Handbook of Statistics

VOLUME 3

Time Series in the Frequency Domain

Edited by

D. R. Brillinger

P. R. Krishnaiah

Time Series in the Frequency Domain

Edited by

D. R. Brillinger

Department of Statistics
University of California, Berkeley, CA, U.S.A.

P. R. Krishnaiah

Center for Multivariate Analysis
University of Pittsburgh, Pittsburgh, PA, U.S.A.

1983

NORTH-HOLLAND
AMSTERDAM · NEW YORK · OXFORD

Preface

Time series analysis is one of the most flourishing of the fields of present day statistics. Exciting developments are taking place: in pure theory and in practice, with broad relevance and with narrow intent, for large samples and for small samples. The flourishing results in part, from the dramatic increase in the availability of computing power for both number crunching and for graphical display and in part from a compounding of knowledge as more and more researchers involve themselves with the problems of the field.

This volume of the *Handbook of Statistics* is concerned particularly with the frequency side, or spectrum, approach to time series analysis. This approach involves essential use of sinusoids and bands of (angular) frequency, with Fourier transforms playing an important role. A principal activity is thinking of systems, their inputs, outputs, and behavior in sinusoidal terms. In many cases, the frequency side approach turns out to be simpler in each of computational, mathematical, and statistical respects. In the frequency approach, an assumption of stationarity is commonly made. However, the essential roles played by the techniques of complex demodulation and seasonal adjustment show that stationarity is far from a necessary condition. So too are assumptions of Gaussianity and linearity commonly made. As various of the papers in this Volume show, nor are these necessary assumptions

The Volume is meant to represent the frequency approach to time series analysis as it is today. Readers working their way through the papers and references included will find themselves abreast of much of contemporary spectrum analysis.

We wish to express our deep appreciation to Professors E. J. Hannan and M. B. Priestley for serving as members of the editorial board. Thanks are due to Professors P. Guttorp, E. J. Hannan, T. Hasan, J. Lillestøl, and M. B. Priestley for refereeing various chapters in the volume. We are most grateful to the authors and North-Holland Publishing Company for their excellent cooperation in bringing out this volume.

D. R. Brillinger
P. R. Krishnaiah

Contributors

R. J. Bhansali, *Statistics Department, University of Liverpool, P.O. Box 147, Liverpool L69 3BX, U.K.* (Ch. 1)

David R. Brillinger, *Department of Statistics, University of California, Berkeley, California 94720* (Ch. 2)

T. C. Chang, *Department of Mathematical Sciences, University of Cincinnati, Cincinnati, Ohio 45221* (Ch. 20)

William S. Cleveland, *Bell Laboratories, Murray Hill, New Jersey 07974* (Ch. 3)

Robert B. Davies, *Department of Scientific and Industrial Research, Applied Mathematics Division, P.O. Box 1335, Wellington, New Zealand* (Ch. 4)

Robert Engle, *Department of Economics, University of California, San Diego, La Jolla, California 92093* (Ch. 5)

C. W. J. Granger, *Department of Economics, University of California, San Diego, La Jolla, California 92093* (Ch. 5)

E. J. Hannan, *Department of Statistics, The Australian National University, Box 4, G.P.O., Canberra, A.C.T., Australia 2600* (Ch. 6)

T. Hasan, *Bell Laboratories, Murray Hill, New Jersey 07974* (Ch. 7)

Melvin J. Hinich, *Department of Government, University of Texas at Austin, Austin, Texas 78712* (Ch. 8)

D. Karavellas, *Statistics Department, University of Liverpool, P.O. Box 147, Liverpool L69 3BX, U.K.* (Ch. 1)

L. H. Koopmans, *Department of Mathematics, University of New Mexico, Albuquerque, New Mexico 87131* (Ch. 9)

P. R. Krishnaiah, *Center for Multivariate Analysis, University of Pittsburgh, Pittsburgh, Pennsylvania 15260* (Ch. 20)

J. C. Lee, *Bell Laboratories, Murray Hill, New Jersey 07974* (Ch. 20)

R. Douglas Martin, *Department of Statistics, University of Washington, Seattle, Washington 98195* (Ch. 10)

Emanuel Parzen, *Institute of Statistics, Texas A & M University, College Station, Texas 77843* (Ch. 11)

J. Pemberton, *Department of Mathematics, University of Salford, Salford, M5 4WT, U.K.* (Ch. 12)

M. B. Priestley, *Department of Mathematics, University of Manchester Institute of Science & Technology, P.O. Box 88, Manchester, M60 1QD, U.K.* (Ch. 13)

