THE FITTING OF TIME-SERIES MODELS

by

James Durbin

University of North Carolina

and

University of London

This research was supported by the Office of Naval Research under Contract No. Nonr-855(09) for research in probability and statistics at Chapel Hill. Reproduction in whole or in part for any purpose of the United States Government is permitted.

Institute of Statistics
Mimeograph Series No. 244
December, 1959

THE FITTING OF TIME-SERIES MODELS [1,2]

by

J. Durbin
University of North Carolina
and London School of Economics

1. Introduction

The purpose of this paper is to review methods of efficient estimation of the parameters in some of the models commonly employed in time-series analysis. The models we shall consider are the following:

The autoregressive model

$$(1)\quad u_t + \alpha_1 u_{t-1} + \dots + \alpha_k u_{t-k} = \varepsilon_t \qquad (t = 1, \dots, n);$$

Regression on fixed x's and lagged y's

$$(2)\quad y_t + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} = \beta_1 x_{1t} + \dots + \beta_q x_{qt} + \varepsilon_t$$
$$(t = 1, \dots, n);$$

Regression on fixed x's with autoregressive disturbances

$$(3)\quad y_t = \beta_1 x_{1t} + \dots + \beta_q x_{qt} + u_t \quad \text{where}$$
$$u_t + \alpha_1 u_{t-1} + \dots + \alpha_p u_{t-p} = \varepsilon_t \qquad (t = 1, \dots, n);$$

The moving-average model

$$(4)\quad u_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_h \varepsilon_{t-h} \qquad (t = 1, \dots, n);$$

[1]This research was supported by the Office of Naval Research under Contract No. Nonr-855(09) for research in probability and statistics at Chapel Hill. Reproduction in whole or in part for any purpose of the United States Government is permitted.

[2]Invited review paper presented at a meeting of the Econometric Society at Washington, D. C. on December 30, 1959.

The autoregressive model with moving-average errors

$$(5)\quad u_t + \gamma_1 u_{t-1} + \cdots + \gamma_p u_{t-p} = \varepsilon_t + \delta_1 \varepsilon_{t-1} + \cdots + \delta_q \varepsilon_{t-q}$$

$$(t = 1, \ldots, n).$$

In all cases $\{\varepsilon_t\}$ is assumed to be a sequence of independently and identically distributed random variables with mean zero, variance σ^2 and with finite moments of all orders. For discussions of efficiency, but only then, the ε_t's will be assumed to be normally distributed. The sequences $u_o, u_{-1}, u_{-2}, \ldots$ and $y_o, y_{-1}, y_{-2}, \ldots$ are regarded as sequences of fixed constants. For model (1) we assume that all the roots of the equation $x^k + \alpha_1 x^{k-1} + \cdots + \alpha_k = o$ have modulus less than one; a similar assumption holds for the equation $x^p + \alpha_1 x^{p-1} + \cdots + \alpha_p = o$ in model (2) and model (3), the equation $x^h + \beta_1 x^{h-1} + \cdots + \beta_h = o$ in model (4) and the equations $x^p + \gamma_1 x^{p-1} + \cdots + \gamma_p = o$ and $x^q + \beta_1 x^{q-1} + \cdots + \beta_q = o$ in model (5). These assumptions are more restrictive than is necessary for certain parts of the exposition; they have been made at the outset in this rather sweeping fashion for the sake of simplicity of presentation later on. Throughout the paper an estimator will be said to be efficient when it is asymptotically efficient.

2. The autoregressive model

There is an underlying unity in the methods of estimation to be discussed in this paper arising from the fact that they all depend fundamentally on the method of least squares. For the autoregressive model

$$u_t + \alpha_1 u_{t-1} + \cdots + \alpha_k u_{t-k} = \varepsilon_t$$

the minimisation of $\sum_{t=1}^{n} (u_t + a_1 u_{t-1} + \cdots + a_k u_{t-k})^2$ leads directly to estimates $a_1, \ldots, a_k$ which are efficient and easy to compute. The sampling properties of $a_1, \ldots, a_k$ were first investigated by Mann and Wald [8] who showed that they are the same asymptotically as those of least-squares estimates of regression coefficients in multivariate normal systems. (Actually Mann and Wald's model contained a constant α_o and therefore differs slightly from (1) but this does not substantially affect the conclusions).

