THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
Series A 8

MARKKU RAHIALA

ON THE IDENTIFICATION AND ESTIMATION OF MULTIPLE INPUT TRANSFER FUNCTION MODELS WITH AUTOCORRELATED ERRORS

Helsinki 1985

ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
The Research Institute of the Finnish Economy
Lönnrotinkatu 4 B 00120 Helsinki 12 Finland
Sarja A 8 Series

Markku Rahiala

ON THE IDENTIFICATION AND ESTIMATION OF MULTIPLE INPUT TRANSFER FUNCTION MODELS WITH AUTOCORRELATED ERRORS

Helsinki 1985
ISBN 951-9206-00-0
ISSN 0356-7435

PREFACE

I wish to express my gratitude to all those who have helped me in the course of this study.

The continuous encouragement by Dr. Timo Teräsvirta and Professor Hannu Niemi and the comments they made on my work were especially valuable. They also read the manuscript and suggested many changes. I am deeply indebted to both of them.

Professor Seppo Mustonen´s impact on my research has also been indespensable. The ideas presented in this dissertation would never have originated without his magnificent statistical data processing systems. It has been a privilege to work as his subordinate.

I am grateful to Dr. Pentti Vartia for his encouragement and for the suggestions he made concerning the empirical part of this study. My gratitude is also due to the Research Institute of the Finnish Economy for letting me spend one month working on this project during my year at the Institute.

Mr. Pentti Saikkonen I wish to thank for the bibliographical references and other valuable help he gave me.

Mrs. Anja Aaltonen revised the English language of this dissertation. Mrs. Riitta Ulmanen typed the text patiently and skilfully, and Mrs. Arja Virtanen drew the figures. I want thank them all.

Finally, I wish to thank the Research Institute of the Finnish Economy for accepting this monograph to its publication series.

Helsinki, March 1985,

CONTENTS

1. INTRODUCTION

The most common form of linear regression models contains the assumption of no correlation between the disturbances in different observations. In connection of time series data, however, one often realizes that the assumption of mutually uncorrelated disturbances is clearly inconsistent with the data. Especially econometricians quite commonly take this to indicate a misspecification of the dynamic effects of the explanatory variables. They often seem to think that if only the effects of the explanatory variables were specified correctly and no relevant explanatory variables were missing, the errors of the model would always behave like white noise. However, in most cases the regressand contains variation that is totally independent of the explanatory variables, but still has a dynamic structure of its own. This is why it seems to be more realistic slightly to reduce the goals of the model building process and to allow some amount of regularity in the behavior of the disturbances. At the same time one should take these regularities into account in the estimation of parameters and especially in hypothesis testing. This is not merely a question of technical perfectionism, but a vital precaution to ensure valid conclusions. The problem of the so-called spurious correlations (cf. Granger and Newbold, 1974) is well known. Time series with internal dynamic mechanisms similar enough often look alike and have high cross correlations, even if the series are actually totally independent of each other.

Thus, there are two parts to be specified in time series models, the structural part describing the way that the regressor variables affect the regressand, and the disturbance part describing the regularity pattern of the errors. If there are no feedback relations between the regressand and the regressors, it will be sufficient to build what is called a single equation transfer function model to desrcibe the relationships between the regressors and the regressand. In this paper we will only consider the case of linear transfer functions. More often than not, the structures of the transfer functions are a priori unknown. Therefore it becomes difficult to identify the form of the disturbance part of the model, because the behavior of the residuals essentially depends on the specification of the structural part. On the other hand, it is difficult to identify the transfer function forms straight away, because the regularity pattern of the disturbance term greatly affects the distributions of most test statistics commonly used in model specification. So far, identification methods have usually started from the specification of the transfer functions. After estimating the parameters of some tentatively identified transfer functions, it is possible to compute a set of preliminary residual estimates. At the next stage, one can correct the test statistics used in the specification of the structural model in accordance with the autocorrelation pattern of these residual estimates. Thus, several iteration rounds are always needed in model identification, and the first tentative choice of the transfer function forms is crucial for the whole procedure.

Provided all the variables are jointly stationary, spectral analysis makes it possible directly to estimate the spectral density and the

autocorrelation pattern of that part of the regressand which is totally uncorrelated with the regressor series. Moreover, it is possible to identify and estimate ARMA-models on the basis of spectral estimates (cf. Rice, 1979, or Nakano, 1982) without using any of the observations explicitly. Thus, it is possible to start the model building process with the identification of the disturbance part. One can first estimate the cross spectral densities between the regressors and the regressand, then the spectral density and the autocorrelations of the disturbance, and finally the parameters of some ARMA-model for the disturbance. The form of the ARMA-model can be identified in the usual manner on the basis of the estimated autocorrelations. If several models are estimated, their goodness of fit can be evaluated visually by comparing the different ARMA-spectra to the previously estimated spectrum of the disturbance. When the disturbance model has been identified and its parameters have been estimated, the observed variables can be filtered or weighted to make the error process of the transformed model approximately white noise. Then, finally, specification tests can be used in the usual manner in the identification of the transfer function forms.

In the second chapter we will define the formal concepts needed hereafter and give a short review of the earlier identification methods of transfer function models. In Chapter 3, we will propose a new identification and estimation method for univariate ARMA-models. The method is based on the maximization of the entropy of the ratio between the estimated and theoretical spectral densities. A new identification method for linear transfer function models is introduced in Chapter 4. The proposed method starts from the disturbance part of the model and follows

the ideas sketched above. The method also yields initial estimates for the parameters of the identified model. The asymptotic behavior of these estimates is studied and an example of the use of the method is given. In Chapter 5, some computational aspects of the proposed method and the method of maximum likelihood are considered. We will describe a computer program for the application of these methods and also discuss the differences between the weighting and filtration of data.

2. PRELIMINARIES

Let us denote the regressand (output variable) by $y(t)$ and the explanatory variables (input variables) by $X(t) = [x_1(t) \ldots x_m(t)]'$.

Suppose it clear a priori that there is no feedback from $y(t)$ to $X(t)$ and that $X(t)$ affects $y(t)$ linearly:

$$y(t) = \mu + v_1(L)x_1(t) + \ldots + v_m(L)x_m(t) + \varepsilon(t), \tag{2.1}$$

where $v_i(L) = \sum_{j=0}^{\infty} v_{ij}L^j$, L is the lag operator, and $\sum_{j=0}^{\infty} |v_{ij}| < \infty$ for all $i = 1,\ldots,m$.

The residual process $\varepsilon(t)$ is supposed to be stationary with finite variance and zero expectation, and totally independent of the $X(t)$-process. We will further assume that $\varepsilon(t)$ behaves like an ARMA(p,q) -process with

$$\phi(L)\varepsilon(t) = \theta(L)\alpha(t),$$

where

$$\phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p ,$$

$$\theta(L) = 1 - \theta_1 L - \ldots - \theta_q L^q ,$$

and $\alpha(t)$ are mutually independent, identically distributed variables for $t = \ldots,-1,0,1,\ldots$.
Furthermore, $\alpha(t)$ are supposed to be normally distributed $N(0,\sigma^2)$.

Some restrictions must be imposed on the $v_i(L)$ -functions to make the number of parameters finite. The most common restrictions are

$$v_i(L) = \frac{\omega_i(L)}{\delta_i(L)} , \tag{2.2}$$

where

$$\omega_i(L) = \omega_{io} - \omega_{i1}L - \ldots - \omega_{ir_i}L^{r_i}$$

$$\delta_i(L) = 1 - \delta_{i1}L - \ldots - \delta_{is_i}L^{s_i} \qquad i = 1,\ldots,m ,$$

or

$$v_i(L) = \sum_{j=0}^{r_i} v_{ij}L^j , \tag{2.3}$$

where

$$v_{ij} = \sum_{k=0}^{d_i} \beta_{ik}j^k \qquad j = 0,\ldots,r_i, \quad i = 1,\ldots,m .$$

Model (2.1) with restriction (2.3) is usually referred to as the Almon model (cf. Almon, 1965).
Both forms (2.2) and (2.3) can of course be used to approximate almost any kinds of $v_i(L)$ -functions.

For a while, let us consider model (2.1) with restriction (2.2). Suppose that we have observed $y(1),\ldots,y(n)$ and $X(1),\ldots,X(n)$. The model can be reformulated into

$$\delta^*(L)y(t) = \mu\delta^*(1) + \sum_{i=1}^{m} \omega_i^*(L)x_i(t) + \frac{\delta^*(L)\theta(L)}{\phi(L)}\alpha(t) ,$$

$$t = t_o+1,\ldots,n ,$$

where

$$\delta^*(L) = \delta_1(L) \ldots \delta_m(L) = 1 - \delta_1^*L - \ldots - \delta_{s^*}^*L^{s^*} ,$$

$$\omega_i^*(L) = \omega_i(L) \prod_{j\neq i} \delta_j(L) = \omega_{io}^* - \omega_{i1}^*L - \ldots - \omega_{i,r_i^*}^*L^{r_i^*} ,$$

$$i = 1,\ldots,m$$

and $t_o = \max(s^*,r_1^*,\ldots,r_m^*)$.

It is quite clear that no useful information can be extracted from the first observations at time points $t = 1,\ldots,t_o$. The likelihood function based on the rest of the observations is of the form

(2.4)
$$L_{y(t_o+1),X(t_o+1),\ldots,y(n),X(n)}(\delta^*,\mu,\omega_1^*,\ldots,\omega_m^*,\phi,\theta,\sigma^2)$$
$$= (\det\Sigma)^{-\frac{n-t_o}{2}} \exp(-\tfrac{1}{2} Z'\Sigma^{-1}Z) ,$$

where

$$Z = [z(t_o+1) \ \ldots \ z(n)]' \ ,$$

$$z(t) = \delta^*(L)y(t) - \mu\delta^*(1) - \sum_{i=1}^{m} \omega_i^*(L)x_i(t) \ , \quad t = t_o+1,\ldots,n \ ,$$

and

$$\Sigma = \mathrm{cov}(Z).$$

It is not possible to optimize (2.4) analytically with respect to the parameters, because Σ depends on ϕ, θ and δ^* in quite a complicated manner (see Chapter 5.2). That is why one has to settle for numerical optimization of (2.4). In Chapter 5.2 we will propose an algorithm for the computation of $Z'\Sigma^{-1}Z$ and $\det\Sigma$. The proposed method resembles closely Dent's algorithm (Dent, 1977). Alternative algorithms have been presented by Ansley (1979), Harvey and Phillips (1979), Ljung and Box (1979), and others.

The amount of calculations needed for the optimization of (2.4) depends primarily on the number of parameters and the number of observations, but also on the goodness of fit and on the specification of the model. If the model is overspecified, the likelihood surface becomes flat in some directions and the optimization algorithms may easily run into trouble. That is why one should start with simple and parsimonious models and enlarge the model only gradually. This fact stresses the importance of the identification stage of model building. The identification problem of model (2.1) with restriction (2.2) has been dealt with fairly extensively in the literature. Most of the suggested methods are appli-

cable only if $y(t)$ and $X(t)$ are jointly stationary. The most important suggestions are the following:

- Box and Jenkins (1970) proposed the following method for the single input case: The input and output series are filtered with the filter that prewhitens the input. By studying the cross correlations between these filtered series one might get an idea of a suitable transfer function form.

  This method cannot be generalized to the case of multiple input series. Further, it only "orthogonalizes the regressors", but does not remove the problems caused by the autocorrelated errors.

- Priestley (1971), Haugh and Box (1977) and Granger and Newbold (1977) have recommended the prewhitening of all the variables separately. By studying the cross correlations between these prewhitened series, one may identify a transfer function between the input and the output innovations. This transfer function can then be translated into a model between the original observed series.

  This method will be worthwhile if there is any doubt about the presence of possible feedback effects, but it often leads to complicated and overstructured models, especially when several input variables are involved.

- The method used in the WMTS program package (cf. Tiao et al., 1979 or Tiao and Box, 1981) can be used to identify transfer function models as special cases of multivariate ARMA-models whenever the input series allow ARMA-representations. The method is based on fitting consecutive AR-models of increasing orders to the data.

- The joint spectral density of the input and output series can also be used as a basis for model identification. One possibility

is the use of Hannan´s efficient method (Hannan, 1963), where the structural parameters are estimated with the weighted least squares method in the frequency domain, using the inverses of square roots of the estimated residual spectral density as weights.

Another possibility (cf. Pukkila, 1978 and 1982) is to calculate the Fourier transform of the spectral transfer function estimate and to assess the significance of different coefficients by comparing them with their theoretical asymptotic standard deviations.

The ordinary least squares technique (OLS) will often do quite well (cf. Harvey, 1981, Ch. 7.5 and Liu and Hanssens, 1982) even if it will be difficult to assess the significance of different terms in the model. We will discuss this problem further in Chapter 4.1.

This method can be applied to nonstationary series as well.

The new method presented in Chapter 4 resembles slightly Hannan´s method, but the residual process will be restricted to a parametric family of models and the weighted least squares fits will be carried out in the time domain.

For some reason, the problem of autocorrelated residuals seems to be mostly ignored in connection with Almon models. The new method can be combined with any of the usual selection methods of Almon models (cf. e.g. Teräsvirta and Mellin, 1983).

3. INITIAL ESTIMATION OF PARAMETERS OF UNIVARIATE ARMA MODELS

3.1 The Maximum Entropy (ME) Procedure

The words "maximum entropy estimation" have previously been used in at least two contexts.

Firstly, Akaike (1977) introduced the entropy maximization principle as a general framework for statistical reasoning and model building. According to the principle, one should strive at maximizing the expected entropy of the estimated distribution of a future observation within a chosen family of models. The method of maximum likelihood can be formulated as a special case of the entropy maximization concept.

Secondly, spectral densities have been estimated by maximizing the entropy of the estimator subject to the condition that m first autocorrelations will coincide with the corresponding sample autocorrelations. This method, however, is identical to the autoregressive estimation method with m as the degree of autoregression. As one of the most recent references, see Parzen (1982). For later reference, we repeat some of the tools and concepts used in the article. The Kullback-Leibler directed information divergence is defined by

$$I(f,g) = \int_{-\pi}^{\pi} -f(\lambda) \log \frac{g(\lambda)}{f(\lambda)} \, d\lambda \ ,$$

where f and g are two spectral density functions. Further, the so-called cross entropy function is defined by

$$H(f,g) = \int_{-\pi}^{\pi} -f(\lambda) \log g(\lambda)\, d\lambda = I(f,fg)$$

satisfying the inequality

$$H(f,f) \leq H(f,g) .$$

Let $G = \{f_\theta \mid \theta \in \Theta\}$ be a parametric family of spectral densities. Whenever $f \in G$, the function $H(f,f_\theta)$ will be minimized with respect to θ at the "true" parameter value corresponding to f. The $H(f,f_\theta)$-function can be estimated from data by $\tilde{H}(f,f_\theta) = H(\hat{f},f_\theta)$, where $\hat{f}$ is some "raw" estimator of the true spectral density.

Let us now return to the question of estimating the parameters of stationary ARMA-models. With the previous notation, let

$$\phi(L)\varepsilon(t) = \theta(L)\alpha(t) , \qquad t = \ldots,-1,0,1,\ldots , \tag{3.1}$$

where $\varepsilon(t)$ is supposed to be stationary and $\alpha(t) \sim NID(0,\sigma^2)$. It will be implicitly understood that $\phi(L)$ and $\theta(L)$ may contain seasonal factors, although they do not appear in the notation explicitly. The first task in the model selection is the choice of the degrees p and q. Several procedures have been suggested, based on the autocorrelation and partial autocorrelation functions (cf. Box and Jenkins, 1970), the so-called S-arrays (cf. Woodward and Gray, 1981), the inverse

autocorrelation function (cf. Cleveland, 1972) and the so-called extended autocorrelation function (cf. Tsay and Tiao, 1982). Beguin et al. (1980) have suggested what is called the corner method. Hokstad (1983) has proposed a method based on the inspection of cross correlations between the observed series and the residuals of different tentative models.

Several general model selection criteria are also available, such as Akaike´s AIC and FPE criteria (cf. Priestley, 1981, pp. 372-373) and Schwarz´s SBIC criterion (cf. Judge et. al., 1984, Ch. 21). The calculation of these criteria requires, however, the estimation of the parameters of all the model candidates. Once all the estimations have been carried out, it is possible to base the model selection on visual goodness of fit evaluation and on more profound diagnostics.

As noted in Chapter 2, the use of the ML estimation principle is often computationally cumbersome. This is why there is a need for computationally easier estimation routines even if they cannot compete with the ML -procedure in small samples. These estimates can also be used as starting values for the iteration in the ML -procedure. Box and Jenkins (1970) suggested a modified method of moments, but the performance of the method is often poor. If there is a multiplicative MA -part in the model, the risk of unstable estimates is obvious. By our experience, the straightforward use of the moment principle does not perform too well either, and in case of mixed models the computations become rather heavy. This leads one to consider taking more than $p+q$ moments (autocorrelations) into account. No ARMA-model will of course have autocorrelations that would exactly match the estimated ones, but one could measure the

"distance" between the theoretical and estimated autocorrelations by some measure and minimize it. It seems natural to weight the first autocorrelations more than the distant ones. This reasoning brings to mind the possibility of minimizing the distance between some window estimator of the spectral density and the theoretical spectra provided by the model. The distance could be measured by any standard measure of distance between distributions. This kind of criteria will be relatively easy to optimize, because the theoretical spectral densities depend on the parameters in a very simple way:

$$f_\varepsilon(\lambda) = \left| \frac{\theta(e^{i\lambda})}{\phi(e^{i\lambda})} \right|^2 f_\alpha(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{\theta(e^{i\lambda})}{\phi(e^{i\lambda})} \right|^2 .$$

Possible choices for the distance measures are for instance the Hellinger distance (cf. Beran, 1977), any of the criteria suggested by Rice (1979), or the estimated cross entropy $\tilde{H}$ of Parzen (1982). Only the first one is strictly speaking a measure of distance, but the latter ones could be used in the same spirit.

