Automatica, Vol. 21, No. 6, pp. 677–696, 1985.
Printed in Great Britain.

0005-1098/85 $3.00 + 0.00
Pergamon Press Ltd.

On the Estimation of Transfer Functions*

LENNART LJUNG†

Connections between frequency and Fourier analysis, spectral analysis and parametric time-domain identification methods are displayed and utilized to provide insight into each method, particularly the role of the noise model in obtaining a fit between the model and the true system.

Key Words—Identification; transfer functions; parameter estimation; Fourier analysis; frequency response; spectral analysis.

Abstract—This paper treats the close conceptual relationships between basic approaches to the estimation of transfer functions of linear systems. The classical methods of frequency and spectral analysis are shown to be related to the well-known time domain methods of prediction error type via a common "empirical transfer function estimate". Asymptotic properties of the estimates obtained by the respective methods are also described and discussed. An important feature that is displayed by this treatment is a frequency domain weighting function that determines the distribution of bias in case the true system cannot be exactly described within the chosen model set. The choice of this weighting function is made in terms of noise models for time-domain methods. The noise model thus has a dual character from the system approximation point of view.

1. INTRODUCTION

THE TRANSFER function is a fundamental concept in linear system theory. Different representations of the transfer function play important roles for description and analysis of the system's properties as well as for the design of controllers. It is quite natural that the problem of determining and estimating the transfer function of a given system has become an important area of research. Many methods have been developed, ranging from straightforward frequency response analysis to spectral analysis and to various sophisticated time domain, parametric identification methods. See, e.g., Jenkins and Watts (1969), Åström and Eykhoff (1971), Eykhoff (1974, 1981), Goodwin and Payne (1977), Rake (1980), Godfrey (1980), Brillinger (1981) and Ljung and Glover (1981).

In this paper we shall consider the different possibilities that are at hand for estimating transfer functions. One might ask what can be added to a subject, already so thoroughly treated in the literature. The current paper will, however, concentrate on three aspects, that in my opinion, have not been sufficiently stressed earlier:

—We shall expose the close kinship between the three basic approaches: frequency analysis, spectral analysis and time domain prediction error methods. Their common denominator is what in this paper is called "the empirical transfer function estimate (ETFE)".

—We shall view the system truly as a black box, meaning that our only ambition is to obtain a good description of its transfer function. Intrinsic variables such as model structures and model orders will be viewed only as vehicles to obtain such a description. This means in particular that we may pose no *a priori* bounds or model orders, but must be prepared to allow them to tend to infinity, if this is found to be beneficial.

—It is a trivial observation that the system typically is considerably more complex than it is suitable to allow the model to be. Here we try to pursue this observation in a consistent manner and expose its consequences. The implications for the choice of several design variables (such as input spectra, prefilters, etc.) are non-trivial.

The above three items can be seen as a program for a discussion of system identification techniques. It is not possible to give a complete account of the program within the format of this article. Complementing treatments are given in Ljung and Yuan (1984, 1985), Yuan and Ljung (1984, 1985), Ljung (1984, 1985), Wahlberg and Ljung (1984), Gevers and Ljung (1985). A comprehensive account is given in Ljung (in press).

Let us state the problem formally. We shall confine ourselves to discrete time single-input–single-output systems. This is for notational convenience and is not essential for the development.

* Received 23 February 1984; revised 19 March 1985. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor G. C. Goodwin under the direction of Editor P. C. Parks.

† Division of Automatic Control, Department of Electrical Engineering, Linköping University, S-58183 Linköping, Sweden.

Consider a linear, time-invariant system with input signal $\{u(t)\}$ and output $\{y(t)\}$, subject to an additive disturbance $\{v(t)\}$:

$$y(t) = \sum_{k=1}^{\infty} g_0(k)u(t-k) + v(t);\ t = 1,2,\ldots. \tag{1.1}$$

This system has the transfer function

$$G_0(e^{i\omega}) = \sum_{k=1}^{\infty} g_0(k)e^{-ik\omega}. \tag{1.2}$$

The additive noise $v(t)$ is assumed to be a stationary zero mean stochastic process with covariance function

$$r_v(\tau) = Ev(t+\tau)v(t) \tag{1.3}$$

and spectral density

$$\Phi_v(\omega) = \sum_{\tau=\infty}^{\infty} r_v(\tau)e^{-i\tau\omega}. \tag{1.4}$$

We shall sometimes represent $v(t)$ also as

$$v(t) = \sum_{k=0}^{\infty} h_0(k)e(t-k);\quad h_0(0) = 1 \tag{1.5}$$

where $\{e(t)\}$ is white noise with variance λ. With

$$H_0(e^{i\omega}) = \sum_{k=0}^{\infty} h_0(k)e^{-ik\omega} \tag{1.6}$$

we can then write v:s spectral density as

$$\Phi_v(\omega) = \lambda|H_0(e^{i\omega})|^2. \tag{1.7}$$

The estimation problem now is the following. Generate or observe an input signal

$$u^N: u(k)\quad k = 1,\ldots N$$

and observe the corresponding output signal

$$y^N: y(k)\quad k = 1,\ldots N.$$

Based on these observations, form an estimate of the frequency transfer function

$$\hat{G}_N(e^{i\omega}) = \hat{G}(e^{i\omega};\ N, y^N, u^N). \tag{1.8}$$

Our aim is that the estimate $\hat{G}_N(e^{i\omega})$ is as close as possible to the true transfer function $G_0(e^{i\omega})$. Since the estimated object is a function over $-\pi \le \omega \le \pi$, the estimation problem is an infinite dimensional one, or "non-parametric".

The topic of the paper is to discuss this estimation problem. The presentation is organized as follows. Section 2 deals with some preliminaries for the input sequence. In Section 3 we briefly review the main approaches to the stated problem. In Section 4 we define the empirical transfer function estimate (ETFE) as a natural development of frequency response analysis. In Section 5 we discuss how to improve this estimate. Approaches leading to spectral analysis are treated in Section 6, while approaches corresponding to time-domain prediction error methods are discussed in Section 7. Insights from the analysis and its implications for time-domain methods are discussed in Section 8.

2. THE INPUT SPECTRUM

Properties of the input signal will play an important role for the discussion here. We shall therefore first give a brief account of some of its assumed properties.

We shall throughout assume that the input sequence is such that the following limit exists:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} u(t-\tau)u(t) = r_u(\tau) \quad \forall\tau. \tag{2.1}$$

This will be the case, e.g. if the input is periodic or a realization of a stationary stochastic process.

We shall also assume that the Fourier transform of $\{r_u(\tau)\}$ exists:

$$\Phi_u(\omega) = \sum_{\tau=-\infty}^{\infty} r_u(\tau)e^{-i\omega\tau},\ -\pi \le \omega < \pi \tag{2.2}$$

where we allow $\Phi_u(\omega)$ to contain delta functions. The function $\Phi_u(\omega)$ is known as the *spectral density* of $\{u(t)\}$.

We shall often work with the following Fourier transform of the input sequence

$$U_N(\omega) = \frac{1}{\sqrt{N}}\sum_{t=1}^{N} u(t)e^{-it\omega} \quad -\pi \le \omega < \pi. \tag{2.3}$$

The sequence

$$U_N(2\pi k/N)\quad k = 0, 1,\ldots N-1$$

$$\left(\text{or } k = \frac{N}{2}+1,\ldots,0,1,\ldots,\frac{N}{2}\right) \tag{2.4}$$

is known as the *discrete Fourier transform* (DFT) of $\{u(t)\ t = 1,\ldots,N)\}$. We note that the values for intermediate frequencies of $U_N(\omega)$ are obtained by trigonometric interpolation of (2.4).

The absolute square of (2.3) is the *periodogram* of

the input:

$$P_N(\omega) = |U_N(\omega)|^2 \quad -\pi \leq \omega \leq \pi \tag{2.5}$$

(see, e.g. Jenkins and Watts, 1969, or Brillinger, 1981) for an account of these concepts).

The periodogram is closely related to the spectral density $\Phi_u(\omega)$. Let us comment on the relationships.

2.1. *The input is a sinusoid*

Suppose that

$$u(t) = A \sin \omega_0 t; \ \omega_0 = 2\pi/N_0. \tag{2.6}$$

We then have

$$r(\tau) = \frac{A^2}{2} \cos \omega_0 \tau \tag{2.7}$$

and

$$\Phi_u(\omega) = \frac{1}{4} A^2(\delta(\omega - \omega_0) + \delta(\omega + \omega_0)) \tag{2.8}$$

where δ is Dirac's delta function. Similarly, straightforward calculations show that if $N = s N_0$ then

$$|U_N(\omega)|^2 = \begin{cases} A^2 N/4 & \text{if } \omega = \pm\omega_0 = 2\pi/N_0 = 2\pi s/N \\ 0 & \text{if } \omega = 2\pi k/N \quad k \neq s. \end{cases} \tag{2.9}$$

We may thus say that $P_N(\omega)$ tends to $\Phi_u(\omega)$ as N tends to infinity in this case.

2.2. *The input signal is periodic with period* N_0

If $u(t) = u(t + N_0) \forall t$, we can, according to Fourier theory, write $u(t)$ as a finite Fourier series over the frequencies $2\pi r/N_0$ $r = -N_0/2 + 1, \ldots, 0, 1, \ldots, N_0/2$ (assuming N_0 to be even)

$$u(t) = \sum_{r=-N_0/2+1}^{N_0/2} A_r e^{2\pi i t r/N_0} \tag{2.10}$$

which is a simple extension of case (2.1) to finitely many sinusoids. W3 then have

$$\Phi_u(\omega) = \sum_{r=-N_0/2+1}^{N_0/2} |A_r|^2 \delta(\omega - 2\pi r/N_0) \tag{2.11}$$

and, for $N = s N_0$

$$|U_N(\omega)|^2 = \begin{cases} s|A_r|^2 & \text{if } \omega = 2\pi r/N_0 \\ & \quad r = -N_0/2 + 1, \ldots, 0, 1, \ldots N_0/2 \\ 0 & \text{if } \omega = 2\pi k/N; \ k \neq rs. \end{cases} \tag{2.12}$$

Here, again, we may say that $P_N(\omega)$ tends to $\Phi_u(\omega)$ as $N \to \infty$.

2.3. *The input is a realization of a stationary stochastic process*

Consider a stationary stochastic process $\{\bar{u}(t)\}$ with zero mean and covariance function

$$r(\tau) = E\bar{u}(t)\bar{u}(t - \tau)$$

and spectral density

$$\Phi_{\bar{u}}(\omega) = \sum_{\tau=-\infty}^{\infty} r(\tau) e^{-i\tau\omega}.$$

Let $u(t)$ be a realization of this stochastic process. Then, under weak conditions, the relations (2.1)–(2.2) will hold with probability one (w.p.1). For the periodogram (2.3), (2.5) the following will hold (see, e.g. Brillinger, 1981):

$$EP_N(\omega) \to \Phi_{\bar{u}}(\omega) \quad \text{as } N \to \infty \tag{2.13}$$

$$E(P_N(\omega_1) - \Phi_{\bar{u}}(\omega_1))(P_N(\omega_2) - \Phi_{\bar{u}}(\omega_2)) =$$

$$= \begin{cases} 0 & \text{if } |\omega_1 - \omega_2| = \dfrac{2\pi k}{N} k = 1, 2, \ldots N - 1 \\ \Phi_{\bar{u}}^2(\omega_1) & \text{if } \omega_1 = \omega_2 \end{cases} \tag{2.14}$$

for $0 \leq \omega_{1,2} \leq \pi$ asymptotically in N.