It is sometimes preferable to work with the asymptotically equivalent values $a_1', \ldots, a_k'$ obtained from the equations

$$
(6) \quad
\begin{aligned}
r_1 + a_1' + r_1 a_2' + \cdots + r_{k-1} a_k' &= o \\
r_2 + r_1 a_1' + a_2' + \cdots + r_{k-2} a_k' &= o \\
&\vdots \\
r_k + r_{k-1} a_1' + \cdots \qquad + \quad a_k' &= o
\end{aligned}
$$

where r_i is the $i^{\underline{th}}$ sample serial correlation coefficient. If desired the r_i's can be replaced by estimates of the serial covariances. For all except small values of k the equations (6) are easier to solve than the least-squares equations owing to their possession of a more symmetric structure. It is found that a pivotal reduction of (6) reduces to the recurrence relations

$$
(7) \quad a_{ss} = - \frac{r_s + a_{s-1.1} r_{s-1} + a_{s-1.2} r_{s-2} + \cdots + a_{s-1.s-1}}{1 + a_{s-1.1} r_1 \quad + \cdots \qquad + a_{s-1.s-1} r_{s-1}}
$$

$(s = 1, \ldots, k)$

(8) $a_{sr} = a_{s-1.r} + a_{ss}a_{s-1.s-r}$ $(r = 1,\ldots,s-1)$,

using $a_{11} = -r_1$ as the starting value. The quantities $a_{s1},\ldots,a_{ss}$ are the coefficients of the best-fitting autoregressive model of order s, while $-a_{22},\ldots,-a_{kk}$ are estimates of the partial correlation coefficients between observations $2,\ldots,k$ time periods apart with intermediate observations held fixed. Apart from yielding this information of incidental interest, the use of (7) and (8) is decidedly more expeditious than a direct solution of (6). The final coefficients $a_{k1},\ldots,a_{kk}$ are identically equal to the values $a_1',\ldots a_k'$, obtained from (6).

3. <u>Regression on fixed x's and lagged y's</u>

The extension of the autoregressive model to include fixed x's appears to have been first considered about 1945 by Cowles Commission writers in connection with the study of simultaneous regression systems (see Koopmans <u>et al</u> [7]). In this paper we are concerned only with the single-equation model

$$y_t + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} = \beta_1 x_{1t} + \cdots + \beta_q x_{qt} + \varepsilon_t \quad (t = 1,\ldots,n)$$

where $\{x_{1t}\},\ldots,\{x_{qt}\}$ are sequences of constants. Putting $y_{t-i} = x_{q+i.t}$ and $\alpha_i = -\beta_{q+i}$ $(i = 1,\ldots,p)$ the model can be written in the form

$$y_t = \beta_1 x_{1t} + \cdots + \beta_{q+p} x_{q+p.t} + \varepsilon_t \qquad (t = 1,\ldots,n),$$

or in an obvious matrix notation,

$$y = X\beta + \varepsilon,$$

where X is a $n x(p+q)$ matrix.

The least-squares estimators of $\beta_1, \ldots, \beta_{q+p}$ are the elements of the vector $b = (X'X)^{-1}X'y$. If X were a matrix of constants, least-squares theory would tell us that the vector of discrepancies $b - \beta$ has vector mean zero and variance matrix $\sigma^2(X'X)^{-1}$. For the present model these results do not hold since some of the elements of X are random variables. However we can obtain analogous results by introducing the matrix $t = \left[E(X'X)\right]^{-1} X'X$, where $E(X'X)$ denotes the matrix whose elements are the expected values of the corresponding elements of $X'X$. The matrix t will usually converge stochastically to the unit matrix as $n \rightarrow \infty$.

It is shown in [4] that $t(b-\beta)$ has vector mean zero and variance matrix $\sigma^2 E(X'X)$. It is also shown that when the ε's are normal this variance matrix is minimal in a certain sense. Letting

$$(9) \quad s^2 = \frac{1}{n-k} \sum_{t=1}^{n} (y_t - b_1 x_{1t} \ldots - b_{q+p,t}\, x_{q+p,t})^2$$

it is shown further that under certain assumptions $E(s^2) = \sigma^2 + O(\frac{1}{n^2})$ and that $b_1, \ldots, b_k$ are asymptotically multinormal. The implication of these results is that least-squares theory applies asymptotically to the model (2).