Another way of looking at the model is to put it into the form

$$\frac{1}{\sigma^2} f_\alpha(\lambda) = \frac{1}{\sigma^2} \frac{f_\varepsilon(\lambda)}{h(\lambda,\phi,\theta)} \equiv \frac{1}{2\pi}$$

for every $\lambda \in [-\pi,\pi]$, where

$$h(\lambda,\phi,\theta) = \left| \frac{\theta(e^{i\lambda})}{\phi(e^{i\lambda})} \right|^2 .$$

Let us denote a window estimator of $f_\varepsilon(\lambda)$ by

$$\hat{f}_\varepsilon(\lambda) = \frac{1}{2\pi} \sum_{j=-M_n}^{M_n} r(j) w_j \cos j\lambda \ ,$$

where $r(j)$ is the j^{th} sample autocorrelation of the $\varepsilon(t)$ -series $(t = 1,\ldots,n)$ and M_n is the truncation point or width of the lag window. Weights

$$w_j = k\left(\frac{j}{M_n}\right) \qquad j = -M_n,\ldots,0,\ldots,M_n$$

are said to constitute the lag window. The function $k:[-1,1] \to \mathbb{R}^+$ is supposed to be even and twice continuously differentiable with $k'(0) = 0$. One way of evaluating the flatness of the normalized estimated residual spectrum

$$\tilde{f}_\alpha(\lambda) = \frac{1}{\sigma^2} \frac{\hat{f}_\varepsilon(\lambda)}{h(\lambda,\phi,\theta)}$$

is to calculate the entropy

$$(3.2) \qquad -\int_{-\pi}^{\pi} \tilde{f}_\alpha(\lambda) \log \tilde{f}_\alpha(\lambda)\, d\lambda$$

subject to the condition

$$\int_{-\pi}^{\pi} \tilde{f}_\alpha(\lambda)\, d\lambda = 1 \ .$$

Even if (3.2) is not strictly speaking a measure of distance, it can of course be interpreted as the complement of the Kullback-Leibler measure

$I(\tilde{f}_\alpha, f)$ of directed divergence between $\tilde{f}_\alpha$ and the flat spectrum

$$f(\lambda) \equiv \frac{1}{2\pi} \qquad \text{for all} \quad \lambda \in [-\pi, \pi] .$$

Note that the maximization of (3.2) with respect to ϕ, θ and σ^2 is equivalent to the minimization of

$$(3.2') \qquad e(\phi,\theta) = \frac{\int_{-\pi}^{\pi} \tilde{f}_\alpha(\lambda) \log \tilde{f}_\alpha(\lambda)\, d\lambda}{\int_{-\pi}^{\pi} \tilde{f}_\alpha(\lambda)\, d\lambda} - \log \int_{-\pi}^{\pi} \tilde{f}_\alpha(\lambda)\, d\lambda$$

$$\geq -\log 2\pi$$

with respect to ϕ and θ. The restriction in (3.2) can be written in the form

$$\sigma^2 = \int_{-\pi}^{\pi} \frac{\hat{f}_\varepsilon(\lambda)}{h(\lambda,\phi,\theta)}\, d\lambda .$$

Now, we suggest the estimation of the parameters ϕ and θ through the maximization of (3.2) or, equivalently, through the minimization of (3.2'). The optimization of (3.2) will mean the maximization of the white noise properties of the residual process $\alpha(t)$. This procedure comes quite close to the method suggested by Nakano (1982) and earlier by Rice (1979), though the motivations of their procedures were totally different. Both of these methods are asymptotically efficient when applied to normal processes. Furthermore, it is easy to visualize the performance of the fitted model by drawing a graph of $\hat{f}_\varepsilon(\lambda)$. In principle, optimizing

(3.2) resembles slightly the so-called MDI estimation method (cf. Kullback, 1959) and is coherent with Akaike's general entropy maximization principle. However, as far as we can see, the theory of neither of these concepts can be utilized here straight away because of the structure of $\hat{f}_\varepsilon(\lambda)$.

Let

$$\psi = [\phi_1 \cdots \phi_p \; \theta_1 \cdots \theta_q]' = [\phi' \quad \theta']' ,$$

$$\xi = [\psi' \quad \sigma^2]'$$

and

$$g_n(\lambda,\xi) = \frac{2\pi}{\sigma^2} \frac{\hat{f}_\varepsilon(\lambda)}{h(\lambda,\psi)} .$$

As an estimator of (3.2') we will use

$$\tilde{e}_n(\psi) = \frac{Q_n(\xi)}{S_n(\xi)} - \log S_n(\xi),$$

where

$$Q_n(\xi) = \frac{\pi}{M_n} \sum_{j=-M_n+1}^{M_n} g_n(\lambda_j,\xi)\log(g_n(\lambda_j,\xi)),$$

$$S_n(\xi) = \frac{\pi}{M_n} \sum_{j=-M_n+1}^{M_n} g_n(\lambda_j,\xi) = \frac{1}{\sigma^2} \tilde{S}_n(\psi)$$

and

$$\lambda_j = j\frac{\pi}{M_n} \qquad j = -M_n,\ldots,0,\ldots,M_n \ .$$

This means that we will use the trapezoidal rule in the integration of (3.2) with as dense a partition as possible without making the estimates $\hat{f}_\varepsilon(\lambda_j)$ functionally dependent of each other.

Because (3.2) involves $\log\hat{f}_\varepsilon(\lambda)$ and negative estimates $\hat{f}_\varepsilon(\lambda)$ would cause additional problems, we assume that the form of the lag window is defined either by

$$k(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{when} \quad x \le \frac{1}{2} \\ 2(1 - |x|)^3 & \text{when} \quad \frac{1}{2} < |x| \le 1 \end{cases}$$

(the Parzen window) or by

$$k(x) = 0.54 + 0.46\cos\pi x, \qquad -1 \le x \le 1$$

(the Tukey-Hamming window).

The values of ϕ and θ that minimize $\tilde{e}_n(\psi)$ will be called the maximum entropy (ME) estimates and denoted by $\tilde{\psi} = [\tilde{\phi}' \quad \tilde{\theta}']'$. Parameter σ^2 will be estimated by

$$\tilde{\sigma}^2 = \frac{1}{2\pi}\tilde{S}_n(\tilde{\psi}) \qquad \text{with} \qquad \tilde{S}_n(\psi) = \sigma^2 S_n(\xi) \ .$$

The value of ξ that minimizes $Q_n(\xi)$ without constraints will be denoted by $\tilde{\xi} = [\tilde{\psi}' \quad \tilde{\sigma}^2]'$. Note that $\tilde{\psi}$ coincides with the previous notation, but $\tilde{\sigma}^2$ and $\bar{\sigma}^2$ will differ from each other.

3.2 Asymptotic Properties of ME-Estimators

In studying the asymptotic properties of the ME-estimators we will make use of a theorem proved by P. Saikkonen. Because his research report (Saikkonen, 1981) has not been published, we repeat the result here, albeit in a less general form than the original:

Theorem (Saikkonen): Suppose $z(t)$ is a normal linear process

$$z(t) = \sum_{j=0}^{\infty} a_j \alpha(t-j)$$

where

$$\sum_{j=0}^{\infty} |a_j| < \infty \qquad \text{and} \qquad \alpha(t) \sim \text{NID}(0,\sigma^2) \qquad t = \ldots,-1,0,1,\ldots$$

Let $\gamma(k) = Ez(t)z(t-k)$ be its covariance function and let $\hat{f}_z$ be either the Parzen or the Tukey-Hamming estimator of the spectral density of $z(t)$. Suppose

$$\sum_{k=-\infty}^{\infty} |k|^{\alpha}|\gamma(k)| < \infty \qquad \text{for some} \quad \alpha \geq \frac{1}{2}$$

and suppose $H:[-\pi,\pi] \to \mathbb{R}^m$ is symmetric and admits the representation

$$H(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} L(k) e^{ik\lambda} ,$$

where

$$\sum_{k=-\infty}^{\infty} |k|^{\delta} |L_j(k)| < \infty$$

for some $\delta \geq \frac{1}{2}$ and all $j = 1,\ldots,m$.

Let

$$\lim_{n\to\infty} \frac{M_n}{\sqrt{n}} = 0 \qquad \text{and} \qquad \lim_{n\to\infty} \frac{\sqrt[4]{n}}{M_n} = 0 .$$

Then

$$\sqrt{n} \left[\frac{\pi}{M_n} \sum_{j=-M_n+1}^{M_n} H(\lambda_j) \hat{f}_z(\lambda_j) - \int_{-\pi}^{\pi} H(\lambda) f_z(\lambda)\, d\lambda \right]$$

$$\lambda_j = j \frac{\pi}{M_n} , \qquad j = -M_n,\ldots,M_n$$

will be asymptotically normal $N_m(0,\Sigma)$ with a covariance matrix

$$\Sigma = 4\pi \int_{-\pi}^{\pi} H(\lambda) H(\lambda)' f_z(\lambda)^2 \, d\lambda .$$

□

Suppose now that we have observations $\varepsilon(1),\ldots,\varepsilon(n)$ from a normal process (3.1). With increasing n, let M_n grow faster than $\sqrt[4]{n}$ but slower than $\sqrt{n}$.

Now, for a fixed ξ

$$\operatorname*{plim}_{n\to\infty} Q_n(\xi) = \int_{-\pi}^{\pi} \frac{\sigma_o^2 h(\lambda,\psi_o)}{\sigma^2 h(\lambda,\psi)} \cdot \log \frac{\sigma_o^2 h(\lambda,\psi_o)}{\sigma^2 h(\lambda,\psi)} \, d\lambda$$

$$= Q(\xi) \ ,$$

where σ_o^2 and ψ_o denote the "true" values of σ^2 and ψ. This is obvious, because the variance of $g_n(\lambda,\xi)$ tends to zero uniformly in λ when ξ is restricted to any closed neighborhood of ξ_o inside the stability and invertibility region.

The limiting function $Q(\xi)$ is clearly continuous. Because no two stationary and invertible ARMA-processes can have exactly the same spectrum, $Q(\xi)$ will have a unique minimum at ξ_o. Now, following the lines of Rice (1979) it is easy to see that the following result holds:

Lemma 3.2.1 $\tilde{\psi}$ and $\tilde{\sigma}^2$ are consistent estimators of ψ and σ^2.

As an immediate consequence of the lemma, we can see that $\bar{\sigma}^2$ is also consistent for σ^2.

Next, we shall discuss the asymptotic distribution of $\tilde{\psi}$.

Theorem 3.2.2 $\sqrt{n}(\tilde{\psi} - \psi_o)$ is asymptotically normal $N_{p+q}(0, I(\psi_o)^{-1})$, where

$$I(\psi) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \text{Dlogh}(\lambda,\psi)' \text{Dlogh}(\lambda,\psi) \, d\lambda$$

and $\mathrm{Dlogh}(\lambda,\psi)$ is the vector of derivatives of $\mathrm{logh}(\lambda,\psi)$ with respect to ψ.

Proof Let $\tilde{g}_n(\lambda,\psi) = g_n(\lambda,\psi,\tilde{\sigma}^2)$ and $\tilde{Q}_n(\psi) = Q_n(\psi,\tilde{\sigma}^2)$ for brevity. All derivatives in this proof will be taken with respect to ψ. The first and second derivatives of $\tilde{Q}_n(\psi)$ will be

$$D\tilde{Q}_n(\psi) = \frac{\pi}{M_n} \sum_{j=-M_n+1}^{M_n} [\log(\tilde{g}_n(\lambda_j,\psi)) + 1] D\tilde{g}_n(\lambda_j,\psi)$$

and

$$D^2\tilde{Q}_n(\psi) = \frac{\pi}{M_n} \sum_{j=-M_n+1}^{M_n} [\log(\tilde{g}_n(\lambda_j,\psi)) + 1] D^2\tilde{g}_n(\lambda_j,\psi)$$

$$+ \frac{\pi}{M_n} \sum_{j=-M_n+1}^{M_n} \frac{1}{\tilde{g}_n(\lambda_j,\psi)} D\tilde{g}_n(\lambda_j,\psi)' D\tilde{g}_n(\lambda_j,\psi) .$$

Now, obviously

$$\operatorname*{plim}_{n\to\infty} \tilde{g}_n(\lambda_j,\psi_o) = 1$$

uniformly in λ_j $(j = -M_n,\ldots,M_n)$, because $f_\varepsilon(\lambda)$ and $h(\lambda,\xi)$ are bounded away from 0 and ∞.

Consequently,

$$\operatorname*{plim}_{n\to\infty} \frac{\pi}{M_n} \sum_{j=-M_n+1}^{M_n} [\log(\tilde{g}_n(\lambda_j,\psi_o)) + 1] D^2\tilde{g}_n(\lambda_j,\psi_o)$$

$$= D^2 \operatorname*{plim}_{n\to\infty} \frac{\pi}{M_n} \sum_{j=-M_n+1}^{M_n} \tilde{g}_n(\lambda_j,\psi_o)$$

$$= D^2 \int_{-\pi}^{\pi} 1 \, d\lambda = 0 .$$

On the other hand,

$$\operatorname*{plim}_{n\to\infty} \frac{\pi}{M_n} \sum_{j=-M_n+1}^{M_n} \frac{1}{\tilde{g}_n(\lambda_j,\psi_o)} D\tilde{g}_n(\lambda_j,\psi_o)' D\tilde{g}_n(\lambda_j,\psi_o)$$

$$= \operatorname*{plim}_{n\to\infty} \frac{\pi}{M_n} \sum_{j=-M_n+1}^{M_n} \tilde{g}_n(\lambda_j,\psi_o) D\log(\tilde{g}_n(\lambda_j,\psi_o))' D\log(\tilde{g}_n(\lambda_j,\psi_o))$$

$$= \int_{-\pi}^{\pi} D\log(\frac{\sigma_o^2}{2\pi} h(\lambda,\psi_o))' D\log(\frac{\sigma_o^2}{2\pi} h(\lambda,\psi_o)) \, d\lambda$$

$$= 4\pi I(\psi_o) .$$

$I(\psi_o)$ is the asymptotic average amount of Fisher information per observation (cf. Parzen, 1971).

Now, according to Saikkonen's theorem,

$$Z_n = \sqrt{n} \frac{\pi}{M_n} \sum_{j=-M_n+1}^{M_n} (D \frac{2\pi}{\sigma_o^2 h(\lambda_j,\psi_o)})' \hat{f}_\varepsilon(\lambda_j)$$

is asymptotically normal with expectation

$$\sqrt{n} \int_{-\pi}^{\pi} \sigma_o^2 h(\lambda,\psi_o)(D \frac{1}{\sigma_o^2 h(\lambda,\psi_o)})' \, d\lambda$$

$$= - \sqrt{n} \int_{-\pi}^{\pi} D\log(h(\lambda,\psi_o))' \, d\lambda$$

$$= - \sqrt{n} \; (D \int_{-\pi}^{\pi} \log(h(\lambda,\psi_o)) \, d\lambda)'$$

$$= 0$$

and the covariance matrix

$$4\pi \int_{-\pi}^{\pi} (D \frac{1}{h(\lambda,\psi_o)})'(D \frac{1}{h(\lambda,\psi_o)}) h(\lambda,\psi_o)^2 \, d\lambda$$

$$= 16\pi^2 I(\psi_o) \; .$$

The conditions in Saikkonen's theorem will be fulfilled, because the fourth order cumulants vanish altogether and because the Fourier transforms of the components of

$$D \frac{1}{h(\lambda,\psi_o)}$$

will be linear combinations of autocorrelations of some ARMA-processes. This becomes obvious if we write

$$D_{\theta_j} h(\lambda,\psi)^{-1} = -\left[- \theta_j e^{ij\lambda} \theta(e^{-i\lambda}) - \theta_j e^{-ij\lambda} \theta(e^{i\lambda}) \right] \frac{|\phi(e^{i\lambda})|^2}{|\theta(e^{i\lambda}) \theta(e^{i\lambda})|^2}$$

$$j = 1,\ldots,q \; .$$

and

$$D_{\phi_j} h(\lambda,\psi)^{-1} = \left[- \phi_j e^{ij\lambda}\phi(e^{-i\lambda}) - \phi_j e^{-ij\lambda}\phi(e^{i\lambda})\right] \frac{1}{|\theta(e^{i\lambda})|^2}$$

$$j = 1,\ldots,p \ .$$

Because $\operatorname{plim}_{n\to\infty} \log(\tilde{g}_n(\lambda,\psi_o)) = 0$ uniformly in λ and $\operatorname{plim}_{n\to\infty} \tilde{\sigma}^2 = \sigma_o^2$, it follows that

$$\operatorname{plim}_{n\to\infty} (Z_n - \sqrt{n}\, D\tilde{Q}_n(\psi_o)') = 0 \ .$$

Thus, Z_n and $\sqrt{n}\, D\tilde{Q}_n(\psi_o)'$ have the same asymptotic distribution.

Consider now the expansion

$$D\tilde{Q}_n(\tilde{\psi})' = D\tilde{Q}_n(\psi_o)' + D^2\tilde{Q}_n(\psi^o)(\tilde{\psi} - \psi_o) = 0 \ ,$$

the components of ψ^o lying between the components of ψ_o and $\tilde{\psi}$. Now

$$\sqrt{n}\,(\tilde{\psi} - \psi_o) = - \sqrt{n}\,\left[D^2\tilde{Q}_n(\psi^o)\right]^{-1} D\tilde{Q}_n(\psi_o)'$$

will be asymptotically normal

$$N_{p+q}(0,(4\pi I(\psi_o))^{-1} 16\,\pi^2 I(\psi_o)(4\pi I(\psi_o))^{-1})$$

$$= N_{p+q}(0,I(\psi_o)^{-1}) \ ,$$

because $\operatorname{plim}_{n\to\infty} \psi^o = \psi_o$.