The periodogram is thus an unbiased estimate of the spectrum, but it does not converge as a function to the spectrum as N tends to infinity. However, as evident from (2.14), the estimates at neighbouring frequencies become uncorrelated. Therefore "smoothed versions" of $P_N(\omega)$ will behave like $\Phi_{\bar{u}}(\omega)$. In particular we have

$$\int_{-\pi}^{\pi} (P_N(\omega) - \Phi_{\bar{u}}(\omega)) f(\omega) d\omega \to 0 \text{ w.p.1} \quad \text{as } N \to \infty \tag{2.15}$$

for any differentiable function $f(\omega)$. Moreover, if $g_d(\omega)$ approaches the δ-function as d increases, i.e. if

$$\int g_d(\omega) f(\omega) d\omega \to f(0) \quad \text{as } d \to \infty \quad \forall \text{ smooth } f$$

then

$$\int_{-\pi}^{\pi} g_{d(N)}(\omega - \omega_0) P_N(\omega) d\omega \to \Phi_u(\omega_0) \quad \text{as } N \to \infty \tag{2.16}$$

provided $d(N)$ increases to infinity sufficiently slowly with N. (See Jenkins and Watts, 1969, or Brillinger, 1981.)

3. APPROACHES TO TRANSFER FUNCTION ESTIMATION

We shall in this section give a brief account of the major approaches that have been taken to the determination or estimation of transfer function

3.1. *Frequency response analysis*

An immediate and well-known consequence of the linear structure (1.1) is that an input sinusoid of a given frequency gives, after a vanishing transient (provided the system is stable), an output sinusoid of the same frequency, but with an amplitude change and a phase shift that is described to be the transfer function/frequency function:

$$\begin{aligned} u(t) &= A\cos\omega_0 t \\ y(t) &= B\cos(\omega_0 t + \varphi) + \text{transient} \end{aligned} \tag{3.1}$$

$$\begin{aligned} \varphi &= \arg G_0(e^{i\omega_0}) \\ B &= |G_0(e^{i\omega_0})|A. \end{aligned} \tag{3.2}$$

Consequently, the transfer function at frequency ω_0 can be determined by inspection of the output from an input sinusoid. This is known as frequency analysis. See, e.g. Rake (1980).

3.2. *Frequency analysis by the correlation method*

It may sometimes be difficult to determine φ and B from inspection of $y(t)$, in particular if the system is subject to disturbances. A commonly applied method is then to correlate the output with $\sin\omega_0 t$ and $\cos\omega_0 t$, respectively:

$$I_c(N) = \sum_{t=1}^{N} y(t)\cos\omega_0 t; \qquad I_s(N) = \sum_{t=1}^{N} y(t)\sin\omega_0 t. \tag{3.3}$$

Then we take

$$\arg\hat{G}_N(\omega_0) = \arctan(I_s(N)/I_c(N)) \tag{3.4a}$$

$$|\hat{G}_N(\omega_0)| = \frac{\sqrt{I_c^2(N) + I_s^2(N)}}{NA/2}. \tag{3.4b}$$

See, e.g. Rake (1980) for a more detailed account. By repeating the procedure (3.1), (3.3)–(3.4) for a number of frequencies, a good picture of the transfer function can be obtained. Equipment that performs frequency analysis by the correlation method is commercially available.

3.3. *Spectral analysis*

Suppose the system is given by (1.1) where the input is a stationary stochastic process, independent of $\{v(t)\}$, with spectrum $\Phi_u(\omega)$. The cross-spectrum between the output $y(t)$ and the input $u(t)$ is defined as

$$\Phi_{yu}(\omega) = \sum_{\tau=-\infty}^{\infty} r_{yu}(\tau)e^{-i\tau\omega} \tag{3.5a}$$

$$r_{yu}(\tau) = Ey(t)u(t-\tau). \tag{3.5b}$$

We then have

$$\Phi_{yu}(\omega) = G_0(e^{i\omega})\Phi_u(\omega). \tag{3.6}$$

(See Jenkins and Watts, 1968, or Åström, 1970). From (3.5) we see that once we obtain estimates of the input spectrum $\Phi_u(\omega)$ and the cross-spectrum $\Phi_{yu}(\omega)$, say $\Phi_u^N(\omega)$ and $\Phi_{yu}^N(\omega)$, we automatically have an estimate of the transfer function:

$$\hat{G}_N(e^{i\omega}) = \hat{\Phi}_{yu}^N(\omega)/\hat{\Phi}_u^N(\omega). \tag{3.7}$$

The question is how to estimate the spectra.

A common method is to smooth the periodogram:

$$\begin{aligned} \Phi_u^N(\omega) &= \int_{-\pi}^{\pi} W_\gamma(\xi - \omega)P_N(\xi)\mathrm{d}\xi \\ &= \int_{-\pi}^{\pi} W_\gamma(\xi - \omega)|U_N(\xi)|^2\mathrm{d}\xi. \end{aligned} \tag{3.8}$$

Here $W_\gamma(\cdot)$ is the weighting function, typically concentrated around the origin, with an adjustable "width", γ, suitably chosen. (See Fig. 1 below.) Similarly, a cross-spectrum estimate is obtained as

$$\Phi_{yu}^N(\omega) = \int_{-\pi}^{\pi} W_\gamma(\xi - \omega)Y_N(\xi)\bar{U}_N(\xi)\mathrm{d}\xi. \tag{3.9}$$

Here U_N is, as before, given by (2.4) while

$$Y_N(\omega) = \frac{1}{\sqrt{N}}\sum_{t=1}^{N} y(t)e^{-it\omega}. \tag{3.10}$$

Overbar here denotes complex conjugate. The product $Y_N(\omega)\bar{U}_N(\omega)$ can be seen as the periodogram estimate of the cross-spectrum. Often the

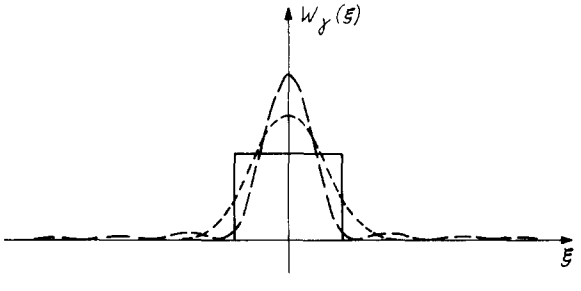

FIG. 1. Typical weighting functions.

convolutions in (3.8), (3.9) are instead performed as multiplications in the time domain, which means that Φ is obtained as the Fourier transform of suitable weighted covariance function estimates. This approach is known as the Blackman–Tukey spectral estimation method, Blackman and Tukey (1958). See, e.g. Jenkins and Watts, (1969, in particular Chapter 10), Godfrey (1980) or Brillinger (1981) for a detailed account of the whole process of spectral analysis.

3.4. *Time domain prediction error methods*

The perhaps most common approach in modern identification is to postulate that the transfer function is to be sought within a certain set:

$$\mathcal{G} = \{G(e^{i\omega}, \theta) \mid \theta \in D_{\mathcal{M}}\}. \tag{3.11}$$

Here $D_{\mathcal{M}}$ typically is a subset of R^d. In order to improve the result, it is customary to also include assumptions about the disturbance spectrum $\Phi_v(\omega)$, see (1.1)–(1.7). It is assumed to belong to a set

$$\begin{aligned} \Phi_v(\omega) &= \lambda |H(e^{i\omega}, \theta)|^2;\ H(e^{i\omega}, \theta) \in \mathcal{H} \\ \mathcal{H} &= \{H(e^{i\omega}, \theta) \mid \theta \in D_{\mathcal{M}}\}. \end{aligned} \tag{3.12}$$

This means that the system is assumed to be described as

$$y(t) = G(q, \theta)u(t) + H(q, \theta)e(t) \tag{3.13}$$

for some $\theta \in D_{\mathcal{M}}$. Here $\{e(t)\}$ is a sequence of independent random variables with zero mean values and variances λ, and G and H are functions of the shift operator q; $q^{-1}u(t) = u(t-1)$, so that

$$G(q, \theta) = \sum_{k=1}^{\infty} g_k(\theta) q^{-k} \tag{3.14a}$$

$$H(q, \theta) = 1 + \sum_{k=1}^{\infty} h_k(\theta) q^{-k}. \tag{3.14b}$$

There are several ways by which the transfer functions in (3.13) can be parametrized. Two common ones are illustrated in the following examples.

Example 3.1.—*State space models.* Suppose that a model of the system is posed in state space, innovations form:

$$\begin{aligned} x(t+1) &= A(\theta)x(t) + B(\theta)u(t) + K(\theta)e(t) \\ y(t) &= C(\theta)x(t) + e(t). \end{aligned} \tag{3.15}$$

Here the matrices A, B, C and K may be parametrized by θ in an arbitrary way. This model corresponds to (3.13) with

$$\begin{aligned} G(q, \theta) &= C(\theta)[qI - A(\theta)]^{-1}B(\theta) \\ H(q, \theta) &= 1 + C(\theta)[qI - A(\theta)]^{-1}K(\theta). \end{aligned} \tag{3.16}$$

Example 3.2—*ARMAX models.* Suppose that the model is chosen as

$$A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})e(t) \tag{3.17}$$

where A, B and C are polynomials in the delay operator. Such a model is known as an ARMAX model. Clearly (3.17) corresponds to (3.13) with

$$G(q, \theta) = B(q^{-1})/A(q^{-1}) \tag{3.18a}$$

$$H(q, \theta) = C(q^{-1})/A(q^{-1}). \tag{3.18b}$$

Here the parameter vector θ consists of the coefficients of the polynomials A, B and C.

Given the model (3.13) and input–output data up to time $t-1$, we can determine the predicted output at time t as follows. Rewrite (3.13) as

$$H^{-1}(q, \theta)y(t) = H^{-1}(q, \theta)G(q, \theta)u(t) + e(t)$$

or

$$\begin{aligned} y(t) = {} & [1 - H^{-1}(q, \theta)]y(t) + H^{-1}(q, \theta)G(q, \theta)u(t) \\ & + e(t). \end{aligned}$$

Since $H^{-1}(q, \theta)$ has an expansion in powers of q^{-1} that starts with a "1" [see (1.5)] and $G(q, \theta)$ contains a delay, the right-hand side of the above expression contains $y(k)$ and $u(k)$ only for $k \leq t-1$. The term $e(t)$ is independent of everything that happened up to time $t-1$. Hence the natural one-step ahead prediction, equal to the conditional expectation of $y(t)$ given previous data is

$$\hat{y}(t|\theta) = (1 - H^{-1}(q, \theta))y(t) + H^{-1}(q, \theta)G(q, \theta)u(t). \tag{3.19}$$

At time t, when $y(t)$ has been recorded we can compute the prediction error that the model (3.13) led to:

$$\varepsilon(t, \theta) = y(t) - \hat{y}(t \mid \theta) = H^{-1}(q, \theta)(y(t) - G(q, \theta)u(t)). \tag{3.20}$$

We may say that the model (3.13) is "good" if the sequence $\varepsilon(t, \theta)$, $t = 1, 2, \ldots, N$ is "small". In a very common class of identification methods, the squared sum of prediction errors is minimized to find the "best" model:

$$\hat{\theta}_N = \arg\min_{\theta \in D_{\mathcal{M}}} \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t, \theta). \tag{3.21}$$

With $\hat{\theta}_N$ determined in this way, the transfer function estimate becomes

$$\hat{G}_N(e^{i\omega}) = G(e^{i\omega}, \hat{\theta}_N). \tag{3.22}$$

Among methods that can be expressed as (3.21) we find the "maximum likelihood method" (assuming Gaussian disturbances), the "least squares method" and others. See Ljung (1978) and Åström (1980) for a further discussion.