4. Regression on fixed x's with autoregressive errors

Of greater relevance in many investigations is the model

$$(10) \quad y_t = \beta_1 x_{1t} + \ldots + \beta_q x_{qt} + u_t \qquad (t = 1, \ldots, n),$$

where the u_t's are autocorrelated. It is well known that a simple least-squares analysis of data from such a model can be seriously misleading owing to the inefficiency of the least-squares estimators of the β's and to the biasedness of their estimates of variance (see for example the discussion by Cochrane and Orcutt [3], Watson [10] and Anderson [1]). It is true that for certain special cases, including regressions on polynomial trends and seasonal constants, the least-squares coefficients have been shown to be asymptotically efficient (Grenander and Rosenblatt [6] and Anderson and Anderson [2]). Nevertheless the least-squares estimator of variance remains biased and the use of analysis-of-variance methods for testing hypotheses and setting confidence limits can be expected to give incorrect results.

It is often reasonable to assume that the u_t's have the autoregressive structure

$$(11) \quad u_t + \alpha_1 u_{t-1} + \cdots + \alpha_p u_{t-p} = \varepsilon_t .$$

(10) may then be transformed to the form

$$(12) \quad y_t + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} = \beta_1 x_{1t} + \alpha_1 \beta_1 x_{1.t-1} + \cdots + \alpha_p \beta_1 x_{1.t-p} + \beta_2 x_{2t} + \alpha_1 \beta_2 x_{2.t-1} + \cdots + \alpha_p \beta_q x_{q.t-p} + \varepsilon_t .$$

This now has the same structure as (2) except that relations exist between the coefficients. Conceptually the simplest approach to the estimation problem would be to minimise $\sum_{t=1}^{n} \varepsilon_t^2$, when this is expressed in terms of the y's and x's, with respect to the α's and β's. However, since some of the coefficients in (12) are quadratic in the unknowns this procedure results in non-linear estimating equations which are usually unmanageable for practical use.

Before going on to discuss general methods we draw attention to an important special case in which simple methods based on least-squares theory do give satisfactory results. This arises when each $x_{i,t-j}$ $(i=1,\ldots,q;\ j=1,\ldots,p)$ can be expressed as a linear function of $x_{1t},\ldots,x_{qt}$. Examples are polynomial trends, seasonal constants and periodic regressions. (12) then reduces to the form (2) with functionally independent coefficients which can be legitimately estimated by least squares. We illustrate the procedure by considering the case of regression on a pure periodic function with first-order autoregressive disturbances, i.e.

$$y_t = \beta_1 \cos\lambda t + \beta_2 \sin\lambda t + u_t, \text{ where } u_t + \alpha u_{t-1} = \varepsilon_t \text{ and}$$

where λ is known.

We have

$$\begin{aligned} y_t+\alpha y_{t-1} &= \beta_1 \cos\lambda t+\alpha\beta_1 \cos\lambda(t-1) + \beta_2 \sin \lambda t+\alpha\beta_2 \sin\lambda(t-1)+\varepsilon_t \\ &= \gamma_1 \cos\lambda t + \gamma_2 \sin\lambda t + \varepsilon_t, \end{aligned}$$

where $\gamma_1 = \beta_1 + \alpha\beta_1 \cos\lambda - \alpha\beta_2 \sin\lambda,$

$$\gamma_2 = \beta_2 + \alpha\beta_1 \sin\lambda + \alpha\beta_2 \cos\lambda.$$

Efficient estimates a, c_1 and c_2 are then obtained by minimising $\sum_{t=1}^{n} (y_t + ay_{t-1} - c_1 \cos\lambda t - c_2 \sin\lambda t)^2$, whence estimates b_1, b_2 of β_1, β_2 result from the equations

$$(1 + a \cos \lambda)\, b_1 - a \sin \lambda b_2 = c_1$$

$$a \sin \lambda\ \ b_1 + (1+a\cos\lambda)b_2 = c_2\ .$$

σ^2 is estimated by $s^2 = \frac{1}{n-3} \sum_{t=1}^{n} (y_t + ay_{t-1} - c_1\cos\lambda t - c_2\sin\lambda t)^2$.

For simplicity of exposition our discussion of the general problem will be based mainly on the two-coefficient model

(13) $y_t = \beta x_t + u_t$, where

(14) $u_t + \alpha u_{t-1} = \varepsilon_t$ $(t=1, . . ., n)$.

It was pointed out by Cochrane and Orcutt [3] that if α were known we could employ an autoregressive transformation $y_t' = y_t + \alpha y_{t-1}$, $x_t' = x_t + \alpha x_{t-1}$ to put the model in the form

$$y_t' = \beta x_t' + \varepsilon_t,$$

to which least squares can be applied quite validly. For the case of unknown α they suggested that one should insert in the autoregressive transformation either a value of α guessed on a priori grounds or a value estimated from the residuals of a fitted least-squares regression, further iterations being carried out if desired. The first of these suggestions is computationally attractive, though inefficient, while the second, though efficient, is computationally burdensome.

