[|\xi^n - \xi| > \varepsilon] \to 0$$

as $n \to \infty$. A sequence of random processes $\{\xi^{(1)}(t)\}, \{\xi^{(2)}(t)\}, \ldots$ is then called *convergent in probability to a limit process* $\{\xi(t)\}$ if, for an arbitrary set $(t) = (t_1, \ldots, t_n)$, the sequence

$$\xi^{(1)}(t_1, \ldots, t_n), \xi^{(2)}(t_1, \ldots, t_n), \ldots$$

is convergent in probability to the limit variable $\xi(t_1, \ldots, t_n)$. This allows Wold (*ibid.*, page 40) to state

Theorem 1.
A necessary and sufficient condition that a sequence $\{\xi^{(1)}(t)\}, \{\xi^{(2)}(t)\}, \ldots$ of random processes be convergent in probability is that, for an arbitrary t, the sequence $\xi^{(1)}(t), \xi^{(2)}(t), \ldots$ be convergent in probability. If the sequence is convergent and if every process $\{\xi^{(n)}(t)\}$ is stationary, the limit process will be stationary.

7.6 Suppose $\xi = [\xi^{(1)}, \ldots, \xi^{[n]}]$ represents an n-dimensional random variable with distribution function $F(u_1, \ldots, u_n)$ and there exists a linear function, say

$$L[x - \mu] = a_1(x^{(1)} - \mu_1) + \cdots + a_n(x^{(n)} - \mu_n)$$

such that

$$P[L[\xi - \mu] \neq 0] = P[a_1(\xi^{(1)} - \mu_1) + \cdots + a_n(\xi^{(n)} - \mu_n) \neq 0] = 0$$

The distribution of ξ is then said to be (*linearly*) *singular* and the variables $\xi^{(i)}$ are said to be connected by the relation $L[\xi - \mu] = 0$. The singularity is of *rank h* if there exist only $n - h$ independent relations between the variables $\xi^{(i)}$, say

$$\begin{aligned} a_{1,h+1}(\xi^{(1)} - \mu_1) + \cdots + a_{n,h+1}(\xi^{(n)} - \mu_n) &= 0 \\ a_{1,n}(\xi^{(1)} - \mu_1) + \cdots + a_{n,n}(\xi^{(n)} - \mu_n) &= 0 \end{aligned}$$

Suppose that the singularity is of the form

$$\xi(t) - \mu + a_1(\xi(t-1) - \mu) + \cdots + a_h(\xi(t-h) - \mu) = 0 \tag{7.8}$$

where $h \leq n$. This is known as a *stochastical difference relation of order h.* Wold (*ibid.*, page 45) was then able to prove the following result.

Theorem 2.
Let $\{\xi(t)\}$ be a discrete stationary process with autocorrelation coefficients r_k. If $\{\xi(t)\}$ is linearly singular then it is a process of superposed harmonics. A necessary condition that $\{\xi(t)\}$ be linearly singular, say on account of the relation $L[\xi(t) - m] = 0$ given by [7.8], *is that r_k satisfies the difference equation $L[r_k] = 0$.*

On defining the *principal correlation determinants*

$$\Delta(r, n) = \begin{vmatrix} 1 & r_1 & r_2 & \ldots & r_n \\ r_1 & 1 & r_1 & \ldots & r_{n-1} \\ r_2 & r_1 & 1 & \ldots & r_{n-2} \\ \vdots & \ldots & \ldots & \ldots & \vdots \\ r_n & r_{n-1} & r_{n-2} & \ldots & 1 \end{vmatrix} \geq 0 \tag{7.9}$$

Wold (*ibid.*, page 47) could then assert

Theorem 3.
Let $\{\xi(t)\}$ be a discrete stationary process with principal correlation determinants $\Delta(r, n)$. A necessary and sufficient condition that $\{\xi(t)\}$ be singular of rank h is that $\Delta(r, h)$ be the first vanishing determinant in the sequence $\Delta(r, 1)$, $\Delta(r, 2), \ldots$.

Wold showed that a stationary process with finite variance which satisfies (7.8) will also satisfy the difference relation

$$\Delta^{2s}\xi(t-s) + h_1\Delta^{2s-2}\xi(t-s+1) + \cdots + h_s[\xi(t) - \mu] = 0 \tag{7.10}$$

Types of discrete stationary processes

7.7 A *purely random process* is a process such that (7.1) has the form

$$F(t_1, \ldots, t_n; u_1, \ldots, u_n) = F(u_1) \ldots F(u_n)$$

On extending the notation to let $\{\xi(t; F)\}$ denote a purely random process defined by a distribution function $F(u)$, the following theorem holds (*ibid.*, page 48)

Theorem 4.
Let $\{\xi^{(1)}(t; F^{(1)})\}$, $\{\xi^{(2)}(t; F^{(2)})\}, \ldots$ *represent independent, purely random processes such that the infinite convolution* $F^{(1)} \otimes F^{(2)} \otimes \cdots$ *is convergent. Then the sum*

$$\{\xi^{(1)}(t; F^{(1)})\} + \{\xi^{(2)}(t; F^{(2)})\} + \cdots$$

will be convergent, and will constitute a purely random process with this convolution for its principal distribution function.

In Theorem 4, the convolution of two distribution functions $F_1(u)$ and $F_2(u)$ is given by

$$G(u) = F_1(u) \otimes F_2(u) = \int_{-\infty}^{\infty} F_1(u - x) \cdot dF_2(x)$$

7.8 From §**7.4**, a stationary process $\{\xi(t)\}$ will be obtained by taking

$$\xi(t) = b_0\eta(t) + b_1\eta(t-1) + \cdots + b_h\eta(t-h) \tag{7.11}$$

Here $\{\eta(t)\}$ represents a purely random process and $(b) = (b_0, b_1, \ldots, b_h)$ an arbitrary sequence of real numbers. Equation (7.11) defines a *process of moving averages* with $\{\eta(t)\}$ known as the *primary* process. Usually the identifying assumption is made that $b_0 = 1$. The principal distribution functions $F_\xi(u)$ and $F_\eta(u)$ are connected by

$$F_\xi(u) = F_\eta(u/b_0) \otimes F_\eta(u/b_1) \otimes \cdots \otimes F_\eta(u/b_h)$$

and, as long as the variance of $\{\eta(t)\}$, $\sigma^2(\eta)$, is finite,

$$\sigma^2(\xi) = (b_0^2 + b_1^2 + \cdots + b_h^2)\sigma^2(\eta)$$

If it is assumed that $E(\eta) = 0$ and, as $h \to \infty$, the real sequence (b) is such that $\sum_{k=0}^{\infty} b_k^2$ is convergent, then (7.11) extends to

$$b_0\eta(t) + b_1\eta(t-1) + b_2\eta(t-2) + \cdots \tag{7.12}$$

It follows from the independence of $\eta(t)$ that the variance of

$$b_n\eta(t-n) + b_{n+1}\eta(t-n-1) + \cdots + b_{n+p}\eta(t-n-p)$$

is given by

$$(b_n^2 + b_{n+1}^2 + \cdots + b_{n+p}^2)\sigma^2(\eta)$$

and thus tends to zero uniformally in p as $n \to \infty$. Accordingly, (7.12) is convergent and, from Theorem 1, the stationary process $\{\xi(t)\}$ may be defined as

$$\xi(t) = b_0\eta(t) + b_1\eta(t-1) + b_2\eta(t-2) + \cdots$$

Wold stated that this is the general formula for a *process of linear regression*.

7.9 Now let $(a) = (a_1, \ldots, a_h)$ be a set of real numbers such that $a_h \neq 0$ and for which the roots of the *characteristic equation*

$$z^h + a_1 z^{h-1} + \cdots + a_{h-1}z + a_h = 0$$

all have modulus less than 1. Let $(b) = (b_1, b_2, \ldots)$ be a sequence such that the difference equation

$$x(t) + a_1 x(t-1) + \cdots + a_h x(t-h) = 0$$

is satisfied when $x_t \equiv b_t$ and where the initial values $b_1, \ldots, b_h$ are solutions of the following system of linear equations

$$\begin{aligned}
&a_1 + b_1 = 0 \\
&a_2 + a_1 b_1 + b_2 = 0 \\
&\vdots \\
&a_{h-1} + a_{h-2}b_1 + \cdots + a_1 b_{h-2} + b_{h-1} = 0 \\
&a_h + a_{h-1}b_1 + \cdots + a_1 b_{h-1} + b_h = 0
\end{aligned} \tag{7.13}$$

The b_i are seen to be real and uniquely determined and, if $\sum_{k=1}^{\infty} b_k^2$ is convergent, a stationary process will be defined by

$$\xi(t) = \eta(t) + b_1\eta(t-1) + b_2\eta(t-2) + \cdots \tag{7.14}$$

for purely random $\{\eta(t)\}$ with finite variance. Since $\{\xi(t)\}$ is stationary, so also will be

$$\zeta(t) = \xi(t) + a_1\xi(t-1) + \cdots + a_h\xi(t-h)$$

Wold (*ibid.*, page 54) then showed that $\{\zeta(t)\}$ and $\{\eta(t)\}$ are equivalent, so that

$$\xi(t) + a_1\xi(t-1) + \cdots + a_h\xi(t-h) = \eta(t) \tag{7.15}$$

which implies that the variables $\xi(t)$, $\xi(t-1), \ldots, \xi(t-h)$ are connected by a 'relation of linear regression', with (7.15) then defining a *process of (linear) autoregression*.

7.10 A stationary and singular process given by

$$\xi(t) - \xi(t-h) = 0$$

is called a *periodic process* and any sample series will be strictly periodic with period h. If $\{\xi^{(1)}(t)\}, \ldots, \{\xi^{(k)}(t)\}$ are independent stationary processes then the sum $\{\xi(t)\} = \{\xi^{(1)}(t)\} + \cdots + \{\xi^{(k)}(t)\}$ will constitute a stationary process. If at least one of the processes $\{\xi^{(i)}(t)\}$ is a periodic process, or a process of superposed harmonics, then $\{\xi(t)\}$ will be called a *process of hidden periodicities*.

7.11 Wold (*ibid.*, pages 60–5) termed a variable $\xi(t, t-1, \ldots)$ connected with a stationary process $\{\xi(t)\}$ *normal* if it had the *characteristic function*

$$f(X_t, X_{t-1}, \ldots) = \exp\left(i\mu \sum_{p=0}^{\infty} X_{t-p} - \frac{\sigma^2}{2} \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} r_{|p-q|} X_{t-p} X_{t-q} \right)$$

where $\mu, \sigma > 0$ and the autocorrelation coefficients r_k are real and satisfy (7.9), i.e., $\Delta(r, n) \geq 0$. This characteristic function has the normal distribution for its principal distribution function.

From Theorem 2, a singular normal process satisfying (7.10) will exist if the autocorrelation coefficients of any superposed harmonic are such that

$$\sum_{p=0}^{n} \sum_{q=0}^{n} r_{|p-q|} X_{t-p} X_{t-q} \geq 0$$

for any n and for the real sequence $(X) = (X_t, X_{t-1}, X_{t-2}, \ldots)$.

Wold used this result to show that Slutzky's Law of the Sinusoidal Limit (as discussed in §**5.16** and extended by another Russian mathematician, Romanovsky, 1932, 1933) may be verified by analogy to the properties of singular normal processes in that certain sections of stationary processes which, in the limit, satisfy singularity restrictions, will approximate superposed harmonics with the sections having the same period but varying amplitudes and phases: see §**7.20** for further details.

Autocorrelation coefficients as Fourier constants

7.12 Wold's next theorem (*ibid.*, page 66) related the autocorrelation coefficients r_k to the Fourier coefficients of a non-decreasing function.