Some extensions. It may often be worth while to consider a modified criterion (3.21) where the prediction errors $\varepsilon(t,\theta)$ (or, equivalently the input–output sequences) first are filtered through a filter $L(q)$:

$$\varepsilon_F(t,\theta) = L(q)\varepsilon(t,\theta). \tag{3.23}$$

This is, however, equivalent to replacing the noise model $H(q,\theta)$ by $H(q,\theta)/L(q)$. See (3.20) and, for a further discussion, Wahlberg and Ljung (1984). Prefiltering the data thus corresponds to selecting another noise model set.

Also the use of k-step ahead predictors in (3.19) might be useful. As elaborated on in Wahlberg and Ljung (1984), k-step ahead prediction methods are equivalent to replacing $H(q,\theta)$ by

$$H(q,\theta)M_k^{-1}(q,\theta) \tag{3.24}$$

where $M_k(q,\theta)$ are the first k terms in the Laurent expansion of $H(q,\theta)$. The use of k-step ahead predictors is thus equivalent to prefiltering ($L(q) = M_k(q,\theta)$) or to selecting another noise model set.

3.5. *Frequency domain interpretation of the prediction error methods*

For future use we shall develop frequency domain expressions for the sum in (3.21). The calculations are closely related to those in Solo (1978) and Ljung and Glover (1981). Since several finite convolutions with non-periodic signals are involved in the expressions, some care must be exercised in the development. For easy reading we allow ourselves to deal with those difficulties somewhat carelessly in this section—a formal treatment is given in the Appendix.

Consider the sum

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon^2(t,\theta). \tag{3.25}$$

Let $E_N(2\pi k/N, \theta)$, $k = 0, 1, \ldots N-1$ be the DFT of $\varepsilon(t,\theta)$ $t = 1, \ldots N$:

$$E_N(2\pi k/N, \theta) = \frac{1}{\sqrt{N}}\sum_{t=1}^{N} \varepsilon(t,\theta)e^{-2\pi ikt/N}. \tag{3.26}$$

Then, by Parseval's relation:

$$V_N(\theta) = \sum_{k=0}^{N-1} |E_N(2\pi k/N, \theta)|^2. \tag{3.27}$$

Now, let

$$z(t,\theta) = G(q,\theta)u(t).$$

Then approximately, the DFT of $z(t,\theta)$ is

$$Z_N(2\pi k/N, \theta) = G(e^{2\pi ik/N}, \theta)U_N(2\pi k/N). \tag{3.28}$$

Then the DFT of $s(t,\theta) = y(t) - z(t,\theta)$ is

$$S_N(2\pi k/N, \theta) = Y_N(2\pi k/N) - G(e^{2\pi ik/N}, \theta)U_N(2\pi k/N). \tag{3.29}$$

Finally, the DFT of

$$\varepsilon(t,\theta) = H^{-1}(q,\theta)s(t,\theta)$$

is, approximately

$$E_N(2\pi k/N, \theta) = H^{-1}(e^{2\pi ik/N}, \theta)S_N(2\pi k/N, \theta).$$

Consequently,

$$\begin{aligned} V_N(\theta) &\approx \frac{1}{N}\sum_{k=0}^{N-1} |H(e^{2\pi ik/N}, \theta)|^{-2}\,|Y_N(2\pi k/N) - G(e^{2\pi ik/N}, \theta)U_N(2\pi k/N)|^2 \\ &\approx \frac{1}{2\pi}\int_0^{2\pi} |H(e^{i\omega}, \theta)|^{-2}\,|Y_N(\omega) - G(e^{i\omega}, \theta)U_N(\omega)|^2 d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} |H(e^{i\omega}, \theta)|^{-2}\,|Y_N(\omega) - G(e^{i\omega}, \theta)U_N(\omega)|^2 d\omega \end{aligned} \tag{3.30}$$

where the last approximation is the approximation of the Riemann sum by the corresponding integral. According to the Appendix the error in the first approximation is bounded by $C/\sqrt{N}$. The last equality follows from the fact that all the functions of frequency are periodic with period 2π.

4. THE EMPIRICAL TRANSFER FUNCTION ESTIMATE (ETFE)

4.1. *Fourier analysis as an extension of frequency analysis*

We shall in this section define a quantity that will turn out to be a common denominator for the different approaches to transfer function estimation. Let us use the simple frequency analysis method as a starting-point. By slightly rearranging the correlation equations (3.3) we find that

$$I_c(N) - iI_s(N) = \sqrt{N}\, Y_N(\omega_0) \tag{4.1}$$

with Y_N defined by (3.10). Then (3.4) can be rearranged as

$$\hat{G}_N(e^{i\omega_0}) = \sqrt{N}\, Y_N(\omega_0)/(N\,A/2). \tag{4.2}$$

Since the input is given by (3.1), we know as in (2.9) that

$$U_N(\omega_0) = \sqrt{N}A/2.$$

Consequently, frequency analysis by the correlation method actually employs the following estimate:

$$\hat{G}_N(e^{i\omega_0}) = Y_N(\omega_0)/U_N(\omega_0). \tag{4.3}$$

In (4.3) ω_0 is the frequency of the input sinusoid. In a linear system, different frequencies pass the system independently of each other. It is therefore quite natural to extend the frequency analysis estimate (4.3) also to the case of multifrequency inputs. That is, we introduce the estimate

$$\hat{\hat{G}}_N(e^{i\omega}) = Y_N(\omega)/U_N(\omega) \tag{4.4}$$

with Y_N and U_N defined by (3.10) and (2.3), respectively. This estimate also results from a direct application of Fourier analysis and will be called the *empirical transfer function estimate* (ETFE), for reasons that we shall discuss shortly. In (4.4) we assume of course that $U_N(\omega) \neq 0$. If this does not hold for some frequencies, we simply regard ETFE as undefined at those frequencies.

We call this estimate *empirical* since no other assumptions have been imposed than linearity of the system. In the case of multifrequency inputs, ETFE consists of $N/2$ essential points. [Recall that frequencies intermediate to the grid $\omega = 2\pi k/N$ $k = 0, 1, \ldots N-1$ are obtained by trigonometrical interpolation. Also since y and u are real we have $G(e^{2\pi ik/N}) = G(e^{2\pi i(N-k)/N})$.]

The original data sequence consisting of $2N$ numbers $y(t)$, $u(t)$, $t = 1, 2, \ldots, N$ has thus been condensed into the N numbers

$$\mathrm{Re}\hat{\hat{G}}_N(e^{2\pi ik/N}), \mathrm{Im}\hat{\hat{G}}_N(e^{2\pi ik/N}) k = 0, 1, \ldots, N/2 - 1.$$

This is quite a modest data reduction, revealing that most of the information contained in the original, data y, u still is quite "raw".

The ETFE can be interpreted in several ways: it is, as we saw, the direct extension of the frequency response analysis to multifrequency inputs; it can be seen as a way of (approximately) solving the set of convolution equations

$$y(t) = \sum_{k=1}^{N} g_k u(t-k) \quad t = 1, 2, \ldots, N \tag{4.5}$$

for g_k $k = 1, 2, \ldots, N$ using Fourier techniques; furthermore it is obtained as a special case of the spectral analysis estimate (3.7) if the DFTs are used for cross spectrum and input spectrum estimates. Let us now turn to the properties of the ETFE.

4.2. *Properties of the ETFE*

We shall in this section investigate the properties of $\hat{\hat{G}}_N$, as defined by (4.4). In order to do that, we have to use some assumptions regarding the properties of the data y^N, u^N, *viz.* that they have been generated by the system (1.1)–(1.7).

It is worth while to point out the difference between these assumptions that we introduce for the analysis, and other hypotheses and "prejudices" about the system that we will introduce (in later sections) as "alibis" to try to improve on the estimate (4.4). The former assumptions can of course never be "verified" for real data and their purpose is rather to make an analytical test of the identification method in question: suppose that we have data with the properties (1.1)–(1.7), what would the properties of the ETFE then be? This is a clear analytic counterpart to simulation studies. From this point of view, the assumptions in question allow for "tests" on a fairly wide range of systems, *viz.* all linear ones with additive, stationary noise.

We can now state the properties of the estimate (4.4). Similar results are also given in Brillinger (1981).

Lemma 4.1. Assume that (1.1)–(1.7) hold with

$$\sum_{k=1}^{\infty} k|g_0(k)| < \infty; \quad \sum_{\tau=1}^{\infty} \tau|r_v(\tau)| < \infty. \tag{4.6}$$

Further assume that $\{u(t)\}$ is a deterministic, bounded sequence. Then the ETFE (4.4) has the following properties:

$$E\hat{G}_N(e^{i\omega}) = G_0(e^{i\omega}) + \rho_1(N)/U_N(\omega) \quad (4.7a)$$

where

$$|\rho_1(N)| \le C_1/\sqrt{N} \quad (4.7b)$$

$$E[\hat{G}_N(e^{i\omega}) - G_0(e^{i\omega})][\hat{G}_N(e^{-i\xi}) - G_0(e^{-i\xi})]$$

$$= \begin{cases} \dfrac{1}{|U_N(\omega)|^2}[\Phi_v(\omega) + \rho_2(N)] & \text{if } \xi = \omega \\ \dfrac{\rho_2(N)}{U_N(\omega)U_N(-\xi)} & \text{if } |\xi - \omega| = 2\pi k/N \end{cases}$$

$$k = 1, 2, \ldots, N-1 \quad (4.8a)$$

where

$$|\rho_2(N)| \le C_2/N. \quad (4.8b)$$

Here U_N is defined by (2.3), and we restrict ourselves to frequencies for which $\hat{G}_N$ is defined. The constants can be taken as

$$C_1 = \left(2\sum_{k=1}^{\infty}|kg_0(k)|\right)\cdot \max|u(t)| \quad (4.9)$$

$$C_2 = C_1^2 + \sum_{\tau=-\infty}^{\infty}|\tau r_v(\tau)|. \quad (4.10)$$

If $\{u(t)\}$ is periodic with period N, then $C_1 = 0$. □

Remark. Note that the input is regarded as a given sequence. Probabilistic quantities, such as "E" ("bias" and "variance") refer to the probability space of $\{v(t)\}$. This does, of course, not exclude that the input may be generated as a realization of a stochastic process, independent of $\{v(t)\}$.