T. Subba Rao, *Department of Mathematics, University of Manchester Institute of Science & Technology, P.O. Box 88, Manchester, M60 1QD, U.K.* (Ch. 14)

Enders A. Robinson, *Department of Geophysics, University of Tulsa, Tulsa, Oklahoma 74104* (Ch. 15)

P. M. Robinson, *Department of Mathematics, University of Surrey, Guildford, Surrey, CU2 5XH, U.K.* (Ch. 16)

M. Rosenblatt, *Department of Mathematics, University of California, San Diego, La Jolla, California 92093* (Ch. 17)

R. H. Shumway, *Division of Statistics, University of California, Davis, California 95616* (Ch. 18)

Tony Thrall, *Systems Applications, Inc., 101 Lucas Valley Rd., San Rafael, California 94903* (Ch. 19)

H. Tong, *Department of Statistics, Chinese University of Hong Kong, New Territories, Hong Kong* (Ch. 12)

此为试读，需要完整PDF请访问：www.ertongbook.com

Table of Contents

此为试读,需要完整PDF请访问: www.ertongbook.com

D. R. Brillinger and P. R. Krishnaiah, eds., *Handbook of Statistics*, Vol. 3
© Elsevier Science Publishers B.V. (1983) 1–19

Wiener Filtering (with emphasis on frequency-domain approaches)

R. J. Bhansali and D. Karavellas

1. Introduction

Let $\{y_t, x_t\}$ $(t = 0, \pm 1, \ldots)$ be a bivariate process. An important class of problems considered in time-series analysis may be formulated in terms of the problem: How can we best predict y_t from $\{x_s, s \leq t\}$? If $y_t = x_{t+\nu}$, $\nu > 0$, then the problem is that of predicting the 'future' of x_t on the basis of its past. If $x_t = \xi_t + \zeta_t$, where ζ_t is 'noise' and ξ_t the 'signal' and $y_t = \xi_{t+\nu}$ then for $\nu = 0$ the problem is that of 'signal extraction', for $\nu > 0$ that of predicting the signal and for $\nu < 0$ that of interpolating the signal, in the presence of noise. If y_t and x_t are arbitrary, then the problem is simply that of predicting one series from another. This last problem is itself of interest in a number of disciplines: for example, in Economics, interest is often centred on obtaining a distributed lag relationship between two economic variables (see, e.g., Dhrymes [11]) such as level of unemployment and the rate of inflation.

A complete solution to the problem of predicting y_t from the past, $\{x_s, s \leq t\}$, of x_t would consist of giving the conditional probability distribution of the random variable y_t when the observed values of the random variables $\{x_s, s \leq t\}$ are given. However, this is seldom practicable as finding such a conditional distribution is usually a formidable problem. A simplifying procedure of taking the mean value of this conditional distribution as the predictor of y_t is also rarely feasible because this mean value is in general a very complicated function of the past x's. Progress may, however, be made if $\{y_t, x_t\}$ is assumed to be jointly stationary and attention is restricted to the consideration of the linear least-squares predictor of y_t, i.e. the best predictor, $\hat{y}_t$, say, of y_t is chosen from the comparatively narrow class of linear functions of $\{x_s, s \leq t\}$,

$$\hat{y}_t = \sum_{j=0}^{\infty} h(j) x_{t-j}, \tag{1.1}$$

the coefficients $h(j)$ being chosen on the criterion that the mean square error of prediction

$$\eta^2 = E(\hat{y}_t - y_t)^2 \tag{1.2}$$

be a minimum.

Formation of $\hat{y}_t$ from the $\{x_s, s \leq t\}$ may be viewed as a filtering operation applied to the past of x_t, and, especially in engineering literature, $\hat{y}_t$ is known as the Wiener filter.

It should be noted that if $\{y_t, x_t\}$ is Gaussian, then the linear least-squares predictor, $\hat{y}_t$, of y_t is also the best possible predictor in the sense that it minimises the mean square error of prediction within the class of all possible predictors of y_t; hence for the Gaussian case the consideration of only linear predictors is not a restriction.