Theorem 5.
Let $r_k(k=0,\pm 1,\pm 2,\dots)$ be an arbitrary sequence of constants. A necessary and sufficient condition that there exists a discrete stationary process with the r_ks for autocorrelation coefficients is that the r_ks are the FOURIER *coefficients of a non-decreasing function, say $W(x)$, such that $W(0)=0$; $W(\pi)=\pi$,*

$$r_k = \frac{1}{\pi}\int_0^{\pi} \cos kx \cdot dW(x)$$

The 'inversion formula' which allows $W(x)$ to be uniquely determined by the autocorrelation coefficients is

$$W(x) = x + 2\sum_{k=1}^{\infty} \frac{r_k}{k} \sin kx$$

This formula, called *the generating function* of the r_k, has a structure given by the following corollary to Theorem 5.

Corollary
Let $\{\xi(t)\}$ be a stationary process with autocorrelation coefficients r_k such that $\sum_{k=1}^{\infty}|r_k|$ is convergent. Then $W(x)$ will be absolutely continuous and will have a derivative $W'(x)$ that is bounded in modulus and given by

$$W'(x) = \sum_{k=-\infty}^{\infty} r_k \cos kx, \quad 0 \le x < \pi$$

Linear autoregression analysis of the discrete stationary process

7.13 Wold (*ibid.*, pages 75–80) showed that the variable $\xi(t)$ connected with the stationary process $\{\xi(t)\}$ may be approximated by $\xi(t-1),\dots,\xi(t-n)$, with the approximating error, termed the *residual*, being given by

$$\eta(t;n) = \xi(t) - \mu - a(1,n)\cdot(\xi(t-1)-\mu) - \cdots - a(n,n)\cdot(\xi(t-n)-\mu)$$

Here $\{\xi(t)\}$ has mean μ and principal correlation determinants $\Delta(r,n)$ given by (7.9). It is also assumed to have finite variance $\sigma^2(\xi)$ and, if $\Delta(r,n-1)\neq 0$, this variance will satisfy the inequalities

$$\sigma^2(\xi) \ge \sigma^2(\eta(n)) = \sigma^2(\xi)\frac{\Delta(r,n)}{\Delta(r,n-1)} \ge 0$$

which implies that

$$1 \ge \frac{\Delta(r,1)}{1} \ge \frac{\Delta(r,2)}{\Delta(r,1)} \ge \cdots \ge \frac{\Delta(r,n)}{\Delta(r,n-1)} \ge 0$$

From the analysis of §**7.6**, this implies that either $\Delta(r,n) > 0$ for all n, or $\Delta(r,n) > 0$ for $n < h$ and $\Delta(r,n) = 0$ for $n \geq h$, where h is the rank of linear singularity. It must therefore be the case that any stationary process belongs to one, and only one, of the following classes:

(I) The process is non-singular, and there exists a positive constant $\chi^2 \leq 1$ such that

$$\frac{\sigma^2(\eta(n))}{\sigma^2(\xi)} = \frac{\Delta(r,n)}{\Delta(r,n-1)} \to \chi^2 \leq 1 \quad \text{as } n \to \infty$$

(II) The process is singular, say of rank h.

(III) The process presents no singularity of finite rank, but

$$\frac{\sigma^2(\eta(n))}{\sigma^2(\xi)} = \frac{\Delta(r,n)}{\Delta(r,n-1)} \to 0 \quad \text{as } n \to \infty$$

in which case the process is termed *singular of infinite rank*.

This led Wold (*ibid.*, pages 81–4) to prove for case (**I**)

Theorem 6.
A residual process $\{\eta(t)\}$ *obtained from a non-singular stationary process* $\{\xi(t)\}$ *is stationary and non-autocorrelated. The variable* $\eta(t)$ *is non-correlated with* $\xi(t-1), \xi(t-2), \ldots,$ *while*

$$r(\xi(t), \eta(t)) = \frac{\sigma(\eta)}{\sigma(\xi)}$$

This theorem also holds for cases (**II**) and (**III**) except that, as the residual variables $\eta(t)$ are then vanishing, their correlation properties will be indeterminate.

A canonical form of the discrete stationary process

7.14 The analysis of §**7.13** enabled Wold (*ibid.*, pages 84–9) to prove the most fundamental theorem in time series analysis, which has since become known as *Wold's Decomposition.*[4]

Theorem 7.
Denoting by $\{\xi(t)\}$ *an arbitrary discrete stationary process with finite dispersion, there exists a three-dimensional stationary process* $\{\psi(t), \zeta(t), \eta(t)\}$ *with the following properties:*

(*A*) $\{\xi(t)\} = \{\psi(t)\} + \{\zeta(t)\}$
(*B*) $\{\psi(t)\}$ *and* $\{\zeta(t)\}$ *are non-correlated.*

(*C*) $\{\psi(t)\}$ *is singular.*
(*D*) $\{\eta(t)\}$ *is non-autocorrelated, and* $E[\eta(t)] = E[\zeta(t)] = 0$.
(*E*) $\{\zeta(t)\} = \{\eta(t)\} + b_1\{\eta(t-1)\} + b_2\{\eta(t-2)\} + \cdots$
where b_n *represent real numbers such that* $\sum b_n^2$ *is convergent.*

Thus, by starting from a purely random process $\{\eta(t)\}$, forming a sum process of the type $\{\zeta(t)\} = \{\eta(t)\} + b_1\{\eta(t-1)\} + b_2\{\eta(t-2)\} + \cdots$ and adding an independent process $\{\psi(t)\}$ ruled by an appropriate stationarity, an arbitrarily prescribed stationary process $\{\xi(t)\}$ is obtained (one of $\{\psi(t)\}$ and $\{\zeta(t)\}$ may be vanishing). The implications and importance of this theorem will be seen throughout later developments.

Stochastical difference equations

7.15 As already stated, Wold referred to equation (7.8) of §**7.6** as a *stochastical difference relation.* A more general representation is the linear autoregression (7.15), rewritten here as

$$\{\xi(t)\} + a_1\{\xi(t-1)\} + \cdots + a_h\{\xi(t-h)\} = \{\eta(t)\} \tag{7.16}$$

which Wold (*ibid.*, page 93) termed a *stochastical difference relation* between the processes $\{\xi(t)\}$ and $\{\eta(t)\}$. If $\{\eta(t)\}$ is known and $\{\xi(t)\}$ is unknown, (7.16) is termed a *stochastical difference equation*.

Equation (7.16) presents obvious analogies with ordinary difference equations of the form

$$x(t) + a_1x(t-1) + \cdots + a_hx(t-h) = y(t) \tag{7.17}$$

If there are no 'external influences' present, so that $y(t) = 0$, the solutions to (7.17) describe how $x(t)$ develops through time from the initial values $x(t-1) = x_{t-1}, \ldots, x(t-h) = x_{t-h}$, say, so that the expected value for $x(t)$ is $-a_1x_{t-1} - \cdots - a_hx_{t-h}$. If there is an external influence this expected value becomes $y(t) - a_1x_{t-1} - \cdots - a_hx_{t-h}$, with $y(t)$ being regarded as functional, i.e., uniquely determined at any future time point.

In contrast, the stochastical approach assumes that the external factors $\{\eta(t)\}$ are only ruled by certain probability laws: although $\{\eta(t)\}$ must follow the consistency relations (7.1) and (7.2), it could be a non-random or even a non-stationary process. If the $\{\eta(t)\}$ process is known, any sample series $(\ldots, \eta_{t-1}, \eta_t, \eta_{t+1}, \ldots)$ will describe an actual realization of the external influence. This will then determine the sample series $(\ldots, \xi_{t-1}, \xi_t, \xi_{t+1}, \ldots)$ of the process $\{\xi(t)\}$. However, typically only probabilistic knowledge of the actual path

$(\ldots, \eta_{t-1}, \eta_t, \eta_{t+1}, \ldots)$ will be available and so only a probability law about the behaviour of $\{\xi(t)\}$ can be reached, so that $(\ldots, \xi_{t-1}, \xi_t, \xi_{t+1}, \ldots)$ forms a solution to the stochastical difference equation (7.16).

The probability laws will provide information on the 'average' behaviour of $\{\xi(t)\}$ but Wold was careful to point out that it will not generally be the case that such behaviour will be identical to the solution of 'functional' difference equations for which there are no external influences.

Wold briefly contrasted the solutions of the stationary linear autoregression (7.16) with those from the process

$$\{\xi(t)\} - \{\xi(t-1)\} = \eta(t)$$

This is referred to as a *discrete homogenous process* which is *evolutive*, having solutions that are oscillatory with amplitude increasing over time.

The process (7.16) will be a stochastical difference equation if the coefficients a_i are real and if $\{\eta(t)\}$ is a discrete random process with finite variance. If $a_h \neq 0$ the equation is said to be of *order h*. If $\{\eta(t)\} = 0$ then, if (7.16) has any solutions, these will be singular and will have (7.8) with $\mu = 0$ as the relation of singularity. There will therefore be a non-vanishing stationary process that satisfies this equation if the related characteristic equation (see §**7.9**) has roots with modulus less than one. It also follows that

Theorem 8.
The series b_i defined by [7.12] *and* [7.13] *does not satisfy any difference equation of lower order than h.*

Theorem 9.
Let [7.16] *be a stochastical difference equation such that all roots of its characteristic equation are of a modulus less than unity, let $\{\eta(t)\}$ be stationary and have finite dispersion, and let the sequence b_i be given by* [7.12] *and* [7.13]. *Then*

$$\lim_{n\to\infty} [\{\eta(t)\} + b_1\{\eta(t-1)\} + b_2\{\eta(t-2)\} + \cdots + b_n\{\eta(t-n)\}]$$

will exist and forms a stationary solution of the equation.

The process of linear autoregression is thus, by construction, a solution of a stochastical difference equation such that $\{\eta(t)\}$ is stationary and all roots of the characteristic equation have modulus less than unity. Theorem 9 states that a mechanism whose intrinsic movements are damped will give rise to stationary oscillations when influenced by stationary external shocks.

7.16 Suppose that $\{\zeta(t)\}$ and $\{\eta(t)\}$ are two stationary processes such that

$$\zeta(t) = \eta(t) + b_1\eta(t-1) + b_2\eta(t-2) + \cdots \tag{7.18}$$

$$\eta(t) = \zeta(t) + a_1\zeta(t-1) + a_2\zeta(t-2) + \cdots \tag{7.19}$$

where

$\{\eta(t)\}$ is non-autocorrelated;
$\sigma^2(\eta, t) > 0$ is finite;
$E[\eta(t)] = 0$;
the sum $\sum b_k^2$ is convergent.

Substituting (7.18) into (7.19) obtains

$$\begin{aligned}\eta(t) = {} & \eta(t) + (a_1 + b_1)\zeta(t-1) + (a_2 + b_1a_1 + b_2)\zeta(t-2) + \cdots \\ & + (a_k + b_1a_{k-1} + \cdots + b_{k-1}a_1 + b_k)\zeta(t-k) + \cdots\end{aligned}$$

so that

$$a_k + b_1a_{k-1} + \cdots + b_{k-1}a_1 + b_k = 0, \quad k = 1, 2, \ldots$$

Writing

$$K^2 = 1 + b_1^2 + b_2^2 + \cdots$$

we thus have

$$\sigma(\zeta) = K \cdot \sigma(\eta)$$

$$K \cdot r(\zeta(t+n); \eta(t)) = b_n$$

$$r_k = r_k(\zeta) = (b_k + b_1b_{k+1} + b_2b_{k+2} + \cdots)/K^2$$

Consider next rewriting (7.18) and (7.19) as

$$\zeta(t+s) = \eta(t+s) + b_1\eta(t+s-1) + b_2\eta(t+s-2) + \cdots$$

$$\zeta(t) + a_1\zeta(t-1) + a_2\zeta(t-2) + \cdots = \eta(t)$$

multiplying them together and taking expectations to obtain, for $s < 0$,

$$r_k + a_1r_{k-1} + \cdots + a_{k-1}r_1 + a_k + a_{k+1}r_1 + a_{k+2}r_2 + \cdots = 0$$

for all $k > 0$. In the same way, for $s \geq 0$,

$$(1 + a_1r_1 + a_2r_2 + \cdots)\sigma^2(\zeta) = \sigma^2(\eta)$$

and

$$r_k + a_1 r_{k+1} + a_2 r_{k+2} + \cdots = b_k/K^2 \quad k \geq 0$$

Forecasting with autoregressions

7.17 Wold (*ibid.*, pages 101–3) considered forecasting using the linear autoregressive process, taking the variable $\zeta(t+k)$ to be conditioned by the development of the processes (7.18) and (7.19) up to time point t inclusive. Thus

$$\zeta(t-k) = \zeta_{t-k}, \quad \eta(t-k) = \eta_{t-k}; \quad k = 0, 1, 2, \ldots$$

where $(\zeta_t, \zeta_{t-1}, \ldots)$ and $(\eta_t, \eta_{t-1}, \ldots)$ are observed sample series, and hence

$$\begin{aligned} \zeta_{t-k} &= \eta_{t-k} + b_1 \eta_{t-k-1} + b_2 \eta_{t-k-2} + \cdots \quad k = 0, 1, 2, \ldots \\ \eta_{t-k} &= \zeta_{t-k} + a_1 \zeta_{t-k-1} + a_2 \zeta_{t-k-2} + \cdots \quad k = 0, 1, 2, \ldots \end{aligned} \tag{7.20}$$