Proof. We have from (3.10) and (1.1)

$$Y_N(\omega) = \frac{1}{\sqrt{N}}\sum_{t=1}^{N} y(t)e^{-it\omega}$$

$$= \frac{1}{\sqrt{N}}\sum_{t=1}^{N}\sum_{k=1}^{\infty} g_0(k)u(t-k)e^{-it\omega} + \frac{1}{\sqrt{N}}\sum_{t=1}^{N} v(t)e^{-it\omega}. \quad (4.11)$$

Denote the last term by $V_N(\omega)$, and consider the first one:

$$\frac{1}{\sqrt{N}}\sum_{t=1}^{N}\sum_{k=1}^{\infty} g_0(k)u(t-k)e^{-it\omega}$$

$$= [\text{change variables: } t-k=\tau]$$

$$= \frac{1}{\sqrt{N}}\sum_{k=1}^{\infty} g_0(k)e^{-ik\omega}\sum_{\tau=-k}^{N-k} u(\tau)e^{-i\tau\omega}.$$

Now,

$$\left|U_N(\omega) - \frac{1}{\sqrt{N}}\sum_{\tau=-k}^{N-k} u(\tau)e^{-i\tau\omega}\right|$$

$$\le \left|\frac{1}{\sqrt{N}}\sum_{\tau=-k}^{0} u(\tau)e^{-i\tau\omega}\right| + \left|\frac{1}{\sqrt{N}}\sum_{N-k+1}^{N} u(\tau)e^{-i\tau\omega}\right| \quad (4.12)$$

$$\le \frac{2}{\sqrt{N}}k\max|u(\tau)|.$$

Hence

$$Y_N(\omega) = G_0(e^{i\omega})\cdot U_N(\omega) + V_N(\omega) + M_N(\omega) \quad (4.13)$$

where

$$|M_N(\omega)|$$

$$= \left|\sum_{k=1}^{\infty} g_0(k)e^{-ik\omega}\left(\frac{1}{\sqrt{N}}\sum_{\tau=-k}^{N-k} u(\tau)e^{-i\tau\omega} - U_N(\omega)\right)\right|$$

$$\le \sum_{k=1}^{\infty}\frac{2k}{\sqrt{N}}\max|u(\tau)||g_0(k)| \le C_1/\sqrt{N} \quad (4.14)$$

using (4.12). Consequently,

$$\hat{G}_N(e^{i\omega}) = G_0(e^{i\omega}) + \frac{V_N(\omega)}{U_N(\omega)} + \frac{M_N(\omega)}{U_N(\omega)}. \quad (4.15)$$

Since $\{v(t)\}$ has zero mean,

$$EV_N(\omega) = 0 \quad \forall\omega$$

and hence (4.7) and (4.9) have been proved. To prove (4.8) we first consider

$$EV_N(\omega)V_N(-\xi)$$

$$= \frac{1}{N}\sum_{r=1}^{N}\sum_{s=1}^{N} E\,v(r)e^{-i\omega r}v(s)e^{+i\xi s}$$

$$= \frac{1}{N}\sum_{r=1}^{N}\sum_{s=1}^{N} e^{i(\xi s-\omega r)} r_v(r-s) = [r-s=\tau] \quad (4.16)$$

$$= \frac{1}{N}\sum_{r=1}^{N} e^{i(\xi-\omega)r} \cdot \sum_{\tau=1-r}^{N-r} r_v(\tau)e^{-i\xi\tau}.$$

Now

$$\sum_{\tau=1-r}^{N-r} r_v(\tau)e^{-i\xi\tau} = \Phi_v(\xi) - \sum_{\tau=-\infty}^{-r} e^{-i\xi\tau} r_v(\tau)$$

$$\cdot - \sum_{\tau=N-r+1}^{\infty} e^{-i\xi\tau} r_v(\tau) \quad (4.17)$$

$$\frac{1}{N}\sum_{r=1}^{N} e^{i(\xi-\omega)r} = \begin{cases} 1 & \text{if } \xi = \omega \\ 0 & \text{if } (\xi-\omega) = \dfrac{k2\pi}{N} \end{cases}$$

$$k = \pm 1, \pm 2, \ldots \pm(N-1). \quad (4.18)$$

Consider

$$\left|\frac{1}{N}\sum_{r=1}^{N} e^{i(\xi-\omega)r} \cdot \sum_{\tau=-\infty}^{-r} e^{-i\xi\tau} r_v(\tau)\right|$$

$$\leq \frac{1}{N}\sum_{r=1}^{N}\sum_{\tau=-\infty}^{-r} |r_v(\tau)|$$

$$\leq \frac{1}{N}\sum_{\tau=-\infty}^{-1} |\tau||r_v(\tau)| \leq C/N \quad (4.19)$$

according to (4.6). Similarly

$$\left|\frac{1}{N}\sum_{r=1}^{N} e^{i(\xi-\omega)r} \cdot \sum_{\tau=N-r+1}^{\infty} e^{-i\xi\tau} r_v(\tau)\right|$$

$$\leq \frac{1}{N}\sum_{\tau=1}^{\infty} \tau\,|r_v(\tau)| \leq C/N. \quad (4.20)$$

Combining (4.16)–(4.20) we find that

$$EV_N(\omega)V_N(-\xi) = \begin{cases} {}_v(\omega) + \rho_2(N) & \text{if } \xi = \omega \\ \rho_2(N) & \text{if } |\xi-\omega| = k2\pi/N, \end{cases}$$

$$k = 1, 2, \ldots, N-1 \quad (4.21)$$

with $|\rho_2(N)| \leq 2C/N$.

Finally, using (4.15)

$$E[\hat{G}_N(e^{i\omega}) - G_0(e^{i\omega})][\hat{G}_N(e^{-i\xi}) - G_0(e^{-i\xi})]$$

$$= \frac{1}{U_N(\omega)U_N(-\xi)}[EV_N(\omega)V_N(-\xi)$$

$$+ M_N(\omega)M_N(-\xi)]$$

since $EV_N(\omega)M_N(\omega) = 0$.

This together with (4.21) and (4.14) proves (4.8) and (4.10). □

The lemma, together with the results of Section 2, tells us the following.

Case 1: The input is periodic with period N_0 and $N = rN_0$.

—The ETFE $\hat{G}_N(e^{i\omega})$ is defined only for frequencies $\omega = 2\pi s/N_0$ $s = 1, 2, \ldots, N_0 - 1$ and such that $A_s \neq 0$ [see (2.10)–(2.12)].

—It is an unbiased estimate of the transfer function at these frequencies.

—The variances of these estimates decay as $\frac{N_0}{N}\Phi_v(\omega)/|A_r|^2$ for $\omega = 2\pi r/N_0$ as N increases. Here A_r is given by (2.10)–(2.12).

—The covariances for different frequency estimates decay as $1/N^2$.

Case 2: The input is a realization of stationary stochastic process.

In this case the periodogram $|U_N(\omega)|^2$ is an erratic function that "on the average" behaves like the input spectrum $\Phi_u(\omega)$. Suppose that

$$0 < m \leq \Phi_u(\omega) \leq M < \infty. \quad (4.22)$$

Then

—The ETFE is an asymptotically unbiased estimate of the transfer function at increasingly many frequencies as N increases.

—The variance of the ETFE does not decrease as N increases and it is given as the noise-to-signal ratio at the frequency in question.

—The estimates at different frequencies are asymptotically uncorrelated.

Remark. Note that although $E\hat{G}_N = G_0$, it does not follow that $E|\hat{G}| = |G_0|$. In fact, for large N

$$E|\hat{G}_N(e^{i\omega})|^2 \cong |G_0(e^{(\omega)}|^2 + \Phi_v(\omega)/|U_N(\omega)|^2.$$

5. IMPROVING ON THE ETFE

We shall in this section mainly discuss Case 2 of the previous section. Then the ETFE, $\hat{G}$ has the nice

feature of providing asymptotically unbiased and uncorrelated estimates of the true transfer function at $N/2$ different frequencies. The disturbing feature is of course that the variance does not decay with N, but remains equal to the noise-to-signal ratio at the corresponding frequency. This latter property makes the empirical estimate useless in most cases in practice.

It is easy to understand the reason why the variance does not decrease with N. We determine as many independent estimates as we have pairs of data points. The information per estimated quantity does not increase as more data is collected. In other words, we have no feature of data and information compression. This in turn is due to the fact that we have only assumed linearity about the true system. Consequently, the system's properties at different frequencies may be totally unrelated.

From this it also follows that the only possibility to increase the information per estimated parameter is to assume that the system's behaviour at one frequency is related to that at another. In other words, that the estimated transfer function at one frequency bears some information about the true transfer function also at other frequencies. In the subsequent sections, we shall discuss in some detail how this can be done, but let us here first dwell on some conceptual issues.

To allow for inference between different frequencies we introduce an *auxiliary hypothesis* about the true system. This could be of the nature (cf. Sections 6 and 7!):

The system is described by (1.1)–(1.2) where $G_0(e^{i\omega})$ is a "fairly smooth function of ω", (5.1)
or
The system is described by (1.1)–(1.2) where the function $G_0(e^{i\omega})$ belongs to a known d-dimensional manifold of functions. A typical example could be the set of all linear systems of order $\leq d/2$. (5.2)

It is true that such auxiliary hypotheses have the character of "prejudices", since they assume something about what we are going to investigate. On the other hand, such prejudices may be well founded in experience, prior knowledge or physical insight. They are the only way of getting further than the ETFE, which is the only "unprejudiced" TFE.

With hypotheses like (5.1) and (5.2) we can form other estimates $\hat{G}$ of the TF from $\hat{\hat{G}}$. Loosely speaking $\hat{G}$ will be some "smoothed" variant of $\hat{\hat{G}}$. The effect of the hypothesis is that the variance of $\hat{G}$ is reduced. A rule of thumb says that under (5.2) the variance reduces by a factor N/d on the average (see Section 7). Now, if the hypothesis is not true, the estimate will be biased. The price for the variance reduction is thus the risk of introducing bias. If the hypothesis is incorrect, but reasonable, the bias will be small. The mean square error (MSE) of the estimate $\hat{G}$ is the sum of its variance and the square of its bias. The MSE may thus very well decrease as a result of decreased variance, even though bias is introduced.

Consequently, a hypothesis (5.1)–(5.2) may provide better (in the MSE sense) TFEs even though it is not correct: "A hypothesis may be wrong but useful". This is why we prefer the term "auxiliary hypothesis", to "assumption". We may introduce and use such a hypothesis without having to believe that it is true.

6. SMOOTHING THE ETFE SPECTRAL ANALYSIS

6.1. *The idea: averaging neighbouring ETFE-values*

Suppose we adopt the prejudice (5.1) that the true transfer function $G_0(e^{i\omega})$ is a smooth function of ω. For a given ω, consider a set of integer values k such that $2\pi k/N \approx \omega$. Then, by the smoothness assumption, the corresponding estimates

$$\hat{\hat{G}}_N(e^{2\pi ik/N}) \quad k \text{ integer}, 2\pi k/N \approx \omega \tag{6.1}$$

are independent, unbiased estimates of roughly the same constant $G_0(e^{i\omega})$, each with a variance of

$$\Phi_v(2\pi k/N)/|U_N(2\pi k/N)|^2.$$

Here we neglected terms that tend to zero as N tends to infinity.

If we assume $G_0(e^{i\omega})$ to be constant over the interval

$$2\pi k_1/N = \omega_0 - \Delta\omega < \omega < \omega_0 + \Delta\omega = 2\pi k_2/N, \tag{6.2}$$

then it is well known that the best way to estimate this constant is to form a weighted average of the measurements (6.1), each measurement weighted according to its inverse variance:

$$\hat{G}_N(e^{i\omega_0}) = \frac{\sum_{k=k_1}^{k_2} \alpha_k \hat{\hat{G}}_N(e^{2\pi ik/N})}{\sum_{k=k_1}^{k_2} \alpha_k} \tag{6.3a}$$

$$\alpha_k = |U_N(2\pi k/N)|^2/\Phi_v(2\pi k/N). \tag{6.3b}$$

For large N we could with good approximation work with the integrals that correspond to the (Riemann) sums in (6.3a):

$$\hat{G}_N(e^{i\omega_0}) = \frac{\int_{\xi=\omega_0-\Delta\omega}^{\omega_0+\Delta\omega} \alpha(\xi)\hat{\hat{G}}_N(e^{i\xi})d\xi}{\int_{\xi=\omega_0-\Delta\omega}^{\omega_0+\Delta\omega} \alpha(\xi)d\xi} \tag{6.4a}$$

$$\alpha(\xi) = |U_N(\xi)|^2/\Phi_v(\xi). \tag{6.4b}$$

If the transfer function G_0 is not constant over the interval (6.2) it is reasonable to use an additional weighting that pays more attention to frequencies close to ω_0:

$$\hat{G}_N(e^{i\omega_0}) = \frac{\int_{-\pi}^{\pi} W_\gamma(\xi - \omega_0)\alpha(\xi)\hat{\hat{G}}_N(e^{i\xi})d\xi}{\int_{-\pi}^{\pi} W_\gamma(\xi - \omega_0)\alpha(\xi)d\xi}. \tag{6.5}$$

Here $W_\gamma(\xi)$ is a function centered around $\xi = 0$ and γ is a "shape parameter" that we shall discuss shortly. Typical forms of $W_\gamma(\xi)$ are shown in Fig. 1.