□

Note that the above reasoning can be extended to other forms of models as well, provided the sequences corresponding to $\int_{-\pi}^{\pi} D \frac{1}{h(\lambda,\psi_o)} e^{ik\lambda} d\lambda$ $k = \ldots,-1,0,1,\ldots$ satisfy the conditions of Saikkonen's theorem and if the correspondence between ψ and $h(\lambda,\psi)$ is one to one and regular enough to preserve consistency. For instance, the unknown component ARIMA (UCARIMA)-models (cf. Nerlove et al., 1979) will usually satisfy these requirements. Thus, the ME -technique can also be used for estimating the parameters of UCARIMA-models.

3.3 Large Sample Tests Based on the ME-Estimators

Theorem 3.2.2 also implies that tests, analogous to the Wald test, the LR-test and Rao's test (LM-test), can be used to test hypotheses concerning $\psi \in \Psi \subset \mathbb{R}^{p+q}$.

At first, let us suppose that the hypothesis to be tested is of the form

$$\psi = \psi_\gamma = G(\gamma), \qquad \gamma \in \Gamma, \quad \dim(\Gamma) = g ,$$

where G is supposed to have a continuous inverse G^{-1} and to be continuously differentiable at all points of the open set Γ.
Suppose $h(\lambda,G(\gamma))$ will meet the requirements in Chapter 3.2. Let γ_o denote the "true" value of γ and $\psi_o = G(\gamma_o)$. Denote the restricted ME -estimator of ψ by $\tilde{\psi}_R = G(\tilde{\gamma})$. Let

$$Z_n = \frac{\sqrt{n}}{\sqrt{4\pi}} \left[D^2_{\psi}\tilde{Q}_n(\psi_o)\right]^{-1/2} D_{\psi}\tilde{Q}_n(\psi_o)'$$

and $A_{\gamma} = I - B_{\gamma}(B'_{\gamma}B_{\gamma})^{-1}B'_{\gamma}$, where $B_{\gamma} = \left[D^2_{\psi}\tilde{Q}_n(\psi)\right]^{1/2} DG(\gamma)$. Now, according to Chapter 3.2, Z_n will be asymptotically normal

$$N_{p+q}(0,I) .$$

Further,

(3.4) $$\sqrt{n}\,(\tilde{\gamma} - \gamma_o) = -\sqrt{n}\left[D^2_{\gamma}\tilde{Q}_n(G(\gamma_o))\right]^{-1} DG(\gamma_o)' D_{\psi}\tilde{Q}_n(\psi_o)' + \sqrt{n}\, o(\tilde{\gamma} - \gamma_o) ,$$

(3.5) $$\sqrt{n}\,(\tilde{\psi}_R - \psi_o) = \sqrt{n}\, DG(\gamma_o)(\tilde{\gamma} - \gamma_o) + \sqrt{n}\, o(\tilde{\gamma} - \gamma_o)$$

and

(3.6) $$\sqrt{n}\,(\tilde{\psi} - \psi_o) = -\sqrt{n}\left[D^2_{\psi}\tilde{Q}_n(\psi_o)\right]^{-1} D_{\psi}\tilde{Q}_n(\psi_o)' + \sqrt{n}\, o(\tilde{\psi} - \psi_o) .$$

By noting that

$$\operatorname*{plim}_{n\to\infty}\left[D^2_{\gamma}\tilde{Q}_n(G(\gamma_o)) - DG(\gamma_o)' D^2_{\psi}\tilde{Q}_n(\psi_o) DG(\gamma_o)\right] = 0$$

and by combining (3.4), (3.5) and (3.6) we get

$$\frac{\sqrt{n}}{\sqrt{4}}\,(\tilde{\psi} - \tilde{\psi}_R) \xrightarrow{L} \left[D^2_{\psi}\tilde{Q}_n(\psi_o)\right]^{-1/2} A_{\gamma_o} Z_n .$$

Furthermore, $D_\psi \tilde{Q}_n(\tilde{\psi}) = 0$ implies

$$\tilde{Q}_n(\tilde{\psi}_R) = \tilde{Q}_n(\tilde{\psi}) + \frac{1}{2}(\tilde{\psi}_R - \tilde{\psi})' D_\psi^2 \tilde{Q}_n(\tilde{\psi})(\tilde{\psi}_R - \tilde{\psi}) + o[(\tilde{\psi}_R - \tilde{\psi})'(\tilde{\psi}_R - \tilde{\psi})]$$

and

$$\frac{n}{2\pi}[\tilde{Q}_n(\tilde{\psi}_R) - \tilde{Q}_n(\tilde{\psi})] \xrightarrow{L} \frac{\sqrt{n}}{\sqrt{4\pi}}(\tilde{\psi}_R - \tilde{\psi})' D_\psi^2 \tilde{Q}_n(\psi_o)(\tilde{\psi}_R - \tilde{\psi}) \frac{\sqrt{n}}{\sqrt{4\pi}}$$

$$\xrightarrow{L} Z_n' A_{\gamma_o} Z_n .$$

Because A_{γ_o} is a projection on a $p+q-g$ dimensional subspace, the asymptotic distribution of

$$\frac{n}{2\pi}[\tilde{Q}_n(\tilde{\psi}_R) - \tilde{Q}_n(\tilde{\psi})] \tag{3.7}$$

will be χ^2_{p+q-g}.

Note that we have made no distinction between $\tilde{Q}_n(\psi)$ and

$$\tilde{Q}_{n,R}(\psi) = \frac{\pi}{M_n} \sum_{j=-M_n+1}^{M_n} \frac{2\pi \hat{f}_\varepsilon(\lambda_j)}{\tilde{\sigma}_R^2 h(\lambda_j,\psi)} \log \frac{2\pi \hat{f}_\varepsilon(\lambda_j)}{\tilde{\sigma}_R^2 h(\lambda_j,\psi)}$$

because under the null hypothesis

$$\operatorname*{plim}_{n\to\infty} \frac{\tilde{\sigma}_R^2}{\sigma_o^2} = 1 .$$

For the same reason, $\bar{\sigma}^2$ can be substituted for $\tilde{\sigma}^2$. The statistic (3.7) can be used as a test statistic corresponding to the usual LR - test.

In a similar way, one can show that

$$R = \frac{n}{16\pi^2} D_\psi \tilde{Q}_n(\tilde{\psi}_R) I(\tilde{\psi}_R)^{-1} D_\psi \tilde{Q}_n(\tilde{\psi}_R)'$$

will be asymptotically distributed as χ^2_{p+q-g}.

If the null hypothesis is of the form

$$H(\psi) = 0 , \qquad \dim(H(\Psi)) = p+q-g ,$$

where $H:\Psi \to \mathbb{R}^{p+q-g}$ is continuously differentiable, then

$$nH(\tilde{\psi})' [DH(\tilde{\psi}) I(\tilde{\psi})^{-1} DH(\tilde{\psi})']^{-1} H(\tilde{\psi})$$

will be asymptotically χ^2_{p+q-g} distributed

The result (3.7) also implies that model selection criteria corresponding to Akaike's AIC criterion and Schwarz's SBIC criterion can be defined on the basis of the ME -estimates. The definitions will be

$$\text{AICME} = \frac{n}{2\pi} \tilde{Q}_n(\tilde{\psi}) + 2(p+q)$$

and

$$\text{SBICME} = \frac{n}{2\pi} \tilde{Q}_n(\tilde{\psi}) + (p+q)\log n .$$

Although we strongly recommend the visual comparison of the spectral densities as the primary means of model selection, these criteria may sometimes be of interest.

3.4 Some Practical Experiences of the ME-Estimation Method

One of the first choices to be made when applying the ME -procedure is the form and width of the lag window. As we already mentioned in Chapter 3.1, we prefer the Parzen or Tukey-Hamming forms, because they never give negative estimates for the spectral densities. With regard to the error variances they compete well with other estimators (cf. Hannan, 1970, Table 1, p.282).

As we saw in Chapter 3.2, the width of the lag window M_n should be increased slower than $\sqrt{n}$ but faster than $\sqrt[4]{n}$ with increasing n to guarantee the asymptotic efficiency of the ME -method. This, however, does not give any directions for the choice of M_n in real estimation situations with for example n less than 100. By experience, we recommend fairly wide windows, for instance up to $M_n = 45$ with $n = 80$. But we want to emphasize that graphs of the time series and its periodogram should be inspected before deciding upon the value of M_n. If there is seasonal variation in the series, the lag window should be wide enough to include sufficiently many seasonal periods.

If a great proportion of the spectral mass is concentrated at low frequencies, one should avoid excessive smoothing. On the other hand,

if the periodogram looks very "edgy", sufficient smoothing is recommended.

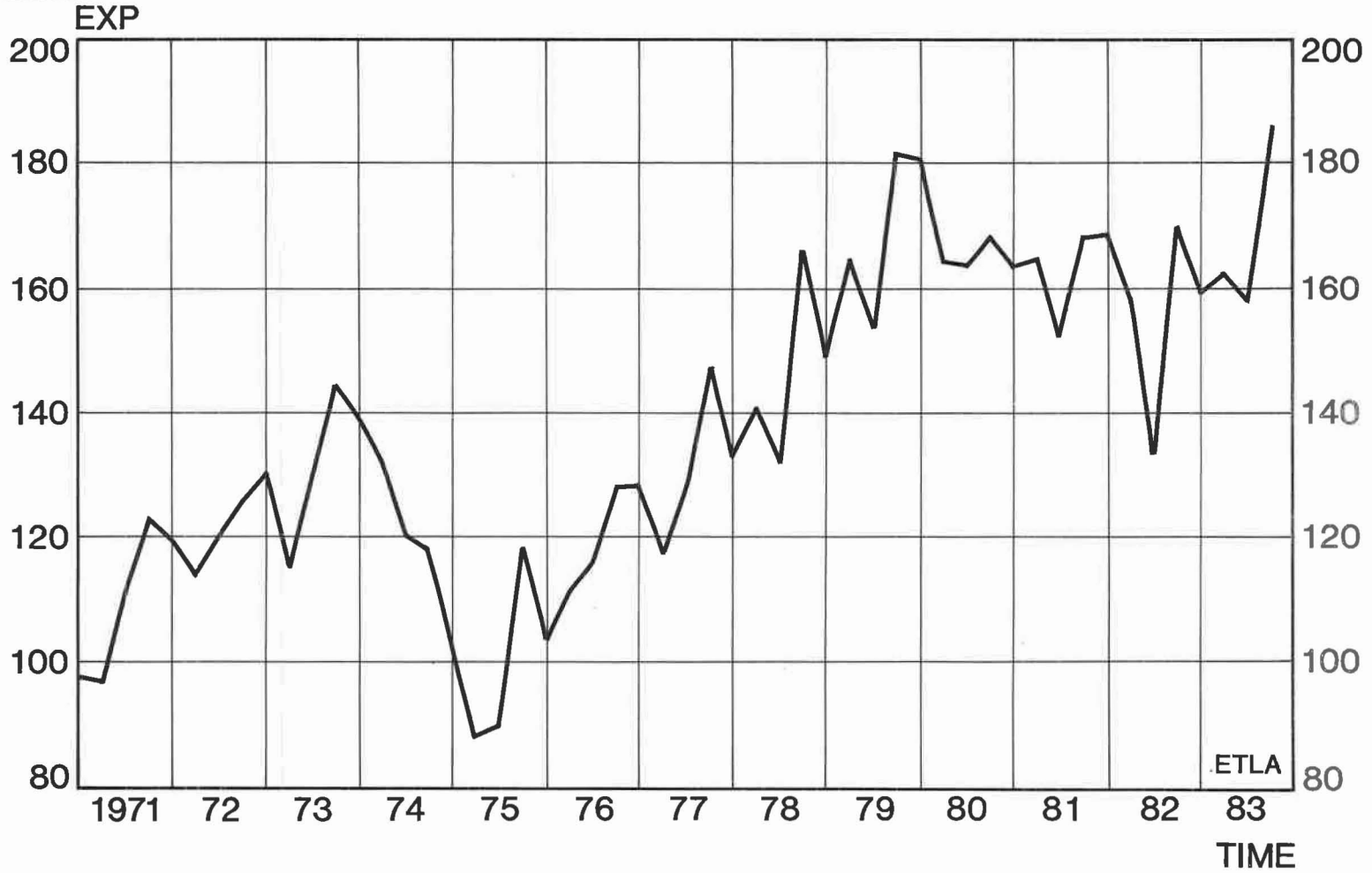
The number of frequency points where the spectral density is estimated can be chosen to equal M_n. If M_n is small, one might wish to use more frequency points, because otherwise the graphs of the spectra will look clumsy. This can be done, since an increase in the number of frequency points seems to affect the ME -estimates relatively little.

Experience has shown that the ME -method has a tendency to underestimate the absolute values of multiplicative seasonal parameters, especially MA -parameters. We tried to improve the method by taking the smoothing into account in the calculation of the theoretical spectra, but managed to sharpen the estimates disappointingly little. It was hardly worth the extra computational effort that was needed to smooth $h(\lambda,\psi)$. Thus, we simply suggest that the user should keep this small sample bias in mind when the model contains multiplicative seasonal MA -parameters.

To illustrate the use of the ME -estimation and identification method, we will consider the following example:

Example 1: For later reference we wanted to build a univariate ARIMA -model for the quarterly series of the volume of Finnish exports to western countries in 1971-1983. The graph of the series is shown in Figure 1. We calculated the annual logarithmic differences of the export volumes denoted by D4LEXP to make the series stationary. In the estimation of the spectral density of D4LEXP we used the Parzen

FIGURE 1 Quarterly volume of Finnish exports to western countries (1975 = 100)



window estimator. Because the number of observations was small, $n = 48$, we chose $M_n = 25$. We estimated the spectral density at 50 equispaced frequencies in the interval $[0,\pi]$.

By means of the autocorrelation and partial autocorrelation coefficients we tentatively identified four models presented in Table 1. In the representation of the orders of the models we have used the notation of Box and Jenkins (1970, Ch. 9). Having estimated the parameters of the four models with the ME -technique, we recorded the following values for the AICME- and SBICME -criteria:

TABLE 1

Model	Model form	AICME	SBICME
1	$(1,2)\times(1,0)_4$	-79.28	-71.71
2	$(1,2)\times(0,1)_4$	-80.25	-72.68
3	$(1,0)\times(0,2)_2$	-80.61	-74.93
4	$(1,0)\times(0,3)_2$	-80.90	-73.34

According to these criteria, models 3 and 4 seem to fit the data best. In Figures 2 and 3 we have illustrated the goodness of fit of models 4 and 3. Model 4 seems to be superior to model 3 at low frequencies.

Finally, we estimated the parameters of model 4 with the ML -method and used the ME -estimates as starting values in the numerical optimization of the likelihood function.

FIGURE 2 Spectral density of D4LEXP and a spectral density of model type 4 (broken line)

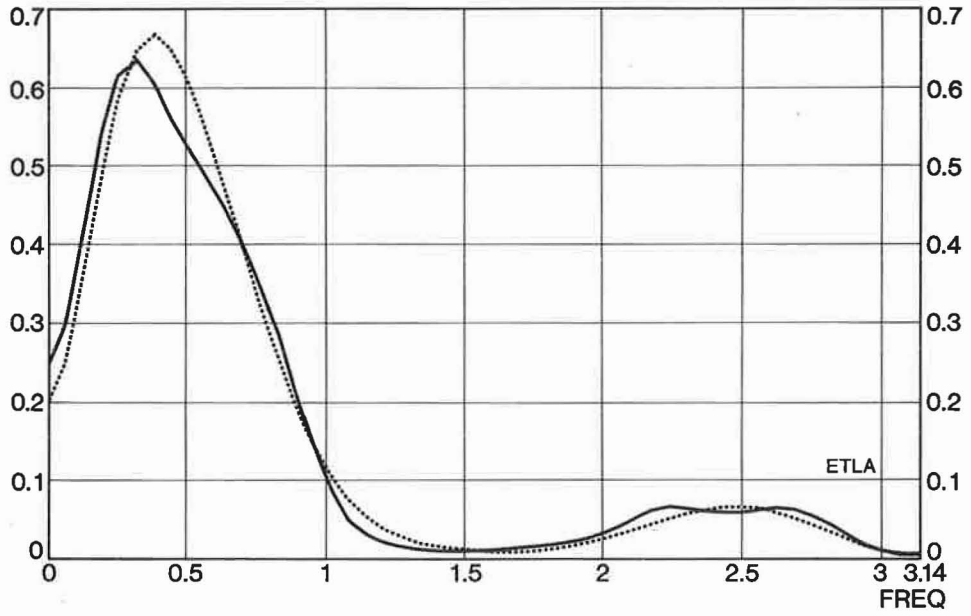


FIGURE 3 Spectral density of D4LEXP and a spectral density of model type 3 (broken line)

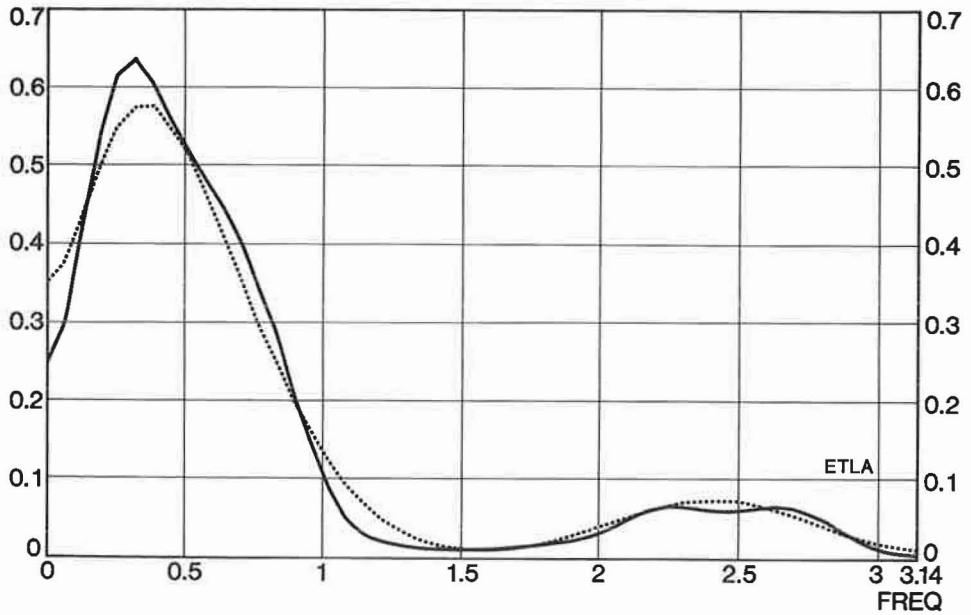


TABLE 2

Parameter	ME-estimate	(std. dev.)	ML-estimate	(std. dev.)
ϕ_1	0.717	(0.091)	0.669	(0.120)
θ_1	-0.238	(0.190)	-0.462	(0.130)
θ_2	0.497	(0.108)	0.696	(0.135)
θ_3	0.230	(0.188)	0.746	(0.141)
σ	0.086		0.066	

Thus, the ME -estimates seemed to provide reasonable starting values for the ML -iteration. The Box-Ljung test statistic (cf. Ljung and Box, 1978) for the eight first autocorrelations of the ML -residual was

$$\text{B-L} = 4.27 ,$$

so the model seemed to fit the data well.