2. Derivation of the filter transfer function and the filter coefficients

Suppose that $\{y_t, x_t\}$ $(t = 0, \pm 1, \ldots)$ is real-valued jointly stationary with zero means, i.e. $Ex_t = Ey_t = 0$. If the means are nonzero, then these may be subtracted out. Let $R_{xx}(u) = E(x_{t+u}x_t)$ and $R_{yy}(u) = E(y_{t+u}y_t)$ denote the autocovariance functions of x_t and y_t, respectively, and let $R_{yx}(u) = Ey_{t+u}x_t$ denote their cross-covariance function. Assume that

$$\sum_{u=-\infty}^{\infty} |R_{xx}(u)| < \infty, \qquad \sum_{u=-\infty}^{\infty} |R_{yy}(u)| < \infty, \qquad \sum_{u=-\infty}^{\infty} |R_{yx}(u)| < \infty$$

and let

$$f_{xx}(\lambda) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} R_{xx}(u) \exp(-iu\lambda),$$

$$f_{yy}(\lambda) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} R_{yy}(u) \exp(-iu\lambda)$$

denote the power spectral density functions of x_t and y_t, respectively, and

$$f_{yx}(\lambda) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} R_{yx}(u) \exp(-iu\lambda)$$

their cross-spectral density function. Assume also that $f_{xx}(\lambda) \neq 0$ $(-\infty < \lambda < \infty)$.

Under these conditions x_t has the one-sided moving average representation (see Billinger [9, p. 78])

$$x_t = \sum_{j=0}^{\infty} b(j)\varepsilon_{t-j}, \quad b(0) = 1, \tag{2.1}$$

and the autoregressive representation

$$\sum_{j=0}^{\infty} a(j)x_{t-j} = \varepsilon_t, \quad a(0) = 1. \tag{2.2}$$

Here ε_t is a sequence of uncorrelated random variables with 0 mean and finite variance σ^2, say, and the $\{b(j)\}$ and $\{a(j)\}$ are absolutely summable coefficients, i.e. they satisfy

$$\sum_{j=0}^{\infty} |b(j)| < \infty, \qquad \sum_{j=0}^{\infty} |a(j)| < \infty.$$

Also, if

$$B(z) = \sum_{j=0}^{\infty} b(j)z^j, \qquad A(z) = \sum_{j=0}^{\infty} a(j)z^j, \tag{2.3}$$

respectively, denote the characteristic polynomials of the $b(j)$ and the $a(j)$, then $B(z) \neq 0$, $A(z) \neq 0$, $|z| \leq 1$ and $A(z) = \{B(z)\}^{-1}$. The transfer functions $B(e^{-i\lambda})$ and $A(e^{-i\lambda})$ of the $b(j)$ and $a(j)$ are denoted by $B(\lambda)$ and $A(\lambda)$ respectively. We have $A(\lambda) = \{B(\lambda)\}^{-1}$ and $f_{xx}(\lambda) = \sigma^2(2\pi)^{-1}|B(\lambda)|^2$.

If $f_{xx}(\lambda)$ is known exactly, then the $\{b(j)\}$ and $\{a(j)\}$ may be determined, by the Wiener–Hopf spectral factorization procedure (Wiener [25, p. 78]). The assumptions made previously on $R_{xx}(u)$ and $f_{xx}(\lambda)$ ensure that $\log f_{xx}(\lambda)$ is integrable and hence has the Fourier series expansion

$$\log f_{xx}(\lambda) = \sum_{v=-\infty}^{\infty} c(v) \exp(-iv\lambda), \tag{2.4}$$

with

$$c(v) = (2\pi)^{-1} \int_{-\pi}^{\pi} \log f_{xx}(\lambda) \exp(iv\lambda)\, d\lambda \tag{2.5}$$

and

$$\sum_{v=-\infty}^{\infty} |c(v)| < \infty.$$

Set

$$B(\lambda) = \exp\left\{\sum_{v=1}^{\infty} c(v) \exp(-iv\lambda)\right\}, \tag{2.6}$$

$$A(\lambda) = \{B(\lambda)\}^{-1} \tag{2.7}$$

and

$$\sigma^2 = 2\pi \exp\{c(0)\}. \tag{2.8}$$

Then

$$b(j) = (2\pi)^{-1} \int_{-\pi}^{\pi} B(\lambda) \exp(ij\lambda)\, d\lambda, \tag{2.9}$$

$$a(j) = (2\pi)^{-1} \int_{-\pi}^{\pi} A(\lambda) \exp(ij\lambda)\, d\lambda, \tag{2.10}$$

and the $\{b(j)\}$ and $\{a(j)\}$ thus obtained are absolutely summable (Brillinger [9,

p. 79]); see also Doob [12, pp. 160–164] and Grenander and Rosenblatt [16, pp. 67–81] for related work.