Clearly (6.4) corresponds to

$$W_\gamma(\xi) = \begin{cases} 1 & |\xi| < \Delta\omega \\ 0 & |\xi| > \Delta\omega. \end{cases} \tag{6.6}$$

Now, if the noise spectrum $\Phi_v(\omega)$ is known, the estimate (6.5) can be realized as written. If $\Phi_v(\omega)$ is not known we could argue as follows.

Suppose that the noise spectrum does not change very much over frequency intervals corresponding to the "width" of the weighting function $W_\gamma(\xi)$:

$$\int_{-\pi}^{\pi} W_\gamma(\xi - \omega_0)\left|\frac{1}{\Phi_v(\xi)} - \frac{1}{\Phi_v(\omega_0)}\right|d\xi = \text{"small"}. \tag{6.7}$$

Then $\alpha(\xi)$ in (6.4b) can be replaced by $\alpha(\xi) = |U_N(\xi)|^2/\Phi_v(\omega_0)$, which means that the constant $\Phi_v(\omega_0)$ cancels when (6.5) is formed. Under (6.7) the estimate

$$\hat{G}_N(e^{i\omega_0}) = \frac{\int_{-\pi}^{\pi} W_\gamma(\xi - \omega_0)\beta(\xi)\hat{\hat{G}}_N(e^{i\xi})d\xi}{\int_{-\pi}^{\pi} W_\gamma(\xi - \omega_0)\beta(\xi)d\xi} \tag{6.8a}$$

$$\beta(\xi) = |U_N(\xi)|^2 \tag{6.8b}$$

is thus a good approximation of (6.4b), (6.5).

Now

$$\beta(\xi)\hat{\hat{G}}_N(e^{i\xi}) = |U_N(\xi)|^2\frac{Y_N(\xi)}{U_N(\xi)} = Y_N(\xi)\bar{U}_N(\xi) \tag{6.9}$$

so comparing with (3.7)–(3.9) we see that (6.8) indeed is the conventional (Blackman–Tukey) spectral analysis estimate of the transfer function! We may remark that if (6.7) does not hold, it might be better to include a procedure where $\Phi_v(\omega)$ is estimated and use that estimate in (6.5).

6.2. *The weighting function* $W_\gamma(\xi)$ (*"frequency window"*)

Let us now discuss the weighting function $W_\gamma(\xi)$. In spectral analysis, it is often called the "frequency window". If this window is "wide" then many different frequencies will be weighted together in (6.5). This should lead to a small variance of $\hat{G}(e^{i\omega_0})$. At the same time, a wide window will involve frequency estimates further away from ω_0, with expected values that may differ considerably from $G_0(e^{i\omega_0})$. This will cause large bias. The width of the window will thus control the trade off between bias and variance. To make this trade-off a bit more formal we shall use the scalar γ to describe the width.

Suppose that we have chosen a basic *shape* of the weighting function $W(\xi)$. Then this shape can be transformed into more narrow ones by

$$W_\gamma(\xi) = \gamma W(\gamma\xi). \tag{6.10}$$

A large value of γ will thus mean a narrow frequency window. (Other ways of scaling the function have also been discussed: see Brillinger, 1981.)

Some typical basic shapes were depicted in Fig. 1. We shall require from basic shapes that

$$\int_{-\infty}^{\infty} W(\xi)d\xi = 1 \qquad \int_{-\infty}^{\infty} \xi W(\xi)d\xi = 0$$

$$\int_{-\infty}^{\infty} \xi^2 W(\xi)d\xi = 1 \tag{6.11a}$$

$$\int_{-\infty}^{\infty} |\xi|^3 W(\xi)d\xi = C_1 \qquad \int_{-\infty}^{\infty} W^2(\xi)d\xi = W/2\pi. \tag{6.11b}$$

Equation (6.11a) means that we can think of $W(\xi)$ as the probability density function of a random variable with zero mean and unit variance. The number W then becomes a parameter that characterizes the basic shape of the weighting function.

6.3. *Asymptotic properties of the smoothed estimate*

The estimate (6.8) has been studied in several treatments of spectral analysis. See, e.g. Chapter 10 in Jenkins and Watts (1969), or Chapter 6 of Brillinger (1981).

Results that are asymptotic in both N and γ can be quoted as follows. Consider the estimate (6.8)–(6.11). Then

Bias:

$$E\hat{G}_N(e^{i\omega}) - G_0(e^{i\omega}) = \frac{1}{\gamma^2}\left[G_0''(e^{i\omega}) + G_0'(e^{i\omega})\frac{\Phi_u'(\omega)}{\Phi_u(\omega)}\right] + \underset{\gamma\to\infty}{\mathcal{O}}(C_1/\gamma^3) + \underset{N\to\infty}{\mathcal{O}}(1/\sqrt{N}). \tag{6.12}$$

Prime and double prime denote differentiation w.r.t. ω, once and twice, respectively. $\mathcal{O}(x)$ is a term that tends to zero at the same rate as x.

Variance:

$$E|\hat{G}_N(e^{i\omega}) - E\hat{G}_N(e^{i\omega})|^2 = \frac{\gamma}{N} W \frac{\Phi_v(\omega)}{\Phi_u(\omega)} + \underset{\gamma, N \to \infty \ \gamma/N \to 0}{o(\gamma/N)}. \tag{6.13}$$

Here $o(x)$ is a term that tends to zero faster than x. We repeat that expectation here is with respect to the noise sequence $\{v(t)\}$ and that the input is supposed to be deterministic, subject to (2.1)–(2.2). [It could very well be a realization of a stochastic process, independent of $\{v(t)\}$.]

The asymptotic expressions (6.12)–(6.13) confirm the trade-off between bias and variance when γ varies. To formalize this trade-off, let us use these expressions to evaluate the mean square error (MSE):

$$E|\hat{G}_N(e^{i\omega}) - G_0(e^{i\omega})|^2 \sim \frac{1}{\gamma^4}|R(\omega)|^2 + \frac{\gamma}{N} W \frac{\Phi_v(\omega)}{\Phi_u(\omega)}. \tag{6.14}$$

Here

$$R(\omega) = G_0''(e^{i\omega}) + G_0'(e^{i\omega}) \frac{\Phi_u'(\omega)}{\Phi_u(\omega)}. \tag{6.15}$$

From (6.14) we see that a desired property of the basic window shape (6.11) is that W should be small. We may also calculate the value of the width parameter γ that minimizes the MSE. Asymptotically in N and γ, it is given by

$$\gamma_{\text{opt}} = \left(\frac{4(R(\omega))^2 \Phi_u(\omega)}{W \Phi_v(\omega)} \right) \cdot N^{1/5} \tag{6.16}$$

This value can of course not be realized by the user, since the constant contains several unknown quantities. We note however, that in any case it increases as $N^{1/5}$ and it should, in principle, be allowed to be frequency-dependent. Recall that (6.16) is based on expressions that are asymptotic as γ tends to infinity, so it may be meaningful only for large N.

The optimal choice of γ leads to a mean square error that decays as

$$\text{MSE} \sim C N^{-4/5} \tag{6.17}$$

6.4. *The idea of prefiltering*

The bias expression (6.12) gives a clue how to improve the spectral analysis estimate. Suppose that we prefilter the input and output sequences by linear filters $L_u(q)$ and $L_y(q)$ respectively:

$$\begin{aligned} u_F(t) &= L_u(q)u(t) \\ y_F(t) &= L_y(q)y(t). \end{aligned} \tag{6.18}$$

Then the true relationship (1.1) becomes

$$y_F(t) = G_0^F(q)u_F(t) + v_F(t) \tag{6.19}$$

with

$$v_F(t) = L_y(q)v(t) \tag{6.20}$$

and

$$G_0^F(q) = \frac{L_y(q)}{L_u(q)} G_0(q). \tag{6.21}$$

Applying the spectral analysis techniques to the filtered sequences gives an estimate $\hat{G}_N^F(q)$ of $G_0^F(q)$ that has properties according to (6.12)–(6.15) with (6.19). We then form

$$\hat{G}_N(e^{i\omega}) = \frac{L_u(e^{i\omega})}{L_y(e^{i\omega})} \hat{G}_N^F(e^{i\omega}). \tag{6.22}$$

It is immediate to verify that the variance of (6.22) is still given by (6.13). The bias (6.12) can however be reduced by selecting L_u and L_y (using *a priori* information) so that the derivatives of $G_0^F(e^{i\omega})$ and $\Phi_{u_F}(\omega) = |L_u(e^{i\omega})|^2 \Phi_u(\omega)$ are small.

6.5. *Other ways of smoothing the ETFE*

There are other ways to smooth the ETFE, than forming weighted averages. One possibility is to split the data set into M batches, each containing R data ($N = RM$) and to compute the ETFEs of these sub-batches. Then form the average over these:

$$\hat{G}_N(e^{i\omega}) = \frac{1}{M} \sum_{k=1}^{M} \hat{G}_R^{(k)}(e^{i\omega}). \tag{6.23}$$

See Brillinger (1981), where this idea is described in detail for spectral analysis.

Another approach is to introduce a "roughness measure" for $\hat{G}_N(e^{i\omega})$, e.g.

$$\kappa(G(e^{i\omega})) = \int_{-\pi}^{\pi} |G'(e^{i\omega})|^2 d\omega \tag{6.24}$$

and minimize

$$V(G) = \int_{-\pi}^{\pi} |\hat{\hat{G}}_N(e^{i\omega}) - G(e^{i\omega})|^2 d\omega + \kappa(G(e^{i\omega})) \tag{6.25}$$

w.r.t. G, giving $\hat{G}_N$. See Gersch and Kitagawa (1985) for further details.

7. APPROXIMATING THE ETFE WITH SMOOTH TRANSFER FUNCTIONS

In this section we shall adopt the prejudice (5.2) that the true transfer function can be well approximated by a member in a prescribed set of functions. Let this set of functions be indexed by a parameter vector θ of dimension d:

$$\mathcal{G} = \{G(e^{i\omega}, \theta) \mid \theta \in D_{\mathcal{M}} \subset R^d\}. \tag{7.1}$$

Occasionally, we shall allow the dimension d, and hence the set $\mathcal{G}$ to depend on N, the number of observed data. The problem that we are faced with is to select one member in $\mathcal{G}$ that resembles the ETFE G_N in some sense. We shall discuss several aspects of this problem here.

The objective of the section is to provide insights. Therefore we shall not dwell on technicalities, such as the exact conditions under which expressions like (2.15) hold.

7.1. *A quadratic norm criterion*

A natural idea is to pick a model TF in $\mathcal{G}$ that is close as possible to the empirical one in some weighted square norm over the frequency range:

$$\hat{G}_N(e^{i\omega}) = G(e^{i\omega}, \hat{\theta}_N) \tag{7.2a}$$

$$\hat{\theta}_N = \arg\min_{\theta \in D_{\mathcal{M}}} \int_{-\pi}^{\pi} |G(e^{i\omega}, \theta) - \hat{\hat{G}}_N(e^{i\omega})|^2 Q_N(\omega) \mathrm{d}\omega. \tag{7.2b}$$

The choice of the weighting function $Q_N(\omega)$ will be discussed later.

A useful way of interpreting the criterion (7.2) is to pose a regression model as follows. From (4.14) and (4.15) we have (asymptotically for large N):

$$\hat{\hat{G}}_N(e^{2\pi ik/N}) = G_0(e^{2\pi ik/N}) + \frac{V_N(2\pi k/N)}{U_N(2\pi k/N)} \qquad k = 0, 1, \ldots, N-1. \tag{7.3}$$

Then consider the nonlinear (in θ) regression model

$$\hat{\hat{G}}_N(e^{2\pi ik/N}) = G(e^{2\pi ik/N}, \theta) + v(k) \qquad k = 0, 1, \ldots, N-1 \tag{7.4}$$

where $\hat{\hat{G}}$ are considered as observations of the unknown transfer function. The weighted least squares (LS) method applied to (7.4) gives the criterion

$$V_N(\theta) = \sum_{k=0}^{N-1} |\hat{\hat{G}}_N(e^{2\pi ik/N}) - G(e^{2\pi ik/N}, \theta)|^2 Q_N(k). \tag{7.5}$$

This LS criterion can be seen as the Riemann sum counterpart of (7.2).