Since the variables $\eta(t)$ are uncorrelated, the linear forecast of $\zeta(t+k)$ made at time t, denoted $F_t[\zeta(t+k)]$, is given by (7.18) as

$$F_t[\zeta(t+k)] = b_k \eta_t + b_{k+1} \eta_{t-1} + b_{k+2} \eta_{t-2} + \cdots \quad k = 1, 2, \ldots \tag{7.21}$$

Equivalently, from (7.19) the forecast can be written, for $k = 1, 2, \ldots$, as

$$\begin{aligned} F_t[\zeta(t+k)] = & -a_1 \cdot F_t[\zeta(t+k-1)] - a_2 \cdot F_t[\zeta(t+k-2)] - \cdots \\ & -a_{k-1} \cdot F_t[\zeta(t+1)] - a_k \zeta_t - a_{k+1} \zeta_{t-1} - a_{k+2} \zeta_{t-2} - \cdots \end{aligned} \tag{7.22}$$

This form shows how successive forecasts may be calculated. It can be shown to be equivalent to (7.21) by writing every $F_t[\zeta(t+k-i)]$ in the above expression in the form implied by (7.22) and expressing every ζ_{t-i} in terms of η_{t-i} by means of (7.20). Alternatively, the forecasts may be expressed in terms of ζ_{t-i}:

$$F_t[\zeta(t+k)] = f_{k,0} \zeta_t + f_{k,1} \zeta_{t-1} + f_{k,2} \zeta_{t-2} + \cdots \tag{7.23}$$

For $k = 1$, (7.22) becomes

$$F_t[\zeta(t+1)] = -a_1 \zeta_t - a_2 \zeta_{t-1} - a_3 \zeta_{t-2} - \cdots$$

so that

$$f_{1,i} = -a_{i+1}$$

Substituting (7.23) into (7.22) obtains

$$\begin{aligned} &f_{k,0} + a_1 f_{k-1,0} + a_2 f_{k-2,0} + \cdots + a_{k-1} f_{1,0} + a_k = 0 \\ &f_{k,1} + a_1 f_{k-1,1} + a_2 f_{k-2,1} + \cdots + a_{k-1} f_{1,1} + a_{k+1} = 0 \\ &\cdots \end{aligned}$$

Thus, after having calculated the coefficients $f_{k-i,j}$ appearing in the forecasts $F_t[\zeta(t+k-i)]$, these relations yield the coefficients $f_{k,j}$ necessary for computing $F_t[\zeta(t+k)]$ in terms of the ζ_{t-i}s.

The relations (7.21)–(7.23) are referred to by Wold as the *forecasting formulae*. Given the sample series $(\zeta_t, \zeta_{t-1}, \zeta_{t-2}, \ldots)$ and/or $(\eta_t, \eta_{t-1}, \eta_{t-2}, \ldots)$, 'these formulae furnish the best linear forecast as to the future development of the series' (*ibid.*, page 102). 'Best' is used in the sense that the expected squared error of the forecast, which, from (7.21), is $(1 + b_1^2 + b_2^2 + \cdots + b_{k-1}^2)\sigma^2(\eta)$, is shown to be a minimum. As $k \to \infty$, this expression tends to $K^2\sigma^2(\eta) = \sigma^2(\zeta)$, so that for large values of k, the forecast $F_t[\zeta(t+k)]$ is approximately of the same efficiency as the trivial forecast $E[\zeta(t+k)] = E[\zeta(t)] = 0$.

Linear autoregressions

7.18 The linear autoregression process $\{\zeta(t)\}$ is

$$\{\zeta(t)\} + a_1\{\zeta(t-1)\} + \cdots + a_h\{\zeta(t-h)\} = \{\eta(t)\} \tag{7.24}$$

where the stationary process $\{\eta(t)\}$ is non-autocorrelated and $E[\eta(t)] = E[\zeta(t)] = 0$. From the expressions at the end of §**7.16**, we have

$$(1 + a_1 r_1 + a_2 r_2 + \cdots + a_h r_h)\sigma^2(\zeta) = \sigma^2(\eta)$$

and three groups of relations involving the autocorrelation coefficients:

$$\begin{cases} \cdots \\ r_k + a_1 r_{k-1} + a_2 r_{k-2} + \cdots + a_{h-1} r_{k-h+1} + a_h r_{k-h} = 0 \\ \cdots \\ r_{h+1} + a_1 r_h + a_2 r_{h-1} + \cdots + a_{h-2} r_3 + a_{h-1} r_2 + a_h r_1 = 0 \\ r_h + a_1 r_{h-1} + a_2 r_{h-2} + \cdots + a_{h-2} r_2 + a_{h-1} r_1 + a_h = 0 \end{cases} \tag{7.25}$$

$$\begin{cases} r_{h-1} + a_1 r_{h-2} + a_2 r_{h-3} + \cdots + a_{h-2} r_1 + a_{h-1} + a_h r_1 = 0 \\ \cdots \\ r_1 + a_1 + a_2 r_1 + \cdots + a_{h-2} r_{h-3} + a_{h-1} r_{h-2} + a_h r_{h-1} = 0 \end{cases} \tag{7.26}$$

$$\begin{cases} 1 + a_1 r_1 + a_2 r_2 + \cdots + a_{h-2} r_{h-2} + a_{h-1} r_{h-1} + a_h r_h = 1/K^2 \\ r_1 + a_1 r_2 + a_2 r_3 + \cdots + a_{h-1} r_h + a_h r_{h+1} = b_1/K^2 \\ \cdots \\ r_k + a_1 r_{k+1} + a_2 r_{k+2} + \cdots + a_{h-1} r_{h+k-1} + a_h r_{h+k} = b_k/K^2 \\ \cdots \end{cases} \tag{7.27}$$

The group (7.25) is given in Walker (1931): see §**6.10**, equation (6.15). The r_k for $k \geq h$ satisfy a difference equation which is the same as that satisfied by the b_k sequence and both evolve as damped oscillations.

The second group (7.26) contains $h-1$ relations and involves $r_1, r_2, \ldots, r_{h-1}$, which may be obtained directly from the a_i by solving the system (7.26), and may be regarded as a corollary to

Theorem 10.
Let $\{\zeta(t)\}$ *be a process of linear autoregression of order h. Then the autocorrelation coefficients of* $\{\zeta(t)\}$ *satisfy no difference equation (cf.* [7.8]) *of lower order than h.*

Since $a_k = 0$ for $k > h$, the forecasting formula (7.22) shows that the forecasts $F_t[\zeta(t+1)], F_t[\zeta(t+2)], \ldots, F_t[\zeta(t+k)], \ldots$ will satisfy a difference equation with respect to k, so that the forecasts will also form a damped oscillation, revealing how the series will evolve from time t if there were no external influences present at the future times $t+1, t+2, \ldots$. Consequently, $F_t[\zeta(t+k)] \to 0 = E[\zeta(t)]$ as $k \to \infty$, in agreement with the concluding remark of §**7.17**.

By referring to Theorem 5 of §**7.12**, Wold next obtained

Theorem 11 (abridged)
The generating function $W(x)$ *of the autocorrelation coefficients in a process* $\{\zeta(t)\}$ *of linear autoregression is absolutely continuous, and has a bounded derivative* $W'(x)$ *given by* $W'(x) = G(x) + G(-x) - 1$, where

$$G(x) = \frac{1 + (a_1 + r_1)e^{ix} + \cdots + (a_{h-1} + a_{h-2}r_1 + \cdots + a_1 r_{h-2} + r_{h-1})e^{i(h-1)x}}{1 + a_1 e^{ix} + a_2 e^{i2x} + \cdots + a_h e^{ihx}}$$

7.19 Wold (*ibid.*, pages 110–21) analysed in detail the autoregression (7.24) when $h = 2$. Denoting the roots of the associated characteristic equation (cf. §**7.9**) as p and q, we then have

$$a_1 = p + q, \quad a_2 = pq, \quad a_n = 0 \quad \text{for } n > 2; \ \ |p| < 1, |q| < 1$$

and thus

$$\zeta(t) - (p+q)\zeta(t-1) + pq\zeta(t-2) = \eta(t) \tag{7.28}$$

As a_1 and a_2 must be real, two possibilities exist: either I: p and q are real, or II: $p = A + iB$, $q = A - iB$, where A and B represent real numbers such that

$$A^2 + B^2 = |p^2| = |q^2| < 1$$

Assuming that $p \neq q$, the general solution of the difference equation obtained from (7.28) with $\eta(t) = 0$ is, for P_1 and P_2 arbitrary,

$$P_1 \cdot p^t + P_2 \cdot q^t \tag{7.29}$$

For case II, the solution is

$$Q_1 \cdot C^t \cos \lambda t + Q_2 \cdot C^t \sin \lambda t$$

where Q_1 and Q_2 are arbitrary and

$$C = +\sqrt{A^2 + B^2} \quad \cos \lambda = A/C, \quad 0 < \lambda < \pi$$

I. p and q are real.

Substituting the general solution (7.29) for b_1 and b_2 into the system (7.13) and solving for P_1 and P_2 obtains

$$b_k = \frac{p}{p-q} \cdot p^k + \frac{q}{q-p} \cdot q^k = \frac{p^{k+1} - q^{k+1}}{p-q} \quad k \geq 0$$

Inserting this result into the expression for K^2 yields

$$K^2 = \frac{\sigma^2(\zeta)}{\sigma^2(\eta)} = \frac{1 + pq}{(1 - p^2)(1 - q^2)(1 - pq)}$$

The system (7.26) reduces to the single equation

$$r_1 + a_1 + a_2 r_1 = 0$$

Solving for

$$r_1 = \frac{p + q}{1 + pq}$$

observing that $r_0 = 1$, and equating these two coefficients to (7.29) for $t = 0$, 1 obtains

$$r_k = \frac{p(1 - q^2)}{(p - q)(1 + pq)} \cdot p^k + \frac{q(1 - p^2)}{(q - p)(1 + pq)} \cdot q^k \quad k \geq 0$$

Two special cases are worth considering. If $q=0$ then the various relations reduce to

$$\zeta(t) - p\zeta(t-1) = \eta(t)$$

$$b_k = r_k = p^k, \quad k \geq 0 \quad \sigma^2(\zeta) = \frac{\sigma^2(\eta)}{1-p^2}$$

These formulae cover the case $h=1$ and were discussed in Walker (1931). If $q=-p$, we have

$$\zeta(t) - p^2\zeta(t-2) = \eta(t)$$

$$b_{2k} = r_{2k} = p^{2k}, \quad b_{2k+1} = r_{2k+1} = 0, \quad k \geq 0 \quad \sigma^2(\zeta) = \frac{\sigma^2(\eta)}{1-p^4}$$

II. p and q are complex conjugates.

$$p = A + iB, \quad q = A - iB$$

Here we have, by a similar analysis,

$$\zeta(t) - 2A \cdot \zeta(t-1) + (A^2 + B^2) \cdot \zeta(t-2) = \eta(t)$$

$$b_k = C^k \cos k\lambda + \frac{A}{B} \cdot \sin k\lambda \tag{7.30}$$

$$r_k = C^k \cos k\lambda + \frac{A}{B} \cdot \frac{1-C^2}{1+C^2} C^k \sin k\lambda \tag{7.31}$$

$$\sigma^2(\zeta) = \frac{1+C^2}{(1-C^2)(1+C^4-2A^2+2B^2)} \cdot \sigma^2(\eta) \tag{7.32}$$

This set-up covers the case of an oscillatory mechanism whose intrinsic oscillations consist of a single damped harmonic with a frequency lying in the interval $0 < \lambda < \pi$ and a damping factor C^t. Wold then considered whether periodogram analysis would accurately uncover λ. If the roots $A \pm iB$ of the characteristic equation lie close to the periphery of the unit circle, so that the intrinsic oscillations are only slightly damped, then periodogram analysis will be able to discover the frequency of the intrinsic oscillation. The more heavily damped the intrinsic oscillation is, however, the larger will the bias be in estimating λ, with periodogram analysis overestimating the intrinsic period if this is above 4 time units and underestimating it if it is between 2 and 4 units. Wold (*ibid.*, page 117) summed up these conclusions by stating that 'the situation may be described by saying that the inference drawn from the characteristic equation of the intrinsic oscillations does not apply directly to the oscillations of the mechanism when influenced by random external factors'.