4. IDENTIFICATION OF MULTIPLE INPUT TRANSFER FUNCTION MODELS

4.1 Some Comments on Earlier Identification Methods

Short descriptions of the most important identification techniques were already given in Chapter 2. We want to emphasize that the parallel use of several identification techniques is often recommendable.

If there is any doubt about possible feedback effects between the input and output variables, one should take the trouble of calculating cross correlations between the prewhitened series. The prewhitening is necessary, because the distributions of sample cross correlations depend on the autocorrelation structures of the variables. For instance, the Box-Ljung test (cf. Ljung and Box, 1978) can only be applied to white noise series. Because the transformations induced by ARIMA-models are "one-sided", the temporal relations between the variables can be inspected on the basis of the ARIMA-residuals as well.

There are, however, many types of variables that do not allow an ARIMA-representation. For instance seasonal dummies and variables measuring real tariffs of some services are slightly problematic in this respect. All variables cannot be made stationary by differencing.

Spectral identification methods naturally require stationarity, too, but these methods do not immediately break down when the variables are

slightly nonstationary. If there are clear trends in the variables, almost all of the spectral mass will concentrate at low frequencies and the smoothing can badly distort the picture. But if the trends do not dominate the variations of the series, it is often worth one´s while to calculate cross spectral densities and to have a look at temporal differences (or differences between phase angles) and coherencies between the variables. Although graphs of temporal differences are often tricky to interpret, they are sometimes useful in the model selection.

If one tries to use the ordinary least squares method in the identification and ignores at first the autocorrelation of the residuals, one should start with fairly a large model including all potentially relevant lags. The total number of parameters should of course be kept within reasonable limits with regard to the number of observations. The OLS -estimates will be unbiased and consistent, but it will be difficult to assess their significance.

If the impulse responses of some explanatory variable seem to die out only gradually, a suitable rational structure for them can be found for example by the corner method (cf. Liu and Hanssens, 1982).

One should be very careful in reducing the model at this stage, because the distribution of the estimates depends heavily on the behavior of the residuals. After fitting an ARMA-model for the OLS residuals one can, however, roughly assess the significance of different test statistics calculated in connection with the OLS estimation. This is possible by means of the following theorem:

Theorem 4.1.1 Let $y(t) = \beta' X(t) + \varepsilon(t)$, $t = 1,\ldots,n$ be a linear regression model with a stationary normal ARMA(p,q)-residual $\varepsilon(t)$ obeying the model

$$\phi(L)\varepsilon(t) = \theta(L)\alpha(t) , \qquad \alpha(t) \sim \text{NID}(0,\sigma^2) , \quad t = \ldots,-1,0,1,\ldots .$$

Let $R\beta = \gamma$ be a testable linear hypothesis, where R is an $r \times m$ -matrix. Let F be the usual (unweighted) F-test statistic for this hypothesis. Let $\xi_i^{(\phi)}$, $i = 1,\ldots,p$, and $\xi_j^{(\theta)}$, $j = 1,\ldots,q$ be the roots of the characteristic equations $\phi(\xi^{-1}) = 0$ and $\theta(\xi^{-1}) = 0$, respectively. Then, asymptotically

$$P\{F > c_1 t_\delta\} \leq \delta$$

and

$$P\{F < \frac{c_2}{c_1} t_\delta\} \leq 1 - \delta$$

irrespective of the behavior of the $X(t)$ -process, where

$$c_1 = \prod_{j=1}^{q} (1+|\xi_j^{(\theta)}|)^2 \Big/ \prod_{i=1}^{p} (1-|\xi_i^{(\phi)}|)^2$$

$$c_2 = \prod_{j=1}^{q} (1-|\xi_j^{(\theta)}|)^2 \Big/ \prod_{i=1}^{p} (1+|\xi_i^{(\phi)}|)^2$$

and t_δ is the 100δ per cent critical value of an F-distribution with r and $n-m$ degrees of freedom.

The proof of the theorem is left to the appendix.

Theorem 4.1.1 motivates the following rule of thumb for inference from the OLS -calculations:

- If the unweighted F-statistic exceeds the usual critical value multiplied by c_1, the null hypothesis can be rejected at the chosen level of significance.
- If F does not exceed the usual critical value multiplied by $\frac{c_2}{c_1}$, the null hypothesis could not be rejected under any circumstances, because the true critical value exceeds $\frac{c_2}{c_1} t_\delta$.

If $r = 1$, the usual critical values of the corresponding t-test statistic should be multiplied by $\sqrt{c_1}$ and $\sqrt{c_2/c_1}$. The values of c_1 and c_2 can be estimated by fitting an ARMA(p,q)-model to the OLS-residuals.

The rule of thumb suggested above is a modification of earlier results of Watson (cf. Watson, 1955 and Watson and Hannan, 1956) and Vinod (cf. Vinod, 1976 and also Sathe and Vinod, 1974 or Kiviet, 1980). We have tried to find a formulation that would be as easy to use as possible when p and q are small. If p and q are large, the discrepancy between $\sigma^2 c_1$ and $\sigma^2 c_2$ and the largest and smallest eigenvalues of $\mathrm{cov}([\varepsilon(1) \ \ldots \ \varepsilon(n)]')$ can become substantial. In that case, the rule of thumb is not of much value.

4.2 The Maximum Entropy - Generalized Least Squares (ME-GLS) Technique

In the previous chapter we saw that residual autocorrelation can badly distort tests of significance based on OLS estimates. That is why we have to get some idea of the form of the residual autocorrelation before we can reliably assess the roles of different terms in the structural model. But before we can calculate any residuals, we must be able to estimate the transfer functions. This vicious circle can be broken by observing that spectral methods make it possible to estimate the spectral density of that part of $y(t)$ that is totally uncorrelated with the explanatory series $x_1(t),\ldots,x_m(t)$. If the observations obey model (2.1), this uncorrelated part will be exactly $\varepsilon(t)$ irrespective of the form of the linear transfer function. The ME -estimation technique, presented in Chapter 3, gives us an opportunity to fit ARMA-models for the residuals without actually computing any of them. When a suitable ARMA-model has been found, we can calculate a weighted covariance matrix of all the potentially relevant lags of the input and output variables using the estimated Σ_n^{-1} (see Chapter 5.2) as weight matrix. With this weighted covariance matrix we can easily compute the generalized least squares (GLS) estimates for the parameters of different transfer functions.

The suggested estimation procedure resembles in many ways the so-called Hannan's efficient method (cf. Hannan, 1963). In our method, the number of (latent) parameters is smaller because we require the residual series to be a member of the ARMA -family. Because of its optimal small sample properties, the use of the GLS -method in time domain seems more natural than the method suggested by Hannan.

Let us now denote

$$\hat{g}_\varepsilon(\lambda) = \hat{f}_y(\lambda) - \overline{\hat{f}_{Xy}(\lambda)'\hat{f}_X(\lambda)^{-1}\hat{f}_{Xy}(\lambda)} \ ,$$

where $\hat{f}_y$, $\hat{f}_{Xy}$ and $\hat{f}_X$ will mean the Parzen or Tukey-Hamming estimators of f_y, f_{Xy} and f_X, respectively. We will use the letter g here to emphasize the difference between $\hat{g}_\varepsilon$ and $\hat{f}_\varepsilon$. Anyway, we can fit different ARMA-spectra for $\hat{g}_\varepsilon$ as well, but it is not immediately clear that the resulting estimates will share the optimality properties in Theorem 3.2.2. This will be shown in the next chapter.

If the data has been differenced and if at the same time the spectral density of the residuals seems to have very low values at the frequencies corresponding to the differentiation span, we suggest that a model connecting the original undifferenced series with each other should be specified. This does not mean, however, that we could not identify the model and estimate its parameters in the differenced form as well, because the MA-part of the residual model does not have to be invertible.

If for instance quarterly data has been seasonally differenced and the estimated residual spectral density has a very low value at frequency $\pi/2$, a seasonal MA-parameter should be included in the residual model. The ME-estimate of the parameter will probably be fairly large (close to unity). It should be replaced by 1 in the calculation of the weighted covariances between the input and output variables. By this means the effects of the differencing will be taken correctly into account and valid conclusions about the form of the structural model can be made.

4.3 The Asymptotic Properties of the ME-GLS-Estimators

We will first consider the ME -estimators for $\psi = [\phi' \quad \theta']$ based on

$$\hat{g}_\varepsilon(\lambda) = \hat{f}_y(\lambda) - \overline{\hat{f}_{xy}(\lambda)'\hat{f}_x(\lambda)^{-1}\hat{f}_{xy}(\lambda)}$$

before discussing the asymptotic properties of the GLS -estimators for the structural parameters.

The results in Sections 3.2 and 3.3 are not directly applicable, because $\hat{g}_\varepsilon$ cannot be interpreted as a window estimator of some observed or unobserved series. Anyway, it turns out that ME -estimator $\tilde{\psi}$ based on $\hat{g}_\varepsilon$ has the same asymptotic distribution as the one based on the window estimator $\hat{f}_\varepsilon$ that could be computed if the error terms $\varepsilon(t)$ were observable.

Define

$$z(t) = \frac{\omega_1(L)}{\delta_1(L)} x_1(t) + \ldots + \frac{\omega_m(L)}{\delta_m(L)} x_m(t)$$

and

$$y(t) = z(t) + \varepsilon(t) \qquad t = 1,\ldots,n \ .$$

The essential assumption here is that the number of structural parameters should be finite. Model (2.3) will not be discussed separately, because all the following statements will apply to it as well.

As before, let us denote

$$\delta^*(L) = \delta_1(L) \ldots \delta_m(L)$$

$$\omega_k^*(L) = \delta^*(L) \frac{\omega_k(L)}{\delta_k(L)} \qquad k = 1,\ldots,m$$

and

$$\omega^*(L) = [\omega_1^*(L) \ldots \omega_m^*(L)]' .$$

Let s denote the maximum of the degrees of the polynomials $\delta^*(L)$, $\omega_1^*(L),\ldots,\omega_m^*(L)$.

Suppose now that starting values $z(0),\ldots,z(-s)$; $x(0),\ldots,x(-s)$ and $\varepsilon(0),\ldots,\varepsilon(-s)$ are available. We will use the following notations:

$$F_x(\lambda) = (2\pi n)^{-1/2} \sum_{t=1}^{n} x(t)e^{it\lambda}$$

and

$$\hat{f}_{xy}(\lambda) = \int_{-\pi}^{\pi} W_n(\lambda - \lambda')F_x(\lambda')\overline{F_y(\lambda')}\, d\lambda' ,$$

where

$$W_n(\lambda) = \frac{1}{2\pi} \sum_{\nu=-M_n}^{M_n} w_\nu e^{i\nu\lambda} .$$

Suppose that the input sequences $X(t)$ satisfy the following conditions:

A1: $X(t)$, though possibly "fixed", will behave like a stationary process with a finite vector-ARMA-representation, i.e., the second order moments will die out exponentially and the fourth order cumulants will constitute an absolutely convergent series (cf. Hannan, 1970, p. 280).

A2: $X(t)$ is bounded and $\| \hat{f}_X(\lambda) \|$ and $\| \hat{f}_X(\lambda)^{-1} \|$ are bounded (cf. Brillinger, 1975, p. 198).

Then the following two lemmas will hold for $X(t)$ and for linear functions of $X(t)$:

Lemma 4.3.1

$$\hat{f}_{x_1,Lx_2}(\lambda) = \hat{f}_{x_1,x_2}(\lambda)e^{-i\lambda} + O(1/M_n^2) + o(M_n/n)$$

uniformly in λ.

Proof Because

$$F_{Lx_2}(\lambda) = (2\pi n)^{-1/2} \sum_{t=1}^{n} x_2(t-1)e^{it\lambda}$$

$$= F_{x_2}(\lambda)e^{i\lambda} + (2\pi n)^{-1/2}[x_2(0) - x_2(n)e^{in\lambda}] ,$$

it follows that

$$\hat{f}_{x_1,Lx_2}(\lambda) = \int_{-\pi}^{\pi} W_n(\lambda - \lambda')F_{x_1}(\lambda')\overline{F_{x_2}(\lambda')}e^{-i\lambda'} \, d\lambda' \tag{4.1}$$

$$+ (2\pi n)^{-1/2} \int_{-\pi}^{\pi} W_n(\lambda - \lambda')F_{x_1}(\lambda')[x_2(0) - x_2(n)e^{-in\lambda'}] \, d\lambda' .$$

On the other hand,

$$\int_{-\pi}^{\pi} W_n(\lambda - \lambda')F_{x_1}(\lambda')e^{-in\lambda'}\, d\lambda'$$

$$= \frac{1}{2\pi} \sum_{\nu=-M_n}^{M_n} w_\nu e^{i\nu\lambda}(2\pi n)^{-1/2} \sum_{t=1}^{n} x_1(t) \int_{-\pi}^{\pi} e^{i(t-(n+\nu))\lambda'}\, d\lambda'$$

$$= O(M_n/\sqrt{n})$$

because $x_1(t)$ is bounded. Thus, the second term of (4.1) will be $O(M_n/n)$. The first term of (4.1) can be written in the form

$$\text{(4.2)} \qquad \int_{-\pi}^{\pi} W_n(\lambda - \lambda')F_{x_1}(\lambda')\overline{F_{x_2}(\lambda')}\, d\lambda' \cdot e^{-i\lambda}$$

$$+ \int_{-\pi}^{\pi} [W_n(\lambda - \lambda')e^{i(\lambda-\lambda')} - W_n(\lambda - \lambda')]F_{x_1}(\lambda')\overline{F_{x_2}(\lambda')}\, d\lambda' \cdot e^{-i\lambda}$$

$$= \hat{f}_{x_1,x_2}(\lambda)e^{-i\lambda} + \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{\nu=-M_n+1}^{M_n} \left[k\left(\frac{\nu-1}{M_n}\right) - k\left(\frac{\nu}{M_n}\right)\right]$$

$$\cdot e^{i(\lambda-\lambda')\nu} F_{x_1}(\lambda')\overline{F_{x_2}(\lambda')}\, d\lambda' \cdot e^{-i\lambda} .$$

Because the k-functions were supposed to be twice continuously differentiable, we can write

$$k\left(\frac{\nu-1}{M_n}\right) - k\left(\frac{\nu}{M_n}\right) = -\frac{\nu}{M_n} k'\left(\frac{\nu}{M_n}\right) + \frac{\nu^2}{M_n^2} k''\left(\frac{\nu}{M_n}\right) + o\left(\frac{\nu^2}{M_n^2}\right) .$$

The second term on the right hand side of (4.2) can thus be written in the form

$$(4.3)\qquad \frac{1}{2\pi}\sum_{\nu=-M_n+1}^{M_n}\left[-\frac{\nu}{M_n}k'\left(\frac{\nu}{M_n}\right)+\frac{\nu^2}{M_n^2}k''\left(\frac{\nu}{M_n}\right)+o\left(\frac{\nu^2}{M_n^2}\right)\right]$$

$$\cdot e^{i\nu\lambda}\int_{-\pi}^{\pi}e^{-i\nu\lambda'}F_{x_1}(\lambda')\overline{F_{x_2}(\lambda')}\,d\lambda'$$

$$=-\frac{1}{M_n}\cdot\frac{1}{2\pi}\sum_{\nu=-M_n+1}^{M_n}\nu k'\left(\frac{\nu}{M_n}\right)e^{i\nu\lambda}r_{x_1,x_2}(\nu)+O(1/M_n^2)$$

because $k''\left(\frac{\nu}{M_n}\right)+M_n^2\,o\left(\frac{\nu^2}{M_n^2}\right)\frac{1}{\nu^2}$ stays bounded. For the Parzen window

$$k'(x)=\begin{cases}-12x+18x^2\operatorname{sgn}(x) & \text{for } |x|\le\frac{1}{2}\\ -6(1-|x|)^2\operatorname{sgn}(x) & \text{for } |x|>\frac{1}{2}\end{cases}$$

and for the Tukey-Hamming window

$$k'(x)=-0.46\pi\sin\pi x\ .$$

Thus, we can write

$$k'(x)=xk^0(x)\ ,$$

where $k^0(x)$ remains bounded for all $-1\le x\le 1$. Because the cross correlations $r_{x_1,x_2}(\nu)$ were supposed to die out exponentially, the series

$$\sum_{\nu=-M_n+1}^{M_n} \nu^2 |r_{x_1,x_2}(\nu)|$$

will converge when $M_n \to \infty$. That is why the whole expression (4.3) will be $O(1/M_n^2)$. This completes the proof of the lemma.