7.2. *Choice of weighting functions*

The close relationship between the criterion (7.2) and the frequency domain expressions for time-domain prediction error methods are obvious from a comparison with (3.30). We shall later in this section make use of these time-domain interpretations. Let us first, however, discuss the criterion (7.2) and what aspects should guide the choice of Q, without reference to such an interpretation.

How does the choice of Q influence the bias and variance of the estimate (7.2)? Let us first consider the bias:

Lemma 7.1. Let $\hat{\theta}_N$ be given by (7.2) where Q_N is chosen as $Q_N(\omega) = |U_N(\omega)|^2/R(\omega)$.

Then, subject to the conditions of Lemma 4.1, and smoothness properties of $G(q, \theta)$ with respect to θ,

$$\hat{\theta}_N \to \theta^* \text{ with probability 1 as } N \to \infty \tag{7.6a}$$

where

$$\theta^* = \arg\min_{\theta \in D_{\mathcal{M}}} \int_{-\pi}^{\pi} |G(e^{i\omega}, \theta) - G_0(e^{i\omega})|^2 \bar{Q}(\omega) \mathrm{d}\omega \tag{7.6b}$$

$$\bar{Q}(\omega) = \Phi_u(\omega)/R(\omega). \tag{7.7}$$

Indication of proof. The lemma is a frequency domain formulation of the convergence results, given e.g. in Ljung (1978) (see that reference for a full, formal proof). It could also be proved by direct means in the frequency domain using (4.15):

$$\int_{-\pi}^{\pi} |\hat{\hat{G}}_N(e^{i\omega}) - G(e^{i\omega}, \theta)|^2 Q_N(\omega) \mathrm{d}\omega$$

$$= \int_{-\pi}^{\pi} |G_0(e^{i\omega}) - G(e^{i\omega}, \theta)|^2 |U_N(\omega)|^2/R(\omega) \mathrm{d}\omega$$

$$+ \int_{-\pi}^{\pi} 2Re\{(G_0(e^{i\omega}) - G(e^{i\omega}, \theta)) V_N(-\omega) U_N(\omega)\}/R(\omega) \mathrm{d}\omega$$

$$+ \mathcal{O}(1/\sqrt{N}) + \theta - \text{independent terms.}$$

The first term converges to the integral in (7.6b)–(7.7), uniformly in θ, in view of (2.15). The second term can be shown to tend to zero, uniformly

in θ, since, analogously to (2.15)

$$\int_{-\pi}^{\pi} V_N(-\omega)U_N(\omega)f(\omega)\mathrm{d}\omega \to \int_{-\pi}^{\pi} \Phi_{uv}(\omega)f(\omega)\mathrm{d}\omega = 0$$

where Φ_{uv} is the cross-spectrum between input and noise (which is zero, since these are independent) and f is a sufficiently smooth function. □

The value θ^* will thus be the a.s. limit of the estimate $\hat{\theta}_N$. For the most common model families (7.1) we also have

$$E\hat{\theta}_N = \theta^* + \underset{N\to\infty}{\mathcal{O}}(1/N), \tag{7.8}$$

see Ljung and Yuan (1985a,b) and Ljung (1984).

From (7.6) we directly notice one aspect on the choice of $\bar{Q}$. If $G_0 \in \mathcal{G}$ then $G(e^{i\omega}, \theta^*) = G_0(e^{i\omega})$ regardless of $\bar{Q}$ at least at those frequencies where $\bar{Q}(\omega) \neq 0$. If $G_0 \notin \mathcal{G}$ then the weighting function $\bar{Q}$ will determine the bias distribution, i.e. the frequency bands at which $G(e^{i\omega}, \theta^*)$ and $G_0(e^{i\omega})$ are close. A high value of $\bar{Q}$ at certain frequencies indicates that we require the bias in $\hat{G}$ to be small at those frequencies.

For the variance aspect of $\hat{\theta}_N$ we have the following result.

Lemma 7.2. Suppose $G(e^{i\omega}, \theta)$ in (7.4) is linear in θ and let $\hat{\theta}_N$ minimize (7.5) for a particular choice of weighting function $Q = \{Q(1), Q(2) \ldots Q(N)\}$. Suppose that the conditions of Lemma 4.1 hold. Then

$$\sqrt{N}(\hat{\theta}_N - E\hat{\theta}_N) \in AsN(O, P(Q)), \tag{7.9}$$

meaning that the random vector on the left converges in distribution to the normal distribution with zero mean and a covariance matrix P, that depends on Q. We then have

$$P(Q) \geq P(Q^*) \quad \forall Q \tag{7.10}$$

where

$$Q^*(k) = \frac{|U_N(2\pi k/N)|^2}{\varphi_v(2\pi k/N)}. \tag{7.11}$$

Indication of Proof. For linear regressions, the result is standard from statistics saying that the minimum variance estimate (BLUE) is obtained when the measurements in the LS-criterion are weighted according to their inverse variances. Notice that the result is true more generally and regardless whether $E\hat{\theta}_N$ corresponds to a "true value". □

These aspects on how the weighting function affects the estimate $\hat{G}_N$ can thus be summarized as follows.

1. Q_N affects the bias-distribution $(G(e^{i\omega}, \theta^*) - G_0(e^{i\omega}))$ over different frequencies and should reflect the user's subjective view on this problem. Let us denote the ideal weighting function according to this by $Q_N^{\mathrm{subj}}(\omega)$. (7.12)
2. Q_N affects the variance of $\hat{\theta}_N$, and hence of $\hat{G}_N$. The smallest variance is obtained for

$$Q_N^{\mathrm{obj}}(\omega) = |U_N(\omega)|^2/\Phi_v(\omega) \tag{7.13}$$

i.e. the frequency dependent signal-to-noise ratio (SNR). This Q reflects the relative reliability of $\hat{G}_N$ at different frequencies, and is as such an objective quantity. We may stress that in order to minimize the variance it is the true noise spectrum $\Phi_v(\omega)$ that should be used in (7.13), even if the system estimate is biased, causing the model error to be different from the noise [see (7.27) below].

A good mean square error

$$|\hat{G}_N(e^{i\omega}) - G_0(e^{i\omega})|^2$$

is obtained as a suitable trade-off between these aspects. Let us first treat some direct consequences.

The *first* aspect, (7.12), is subjective and not directly related to properties of the data. The ideal choice Q^{subj} from this point of view should be possible to reach *a priori*, based on the intended use of the resulting model.

The *second* aspect (7.13) is related to the SNR. While U_N is a known quantity, the noise spectrum $\Phi_v(\omega)$ is typically unknown. The weighting function is from this point of view "a noise model", see (7.13). An alternative would be to estimate the noise properties from data and use that in (7.13).

7.3. *Estimating the noise spectrum*

Suppose that we disregard the aspect (7.12) (for example by assuming that $G_0 \in \mathcal{G}$) and concentrate on trying to realize (7.13). We would thus use

$$Q_N(\omega) = |U_N(\omega)|^2/R(\omega) \tag{7.14}$$

where we want to have $R(\omega)$ as (an estimate of) $\Phi_v(\omega)$. How can this be achieved?

Let a function $R(\omega, \eta)$ be parametrized by a parameter vector η and consider

$$\min_{\eta \in D_\eta} \int_{-\pi}^{\pi} \Phi_v(\omega)/R(\omega, \eta)\mathrm{d}\omega \tag{7.15}$$

where $\Phi_v(\omega)$ is a given positive function that can be represented as

$$\Phi_v(\omega) = |H_0(e^{i\omega})|^2. \tag{7.16}$$

Clearly the function $R(\omega, \eta)$ must be normalized in some way, so as to make the problem (7.15) meaningful. R is a positive function, so we may choose the following normalization:

$$R(\omega, \eta) = |\Gamma(e^{i\omega}, \eta)|^2 \tag{7.17}$$

where

$$\Gamma(z, \eta) = \gamma_0 + \sum_{k=1}^{\infty} \gamma_k(\eta) z^k \tag{7.18}$$

is a well-defined expansion. The important normalization here is that γ_0 does not depend on η.

Now define

$$\frac{H_0(z)}{\Gamma(z, \eta)} = \bar{\gamma}_0 + \sum_{k=1}^{\infty} \delta_k(\eta) z^k \tag{7.19}$$

Then the minimization problem (7.15) can be rewritten as

$$\begin{aligned} \min_{\eta} \int_{-\pi}^{\pi} \Phi_v(\omega)/R(\omega, \eta) \mathrm{d}\omega \\ = \min_{\eta} \int_{-\pi}^{\pi} \left| \frac{H_0(e^{i\omega})}{\Gamma(\eta, e^{i\omega})} \right|^2 \mathrm{d}\omega \\ = \min_{\eta} \left\{ \bar{\gamma}_0^2 + \sum_{k=1}^{\infty} \delta_k^2(\eta) \right\} \end{aligned} \tag{7.20}$$

The minimization problem in (7.15) is consequently to make the squared δ_k:s in (7.19) as small as possible. In terms of the original problem (7.15) we may say that $R(\omega, \eta)$ should be as close as possible to $\Phi_v(\omega)$ (up to scaling constant).

This shows that (7.15), together with (7.17), (7.18) is a well-posed problem for our purpose.

Now $\Phi_v(\omega)$ is not available, but we have as in (2.15) that

$$\begin{aligned} \int_{-\pi}^{\pi} |V_N(\omega)|^2/R(\omega, \eta) \mathrm{d}\omega \\ \to \int_{-\pi}^{\pi} \Phi_v(\omega)/R(\omega, \eta) \mathrm{d}\omega \quad \text{as } N \to \infty. \end{aligned} \tag{7.21}$$

A reasonable estimate thus is

$$\hat{\eta}_N = \arg\min_{\eta} \int_{-\pi}^{\pi} |V_N(\omega)|^2/R(\eta, \omega) \mathrm{d}\omega \tag{7.22}$$

with V_N given as the DFT of the noise sequence.