An approach will now be outlined which leads to estimates which are efficient and which are not too onerous to compute. From (13) and (14) we have

(15) $y_t + \alpha y_{t-1} = \beta x_t + \gamma x_{t-1} + \varepsilon_t$,

where $\gamma = \alpha\beta$. If we were to ignore the restriction $\gamma = \alpha\beta$ and regard γ as a free parameter, (15) would have the form (2) so that the least-squares estimators a, b, c obtained by minimising $\sum_{t=1}^{n}(y_t+ay_{t-1}-bx_t-cx_{t-1})^2$

would be efficient estimators of α, β, γ. To obtain efficient estimators of α and β we need only therefore consider the joint distribution of a, b, c.

Now $y_t + ay_{t-1} - bx_t - cx_{t-1} = y_t + au_{t-1} - bx_t - (c - a\beta)\, x_{t-1}$ since $y_{t-1} = \beta x_{t-1} + u_{t-1}$. Consequently a, b and $c - a\beta$ are the least-squares coefficients of regression of y_t on $-u_{t-1}$, x_t and x_{t-1}. The corresponding true coefficients are α, β and zero in virtue of the relation

$$(16) \qquad y_t + \alpha u_{t-1} = \beta x_t + \epsilon_t .$$

By a slight extension of the results of the previous section we know that least-squares regression theory applies asymptotically to (16). Consequently the quantities $a - \alpha$, $b - \beta$ and $c - a\beta$ are asymptotically normally distributed with zero means and variance matrix $\sigma^2 A^{-1}$, where A is the expected value of the matrix

$$\begin{bmatrix} \Sigma\, u_{t-1}^2 & \Sigma u_{t-1}x_t & \Sigma\, u_{t-1}x_{t-1} \\ \Sigma\, u_{t-1}x_t & \Sigma\, x_t^2 & \Sigma\, x_t x_{t-1} \\ \Sigma\, u_{t-1}x_{t-1} & \Sigma\, x_t x_{t-1} & \Sigma\, x_{t-1}^2 \end{bmatrix} .$$

Since $E(\Sigma\, u_{t-1}x_t) = E(\Sigma\, u_{t-1}x_{t-1}) = 0$ it follows that the asymptotic distribution of a, b and $c - a\beta$ has a density which is the limit as $n \to \infty$ of the expression

$$(17)\quad \text{constant} \times \exp\left[-\frac{1}{2\sigma^2}\left\{(a-\alpha)^2\, E(\Sigma u_{t-1}^2) + (b-\beta)^2\, \Sigma x_t^2 + 2(b-\beta)(c-a\beta)\, \Sigma x_t x_{t-1} + (c-a\beta)^2\, \Sigma x_{t-1}^2\right\}\right].$$

On maximising the exponent of (17) with respect to α and β we find thgt their efficient estimates are

$$(18)\qquad \hat{\alpha} = a, \qquad \hat{\beta} = \frac{\Sigma(x_t + ax_{t-1})(y_t + ay_{t-1})}{\Sigma(x_t + ax_{t-1})^2}.$$

It is remarkable that $\hat{\beta}$ is precisely the same estimator as is obtained by using a as an estimator of α in an autoregressive transformation. The same procedure was arrived at earlier by the author [4] using a rather different approach.

Ignoring the difference between $\frac{1}{n}\Sigma x_t^2$ and $\frac{1}{n}\Sigma x_{t-1}^2$, as is legitimate to the order of accuracy considered here, we find that (18) reduces to

$$(19)\qquad \hat{\beta} = \frac{(1 + ar)b + (a + r)c}{1 + 2ar + a^2},$$

where $r = \Sigma x_t x_{t-1} / \Sigma x_t^2$. Note that a and $\hat{\beta}$ are asymptotically independently distributed.

The treatment of the general model (3) follows along similar lines. We shall confine ourselves here to the presentation of a brief summary of the computing routine, referring the reader to [4] for further theoretical discussion. For simplicity of exposition let us suppose that the variables $x_{1t}, x_{1,t-1}, \ldots, x_{qt}, \ldots, x_{q,t-p}$ are linearly independent. (The outstanding case to the contrary is the common one in which the model contains a constant term, i.e. x_{it} equals unity for some i and all t; this, however, is easily dealt with by working throughout with deviations from sample means as in ordinary regression analysis. Modifications for other cases are easily worked out ad hoc).