□

Because $\hat{f}_{x_1,x_2}$ is bilinear with respect to x_1 and x_2, this immediately implies the following lemma:

Lemma 4.3.2

$$\hat{f}_{\delta_1(L)x_1,\delta_2(L)x_2}(\lambda) = \delta_1(e^{i\lambda})\hat{f}_{x_1,x_2}(\lambda)\delta_2(e^{-i\lambda}) + O(1/M_n^2) + O(M_n/n) ,$$

where $\delta_1(L)$ and $\delta_2(L)$ are any lag polynomials of finite degrees.

After these preliminaries, let us define

$$(4.4) \qquad z^*(t) = \delta^*(L)z(t) - \omega^*(L)'X(t) \qquad t = 1,\ldots,n .$$

Because $z^*(t) \equiv 0$ for all $t = 1,\ldots,n$, we have

$$\hat{f}_{z^*}(\lambda) = \hat{f}_{z^*,z^*}(\lambda) \equiv 0 .$$

On the other hand, Lemma 4.3.2 implies

$$(4.5)\qquad \hat{f}_{z^*}(\lambda) = \delta^*(e^{i\lambda})\hat{f}_z(\lambda)\delta^*(e^{-i\lambda}) - \delta^*(e^{i\lambda})\hat{f}_{z,X}(\lambda)\omega^*(e^{-i\lambda})$$

$$- \omega^*(e^{i\lambda})'\overline{\hat{f}_{z,X}(\lambda)}'\delta^*(e^{-i\lambda}) + \omega^*(e^{i\lambda})'\hat{f}_X(\lambda)\omega^*(e^{-i\lambda})$$

$$+ O(1/M_n^2) + O(M_n/n) .$$

Because substituting

$$\begin{bmatrix} 1 \\ \hat{f}_X(\lambda)^{-1}\hat{f}_{X,z}(\lambda) \end{bmatrix} \quad \text{for} \quad \begin{bmatrix} 1 \\ \delta^*(e^{-i\lambda})^{-1}\omega^*(e^{-i\lambda}) \end{bmatrix}$$

will minimize the nonnegative quadratic form, constituted by the first four terms on the right hand side of (4.5), this implies

$$(4.6)\qquad \hat{f}_z(\lambda) - \overline{\hat{f}_{X,z}(\lambda)}'\hat{f}_X(\lambda)^{-1}\hat{f}_{X,z}(\lambda) = O(1/M_n^2) + O(M_n/n)$$

uniformly in λ.

On the other hand, (4.4), Lemma 4.3.2 and the form of the model imply

$$(4.7)\qquad 0 \equiv \hat{f}_{z^*,\varepsilon}(\lambda) = \delta^*(e^{i\lambda})\hat{f}_{z,\varepsilon}(\lambda) - \omega^*(e^{i\lambda})'\hat{f}_{X,\varepsilon}(\lambda)$$

$$+ O(1/M_n^2) + O(M_n/n)$$

and

$$(4.8) \qquad 0 \equiv \hat{f}_{z^*,X}(\lambda) = \delta^*(e^{i\lambda})\hat{f}_{z,X}(\lambda) - \omega^*(e^{i\lambda})'\hat{f}_X(\lambda) + O(1/M_n^2) + O(M_n/n) .$$

Because $\varepsilon(t)$- and $X(t)$ -processes are independent of each other, $f_{X,\varepsilon}(\lambda) \equiv 0$ and thus

$$(4.9) \qquad \hat{f}_{X,\varepsilon}(\lambda) = O((M_n/n)^{1/2}) .$$

By combining this with (4.7) and (4.8) we get

$$(4.10) \qquad \hat{f}_{z,\varepsilon}(\lambda) - \overline{\hat{f}_{X,z}(\lambda)'\hat{f}_X(\lambda)^{-1}}\hat{f}_{X,\varepsilon}(\lambda)$$

$$= \hat{f}_{z,\varepsilon}(\lambda) - \delta^*(e^{i\lambda})^{-1}\omega^*(e^{i\lambda})\hat{f}_{X,\varepsilon}(\lambda)$$

$$+ [\delta^*(e^{i\lambda})^{-1}\omega^*(e^{i\lambda})' - \overline{\hat{f}_{X,z}(\lambda)'\hat{f}_X(\lambda)^{-1}}]\hat{f}_{X,\varepsilon}(\lambda)$$

$$= O(1/M_n^2) + O(M_n/n) + [O(1/M_n^2) + O(M_n/n)]O((M_n/n)^{1/2})$$

$$= O(1/M_n^2) + O(M_n/n)$$

because $\| \hat{f}_X(\lambda)^{-1} \|$ and $\| \hat{f}_X(\lambda) \|$ are by Assumption A2 uniformly bounded. Equation (4.9) also implies

$$(4.11) \qquad \overline{\hat{f}_{X,\varepsilon}(\lambda)'\hat{f}_X(\lambda)^{-1}}\hat{f}_{X,\varepsilon}(\lambda) = O(M_n/n) .$$

Finally, by combining (4.6), (4.10) and (4.11) we get

$$\hat{g}_\varepsilon(\lambda) - \hat{f}_\varepsilon(\lambda)$$

$$= \hat{f}_z(\lambda) + \hat{f}_{z,\varepsilon}(\lambda) + \overline{\hat{f}_{z,\varepsilon}(\lambda)} - \overline{\left[\hat{f}_{\chi,z}(\lambda) + \hat{f}_{\chi,\varepsilon}(\lambda)\right]'} \hat{f}_\chi(\lambda)^{-1}$$

$$\cdot \left[\hat{f}_{\chi,z}(\lambda) + \hat{f}_{\chi,\varepsilon}(\lambda)\right]$$

$$= O(1/M_n^2) + O(M_n/n) .$$

When $H(\lambda)$ is any bounded, continuous function, it follows now that

$$\operatorname*{plim}_{n\to\infty} \sqrt{n} \left\| \frac{1}{M_n} \sum_{j=-M_n}^{M_n} H(\lambda_j)\hat{g}_\varepsilon(\lambda_j) - \frac{1}{M_n} \sum_{j=-M_n}^{M_n} H(\lambda_j)\hat{f}_\varepsilon(\lambda_j) \right\| = 0$$

because $\lim_{n\to\infty} \sqrt{n}/M_n^2 = 0$ and $\lim_{n\to\infty} M_n/\sqrt{n} = 0$.

By combining this result with Theorem 3.2.2 we have proved the following theorem:

Theorem 4.3.3 Under conditions A1, A2 and the conditions of Theorem 3.2.2, the ME -estimators $\tilde{\psi} = [\tilde{\phi}' \quad \tilde{\theta}']'$, based on $\hat{g}_\varepsilon(\lambda)$, are asymptotically normal and $\sqrt{n}\,(\tilde{\psi} - \psi)$ has the asymptotic distribution $N_{p+q}(0, I(\psi)^{-1})$, where the form of $I(\psi)$ is given in Theorem 3.2.2.

Let us now turn to the problem of estimating the structural parameters of the model (2.1). Suppose for a moment that $\delta^*(L) \equiv L^0$ or that the model is of the Almon type (2.3). Then, the model (2.1) can be put in the standard regression form

$$Y = X\omega + \varepsilon, \qquad \varepsilon \sim N_{n-s}(0,\sigma^2\Sigma(\psi)),$$

where

$$Y = [y(s+1) \ \ldots \ y(n)]' \ ,$$

$$X = \begin{bmatrix} x_1(s+1) & \ldots & x_1(1) & \ldots & x_m(s+1) & \ldots & x_m(1) \\ \vdots & & \vdots & & \vdots & & \vdots \\ x_1(n) & \ldots & x_1(n-s) & \ldots & x_m(n) & \ldots & x_m(n-s) \end{bmatrix} \ ,$$

$$\omega = [\omega_{1o} \ \ldots \ -\omega_{1s} \ \ldots \ \omega_{mo} \ \ldots \ -\omega_{ms}]' \ ,$$

$$\psi = [\phi' \quad \theta']' \qquad \text{and} \qquad \Sigma(\psi) = \sigma^{-2}\text{cov}(\varepsilon) \ .$$

An explicit formula (5.3) for $\sigma^2\Sigma(\psi) = \text{cov}(\varepsilon)$ will be derived in Chapter 5.2.

Because $\tilde{\psi}$ is consistent for ψ, it is easy to see from the corresponding inverted form (5.4) that

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} X'\Sigma(\tilde{\psi})^{-1}X = \lim_{n\to\infty} \frac{1}{n} X'\Sigma(\psi)^{-1}X$$

whenever the latter exists.

Because the generalized least squares estimator

$$\hat{\omega} = (X'\Sigma(\tilde{\psi})^{-1}X)^{-1}X'\Sigma(\tilde{\psi})^{-1}Y$$

has a conditional distribution

$$\hat{\omega}|\tilde{\psi} \sim N_{m(s+1)}(\omega,(X'\Sigma(\tilde{\psi})^{-1}X)^{-1}X'\Sigma(\psi)^{-1}X(X'\Sigma(\tilde{\psi})^{-1}X)^{-1}\sigma^2) , \tag{4.12}$$

the asymptotic distribution of $\sqrt{n}\ (\hat{\omega} - \omega)$ will be

$$N_{m(s+1)}(0,\sigma^2(\lim_{n\to\infty} \frac{1}{n} X'\Sigma(\psi)^{-1}X)^{-1}) . \tag{4.13}$$

Note that $\hat{\omega}$ is unbiased regardless of the value of $\tilde{\psi}$, and thus

$$\begin{aligned} \operatorname{cov}(\hat{\omega}) &= \underset{\tilde{\psi}}{E} \operatorname{cov}(\hat{\omega}|\tilde{\psi}) + \underset{\tilde{\psi}}{\operatorname{cov}} E(\hat{\omega}|\tilde{\psi}) \\ &= \underset{\tilde{\psi}}{E}(X'\Sigma(\tilde{\psi})^{-1}X)^{-1}X'\Sigma(\psi)^{-1}X(X'\Sigma(\tilde{\psi})^{-1}X)^{-1}\sigma^2 . \end{aligned}$$

Now, because the conditional distribution (4.12) is normal and the asymptotic distribution of $\tilde{\psi}$ is normal, the joint asymptotic distribution of $\hat{\omega}$ and $\tilde{\psi}$ will be normal, too.

On the other hand,

$$\operatorname{cov}(\hat{\omega},\tilde{\psi}) = \underset{\tilde{\psi}}{E}\left[E((\hat{\omega} - \omega)(\tilde{\psi} - \psi)'|\tilde{\psi})\right] = 0 \qquad \text{for all } n$$

because of the unbiasedness of $\hat{\omega}$. Thus, $\sqrt{n}\ (\hat{\omega} - \omega)$ and $\sqrt{n}\ (\tilde{\psi} - \psi)$ will be asymptotically independent of each other and their asymptotic distributions will be given by (4.13) and Theorem 4.3.3. This means that $\hat{\omega}$ and $\tilde{\psi}$ are jointly asymptotically efficient. Note that the information matrix will be of block diagonal form.

This result does not of course guarantee the "optimality" of the proposed estimators. It merely states that the proposed method does not systematically waste information. Although the asymptotic distribution of the estimators is identical with that of the ML -estimators, we believe that in finite samples the ML -method is still to be preferred. On the other hand, the ML -method cannot be used to solve ill-defined problems (i.e., if the models are overspecified or badly misspecified) and it is computationally cumbersome. These facts underline how important it is to identify the model form correctly before the ML -estimation. The ME -method combined with the GLS -procedure seems to be very well suited for identification purposes, because it is possible to choose the forms of the structural model and the residual model one at a time. The choice of the residual model can be made either visually or by means of the AICME or SBICME criteria as described in Chapter 3. The form of the structural model can be chosen with regard to the usual t- and F -statistics, to the signs and magnitudes of the parameter estimates, etc.

It should be noted that if a linear restriction

$$R\omega = \gamma$$

is added to the model, the corresponding least squares estimator

$$\hat{\omega}_R = \hat{\omega} - (X'\Sigma(\tilde{\psi})^{-1}X)^{-1}R'(R(X'\Sigma(\tilde{\psi})^{-1}X)^{-1}R')^{-1}(R\hat{\omega} - \gamma)$$

will be obtained from the weighted moment matrix

$$(4.14) \qquad \frac{1}{n} [Y \quad X]' \Sigma(\tilde{\psi})^{-1} [Y \quad X]$$

by the usual calculations. Thus, once (4.14) is calculated, the choice of the model form will be as easy (or as difficult) as the choice of a normal regression model. Ridge estimators, selection criteria for Almon models, estimators with "end conditions" in Almon models, etc., can be obtained from (4.14) in the usual manner.

Only if the initial guess for s is too large, it might be necessary to recalculate (4.14) in order to save observations at the beginning of the observation period.

If $\delta^*(L) \not\equiv L^0$, the identification problem is more difficult. To identify the forms of the transfer functions we suggest the calculation of (4.14) with fairly large s. After estimating all possible linear coefficients up to lag s by the GLS -method one can norm the coefficients by the largest absolute value of the coefficient estimates for each input variable. The resulting sequences can be taken as "autocorrelation sequences" and suitable rational forms can be identified for instance by the corner method of Beguin et al. (1980). For a short exposé of this method, see Liu and Hanssens (1982). We cannot claim the estimators of the structural parameters to be asymptotically efficient in this case. Anyway, the residual parameters will be estimated asymptotically efficiently by the ME -method.

4.4 An Example of the Use of the ME-GLS Method

Example 2: In Example 1 we estimated an ARIMA-model for the volume of Finnish exports to western countries. In this chapter we will try to include two explanatory variables in the model. The first one (called SEM) measures the size of the western market for Finnish export commodities and it is constructed as a weighted average of the indices of industrial production in the 11 most important OECD countries, using the values of Finnish exports to these countries as weights. The graph of the series has been presented in Figure 4. The second one (called WEP) is a ratio between the price index of Finnish exports and the price index of competing exports from other countries (see Figure 5). The latter index is constructed as a weighted average of the import price indices of the 11 OECD countries with the weight structure described above. The import price indices naturally contain commodities which Finland does not export, but the construction of a better price index for competing exports would be a cumbersome task. All variables were made stationary by taking logarithms and by computing their annual differences. (A prefix D4L will be attached to each variable symbol to denote this transformation.) We calculated all spectral and cross spectral densities between the three series at 50 frequencies with a Parzen window estimator of width $M_n = 25$. The corresponding residual spectral density $\hat{g}_\varepsilon$ (see Chapter 4.2) was estimated. Tentative models for the disturbances were identified by interpreting the Fourier transform of $\hat{g}_\varepsilon$ as an autocorrelation function. The 10 first autocorrelations and the corresponding partial autocorrelations were

FIGURE 4 Weighted average of quarterly indices of industrial production of 11 OECD countries (1980 = 100)