How do we find V_N? From (4.15) we have

$$V_N(\omega) = (\hat{\bar{G}}_N(e^{i\omega}) - G_0(e^{i\omega}))U_N(\omega). \tag{7.23}$$

Assuming that $G_0 \in \mathscr{G}$ so that, whatever the choice of $Q_N(\omega)$ in (7.2) we have

$$\hat{G}_N(e^{i\omega}) \approx G_0(e^{i\omega})$$

we could use $\hat{G}_N$ for G_0 in (7.23), and obtain from (7.22)

$$\begin{aligned} \hat{\eta}_N = \arg\min_{\eta} \int_{-\pi}^{\pi} |\hat{\bar{G}}_N(e^{i\omega}) - G(e^{i\omega}, \hat{\theta}_N)|^2 \\ |U_N(\omega)|^2/R(\omega, \eta) \mathrm{d}\omega. \end{aligned} \tag{7.24}$$

Now, so far we have assumed $\hat{\theta}_N$ to be available when we estimate $\hat{\eta}_N$. In practice this is not the case, and the problem is indeed one of jointly estimating η and θ. The problems (7.24) and (7.2b) with (7.14) can be combined into

$$\begin{aligned} \begin{pmatrix} \hat{\theta}_N \\ \hat{\eta}_N \end{pmatrix} = \arg\min_{\substack{\theta \in D_{\mathscr{M}} \\ \eta \in D_\eta}} \int_{-\pi}^{\pi} |\hat{\bar{G}}_N(e^{i\omega}) \\ - G(e^{i\omega}, \theta)|^2 Q_N(\omega, \eta) \mathrm{d}\omega \end{aligned} \tag{7.25a}$$

$$Q_N(\omega, \eta) = |U_N(\omega)|^2/R(\omega, \eta). \tag{7.25b}$$

Let us recapitulate that we have been led to the problem formulation (7.25) from (7.2) by our wish to realize (7.13). We have used a prejudice that the true system belongs to the model set $\mathscr{G}$ at two instances to motivate (7.25):

1. We have neglected the subjective aspect (7.12).
2. We have replaced G_0 by $\hat{G}_N$ when evaluating (7.23).

Note the significance of the latter point! If $\hat{\theta}_N \to \theta^*$ and $G(e^{i\omega}, \theta^*) \neq G_0(e^{i\omega})$, then it follows from (2.15) that the integral in (7.24) tends to

$$\begin{aligned} \int_{-\pi}^{\pi} \{|G_0(e^{i\omega}) - G(e^{i\omega}, \theta^*)|^2 \Phi_u(\omega) \\ + \Phi_v(\omega)\}/R(\omega, \eta) \mathrm{d}\omega \end{aligned} \tag{7.26}$$

and $R(\omega, \hat{\eta}_N)$ will approximate the error spectrum

$$\Phi_{ER}(\omega) = |G_0(e^{i\omega}) - G(e^{i\omega}, \theta^*)|^2 \Phi_u(\omega) + \Phi_v(\omega) \tag{7.27}$$

instead of the noise spectrum $\Phi_v(\omega)$. Compare with the remark following (7.13)!

7.4. *Relation to prediction error methods*

Now recall the prediction error method (3.21), (3.30):

$$\hat{\theta}_N = \arg\min_{\theta \in D_{\mathcal{M}}} \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t,\theta)$$

$$\approx \arg\min_{\theta \in D_{\mathcal{M}}} \int_{-\pi}^{\pi} |H(e^{i\omega},\theta)|^{-2} |Y_N(\omega) - G(e^{i\omega},\theta)U_N(\omega)|^2 d\omega$$

$$= \arg\min_{\theta \in D_{\mathcal{M}}} \int_{-\pi}^{\pi} \left| \frac{Y_N(\omega)}{U_N(\omega)} - G(e^{i\omega},\theta) \right|^2 \frac{|U_N(\omega)|^2}{|H(e^{i\omega},\theta)|^2} d\omega. \quad (7.28)$$

If we compare (7.28) with (7.25), (7.17), (4.4) (incorporating the parameter vector η into θ), we see that these criteria are identical with the noise model

$$\Gamma(e^{i\omega},\theta) = H(e^{i\omega},\theta). \quad (7.29)$$

It is instructive to compare the two different ways of deriving this criterion. The frequency domain one—smoothing the empirical TF estimate—stresses the role of the "noise model" as a weighting function and it also displays some underlying assumptions that support the validity of such a criterion. We shall discuss these aspects in more detail in Section 8.

7.5. *A direct link between prediction error methods and spectral analysis*

We have now established a link between the time-domain method (3.21), (3.30) and spectral analysis—they represent two different approaches to smooth the ETFE $\hat{\hat{G}}_N$. In fact, a more explicit link can also be obtained as follows. Consider the particular model parametrization

$$\theta = (\theta_1, \theta_2, \ldots, \theta_d)^T$$
$$G(e^{i\omega},\theta) = \theta_k \quad \omega_{k-1} \le \omega < \omega_k; \quad \omega_k = k\Delta\omega$$
$$G(e^{-i\omega},\theta) = \bar{\theta}_k \quad (7.30)$$
$$H(e^{i\omega},\theta) \equiv H_*(e^{i\omega}) \quad \text{(fixed noise model)}$$

where we let θ be complex valued.

Here

$$\dim \theta = d, \Delta\omega = \pi/d. \quad (7.31)$$

The prediction error estimate is given by (7.28). Differentiating the criterion function w.r.t. θ_k gives

$$\frac{\partial}{\partial\theta_k} V_N(\theta) = \frac{\partial}{\partial\theta_k} \frac{1}{\pi} \int_0^{\pi} |\hat{\hat{G}}_N(e^{i\omega}) - G(e^{i\omega},\theta)|^2 \frac{|U_N(\omega)|^2}{|H_*(e^{i\omega})|^2} d\omega = -\frac{2}{\pi} \int_{\omega_{k-1}}^{\omega_k} (\hat{\hat{G}}_N(e^{i\omega}) - G(e^{i\omega},\theta)) \frac{|U_N(\omega)|^2}{|H_*(e^{i\omega})|^2} d\omega.$$

Note that $G(e^{i\omega},\theta) = \theta_k$ over the integration interval. Now, suppose that $|H_*(e^{i\omega})|$ does not change very much over $\omega_{k-1} \le \omega \le \omega_k$. Then

$$\frac{\partial}{\partial\theta_k} V_N(\theta) = 0 \Rightarrow \omega_{k-1} \le \omega < \omega_k:$$

$$\hat{G}_N(e^{i\omega}) = \theta_k = \frac{\int_{\omega_{k-1}}^{\omega_k} \hat{\hat{G}}_N(e^{i\omega}) |U_N(\omega)|^2 d\omega}{\int_{\omega_{k-1}}^{\omega_k} |U_N(\omega)|^2 d\omega}$$

which coincides with (6.8) with the window (6.6)!

Clearly, the same result is achieved for any smooth parametrization of $H(e^{i\omega},\theta)$. Also, other and smoother parametrizations of $G(e^{i\omega},\theta)$ as a function of ω than the piecewise constant one (7.30), will lead approximately to estimates (6.8) with softer windows.

The spectral analysis estimates can thus be viewed as prediction error estimates for model structures $G(q,\theta)$ that are parametrized directly as frequency functions.

7.6. *Asymptotic properties*

The asymptotic properties of the estimate (7.25) or (3.21) when the system does not belong to the model set have been studied, e.g. in Ljung (1978), Caines (1978), Ljung and Caines (1979), Kabaila and Goodwin (1980). Results for the case where the number of estimated parameters (the parameter d) may be a function of the number of data (the parameter N) and increase to infinity are given in Ljung and Yuan (1984, 1985), Ljung (1984). Results on asymptotic expressions for variances are derived in Ljung (1985). We shall now quote some of these results.

Bias:

When N tends to infinity, the estimate (7.25), corresponding to independently parametrized transfer function and noise models will obey

$$\begin{pmatrix} \hat{\theta}_N \\ \hat{\eta}_N \end{pmatrix} \to \begin{pmatrix} \theta^* \\ \eta^* \end{pmatrix} \text{ with probability one as } N \to \infty \quad (7.32a)$$

$$\theta^* = \arg\min_{\theta} \int_{-\pi}^{\pi} |G_0(e^{i\omega}) - G(e^{i\omega},\theta)|^2 \bar{Q}(\omega,\eta^*) d\omega \quad (7.32b)$$

$$\eta^* = \arg\min_{\eta} \int_{-\pi}^{\pi} \{|G_0(e^{i\omega}) - G(e^{i\omega}, \theta^*)|^2 \Phi_u(\omega) + \Phi_v(\omega)\}/|H(e^{i\omega}, \eta)|^2 d\omega \quad (7.32c)$$

$$\bar{Q}(\omega, \eta) = \Phi_u(\omega)/|H(e^{i\omega}, \eta)|^2. \quad (7.32d)$$

Equation (7.32) is a simultaneous minimization problem that of course could be written as a joint one.

In case the transfer function model and the noise model have common parameters θ (as in (3.13)) the limiting estimate θ^* will be expressed as

$$\theta^* = \arg\min_{\theta} \int_{-\pi}^{\pi} [|G_0(e^{i\omega}) - G(e^{i\omega}, \theta)|^2 \Phi_u(\omega) + \Phi_v(\omega)]/|H(e^{i\omega}, \theta)|^2 d\omega. \quad (7.33)$$

In this case it is not possible to interpret $G(e^{i\omega}, \theta^*)$ as the best approximation of $G_0(e^{i\omega})$ in some (possibly *a priori* unknown) quadratic norm. Instead, the value of θ^* is a compromise between the fit of $G(e^{i\omega}, \theta^*)$ to G_0 and the fit of $H(e^{i\omega}, \theta^*)$ as an estimate of the residual spectrum.

Variance:

While the general expressions for the variance of $\hat{\theta}_N$ are fairly complex, it is possible to derive nice expressions for the variance of $\hat{G}_N(e^{i\omega})$ that are asymptotic in both N and d ($d = \dim \theta$), subject to certain conditions on the model structure.

A basic result is the following one, see Ljung (1985). Let $\hat{G}_N(e^{i\omega}, n)$ be the estimate of a transfer function, based on N data and obtained in a set of nth order models. Then, under certain regularity conditions

$$\sqrt{N}\,[\hat{G}_N(e^{i\omega}, n) - E\hat{G}_N(e^{i\omega}, n)] \in \underset{\text{as } N\to\infty}{AsN}(O, P_n(\omega)) \quad (7.34)$$

where

$$\lim_{n\to\infty} \frac{1}{n} P_n(\omega) = \frac{\Phi_v(\omega)}{\Phi_u(\omega)}. \quad (7.35)$$

Here (7.33) means that the random variable on the left converges in distribution to the normal distribution with zero mean and variance $P_n(\omega)$. Heuristically, we could formulate the result as

$$\operatorname{Cov} \hat{G}_N(e^{i\omega}) \sim \frac{n}{N} \cdot \frac{\Phi_v(\omega)}{\Phi_u(\omega)}. \quad (7.36)$$

Since $\hat{G}$ is a complex valued quantity, we define the covariance as

$$\operatorname{Cov} \hat{G}_N(e^{i\omega}) = E\,|\hat{G}_N(e^{i\omega}) - E\hat{G}_N(e^{i\omega})|^2. \quad (7.37)$$

A more general result is given in Ljung (1985) where the joint distribution of $\hat{G}_N$ and $\hat{H}_N$ is considered, and closed-loop identification configurations are allowed.

A particular aspect of (7.36) should be noted: asymptotically in n, the variance is independent of the noise model or weighting function. The variance is proportional to the noise-to-signal ratio for all choices of noise models. The result of Lemma 7.2 that the variance of $\hat{\theta}$ is minimized for a certain weighting function thus becomes insignificant when the model order increases.

The mean square error

Combining the expressions for bias and variance we obtain

$$E|\hat{G}_N(e^{i\omega}, n) - G_0(e^{i\omega})|^2 \sim |G_*(e^{i\omega}, n) - G_0(e^{i\omega})|^2 + \frac{n}{N} \Phi_v(\omega)/\Phi_u(\omega) \quad (7.38)$$

where we wrote

$$G_*(e^{i\omega}, n) = G(e^{i\omega}, \theta^*(n))$$

(n denoting model order).

The bias term is here implicitly defined by [see (7.32b)]

$$G_*(e^{i\omega}, n) = \arg\min_{G\in\mathcal{G}_n} \int_{-\pi}^{\pi} |G(e^{i\omega}) - G_0(e^{i\omega})|^2 \bar{Q}(\omega, \eta^*) d\omega. \quad (7.39)$$

It is instructive to compare the expression (7.38) to the corresponding one for spectral analysis; (6.14). The variance contributions to the mean squares errors are very similar; the noise to signal ratio multiplied by n/N and $\gamma W/N$ respectively. We see that the "window size"

$$\gamma W = 2\pi \int_{-\infty}^{\infty} W_\gamma^2(\xi) d\xi$$

plays the role of model order n. The expressions for the decrease of bias error as n or γ increases are however different. Intuitively, these expressions measure how "wrong" our corresponding prejudices were. In (6.14), $|R(\omega)|$ describes how fast the true transfer function varies, i.e. how much (5.1) is violated. In (7.38)–(7.39) the first term describes the distance from the true system to the set where we postulate that it would be found, cf. (5.2).