Suppose that the normal equations for the least-squares fitting of the regression of y_t on x_{1t}, $x_{1.t-1}, \ldots, x_{1.t-p}, \ldots, x_{qt}, \ldots,$ $x_{q.t-p}$, $-y_{t-1}, \ldots, -y_{t-p}$, the variables being taken in this order, are denoted by

$$A_1 b = c_1,$$

where A_1 is the $(p + q + pq) \times (p + q + pq)$ matrix of sums of squares and products of the variables $x_{1t}, \ldots, x_{q.t-p}$, $-y_{t-1}, \ldots, -y_{t-p}$, where c_1 is the vector of sums of products of these variables with y_t, and where b is the vector of regression coefficients. If the equations are solved by a method such as the abbreviated Doolittle method note that it is only necessary to carry the back solution far enough to give the coefficients $a_1, \ldots, a_p$ of $-y_{t-1}, \ldots, -y_{t-p}$.

Form a new matrix A_2 whose first row is obtained by multiplying the first p+1 rows of A_1 by $1, a_1, \ldots, a_p$ respectively and adding, whose second row is obtained by taking the second group of p+1 rows of A_1, multiplying by $1, a_1, \ldots, a_p$ respectively and adding. Continue in this way until A_2 has q rows. Repeat the process on c_1 to give a new vector c_2 containing q elements. Repeat the process on the columns of A_2 to give a new matrix A_3 with q rows and columns. Let c_3 be the vector obtained by subtracting from c_2 the sum of a_1 times the last column of A_2, a_2 times the second last column, . . ., a_p times the $p^{\underline{th}}$ column of A_2 counting backwards from the last column. Then the solution $\hat{\beta}$ of the equation

$$(20) \qquad A_3 \hat{\beta} = c_3$$

is the vector of efficient estimators of $\beta_1, \ldots, \beta_q$. Its estimated variance matrix is $s^2 A_3^{-1}$, where

$$s^2 = \frac{1}{n-p-q}\left[\sum_{t=1}^{n} (y_t + a_1 y_{t-1} + \ldots + a_p y_{t-p})^2 - \hat{\beta}' c_3\right].$$

5. <u>The moving-average model</u>

The special problems of fitting moving-average models can be appreciated from a consideration of the first-order model

$$(21) \qquad u_t = \varepsilon_t + \beta\,\varepsilon_{t-1} \qquad (t = 1, \ldots, n).$$

A simple estimator of β can be obtained by equating the theoretical value of the first serial correlation, namely $\beta/(1+\beta^2)$, to the sample value r_1. However, the estimator was shown to be inefficient by Whittle [11,12] who proposed the use of an approximate maximum-likelihood estimator equivalent to the solution to the equation

$$\frac{\partial}{\partial\beta}\left[\frac{1}{1-\beta^2}(1 - 2\beta r_1 + 2\beta^2 r_2 - \ldots)\right] = 0$$

where r_i is the $i^{\underline{th}}$ sample serial correlation.

Although efficient, this estimator is difficult to calculate and the method does not easily extend to higher-order models. The author [5] has therefore suggested a different method in which a $k^{\underline{th}}$ - order autoregressive model is first fitted to the data, k being taken to be large. It is shown in [5] that the fitted coefficients $a_1, \ldots, a_k$ have an asymptotic distribution with the approximate density

$$\text{constant x } (1-\beta^2)^{-1/2} \exp\left[-\frac{n}{2}\left\{\sum_{i=0}^{k-1}(a_{i+1}+\beta a_i)^2+\beta^2 a_k^2\right\}\right] .$$

Neglecting the factor $(1-\beta^2)^{-1/2}$ since this is of small order in n compared with the remainder and maximising the exponent with respect to β we obtain as our estimator of β,

$$(22) \qquad b = -\frac{\sum_{i=0}^{k-1} a_i\, a_{i+1}}{\sum_{i=0}^{k} a_i^2} \qquad (a_0 = 1),$$

the efficiency of which can be made as close to unity as desired by taking k sufficiently large.

For the general model

$$u_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \ldots + \beta_h \varepsilon_{t-h} \qquad (t = 1, \ldots, n)$$

the same approach yields estimators $b_1, \ldots, b_h$ which are obtained as the solution to the linear equations

$$(23) \qquad \sum_{i=1}^{h} A_{|r-j|}\, b_j = -A_r \qquad (r = 1, \ldots, h),$$

where $A_r = \sum_{i=0}^{k-r} a_i a_{i+r}$, the a_i's being as before.