7.20 Wold used the linear autoregression of order two to reveal a connection with the law of the sinusoidal limit (see §**5.15** and §**7.11**). Let

$$L(x) = x^2 - 2Ax + 1 = 0 \quad -1 < A < 1$$

be the characteristic equation of the simple harmonic

$$P_1 \cos \lambda_1 t + P_2 \sin \lambda_2 t$$

and let $\{\zeta^{(1)}(t)\}, \{\zeta^{(2)}(t)\}, \ldots$ be a sequence of autoregressions of the form

$$\zeta^{(p)}(t) - 2A_p \cdot \zeta^{(p)}(t-1) + C_p^2 \cdot \zeta^{(p)}(t-2) = \eta^{(p)}(t)$$

where

$$\lim_{p\to\infty} A_p = A, \quad \lim_{p\to\infty} C_p = 1$$

Using (7.32),

$$\sigma^2(\zeta^{(p)}) = K_p^2 \cdot \sigma^2(\eta^{(p)}) = \frac{1 + C_p^2}{(1 - C_p^2)(1 + C_p^4 - 2A_p^2 + 2B_p^2)} \cdot \sigma^2(\eta^{(p)})$$

where $K_p^2 = 1 + (b_1^{(p)})^2 + (b_2^{(p)})^2 + \cdots$. Since $1 - C_p^2$ tends to zero as $p \to \infty$, it follows that $K_p \to \infty$ as $p \to \infty$. Since, from (7.30), $b_k^{(p)}$ is bounded,

$$\lim_{p\to\infty} \frac{b_k^{(p)}}{K_p^2} = 0$$

Thus, the systems of equations (7.25–7.27) imply that

$$\lim_{p\to\infty} L(r_k^{(p)}) = 0, \quad -\infty < k < \infty$$

which, in turn, implies that the sequence $\{\zeta^{(p)}(t)\}$ is ruled by the singularities that embody the sinusoidal limit theorem (see §**7.11**). Thus, if we approximate an arbitrary sample series $(\zeta) = (\zeta_1^{(p)}, \ldots, \zeta_n^{(p)})$ by a simple harmonic with frequency λ, say $x_p(t, \zeta)$, then, holding n fixed, it follows that, for every $\varepsilon > 0$

$$\lim_{p\to\infty} P[|x_p(1,\zeta) - \zeta_1^{(p)}| < \varepsilon, \ldots, |x_p(n,\zeta) - \zeta_n^{(p)}| < \varepsilon] = 1$$

This result generalizes to a relation $L(x) = 0$ of arbitrary order, so that the process of linear autoregression 'forms a convenient starting point for the construction of sequences covered by the sinusoidal limit theorems' (*ibid.*, page 121).

Moving average processes

7.21 The general moving average of order h is

$$\{\zeta(t)\} = \{\eta(t)\} + b_1\{\eta(t-1)\} + \cdots + b_h\{\eta(t-h)\} \tag{7.33}$$

where $\{\eta(t)\}$ is purely random or, more generally, non-autocorrelated, and the sequence $(b) = (b_1, \ldots, b_h)$ is real. We continue to assume that $\sigma^2(\eta)$ is finite and that $E(\eta_t) = 0$. Unlike the process of linear autoregression, where the sequences of autocorrelation coefficients and forecasts follow damped harmonic processes, only the first h elements of these sequences are non-zero. The variance of $\{\zeta(t)\}$ is

$$\sigma^2(\zeta) = (1 + b_1^2 + b_2^2 + \cdots + b_h^2) \cdot \sigma^2(\eta)$$

while the autocorrelations are given by

$$r_k(\zeta) = \begin{cases} (b_k + b_1 b_{k+1} + \cdots + b_h b_{h+k})/(1 + b_1^2 + \cdots + b_h^2) & \text{for } k \le h \\ 0 & \text{for } k > h \end{cases} \tag{7.34}$$

where $k \ge 0$ and $b_0 = 1$. Specializing the forecast formula (7.21) gives

$$F_t[\zeta(t+k)] = \begin{cases} b_k \eta_t + b_{k+1}\eta_{t-1} + \cdots + b_h \eta_{t-h+k} & \text{for } 0 \le k \le h \\ 0 & \text{for } k > h \end{cases}$$

Given the moving average process (7.33), does a primary process of the form (7.19)

$$\eta(t) = \zeta(t) + a_1\zeta(t-1) + a_2\zeta(t-2) + \cdots$$

exist and, if so, how can the coefficients (a) be obtained? Consider the characteristic equation

$$x^h + b_1 x^{h-1} + \cdots + b_{h-1}x + b_h = 0 \tag{7.35}$$

If all the roots of (7.35) have modulus less than unity, Theorem 9 states that an infinite sequence $(a) = (a_1, a_2, \ldots)$ such that (7.19) holds is given by the system (7.13) and the difference relation of §**7.9** on replacing the a_is by the b_is. These relations constitute a difference equation of order h satisfied by the sequence (a), which forms a damped harmonic. Under these circumstances the relations of §**7.16** hold, and take the following form:

$$a_{2h+k}r_h + a_{2h+k-1}r_{h-1} + \cdots + a_{h+k+1}r_1 + a_{h+k} + a_{h+k-1}r_1 + \cdots + a_{k+1}r_{h-1} + a_k r_h = 0 \tag{7.36}$$

$$\begin{cases} a_{2h}r_h + a_{2h-1}r_{h-1} + \cdots + a_{h+1}r_1 + a_h + a_{h-1}r_1 + \cdots + a_1 r_{h-1} + r_h = 0 \\ a_{2h-1}r_h + a_{2h-2}r_{h-1} + \cdots + a_h r_1 + a_{h-1} + a_{h-2}r_1 + \cdots + a_1 r_{h-2} + r_{h-1} = 0 \\ \cdots \\ a_{h+1}r_h + a_h r_{h-1} + \cdots + a_2 r_1 + a_1 + r_1 = 0 \end{cases} \tag{7.37}$$

$$\begin{cases} a_h r_h + a_{h-1}r_{h-1} + \cdots + a_1 r_1 + 1 = 1/K^2 \\ a_{h-1}r_h + \cdots + a_1 r_2 + r_1 = b_1/K^2 \\ \cdots \\ a_2 r_h + a_1 r_{h-1} + r_{h-2} = b_{h-2}/K^2 \\ a_1 r_h + r_{h-1} = b_{h-1}/K^2 \\ r_h = b_h/K^2 \end{cases} \tag{7.38}$$

Thus, if (7.35) has no root x_k falling outside the unit circle, a set of well-defined linear operations on the moving average (7.33) will yield the primary process $\{\eta(t)\}$ given by (7.19): if $|x_k| \leq 1$ for all k, the sequence (b) and the process $\{\zeta(t)\}$ is termed *regular*.

Wold then uses the generating function $W(x)$ of §**7.12** to obtain the fundamental identity

$$\begin{aligned} &\frac{1}{K^2}(x^h + b_1 x^{h-1} + \cdots + b_{h-1}x + b_n)(b_h x^h + b_{h-1}x^{h-1} + \cdots + b_1 x + 1) \\ &= r_h x^{2h} + r_{h-1}x^{2h-1} + \cdots + r_1 x^{h+1} + x^h + r_1 x^{h-1} + \cdots + r_{h-1}x + r_h \end{aligned} \tag{7.39}$$

Since the zeros of the factor $b_h x^h + b_{h-1}x^{h-1} + \cdots + b_1 x + 1$ will be x_k^{-1}, the zeros of the right-hand side of (7.39) may be denoted $x_1, x_2, \ldots, x_{2h-1}, x_{2h}$, where

$$x_k = x_{2h+1-k}^{-1}, \quad 0 < |x_1| \leq |x_2| \leq \cdots \leq |x_h| \leq 1 \leq |x_{h+1}| \leq \cdots \leq |x_{2h}|$$

It then follows that if there exists another sequence, say $(1, b_1^{(i)}, \ldots, b_h^{(i)})$, such that the associated moving average has autocorrelation coefficients coinciding with those of (7.33), then one zero of the polynomial $x^h + b_1^{(i)}x^{h-1} + \cdots + b_{h-1}^{(i)}x + b_h^{(i)}$ will equal either x_1 or x_1^{-1}, another either x_2 or x_2^{-1}, etc. There will be at most 2^h real sequences of this type, say $(b_k^{(0)}), \ldots, (b_k^{(s)})$. If $(b_k^{(0)})$ represents the regular sequence then all other sequences are non-regular.

Letting $(b_k^{(i)})$ be a group of sequences attached to the regular sequence $(b_k^{(0)})$ and writing

$$(K^{(i)})^2 = 1 + (b_i^{(1)})^2 + (b_i^{(2)})^2 + \cdots + (b_i^{(h)})^2$$

we can then define a group of moving averages as

$$\zeta^{(i)}(t) = \frac{K^{(0)}}{K^{(i)}}[\eta(t) + b_1^{(i)}\eta(t-1) + \cdots + b_h^{(i)}\eta(t-h)] \quad i = 1, \ldots, s < 2^h \tag{7.40}$$

It follows from its construction that this group will contain one, and only one, regular process and that

$$\sigma^2(\zeta^{(i)}) = \sigma^2(\zeta^{(j)}); \quad r_k^{(i)} = r_k^{(j)}, \quad k = 0, \pm 1, \pm 2, \dots \quad i, j = 1, \dots, s < 2^h$$

If all the roots of (7.35) lie of the periphery of the unit circle, i.e., $x_k = 1$ for all k, then the group will contain just the process $\{\zeta^{(0)}(t)\} = \{\zeta(t)\}$; otherwise the group will contain at most 2^h. If we denote by $\{\eta^{(i)}(t)\}$ the residuals of the non-regular processes, then these are given by

$$\begin{aligned}\frac{K^{(i)}}{K^{(0)}}\eta^{(i)}(t) &= \eta(t) + (a_1 + b_1^{(i)}) \cdot \eta(t-1) + (a_2 + a_1 b_1^{(i)} + b_2^{(i)}) \cdot \eta(t-2) + \cdots \\ &\quad + (a_h + a_{h-1} b_1^{(i)} + \cdots + b_h^{(i)}) \cdot \eta(t-h) \\ &\quad + (a_{h+1} + a_h b_1^{(i)} + \cdots + a_1 b_h^{(i)}) \cdot \eta(t-h-1) + \cdots\end{aligned} \tag{7.41}$$

and we have

$$\{\zeta^{(i)}(t)\} = \{\eta^{(i)}(t)\} + b_1\{\eta^{(i)}(t-1)\} + \cdots + b_h\{\eta^{(i)}(t-h)\} \tag{7.42}$$

for all processes in the group $\{\zeta^{(i)}\}$.

These ideas may be illustrated by the following examples.