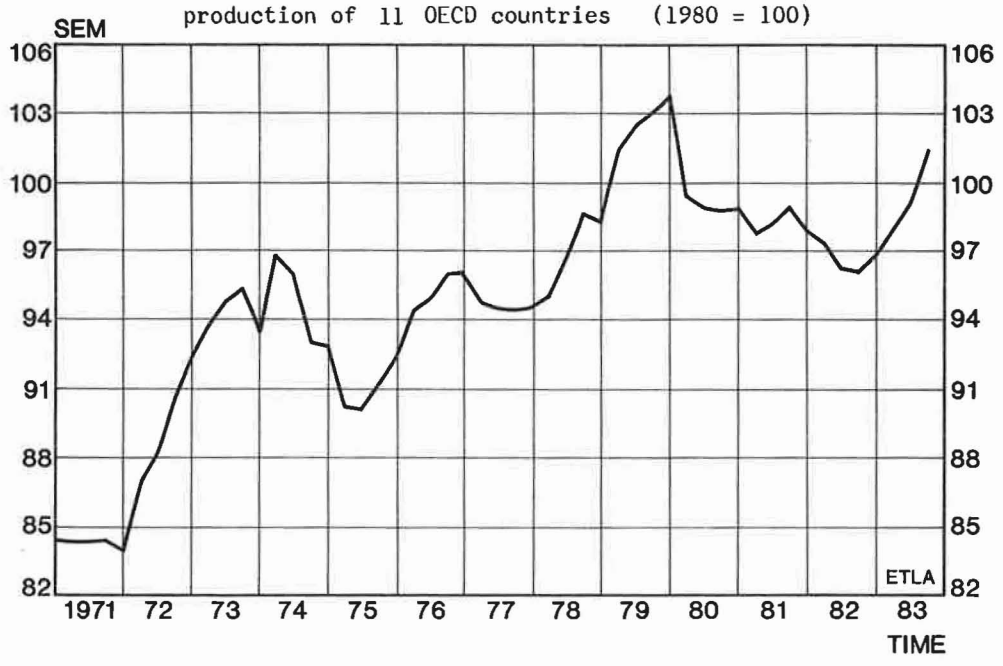


FIGURE 5 Relative price of Finnish exports commodities

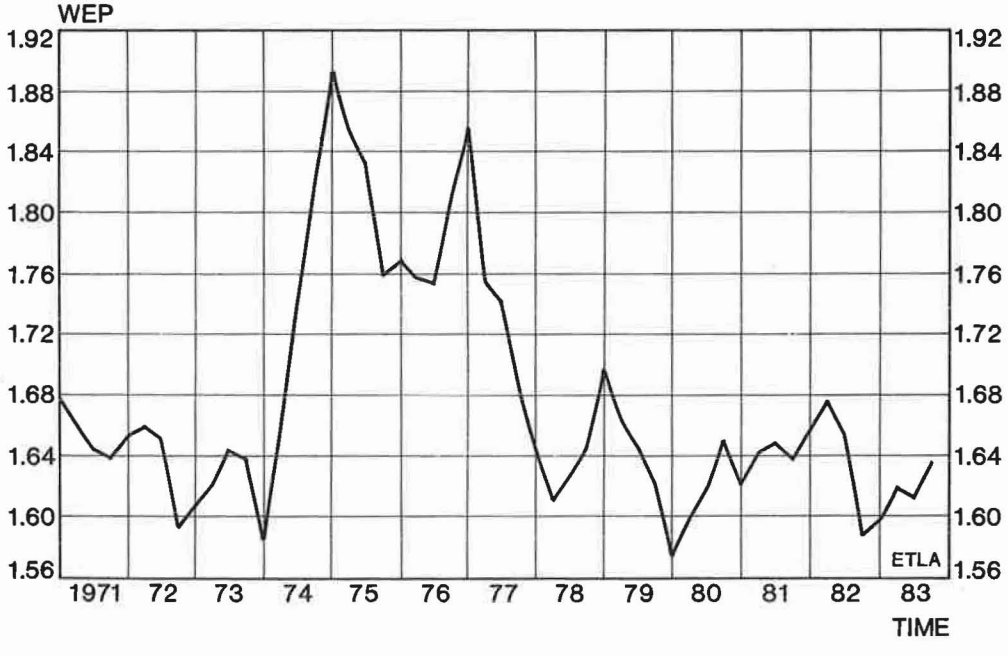
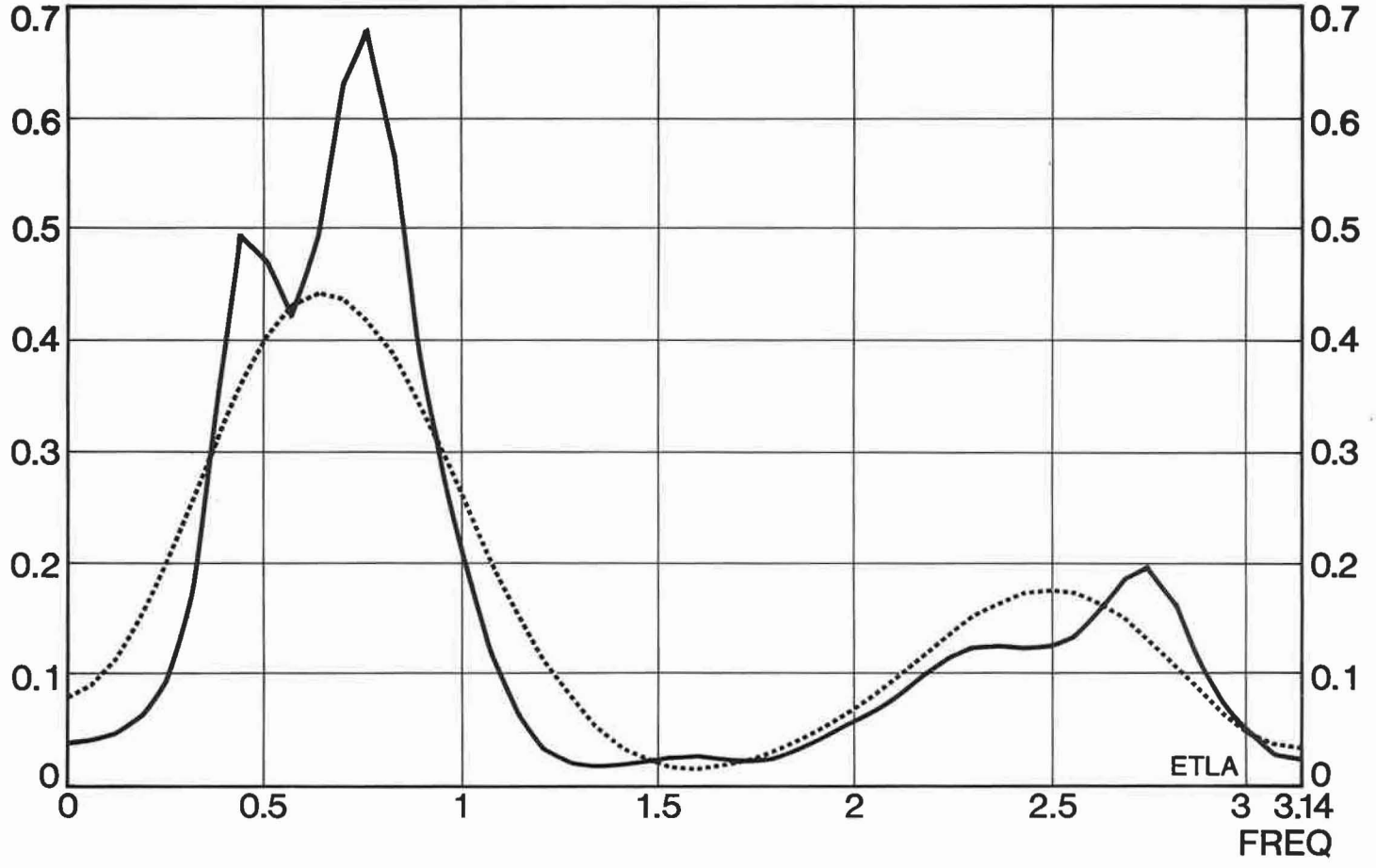


FIGURE 6 Residual spectral density of D4LEXP and a spectral density of model type 2 (broken line)



Lag	Autocorrelation coefficient	Partial autocorr. coefficient
1	0.336	0.336
2	0.199	0.097
3	-0.201	-0.334
4	-0.475	-0.429
5	-0.284	0.073
6	-0.197	0.046
7	0.056	-0.041
8	0.067	-0.234
9	0.109	-0.021
10	-0.021	-0.081

The following AICME and SBICME criteria were recorded for the tentatively identified models:

Model	Model form	AICME	SBICME
1	$(0,2)\times(1,0)_4$	-78.98	-73.31
2	$(0,2)\times(0,1)_4$	-80.06	-74.38
3	$(0,1)\times(0,2)_2$	-76.60	-70.92

Figure 6 shows the estimated residual spectral density together with the most resembling ARMA$(0,2)\times(0,1)_4$ -spectrum. The ME -estimates of the MA-parameters are represented in Table 3. A weighted correlation matrix for the three variables and their lags up to 7 was computed using $\Sigma_{n-7}^{-1}(\tilde{\theta})$ as weighting matrix. The most plausible model had D4LSEM with lags 0, 4, 5 and 7 and D4LWEP with lags 0 and 4 as explanatory variables. Almon models did not seem to be of any help. The GLS -estimates for the parameters are shown in Table 3 together with the ML -estimates.

TABLE 3

Variable	Parameter	ME -estimate	(std. dev.)	ML -estimate	(std. dev.)
D4LSEM	ω_o	1.898	(0.369)	1.841	(0.354)
"	ω_4	0.954	(0.453)	0.924	(0.436)
"	ω_5	-1.301	(0.497)	-1.275	(0.499)
"	ω_7	1.078	(0.412)	1.059	(0.371)
D4LWEP	ω_o	-1.285	(0.233)	-1.282	(0.199)
"	ω_4	0.447	(0.234)	0.454	(0.210)
	constant	0.024	(0.010)	0.025	(0.009)
	θ_1	-0.277	(0.278)	-0.213	(0.166)
	θ_2	-0.335	(0.246)	-0.327	(0.152)
	Θ_1	0.475	(0.237)	0.460	(0.170)
	σ	0.049		0.045	

5. COMPUTATIONAL ASPECTS OF THE ME- AND ML-ESTIMATION METHODS

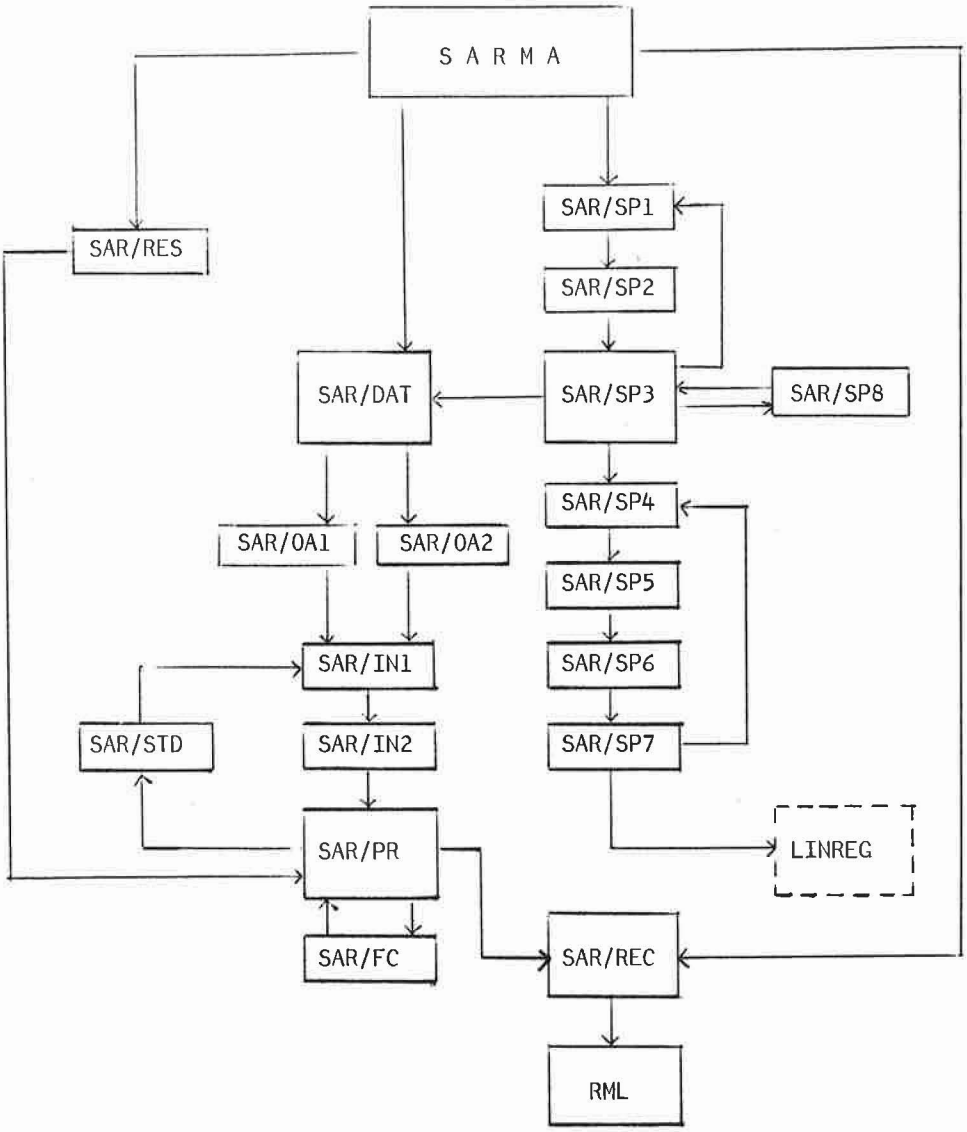
5.1 Description of the Computer Program

The ME-, ML- and ME-GLS -methods have been implemented in a computer program called SARMA. It has been integrated into the SURVO-76 system (cf. Mustonen, 1977) on Wang 2200 computers. SURVO-76 is a statistical data processing system with a data storage arrangement that is common to all of the more than 50 programs integrated into the system. The unified data storage system allows the user to link different programs together flexibly by saving intermediate results in a disk file and using these results as input data for the next program in the runstream. Larger programs may consist of several modules because of the limitations of the core memory size. The execution of different operations can be started with the so-called F-keys on the keyboard. The roles of these keys are redefined by each program module. A basic SURVO module surveying data transitions and program execution always occupies a part of the core memory. The range of available programs covers most fields of applied statistics.

The SARMA program consists of 20 modules that can be used in sequences according to Chart 1. There are three main lines of operations:

1. The first one computes exact maximum likelihood estimates for models (2.1) with restriction (2.2) and at most two input variables. Module SAR/DAT loads the data into the core memory and allows the

CHART 1


S A R M A
SAR/RES
SAR/SP1
SAR/SP2
SAR/DAT
SAR/SP3
SAR/SP8
SAR/SP4
SAR/OA1
SAR/OA2
SAR/SP5
SAR/IN1
SAR/SP6
SAR/STD
SAR/IN2
SAR/SP7
LINREG
SAR/PR
SAR/FC
SAR/REC
RML

user to specify the model. The optimization of the likelihood is carried out either by SAR/OA1 or by SAR/OA2 depending on the form of the model. The final module of this line is SAR/PR that will print the estimates and save the results, residuals and forecasts on disk. There is also a direct link to recursive estimation.

2. The second line of operations computes ME -estimates for the parameters of ARMA-models. A spectrum file is needed as a prerequisite. It can be computed and saved on disk with the program SPECTRUM in the SURVO-76 system. At most five variables (one output and four inputs) can be inspected simultaneously. Module SAR/SP2 will compute residual spectral densities when the shares of selected inputs have been eliminated from the output spectra. The ME -estimates will be computed in SAR/SP3. If the estimates are to be used as initial values for the ML -procedure, one can use the direct link to SAR/DAT.

To facilitate the choice of the residual model, it is possible to calculate autocorrelations and partial autocorrelations connected with the estimated residual spectrum or its inverse. (For the interpretations of inverse autocorrelations, see Cleveland, 1972, and Bhansali, 1983.) The final choice of the model should be made on visual grounds possibly aided by AICME- or SBICME -criteria. To draw the graphs of different spectral densities, SAR/SP3 makes use of the same submodules as the PLOT program in SURVO-76.

When a suitable model has been found and the estimation of parameters has been carried out, weighted correlation matrices of selected lags of the input and output variables can be computed. It is also possible to include variables incorporating polynomially distributed lags. The GLS -stage of the transfer function identification

can be carried out with the programs LINREG or STEPREG.

3. The third line of operations is for recursive estimation of the parameters of models (2.1) with polynomial transfer functions $v_i(L)$, $i = 1,\ldots,m$. Good estimates (preferably ML -estimates) for some short initial period and the corresponding residuals are required. The recursive residuals and the series of estimates of selected parameters will be saved on disk for later inspection.

   This line will not be discussed here any further. A paper representing the details of the estimation procedure and the potential use of the recursive residuals will be published later.

Directions for the Use of the Program:

The following symbols and phrases are used in the specification of the form of the model:

P = degree of the non-seasonal AR-part of the residual

PS = degree of the seasonal AR-part of the residual

Q and QS will denote the degrees of the non-seasonal and seasonal MA-parts, respectively.

In the ML -estimation routine transfer functions are supposed to be of the form (2.2). Symbols OMEGA(J) and DELTA(K) will refer to the notation in (2.2).

At most two inputs are allowed for.

The inputs do not have to be stationary, but if nonstationary series are used, the initial estimates ought to be good, because numerical difficulties may arise if the transfer functions are poorly specified.

Restrictions (partition size 30.5 K):

The length of the time series must not exceed 220 observations. The maximal degree of all lag polynomials is 3. This restriction does not apply to the numerators of transfer functions, where the maximal degrees are not limited. Only 5 parameters can be attached to each numerator.

The maximal number of parameters altogether is 10.

No lags in the ARMA-structures are allowed to exceed 15.

F -functions:

[F1]: INITIATING THE PROGRAM (NO ESTIMATES PREVIOUSLY COMPUTED)

[F2]: RECALLING ESTIMATES FROM A DISK FILE

[F3]: COMPUTING INITIAL ESTIMATES, BASED ON THE ESTIMATED SPECTRUM

[F4]: RECURSIVE ESTIMATION

[F1]: Initiates the ML -estimation. It is possible to keep watch on the iteration process on the screen. The iteration can be stopped manually by pressing the full-stop key. The terminal may be released to other purposes by pressing the R -key.

NOTE: During the iteration, a set of transformed parameters is used (cf. Osborne, 1976) to make sure that all polynomials in the

model will meet the stability requirements. This is why one should not try to interpret the parameter values displayed on the screen during the iterations.

When an optimum has been found or the iteration has been stopped, the following F -functions will be available:

FØ: PRINTING THE ESTIMATES (WITHOUT STANDARD DEVIATIONS)
F1: COMPUTING THE STANDARD DEVIATIONS
F2: RESTARTING THE ITERATION (AFRESH)
F3: CONTINUING THE ITERATION

If either FØ or F1 is pressed, module SAR/PR will be called to the core memory. The new F -definitions will be:

F1: PRINTING THE ESTIMATES OF THE PARAMETERS
F2: SAVING THE RESULTS ON A DISK FILE
F3: SAVING THE RESIDUALS ON A SURVO FILE
F4: CALCULATING THE COVARIANCES OF THE ESTIMATES
F5: FORECASTING
F6: CONTINUATION TO RECURSIVE ESTIMATION (RESIDUALS MUST BE SAVED BY F3)

F2 is used if one wants to recall the estimated model later for example to generate forecasts. The results can be recalled by using [F2].
F3 can be used for model diagnostics etc. . F5 is pressed when forecasts are required. If the forecasting model includes a transfer

function part, one has to make sure that all necessary values of the input variables are available in the data file.

If the data has been differenced before modelling, forecasts for the original variables can be calculated with module LAG.