6. The autoregressive model with moving-average errors.

This model, which has the generating equation

$$(24)\quad u_t + \gamma_1 u_{t-1} + \cdot\ \cdot\ \cdot + \gamma_p u_{t-p} = \varepsilon_t + \delta_1 \varepsilon_{t-1} + \cdot\ \cdot\ \cdot + \delta_q \varepsilon_{t-q} \qquad (t=1, \ldots, n),$$

has greater theoretical importance than the attention paid to it in time-series literature would appear to indicate. Firstly, it is the general model of which the autoregressive and moving-average models are special cases. Secondly, when $q=p-1$ it is the only one of the three models whose structure is invariant under changes in the time-period between successive observations, a fact pointed out by Quenouille [9]. Thirdly, equi-spaced observations from a continuous stochastic process generated by a linear stochastic differential equation, or having a rational spectral density, conform to a discrete model (24) with $q = p-1$. Consequently a solution to the problem of efficient fitting of (24) also gives as a by-product the solution to the problems of fitting stochastic differential equation models and of estimating rational spectral densities from discrete data. Yet in spite of the theoretical importance of the model only Quenouille [9] appears to have considered the fitting problem; however Quenouille did not attempt to discuss efficient methods of estimation.

Two methods of fitting will now be described. The first is non-iterative but is not fully efficient. The second is an iterative method in which the autoregressive and moving-average parameters are estimated alternately. It is hoped to investigate the performance of both methods by means of sampling experiments and to publish the results later.

Let us begin by considering the first-order model

$$(25) \qquad u_t + \gamma u_{t-1} = \varepsilon_t + \delta\varepsilon_{t-1} \qquad (t=1,\ldots,n).$$

Let $a_1,\ldots,a_k$ denote the coefficients of a fitted autoregressive model of large order k and let e_t denote the residual $u_t + a_1u_{t-1} +\ldots+ a_ku_{t-k}$. In the first method of estimation we replace ε_{t-1} in (25) by e_{t-1} and estimate γ and δ by the values of c and of d which minimise $\Sigma\,(u_t + cu_{t-1} - de_{t-1})^2$. This leads to the equations

$$c\,\Sigma\, u_{t-1}^2 - d\,\Sigma\, u_{t-1}e_{t-1} = -\Sigma u_tu_{t-1}$$

$$c\,\Sigma\, u_{t-1}e_{t-1} - d\,\Sigma\, e_{t-1}^2 = -\,\Sigma u_te_{t-1},$$

the solution of which is asymptotically equivalent to the expressions

$$(26) \qquad c = -\frac{a_1r_2 + a_2r_3 +\ldots+ a_kr_{k+1}}{a_1r_1 + a_2r_2 +\ldots+ a_kr_k},$$

$$(27) \qquad d = c + \frac{r_1 + a_1r_2 +\ldots+ a_kr_{k+1}}{1 + a_1r_1 +\ldots+ a_kr_k}.$$

The method is readily extended to cover higher-order systems and can be used to give starting values for the second method, which we now describe.

First let us consider the estimation of δ for a given value of γ. Suppose that the true values of the first k autoregressive coefficients are $\alpha_1, \ldots, \alpha_k$, the fitted values being denoted by $a_1, \ldots, a_k$ as before. Using Mann and Wald's results [8] we know that for large k the asymptotic distribution of $a_1, \ldots, a_k$ is normal with density

$$(28) \quad \text{constant x } \exp\left[-\frac{n}{2\sigma^2} \sum_{i,j=1}^{k} (a_i - \alpha_i)(a_j - \alpha_j) E(u_{t-i} u_{t-j})\right].$$

The quadratic form in the exponent can be represented operationally as

$$(29) \qquad Q = \mathcal{E}\left(\sum_{i=1}^{k} (a_i - \alpha_i) u_{t-i}\right)^2$$

where $\mathcal{E}$ denotes the operation of taking an expectation over variation of the u's, the a's being regarded as fixed constants.