Example 1. Let $h = 1$ with $b_1 = 2$. Then (7.39) is

$$0.2(x+2)(2x+1) = 0.4x^2 + x + 0.4$$

Hence $r_1 = 0.4$ and $r_k = 0$ for $k > 1$. The characteristic equation is $(x+2)(2x+1) = 0$, which gives two sequences $b^0 = (1, 0.5)$ and $b^1 = (1, 2)$. The system (7.13) gives $a_1 = -0.5$ and, in general, $a_k = -0.5^k$. Thus, for the regular process b^0,

$$\{\zeta(t)\} = \{\eta(t)\} + 0.5\{\eta(t-1)\}$$

and it follows that

$$\{\eta(t)\} = \{\zeta(t)\} - 0.5\{\zeta(t-1)\} + 0.5^2\{\zeta(t-2)\} - 0.5^3\{\zeta(t-3)\} + \cdots$$

Since $K^2 = 1.25$ and $[K^{(1)}]^2 = 5$, (7.40) gives

$$\{\zeta^{(1)}(t)\} = 0.5\{\eta(t)\} + \{\eta(t-1)\}$$

while (7.41) gives

$$\eta^{(1)}(t) = \tfrac{1}{2}\eta(t) + \tfrac{3}{4}\eta(t-1) - \tfrac{3}{8}\eta(t-2) + \tfrac{3}{16}\eta(t-3) - \cdots$$

Thus, using (7.42), we have

$$\zeta^{(1)}(t) = \{\eta^{(1)}(t)\} + 0.5\{\eta^{(1)}(t-1)\} = 0.5\{\eta(t)\} + \{\eta(t-1)\}$$

from which it can easily be verified that $\sigma^2(\eta^{(1)}(t)) = \sigma^2(\eta(t))$ and that $r_k(\eta^{(1)}) = 0$ for $k \neq 0$, i.e., $\{\eta^{(1)}(t)\}$ is non-autocorrelated.

Example 2. Here we suppose that $r_1 = \frac{1}{6}$, $r_2 = -\frac{1}{3}$ and $r_k = 0$ for $k > 2$. The fundamental identity (7.39) reads

$$\tfrac{2}{3}(x^2 + 0.5x - 0.5)(-0.5x^2 + 0.5x + 1) = -\tfrac{1}{3}x^4 + \tfrac{1}{6}x^3 + x^2 + \tfrac{1}{6}x - \tfrac{1}{3}$$

Noting that $x^2 + 0.5x - 0.5 = (x-0.5)(x+1)$, so that neither root lies outside the unit circle, there are two sequences, $b^0 = (1, 0.5, -0.5)$, which is regular, and $b^1 = (1, -1, -2)$. Thus the regular process is defined as

$$\{\zeta(t)\} = \{\eta(t)\} + 0.5\{\eta(t-1)\} - 0.5\{\eta(t-2)\}$$

from which we obtain the (a) sequence as

$$a_k = \tfrac{1}{3}\left(\tfrac{1}{2}\right)^k + \tfrac{2}{3}(-1)^k$$

The non-regular process is given by

$$\begin{aligned}\{\zeta^{(1)}(t)\} &= 0.5\{\eta(t)\} - 0.5\{\eta(t-1)\} - \{\eta(t-2)\} \\ &= \{\eta^{(1)}(t)\} + 0.5\{\eta^{(i)}(t-1)\} - 0.5\{\eta^{(i)}(t-2)\}\end{aligned}$$

with the non-autocorrelated residual being

$$\eta^{(1)}(t) = \tfrac{1}{2}\eta(t) - \tfrac{3}{4}\eta(t-1) - \tfrac{3}{8}\eta(t-2) - \tfrac{3}{16}\eta(t-3)$$

Example 3. Finally, suppose $r_1 = -0.5$ and $r_k = 0$ for $k > 1$. Now (7.39) reads

$$0.5(x-1)(-x+1) = 0.5x^2 + x - 0.5$$

and we conclude that there is just one sequence $b = (1, -1)$, associated with the process

$$\{\zeta(t)\} = \{\eta(t)\} - \{\eta(t-1)\}$$

Here (7.35) has just one root, which falls on the unit circle. In these circumstances, Wold (*ibid.*, pages 124–6) showed that a limit process of the form

$$\{\eta(t)\} = \lim_{i\to\infty}[\{\zeta(t)\} + a_1^{(i)}\{\zeta(t-1)\} + a_2^{(i)}\{\zeta(t-2)\} + \cdots]$$

exists where, in this case $a_k^{(i)} = (1-\varepsilon)^k$, with $0 < \varepsilon \to 0$ as $i \to \infty$. Setting $\varepsilon = 10^{-i}$, then

$$\{\eta(t)\} = \lim_{i\to\infty}[\{\zeta(t)\} + (1-10^{-i})\{\zeta(t-1)\} + (1-10^{-i})^2\{\zeta(t-2)\} + \cdots]$$

which we see approaches the process

$$\{\eta(t)\} = \{\zeta(t)\} + \{\zeta(t-1)\} + \{\zeta(t-2)\} + \cdots$$

Some applications of stationary processes

7.22 In §§**6.12–6.13** we discussed Walker's (1931) analysis of the Port Darwin air pressure data, focusing on the complete 'correlation periodogram', which Wold (*ibid.*, page 135) referred to more succinctly as the correlogram – 'for the sake of brevity in writing, the graphs of serial and autocorrelation coefficients will be termed *correlograms* (*empirical* and *hypothetical* respectively)'. Wold focused attention on Walker's preliminary analysis of the first 40 serial coefficients, which showed that the r_k, $0 \le k \le 40$, could be approximately represented by the function

$$r_k = 0.19 \cdot 0.96^k \cos \pi k/6 + 0.15 \cdot 0.98^k + 0.66 \cdot 0.71^k \tag{7.43}$$

This function has a damped harmonic with a period of 12 quarters and satisfies the difference equation

$$r_k - 3.35r_{k-1} + 4.43r_{k-2} - 2.71r_{k-3} + 0.64r_{k-4} = 0$$

Walker then used the argument that, since (7.25) implies that

$$r_k + a_1 r_{k-1} + \cdots + a_h r_{k-h} = 0, \quad k \ge h$$

then the empirical series, $\ldots, \bar{\zeta}_{t-1}, \bar{\zeta}_t, \bar{\zeta}_{t+1}, \ldots$, follows the autoregression

$$\bar{\zeta}_t + a_1\bar{\zeta}_{t-1} + \cdots + a_h\bar{\zeta}_{t-h} = \bar{\eta}_t \tag{7.44}$$

i.e.,

$$\bar{\zeta}_t - 3.35\bar{\zeta}_{t-1} + 4.43\bar{\zeta}_{t-2} - 2.71\bar{\zeta}_{t-3} + 0.64\bar{\zeta}_{t-4} = \bar{\eta}_t \tag{7.45}$$

Wold (*ibid.*, pages 144–5) pointed out that such an argument was, in fact, invalid, as the autocorrelation coefficients not only satisfy (7.25) but also the systems (7.26) and (7.27): in fact, the coefficients $r_1, r_2, \ldots, r_{h-1}$ will be uniquely determined by (7.26) in terms of the a_is. Thus it is not certain that the autocorrelation coefficients corresponding to (7.45) will be given by the function (7.43). Wold showed that the system (7.26) corresponding to (7.45) gives the values $r_1 = 0.93$, $r_2 = 0.72$ and $r_3 = 0.43$, rather than the values 0.75, 0.55 and 0.35 given by Walker. Wold also showed that the relationship between the variance of the disturbances and the observed series given by Walker was incorrect, so that all the parameters of Walker's model required modification.

In fact, Wold suggested that the simpler model

$$\bar{\zeta}_t - 0.73\bar{\zeta}_{t-1} = \bar{\eta}_t \tag{7.46}$$

gave a good fit to the first few serial coefficients. Since this model is an example of case I of §**7.19** with $q = 0$, it gives the sequence of correlations $r_1 = 0.73$, $r_2 = 0.73^2 = 0.53$, $r_3 = 0.73^3 = 0.39$, $r_4 = 0.73^4 = 0.28$, compared to the actual air pressure serial coefficients of 0.76, 0.56, 0.36 and 0.18 respectively.

Turning his attention to the empirical correlogram of air pressure, shown in Figure 6.10, Wold remarked that

> the serial coefficients show rather small deviations from zero in the interval $3 < k < 40$. On the other hand, the increase in amplitude for certain k-values > 40 might be due to the successive reduction in the number of correlates. Perhaps this argument is sufficient to explain also why the fluctuations are somewhat larger in that alternative variant of a correlogram given by Walker, where all serial coefficients are based on 77 pairs of correlates. As the fluctuations, furthermore, seem rather irregular and aperiodic – at least to my eye – it is doubtful whether it would be possible to improve sensibly the approach [7.46] by taking into account more distant elements $\bar{\zeta}_{t-2}, \bar{\zeta}_{t-3}$, etc. In this connexion it is rather interesting to notice that according to the general analysis there exists no process of linear autoregression having [7.43] for autocorrelation coefficients. Another reason for resting satisfied with the simple approach [7.46] is that the ordinates of the periodogram presented [in Figure 6.9] are all lying on about the same level – this periodogram does not, like that of the sunspots, suggest a scheme of linear autoregression with a tendency to periodicity. (*ibid.*, pages 145–6)

Thus Wold suggested that a simpler, first-order linear autoregression presented the best fit to the Port Darwin air pressure data: a view that would hold that any tendency to periodicity in the series was of a spurious nature.[5]

7.23 A function of the form

$$y(t) = \mu + \sum_{k=1}^{s} C_k \cos(\lambda_k + \varphi_k) = \mu + \sum_{k=1}^{s} (A_k \cos \lambda_k t + B_k \sin \lambda_k t) \tag{7.47}$$

is referred to by Wold as a *composed harmonic*. Suppose that $\zeta(t) = y(t) + \eta(t)$, where $\eta(t)$ is purely random with variance $\sigma^2(\eta)$. This is known as the *scheme of hidden periodicities*. The autocorrelations of $\zeta(t)$ are, for $k \neq 0$, given by (Wold, equation 46)

$$r_k = \frac{\sum_{i=1}^{s} C_i^2 \cos \lambda_i k}{2\sigma^2(\eta) + \sum_{i=1}^{s} C_i^2} \tag{7.48}$$

so that r_k is also a composed harmonic and there exist arbitrarily large k-values such that

$$r_k \approx r_0 = \frac{\sum_{i=1}^{s} C_i^2}{2\sigma^2(\eta) + \sum_{i=1}^{s} C_i^2}$$

The implication of this is that even if an observed time series clearly shows a cyclical character but has serial coefficients that are gradually vanishing, then the scheme of hidden periodicities is inappropriate.

In contrast to (7.48), the correlogram of a linear autoregression will form a damped harmonic while that for a moving average will cut off beyond a certain k-value. These possibilities are illustrated in Figure 7.1, in which the

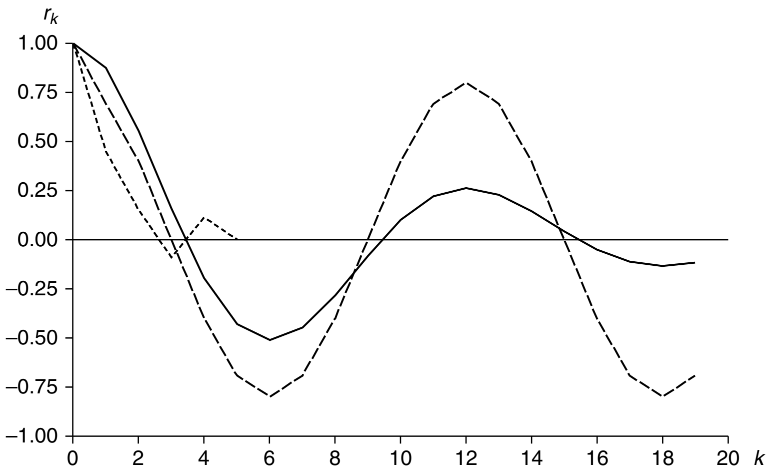

Figure 7.1 Correlograms illustrating the schemes of hidden periodicities (dashed line), linear autoregression (unbroken line), and moving average (dotted line)

correlograms are constructed in the following way. For the scheme of hidden periodicities, (7.48) was used with $s = C_1 = 1$, $\sigma^2(\eta) = 1.25$ and $\lambda_1 = \pi/6$. The linear autoregression uses (7.31) with $A = 0.8$, $B = 0.4$ and $\lambda = \pi/6$, while the moving average uses (7.34) with $b_1 = 0.7$, $b_2 = 0.4$, $b_3 = -0.3$ and $b_4 = 0.2$. Given the very different behaviour of the alternative schemes, Wold (*ibid.*, page 147) argued that 'it may be expected that we would obtain useful suggestions by inspecting the empirical correlogram when searching for an adequate scheme to be applied to an observational time series. For this reason, the construction of an empirical correlogram is taken as (a) starting point in ... applications'.