F2 : Recalls the results of some earlier ML -estimation from disk and goes straight to SAR/PR.

F3 : Starts the ME -estimation procedure. If there are several variables in the spectral disk file (created with SPECTRUM), one of them will be chosen as an output variable, and the influences of the other variables can be removed from the output spectrum.

The F -functions will be

F1: STARTING THE ITERATION

F2: PLOTTING THE FITTED ARMA-SPECTRUM

F3: DIRECT CONTINUATION TO ML -ESTIMATION

F4: PRINTING THE ESTIMATES

F5: PLOTTING THE SPECTRAL DENSITY TO BE FITTED

F6: COMPUTING WEIGHTED CORRELATIONS BETWEEN SELECTED LAGS OF SELECTED SERIES (THE WEIGHTS WILL CORRESPOND TO THE ARMA-STRUCTURE OF THE RESIDUAL)

F7: RESTARTING THE INITIAL ESTIMATION ROUTINE

F8: ESTIMATING THE ACF OF THE RESIDUAL

F9: ESTIMATING THE INVERSE ACF OF THE RESIDUAL

When F1 is pressed and the form of the ARMA-model has been

given, the following text will appear on the screen: PRESS 'RETURN' TO START THE ITERATION... .

If you do not want to start the ME -estimation, press FØ.

After the ME -estimation, the estimates can be printed by F4. It usually takes the computer some time to calculate the approximate standard deviations of the estimates. This is why the printing proceeds rather slowly.

It is advisable to plot the fitted spectrum (F2) in the same graph with the residual spectrum (F5) in order to judge the goodness of fit visually.

By pressing F6, one can compute weighted covariances between chosen variables. The user can choose the lags of potential relevance for each variable by entering the desired lags as a sequence (separated by commas) after each variable name, for instance

DLOGP: 0,1,4,8,12

If Almon models are desired, the previous list can be augmented for example with symbols A3/16 which will mean an Almon polynomial of degree 3 with a maximal lag length 16. All the four Almon variables will then be generated automatically.

F4: The use of the recursive estimation module will not be discussed here, because recursive estimation lies outside the scope of this paper.

5.2 Inverting Covariance Matrices of Observations from ARMA-Processes

In Chapter 2, it was seen that, in order to compute the values attained by the likelihood (2.4), we need a method for inverting efficiently matrices

$$\mathrm{cov}([z(t_o+1) \ \ldots \ z(n)]') \ ,$$

where

$$z(t) = \frac{\delta^*(L)\theta(L)}{\phi(L)}\alpha(t) \qquad t = t_o+1,\ldots,n \ ,$$

and $\alpha(t) \sim \mathrm{NID}(0,\sigma^2)$.

To simplify the notation, let us consider a general stationary ARMA(p,q)-process

$$\phi(L)\varepsilon(t) = \theta(L)\alpha(t) \ , \qquad \alpha(t) \sim \mathrm{NID}(0,\sigma^2) \quad t = \ldots,-1,0,1,\ldots \ .$$

Let

$$\gamma(t) = \frac{1}{\phi(L)}\alpha(t) \qquad t = \ldots,-1,0,1,\ldots$$

$$\Sigma_n = \mathrm{cov}([\varepsilon(1) \ \ldots \ \varepsilon(n)]') = \mathrm{cov}([\varepsilon(n) \ \ldots \ \varepsilon(1)]') \ ,$$

$$s = \max(p,q) \ ,$$

$$\Phi_n = \begin{bmatrix} 1 & -\phi_1 & \cdots & -\phi_s & 0 & \cdots & 0 \\ 0 & 1 & & & & & 0 \\ \vdots & & \ddots & & & & -\phi_s \\ & & & \ddots & & & \vdots \\ & & & & \ddots & & -\phi_1 \\ 0 & \cdots & & & & & 1 \end{bmatrix} \qquad (n \times n)$$

$$\phi = \begin{bmatrix} \phi_s & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ \phi_1 & & \cdots & \phi_s \end{bmatrix} \qquad (s \times s)$$

$$D_\phi = \Phi_s^{-1}\phi,$$

and

$$\Omega_\phi = \frac{1}{\sigma^2} \operatorname{cov}([\gamma(s) \ \ldots \ \gamma(1)]') \ .$$

Matrices Θ_n, θ, D_θ and Ω_θ are defined in a similar fashion. Finally, let

$$\Theta^* = \Theta_n^{-1} \begin{bmatrix} 0 \\ I \end{bmatrix} \qquad (n \times s)$$

and

$$\Gamma = \frac{1}{\sigma^2} \operatorname{cov}\left(\phi \begin{bmatrix} \varepsilon(0) \\ \vdots \\ \varepsilon(-s+1) \end{bmatrix} - \theta \begin{bmatrix} \alpha(0) \\ \vdots \\ \alpha(-s+1) \end{bmatrix} \right) .$$

Now, obviously

$$\Phi_s'\Omega_\phi\Phi_s = I + \phi'\Omega_\phi\phi \tag{5.1}$$

$$= I + D_\phi'D_\phi + \ldots + D_\phi'^{k}D_\phi^{k} + D_\phi'^{k}\phi'\Omega_\phi\phi D_\phi^{k}$$

for any $k \geq 1$.

According to Ali (1977),

$$\Omega_\phi = (\Phi_s'\Phi_s - \phi\phi')^{-1} . \tag{5.2}$$

The above definitions and the Yule-Walker equations imply

$$\Sigma_s = \sigma^2[\Theta_s - \theta]\begin{bmatrix} \Omega_\phi & D_\phi\Omega_\phi \\ \Omega_\phi D_\phi' & \Omega_\phi \end{bmatrix}\begin{bmatrix} \Theta_s' \\ -\theta' \end{bmatrix}$$

$$= \sigma^2\{\Theta_s(D_\phi - D_\theta)\Omega_\phi(D_\phi - D_\theta)'\Theta_s' + \Theta_s\Phi_s^{-1}\Phi_s^{-1\prime}\Theta_s\} .$$

Because the equation

$$\Phi_n\Sigma_n\Phi_n' = \sigma^2(\Theta_n\Theta_n' + \begin{bmatrix} 0 & 0 \\ 0 & \Gamma \end{bmatrix})$$

holds for all $n \geq s$ and because Φ_n and Θ_n commute, we have

$$\Gamma = \Phi_s\Theta_s^{-1}(D_\phi - D_\theta)(\Phi_s'\Phi_s - \phi\phi')^{-1}(D_\phi - D_\theta)'\Theta_s^{-1\prime}\Phi_s' .$$

Further,

$$\Sigma_n = \sigma^2 \Phi_n^{-1} \Theta_n (I + \Theta^* \Gamma \Theta^{*\prime}) \Theta_n' \Phi_n^{-1\prime} \quad . \tag{5.3}$$

By applying the Sherman-Morrison-Woodbury formula to the inverse of the matrix $I + \Theta^* \Gamma \Theta^{*\prime}$ we get

$$\Sigma_n^{-1} = \frac{1}{\sigma^2} \Phi_n' \Theta_n^{-1\prime} [I - \Theta^* (I + \Gamma \Theta^{*\prime} \Theta^*)^{-1} \Gamma \Theta^{*\prime}] \Theta_n^{-1} \Phi_n \quad . \tag{5.4}$$

If $\theta(L)$ fulfils the invertibility conditions, equations (5.1) and (5.2) provide an efficient way of calculating

$$\Theta^{*\prime} \Theta^* = \Theta_s^{-1\prime} (I + \ldots + (D_\theta')^{k-1} D_\theta^{k-1}) \Theta_s^{-1} \tag{5.5}$$

whenever $n = ks$.

Because $\det \Phi_n = \det \Theta_n = 1$, (5.3) further implies

$$\det \Sigma_n = \det(I + \Gamma \Theta^{*\prime} \Theta^*) \sigma^{2n} \tag{5.6}$$

This way of calculating Σ_n^{-1} and $\det \Sigma_n$ resembles closely Dent's method (cf. Dent, 1977).

It is to be noted that the MA-part of the residual does not have to be invertible to enable one to apply (5.4). The only thing that breaks down when $\theta(L)$ is noninvertible is the shortcut in calculating $\Theta^{*\prime} \Theta^*$. It can, however, be computed in accordance with (5.5) as well. This means that it is possible to take superfluous differencing of the data into account at the estimation stage.

5.3 Some Comments about the Stationarity Assumption of an Unobserved Disturbance

When one is estimating the parameters of model (2.1) with polynomial transfer functions $v_i(L)$ $(i = 1,\ldots,m)$ and an autoregressive residual

$$\varepsilon(t) = \phi(L)^{-1}\alpha(t) \qquad \text{where } \alpha(t) \sim \text{NID}(0,\sigma^2)$$
$$t = \ldots -1,0,1,\ldots$$

it seems tempting to filtrate the $y(t)$- and $x_i(t)$ -series $(i = 1,\ldots,m)$ with the filter $\bar{\phi}(L)$, where $\bar{\phi}$ denotes some preliminary estimator of the ϕ -parameters. Then, one could use the ordinary least squares method in the estimation of the transfer function parameters and obtain new estimates for ϕ on the basis of the least squares calculations. These kinds of procedures have been suggested among others by Cochrane and Orcutt (1949) and by Hatanaka (1976). Both of these procedures are sometimes used iteratively (cf. Harvey, 1981, p. 192, Luukkonen, 1983 or Mellin, 1983), although Hatanaka introduced his method only as a two-step procedure. The iterative versions of both procedures are numerically asymptotically equivalent to the ML -estimation method, and are thus often considered "almost equivalent" to the ML -method. In small samples, however, the difference can be substantial, because only the ML -procedure has the requirement of the stationarity of the *unobservable* $\varepsilon(t)$ imbedded. The difference is perhaps best illustrated by a practical example:

Example 3: Let us consider a simple double logarithmic demand model for the annual consumption of furniture, furnishings and household equipment in Finland:

$$\log q(t) = \beta_o + \beta_1 \log p(t) + \beta_2 \log Q(t) + \varepsilon(t) ,$$

where

$q(t)$ = the consumption of furniture, furnishings and household equipment in the year t, fixed prices

$p(t)$ = the relative price index of the commodities mentioned above

$Q(t)$ = total private consumption expenditure in Finland, fixed prices

The estimation period is 1950 - 1979.

OLS -estimates of the parameters are:

Parameter	Estimate	Std. error
β_o	-2.609	0.324
β_1	-0.848	0.288
β_2	0.990	0.031

$R^2 = 0.975$ D-W = 0.443

Because of the low value of the Durbin-Watson test statistic, it seems

natural to use Cochrane-Orcutt technique in the estimation and augment the model with the assumption

$$\varepsilon(t) - \phi\varepsilon(t-1) = \alpha(t) \sim \text{NID}(0,\sigma^2), \qquad t = 1,\ldots,n \ .$$

The resulting iterated estimates are

Parameter	Estimate	Std. error
β_0	-5.067	1.181
β_1	0.228	0.174
β_2	1.217	0.109
ϕ	0.881	0.088

The nine first autocorrelations of the estimated $\alpha(t)$ -residuals are

lag	acf
1	0.17
2	-0.19
3	-0.16
4	-0.03
5	0.26
6	-0.01
7	-0.12
8	0.05
9	-0.10

resulting in a Box-Ljung statistic (cf. Ljung and Box, 1978)

$$\text{B-L} = 5.26 \ .$$

Apart from the positive price elasticity, the model seems to fit the data and everything seems to be in order. A graph of the fitted and observed values (Figure 7) reveals, however, that the model is unacceptable because the $\varepsilon(t)$ -residuals do not even look stationary.

The corresponding ML -estimates are

Parameter	Estimate	Std. error
β_0	-2.821	0.721
β_1	-0.140	0.226
β_2	1.013	0.068
ϕ	0.882	0.081

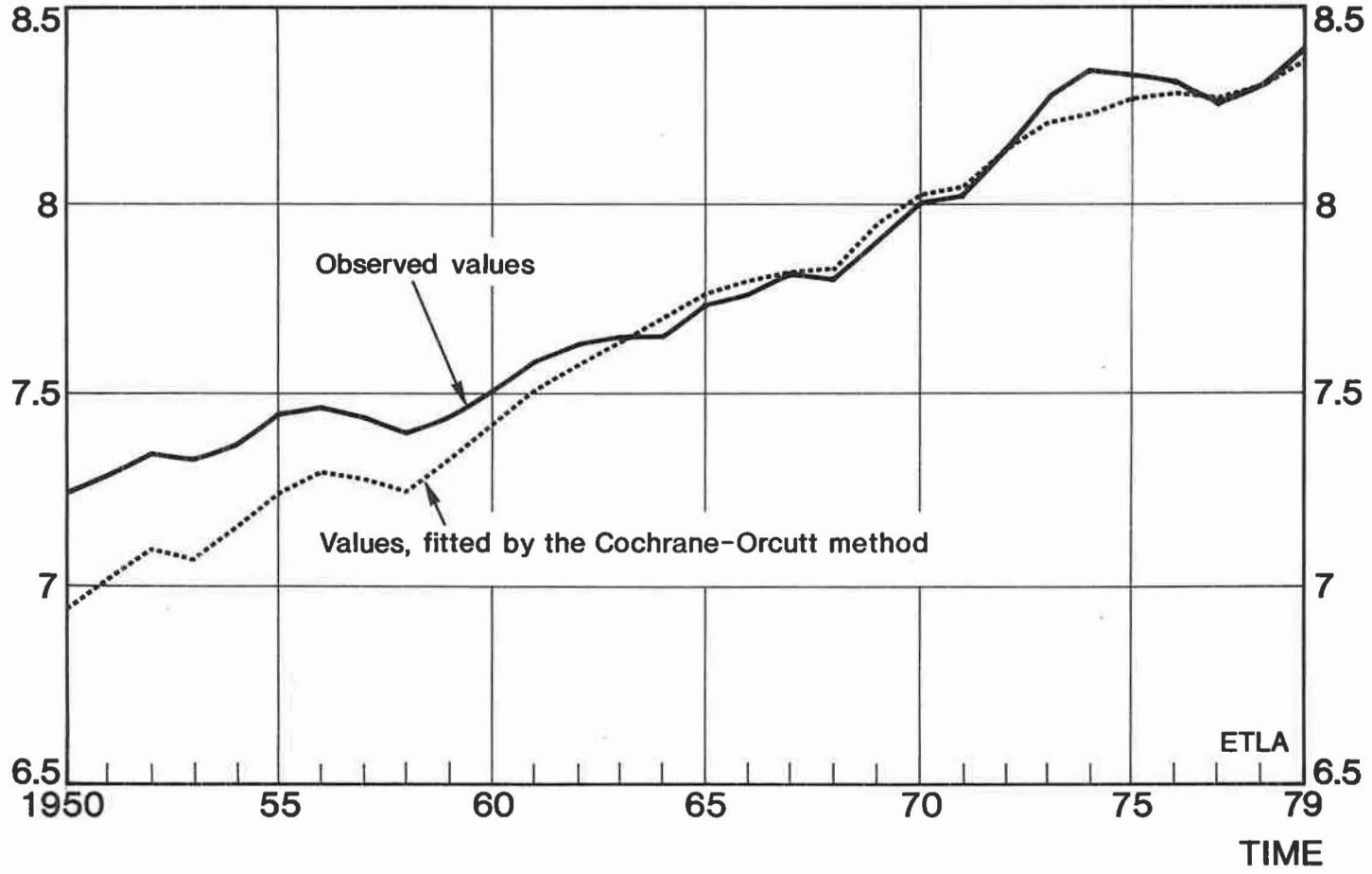
The Box-Ljung statistic, calculated from nine residual autocorrelations, is

$$B\text{-}L = 5.49 .$$

OLS seems to overestimate the importance of price changes, but the ML estimates seem sensible. The results of the iterative Cochrane-Orcutt method are not of much value.

There are other, even more striking examples where the trajectories of the observed and fitted values do not even cross each other during the sample period.

FIGURE 7 Logarithmic volume of the consumption of furniture, furnishings and household equipment in Finland



The previous example is meant to illustrate the difference between *filtration* of the observed data in order to whiten the residual of a linear model and the corresponding proper *weighting* procedure (GLS -procedure). The differences between these methods are not only due to the different uses of the information contained in the p first observations, but also to the fact that shere filtration of the data does not take the stationarity of $\varepsilon(t)$ into account. Yet, the stationarity of the disturbance is one of the most relevant assumptions in model (2.1).

In non-iterative methods, the difference between weighting and filtration is not quite as clear, but we feel that one should be prepared to pay a small computational price for the soundness of the estimation principle. That is why we have suggested in Chapter 4 that the GLS method should be used in the identification of models (2.1) instead of the filtration of $y(t)$ and $X(t)$, which has been proposed by some other authors (cf. e.g. Edlund, 1984).

6. SUMMARY

There are two stages in the selection of a transfer function model, namely, the specification of the transfer functions and the description of the regularities of the disturbances by means of a parametric model. Because the disturbance model is usually of the ARMA type, we have in this paper discussed the identification of univariate ARMA-models. The methods based on the estimated autocorrelations usually suggest several possible models whose superiority can only be evaluated by estimating the parameters of the different models, and by inspecting how well they fit the data. Thus, there is a need for a fairly quick estimation procedure and a method for visualizing the goodness of fit of the estimated models. The so-called ME -method suggested in this paper satisfies these requirements. Because of the asymptotic efficiency of the method, it is also possible to build an asymptotic test theory on it.

Simultaneously, the ME -estimation principle makes it possible to start the identification of a transfer function model with the choice of the disturbance model. In this way, one can weight the observations according to the disturbance autocorrelation already at the first stage of transfer function specification. Proper weighting usually reduces the number of iteration stages in the model selection, because the results of different specification tests are fairly reliable already at the first stage. Furthermore, different models can be tentatively estimated with very little computational effort. The parameter estimates produced by the generalized least squares method are asymptotically efficient,

and they constitute good starting values for the maximum likelihood method. Computationally, the ME-GLS -procedure is suitable even for micro computers.