Suppose that the true value of γ were known and the following transformation from $a_1, \ldots, a_k$ and $\alpha_1, \ldots, \alpha_k$ to $\ell_1, \ldots, \ell_k$ and $\lambda_1, \ldots, \lambda_k$ were made,

$$\begin{array}{ll} a_1 = \ell_1 + \gamma & \alpha_1 = \lambda_1 + \gamma \\ a_2 = \ell_2 + \gamma\ell_1 & \alpha_2 = \lambda_2 + \gamma\lambda_1 \\ \vdots & \vdots \\ a_k = \ell_k + \gamma\ell_{k-1} & \alpha_k = \lambda_k + \gamma\lambda_{k-1} . \end{array}$$

Substituting in (29) we have

$$(30) \quad Q = \mathcal{E}\left\{(\ell_1-\lambda_1)(u_{t-1}+\gamma u_{t-2})+\ldots+(\ell_{k-1}-\lambda_{k-1})(u_{t-k+1}+\gamma u_{t-k}) + (\ell_k-\lambda_k)\, u_{t-k}\right\}^2 .$$

Now $u_t + \lambda u_{t-1} = \varepsilon_t + \delta\varepsilon_{t-1}$. Consequently, on putting $z_t = u_t + \gamma u_{t-1}$ we see that z_t satisfies the moving-average process $z_t = \varepsilon_t + \delta\varepsilon_{t-1}$. Moreover $u_{t-k} = z_{t-k} - \gamma z_{t-k-1} + \gamma^2 z_{t-k-2} + \ldots$. Consequently on putting $\ell_{k+r} = (-\gamma)^r \ell_k$ and $\lambda_{k+r} = (-\gamma)^r \lambda_k$ $(r = 1,2,\ldots)$, (30) gives

$$(31) \qquad Q = \mathcal{E}\left\{\sum_{i=1}^{\infty} (\ell_i - \lambda_i)\, z_{t-i}\right\}^2 .$$

From the fact that the true autoregressive coefficients are generated by the relation $1+\alpha_1 z+\alpha_2 z^2+\ldots = (1+\gamma z)(1+\delta z)^{-1}$, it follows that for large k, $\lambda_i = (-\delta)^i$ $(i = 1,\ldots,k)$ as accurately as desired. Since λ_k can be made arbitrarily small the error committed by taking $\lambda_{k+r} = (-\delta)^r \lambda_k$ in (30) in place of $\lambda_{k+r} = (-\gamma)^r \lambda_k (r=1,2,\ldots)$ can be made arbitrarily small. Thus to a high degree of accuracy (31) holds with $z_{t-1}, z_{t-2}, \ldots$ corresponding to a moving-average model for which $\lambda_1, \lambda_2, \ldots$ are the true autoregressive coefficients.

Comparing (31) with (29) we see that $\ell_1, \ell_2, \ldots$ behave like autoregressive coefficients fitted to data corresponding to the model

$$z_t = \varepsilon_t + \delta\varepsilon_{t-1} \qquad (t=1,\ldots,n) .$$

Consequently it follows from (22) that the efficient estimator of

δ is

$$(32) \qquad d = -\frac{\sum_{i=0}^{\infty} \rho_i \rho_{i+1}}{\sum_{i=0}^{\infty} \rho_i^2} \qquad (\rho_0 = 1) .$$

In practice it will probably suffice to terminate the summations in this expression at about $i = k$. Note that we need not take explicit account of the fact that the "constant" in (28) depends on the covariance determinant of $a_1, \ldots, a_k$, which in turn depends on the unknown parameter δ, since the determinant is of small order in n compared with the exponent of (28).

Let us now consider the converse problem, i.e. the estimation of γ given the true value of δ. Define $w_1, \ldots, w_n$ by the relation

$$u_t = w_t + \delta w_{t-1} \qquad (t = 1, \ldots, n)$$

where w_o is either defined arbitrarily or taken equal to $u_o - \delta u_{-1} + \delta^2 u_{-2} \ldots$. The w's then satisfy the autoregressive model

$$w_t + \gamma w_{t-1} = \varepsilon_t \quad .$$

Consequently γ is efficiently estimated by the expression

$$(33) \qquad c = -\frac{\Sigma' w_t w_{t+1}}{\Sigma' w_t^2} ,$$

where Σ' denotes summation over the range of possible values of t divided by the number of terms summed.

Let

$$s_r = \Sigma' u_t u_{t+r}$$

$$p_r = \Sigma' u_t w_{t+r}$$

$$q_r = \Sigma' w_t w_{t+r}$$

It is easy to verify that to a good approximation we have the relations

$$(34) \qquad p_r = s_r - \delta\, p_{r-1} \qquad (r = -k+1, -k+2, \ldots, k)$$

$$(35) \qquad q_r = p_r - \delta\, q_{r+1} \qquad (r = k-1, k-2, \ldots, o) \;.$$

Applying (34) recursively taking $p_{-k} = s_{-k}$, and then (35) recursively taking $q_k = p_k$, we obtain q_1 and q_o from which we obtain the estimate of γ as

$$(36) \qquad c = -\frac{q_1}{q_o} \;.$$

By applying (32) and (36) alternately we obtain an iterative method of estimating γ and δ. (26) or (27) can be used to provide a starting point.