While the bias contribution decays as $1/\gamma^4$ for spectral analysis, the decrease of bias could be faster

in (7.38). This is of course due to a great advantage of prediction error methods over spectral analysis: to be able to choose specifically tailored model sets to a particular application so that the bias may decrease rapidly with n. For example, if the system is known to be exponentially stable then with a straightforward linear model, the bias term will decay as

$$C\ \lambda^n \quad \lambda < 1,$$

where λ is the base of the exponential decay of the impulse response. This gives that an optimal choice of n should behave as

$$n \sim \log N, \tag{7.40}$$

which means that the MSE may decay as

$$\text{MSE} \sim \frac{\log N}{N}.$$

Compare this with (6.17)! The prediction error approach thus allows (asymptotically) a smaller MSE for a given length of the data set. In return, we have assumed a bit more, *viz.* the true system being exponentially stable, while the spectral analysis result assumes only the true frequency function $G_0(e^{i\omega})$ to be twice differentiable.

8. TIME DOMAIN METHODS REVISITED

It is interesting that the prediction error criterion (3.30) = (7.25) can be derived via a completely different route as a smoothing operation of the ETFE. The derivation would however be little more than a formal exercise, if it did not enhance any new and useful aspects of the time-domain identification method. This is indeed the case and we shall display some such aspects in this section.

The most important feature is the dual role of the weighting function (= noise model) as a bias distributor and as a distributor of relative reliability: see Section 7.2. The subjective aspect of bias distribution, (7.12), is basically neglected in the derivation of the criteria.

Remark. It is true that the limiting estimate θ^* obtained with the time domain criterion (3.30) distributes the bias so that the model is as good as possible when applied to the problem of predicting the next output (for the particular input used at the identification experiment). Indeed, if this is our intended application, Q_N^{subj} should be chosen as (7.13). □

The reason why (7.12) is neglected is, to some extent, that most analysis has been performed under the hypothesis that $G_0 \in \mathcal{G}$. Then (7.12) is insignificant. Let us illustrate the consequences that may result otherwise.

Suppose that we are interested in the low-frequency behaviour of a system. A standard model of the type

$$A(q^{-1})y(t) = B(q^{-1})u(t) + v(t) \tag{8.1}$$

is used (compare Example 3.2). The noise model is in this case

$$H(q, \theta) = 1/A(q^{-1}).$$

The weighting function in the prediction error criterion (in this case the conventional least squares method) is thus

$$\Phi_u(\omega)|A(e^{i\omega})|^2. \tag{8.2}$$

Now, $1/A$ is typically of low pass character. Hence the weighting function (8.2) will be high pass. The fit at low frequencies is thus de-emphasized and we may get a model that is bad at describing the low frequency properties, such as the steady state gain. This fact is well known from experiences with the least squares method.

Now, in time domain identification methods the choice of weighting function

$$\Phi_u(\omega)/|H(e^{i\omega}, \eta^*)|^2 \tag{8.3}$$

is not explicit. Rather, it is obtained as a consequence of other choices. These are:

1. Choice of input spectrum $\Phi_u(\omega)$. (8.4)

2. Choice of noise model $H(q, \eta)$ (8.5)
   Notice that when the noise model contains adjustable parameters we may not know *a priori* what the resulting function (8.3) will be.

3. Prefiltering of data (8.6)
   If the input and the output are filtered as
   $$y^F(t) = L(q)y(t)$$
   $$u^F(t) = L(q)u(t)$$
   before they are used in the identification method, this has the effect that the weighting function is changed to
   $$|L(e^{i\omega})|^2 \cdot \Phi_u(\omega)/|H(e^{i\omega}, \eta^*)|^2,$$
   see (3.23).

   Note:
   (i) This is equivalent to changing the noise model.

(ii) The noise model may "counteract" by selecting another η^*.

4. Choice of prediction horizon (8.7)
The method (3.21) is based on the one-step ahead prediction (3.19). The same approach could also be applied to k-step ahead predictions. In the frequency domain, the effect is that the weighting function is changed according to (3.24).

5. Choice of sampling interval (8.8)
The choice of the sampling interval T will also affect the resulting weighting function, and the frequency interval over which the fit is measured.

The practical consequences of these ways of affecting the weighting function, and possibly modifying it with the subjective aspect (7.12) in mind are discussed in some companion papers. The choice (8.4) is treated in Yuan and Ljung (1984, 1985) for open-loop experiments and in Gevers and Ljung (1985) for closed-loop ones. The related choices (8.5)–(8.6) and (8.7)–(8.8) are studied in Wahlberg and Ljung (1984).

9. CONCLUSIONS

One of the results of this paper is that we have displayed the basic relationship between the classical methods of frequency and spectral analysis and the time domain methods of prediction error type. They have here been interpreted as two different, but conceptually related approaches to smoothing "the empirical transfer function estimate". The kinship between the methods is also manifested by the quite similar asymptotic expressions for the estimate mean square error, viz. (6.14) and (7.38).

The relationships will of course give a deeper insight into each of the methods. A particular insight that we have stressed here is the role of the noise model as a frequency domain weighting function for the fit between the model and the true system. When the system does not belong to the model set, this function plays an important role for the bias distribution over different frequencies. This fact means that additional care must be exercised when selecting several quantities in the identification procedure, such as (8.4)–(8.8). This study, which has basically been one of theoretical and formal aspects, thus also offers some quite nontrivial aspects for practical applications of identification.

APPENDIX—ON THE DERIVATION OF (3.30)

We shall in this Appendix discuss the approximations involved in the step from (3.21) to (3.30). For simplicity we will assume that (4.6)–(4.7) holds, that

$$H^{-1}(\theta, q) = \Sigma h_k(\theta) q^{-k}; \quad \sum_1^\infty k|h_k(\theta)| < \infty \quad \text{and that } |v(t)| \le C. \tag{A.1}$$

With a little more work (A.1) could be relaxed to conditions on the moments, but it is beyond the point of this paper to become involved in those calculations. The approximation involved in the derivation of (3.30) is that the DFT of a convolution is replaced by the product of the corresponding DFTs as in (3.28). This is exactly correct only if the input to the convolution is periodic. Otherwise a remainder term like $M_N(\omega)$ in (4.13) is obtained. That means that (3.29) should read

$$S_N(\omega, \theta) = Y_N(\omega) - G(e^{i\omega}, \theta) U_N(\omega) + R_N(\omega)$$
$$|R_N(\omega)| \le C_1/\sqrt{N}.$$

Similarly the exact expression for E_N is

$$E_N(\omega) = H^{-1}(e^{i\omega}, \theta) S_N(\omega, \theta) + \tilde{R}_N(\omega)$$
$$|\tilde{R}_N(\omega)| \le C_2/\sqrt{N}.$$

Inserting these expressions into $V_N(\theta)$ gives

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon^2(t, \theta) = \frac{1}{N}\sum_{k=0}^{N-1} |E_N(2\pi k/N)|^2$$
$$= \frac{1}{N}\sum_{k=0}^{N-1} |H(\exp(2\pi i k/N), \theta)|^{-2} |Y_N(2\pi k/N) - G(\exp(2\pi i k/N), \theta) U_N(2\pi k/N)|^2 + \bar{R}_N(\theta)$$

where

$$|\bar{R}_N(\theta)| \le C/\sqrt{N}.$$

This clarifies the approximation involved in (3.30). □

Acknowledgement—This work was supported by the National Swedish Board for Technical Development (STU) under contract No. 84-3837. I am also very grateful for the many detailed comments provided by the reviewers. These were very useful when revising the manuscript.

REFERENCES

Åström, K. J. and P. Eykhoff (1971). System identification—a survey. *Automatica*, **7**, 123.
Åström, K. J. (1970). *Introduction to Stochastic Control Theory*. Academic Press, New York.
Åström, K. J. (1980). Maximum likelihood and prediction error methods. *Automatica*, **16**, 551.
Blackman, R. B. and J. W. Tukey (1958). *The Measurement of Power Spectra*, Dover, New York.
Brillinger, D. R. (1981). *Time Series. Data Analysis and Theory*. Holden Day, San Francisco.
Caines, P. E. (1978). Stationary linear and nonlinear system identification and predictor set completeness. *IEEE Trans. Aut. Control*, **AC-23**, 583.
Eykhoff, P. (1974). *System Identification*. Wiley, London.
Eykhoff, P. (Ed.) (1981). *Trends and Progress in System Identification*. Pergamon Press, Oxford.
Gersch, W. and G. Kitagawa (1985). A smoothness priors method for transfer function estimation. Submitted for publication.
Gevers, M. and L. Ljung (1985). Benefits of feedback in experiment design. 7th IFAC Symposium on Identification, York.
Godfrey, K. R. (1980). Correlation methods. *Automatica*, **16**, 527.
Goodwin, G. C. and R. L. Payne (1977). *Dynamic System Identification: Experiment Design and Data Analysis*. Academic Press, New York.
Jenkins, G. M. and D. G. Watts (1969). *Spectral Analysis and Its Applications*. Holden-Day, San Francisco.
Kabaila, P. V. and G. C. Goodwin (1980). Estimation of the

parameters of an optimal interpolator when the class of interpolators is restricted. *SIAM J. Control*, **18**, 121.

Ljung, L. (1978). Convergence analysis of parametric identification methods. *IEEE Trans. Aut. Control*, **AC-23**, 770.

Ljung, L. (1984). Asymptotic properties of the least squares method for estimating transfer functions and disturbance spectra. Report LiTH-ISY-I-0709, Department of Electrical Engineering, Linköping University, Linköping, Sweden.

Ljung, L. (1985). Asymptotic variance expressions for identified black-box transfer function models. *IEEE Trans. Aut. Control*, **AC-30**, 837.

Ljung, L. *System Identification—Theory for the User*. Prentice-Hall, Englewood Cliffs, New Jersey (in press).

Ljung, L. and P. E. Caines (1979). Asymptotic normality of prediction error estimation for approximate system models. *Stochastics*, **3**, 29.

Ljung, L. and K. Glover (1981). Frequency domain versus time domain methods in system identification. *Automatica*, **17**, 71.

Ljung, L. and Z. D. Yuan (1984). Black-box identification of transfer functions: Asymptotic results for increasing model order and data records. *IMA J. Math. Control Inform.* **1**, 323.

Ljung, L. and Z. D. Yuan (1985). Asymptotic properties of black-box identification of transfer functions. *IEEE Trans. Aut. Control*, **AC-30**, 514.

Rake, H. (1980). Step response and frequency response methods. *Automatica*, **16**, 519.

Söderström, T. and P. G. Stoica (1983). *Instrumental Variable Methods for System Identification*. Lecture Notes in Control and Information Sciences No. 57. Springer Verlag, Berlin.

Solo, V. (1978). Time series recursions and stochastic approximation. Ph.D. dissertation, The Australian National University, Canberra, Australia.

Wahlberg, B. and L. Ljung (1984). Design variables for bias distribution in transfer function estimation. 23rd IEEE Conference on Decision and Control, Las Vegas, December. Also *IEEE Trans. Aut. Control*, **AC-31** (to appear).

Yuan, Z. D. and L. Ljung (1984). Black-box identification of multivariable transfer functions—asymptotic properties and optimal input design. *Int. J. Control*, **40**, 223.

Yuan, Z. D. and L. Ljung (1985). Unprejudiced optimal open-loop input design for identification of transfer functions. *Automatica*, **21**, (this issue).