There are many subjective elements involved in the method, such as the choice of the window form and width for the spectral estimator, and the choice of the frequency points. Because the ME-GLS -procedure should be primarily regarded as an identification method and because it contains subjective elements, we have not tried to investigate the small sample properties of the method by simulation.

The question naturally arises, whether the same idea could be implemented to the identification of vector ARMA-models with exogenous explanatory variables. As far as we can see, the answer is negative, because the spectral theory seems to be better suited for the analysis of one-sided relationships than of feedback relations.

REFERENCES

Akaike, H. (1977): On entropy maximization principle. *Applications in Statistics* (P. R. Krishnaiah, editor). North-Holland, Amsterdam.

Ali, M. M. (1977): Analysis of autoregressive-moving average models: Estimation and prediction. *Biometrika* 64, pp. 535-545.

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica* 33, pp. 178-196.

Anderson, T. W. (1971): *The Statistical Analysis of Time Series*. Wiley, New York.

Ansley, C. F. (1979): An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrika* 66, 59-65.

Beguin, J.-M., Gourieroux, C. and Monfort, A. (1980): Identification of a mixed autoregressive-moving average process: The corner method. *Time Series* (O. D. Anderson, editor). North-Holland, Amsterdam.

Beran, J. (1977): Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* 5, pp. 445-463.

Bhansali, R. J. (1983): The inverse partial correlation function of a time series and its applications. *J. Multivariate Anal.* 13, pp. 310-327.

Bloomfield, P. and Watson, G. S. (1975): The inefficiency of least squares. *Biometrika* 62, pp. 121-128.

Boos, D. D. (1981): Minimum distance estimators for location and goodness of fit. *J. Amer. Statist. Assoc.* 76, pp. 663-670.

Box, G. E. P. and Jenkins, G. M. (1970): *Time Series Analysis: Forecasting and Control*. Holden-Day, San Fransisco.

Box, G. E. P. and Tiao, G. C. (1975): Intervention analysis with applications to environmental and economic problems. *J. Amer. Statist. Assoc.* 70, pp. 70-79.

Brillinger, D. R. (1975): *Time Series. Data Analysis and Theory*. Holt, Rinehart and Winston, New York.

Cleveland, W. S. (1972): The inverse autocorrelations of a time series and their applications. *Technometrics* 14, pp. 277-293.

Cochrane, D. and Orcutt, G. H. (1949): Application of least squares regression to relationships containing autocorrelated error terms. *J. Amer. Statist. Assoc.* 44, pp. 32-61.

Dent, W. (1977): Computation of the exact likelihood function of an ARIMA-process. *J. Statist. Comput. Simul.* 5, pp. 193-202.

Dudewicz, E. J. and Van der Meulen, E. C. (1981): Entropy-based tests of uniformity. *J. Amer. Statist. Assoc.* 76, pp. 967-974.

Edlund, P.-O. (1984): Identification of the multi-input Box-Jenkins transfer function model. *J. Forecasting* 3, pp. 297-308.

Granger, C. W. and Newbold, P. (1974): Spurious regression in econometrics. *J. Econometrics* 2, pp. 111-120.

Granger, C. W. and Newbold, P. (1977): *Forecasting Economic Time Series*. Academic Press, New York.

Gray, H. L., Kelley, G. D. and MicIntire, D. D. (1978): A new approach to ARMA modelling. *Commun. Statist. - Simul. Comp.*, B 7(1), pp. 1-77.

Grenander, U. and Szegö, G. (1958): *Toeplitz Forms and Their Application*. University of California Press, Berkeley.

Hannan, E. J. (1963): Regression for time series. *Time Series Analysis* (M. Rosenblatt, editor). Wiley, New York.

Hannan, E. J. (1965): The estimation of relationships involving distributed lags. *Econometrica* 33, pp. 206-224.

Hannan, E. J. (1970): *Multiple Time Series*. Wiley, New York.

Hannan, E. J. (1980): Recursive estimation based on ARMA models. *Ann. Statist.* 8, pp. 762-777.

Harvey, A. C. (1981): *The Economic Analysis of Time Series*. Philip Allan, Oxford.

Harvey, A. C. and Phillips, G. D. A. (1979): Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika* 66, pp. 49-58.

Hatanaka, M. (1976): Several efficient two-step estimators for the dynamic simultaneous equations model with autoregressive disturbances. *J. Econometrics* 4, pp. 189-204.

Haugh, L. D. and Box, G. E. P. (1977): Identification of dynamic regression (distributed lag) models connecting two time series. *J. Amer. Statist. Assoc.* 72, pp. 121-130.

Hillmer, S. C. and Tiao, G. C. (1979): Likelihood function of stationary multiple autoregressive moving average models. *J. Amer. Statist. Assoc.* 74, pp. 652-660.

Hokstad, P. (1983): A method for diagnostic checking of time series models. *J. Time Ser. Anal.* 4, pp. 177-183.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. and Lee, T.-C. (1984): *The Theory and Practice of Econometrics*, 2nd edition. Wiley, New York.

Kiviet, J. F. (1980): Effects of ARMA errors on tests for regression coefficients: Comments on Vinod's article; Improved and additional results. *J. Amer. Statist. Assoc.* 75, pp. 353-358.

Kullback, S. (1959): *Information Theory and Statistics*. Wiley, New York.

Liu, L.-M. and Hanssens, D. M. (1982): Identification of multiple input transfer function models. *Commun. Statis. - Theor. Meth.* A 11(3), pp. 297-314.

Ljung, G. M. and Box, G. E. P. (1978): On a measure of lack of fit in time series models. *Biometrika* 65, pp. 297-303.

Ljung, G. M. and Box, G. E. P. (1979): The likelihood function of stationary autoregressive-moving average models. *Biometrika* 66, pp. 265-270.

Ljung, L. (1977): Analysis of recursive stochastic algorithms. *IEEE Trans. Automatic Control.* 22, pp. 551-575.

Luukkonen, R. (1983): SURVO-76. Two programs for time series analysis. *Research Report* No 40, Department of Statistics, University of Helsinki.

Mellin, I. (1983): Simple static and dynamic regression models: An application to consumption expenditure in Finland. *Research Report* No 42, Department of Statistics, University of Helsinki.

Mustonen, S. (1977): SURVO-76. A statistical data processing system. *Research Report* No 9, Department of Statistics, University of Helsinki.

Mäkeläinen, T. (1970): A specification analysis of the general linear model. *Comment. Physico-Mathematicae, Soc. Sci. Fenn.* 38, No 5.

Nakano, J. (1982). Parameter estimation of an autoregressive moving average model. *Ann. Inst. Statist. Math.* 34 A, pp. 83-90.

Nerlove, M., Grether, D. M. and Carvalho, J. L. (1979): *Analysis of Economic Time Series. A Synthesis.* Academic Press, New York.

Newbold, P. (1974): The exact likelihood function for a mixed autoregressive-moving average process. *Biometrika* 61, pp. 423-426.

Nicholls, D. F. and Hall, A. D. (1979): The exact likelihood function of multivariate autoregressive-moving average models. *Biometrika* 66, pp. 259-264.

Osborn, D. R. (1976): Maximum likelihood estimation of moving average processes. *Ann. Econ. Soc. Measur.* 5, pp. 75-87.

Parr, W. C. (1981): Minimum distance estimation: A bibliography. *Commun. Statist. - Theor. Meth.* A 10(12), pp. 1205-1224.

Parzen, E. (1971): Efficient estimation of stationary time series. Mixed schemes. *Bulletin of the ISI; Proceedings of the 38th Session,* 44:2, pp. 315-319.

Parzen, E. (1982): Maximum entropy interpretation of autoregressive spectral densities. *Statist. Prob. Letters* 1, pp. 7-11.

Praetz, P. A. (1981): A note on the effect of autocorrelation on multiple regression statistics. *Austral. J. Statist.* 23, pp. 309-313.

Priestley, M. B. (1971): Fitting relationships between time series. *Bulletin of the ISI, Proceedings of the 38th Session*, 44 (invited papers), pp. 295-324.

Priestley, M. B. (1981): *Spectral Analysis and Time Series, Vol. 1, Univariate Series*. Academic Press, London.

Pukkila, T. (1978): On the identification of transfer function noise models with several correlated input processes using frequency domain tools. *Report No A29*, Department of Mathematical Sciences, University of Tampere.

Pukkila, T. (1982): On the identification of transfer function noise models with several correlated inputs. *Scand. J. Statist.* 9, pp. 139-146.

Rice, J. (1979): On the estimation of the parameters of a power spectrum. *J. Multivariate Anal.* 9, pp. 378-392.

Saikkonen, P. (1981): Estimates for spectral averages. *Research Report* No 25, Department of Statistics, University of Helsinki.

Sathe, S. T. and Vinod, H. D. (1974): Bounds on the variance of regression coefficients due to heteroscedastic autoregressive errors. *Econometrica* 42, pp. 333-340.

Schwarz, G. (1978): Estimating the dimension of the model. *Ann. Statist.* 6, pp. 461-464.

Teräsvirta, T. and Mellin, I. (1983): Estimation of polynomial distributed lag models. *Research Report* No 41, Department of Statistics, University of Helsinki.

Tiao, G. C., Box, G. E. P., Grupe, M. R., Hudak, G. B., Bell, W. R. and Chang, I. (1979). The Wisconsin multiple time series (WMTS-1) program. A preliminary guide. Department of Statistics, University of Wisconsin, Madison.

Tiao, G. C. and Box, G. E. P. (1981): Modelling multiple time series with applications. *J. Amer. Statist. Assoc.* 76, pp. 802-816.

Tsay, R. S. and Tiao, G. C. (1982): Consistent estimates of autoregressive parameters and extended sample autocorrelation function for sta-

tionary and non-stationary ARMA modles. *Technical Report* No 683, Department of Statistics, University of Wisconsin.

Vinod, H. D. (1976): Effect of ARMA errors on the significance tests for regression coefficients. *J. Amer. Statist. Assoc.* 71, pp. 929-933.

Walker, A. M. (1962): Large sample estimation of parameters for autoregressive processes with moving-average residuals. *Biometrika* 49, pp. 117-131.

Walker, A. M. (1964): Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time series. *J. Austral. Math. Soc.* 4, pp. 363-384.

Watson, G. S. (1955): Serial correlation in regression analysis I. *Biometrika* 42, pp. 327-341.

Watson, G. S. and Hannan, E. J. (1956): Serial correlation in regression analysis II. *Biometrika* 43, pp. 436-448.

Whittle, P. (1951): *Hypothesis Testing in Time Series Analysis*. Almqvist & Wicksell, Uppsala.

Woodward, W. A. and Gray, H. L. (1981): On the relationship between the S array and the Box-Jenkins method of ARMA model identification. *J. Amer. Statist. Assoc.* 76, pp. 579-587.

Zellner, A. and Palm, F. (1974): Time Series analysis and simultaneous equation econometric models. *J. Econometrics* 2, pp. 17-54.

APPENDIX

Proof of Theorem 4.1.1

First, we need the following lemma:

Lemma Let $Z \sim N_n(0,I)$ and $\det\Gamma = 1$, where Γ is a positive definite $n\times n$ -matrix. Then

$$P\{Z'\Gamma Z \leq t\} \leq P\{Z'Z \leq t\} .$$

Proof Let

$$C_1 = \{Z \in \mathbb{R}^n \mid Z'\Gamma Z \leq t,\ Z'Z > t\}$$
$$C_2 = \{Z \in \mathbb{R}^n \mid Z'Z \leq t,\ Z'\Gamma Z > t\}$$

and

$$C_3 = \{Z \in \mathbb{R}^n \mid Z'Z \leq t,\ Z'\Gamma Z \leq t\} .$$

The Lebesgue-measures of C_1 and C_2 are equal, because $C_1 \cap C_3 = C_2 \cap C_3 = \emptyset$ and $C_1 \cup C_3$ and $C_2 \cup C_3$ have the same measure. Further,

$$p_Z(Z_1) \leq p_Z(Z_2)$$

for all $Z_1 \in C_1$ and $Z_2 \in C_2$. Thus

$$\int_{C_1} p_Z(Z)\, dZ \leq \int_{C_2} p_Z(Z)\, dZ$$

and

$$P\{Z'\Gamma Z \leq t\} = \int_{C_3 \cup C_1} p_Z(Z)\, dZ$$

$$\leq \int_{C_3 \cup C_2} p_Z(Z)\, dZ$$

$$= P\{Z'Z \leq t\}$$

□

Proof of the Theorem Let $\lambda_1 \geq \ldots \geq \lambda_n > 0$ be the eigenvalues of $\Sigma_n = \mathrm{cov}([\varepsilon(1) \ldots \varepsilon(n)]')$ and let $Z = [z_1 \ldots z_n]' \sim N_n(0,I)$. According to Mäkeläinen (1970),

$$P\{F > c\} \leq P\left\{\frac{n-m}{r} \cdot \frac{\lambda_1 z_1^2 + \ldots + \lambda_r z_r^2}{\lambda_{m+1} z_{m+1}^2 + \ldots + \lambda_n z_n^2} > c\right\}$$

and

$$P\{F > c\} \geq P\left\{\frac{n-m}{r} \cdot \frac{\lambda_{n-r+1} z_{n-r+1}^2 + \ldots + \lambda_n z_n^2}{\lambda_1 z_1^2 + \ldots + \lambda_{n-m} z_{n-m}^2} > c\right\}.$$

By combining this with the previous Lemma we get

$$P\{F > c\} \leq P\left\{\frac{n-m}{r} \cdot \frac{z_1^2 + \ldots + z_r^2}{z_{m+1}^2 + \ldots + z_n^2} > c\, \frac{\sqrt[n-m]{\lambda_{m+1} \cdots \lambda_n}}{\lambda_1}\right\}.$$

Further,

$$P\{F > c\} \geq P\left\{\frac{n-m}{r}\cdot\frac{z_{n-r+1}^2 + \ldots + z_n^2}{z_1^2 + \ldots + z_{n-m}^2} > c\,\frac{\lambda_1}{\lambda_n}\right\}.$$

It has been shown (cf. Grenander and Szegö, 1958, pp. 64-65) that

$$\lambda_1 \leq \max_{\lambda} \sigma^2 \left|\frac{\theta(e^{i\lambda})}{\phi(e^{i\lambda})}\right|^2 \leq \sigma^2 c_1 \quad \text{and}$$

$$\lambda_n \geq \min_{\lambda} \sigma^2 \left|\frac{\theta(e^{i\lambda})}{\phi(e^{i\lambda})}\right|^2 \geq \sigma^2 c_2 .$$

On the other hand,

$$\lim_{n\to\infty} \sqrt[n-m]{\lambda_{m+1}\ldots\lambda_n} = \lim_{n\to\infty} \left(\frac{\det\Sigma_n}{\lambda_1\ldots\lambda_m}\right)^{1/(n-m)} = \sigma^2$$

because $\sigma^{-2n}\det\Sigma_n$ tends to a fixed limit when $n \to \infty$ (see formula 5.6 or formula 12 in Grenander and Szegö, 1958, p. 65).

So, asymptotically,

$$P\{F > c_1 t_\delta\} \leq \delta$$

and

$$P\{F < \frac{c_2}{c_1} t_\delta\} \leq 1 - \delta$$

irrespective of $X(t)$, $t = 1,\ldots,n$.

□

LIST OF SYMBOLS AND NOTATIONS

Matrices and most vectors will be denoted by capital letters, real or complex scalars usually by lower case letters.

$\bar{z}$: Conjugate of a complex number z.

$\| A \|$: Euclidean norm of a vector (or matrix) A.

$DG(X)$: Derivative matrix of a differentiable function $G: \mathbb{R}^k \to \mathbb{R}^p$ at point $X \in \mathbb{R}^k$; $\left[DG(X)\right]_{ij} = \partial G_i(X)/\partial x_j$. To avoid confusions, a subsrcipt is sometimes used to specify the variables that the derivative operator D refers to.

p_X : Probability density function of a vector random variable X.

f_X : Spectral density matrix of a stationary random sequence $X(t)$.

$r_{x_1,x_2}(\nu)$: Sample cross correlation of sequences $x_1(1),\ldots,x_1(n)$ and $x_2(1),\ldots,x_2(n)$ defined by

$$r_{x_1,x_2}(\nu) = \frac{\sum_{t=1}^{n-|\nu|} (x_1(t) - \bar{x}_1)(x_2(t+\nu)-\bar{x}_2)}{\left[\sum_{t=1}^{n} (x_1(t)-\bar{x}_1)^2 \cdot \sum_{t=1}^{n} (x_2(t)-\bar{x}_2)^2\right]^{1/2}}$$

$r_x(\nu)$: Sample autocorrelation of sequence $x(1),\ldots,x(n)$; $r_x(\nu) = r_{x,x}(\nu)$.

$o(x)$: Generic expression of a term of smaller order than x; $\lim_{x\to 0} \frac{1}{x} o(x) = 0$.

$O(x)$: A term of the same order as x; $\lim_{x\to 0} \frac{1}{x} O(x) = \text{const.} > 0$.

$X(t) \sim \text{NID}_p(\nu,\Sigma)$: p-dimensional random variables $X(t)$ are independent of each other and normally distributed with expectation μ and covariance matrix Σ; $t = 1,2,\ldots$.

$\overset{L}{x = y}$: Random variables x and y have a common distribution.

$x(t) \overset{L}{\to} y$: Sequence $x(t)$ $(t = 1,2,\ldots)$ of random variables tends to a limiting random variable that has the same distribution as y.

LS : Least squares method.

OLS : Ordinary least squares method.

GLS : Generalized least squares method.