Next, consider prediction of y_t from the past, $\{x_s, s \leq t\}$, of x_t, and in particular the determination of the filter coefficients $h(j)$ of the linear least-squares predictor $\hat{y}_t$ of y_t. The mean square error of prediction η^2 is given by

$$\eta^2 = R_{yy}(0) - 2\sum_{j=0}^{\infty} h(j)R_{yx}(j) + \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} h(j)h(k)R_{xx}(k-j). \tag{2.11}$$

If the $h(j)$ minimise η^2, then we must have $\partial\eta^2/\partial h(j) = 0$ $(j = 0, 1, \ldots)$. This requirement leads to the equations

$$\sum_{k=0}^{\infty} h(k)R_{xx}(k-j) = R_{yx}(j) \quad (j = 0, 1, \ldots). \tag{2.12}$$

That the $h(k)$ satisfying (2.12) also minimise η^2 may be established by using an argument analogous to that given, for example, by Jenkins and Watts [18, pp. 204–205].

Equations (2.12) provide discrete analogues of the Wiener–Hopf integral equations (Wiener [25, p. 84]). As their left-hand side is of the form of a convolution, the use of Fourier series techniques is a natural approach to adopt for solving them. However, as discussed by N. Levinson (see [25, p. 153]) a direct use of the Fourier series techniques for obtaining the $h(j)$ is not feasible as well, because (2.12) is valid only for $j \geq 0$. Therefore, a somewhat indirect approach is adopted for expressing $h(j)$ in terms of $f_{yx}(\lambda)$ and $f_{xx}(\lambda)$.

The representation (2.1) implies that

$$R_{xx}(u) = \sigma^2 \sum_{s=0}^{\infty} b(s)b(s+u) \quad (u = 0, 1, \ldots). \tag{2.13}$$

Put

$$D(\lambda) = f_{yx}(\lambda)\overline{A(\lambda)} = \sum_{u=-\infty}^{\infty} d(u)\, e^{-iu\lambda}, \tag{2.14}$$

and

$$[D(\lambda)]_+ = \sum_{u=0}^{\infty} d(u)\exp(-iu\lambda), \tag{2.15}$$

where

$$d(u) = (2\pi)^{-1}\int_{-\pi}^{\pi} f_{yx}(\lambda)\overline{A(\lambda)}\exp(iu\lambda)\, d\lambda \tag{2.16}$$

and

$$\sum_{u=-\infty}^{\infty} |d(u)| < \infty.$$

Note that $2\pi d(u) = E(y_t\varepsilon_{t-u})$ and $D(\lambda)$ gives the cross-spectral density function of y_t and ε_t.

From (2.14), we get

$$R_{yx}(j) = 2\pi \sum_{s=0}^{\infty} b(s)d(j+s)\,. \tag{2.17}$$

Hence, (2.12) may be rewritten as

$$\sum_{s=0}^{\infty} b(s)d(j+s) = \frac{\sigma^2}{2\pi} \sum_{s=0}^{\infty} \sum_{k=0}^{\infty} h(k)b(s)b(s+j-k) \quad (j = 0, 1, \ldots),$$

or, as

$$d(v) = \frac{\sigma^2}{2\pi} \sum_{k=0}^{\infty} h(k)b(v-k) \quad (v = 0, 1, \ldots)\,. \tag{2.18}$$

Since, $b(v) = 0$, $v < 0$, (2.18) may be solved by the Fourier series techniques. On multiplying both the sides of (2.18) by $e^{-iv\lambda}$ and summing for all $v \geqslant 0$, we get

$$\begin{aligned} H(\lambda) &= \sum_{k=0}^{\infty} h(k) \exp(-ik\lambda) \\ &= \frac{2\pi}{\sigma^2} B(\lambda)^{-1}[D(\lambda)]_+ = \frac{2\pi}{\sigma^2} A(\lambda)[f_{yx}(\lambda)\overline{A(\lambda)}]_+ \end{aligned} \tag{2.19}$$

and

$$h(j) = (2\pi)^{-1} \int_{-\pi}^{\pi} H(\lambda) \exp(ij\lambda)\, d\lambda\,. \tag{2.20}$$