This led Wold to recommend the following approach to applying the theory developed above.

> If the empirical correlogram suggests a scheme of hidden periodicities, the next step would be to construct a periodogram for a more detailed analysis of possible periodicities in the material under investigation.
>
> Next, if the correlogram suggests a scheme of linear autoregression, our first problem is to find a scheme [7.15] such that the corresponding hypothetical correlogram will fit the empirical one. The chief difficulty is to derive suitable values for the coefficients (a) – when having arrived at a set of coefficients (a), the corresponding autocorrelation coefficients will be uniquely determined by the system [7.25–7.26], and the residuals $\overline{\eta}_t$ by the relations [7.44]. It is further a desideratum that these residuals be as small as possible. Having seen above that these problems are more intricate than emphasized in earlier studies of the graph of serial coefficients, it will be found that an empirical autoregression analysis as proposed by Yule (1927) will be useful in this connection.
>
> Finally, it may happen that the empirical correlogram will suggest a scheme of moving averages. As far as I know, the problem of fitting this scheme to observational data has not been attacked.... It will be seen that the relation [7.39] gives a starting point for attacking this problem. (*ibid.*, page 148)

7.24 Before embarking on applications, Wold took great pains to point out various limitations of the methodology. A major drawback was the lack of an inferential framework within which any results might be assessed –

> in time series analysis, significance problems are extremely intricate.... Consequently, all questions about the significance and the interpretation of the quantitative results fall outside the scope of this study, and again an explicit warning is given against attaching importance to the numerical values found for the parameters of the different models fitted to the observational data. (*ibid.*, pages 148–9)

A second question was that of identification: it will not be possible to distinguish between different schemes that give rise to the same set of autocorrelation coefficients. As an example of this, Wold considered the nonlinear function

$$\xi(t) = \eta(t) \cdot \eta(t-1)$$

in which $\eta(t)$ is, as usual, a zero mean random process. It is clear that $\xi(t)$ will be non-autocorrelated and hence indistinguishable from $\eta(t)$.

> Thus, if we have found a hypothetical scheme that fits well to an empirical correlogram, it is perfectly possible that there are other schemes which yield an equally close approximation. When it is necessary to choose between different schemes, it may happen that theoretical arguments will speak in favour of one of the schemes.. .. (T)he schemes of linear regression often seem plausible from theoretical viewpoints, at least to a first approximation. On the other hand, a rational choice between different schemes may be alternatively based on an examination of other structural properties of the time series than its serial coefficients. (*ibid.*, pages 149–50)

Moreover, if a process is actually generated by a nonlinear function, then restricting analysis to only linear autoregressions may lead to unduly complicated processes being arrived at.

An application of moving averages

7.25 Wold's first application was to analyse Beveridge's Index of Fluctuation, the periodogram of which was constructed in §**3.8**, in which he focused attention on the last 100 years of observations from 1770 to 1869. The correlogram for $0 \leq k \leq 15$ is shown in Figure 7.2, where it is observed that $\bar{r}_1 \approx 0.6$ and all following serial coefficients lie in the interval $-0.16 < \bar{r}_k < 0.13$, i.e., they are all rather close to zero, allowing Wold (*ibid.*, pages 151–2) to conclude that '(t)o my eye, the correlogram definitely suggests a scheme of moving averages', leading him to set out the following problem:

> A set of numbers $u_1, u_2, \ldots, u_h$ being given, does there exist a moving average [7.32] with autocorrelation coefficients r_k such that $r_k = u_k$ for $1 \leq k \leq h$? If the answer is in the affirmative, we know from [§**7.21**] that there in general will exist a finite group of moving averages with the prescribed autocorrelation coefficients, and we are also in possession of a direct method for determining the coefficients (b) of these moving averages. (*ibid.*, page 152)

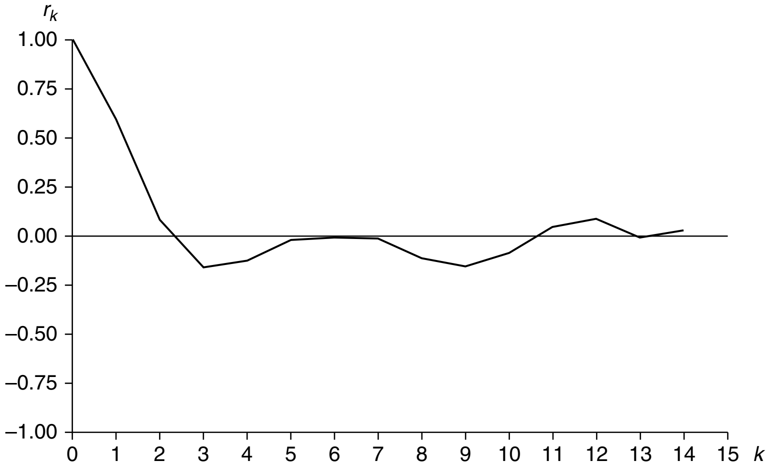

Figure 7.2 Correlogram of Beveridge's Index of Fluctuation, 1770–1869

Under these conditions (7.39) becomes

$$\begin{aligned} u(x) &= u_h x^h + u_{h-1}x^{h-1} + \cdots + u_1 x + 1 + \frac{u_1}{x} + \cdots + \frac{u_{h-1}}{x^{h-1}} + \frac{u_h}{x^h} \\ &= \frac{1}{K^2}(x^h + b_1 x^{h-1} + \cdots + b_{h-1}x + b_h)\left(b_h + \frac{b_{h-1}}{x} + \cdots + \frac{b_1}{x^{h-1}} + \frac{1}{x^h}\right) \end{aligned} \tag{7.49}$$

If x_0 is a root of $u(x)=0$ then so will be x_0^{-1}. Thus the substitution $z=x+x^{-1}$ will transform $u(x)$ to

$$v(z) = v_0 z^h + v_1 z^{h-1} + \cdots + v_{h-1}z + v_h \tag{7.50}$$

For example, with $h=3$, (7.49) becomes

$$\begin{aligned} u(x) &= u_3(x^3 + x^{-3}) + u_2(x^2 + x^{-2}) + u_1(x + x^{-1}) + 1 \\ &= u_3(z^3 - 3z) + u_2(z^2 - 2) + u_1 z + 1 \\ &= u_3 z^3 + u_2 z^2 + (u_1 - 3u_3)z + (1 - 2u_2) \end{aligned}$$

If z is a root of $v(z)=0$ then two roots of $u(x)=0$ will be obtained from the equation

$$P(x,z) = z - x - x^{-1} = x^2 - zx + 1 = 0$$

The roots of this equation are given by

$$\frac{z}{2} \pm \sqrt{\frac{z^2}{4} - 1}$$

Since the products of these roots is unity then, unless both roots have modulus unity, one of them must be situated inside, and the other outside, the unit circle.

If z is a complex root of $v(z) = 0$, then another root will be the complex conjugate of z, which we denote z^*. If $P(x, z) = 0$ has the roots x and x^{-1}, then $P(x, z^*) = 0$ will have the roots x^* and $(x^*)^{-1}$, in which case one of the real polynomials $(x - x_1)(x - x_1^*)$ and $(x - (x_1)^{-1})(x - (x_1^*)^{-1})$ must be a factor in the polynomial

$$b(x) = x^h + b_1 x^{h-1} + \cdots + b_{h-1} x + b_h$$

appearing in (7.49).

If z_0 is a real root of $v(z) = 0$, two cases need to be distinguished. If $|z_0| \geq 2$, both x_1 and x_2 must be real and this will correspond to real roots in $b(x) = 0$. If, on the other hand, $|z_0| < 2$ then x_1 and x_2 will be complex conjugates of modulus unity, and both $(x - x_1)$ and $(x - x_2)$ must be contained in $b(x)$. Since one zero of $u(x)$ corresponds to one zero of $v(z)$, this is impossible unless z_0 is a root of even multiplicity of $v(z) = 0$. The following theorem is thus obtained.

Theorem 12.
A necessary and sufficient condition that there exists a moving average [7.33] *with autocorrelation coefficients r_k equalling u_k for $1 \leq k \leq h$ is that the auxiliary polynomial $v(z)$ defined by* [7.50] *has no zero z_0 of odd multiplicity in the real interval $-2 < z_0 < 2$.*

If this condition is satisfied, the sequences (b) sought for will be given by the real polynomials $b(x)$ satisfying (7.49). There will at most be 2^h of these sequences and the polynomials $b(x)$ may be written in the form $(x - x_1)(x - x_2) \ldots (x - x_h)$, where the real or complex quantity x_i is a root of $P(x, z_i) = 0$, where the z_i, $i = 1, \ldots, h$, are the roots of $v(z) = 0$.

Thus, returning to the correlogram of the Index of Fluctuation, Wold assumed that the small deviations from zero of r_k for $k > 1$ were merely the product of chance fluctuations and asked whether there existed a moving average $\eta(t) + b_1\eta(t - 1)$ with autocorrelation coefficient r_1 equalling 0.595.[6] Putting $h = 1$ and $u_1 = 0.595$ into (7.49) obtains $u(x) = 0.595x + 1 + 0.595x^{-1}$ and $v(z) = 0.595z + 1$. Since the root $-0.595^{-1} = -1.68$ of $v(z) = 0$ lies in the critical interval $-2 < z < 2$, it must be concluded from Theorem 12 that there exists no moving average with $r_1 = 0.595$ and $r_k = 0$ for $k > 1$.

For z to lie outside the critical interval $-2 < z < 2$ it must therefore be the case that $-0.5 \leq r_1 \leq 0.5$: i.e., all moving averages of the form $\eta(t) + b_1\eta(t - 1)$

have $|r_1| \le 0.5$ and there will only be one moving average of this form for which $r_1 = 0.5$, namely

$$\zeta(t) - \mu = \eta(t) + \eta(t-1)$$

Consequently, this moving average will yield the closest fit to the prescribed value of 0.595, in which case the deviations of the serial correlation coefficients shown in Figure 7.2 from the values $r_1 = 0.5$, $r_2 = r_3 = \cdots = 0$ must be ascribed to pure chance.

To obtain a better fit, higher-order moving averages must be considered. Using the first two serial correlations, $u_1 = 0.595$ and $u_2 = 0.081$, gives

$$u(x) = 0.081x^2 + 0.595x + 1 + 0.595x^{-1} + 0.081x^{-2}$$

and

$$v(z) = 0.081z^2 + 0.595z + 0.838$$

Since the roots of $v(z) = 0$ are $z_1 = -1.90$ and $z_2 = -5.45$, it follows from Theorem 12 that no moving average of order 2 exists with these autocorrelation coefficients. To remove z_1 from the critical interval, u_2 will need to be modified. The general expression for $v(z_1)$ is

$$v(z_1) = u_2 z_1^2 + u_1 z_1 + (1 - 2u_2)$$

Putting $z_1 = -2$ into $v(z_1) = 0$ along with $u_1 = r_1 = 0.595$ yields the solution $u_2 = r_1 - \frac{1}{2} = 0.095$ with corresponding function $0.095v^2 + 0.595z + 0.810 = 0$, from which $z_1 = -2.0$ and $z_2 = -4.263$.