The treatment of the general model

$$u_t + \gamma_1 u_{t-1} + \ldots + \gamma_p u_{t-p} = \varepsilon_t + \delta_1 \varepsilon_{t-1} + \ldots + \delta_q \varepsilon_{t-q} \quad (t=1, \ldots, n)$$

follows similar lines. The fitted autoregressive constants $a_1, \ldots, a_k$ are transformed to $\ell_1, \ldots, \ell_k$ by the relations

$$a_1 = \ell_1 + \gamma_1$$

$$a_2 = \ell_2 + \gamma_1 \ell_1 + \gamma_2$$

$$\vdots$$

$$a_k = \ell_k + \gamma_1 \ell_{k-1} + \cdots + \gamma_p \ell_{k-p}$$

and further ℓ's are obtainable from the expression

$$\ell_r + \gamma_1 \ell_{r-1} + \cdots + \gamma_p \ell_{r-p} = 0 \qquad (r = k+1, k+2, \ldots).$$

The ℓ's behave approximately like autoregressive coefficients fitted to data generated by the moving-average model

$$z_t = \varepsilon_t + \delta_1 \varepsilon_{t-1} + \cdots + \delta_q \varepsilon_{t-q} \ .$$

$\delta_1, \ldots, \delta_q$ can therefore be estimated from equations (23) taking $A_r = \sum_{i=o}^{\infty} \ell_i \ell_{i+r}$ $(\ell_o = 1)$. In practice the summation can probably be truncated at about $i = k$.

To estimate $\gamma_1, \ldots \gamma_p$ for given $\delta_1, \ldots, \delta_q$ we use in place of (34) and (35) the expressions

(37) $$p_r + \delta_1 p_{r-1} + \cdots + \delta_q p_{r-q} = s_r \qquad (r = -k+q, \ldots, k)$$

(38) $$q_r + \delta_1 q_{r+1} + \cdots + \delta_q q_{r+q} = p_r \qquad (r = k-q, \ldots, o) \ .$$

Estimates $c_1, \ldots, c_p$ are then obtained from the equations

(39) $$\begin{aligned} q_o c_1 + q_1 c_2 + \cdots + q_{p-1} c_p &= -q_1 \\ q_1 c_1 + q_o c_2 + \cdots + q_{p-2} c_p &= -q_2 \\ &\vdots \\ q_{p-1} c_1 + \cdots + q_o c_p &= -q_p \ . \end{aligned}$$

References

[1] Anderson, R. L. : "The Problem of Autocorrelation in Regression Analysis," Journal of the American Statistical Association, Vol. 49 (1954), 113 - 129.

[2] Anderson, R. L. and Anderson, T. W. : "Distribution of the Circular Serial Correlation Coefficient for Residuals from Fitted Fourier Series," Annals of Mathematical Statistics, Vol. 21 (1950), 59 - 81.

[3] Cochrane, D. and Orcutt, G. H. : "Application of Least-Squares Regression to Relationships Containing Autocorrelated Error Terms," Journal of the American Statistical Association, Vol. 44 (1949), 32 - 61.

[4] Durbin, J. : "Estimation of Parameters in Time-Series Regression Models," Journal of the Royal Statistical Society, Series B, (1960). In the press.

[5] ________, : "Efficient Estimation of Parameters in Moving-Average Models," Biometrika, Vol. 46 (1959), In the press.

[6] Grenander, U. and Rosenblatt, M.: Statistical Analysis of Stationary Time Series. New York: John Wiley and Sons, 1957.

[7] Koopmans, T. (Editor): Statistical Inference in Dynamic Economic Models. New York: John Wiley and Sons, 1950.

[8] Mann, H. B. and Wald, A. : "On the Statistical Treatment of Linear Stochastic Difference Equations," Econometrica, Vol. 11 (1943), 173 - 220.

[9] Quenouille, M. H. : "Discrete Autoregressive Schemes with Varying Time-Intervals," Metrika, Vol. 1 (1958), 21 - 27.

[10] Watson, G. S. : "Serial Correlation in Regression Analysis I," Biometrika, Vol. 42 (1955), 327 - 341.

[11] Whittle, P. : Hypothesis Testing in Time Series Analysis. Uppsala: Almgvist and Wiksells, 1951.

[12] ________, : "Estimation and Information in Stationary Time Series," Arkiv för Matematik, Vol. 2 (1953), 423 - 434.