Since the $d(u)$ given by (2.16) and the $a(j)$ given by (2.9) are absolutely summable, so are the $h(j)$ (see, e.g., Fuller [14, p. 120]). Thus, the $h(j)$'s satisfy

$$\sum_{j=0}^{\infty} |h(j)| < \infty\,.$$

The mean square error of prediction η^2 is

$$\begin{aligned} \eta^2 &= E\{(y_t - \hat{y}_t)^2\} = R_{yy}(0) - \sum_{j=0}^{\infty} h(j)R_{yx}(j) \\ &= \int_{-\pi}^{\pi} \left[f_{yy}(\lambda) - \left\{ \frac{2\pi}{\sigma^2} |[D(\lambda)]_+|^2 \right\} \right] d\lambda = R_{yy}(0) - \frac{4\pi^2}{\sigma^2} \sum_{j=0}^{\infty} d^2(j)\,. \end{aligned} \tag{2.21}$$

Equations (2.19) and (2.21) are consistent with the results of Whittle [24, pp. 66–68], but note that a dividing factor of σ^2 is missing in equation (3.7.2) of Whittle [24, p. 42]; see also Bhansali [3].

It is instructive to compare the 'one-sided' predictor (2.10) with the corresponding 'two-sided' predictor of y_t obtained by assuming that the complete past, present and the complete future of x_t is known. Let

$$\tilde{y}_t = \sum_{j=-\infty}^{\infty} g(j)x_{t-j} \tag{2.22}$$

be the 'two-sided' linear least-squares predictor of y_t. Then, as in (2.12), the $g(j)$ are the solutions of the equations

$$R_{yx}(u) = \sum_{j=-\infty}^{\infty} g(j)R_{xx}(u-j) \quad (u = 0, \pm 1, \ldots). \tag{2.23}$$

Since these equations are two-sided and are valid for all integral values of u, they may be solved by the Fourier series techniques. On multiplying both the sides of (2.23) by $(2\pi)^{-1}\,\mathrm{e}^{-iu\lambda}$ and summing over u, we have

$$\Gamma(\lambda) = \sum_{j=-\infty}^{\infty} g(j)\exp(-\mathrm{i}j\lambda) = f_{yx}(\lambda)/f_{xx}(\lambda) \tag{2.24}$$

and

$$g(j) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Gamma(\lambda)\exp(\mathrm{i}j\lambda). \tag{2.25}$$

Let $\tau^2 = E\{(y_t - \tilde{y}_t)^2\}$ be the corresponding mean square error of prediction. We have

$$\tau^2 = \int_{-\pi}^{\pi}\left\{f_{yy}(\lambda) - \frac{|f_{yx}(\lambda)|^2}{f_{xx}(\lambda)}\right\}\mathrm{d}\lambda = \int_{-\pi}^{\pi}\{1 - C_{yx}^2(\lambda)\}f_{yy}(\lambda)\,\mathrm{d}\lambda, \tag{2.26}$$

where $C_{yx}(\lambda) = |f_{yx}(\lambda)|/\{f_{yy}(\lambda)f_{xx}(\lambda)\}^{1/2}$ is called the coherence between y_t and x_t. Note that $0 \leq C_{yx}(\lambda) \leq 1$, all λ. Expression (2.26) therefore shows that if $C_{yx}(\lambda)$ is close to 1 at all frequencies, then τ^2 is close to 0, and one would expect to obtain a close linear fit between y_t and x_t. In this sense, $C_{yx}(\lambda)$ may be interpreted as a correlation coefficient 'in the frequency domain' (see, e.g., Priestley [22] and Granger and Hatanaka (15]).

On using (2.14)–(2.16), (2.21) may be rewritten as (see Whittle [24, p. 69])

$$\eta^2 = \tau^2 + \frac{4\pi^2}{\sigma^2}\sum_{j=-\infty}^{-1} d^2(j). \tag{2.27}$$

The second term to the right of this expression gives the increase in mean square error due to the restriction that only the 'past' of x_t may be used for predicting y_t. In general, therefore, $\eta^2 \geq \tau^2$.

There is, however, one important situation in which $\eta^2 = \tau^2$. This occurs when x_t is the input to, and y_t the output of, a physically realizable linear time-invariant filter with uncorrelated noise, i.e. when,

$$y_t = \sum_{j=0}^{\infty} l(j)x_{t-j} + z_t, \tag{2.28}$$

$\{z_t\}$ is a stationary process uncorrelated with x_t and $\Sigma\,|l(j)| < \infty$.

此为试读，需要完整PDF请访问： www.ertongbook.com