We next solve $P(x, 2) = x^2 + 2x + 1 = 0$, which gives the double root $x = -1$, and $P(x, -4.263) = x^2 + 4.263x + 1 = 0$, which gives the real roots $x = -0.2491$ and $x = -4.0139$. It then follows that there exist two functions which satisfy the conditions

$$\begin{aligned} b_1(x) &= (x+1)(x+0.2491) = x^2 + 1.2491x + 0.2491 \\ b_2(x) &= (x+1)(x+4.0139) = x^2 + 5.0139x + 4.0139 \end{aligned}$$

The function $b_1(x)$ gives rise to the regular moving average

$$\zeta_1(t) - \mu = \eta(t) + 1.2491\eta(t-1) + 0.2491\eta(t-2)$$

while the function $b_2(x)$ yields

$$\begin{aligned} \zeta_2(t) - \mu &= \frac{K_1}{K_2}(\eta(t) + 5.0139\eta(t-1) + 4.0139\eta(t-2)) \\ &= 0.2491\eta(t) + 1.2491\eta(t-1) + \eta(t-2) \end{aligned}$$

on using $K_1^2 = 1 + 1.2491^2 + 0.2491^2$ and $K_2^2 = 1 + 5.0139^2 + 4.0139^2$. Alternatively, because of the symmetry of the two roots, $\zeta_2(t)$ may be written down immediately once $\zeta_1(t)$ has been obtained.

Wold argued, after using a second example in which u_2 was further adjusted to make the roots z_1 and z_2 coincide, that even small changes in autocorrelations would lead to substantial alterations in the values taken by the moving average coefficients.

A further example was considered in detail by Wold. The values $u_1 = 0.60$, $u_2 = 0.09$, $u_3 = -0.15$ and $u_4 = -0.10$ closely approximate the first four serial coefficients, which we estimate as 0.595, 0.081, −0.161 and −0.126 respectively. These values yield

$$-10^3 v(z) = 10z^4 + 15z^3 - 49z^2 - 105z - 62$$

and on solving $v(z) = 0$ we obtain

$$z_1 = -2.1272 \quad z_2 = 2.5103 \quad z_3, z_4 = -0.9415 \pm 0.5240i$$

From Theorem 12 there will therefore exist a group of moving averages with the prescribed correlogram, and this group will consist of eight processes. Solving $P(x, z_i) = 0$, $i = 1, \ldots, 4$, gives the following solutions

$$\begin{array}{ll}
x_{11} = -0.7013 & x_{12} = -1.4259 \\
x_{21} = 0.4966 & x_{22} = 2.0137 \\
x_{31} = -0.3381 - 0.6679i & x_{32} = -0.6034 + 1.1919i \\
x_{41} = -0.3381 + 0.6679i & x_{42} = -0.6034 - 1.1919i
\end{array}$$

Writing

$$B(x) = (x + 0.3381 - 0.6679i)(x + 0.3381 + 0.6679i) = x^2 + 0.6762x + 0.5604$$

the regular moving average will be obtained from

$$\begin{aligned}
b(x) &= (x + 0.7013)(x - 0.4966) \cdot B(x) \\
&= x^4 + 0.8809x^3 + 0.3505x^2 - 0.1208x - 0.1952
\end{aligned}$$

i.e., as

$$\eta(t) + 0.8809\eta(t-1) + 0.3505\eta(t-2) - 0.1208\eta(t-3) - 0.1952\eta(t-4)$$

with $K^2 = 1.9515$. A second moving average with the same correlogram will be delivered by

$$\begin{aligned} b_1(x) &= (x + 1.4259)(x - 0.4966) \cdot B(x) \\ &= x^4 + 1.6055x^3 + 0.4807x^2 + 0.0420x - 0.3968 \end{aligned}$$

with $K_1^2 = 3.9679$. Multiplying $b_1(x)$ by $K/K_1 = 0.7013$ yields the second moving average

$$\eta(t) + 1.1259\eta(t-1) + 0.3371\eta(t-2) + 0.0294\eta(t-3) - 0.2783\eta(t-4)$$

Proceeding in analogous fashion, the third and fourth moving averages are obtained from

$$b_2(x) = (x + 0.7013)(x - 2.0137) \cdot B(x)$$

and

$$b_3(x) = (x + 1.4259)(x - 2.0137) \cdot B(x)$$

to yield

$$\eta(t) - 0.3159\eta(t-1) - 0.8637\eta(t-2) - 0.8395\eta(t-3) - 0.3930\eta(t-4)$$

and

$$\eta(t) + 0.0308\eta(t-1) - 0.9432\eta(t-2) - 0.7909\eta(t-3) - 0.5604\eta(t-4)$$

The four remaining moving averages correspond to the complex roots $x = -0.6034 \pm 1.1919i$ of $u(x) = 0$. Due to symmetry, these processes can be obtained directly from the four processes above by reversing the order of the coefficients. For example, the regular moving average gives

$$-0.1952\eta(t) - 0.1208\eta(t-1) + 0.3505\eta(t-2) + 0.8809\eta(t-3) + \eta(t-4)$$

7.26 If $\{\zeta(t)\}$ is a regular moving average then the primary process $\{\eta(t)\}$ will be given either by (7.19) or by its limiting counterpart of Example 3 of §**7.21** when the characteristic equation has a root with modulus unity. Thus, consider the regular moving average

$$\zeta(t) = \eta(t) + 0.8809\eta(t-1) + 0.3505\eta(t-2) - 0.1208\eta(t-3) - 0.1952\eta(t-4)$$

Using the system (7.13) obtains $a_1 = -0.8809$, $a_2 = 0.4255$, $a_3 = 0.0548$ and $a_4 = 0.1043$, after which the (a) coefficients are given by the difference relation

$$a_k = -0.8809a_{k-1} - 0.3505a_{k-2} + 0.1208a_{k-3} + 0.1952a_{k-4} \quad k > 4$$

The primary process associated with a non-regular moving average may be obtained in a similar fashion as outlined by Wold (*ibid.*, pages 160–2). The moving averages in a group will, by construction, present the same correlogram and same variance, so that the autocorrelation properties of the corresponding series $\overline{\eta}_t$ will provide no basis for deciding which of the moving averages should be preferred.

In terms of forecasting, Wold argued that the forecast for which the expected squared deviation from the future path of $\{\zeta(t)\}$ was minimized is given by

$$F_t[\zeta(t+k)] = b_h\eta_t^{(i)} + b_{k+1}\eta_{t-1}^{(i)} + \cdots + b_h\eta_{t-h+k}^{(i)}$$

where $\{\eta_t^{(i)}\}$ is the primary process constructed from the ith non-regular moving average in the group and the sequence (b) is that for the regular moving average. In other words, the different moving averages in a group will give rise to the same sequence of optimal forecasts and thus in general

$$F_t[\zeta(t+k)] = \overline{\mu} + b_h\overline{\eta}_t + b_{k+1}\overline{\eta}_{t-1} + \cdots + b_h\overline{\eta}_{t-h+k}$$

where $\overline{\mu}$ is the sample average of $\overline{\zeta}_t$. From this formula it is seen that forecasts beyond the next h observations reduce to the sample average of the data. The squared deviation of errors of these forecasts are given by

$$(1 + b_1^2 + b_2^2 + \cdots + b_{h-1}^2)\sigma^2(\eta)$$

so that the efficiency of the forecasts decreases gradually as the number of periods being forecasted is extended, leading Wold to the opinion that

> especially in view of economic time series, the type of forecast delivered by the scheme of moving averages seems *a priori* more realistic, seems to correspond better to what might be reasonably possible to find out from the past development. Further, considering the forecasts over a short period, the prognosis given by the scheme of moving averages is, as a rule, rather efficient. In my opinion, this is a circumstance of central importance, for often the main interest is concentrated upon the prognosis concerning the near future. (*ibid.*, page 168)

An application of linear autoregression

7.27 Wold's second major application was to consider the Swedish cost of living index between 1840 and 1913 after he had removed a trend in the data

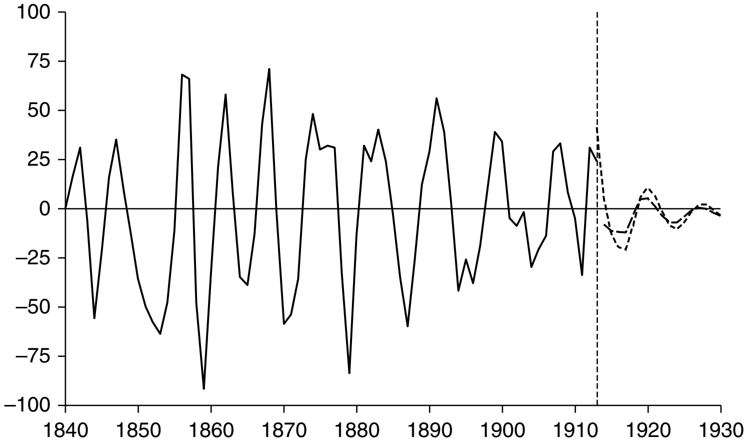

Figure 7.3 Swedish Cost of Living Index, 1840–1913, with forecasts out to 1930

to induce stationarity. This index is shown in Figure 7.3 and 'is seen to reflect clearly changes between economic expansion and contraction. A certain regularity seems to be present in the movement up and down, but the distance between two adjacent maxima is rather inconstant, varying between some 5 and 10 years' (*ibid.*, page 177). The correlogram is shown in Figure 7.4:

> The correlogram looks rather like a simple damped oscillation, say $C \cdot q^k \cdot \cos(\lambda k + \varphi)$. An inspection of the graph shows that in approximating the correlogram by such a function we would have to take the period $p = 2\pi/\lambda$ to be about 7 or 8 years, the phase φ to be approximately vanishing, and $q^7 \sim 1/2$, the latter relation corresponding to a damping of some 50% in the duration of one period. (*ibid.*, page 177)

This led Wold to consider a linear autoregression of order two, since 'this will present a correlogram forming a simple damped harmonic' (*ibid.*, page 177):

$$\zeta(t) + a_1\zeta(t-1) + a_2\zeta(t-2) = \eta(t) \qquad (7.51)$$

Wold preferred such a process to a scheme of hidden periodicities since the latter model, because each harmonic component will produce an undamped harmonic in the correlogram, would require at least two superposed harmonics to adequately represent its shape. Such a scheme would therefore involve at least six parameters rather than just the two required by (7.51).

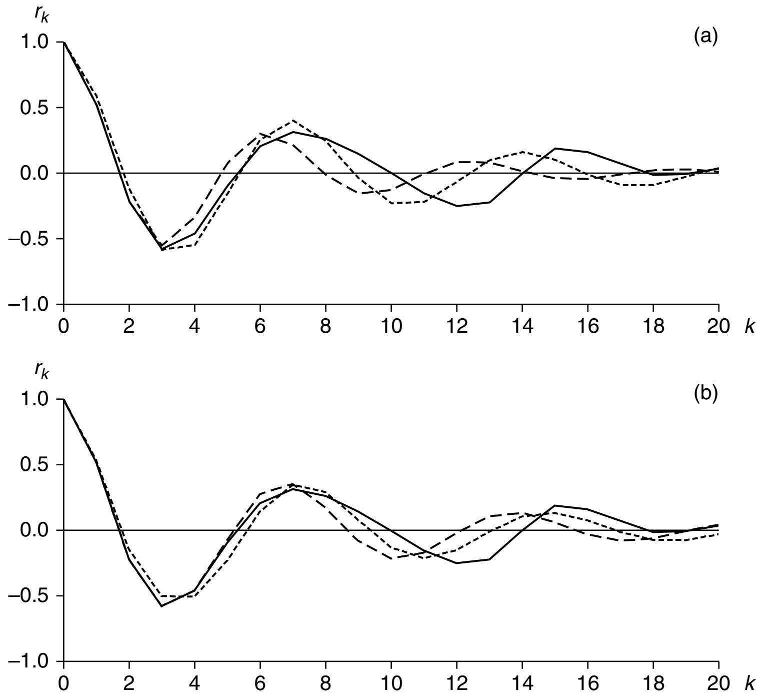

Figure 7.4 Correlogram of the cost of living index (unbroken line) with the hypothetical correlograms from equations (7.54) (dashed line) and (7.55) (dotted line) in panel (a), and from equations (7.56) (dashed line) and (7.57) (dotted line) in panel (b)

7.28 In §**7.18**, the system of equations (7.26) with coefficients $a_1, \ldots, a_h$ will deliver the autocorrelation coefficients $r_1, \ldots, r_{h-1}$ required for deriving the following coefficients $r_h, r_{h+1}, \ldots$ from the difference relations (7.25). In searching for an adequate autoregressive process of the form (7.51), the 'inverse problem' has to be confronted, i.e., that of finding a set of coefficients $a_1, \ldots, a_h$ given a set of serial coefficients. Wold thus suggested replacing the r_k with the corresponding serial coefficients $\bar{r}_k$ in the system (7.26) and the last relation in (7.25) and solving the following system of equations for a 'trial' set of (a) coefficients:

$$\begin{cases} \bar{r}_1 + a_1 + a_2\bar{r}_1 + a_3\bar{r}_2 + \cdots + a_h\bar{r}_{h-1} = 0 \\ \bar{r}_2 + a_1\bar{r}_1 + a_2 + a_3\bar{r}_1 + \cdots + a_h\bar{r}_{h-2} = 0 \\ \cdots \\ \bar{r}_h + a_1\bar{r}_{h-1} + a_2\bar{r}_{h-2} + \cdots + a_h = 0 \end{cases} \tag{7.52}$$

If the roots of the characteristic equation associated with (a) lie in the unit circle, these coefficients will define a linear autoregression of the form (7.24).

By construction, the first h autocorrelation coefficients of this process will coincide with the serial coefficients $\overline{r}_1, \ldots, \overline{r}_h$ and the subsequent coefficients can be obtained using the difference relations (7.25) and (a).

Having thus derived the correlogram of the hypothetical process defined by (a), this can then be compared with the empirical correlogram. If the fit appears satisfactory then the analysis can be carried further using (a), but if the deviations between the hypothetical and empirical correlograms are deemed to be too large, some adjustment of the (a) coefficients would then be required.

Given the h coefficients (a), the primary series $\overline{\eta}_t$ may be constructed from the observed values $\overline{\zeta}_t$ as

$$\overline{\eta}_t = \overline{\zeta}_t - \overline{\mu} + a_1(\overline{\zeta}_{t-1} - \overline{\mu}) + \cdots + a_h(\overline{\zeta}_{t-h} - \overline{\mu})$$

The method employed by Yule (1927) to obtain the (a) coefficients was least squares (see §**6.3**), which chooses (a) to minimize $\sum \overline{\eta}^2$. In fact, this approach closely approximates solving the system (7.52). In this case,

$$\sigma^2(\overline{\eta}) \approx (1 + a_1 r_1 + a_2 r_2 + \cdots + a_h r_h) \cdot \sigma^2(\overline{\zeta}) \qquad (7.53)$$

where the $\approx$ sign conveys the approximation produced by having to discard the first h terms of $\overline{\zeta}_t$ for which the corresponding values of $\overline{\eta}_t$ cannot be calculated. In other words, the variance of $\overline{\eta}_t$ will approximate the hypothetical variance $\sigma^2(\eta)$, although this will not be the case if the trial set (a) is determined otherwise than by (7.52).

Consequently, (7.52) gives a set of coefficients $a_1, \ldots, a_h$ that minimize the variance of the residuals $\overline{\eta}_t$. The first h autocorrelation coefficients will coincide with the corresponding serial coefficients but there is no guarantee that the complete hypothetical correlogram will provide a good fit to the empirical correlogram throughout its whole range. In practice, a compromise needs to be met 'between the two desiderata of obtaining small residuals $\overline{\eta}_t$ and small deviations between the correlograms, and besides try to satisfy the relation $\sigma^2(\eta) \sim \sigma^2(\overline{\eta})$.

7.29 Applying this approach to the cost of living index, Wold used the values $\overline{r}_1 = 0.5216$ and $\overline{r}_2 = -0.2240$ so that, with $h = 2$, the system (7.52) becomes[7]

$$\begin{aligned} 0.5216 + a_1 + 0.5216a_2 &= 0 \\ -0.2240 + 0.5216a_1 + a_2 &= 0 \end{aligned}$$

with the solution $a_1 = -0.8771$ and $a_2 = 0.6815$. The roots of the characteristic equation $z^2 + a_1 z + a_2 = 0$ are $0.4385 \pm 0.6994i$ and are thus less than unity in modulus, so that the relation

$$\zeta(t) - 0.8771\zeta(t-1) + 0.6815\zeta(t-2) = \eta(t) \qquad (7.54)$$

defines a process of linear autoregression. By construction, the first two autocorrelations of the process $\{\zeta(t)\}$ will be $r_1 = \bar{r}_1 = 0.5216$ and $r_2 = \bar{r}_2 = -0.2240$, with subsequent autocorrelations being obtained recursively from the difference relation $r_k - 0.8871 r_{k-1} + 0.6815 r_{k-2} = 0$, $k \geq 3$. The resulting correlogram is also shown in Figure 7.4(a), prompting Wold to argue that

> (c)omparing with the empirical correlogram, it is seen that the period in the hypothetical correlogram is too short, and that the damping is a little too heavy. ... (T)he damping factor equals $\sqrt{a_2}$, while the period is given by $p = 2\pi/\lambda$, where $\cos\lambda = -a_1/2\sqrt{a_2}$. Thus, an increase in a_2 will bring on a slighter damping. Further, reducing λ we obtain a longer period. However, ... we cannot conclude without further evidence that it will be possible to improve the fit – the coefficients a_1 and a_2 determine also the constant factor and the phase of the damped harmonic, and it might happen that an adjustment in a_1 and a_2 would cause such a change, e.g. in the phase, that the total result of the adjustment would be a poorer fit. (*ibid.*, page 180)

The hypothetical correlogram of (7.54) has a period of 6.22 years. To achieve a period close to seven years with reduced damping, Wold adjusted the coefficients to $a_1 = -1.10$ and $a_2 = 0.77$. The correlogram of the process defined by

$$\zeta(t) - 1.10\zeta(t-1) + 0.77\zeta(t-2) = \eta(t) \tag{7.55}$$

is also shown in Figure 7.4(a): 'up to r_8 and r_9, the hypothetical correlogram seems to fit rather well. Beyond this point, the fit is less satisfactory, partly because the graph of the serial coefficients presents a slow descent to the minimum in $k \sim 12.5$, and a rapid rise to the next maximum' (*ibid.*, page 181).

Substituting the appropriate values from the model (7.54) into (7.53) obtains $\sigma^2(\bar{\eta}) \sim \sigma^2(\eta) = 0.390\sigma^2(\zeta)$. Wold showed that (7.55) led to a larger residual variance, leading him to conclude that 'all in all, neither of the schemes seems adequate ... It seems as if we cannot find a satisfactory approach without taking into account more distant elements $\bar{\zeta}_{t-3}$, $\bar{\zeta}_{t-4}$, etc.' (*ibid.*, page 182). Wold thus extended the model by taking $h = 4$, arriving at the process

$$\zeta(t) - 0.8100\zeta(t-1) + 0.7452\zeta(t-2) - 0.0987\zeta(t-3) + 0.2101\zeta(t-4) = \eta(t) \tag{7.56}$$

The correlogram of this process is shown in Figure 7.4(b) and is seen to be almost identical to that from (7.55) although here, of course, $r_k = \bar{r}_k$ for $k \leq 4$. The roots of (7.56) are $0.5385 \pm 0.6814i$ and $-0.1335 \pm 0.5106i$, so that two of the roots are reasonably close to those of (7.55). Wold found that an improved fit to the empirical correlogram was obtained by adjusting the roots to $0.5888 \pm 0.6540i$

and $-0.20 \pm 0.58i$, leading to

$$\zeta(t) - 0.7776\zeta(t-1) + 0.6797\zeta(t-2) - 0.1342\zeta(t-3) + 0.2914\zeta(t-4) = \eta(t) \quad (7.57)$$

This is also shown in Figure 7.4(b) and Wold regarded the general shape as being 'rather satisfactory'. For (7.56), the relation (7.53) gives $\sigma^2(\eta) = 0.371\sigma^2(\zeta)$, which represents a slight increase in efficiency over (7.54) as compensation for introducing two further parameters.

7.30 Forecasts of the cost of living index may be calculated using equation (7.22). For a model of the type (7.57), these forecasts are built up as

$$\begin{aligned} F_t[\zeta(t+1)] &= (1 + a_1 + a_2 + a_3 + a_4)\overline{\mu} - a_1\overline{\zeta}_t - a_2\overline{\zeta}_{t-1} - a_3\overline{\zeta}_{t-2} - a_4\overline{\zeta}_{t-3} \\ F_t[\zeta(t+2)] &= (1 + a_1 + a_2 + a_3 + a_4)\overline{\mu} - a_1F_t[\zeta(t+1)] - a_2\overline{\zeta}_t - a_3\overline{\zeta}_{t-1} - a_4\overline{\zeta}_{t-2} \end{aligned}$$

and so on. These forecasts, calculated as $F_{1912}[\zeta(1913)] = 41.4$, $F_{1912}[\zeta(1914)] = 5.2$, etc., are shown in Figure 7.3 up to 1930, i.e., for k up to 18. Also shown are the set of forecasts $F_{1913}[\zeta(1914)]$, etc., up to 1930.

> The two forecasts curves in [Figure 7.3] yield a good illustration of the prognosis situation in an approach of linear autoregression. Firstly, while a forecast $F_t[\zeta(t+k)]$ is often rather efficient for small k-values, the efficiency vanishes asymptotically as k increases. Further, as soon as we are in a position to take a new observation $\overline{\zeta}_{t+1}$ into consideration when forming the prognosis, the new forecast curve is often substantially modified; how much, will depend on the residual $\overline{\eta}_{t+1} = \overline{\zeta}_{t+1} - F_t[(t+1)]$. – Summing up, it is the short forecasts that are efficient. In this respect, we meet the same situation as in the scheme of moving averages, and the same contrast to the scheme of hidden periodicities. On the other hand, under special circumstances the oscillations in a scheme of linear regression are nearly functional, viz. nearly strictly periodic – as remarked in discussing the sinusoidal limit theorem …, processes of hidden periodicities can be obtained as limit cases of the schemes of linear autoregression. (*ibid.*, page 187)

7.31 Wold finally considered how linear autoregressions were formed by the *complete systems* analysed in economics by Frisch (1933) and Tinbergen (1937). A simple example of a complete system is given by

$$\begin{aligned} \xi(t) &= c_1\zeta(t-1) + \eta'(t) \\ \zeta(t) &= d_0\xi(t) + d_1\xi(t-1) + \eta''(t) \end{aligned}$$

Such a system may be rewritten as

$$\zeta(t) = d_0c_1\zeta(t-1) + d_1c_1\zeta(t-2) + (d_0+d_1)\eta'(t) + \eta''(t)$$

i.e., as (7.51) with $a_i = -d_ic_1$, $i = 0, 1$, and $\eta_t = (d_0 + d_1)\eta'(t) + \eta''(t)$.

7.32 Wold's monograph was rightly hailed as a major contribution to the foundations of time series analysis, fusing together the intuitive autoregressive and moving average models of Yule, Slutzky and Walker, developed in response to observing physical and economic phenomena, with the advances in probability theory made by the Russian mathematicians Kolmogorov and Kinchine. But there is more to Wold's contribution, for he also introduced the first formal concepts in the theory of forecasting and made suggestions as to how these models may be arrived at by examining the contrast between empirical and hypothetical correlograms.

Wold, however, was acutely aware of the limitations of his framework, most notably in the absence of an inferential framework to bring to bear on the model selection process – a subject that, unsurprisingly, would quickly engage the attention of the new breed of mathematical statisticians encouraged by the work of Wold to research in the area of time series. Wold was also concerned with two other problems that had been avoided by focusing attention just on stationary time series – the necessity for detrending an 'evolutive' series before this modelling framework could be employed and the possibility that observed time series might be generated by a nonlinear process. Again, these were to become major research agendas in subsequent developments in time series modelling.